Construction and Simulation of Correlated Multivariate Poisson Processes

by

Michael Tsz-Chun Chiu

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto

Abstract

Construction and Simulation of Correlated Multivariate Poisson Processes

Michael Tsz-Chun Chiu
Doctor of Philosophy
Graduate Department of Computer Science
University of Toronto
2026

Poisson processes are the canonical stochastic processes for modelling counts. In practice, counting phenomena are primarily multivariate in nature. A motivating example for this thesis is the modelling of operational events of different business lines within a financial institution as a multivariate Poisson process where the observed correlation matrices have been found to exhibit *negative* correlation; similar behavior has been observed in geophysics. Standard approaches to constructing multivariate Poisson processes are unable to correctly model the observed negative dependence between the components of the multivariate Poisson process. We extend an approach introduced by Kreinin consisting of two pillars: Extreme Joint Distributions (EJD) and Backward Simulation (BS).

The EJD approach is a pure probabilistic-based approach for constructing joint distributions satisfying given marginals and possessing extreme dependence between its components, known as extreme joint distributions, which are used in calibrating a (bivariate) joint distribution satisfying given marginal constraints to a desired correlation matrix. We extend the EJD approach to the general d-dimensional setting.

Backward Simulation exploits the conditional uniformity property of Poisson processes for its simulation within an interval $[0, T]$. Given the number of the terminal events of the Poisson process, n, the arrival moments are uniformly distributed and can be obtained by sampling. Backward Simulation is an attractive simulation method for multivariate Poisson processes since, given a vector of terminal events, Backward Simulation of the multivariate Poisson process consists of applying Backward Simulation to each univariate component using the corresponding terminal event. The EJD approach is crucial for constructing the appropriate multivariate joint distributions with the desired correlation matrix. We also introduce a methodology for extending a multivariate Poisson process simulated within an interval $[0, T]$ using Backward Simulation to any subsequent interval $[mT, (m+1)T]$ for any integer $m \geq 1$. We extend Backward Simulation to mixed Poisson and compound Poisson processes.

To my children, Gabriel and Penelope.

Acknowledgements

First and foremost, I must express my deepest gratitude for my supervisor, Ken Jackson, to whom I am greatly indebted for his tutelage, dedication, time, and generosity. It has been a privilege for me to have studied under not only a sharp mind but a kind soul from whom I have learnt many invaluable skills. I hope that you found the work in this thesis both meaningful and worthwhile and that we may continue to work on interesting problems in the future.

I am also grateful to my co-supervisor, Alex Kreinin, for introducing me to the problem that has become the starting point of this thesis and for teaching me *how* to think about and approach problems in probability. I am also grateful for the internship opportunity at Algorithmics where I learnt a lot and met many wonderful people: I must thank Todd Goodwin for mentoring me in C++ and teaching me compilers and Asif Lakhany for many insightful discussions on Scientific Computing.

I must also thank my committee members, Christina Christara and Sebastian Jaimungal for being accommodating, encouraging, and for their helpful comments that have improved this thesis and my presentation skills.

It has been a privilege to be a part of the Numerical Analysis group at DCS. There, I must thank Jonathan Calver, Nat Leung, and Vida Heidarpour for making the NA Lab a warm place and for the time we spent together.

Special acknowledgement and thanks goes to Uroš Kalabić, my brother from another mother, who has given me much invaluable advice through many personal and professional difficulties and for introducing me to the Serbian way of life that has now become part of my own as well. I am grateful for the time that we have spent together both personally and professionally and am indebted for the many things that he has taken the time to teach me. I am particularly grateful and fond of the writing bootcamp back in depths of 2020. What a journey it has been since.

Acknowledgment and thanks also goes out to Peter McCormick, to whom I am grateful for our many discussions on computing and faith and for teaching me much of what I know about Linux and computer networking.

To my now extended family, my heartfelt gratitude goes first and foremost to my mother for the numerous sacrifices she has made time and time again, without which I would not be where I am today. I am also grateful to: my Aunt Sandy and Uncle Simon for their unconditional love and support; my Godmother also for her love and support; Sharon Lee—the sister I never had—and Chris Leung, for always supporting me personally without a second thought and for keeping me fed; Peter Ho for his friendship throughout all these years.

Last, but certainly not least, I must thank past-Michael for persevering, even though you were slow.

Contents

1 Introduction **1**

1.0.1 Outline . . . 3

1.1 Operational Risk . . . 3

1.2 Poisson Processes . . . 4

1.2.1 Mixed Poisson Processes . . . 5

1.2.2 Compound Poisson Processes . . . 6

1.3 Correlated Multivariate Poisson Processes . . . 7

1.4 Extreme Joint Distributions . . . 8

1.5 Simulation of Poisson Processes . . . 10

1.6 Thesis Outline . . . 11

1.7 Summary of Contributions . . . 12

2 Extreme Joint Distributions **13**

2.0.1 Outline . . . 16

2.1 Notation . . . 17

2.2 Extreme Measures in two-dimensions . . . 20

2.2.1 Monotone Sets and Distributions . . . 21

2.2.2 Extreme Joint Distributions in two-dimensions . . . 31

2.2.3 Equivalence to the Fréchet-Hoeffding Theorem . . . 37

2.3 The EJD Algorithm in two-dimensions . . . 39

2.3.1 The Comonotone Case . . . 40

2.3.2 The Antimonotone Case . . . 42

2.4 Extreme Measures in d-dimensions . . . 47

2.4.1 Monotonicity Structures . . . 49

2.4.2 Extreme Joint Distributions in d-dimensions . . . 51

2.4.3 Monotonicity in Higher Dimensions . . . 56

2.5 The EJD Algorithm in d-dimensions . . . 58

2.6 The Infinite Dimensional Case . . . 68

2.7 Numerical Example . . . 72

2.7.1 Support . . . 72

2.7.2 Marginal Distributions and the Joint Partition . . . 73

2.7.3 Probability Weights . . . 73

2.7.4 Extreme Correlations . . . 76

2.8 Calibration . . . 77
2.8.1 A Linear Algebraic Approach . . . 78
2.8.2 The Independent Case . . . 79
2.8.3 Admissible Correlations . . . 80
2.8.4 Calibration Algorithm . . . 82
2.9 Sampling from Multivariate Extreme Measures . . . 85
2.9.1 Sampling from Calibrated Measures . . . 85
2.10 Computational Complexity . . . 87
2.10.1 EJD Construction . . . 87
2.10.2 Extreme Correlation Matrix $C^{(j,d)}$ Construction . . . 87
2.10.3 Calibration . . . 88
2.11 Summary . . . 89

3 Backward Simulation of Poisson Processes 90
3.0.1 Outline . . . 91
3.0.2 Notation . . . 93
3.1 Backward Simulation . . . 94
3.1.1 Backward Simulation Algorithm in d-dimensions . . . 101
3.2 Correlation Structure . . . 102
3.3 Forward Continuation of the Backward Simulation . . . 108
3.3.1 Forward Correlation Structure . . . 108
3.3.2 Forward-Backward Simulation Algorithm in d-dimensions . . . 113
3.4 Forward versus Backward Simulation . . . 114
3.4.1 Forward Simulation . . . 114
3.4.2 Range of Correlations Restricted under FS . . . 115
3.5 Computational Complexity . . . 117
3.5.1 Backwards Simulation . . . 117
3.6 Summary . . . 118

4 Backward Simulation of Mixed Poisson Processes 119
4.0.1 Outline . . . 119
4.0.2 Notation . . . 120
4.1 Mixed Poisson Process . . . 121
4.2 Backward Simulation . . . 123
4.2.1 Backward Simulation Algorithm for mixed Poisson processes in d-dimensions . . . 127
4.3 Correlation Structure . . . 128
4.4 Forward Continuation of the Backward Simulation for Mixed Poisson Processes . . . 131
4.4.1 Forward Correlation Structure . . . 131
4.4.2 Forward-Backward Simulation Algorithm in d-dimensions . . . 135
4.5 Summary . . . 136

5 Backward Simulation of Compound Poisson Processes **138**
5.0.1 Outline . . . 138
5.0.2 Notation . . . 139
5.1 Compound Poisson Processes . . . 140
5.2 Backward Simulation . . . 142
5.2.1 Backward Simulation Algorithm for compound Poisson processes in d-dimensions . 148
5.3 Correlation Structure . . . 149
5.4 Forward Continuation of the Backward Simulation . . . 153
5.4.1 Forward Correlation Structure . . . 153
5.4.2 Forward-Backward Simulation Algorithm in d-dimensions . . . 157
5.5 Summary . . . 158

6 Conclusions and Future Work **159**
6.1 Summary and Conclusions . . . 159
6.1.1 Our Contributions . . . 161
6.2 Future Work . . . 162

Bibliography **163**

Appendices **171**

A Chapter 2.3 Proofs **172**
A.1 Proof of Theorem 3 . . . 172
A.2 Proof of Proposition 2 . . . 181

B Chapter 2.5 Proofs **184**
B.1 Proof of Lemma 16 . . . 184
B.2 Proof of Lemma 17 . . . 185
B.3 Proof of Lemma 23 . . . 185
B.4 Proof of Corollary 2 . . . 187
B.5 Proof of Lemma 22 . . . 190
B.6 Proof of Lemma 21 . . . 194
B.7 Lemma 24 . . . 195
B.7.1 Proof of Lemma 24 . . . 199

List of Tables

2.1 Joint partition of the extreme measure corresponding to comonotone case and the associated support and marginal cdf for each point. 73
2.2 Support and probabilities of three-dimensional extreme measures having Poisson marginals with parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$ corresponding to the monotone structures $(0, 0, 0)$ and $(0, 1, 0)$. 74
2.3 Support and probabilities of three-dimensional extreme measures having Poisson marginals with parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$ corresponding to the monotone structures $(0, 0, 1)$ and $(0, 1, 1)$. 75

List of Figures

1.1 Realized correlations (in the main body of the table above) and estimated intensities (in the right most column of the table above) associaed with the events listed in the top row and left most column of the table above. [32]. 1

2.1 Support of a bivariate comonotone distribution (left) and a bivariate antimonotone distribution (right). 29
2.2 Partitions of the unit interval in the comonotone case. 41
2.3 Monotonicity structures of extreme measures. The direction of the arrows indicates the direction of monotonically increasing support. 50
2.4 Different perspectives of the support of the 4 extreme measures for a three-dimensional joint distribution. 57
2.5 Partitions of the unit interval in the multivariate case 60
2.6 Supports of multivariate ($d = 3$) extreme measures with Poisson marginals having parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$ under all possible combinations of extremal dependence between the coordinates. The binary vector in square brackets on top of each figure indicates the corresponding monotonicity structure. 72
2.7 Extreme correlation matrices $\hat{C}^{(j,d)}$ corresponding to extreme measures $\hat{P}^{(j,d)}$ with given Poisson marginals having parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$. 76

3.1 The dashed black lines depict the correlation structure for two bivariate Poisson process, each with intensities 3 and 5, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.9$, respectively, computed by Backward Simulation. The blue circles depict the theoretical values according to Theorem 7. 107
3.2 The dotted black lines depict the correlation structure for a bivariate Poisson process with intensities 3 and 5, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.9$, respectively, computed by the Forward-Backward Simulation approach. The blue circles depict the theoretical values according to (3.45). 112

4.1 The dashed black lines depict the correlation structure for two bivariate Negative Binomial process, with means 3 and 5 and variances 5 and 7, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.9$, respectively, computed by Backward Simulation. The blue circles depict the theoretical values according to (4.17). The dashed red lines depict the Poisson case: two bivariate Poisson processes with intensities 3 and 5, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.9$, respectively, computed by Backward Simulation. 130

4.2 The dashed black lines depict the correlation structure for two bivariate Negative Binomial processes with means 3 and 5 and variances 5 and 7, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.8$, respectively, computed by the Forward-Backward Simulation approach. The blue circles depict the theoretical values according to (4.28). The dashed red lines depict the Poisson case: two bivariate Poisson processes with intensities 3 and 5, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.8$, respectively, computed by the Forward-Backward Simulation approach. 137

5.1 The dashed black lines depict the correlation structure, $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$, for a bivariate compound Poisson process, $Y_t = (Y_t^{(1)}, Y_t^{(2)})$, having a primary processes, $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$, a bivariate Poisson process with intensities 3 and 5, calibrated to $\hat{\rho}(5) = 0.7$ and $\hat{\rho}(5) = -0.9$, respectively, computed by Backward Simulation. The jump sizes $F_{Z^{(1)}}$ and $F_{Z^{(2)}}$ are lognormally distributed with parameters $(2.1235, 0.5)$ and $(1.9449, 0.75)$, respectively. The parameters for the jump distributions were selected for illustrative purposes. The blue circles depict the theoretical values for $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$ according to Theorem 13. The dashed red lines depict the correlation structure, $\hat{\rho}(t) = \text{corr}(X_t^{(1)}, X_t^{(2)})$, of the primary process, i.e., the bivariate Poisson process $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$. 152

5.2 The dashed black lines depict the correlation structure, $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$, for a bivariate compound Poisson process, $Y_t = (Y_t^{(1)}, Y_t^{(2)})$, computed by the Forward-Backward Simulation approach. The primary process $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$, is a bivariate Poisson process with intensities 3 and 5, calibrated to $\hat{\rho}(1) = 0.7$ and $\hat{\rho}(1) = -0.9$, respectively. The secondary distributions $F_{Z^{(1)}}$ and $F_{Z^{(2)}}$ are lognormal distributions with parameters $(2.1235, 0.5)$ and $(1.9449, 0.75)$, respectively. The blue circles depict the theoretical values for $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$ according to (5.40). The dashed red line depicts the correlation structure, $\hat{\rho}(t) = \text{corr}(X_t^{(1)}, X_t^{(2)})$, of the primary process, $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$. 156

List of Algorithms

1 Extreme Joint Distribution Algorithm in two-dimensions . . . 45
2 Subroutine: Reverse Support . . . 45
3 Subroutine: Construct Marginal CDFs . . . 46
4 Generation of monotonicity structures . . . 51
5 Extreme Joint Distribution Algorithm in d-dimensions . . . 70
6 Subroutine: Reverse support in d-dimensions . . . 71
7 Subroutine: Construct Marginal CDFs in d-dimensions . . . 71
8 Subroutine: Partition of Unity . . . 71
9 Calibration of Extreme Measures . . . 82
10 Subroutine: Flatten entries above diagonal of A . . . 82
11 Compute Correlation Matrix from a d-dimensional Extreme Measures . . . 83
12 Marginalize d-dimensional Extreme Measure into a d'-dimensional Extreme Measure . . . 83
13 Subroutine: Correlation Between Components of a Bivariate Extreme Measure . . . 84
14 Subroutine: Expectation of a Bivariate Extreme Measure . . . 84
15 Subroutine: Variance of a Bivariate Extreme Measure . . . 84
16 Subroutine: Dot Product Specialized to Birvariate Measures . . . 84
17 Sampling from Multivariate Extreme Measures . . . 85
18 Sampling from Calibrated Measures . . . 86
19 Backward Simulation of correlated multivariate Poisson processes . . . 101
20 Forward-Backward Simulation of correlated multivariate Poisson processes . . . 113
21 Backward Simulation of correlated multivariate mixed Poisson processes . . . 127
22 Forward-Backward Simulation of correlated multivariate mixed Poisson processes . . . 135
23 Backward Simulation of correlated multivariate compound Poisson Processes . . . 148
24 Forward-Backward Simulation of correlated multivariate compound Poisson processes . . . 157
25 Subroutine: Extend sets constructed in the $l^2_{\max}$ setting to $l^{(j,d)}_{\max}$ . . . 195
26 Subroutine: Construct $\hat{z}^{(j,d;uv)}$ from $z^{(j,d)}$ . . . 196

Chapter 1

Introduction

Poisson processes are the canonical stochastic processes used to model counts of random events and are broadly used in fields such as Insurance, Finance, Machine Learning, Operational Risk, Extreme Value Theory, Geophysics, and many others, where they are most commonly used in a univariate setting. Poisson processes are also often used in multivariate settings without correlation, i.e., multivariate Poisson processes with independent components. There are, however, many instances, in fields such as Geophysics [7] and Operational Risk [32], where events have been found to be correlated—even negatively correlated—with each other. An example of a correlation matrix exhibiting negative correlation that arises in Operational Risk [32] is reproduced below in Figure 1.1. This correlation matrix describes the realized dependence between operational events within a financial institution; note the strong negative correlations highlighted in red. Negative correlations between events have also been observed in earthquakes [7].

	CB-CP	CB-EF	CB-IF	RB-CP	RB-EF	RB-IF	RK-CP	RK-EF	RK-IF	λ
CB-CP	1									7.31
CB-EF	0.14	1								13.73
CB-IF	0.29	0.55	1							11.62
RB-CP	0.32	-0.12	0.11	1						16.59
RB-EF	0.15	0.49	0.27	-0.12	1					14.38
RB-IF	0.16	0.52	0.17	-0.23	0.49	1				22.68
RK-CP	0.03	-0.16	-0.31	0.19	-0.17	-0.02	1			14.78
RK-EF	0.05	0.2	0.05	-0.18	0.44	0.13	0.32	1		1.1
RK-IF	-0.06	0.02	0.08	-0.11	-0.03	0.29	0.5	0.16	1	8.17

Figure 1.1: Realized correlations (in the main body of the table above) and estimated intensities (in the right most column of the table above) associaed with the events listed in the top row and left most column of the table above. [32].

Multivariate models arise from modeling random vectors and are completely described by their joint distributions. For example, in the case of Operational Risk, losses are categorized by the tuple (event type, business line) and are known as risk cells. Let d be the number of Operational Risks that an institution is susceptible to. Then, we can denote the loss at each time t due to the k^{th} risk cell, by $L_t^{(k)}$, where $k = 1, \ldots, d$. The total losses, at time t, of the d possible Operational Risks can be modeled as a d-dimensional vector $\mathbf{L}_t = (L_t^{(1)}, \cdots, L_t^{(d)})$. Summing the component losses of the total Operational Risk random vector, $\mathbf{L}_t$, gives the aggregate loss $\sum_{k=1}^{d} L_t^{(k)}$ at time t [96].

In general, correctly estimating the joint distribution of a multivariate model given empirical data, also known as *calibration*, is a difficult problem [113]. One approach in determining a joint distribution that corresponds to some given data is to estimate the marginal distributions and the dependence structure[1] separately. Then, suitable joint distributions satisfying the given marginals and possessing the desired dependence structure can be determined by methods such as Mathematical Programming. (See, for example, Section 2.8.) Note that there are many possible joint distributions even when the marginals are given and the dependence structure—usually measured by a correlation matrix—is specified, since higher order interactions and moments are not specified [53, 109, 119].

A natural question to ask is then:

Is it possible to construct correlated multivariate Poisson processes with given marginals and a specified dependency structure, including negative correlations?

The answer to this question is—thankfully—a yes. Previous work in the literature on this topic include: the works by Fréchet [49] and Hoeffding [64] from which the EJD algorithm was inspired; the works by Tiit [121, 122, 124]; the NORmal To Anything (NORTA) approach [54]; the approach by Li et al. [78]; and the approach by Karlis [70]. This problem, more generally, has a long history, being related to what has classically been known as problems of "Distributions with Given Marginals" [64, 109, 121, 124, 128] or the problem of "Extremal Distributions" [122] in the literature and also has rich connections to Optimal Transport[2] and Mathematical Programming [105].

This thesis introduces a novel approach for the construction and simulation of correlated multivariate Poisson processes and is based on two pillars: the Extreme Joint Distribution (EJD) approach (Chapter 2) and Backward Simulation (Chapters 3-5).

The EJD approach, first introduced in the bivariate setting [74] and subsequently extended to the multivariate setting [17], is a pure probabilistic approach for the computation of discrete extreme measures—probability measures having specified marginals that exhibit extremal dependency in the form of extreme values for the correlation coefficient; indeed there is a one-to-one relationship between an extreme measure and an extreme dependence structure (correlation coefficient in the bivariate setting and correlation matrix in the multivariate setting). Convex combinations of extreme measures are used to construct a multivariate measure having specified marginals and a specified correlation coefficient in the bivariate setting or a specified correlation matrix in the multivariate setting. The EJD approach provides an efficient algorithm for computing extreme measures.

Backward Simulation is a method of simulating Poisson processes that, at a high level, starts from the end of interval $[0, T]$ and works backwards in time to simulate the Poisson process. This is possible by exploiting the well-known *conditional uniformity* property of univariate Poisson processes: conditional on the total number of events n at time T, the n arrival moments are uniformly distributed in the interval $[0, T]$. Crucially, Backward Simulation extends straightforwardly to the multivariate setting since Backward Simulation of a multivariate Poisson process consists of applying Backward Simulation to each univariate Poisson process component of the multivariate Poisson process. This requires a vector of terminal events at time T, which is sampled from a suitable joint distribution; the EJD approach is used to construct the joint distribution with given marginals and a specified dependence structure.

[1] If there is no dependence structure, i.e., the components of the multivariate process are independent, the joint distribution factorizes multiplicatively and is completely specified by the product of the marginal distributions.

[2] We only mention this in passing for the sake of completeness; this connection is not the focus of our thesis.

1.0.1 Outline

The outline for this chapter is as follows. We begin in Section 1.1 by reviewing the motivating problem originating from Operational Risk. In Section 1.2, we briefly recapitulate the basics of Poisson processes to review notations and definitions. Section 1.3 reviews some constructions of correlated Poisson processes. Section 1.4 introduces the Extreme Joint Distribution approach. Section 1.5 discusses the general approaches to the simulation of Poisson processes. Section 1.6 provides an outline of the thesis. Section 1.7 summarizes the contributions of this thesis.

1.1 Operational Risk

Operational Risk (OpRisk) within a financial institution is broadly defined as the risk of loss resulting from inadequate or failed internal processes, people, systems, or from external events [21, 94, 96, 116]. Historically, OpRisk was managed implicitly, often subsumed under credit or market risk, until the introduction of the Basel II regulatory framework [94], which required financial institutions to explicitly quantify and provision capital against such risks.

A widely adopted modeling framework for OpRisk is the Loss Distribution Approach (LDA) [4]. In the LDA, total losses over a fixed horizon are modeled as the aggregation of losses across risk cells, where the loss in each cell is represented as a compound-type process consisting of a frequency component and a severity component. Specifically, the number of events in each risk cell is modeled as a counting process—most commonly a Poisson or mixed Poisson process—while individual loss amounts are modeled separately via a severity distribution. The frequency and severity components are typically assumed to be independent and are calibrated independently. More formally, let

$$L_t = L_t^{(1)} + \cdots + L_t^{(d)} \tag{1.1}$$

where the $L_t^{(j)}$, for $j = 1, 2, \ldots, d$, model the annual loss at time t in risk cell j. Each $L_t^{(j)}$ is given by

$$L_t^{(j)} = \sum_{i=1}^{N_t^{(j)}} Z_i^{(j)} \tag{1.2}$$

where each $Z_i^{(j)}$, random variables drawn from a corresponding severity distribution, models the i^{th} loss in risk cell j. The $Z_i^{(j)}$ are usually assumed to be iid. $N_t^{(j)}$ describes the number of events occurring within risk cell j up until time t and is known as the *frequency process*, typically modelled as a Poisson process or a Negative Binomial process [96]. The random variables $Z_i^{(j)}$ and the frequency processes $N_t^{(j)}$ are generally assumed to be independent and are estimated separately.

From the perspective of frequency modeling, the LDA reduces the OpRisk problem to the construction and simulation of multivariate counting processes with given marginal distributions and a prescribed dependence structure across components. Dependence between risk cells enters primarily through the joint behavior of the frequency processes, and it is precisely at this level that negative correlations have been observed empirically. Accurately modeling such dependence is critical, as it directly impacts the distribution of aggregate losses obtained by summing across risk cells.

While the LDA provides a concrete applied context in which correlated multivariate Poisson processes naturally arise, the scope of this thesis is not restricted to Operational Risk modeling. The constructions

and simulation techniques developed herein address the general problem of generating multivariate count processes with specified marginals and flexible dependence structures, including negative correlations. As such, the OpRisk and LDA frameworks serve primarily as motivating examples, illustrating the practical relevance of the theoretical developments, rather than as recurring applications in the subsequent chapters.

1.2 Poisson Processes

We briefly review basics properties of the Poisson processes discussed in this thesis—Poisson, mixed Poisson, and compound Poisson—and introduce some related notation. See standard references such as [23, 24, 51, 58, 71, 75] for more in-depth discussions on Poisson processes. We begin by recalling the definition of a counting process.

Definition 1 (Counting Process [44]). *A counting process* $\{N_t; t \geq 0\}$ *is a stochastic process such that*

1. $N_t \geq 0$,
2. N_t *is an integer,*
3. *If* $s \leq t$ *then* $N_s \leq N_t$.

Counting processes are a very general class of integer-valued non-decreasing processes that are well suited for modelling counts of random events. One of the most important counting processes is the Poisson process.

Definition 2 (Poisson Distribution). *A random variable* N *is said to be Poisson distributed with parameter* λ, $\text{Pois}(\lambda)$, *if*

$$\mathbb{P}(N = n) = e^{-\lambda}\frac{\lambda^n}{n!} \qquad \text{for} \quad n = 0, 1, 2, \ldots \quad \text{and} \quad \lambda \geq 0. \tag{1.3}$$

Definition 3 (Poisson Process [51]). *A counting process* $\{N_t; t \geq 0\}$ *is a Poisson process with intensity* λ *if*

1. $N_0 = 0$,
2. N_t *has independent increments,*
3. $N_t - N_s \sim \text{Pois}(\lambda(t-s))$ *for* $s < t$.

where $\text{Pois}(\lambda(t-s))$ denotes a Poisson distribution with parameter $\lambda(t-s)$, defined above. The conditions listed in Definition 3 are one characterization of Poisson processes and are derived from the more general theory of Point processes [23, 24].

Remark 1. *Generalizations of Poisson processes, such as the mixed Poisson and compound Poisson processes, can be characterized similarly to Definition 3, but the increments are mixed Poisson and compound Poisson distributed, respectively. See Sections 1.2.1 and 1.2.2 for more details.*

Poisson processes posses an important property known as *conditional uniformity*.

Proposition 1 (Conditional Uniformity [51]). *The joint probability density function, f, of the arrival moments $T_1, T_2, \ldots, T_n$ of the Poisson process, N_t, conditioned on the event $N_t = n$, is given by*

$$f_{(T_1,\ldots,T_n \mid N_t=n)}(x_1, x_2, \ldots, x_n) = \begin{cases} \frac{n!}{t^n} & \text{if} \quad 0 \le x_1 \le x_2 \le \cdots \le x_n \le t \\ 0 & \text{otherwise.} \end{cases} \tag{1.4}$$

The conditional uniformity property is what enables the Backwards Simulation of Poisson processes; mixed Poisson and compound Poisson processes also posses this crucial property [58].

There are many ways to construct and define multivariate Poisson distributions [53, 69, 120]. We define the extensions of Definitions 2 and 3 to the multivariate setting below.

Definition 4 (Multivariate Poisson Distribution). *A multivariate random variable $\mathbf{N} = (N^{(1)}, ..., N^{(d)})$ is said to be Poisson distributed with parameter $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ if every component $N^{(k)}$ for $1 \le k \le d$ is a Poisson random variable with parameter λ_k.*

Remark 2. *Definition 4 implies that the d-dimensional Poisson probability function $P^{(d)} \sim \mathbf{N}$ satisfies marginal constraints of the form*

$$\sum_{v \in \mathcal{I}_k} \sum_{i_v=0}^{\infty} P^{(d)}_{i_1,\ldots i_{k-1}, i_k, i_{k+1},\ldots,i_d} = Q^{(k)}_{i_k} \tag{1.5}$$

where $k = 1, \ldots, d$, $i_k = 0, 1, \ldots$, $\mathcal{I}_k = \{r : 1 \le r \le d, r \ne k\}$, and $Q^{(k)}_{i_k}$ is element i_k of the k^{th} Poisson probability distribution $Q^{(k)} \sim N^{(k)}$.

Definition 5 (Multivariate Poisson Process). *A multivariate Poisson process $\boldsymbol{N}_t = (N^{(1)}_t, \ldots, N^{(d)}_t)$ is a stochastic process such that every component $N^{(k)}_t$ for $1 \le k \le d$ is a univariate Poisson process having intensity λ_k and that, at each time t, $\boldsymbol{N}_t$ is multivariate Poisson distributed.*

1.2.1 Mixed Poisson Processes

The mixed Poisson process is one generalization of the Poisson process used in many settings where the empirical data displays overdispersion.

Definition 6 (Mixed Poisson Distribution [58]). *A discrete random variable $\hat{N}$ is said to be mixed Poisson distributed, MP(U), with structure distribution U, if*

$$\begin{aligned} \mathbb{P}(\hat{N} = n) &= \mathbb{E}\left[\frac{(\Lambda)^n}{n!} e^{-\Lambda}\right] \\ &= \int_{0-}^{\infty} \frac{(\lambda)^n}{n!} e^{-\lambda} dU(\lambda), \quad n = 0, 1, 2, \ldots \end{aligned} \tag{1.6}$$

where Λ is a random variable distributed according to U.

Definition 7 (Mixed Poisson Process). *$\hat{N}_t$ is a mixed Poisson process if it is $\mathrm{MP}(t, U)$-distributed for all $t \ge 0$. That is,*

$$\mathbb{P}(\hat{N}_t = n) = \int_{0-}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} dU(\lambda), \quad n = 0, 1, 2, \ldots . \tag{1.7}$$

The mixed Poisson process is a Poisson process with a non-negative random intensity.

Definition 8 (Multivariate Mixed Poisson Distribution)**.** *A multivariate mixed Poisson distribution* $\hat{\boldsymbol{N}} = (\hat{N}^{(1)}, \ldots, \hat{N}^{(d)})$ *is a multivariate distribution such that every component* $\hat{N}^{(k)} \in \mathrm{MP}(U^{(k)})$ *for* $1 \leq k \leq d$ *is a univariate mixed Poisson distribution having structure distribution* $U^{(k)}$*.*

Remark 3. *Definition 8 implies that the d-dimensional mixed Poisson distribution, similar to Remark 2, satisfies marginal constraints of the form (1.5) where the marginals* $Q^{(k)}$ *on the right side of (1.5) are now mixed Poisson distributions rather than Poisson distributions.*

Definition 9 (Multivariate Mixed Poisson Process)**.** *A multivariate mixed Poisson process* $\hat{\boldsymbol{N}}_t = (\hat{N}_t^{(1)}, \ldots, \hat{N}_t^{(d)})$ *is a stochastic process such that every component* $\hat{N}_t^{(k)} \in \mathrm{MP}(t, U^{(k)})$ *for* $1 \leq k \leq d$ *is a univariate mixed Poisson process with structure distribution* $U^{(k)}$ *and that, at each time* t*,* $\hat{\boldsymbol{N}}_t$ *is multivariate mixed Poisson distributed.*

1.2.2 Compound Poisson Processes

Definition 10 (Compound Poisson Distribution [40])**.** *The random variable*

$$S = Z_1 + Z_2 + \cdots + Z_N$$

is said to be compound Poisson distributed if $N \sim \mathrm{Pois}(\lambda)$ *and the random variables* $\{Z_i\}_{i=1}^{\infty}$ *are identically and independently distributed having the common distribution* F_Z *which is independent of* N *and has finite expectation and finite variance.* N *is known as the primary random variable having, in this case,* $\mathrm{Pois}(\lambda)$ *as the primary distribution and* F_Z *is known as the secondary distribution.*

Definition 11 (Compound Poisson Process [40])**.** *The process*

$$L_t = \sum_{i=1}^{N_t} Z_i \tag{1.8}$$

is said to be a compound Poisson process if N_t*, known as the primary process, is a Poisson process with intensity* λ *and the random variables* Z_i *are identically and independently distributed having the common distribution* F_Z *which has finite expectation and finite variance and is independent of the primary process* N_t*.*

Definition 12 (Multivariate Compound Poisson Distribution)**.** *A multivariate compound Poisson distribution* $\mathbf{S} = (S^{(1)}, \ldots, S^{(d)})$ *is a multivariate distribution with univariate compound Poisson marginals having Poisson primary distributions with parameters* $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$ *and secondary distribution* F_Z*.*

Remark 4. *Definition 12 implies that the d-dimensional compound Poisson distribution, similar to Remark 2, satisfies marginal constraints of the form (1.5) where the marginals* $Q^{(k)}$ *on the right side of (1.5) are now compound Poisson distributions rather than Poisson distributions.*

Definition 13 (Multivariate Compound Poisson Process)**.** *A multivariate compound Poisson process* $\boldsymbol{N}_t = (N_t^{(1)}, \ldots, N_t^{(d)})$ *is a stochastic process such that every component* $N_t^{(i)}$ *for* $1 \leq i \leq d$ *is a univariate compound Poisson process having a Poisson primary process with intensity* λ_i *and secondary distribution* F_Z *and that, at each time* t*,* $\boldsymbol{N}_t$ *is multivariate compound Poisson distributed.*

1.3 Correlated Multivariate Poisson Processes

In this thesis, we use the Pearson correlation as the measure of dependency in the multivariate setting. The Pearson correlation for a bivariate Poisson process, $(N_t^{(1)}, N_t^{(2)})$, with intensities (λ_1, λ_2) at time t is

$$\rho(t) = \frac{\mathbb{E}\left[N_t^{(1)} N_t^{(2)}\right] - \mathbb{E}\left[N_t^{(1)}\right]\mathbb{E}\left[N_t^{(2)}\right]}{\sigma(N_t^{(1)}) \cdot \sigma(N_t^{(2)})}. \tag{1.9}$$

Here, $\sigma^2(X)$ denotes the variance of the random variable X, defined by $\sigma^2(X) = \mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])^2\right]$. In particular, for a Poisson process N_t with intensity λ, $\sigma^2(N_t) = \lambda t$. We can see from (1.9) that the only term that is not determined by the distributional parameters of the Poisson processes is the joint expectation, $\mathbb{E}\left[N_t^{(1)} N_t^{(2)}\right]$. Therefore, extremizing $\rho(t)$ is equivalent to extremizing the joint expectation, $\mathbb{E}\left[N_t^{(1)} N_t^{(2)}\right]$. This realization guides the choice of the objective function in the optimization problems (2.3) in the two-dimensional case and (2.37) in the d-dimensional case.

One of the early attempts in introducing dependency between Poisson processes originated from actuarial modelling, known as the Common Shock Model (CSM) [82, 102, 116], wherein a third Poisson process is used to couple two otherwise independent Poisson processes. To illustrate this, let $(\nu_t^{(1)}, \nu_t^{(2)}, \nu_t^{(3)})$ be three independent Poisson processes with intensities $(\lambda_1, \lambda_2, \lambda_3)$ We can obtain two correlated Poisson processes from the three independent Poisson processes through superposition:

$$N_t^{(1)} := \left(\nu_t^{(1)} + \nu_t^{(2)}\right)$$

and

$$N_t^{(2)} := \left(\nu_t^{(2)} + \nu_t^{(3)}\right)$$

having intensities $\mu_1 = \lambda_1 + \lambda_2$ and $\mu_2 = \lambda_2 + \lambda_3$. The correlation coefficient between the Poisson processes $N_t^{(1)}$ and $N_t^{(2)}$ in the CSM satisfies [82]

$$\rho = \frac{\lambda_2}{\sqrt{\mu_1 \cdot \mu_2}}.$$

The latter relation immediately implies that

$$0 \leq \rho \leq \sqrt{\frac{\min(\mu_1, \mu_2)}{\max(\mu_1, \mu_2)}}.$$

It is clear that, in the CSM, correlations do not depend on time and do not allow for negative correlations. Therefore the CSM cannot reproduce the negative correlations exhibited in Figure 1.1. In addition, the upper bound on the ρ, above, is in some cases less than that attainable by our method (see Chapter 3). For these reasons, the CSM has significant deficiencies as a method for constructing correlated multivariate Poisson processes that often appear in practice.

We note that our approach to constructing multivariate distributions is not a Copula based approach. Our Extreme Joint Distribution approach (see the next section for a brief introduction and Chapter 2 for a detailed exposition) is a direct approach that generates joint distributions with prescribed marginal distributions (that are not uniform distributions as would be the case for a Copula-based approach) and prescribed correlations using Mathematical Programming.

1.4 Extreme Joint Distributions

First introduced by Kreinin [74] for the bivariate setting, the Extreme Joint Distribution (EJD) methodology is a pure probabilistic based approach for constructing discrete joint distributions having given marginals and extreme correlations between their components. In the two-dimensional case, extreme joint distributions are solutions to an infinite-dimensional Linear Program (2.3) where the objective function is the bivariate joint expectation (2.3e) and the constraints are the marginals. Recall from the discussion in the previous section that extremizing the bivariate joint expectation is equivalent to extremizing the correlation coefficient ρ between two random variables $X^{(1)}$ and $X^{(2)}$. Extreme joint distributions are useful for constructing multivariate discrete probability measures with given marginals and having a prescribed correlation matrix. The intuition behind this is best illustrated in the bivariate setting. To that end, note that in the two-dimensional case there are only two extremal bivariate distributions $\hat{P}^{(1)}$ and $\hat{P}^{(2)}$, each having the specified marginals $Q^{(1)}$ and $Q^{(2)}$, $\hat{P}^{(1)}$ has the extremal correlation $\hat{C}^{(1)}$ and $\hat{P}^{(2)}$ has the extremal correlation $\hat{C}^{(2)}$, where $-1 \leq \hat{C}^{(1)} \leq \hat{C}^{(2)} \leq 1$, corresponding to the extremal positive and extremal negative correlations attainable between two random variables $X^{(1)}$ and $X^{(2)}$ distributed according to the specified marginals $Q^{(1)}$ and $Q^{(2)}$, respectively. The importance of extremal correlations[3] stems from the fact that, for any correlation $C \in [\hat{C}^{(1)}, \hat{C}^{(2)}]$, we can easily find a $w \in [0, 1]$ such that

$$C = w\,\hat{C}^{(1)} + (1 - w)\,\hat{C}^{(2)}. \tag{1.10}$$

By itself, this convex combination is not significant. However, the w that satisfies (1.10) can be substituted into

$$P = w\,\hat{P}^{(1)} + (1 - w)\,\hat{P}^{(2)} \tag{1.11}$$

to obtain a bivariate probability measure P having the given marginals $Q^{(1)}$ and $Q^{(2)}$ and the associated correlation coefficient C.

In the d-dimensional setting, extreme joint distributions are solutions to a multi-objective infinite-dimensional Linear Program (2.37) and our approach to (1.10) and (1.11) holds with some modification. Since there are $n = 2^{d-1}$ extreme measures (see Section 2.4), (1.10) and (1.11) generalize to

$$C = w_1\hat{C}^{(1)} + \cdots + w_n\hat{C}^{(n)} \tag{1.12}$$

and

$$P^{(C)} = w_1\hat{P}^{(1)} + \cdots + w_n\hat{P}^{(n)}, \tag{1.13}$$

respectively, where $w_i \in [0, 1]$ for all $i = 1, 2, \ldots, n$ and $\sum_{i=1}^{n} w_i = 1$. The solution $(w_1, \ldots, w_n)$ to (1.12) can be substituted in (1.13) to obtain a multivariate probability measure $P^{(C)}$ having given marginals $Q^{(1)}, \ldots, Q^{(d)}$ and the associated correlation matrix C. Note that, in the d-dimensional setting, (1.12) is system of linear equations having constraints on the $\{w_i : i = 1, 2, \ldots, n\}$ that can be solved using Mathematical Programming. (See Section 2.8.1.) If there is no solution to (1.12), then C is said to be *inadmissible*; otherwise, it is *admissible*. (See Section 2.8.3.)

Thus, the EJD methodology can be used to construct joint distributions having an admissible prescribed dependence structure and satisfy given marginal distributions $Q^{(1)}, \ldots, Q^{(d)}$. As explained in Section 2.9, for a multivariate Poisson process having univariate Poisson marginals, $P^{(C)}$ can be used

[3] The method still works if $\hat{C}^{(1)}$ and $\hat{C}^{(2)}$ were not extreme, but it would not be possible to calibrate to the full range of admissible correlations. That is some admissible correlations would be considered as inadmissible.

to sample the (correlated) events from a multivariate Poisson distribution with correlation matrix C in order to obtain the vector of events at terminal time T. This is used by Backward Simulation to exploit the conditional uniformity property to generate the sample paths of a correlated multivariate Poisson process. (See Chapters 3, 4, and 5.)

1.5 Simulation of Poisson Processes

There are two general approaches to the simulation of Poisson processes within a simulation interval $[0, T]$: the forward approach and the backwards approach. To begin, we describe below both Forward Simulation and Backward simulation in one dimension. Forward Simulation is the more intuitive approach of starting at time $t = 0$ and simulating the (exponentially-distributed) interarrival times until the end of the simulation interval, $t \geq T$, is reached. In contrast, Backward Simulation relies on the conditional uniformity property (Proposition 1), also known as the order statistic property, of Poisson processes: given the number of events, n, at terminal simulation time, T, the arrival moments of the Poisson process are uniformly distributed. Thus, Backward Simulation consists of simulating the number of terminal events, n, from a suitable distribution at the end of the simulation interval $t = T$, and then sampling n uniform random variables in the interval $[0, T]$. Sorting the n uniform random variables gives the arrival moments of the Poisson process within the simulation interval $[0, T]$[4]. In the univariate setting, the Forward Simulation and Backward Simulation are quite similar in terms of the simplicity of the methodology and the ease of implementation.

In multivariate settings, however, the two approaches differ significantly, with the backward approach exhibiting many advantages over the forward approach. For example, in the bivariate setting, to the best of our knowledge, only independence or extreme correlations are possible under Forward Simulation, whereas, all admissible correlations can be obtained under the Backward Simulation approach. (See Section 3.4 for a discussion comparing Forward and Backward Simulation.) Moreover, to the best of our knowledge, no Forward approach is capable of generating correlated multivariate Poisson processes for dimensions $d > 2$. In contrast, Backward Simulation extends naturally to the general d-dimensional setting, for $d \geq 2$, since Backward Simulation of a d-dimensional Poisson process simply consists of Backward Simulation of each univariate component with the number of terminal events obtained from a suitable joint distribution at terminal simulation time T.

This thesis explores the extension of the Backward Simulation approach to the general setting where $d \geq 2$ and also to more general Poisson process such as the mixed Poisson process (Chapter 4) and the compound Poisson process (Chapter 5). A key property of Poisson processes under Backward Simulation is that the correlation structure is a linear function of time

$$\rho(t) = \frac{t}{T} \cdot \rho(T)$$

for $t \in [0, T]$. Moreover, the thesis also introduces a methodology for extending a Poisson process simulated by Backward Simulation within the interval $[0, T]$ to intervals $[mT, (m+1)T]$ for any positive integer m, where the correlation exhibits the asymptotic behavior

$$\rho(mT + \tau) \approx \rho(T)$$

for m sufficiently large and all $\tau \in [0, T]$. This extension is known as the Forward Continuation of Backward Simulation. Note that the first few intervals of the process simulated under the Forward Continuation of Backward Simulation can be discarded, similar to a burn-in phase, so that the process exhibits a near constant correlation structure.

[4] Ordered uniform random variables can be simulated directly using order statistics; see [3].

1.6 Thesis Outline

This thesis develops a unified framework for constructing and simulating correlated multivariate Poisson processes that attain the full range of dependence structures, including extreme negative correlation. We integrate the two pillars—Extreme Joint Distributions (EJD) and Backward Simulation (BS)—and extend them substantially in both theoretical scope and computational applicability.

- Chapter 2 develops the theory of Extreme Joint Distributions. We extend the EJD approach from the bivariate setting in [74] to the general d-dimensional case, characterize the associated monotonicity structures, and construct extremal multivariate measures with Poisson marginals.
- Chapter 3 introduces Backward Simulation for Poisson processes. Using the extreme measures constructed in Chapter 2, we calibrate joint distributions to target correlation matrices—including those with negative entries—and derive the resulting correlation structures under BS and the Forward Continuation of Backward Simulation.
- Chapter 4 generalizes the Backward Simulation-based framework to mixed Poisson processes and characterizes the resulting correlation structures.
- Chapter 5 extends the Backward Simulation framework further to compound Poisson processes. We establish how the dependence structure of the primary process interacts with the secondary distribution and derive correlation structures under both Backward Simulation and Forward Continuation.
- Chapter 6 summarizes the theoretical developments, revisits the contributions of the thesis and outlines directions for future work.

1.7 Summary of Contributions

The main contributions of this thesis are as follows:

1. A general d-dimensional EJD framework.

   We develop extremal joint distributions for discrete distributions in the general d-dimensional setting, providing the foundation for modeling extreme dependence in multivariate Poisson processes.

2. A calibration method for multivariate discrete distributions using EJD.

   The proposed approach constructs joint distributions having prescribed discrete marginal distributions and admissible correlation matrices. In the case of Poisson marginals, this method overcomes previous limitations of existing constructions that cannot accommodate negative dependence.

3. Extended Backward Simulation for multivariate mixed and compound Poisson processes.

   We prove that mixed and compound Poisson processes simulated via Backwards Simulation remain said processes. We show how Backwards Simulation can be applied component-wise, once an appropriate joint distribution has been calibrated, while preserving the dependence structure.

4. A Forward-Backward Simulation scheme enabling process continuation.

   After the simulation of Poisson processes within a target interval $[0, T]$, it is natural to consider extending the process to subsequent intervals $[nT, (n+1)T]$. We propose a method, the Forward Continuation, that extends the process by keeping the distribution of the increments the same.

5. Theoretical analysis and numerical validation.

   We establish correctness results not just for the EJD Theorem in d-dimension, but also for the EJD algorithms. We numerically validate the behavior of the correlations described by the theory using Monte Carlo analysis.

A more extensive list of contributions, with references to sections of the thesis, can be found in Section 6.1.1

Chapter 2

Extreme Joint Distributions

This chapter forms one of the pillars of the thesis and describes, in detail, an important part in the construction of multivariate Poisson processes with correlation—negative correlation in particular—between their components. A necessary capability for calibrating a multivariate Poisson process at a certain point in time to some data, is the ability to obtain a joint distribution that has prescribed marginal distributions and a correlation structure consistent with the data. A calibrated joint distribution enables the calculation of various statistics, but more importantly, it enables the generation of random vectors, which are necessary for our Backward Simulation method (see Chapters 3, 4, and 5).

The problem of constructing joint distributions having prescribed marginal distributions and possessing a desired dependence structure has been given much consideration in the literature. Two of the pioneering works are by Fréchet [49] and Hoeffding [64] who introduced the extremal cdfs H^* and H_* having maximal and minimal correlation coefficient, respectively, within the set $\Pi(F, G)$ of all cdfs H on $\mathbb{R}^2$ having F and G as marginal cdfs where F and G have positive finite variance. Later, Whitt [128] studied extremal bivariate distributions and their correlations in a more general setting and Strassen [119] considered the problem of existence of the bivariate problem but for general probability measures. Hill et al. [63] continues the theme of partially specifying joint distributions through prescribed marginal distributions and a prescribed correlation structure. Similar to Fréchet [49] and Whitt [128], Hill et al. [63] is close in spirit to our Extreme Joint Distribution (EJD) method as they present a method for determining the mixing probabilities of extremal joint probability mass functions (pmfs) to achieve a certain correlation structure. Unlike our work, however, the work of Hill et al. [63] is limited to the lower-dimensional setting $d \leq 3$ and is unable to generate *all* feasible correlation matrices for a set of marginals [53].

A separate but related body of work focuses on obtaining random vectors satisfying prescribed marginal distributions and correlation structure directly, without explicitly specifying the joint distribution due to its difficulty; this is especially true for cases where the marginals do not come from the same parametric family [27, 67]. Early works such as Li and Hammond [78] specified a method of constructing a joint distribution that satisfies prescribed marginal distributions and correlations by applying nonlinear transformations to normally distributed random variables. However, this method is computationally inefficient since it requires the inversion of a double integral where the integrand itself must be numerically approximated. The nonlinear transformations make it difficult to specify an input correlation matrix that will remain positive semidefinite after the transformation. Lurie and Goldberg [85] introduce a

nonlinear optimization step in order to alleviate this very issue. A more recent and popular body of work on the same theme is the NORmal To Anything (NORTA) method [14, 53, 54, 55] which generates iid variates of a random vector $X = (X^{(1)}, \ldots, X^{(d)})$ where each component is distributed according to prescribed marginal distributions $Q^{(j)}$ for $j = 1, \ldots, d$ and a prescribed correlation structure in the form of a correlation matrix C. The NORTA method accomplishes this by first sampling a random vector $Z = (Z_1, \ldots, Z_d)$ that is normally distributed with mean vector $\mathbf{0}$ and covariance matrix C^Z. Then, the nonlinear transformation $X_i = F_i^{-1}(\phi(Z_i))$ is applied to each component, where, $\phi(\cdot)$ is the distribution function of a standard normal random variable, $F^{(i)}$ is the cumulative distribution function corresponding to the marginal distribution $Q^{(i)}$ and

$$F_i^{-1}(u) = \inf\{x : F^{(i)}(x) \geq u\}$$

is the quantile function[1] corresponding to marginal distribution $Q^{(i)}$. While the NORTA method is simple and straightforward, it has many shortcomings. An example of this is that, similar to earlier work in the literature [78, 85], the intermediate correlation matrix C^Z must be chosen such that the resulting correlation remains positive semidefinite, after the nonlinear transformations are applied. A remedy for this was suggested in Ghosh and Henderson [54] by adding a Semidefinite Programming (SDP) step in the initialization of the NORTA method; an obvious downside to this is the added computational cost. A more serious shortcoming is that NORTA fails for the case $d > 3$ in that, as the dimensionality of the random vector grows, NORTA becomes increasingly unable to match feasible correlation matrices corresponding to a given set of prescribed marginal distributions; in particular, they become increasingly unable to attain extreme correlations as the dimension increases [53, 55].

A result worth noting is that bivariate Poisson processes exhibiting negative correlations are not infinitely divisible [59]. Thus, while bivariate Poisson distributions constructed using a "random elements in common" decomposition [34], where (X, Y) is a bivariate Poisson distribution such that $X = U + V$ and $Y = V + W$ where U, V, W are three independent Poisson random variables[2] are infinitely divisible, the extreme distributions constructed by the EJD method exhibiting negative correlations are not.

Let us now introduce the intuition behind our approach. Consider first the bivariate setting and suppose we are given the extremal correlations $\hat{C}^{(1)}$ and $\hat{C}^{(2)}$ where $-1 \leq \hat{C}^{(1)} \leq \hat{C}^{(2)} \leq 1$[3] corresponding to the extremal positive and extremal negative correlations attainable between two random variables $X^{(1)}$ and $X^{(2)}$ distributed according to $Q^{(1)}$ and $Q^{(2)}$, respectively. The extremal correlations determine an admissible range $[\hat{C}^{(1)}, \hat{C}^{(2)}]$ from which any correlation, C, within the admissible range can be attained by first solving the simple linear equation

$$C = w\,\hat{C}^{(1)} + (1 - w)\,\hat{C}^{(2)} \tag{2.1}$$

for $w \in [0, 1]$. Crucially, (2.1) relies on being able to generate the extreme correlations $\hat{C}^{(k)}$, for $k \in \{1, 2\}$, which can be obtained from the corresponding extreme distributions, $\hat{P}^{(k)}$, given by either the Fréchet-Hoeffding theorem [49, 64] in the two-dimensional case or by our EJD method in Section 2.2. Each two-dimensional extreme distribution $\hat{P}^{(k)}$, for $k = 1$ or 2, has correlation $\hat{C}^{(k)}$ and marginals

[1] This is also known as a generalized inverse; see [38] for details.
[2] This is the distribution version of the Common Shock Model (CSM) [82] which applies to processes.
[3] Note that there are cases where $-1 < \hat{C}^{(1)}$ and $\hat{C}^{(2)} < 1$.

$Q^{(1)}$ and $Q^{(2)}$. Key to our approach is that the solution w to (2.1), when plugged into

$$P = w\hat{P}^{(1)} + (1-w)\hat{P}^{(2)}, \tag{2.2}$$

specifies a probability distribution P having the marginal distributions $Q^{(1)}$ and $Q^{(2)}$ and correlation C.

The approach outlined in the paragraph above can be extended to higher dimensions. The $\hat{C}^{(k)}$ are no longer scalar values, but are (extreme) correlation matrices. There are $n = 2^{d-1}$ extreme correlation matrices in the d-dimensional case—see Lemma 15 in Section 2.4. Therefore, in the d-dimensional setting (2.1) becomes

$$C = w_1\hat{C}^{(1)} + \cdots + w_n\hat{C}^{(n)}$$

with the constraints $w_j \geq 0$ for $j = 1, \ldots, n$ and $\sum_j w_j = 1$. We discuss ways to solve this equation for the weights $(w_1, \ldots, w_n)$ in Section 2.8. The extreme correlation matrices $\hat{C}^{(k)}$ are unique and each $\hat{C}^{(k)}$ is associated with a d-dimensional extreme probability measure $\hat{P}^{(k)}$ satisfying the marginal distributions $(Q^{(1)}, \ldots, Q^{(d)})$. A natural question that arises is how to define and compute $\hat{P}^{(k)}$; this is the topic of Section 2.4. In the d-dimensional case, similar to the two-dimensional case, the weight vector $(w_1, \ldots, w_n)$ when substituted into the multidimensional analogue of (2.2),

$$P = w_1\hat{P}^{(1)} + \cdots + w_n\hat{P}^{(n)},$$

specifies the (multivariate) probability distribution P having marginal distributions $Q^{(1)}, \ldots, Q^{(d)}$ and correlation matrix C.

Kreinin in [74] introduced the EJD approach for the bivariate case, a pure probabilistic approach for the computation of extreme measures and showed that it is equivalent to the Fréchet-Hoeffding theorem. Our main contribution in this chapter is the extension of the EJD method from the bivariate setting to the general d-dimensional setting for any $d \geq 3$, which improves on many of the previous methods and alleviates some of their shortcomings. In particular, our approach is able to:

1. Attain *all* admissible correlations corresponding to a set of marginal distributions
2. Efficiently compute extreme measures given marginal distributions
3. Compute d-dimensional extreme distributions for $d \leq 51$, where the limitation on d is due to memory [86]
4. Efficiently sample multivariate distributions constructed using the EJD approach

The closest body of work in the literature to ours is that of Tiit [122, 123, 124] who also considers the idea of extremal joint distributions and their mixtures. However, we differ from Tiit in that we provide an algorithm that constructs an extreme joint distribution directly without resorting to the construction of conditional distributions and we provide a simple, more natural concept for keeping track of the possible extremal dependencies.

Although the main impetus for the development of the EJD method stemmed from the need to construct multivariate Poisson distributions, the EJD method is very general and applies to any discrete marginal distribution with finite variance resulting in a general discrete multivariate distribution with some extremal dependence structure between its components. A minor but important assumption we make throughout this chapter is that the marginal distributions $Q^{(j)}$ have $Q_k^{(j)} > 0$ for $k = 0, 1, \ldots, n$

in the finite-dimensional case and $k = 0, 1, \ldots$ in the infinite-dimensional case; the extension to more general marginal distributions that may have $Q_k^{(j)} = 0$, for some k, is straightforward, but makes the statement and proof of our results a little messy.

2.0.1 Outline

The outline for this chapter is as follows. We begin, in Section 2.2, by discussing the problem setting and introducing the EJD method in two-dimensions and the extreme measures it generates; much of the intuition in the bivariate setting also holds true in the general setting. Section 2.3 describes in detail the EJD algorithm, an efficient algorithm that simultaneously computes an extreme measure and its support in the two-dimensional case. In preparation for extending the EJD method to the general d-dimensional case, Section 2.4.1 introduces the notion of monotonicity structures which describe the associated extreme measures and provide an ordering for them. Extreme measures in d-dimensions are discussed in Section 2.4. We extend the EJD algorithm to the general case in Section 2.5. This is followed by a numerical example in Section 2.7. Calibration of multivariate extreme distributions is discussed in Section 2.8. Finally, sampling from extreme measures and multivariate extreme distributions constructed from extreme measures is discussed in Section 2.9.

2.1 Notation

We make use of the following notation in this chapter.

Symbol	Definition
d	Dimensionality of a multivariate distribution
$n = 2^{d-1}$	Number of extreme measures and correlation matrices for a d-dimensional distribution
$m = d(d-1)/2$	Number of entries in the strictly upper triangular part of a $d \times d$ correlation matrix
$k \in \{1.\ldots, d\}$	Index of the dimension of the problem
$j \in \{1, \ldots, n\}$	Index of the extreme points of the problem
P	General probability measure
$P_{i,j}$	$(i,j)^{\text{th}}$ element of P in the two-dimensional case
$\mathbb{N}_0$	Set of non-negative integers {0,1,2,...}
X	General discrete valued random variable
$h(P) := \mathbb{E}[X^{(1)}X^{(2)}]$	Joint expectation of the discrete bivariate probability measure P in the two-dimensional case
$\mathcal{I}_k = \{j : 1 \le j \le d, j \ne k\}$	Set of indices not equal to index k in the d-dimensional case
$\mathcal{I}_{k,l} = \{j : 1 \le j \le d, j \ne k, j \ne l\}$	Set of indices not equal to index k or index l in the d-dimensional case
$\mathcal{R}_+$	The set $\{(x,y) \in \mathbb{R}^2 : x \cdot y \ge 0\}$
$\mathcal{R}_-$	The set $\{(x,y) \in \mathbb{R}^2 : x \cdot y \le 0\}$
$Q_i^{(k)}$	i^{th} element of the k^{th} marginal probability distribution $Q^{(k)}$
$F^{(k)}$	Cumulative distribution function associated with $Q^{(k)}$
$\bar{F}_i^{(k)} = 1 - F_i^{(k)}$	i^{th} element of $\bar{F}^{(k)}$, the reversed cdf associated with $F^{(k)}$, in the two-dimensional case where $k \in \{1,2\}$
$l_{\max}$	The final iteration of an algorithm
$l \in \{0, 1, \ldots, l_{\max}\}$	Iteration index of an algorithm
$\hat{P}^{(j)}$	Extreme measure in the two-dimensional case where $j = \{1,2\}$
$\hat{P}_{s_l}^{(j)} = \hat{P}^{(j)}_{s_l^{(1)}, s_l^{(2)}}$	Probability corresponding to extreme measure j at the $(l+1)^{\text{st}}$ point of the $\mathcal{S}$-path where $j \in \{1,2\}$ in the two-dimensional case
$s_l = (s_l^{(1)}, s_l^{(2)})$	$(l+1)^{\text{st}}$ point belonging to the $\mathcal{S}$-path corresponding to $\hat{P}^{(1)}$ in the two-dimensional case
$\bar{s}_l = (\bar{s}_l^{(1)}, \bar{s}_l^{(2)})$	$(l+1)^{\text{st}}$ point belonging to the $\mathcal{S}$-path corresponding to $\hat{P}^{(2)}$ in the two-dimensional case
$i_{\max}^{(k)}$	Smallest integer such that $Q_i^{(k)} = 0$ for all $i > i_{\max}^{(k)}$ in the two-dimensional case
$\tilde{Q}^{(k)}$	Finite dimensional approximation of $Q^{(k)}$
$\breve{\mathbf{Q}}^{(2)}$	$\{Q_{j_{\max}}^{(2)}, Q_{j_{\max}-1}^{(2)}, \ldots, Q_0^{(2)}\}$ in the two-dimensional case
$\breve{\mathbf{F}}^{(2)}$	Vector containing cumulative sum of $\breve{\mathbf{Q}}^{(2)}$ in the two-dimensional case

Symbol	Definition
$H^{(1)}_{i,j} = \mathbb{P}(X^{(1)} \leq i, X^{(2)} \leq j)$	Bivariate cdf for the general random variables $X^{(1)}$ and $X^{(2)}$ in the two-dimensional case
$\Pi_{X^{(k)}} = \{F^{(k)}_0, \cdots, F^{(k)}_{i^{(k)}_{\max}}\}$	Partition of the unit interval by the k^{th} marginal cdf values
$\Pi_Z = \{z_0, \ldots, z_{l_{\max}}\}$	Partition of the unit interval by the unique and ordered values of the marginal cdfs $F^{(1)}$ and $F^{(2)}$ in the two-dimensional case
$\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$	Probability distribution computed in the comonotone case of Algorithm 1 using $\breve{\mathbf{F}}^{(2)}$ instead of $F^{(2)}$ in the two-dimensional case
$\hat{P}^{(1)}_{s_l \mid \breve{\mathbf{F}}^{(2)}}$	Value of $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$ at the support point s_l in the two-dimensional case
$(X^{(1)}, \ldots, X^{(d)})$	Random vector with associated distributions $(Q^{(1)}, \ldots, Q^{(d)})$ in the d-dimensional case
$P^{(d)}$	General probability measure in d-dimensions
$P^{(d;u,v)}$	Projection of $P^{(d)}$ onto the u^{th} and v^{th} coordinates
$P^{(d;v)}$	Projection of $P^{(d)}$ onto the v^{th} coordinate
$\hat{P}^{(j,d)}$	d-dimensional extreme measure with monotone structure $\mathbf{e}^{(j,d)}$
$\hat{P}^{(j,d)}_{i_1,\ldots,i_d}$	$(i_1, \ldots, i_d)^{\text{th}}$ element of $\hat{P}^{(j,d)}$
$Q^{(d;u)}$	u^{th} marginal distribution in the d-dimensional case
$F^{(d;u)}$	u^{th} marginal cdf in the d-dimensional case
$F^{(d;u)}_{s^{(j,d;u)}_{l^{(j,d)}}}$	Value of $F^{(d;u)}$ at $s^{(j,d;u)}_{l^{(j,d)}}$
$h^{(j,d)}_{u,v}(P^{(d)}) := \mathbb{E}[X^{(u)}X^{(v)}]$	Joint expectation of the u^{th} and v^{th} coordinate of $P^{(d)}$
$\mathbf{e}^{(j,d)} = (e^{(j,d)}_1, \ldots, e^{(j,d)}_d)$	j^{th} monotone structure of a d-dimensional extreme measure
$\mathbf{e}^{(j,d;u,v)} = (e^{(j,d)}_u, e^{(j,d)}_v)$	Projection of $\mathbf{e}^{(j,d)}$ onto the u^{th} and v^{th} coordinates
$l^{(j,d)}_{\max}$	The final iteration of an algorithm in the d-dimensional case
$l^{(j,d)} \in \{0, 1, \ldots, l^{(j,d)}_{\max}\}$	Iteration index of an algorithm in the d-dimensional case
$\mathbf{s}^{(j,d)} = (s^{(j,d)}_0, \ldots, s^{(j,d)}_{l^{(j,d)}_{\max}})$	d-dimensional support corresponding to $\hat{P}^{(j,d)}$
$s^{(j,d)}_{l^{(j,d)}} = (s^{(j,d;1)}_{l^{(j,d)}}, \ldots, s^{(j,d;d)}_{l^{(j,d)}})$	$(l^{(j,d)}+1)^{\text{st}}$ d-dimensional support point corresponding to $\hat{P}^{(j,d)}$
$\hat{P}^{(j,d)}_{l^{(j,d)}} = \hat{P}^{(j,d)}_{s^{(j,d)}_{l^{(j,d)}}}$	$(l^{(j,d)}+1)^{\text{st}}$ element of $\hat{P}^{(j,d)}$
$\Pi_{Z^{(j,d)}} = \{z^{(j,d)}_0, \ldots, z^{(j,d)}_{l^{(j,d)}_{\max}}\}$	Partition of the unit interval by the unique and ordered values of the marginal cdfs $F^{(j,1)}, \ldots, F^{(j,d)}$ in the d-dimensional case
$\hat{P}^{(j,d;u,v)}$	Projection of $\hat{P}^{(j,d)}$ onto the u^{th} and v^{th} coordinates
$\mathbf{s}^{(j,d;u,v)} = (s^{(j,d;u,v)}_0, \ldots, s^{(j,d;u,v)}_{l^{(j,d)}_{\max}})$	two-dimensional support corresponding to $\hat{P}^{(j,d;u,v)}$ possibly having duplicates in the set
$Q^{(j,d;k)}$	k^{th} marginal distribution in the d-dimensional case taking into account its corresponding monotone structure $\mathbf{e}^{(j,d)}$
$F^{(j,d;k)}$	k^{th} marginal cdf in the d-dimensional case taking into account its corresponding monotone structure $\mathbf{e}^{(j,d)}$
$\tilde{P}^{(j,d)}$	Probability measure computed by Algorithm 5
$\hat{\mathbf{s}}^{(j',2)} = \{\hat{s}^{(j',2)}_0, \ldots, \hat{s}^{(j',2)}_{l^{(j,d)}_{\max}}\}$	Set constructed by duplicating points in $\{s^{(j',2)}_0, \ldots, s^{(j',2)}_{l^2_{\max}}\}$ according to Algorithm 25

Symbol	Definition
$\hat{\mathbf{z}}^{(j,2)} = \{\hat{z}_0^{(j,2)}, \ldots, \hat{z}_{l_{\max}^{(j,d)}}^{(j,2)}\}$	Set constructed by duplicating points in $\{z_0^{(j,2)}, \ldots, z_{l_{\max}^2}^{(j,2)}\}$ according to Algorithm 25
$\{\hat{z}_0^{(j,d;u,v)}, \ldots, \hat{z}_{l_{\max}^{(j,d)}}^{(j,d;u,v)}\}$	Set constructed by Algorithm 26 containing only values from $F^{(j,d;u)}$ and $F^{(j,d;v)}$
$\{z_0^{(j,2;u,v)}, \ldots, z_{l_{\max}^{(j,d;u,v)}}^{(j,2;u,v)}\}$	Set constructed by taking only the unique elements of $\{\hat{z}_0^{(j,d;u,v)}, \ldots, \hat{z}_{l_{\max}^{(j,d)}}^{(j,d;u,v)}\}$
$\underline{\mathbf{s}}^{(j,d;u,v)} = (\underline{s}_0^{(j,d;u,v)}, \ldots, \underline{s}_{l_{\max}^{(j,d;u,v)}}^{(j,d;u,v)})$	Projection of $s_l^{(j,d)}$ onto the u^{th} and v^{th} coordinates with duplicates eliminated
$\tilde{\mathbf{s}}^{(j,d;u,v)} = (\tilde{s}_0^{(j,d;u,v)}, \ldots, \tilde{s}_{l_{\max}^{(j,d;u,v)}}^{(j,d;u,v)})$	Support corresponding to $\tilde{P}^{(j,d;u,v)}$ before being post-processed by Algorithm 6
$\tilde{\underline{\mathbf{s}}}^{(j,d;u,v)} = (\tilde{\underline{s}}_0^{(j,d;u,v)}, \ldots, \tilde{\underline{s}}_{l_{\max}^{(j,d;u,v)}}^{(j,d;u,v)})$	Set $\{\tilde{s}_0^{(j,d;u,v)}, \ldots, \tilde{s}_{l_{\max}^{(j,d)}}^{(j,d;u,v)}\}$ with duplicates removed
$\mathcal{A} = \{p : Aw = b,\ w \geq 0,\ p = Pw\}$	Set of discrete probability measures having correlation vector b that are a convex combination of extreme measures P
C	General correlation matrix
$\hat{C}^{(1)}$	Extremal positive correlation (scalar) in the two-dimensional case
$\hat{C}^{(2)}$	Extremal negative correlation (scalar) in the two-dimensional case
$\hat{C}^{(j,d)}$	j-th extreme correlation matrix corresponding to $\hat{P}^{(j,d)}$ in the d-dimensional case
A_k	Column vector containing the m entries of the upper triangular part of $\hat{C}^{(j)}$ plus an additional element equal to 1 at the bottom of A_k
A	Matrix of extreme correlation vectors having columns A_k
b	Correlation vector containing the m entries of the upper triangular part of C plus an additional element equal to 1 at the bottom of b
w_k	Weights of a convex combination, that must be solved for in the calibration problem (2.106)

2.2 Extreme Measures in two-dimensions

In this section we define extreme measures in two-dimensions and describe how they can be obtained using the Extreme Joint Distribution (EJD) approach. We introduce two special classes of discrete, bivariate distributions known as comonotone and antimonotone[4] distributions and show that extreme measures belong to these classes. The bivariate case is illustrative; much of the intuition also holds for the general d-dimensional case (Section 2.4). Importantly, the EJD theorem in two-dimensions has been shown to be equivalent to the Fréchet-Hoeffding theorem, a well known result. We note that the results in the subsequent section (Section 2.2.1) are primarily for proving the correctness of the EJD algorithm in Section 2.3.

In this section, we consider bivariate random vectors $(X^{(1)}, X^{(2)})$ on the positive quadrant having associated discrete probability distributions $Q_i^{(1)}$ and $Q_j^{(2)}$ with finite variance supported on $\mathbb{N}_0^2$, where $i = 0, 1, \ldots$ and $j = 0, 1, \ldots$ and $\mathbb{N}_0^2$ is the set of non-negative integers on the bivariate lattice.

We begin by defining extreme measures.

Definition 14 (Extreme Measures in two-dimensions)**.** *Extreme Measures are solutions to the following infinite-dimensional Linear Program (LP)*

$$\text{extremize} \quad h(P) \tag{2.3a}$$

$$\textit{subject to} \quad \sum_{j=0}^{\infty} P_{i,j} = Q_i^{(1)}, \quad i = 0, 1, \ldots \tag{2.3b}$$

$$\sum_{i=0}^{\infty} P_{i,j} = Q_j^{(2)}, \quad j = 0, 1, \ldots \tag{2.3c}$$

$$P_{i,j} \geq 0 \quad i, j = 0, 1, \ldots \tag{2.3d}$$

where $Q_i^{(1)} \geq 0$ *for* $i = 0, 1, 2, \ldots$, $Q_j^{(2)} \geq 0$ *for* $j = 0, 1, 2, \ldots$, *and* $\sum_{i=0}^{\infty} Q_i^{(1)} = \sum_{j=0}^{\infty} Q_j^{(2)} = 1$. *Extremize* denotes either max or min and the objective function is

$$h(P) := \mathbb{E}[X^{(1)} X^{(2)}] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} ij \, P_{i,j} \tag{2.3e}$$

where $P_{i,j} = \mathbb{P}(X^{(1)} = i, X^{(2)} = j)$

Remark 5. *Definition 14 is consistent with finite-dimensional* $Q^{(1)}$ *and* $Q^{(2)}$ *having elements* $Q_0^{(1)}, \ldots, Q_{i_{\max}}^{(1)}$ *and* $Q_0^{(2)}, \ldots, Q_{j_{\max}}^{(2)}$, *since they can be simply extended to the infinite-dimensional case by setting* $Q_i^{(1)} = 0$ *for* $i = i_{\max} + 1, i_{\max} + 2, \ldots$ *and* $Q_j^{(2)} = 0$ *for* $j = j_{\max} + 1, j_{\max} + 2, \ldots$, *respectively.*

Remark 6. *The optimization problem (2.3) is a semi-infinite linear program for which existence results for a solution are well known [114].*

Extreme measures are probability distributions that either maximize or minimize the joint expectation (2.3e), which is the only quantity within the calculation of the Pearson correlation coefficient that is not determined by the given marginal distributions. Therefore, the same distribution P that maximizes (minimizes) $h(P)$ also maximizes (minimizes) the correlation coefficient associated with P.

[4] Antimonotonicity is also known—more widely, in fact—as countermonotonicity in the literature. However, in this thesis we will stick with the use of countermonotonicity.

While extreme measures can be obtained by directly solving the LP (2.3), this approach is inefficient compared to the EJD algorithm, as explained in Section 2.3.

Moreover, it follows from the discussion in Section 2.4 that the higher-dimensional analogue of (2.3) precludes solving the optimization problem with standard software due to the complexity of converting the problem to a form that modern solvers would accept. We mention that the LP (2.3) is also a Monge-Kantorovich transportation problem (MKP) and thus the EJD method is a novel approach to solving a special class of MKP; extreme measures in two-dimensions are *optimal couplings* [57]. The connections to the MKP are not the focus of the thesis and will not be explored further.

2.2.1 Monotone Sets and Distributions

In this subsection, we introduce a special class of discrete bivariate distributions with given marginals known as monotone distributions [49, 112] and explore some of their properties. In the next subsection, we show that the distributions that maximize and minimize the $h(P)$ given in (2.3e) are monotone distributions. These bivariate discrete distributions P take the form

$$P_{i,j} \quad \textit{for} \quad i = 0, 1, 2, \dots \quad \text{and} \quad j = 0, 1, 2, \dots$$

and satisfy the marginal equations (2.3b) and (2.3c)

To ease the burden in proving the results in the subsection, we make the following assumption throughout this subsection.

Assumption 1. *The distribution P has marginals $Q^{(1)}$ and $Q^{(2)}$, discrete distributions with finite variance, satisfying the equations (2.3b) and (2.3c) and $Q_i^{(k)} > 0$ for $i = 0, 1, 2, \dots$ and $k \in \{1, 2\}$.*

Remark 7. *The results in this subsection also hold for the assumption that the distribution P has marginals $Q^{(1)}$ and $Q^{(2)}$, discrete distributions with finite variance, satisfying the equations (2.3b) and (2.3c). Moreover, $Q_i^{(1)} > 0$ for $i = 0, 1, 2, \dots, i_{\max}$ where $i_{\max}$ is some finite integer, $Q_i^{(1)} = 0$ for all $i > i_{\max}$. Similarly for $Q_j^{(2)}$ and $j_{\max}$.*

That is, the results in this subsection hold true for the case that the marginals $Q^{(1)}$ and $Q^{(2)}$ are finite-dimensional. This can be seen from the fact that the infinite sums of the bivariate discrete distributions $P_{i,j}$ in this subsection can instead be replaced by finite sums where the upper bound of summation is $i_{\max}$ if i is the index of summation; and similarly $j_{\max}$ when j is the index of summation. Note that $i_{\max}$ and $j_{\max}$ are the greatest integers such that $Q_{i_{\max}}^{(1)} > 0$ and $Q_{j_{\max}}^{(2)} > 0$, respectively.

Remark 8. *We believe the results in this subsection can be extended to the case where the distribution $\hat{P}$ has marginals $Q^{(1)}$ and $Q^{(2)}$, discrete distributions with finite variance, satisfying the equations (2.3b) and (2.3c) and $Q_i^{(k)} > 0$ for $i \in \mathcal{I}^{(k)} \subset \mathbb{N}_0$ and $Q_i^{(k)} = 0$ for $i \in \mathbb{N}_0 \setminus \mathcal{I}^{(k)}$ and $k \in \{1, 2\}$. However, note that this extensions is not needed in this thesis, since the Poisson and mixed Poisson cases we focus on in Chapters 3, 4, and 5 have $Q_i^{(1)} > 0$ for $i = 0, 1, 2, \dots$ and $Q_j^{(2)} > 0$ for $j = 0, 1, 2, \dots$. Moreover, the extension would complicate the proofs in this subsection significantly and detract from the main ideas.*

Definition 15 (Monotone Sets)**.** *A set of points $\mathcal{S} = \{s_n : n \in \mathbb{N}_0\}$, where $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ and $s_n = (x_n, y_n) \in \mathbb{R}^2$, is called comonotone if $\forall i, j \in \mathbb{N}_0$, the vector $s_j - s_i \in \mathcal{R}_+$ where $\mathcal{R}_+ = \{(x, y) \in \mathbb{R}^2 : x \cdot y \geq 0\}$. A set $\mathcal{S}$ is called antimonotone, if $\forall i, j \in \mathbb{N}_0$, the vector $s_j - s_i \in \mathcal{R}_-$, where $\mathcal{R}_- = \{(x, y) \in \mathbb{R}^2 : x \cdot y \leq 0\}$.*

Definition 16 (Monotone Distributions)**.** *A distribution P is said to be comonotone (antimonotone[5]) if its support,* $\text{supp}(P)$*, is a comonotone (antimonotone) set.*

Comonotone Distributions

Lemma 1. *Assume the distribution P is comonotonic and satisfies Assumption 1. Then the origin is in the support of P.*

PROOF: We show, by contradiction, that $(0,0)$ is in the support of P. To this end, suppose $(0,0)$ is not in the support of the comonotonic distribution P. Then, there must be a point $(0,n)$ with $n>0$ in the support of P, since the marginal assumption

$$\sum_{j=0}^{\infty} P_{0j} = Q_0^{(1)} > 0$$

implies that at least one $P_{0n}>0$ and the assumption that $(0,0)$ is not in the support of P (i.e., $P_{00}=0$) ensures that $n>0$. Similarly, there must be a point $(m,0)$ with $m>0$ is in the support of P. However, the point

$$s = (0,n) - (m,0) \notin \mathcal{R}_+.$$

Therefore, P is not a comonotone distribution, contradicting our assumption that P is a comonotone distribution. Hence, $(0,0)$ must be in the support of P. □

Remark 9. *In fact, a comonotonic distribution in higher dimensions must include the origin. The extension of Lemma 1 to the multivariate setting is straightforward.*

Lemma 2. *Assume the distribution P is comonotonic and satisfies Assumption 1. If (m,n) is in the support of P, then at least one of $(m,n+1),(m+1,n),(m+1,n+1)$ is also in the support of P. In addition, $(m,n+1)$ and $(m+1,n)$ cannot both be in the support of P.*

PROOF: We begin by showing, using proof by contradiction, the first part of the lemma, namely that, if (m,n) is in the support of P, then at least one of $(m,n+1),(m+1,n),(m+1,n+1)$ must also be in the support of P. To that end, suppose that none of $(m,n+1),(m+1,n),(m+1,n+1)$ are in the support of P. We first show that there must be a point $(m+1,n')$ where $n'>n+1$ in the support of P. To see this, note that the marginal assumption

$$\sum_{j=0}^{\infty} P_{m+1,j} = Q_{m+1}^{(1)} > 0$$

implies that at least one $P_{m+1,n'}>0$. The assumption that both $(m+1,n)$ and $(m+1,n+1)$ are not in the support of P implies that either $n'>n+1$ or $n'<n$. The latter cannot occur since, if $(m+1,n')$ is in the support of P with $n'<n$, then the point $s=(m,n)-(m+1,n')\notin\mathcal{R}_+$, contradicting our assumption that P is a comonotone distribution.

In a similar manner, there must be a point $(m',n+1)$ where $m'>m+1$ in the support of P. However, the point

$$s = (m+1,n') - (m',n+1) \notin \mathcal{R}_+$$

[5]Also known as countermonotonicity in some of the literature

if $m' > m+1$ and $n' > n+1$, which again contradicts our assumption that P is comonotone. Therefore, at least one of $(m, n+1), (m+1, n), (m+1, n+1)$ must be in the support of P.

That $(m, n+1)$ and $(m+1, n)$ cannot both be in the support of P can be shown, by contradiction, in a similar manner. To this end, suppose that $(m, n+1)$ and $(m+1, n)$ are both in the support of P. The point

$$s = (m, n+1) - (m+1, n) \notin \mathcal{R}_+,$$

again contradicting our assumption that P is comonotone. Thus, $(m, n+1)$ and $(m+1, n)$ cannot both be in the support of P. □

Remark 10. *If the marginal distributions $Q^{(1)}$ and $Q^{(2)}$ are finite-dimensional, then there exist integers m and n such that $Q^{(1)}_{m'} = 0$ for any $m' > m$ and $Q^{(2)}_{n'} = 0$ for any $n' > n$. Therefore, no point of the form (m', n') for any $m' > m$ or any $n' > n$ belongs to the support of P.*

Lemma 3. *Assume the distribution P is comonotonic and satisfies Assumption 1. If (m, n) is in the support of P and $(m, n) \neq (0, 0)$, then at least one of the points $(m, n-1), (m-1, n), (m-1, n-1)$ is also in the support of P. In addition, $(m, n-1)$ and $(m-1, n)$ cannot both be in the support of P.*

Remark 11. *In the special case that $m = 0$ and $n > 0$ the points $(m-1, n) = (-1, n)$ and $(m-1, n-1) = (-1, n-1)$ are not in the domain of P and so cannot be support points of P. However, in this case, the point $(m, n-1) = (0, n-1)$ is always a support point of P. So, the conclusion of Lemma 3 remains true.*

Similarly, if $m > 0$ and $n = 0$, the points $(m, n-1) = (m, -1)$ and $(m-1, n-1) = (m-1, -1)$ are not in the domain of P and so cannot be support points of P. However, in this case, the point $(m-1, n) = (m-1, 0)$ is always a support point of P. Again, the conclusion of Lemma 3 remains true.

Proof: We show that the lemma holds true for the following three cases:

1. both $m > 0$ and $n > 0$
2. $m = 0$ and $n > 0$
3. $m > 0$ and $n = 0$.

Case 1: both $m > 0$ and $n > 0$.

We use proof by contradiction to prove the first part of this lemma. To that end, suppose that none of $(m, n-1), (m-1, n), (m-1, n-1)$ are in the support of P. We first show that this implies that there must be a point $(m-1, n')$ where $n' < n-1$ in the support of P. To this end, note that the marginal assumption

$$\sum_{j=0}^{\infty} P_{m-1,j} = Q^{(1)}_{m-1} > 0$$

implies that at least one $P_{m-1,n'} > 0$. The assumption that both $(m-1, n)$ and $(m-1, n-1)$ are not in the support of P ensures that either $n' > n$ or $n' < n-1$. The former cannot occur since, if $(m-1, n')$ is in the support of P and $n' > n$, then the point $s = (m, n) - (m-1, n') \notin \mathcal{R}_+$, which contradicts our assumption that P is a comonotone distribution. Similarly, there must be a point $(m', n-1)$ where $m' < m-1$ in the support of P. However, if $n' < n-1$ and $m' < m-1$, then the point

$$s = (m-1, n') - (m', n-1) \notin \mathcal{R}_+$$

contradicts, again, our assumption that P is a comonotone distribution. Therefore, if (m, n) is in the support of P, then at least one of the points $(m, n-1), (m-1, n), (m-1, n-1)$ is also in the support of P.

The final statement of the lemma can also be shown by contradiction. To this end, suppose both $(m-1, n)$ and $(m, n-1)$ are in the support of P. Then the point $s = (m-1, n) - (m, n-1) \notin \mathcal{R}_+$ contradicting our assumption that P is a comonotone distribution. Hence, $(m, n-1)$ and $(m-1, n)$ cannot both be in the support of P.

Case 2: $m = 0$ and $n > 0$.

If $(m, n) = (0, n)$ where $n > 0$, the points $(m-1, n) = (-1, n)$ and $(m-1, n-1) = (-1, n-1)$ do not belong to the domain of P and thus cannot be in its support. Thus, the second part of the lemma follows immediately. Hence, all that remains to show is that, if the point $(0, n)$ is in the support of P, then the point $(0, n-1)$ must also be in the support of P.

We prove this result by contradiction. To this end, suppose there exists an $n \geq 1$ such that $(0, n)$ is in the support of P, but $(0, n-1)$ is not in the support of P. There must be a point $(m', n-1)$ for $m' > 0$ in the support of P since the marginal assumption

$$\sum_{i=0}^{\infty} P_{i,n-1} = Q^{(2)}_{n-1} > 0$$

implies that at least one $P_{m',n-1} > 0$. That $(0, n-1)$ is not in the support of P ensures that $m' > 0$. However, the point $s = (0, n) - (m', n-1) \notin \mathcal{R}_+$ for $m' > 0$, contradicting our assumption that P is a comonotone distribution. Therefore, if $(0, n)$ is in the support of P, then $(0, n-1)$ must also be in the support of P.

Case 3: $m > 0$ and $n = 0$.

If $(m, n) = (m, 0)$ where $m > 0$, the points $(m, n-1) = (m, -1)$ and $(m-1, n-1) = (m-1, -1)$ do not belong to the domain of P and thus cannot be in its support. Thus, the second part of the lemma follows immediately. An argument similar to the one employed in Case 2 can be used to show that, if $(m, 0)$ is in the support of P, then $(m-1, 0)$ is in the support of P as well.

□

Lemma 1 tells us that comonotonic distributions P must start at the origin. Lemma 2 tells us that given a point (m, n) belonging to the support of a comonotonic distribution P, then at least one of the points $(m, n+1), (m+1, n), (m+1, n+1)$ is also in the support of P, but not both $(m, n+1)$ and $(m+1, n)$. Note that it is possible that if (m, n) is in the support, then $(m, n+1)$ and $(m+1, n+1)$ or $(m+1, n)$ and $(m+1, n+1)$ are both in the support of P. This allows us to introduce, in a well-defined manner, the "subsequent point" of a point (m, n) in the support of P.

Definition 17 (Subsequent and Antecedent Points)**.** *Given some i^{th} point $s_i = (m, n)$ in the support of a comonotonic distribution P, the subsequent point in the support of P, s_{i+1}, is defined to be one of the points $(m, n+1), (m+1, n), (m+1, n+1)$ in the support of P with the smallest Euclidean distance to s_i. Similarly, the antecedent point s_{i-1} of s_i is defined to be one of the points $\{(m-1, n), (m-1, n-1), (m, n-1)\}$ in the support of P with the smallest Euclidean distance to s_i.*

Remark 12. *The subsequent point is well-defined because Lemma 2 ensures that the points $(m, n+1)$ and $(m+1, n)$ cannot both be in the support of P. Hence, if either of these points is in the support of*

P, it is the subsequent point, since its Euclidean distance to the point (m,n) is less than the Euclidean distance of $(m+1,n+1)$ to (m,n). If neither $(m,n+1)$ nor $(m+1,n)$ is in the support of P, then Lemma 2 ensures us that $(m+1,n+1)$ is in the support of P and so it is the subsequent point.

Remark 13. *Note that in the case where the marginal distributions $Q^{(1)}$ and $Q^{(2)}$ are finite, then there exists a point (m,n) in the support of P for which there is no subsequent point; this point is known as the terminal point of the support.*

Now we have all the ingredients to characterize the support of a comonotonic distribution P.

Definition 18 ($\mathcal{S}$-path)**.** *The $\mathcal{S}$-path is a directed path of support points of P starting with $s_0 = (0,0)$. Moreover, if the point $s_i = (s_i^{(1)}, s_i^{(2)})$ is on the $\mathcal{S}$-path for some $i \geq 0$, then the next point, $s_{i+1} = (s_{i+1}^{(1)}, s_{i+1}^{(2)})$, on the path is the subsequent point of s_i.*

Lemma 4. *Suppose that the distribution P is comonotonic and satisfies Assumption 1. There is no point (m,n) belonging to the support of P that is not on the $\mathcal{S}$-path.*

PROOF: We prove this result by contradiction. To this end, suppose that there is a support point (m,n) of P that is not on the $\mathcal{S}$-path. Lemma 3 ensures that we can construct a path from (m,n) that must lead back to the origin. However, by Lemmas 1, 2 and the definition of the $\mathcal{S}$-path, the $\mathcal{S}$-path begins at the origin and must be unique. So, the support path starting from (m,n) and proceeding backwards must be part of the $\mathcal{S}$-path, contradicting our assumption that (m,n) is not on the $\mathcal{S}$-path. □

Remark 14. *The support of comonotone distributions, by Definition 17 and Lemma 4, can also be characterized by their monotonically increasing coordinates.*

Remark 15 (Sparsity of Comonotone Distributions)**.** *Comonotone $\mathcal{S}$-paths are directed paths that, by Lemmas 1-4, are sparse. Every point on the comonotone $\mathcal{S}$-path except the starting point (and the terminal point in the case of finite marginals) has an in-degree and out-degree of 1.*

Antimonotone Distributions

Lemma 5. *Assume the distribution P is antimonotonic and satisfies Assumption 1. Then, there exists integers $m \geq 0$ and $n \geq 0$ such that the points $(0,n)$ and $(m,0)$ are in the support of P.*

PROOF: From Assumption 1, we know

$$\sum_{j=0}^{\infty} P_{0,j} = Q_0^{(1)} > 0.$$

Therefore, there must be an integer $n \geq 0$ such that $P_{0,n} > 0$. Hence, $(0,n)$ is in the support of P. Similarly, there must be an integer $m \geq 0$ such that $(m,0)$ is in the support of P. □

Lemma 6. *Assume the distribution P is antimonotonic and satisfies Assumption 1. If (m,n) is in the support of P, then at least one of the points $(m+1,n),(m,n-1),(m+1,n-1)$ is also in the support of P. In addition, $(m+1,n)$ and $(m,n-1)$ cannot both be in the support of P.*

Remark 16. *Note that in the case $(m,n) = (m,0)$, the points $(m,n-1) = (m,-1)$ and $(m+1,n-1) = (m+1,-1)$ are not in the domain of P and thus cannot be in its support. However, in this case, the point $(m+1,n) = (m+1,0)$ is always in the support of P. So, the conclusion of Lemma 6 remains true.*

PROOF:

Case 1: $n > 0$. We use proof by contradiction to prove the first part of the result. To this end, suppose that none of the points $(m+1, n), (m, n-1), (m+1, n-1)$ are in the support of P. We first show that this implies that there must be a point $(m+1, n')$ where $n' < n-1$ in the support of P. To this end, note that since the marginal assumption

$$\sum_{j=0}^{\infty} P_{m+1,j} = Q_{m+1}^{(1)} > 0$$

implies that at least one $P_{m+1,n'} > 0$. The assumption that both $(m+1, n)$ and $(m+1, n-1)$ are not in the support of P ensures that either $n' > n$ or $n' < n-1$. The former cannot occur since the point $s = (m, n) - (m+1, n') \notin \mathcal{R}_-$ for any $n' > n$, contradicting our assumption that P is an antimonotone distribution. Similarly, there must be a point $(m', n-1)$ where $m' > m+1$ in the support of P. To see this, first note that the marginal assumption

$$\sum_{i=0}^{\infty} P_{i,n-1} = Q_{n-1}^{(2)} > 0$$

implies that at least one $P_{m',n-1} > 0$. The assumption that both $(m, n-1)$ and $(m+1, n-1)$ are not in the support of P ensures that either $m' < m$ or $m' > m+1$. Here, there are two cases to consider: $m = 0$ and $m > 0$. If $m > 0$, then $m' < m$ cannot occur since the point $s = (m, n) - (m', n-1) \notin \mathcal{R}_-$ for any $m' < m$, contradicting our assumption that P is an antimonotone distribution. In the case that $m = 0$, there also cannot be a point $(m', n-1)$ where $m' < 0$ belonging to the support since it is not even in the domain of P. Therefore $m' > m+1$. However, the point

$$s = (m+1, n') - (m', n-1) \notin \mathcal{R}_-$$

for $n' < n-1$ and $m' > m+1$, contradicting again our assumption that P is antimonotone. Therefore, at least one of the points $(m+1, n), (m, n-1), (m+1, n-1)$ is in the support of P.

Finally, we use proof by contradiction again to show that $(m+1, n)$ and $(m, n-1)$ cannot both be in the support of P. To this end, suppose $(m+1, n)$ and $(m, n-1)$ are both in the support of P. Then the point $s = (m+1, n) - (m, n-1) \notin \mathcal{R}_-$, contradicting again our assumption that P is antimonotone. Therefore, $(m+1, n)$ and $(m, n-1)$ cannot both be in the support of P.

Case 2: $n = 0$.

Then $(m, n) = (m, 0)$ and the points $(m, n-1) = (m, -1)$ and $(m+1, n-1) = (m+1, -1)$ do not belong in the domain of P and therefore cannot belong to the support of P. Therefore, the second part of the lemma follows immediately.

We prove the first part of the lemma by contradiction. To this end, suppose that the point $(m+1, n) = (m+1, 0)$ does not belong in the support of P. Then there must be a point $(m+1, n')$ where $n' > 0$ in the support of P since the marginal assumption

$$\sum_{j=0}^{\infty} P_{m+1,j} = Q_{m+1}^{(1)} > 0$$

implies that at least one $P_{m+1,n'} > 0$. The assumption that $(m+1, 0)$ is not in the support of P ensures that $n' > 0$. However, the point $s = (m, 0) - (m+1, n') \notin \mathcal{R}_-$ for $n' > 0$, contradicting our assumption that P is antimonotone. Therefore, the point $(m+1, n) = (m+1, 0)$ belongs to the support of P.

□

Lemma 7. *Assume the distribution P is antimonotonic and satisfies Assumption 1. If (m, n) is in the support of P, then at least one of the points $(m-1, n), (m-1, n+1), (m, n+1)$ is also in the support of P. In addition, $(m, n+1)$ and $(m-1, n)$ cannot both be in the support of P.*

The proof of Lemma 7 is similar to that of Lemma 6 and is omitted.

Remark 17. *Note that in the case $(m, n) = (0, n)$, the points $(m-1, n) = (-1, n)$ and $(m-1, n+1) = (-1, n+1)$ are not in the domain of P and thus cannot be in its support. However, in this case, the point $(m, n+1) = (0, n+1)$ is always in the support of P. Hence, the conclusion of Lemma 7 remains true.*

Remark 18. *Lemma 5 says that both $(m, 0)$ and $(0, n)$ must belong to the support of the antimonotone distribution P,* $\text{supp}(P)$. *Lemmas 6 and 7 tell us that if $(m, 0) \in \text{supp}(P)$ then $(m+1, 0) \in \text{supp}(P)$ and that if $(0, n) \in \text{supp}(P)$ then $(0, n+1) \in \text{supp}(P)$, respectively. If the marginal distributions $Q^{(1)}$ and $Q^{(2)}$ are finite, that is, there exists m' and n' such that $Q^{(1)}_{m''} = 0$ for all $m'' > m'$ and $Q^{(2)}_{n''} = 0$ for all $n'' > n'$, then the points $(m'', 0)$ and $(0, n'')$ for $m'' > m'$ and $n'' > n'$ cannot belong in* $\text{supp}(P)$ *in order for the marginal assumptions to be satisfied. However, note that the points $(m', 0)$ and $(0, n')$ belong to the support of P.*

We can now, in a manner similar to the comonotonic case, characterize the support of an antimonotonic distribution P.

Definition 19 (Subsequent and Antecedent Points)**.** *Given a point $s_i = (m, n)$ in the support of an antimonotonic distribution P, the subsequent point in the support of P, s_{i+1}, is defined to be one of the points $(m+1, n), (m, n-1), (m+1, n-1)$ in the support of P with the smallest Euclidean distance to s_i. Similarly, the antecedent point s_{i-1} of s_i is defined to be one of the points $(m-1, n), (m-1, n+1), (m, n+1)$ in the support of P with the smallest Euclidean distance to s_i.*

Remark 19. *As in the case of the comonotone distribution, both the subsequent point and the antecedent point for an antimonotone distribution are well-defined. The proof of this is very similar to the proof given for the comonotone case in Remark 12.*

Definition 20 ($\mathcal{S}$-path)**.** *Lemma 5 ensures that there is an integer $n \geq 0$ such the point $(0, n)$ is in the support of the antimonotonic distribution P. Let n' be the smallest integer such that $(0, n')$ is in the support of P. The $\mathcal{S}$-path is a directed path of support points of P containing the support point $(0, n')$. Moreover, if s_i is a point in the $\mathcal{S}$-path, then the subsequent point, s_{i+1}, of s_i is also a point in the $\mathcal{S}$-path. Similarly, if s_i is a point in the $\mathcal{S}$-path, then the antecedent point, s_{i-1}, of s_i is also a point in the $\mathcal{S}$-path.*

Remark 20. *Note that in the infinite-dimensional case, $Q^{(j)}_i > 0$ for all $i = 0, 1, 2, \ldots$ and $j \in \{1, 2\}$, there is no first point on the $\mathcal{S}$-path and we, therefore, cannot number the support points starting from 0. Note, however, that we can always re-number the support points.*

Lemma 8. *Suppose that P is an antimonotone distribution and satisfies Assumption 1. There is no point (m, n) belonging to the support of P that is not on the $\mathcal{S}$-path.*

PROOF: The proof is by contradiction again. To that end, suppose that there is a point (m, n) in the support of P that is not on the $\mathcal{S}$-path. Using antecedent points, we can recursively construct a unique path backwards from (m, n) as follows. The point $s_0 = (m, n)$ is on this path and, if a point s_i is on the path, then the antecedent point, s_{i-1} of s_i, is also on the path (Lemma 7). Note that this process specifies a unique path, since the antecedent point, s_{i-1} of s_i is unique (Definition 19).

There are two cases to consider:

1. The path starting from (m, n) eventually reaches a point $(0, n'')$ for some $n'' \geq 0$,
2. The path starting from (m, n) never reaches a point $(0, n'')$ for any $n'' \geq 0$.

Case 1: We can further break Case 1 into two subcases:

(1a) $n'' < n'$,

(1b) $n'' \geq n'$,

where n' is the smallest integer such that $(0, n')$ is on the $\mathcal{S}$-path (Lemma 5). Consider case (1a) first. Since $n'' < n'$, we can extend the path that we constructed backwards starting from (m, n) further backwards. In doing so, note that the antecedent point of $(0, n'')$ must be $(0, n'' + 1)$. Repeating this observation, we see that the path starting from (m, n) must include the points $(0, n''), (0, n'' + 1), (0, n'' + 2), \ldots$ Hence, the path starting at (m, n) and proceeding backwards using antecedent points at each step must eventually reach $(0, n')$. Therefore, the $\mathcal{S}$-path and the path starting at (m, n) and moving backwards using antecedent points at each step must have a point in common.

Next consider case (1b). Since $n'' \geq n'$ and the $\mathcal{S}$-path includes the points $(0, n'), (0, n' + 1), (0, n' + 2), ...$, the $\mathcal{S}$-path must also include the point $(0, n'')$ (Lemma 6). Again, we see that the $\mathcal{S}$-path and the path starting at (m, n) and moving backwards using antecedent points at each step must have a point in common.

So, we have shown that in case (1), the $\mathcal{S}$-path and the path starting at (m, n) and moving backwards have a point in common. Call this common point $(0, \hat{n})$. Now, if we start at $(0, \hat{n})$ and construct a path moving forwards using subsequent points, this path must include the path segment that we constructed starting at (m, n) and moving backwards using antecedent points, since (m', n') is the antecedent point of (m'', n'') if and only if (m'', n'') is the subsequent point of (m', n'). Similarly, the path starting at $(0, \hat{n})$ moving forwards using subsequent points must be a segment of the $\mathcal{S}$-path. However, this process of constructing a path by moving forwards using subsequent points specifies a unique path, since each subsequent point is unique. Hence, the path segment that we constructed starting at (m, n) and moving backwards using antecedent points at each step must be a segment of the $\mathcal{S}$-path. Therefore, (m, n) is on the $\mathcal{S}$-path, contradicting our original assumption that (m, n) is not on the $\mathcal{S}$-path. Therefore, we have shown in Case 1 that there is no point (m, n) belonging to the support of P that is not on the $\mathcal{S}$-path.

Case 2: Each time we move backwards on the path starting at (m, n) using antecedent points at each step, we must either

(a) decrease the first coordinate by 1, leaving the second coordinate unchanged,

(b) leave the first coordinate unchanged, increase the second coordinate by 1,

(c) decrease the first coordinate by 1 and increase the second coordinate by 1.

Note that we can execute (a) and (c) at most m-1 times, since, if we were to execute them m times we would reach a point $(0, n'')$ for some $n'' \geq n$, which is not permitted in case (2). Therefore, if we take at least $\max(m-1+n-n'+1, 0)$ steps on the path starting from (m, n) and, proceeding backwards using antecedent points at each step, we must reach a point (m'', n'') with $0 < m'' \leq m$ and $n'' > n'$, since our assumption in case (2) is that all the points on this path have $m'' > 0$. Now note that both $(0, n')$ and (m'', n'') are support points of P and that $(m'', n'') - (0, n') \notin \mathcal{R}_-$ contradicting our assumption that P is an antimonotone distribution. Therefore, we have shown in Case 2 that there is no point (m, n) belonging to the support of P that is not on the $\mathcal{S}$-path.

□

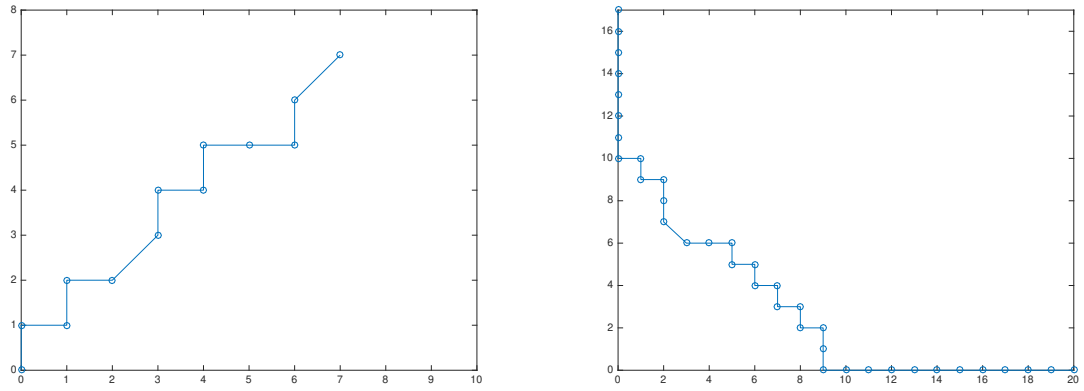

Figure 2.1: Support of a bivariate comonotone distribution (left) and a bivariate antimonotone distribution (right).

Figure 2.1 provides an illustration of the support of a bivariate comonotonic distribution and a bivariate antimonotonic distribution. The characterization of the properties of these supports are detailed in Lemmas 1-8.

Remark 21. *If P is an antimonotone distribution with finite-dimensional marginal distributions $Q^{(1)}$ and $Q^{(2)}$ that have support $\{0, 1, ..., i_{\max}\}$ and $\{0, 1, ..., j_{\max}\}$, respectively, then $(0, j_{\max})$ and $(i_{\max}, 0)$ are always supports points of P. Moreover, we can take $(0, j_{\max})$ to be the first point and $(i_{\max}, 0)$ to be the last point of the $\mathcal{S}$-path associated with P.*

Remark 22. *By Definition 19 and Lemma 8, a graph of the support points of an antimonotone distribution is monotonically decreasing (but not necessarily strictly monotonically decreasing).*

Remark 23 (Sparsity of Antimonotone Distributions)**.** *Antimonotone $\mathcal{S}$-paths are directed paths that, by Lemmas 5-8, are sparse. Every point on the antimonotone $\mathcal{S}$-path (except the points $(i_{\max}, 0)$ and $(0, j_{\max})$ in the case of finite marginals) has an in-degree and out-degree of 1.*

More importantly, the properties of the support of a monotone distribution extend to their random samples.

Lemma 9 (Monotonicity of the samples [74]). *Consider a finite random sample $Z = \{Z_n\}_{n=1}^N$ of independent two-dimensional vectors $Z_n = (X_n^{(1)}, X_n^{(2)})$ from a bivariate comonotone (antimonotone) distribution P. Then, Z is a comonotone (antimonotone) set. Conversely, if a random sample of N independent two-dimensional vectors, Z, is a comonotone (antimonotone) set for any integer $N \geq 2$, then Z is a sample from a comonotone (antimonotone) distribution, almost surely.*

PROOF: Consider a finite random sample, $Z = \{Z_n\}_{n=1}^N$, of independent two-dimensional vectors $Z_n = (X_n^{(1)}, X_n^{(2)})$ from a bivariate comonotone distribution P. Since P is a comonotone distribution, the support of P is a comonotone set and any two elements $s_k = (m_k, n_k)$ and $s_l = (m_l, n_l)$ of the support satisfy the property $s_k - s_l \in \mathcal{R}_+$. Now note that there exists a permutation π ordering the first component $X^{(1)}$ of the sample Z in a monotonically increasing order

$$X_{\pi(1)}^{(1)} \leq X_{\pi(2)}^{(1)} \leq \cdots \leq X_{\pi(N)}^{(1)}$$

where the subscript $\pi(n)$, for $n = 1, \ldots, N$, denotes the reordering of the samples under the permutation π. If the permutation π, applied to the second coordinate $X^{(2)}$, reorders it such that it is also non-decreasing, then we are done, since any two elements of the sample, $Z_k = (X_{\pi(k)}^{(1)}, X_{\pi(k)}^{(2)})$ and $Z_l = (X_{\pi(l)}^{(1)}, X_{\pi(l)}^{(2)})$, satisfy the property that $Z_k - Z_l \in \mathcal{R}_+$. Suppose, however, that there are some indices $i < j$ such that $X_{\pi(i)}^{(2)} > X_{\pi(j)}^{(2)}$. Then, the first coordinates of the sample corresponding to these indices must satisfy $X_{\pi(i)}^{(1)} = X_{\pi(j)}^{(1)}$.

To see why we must have $X_{\pi(i)}^{(1)} = X_{\pi(j)}^{(1)}$, suppose this were not the case. Then we must have $X_{\pi(i)}^{(1)} < X_{\pi(j)}^{(1)}$, since we chose π so that $X_{\pi(i)}^{(1)} \leq X_{\pi(j)}^{(1)}$. In addition, since $(X_{\pi(i)}^{(1)}, X_{\pi(i)}^{(2)})$ and $(X_{\pi(j)}^{(1)}, X_{\pi(j)}^{(2)})$ occur in the sample, we must also have $P_{X_{\pi(i)}^{(1)}, X_{\pi(i)}^{(2)}} > 0$ and $P_{X_{\pi(j)}^{(1)}, X_{\pi(j)}^{(2)}} > 0$. Thus, $(X_{\pi(i)}^{(1)}, X_{\pi(i)}^{(2)})$ and $(X_{\pi(j)}^{(1)}, X_{\pi(j)}^{(2)})$ are in the support of P. However, $(X_{\pi(i)}^{(1)}, X_{\pi(i)}^{(2)}) - (X_{\pi(j)}^{(1)}, X_{\pi(j)}^{(2)}) \notin R_+$, contradicting our assumption that P is comonotone. Therefore, as noted above, if $X_{\pi(i)}^{(2} > X_{\pi(j)}^{(2)}$, we must have $X_{\pi(i)}^{(1)} = X_{\pi(j)}^{(1)}$. Consequently, we introduce a new permutation $\hat{\pi}$ such that

$$\begin{aligned} \hat{\pi}(i) &= \pi(j), \\ \hat{\pi}(j) &= \pi(i), \\ \hat{\pi}(k) &= \pi(k) \quad \text{if} \quad k \neq i \quad \text{and} \quad k \neq j. \end{aligned}$$

Reordering $(X_{\hat{\pi}(i)}^{(1)}, X_{\hat{\pi}(i)}^{(2)})$ and $(X_{\hat{\pi}(j)}^{(1)}, X_{\hat{\pi}(j)}^{(2)})$ such that both coordinates are non-decreasing. This can be repeated for a finite number of times to obtain a permutation τ that reorders the second coordinate to ensure that it is non-decreasing while preserving the monotonicity of the first coordinate. Therefore, Z is a comonotone set since any two samples, $Z_{\tau(i)} = (X_{\tau(i)}^{(1)}, X_{\tau(i)}^{(2)})$ and $Z_{\tau(j)} = (X_{\tau(j)}^{(1)}, X_{\tau(j)}^{(2)})$, satisfy $Z_{\tau(i)} - Z_{\tau(j)} \in R_+$.

We use proof by contradiction to prove the converse statement. To this end, we assume that the support of the bivariate distribution is not comonotone. This implies that there exists two support points (m, n) and (p, q) such that $(m, n) - (p, q) \notin \mathcal{R}_+$. Since these two support points are associated with positive probabilities $P_{m,n} > 0$ and $P_{p,q} > 0$, there exists N sufficiently large such that $Z_k =$

$(X_k^{(1)}, X_k^{(2)}) = (m, n)$ and $Z_l = (X_l^{(1)}, X_l^{(2)}) = (p, q)$ is contained in an independent sampling of the distribution, almost surely. Then, Z is not a comonotone set since $Z_k - Z_l = (m, n) - (p, q) \notin \mathcal{R}_+$.

Similar arguments hold for the antimonotone case. □

2.2.2 Extreme Joint Distributions in two-dimensions

In the bivariate setting, there is a unique solution to the max (min) version of the extremal problem (2.3) and that solution is a comonotonic (antimonotonic) distribution. As explained earlier, the max (min) solution to problem (2.3) maximizes (minimizes) the correlation associated with the discrete distribution, P, over all choices of P that satisfy the marginal equations (2.3*b*) and (2.3*c*). We show in Section 2.4 that the results for bivariate distributions can be extended to multivariate distributions.

In this section, we show that solutions of the extremal problem (2.3) must be comonotone (antimonotone) in case of maximization (minimization). We introduce the semi-analytic equations that comprise the EJD theorem and show that any joint probability distribution that is a solution to the optimization problem (2.3) must be a comonotone or antimonotone distribution, respectively, in the maximization or minimization versions of (2.3). We take a probabilistic approach to proving the EJD theorem. Finally, we mention the equivalence of the EJD theorem to a famous result by Fréchet and Hoeffding [49, 64].

Lemma 10. *Suppose that the distribution P solves the optimization problem (2.3) for the case where extremize denotes* max*, then P is comonotone.*

PROOF: We prove this result by contradiction. To that end, suppose that P solves the optimization problem (2.3) for the case where extremize denotes max but P is not comonotone. Since P is not comonotone, there must exist a pair of points (i_1, j_1) and (i_2, j_2) belonging to the support of P for which $(i_1, j_1) - (i_2, j_2) \notin R_+$, whence $(i_1 - i_2)(j_1 - j_2) < 0$. So, we must have either

(a) $i_1 > i_2$ and $j_1 < j_2$, or

(b) $i_1 < i_2$ and $j_1 > j_2$.

In the proof below, we assume $i_1 > i_2$ and $j_1 < j_2$. The proof for $i_1 < i_2$ and $j_1 > j_2$ is similar; therefore, for the sake of brevity, we omit it. Hence, assuming $i_1 > i_2$ and $j_1 < j_2$, we can take $i_2 = i$ and $i_1 = i_2 + (i_1 - i_2) = i + m$, where $i = i_2 \geq 0$ and $m = i_1 - i_2 \geq 1$. Similarly, we can take $j_1 = j$ and $j_2 = j_1 + (j_2 - j_1) = j + n$, where $j = j_1 \geq 0$ and $n = j_2 - j_1 \geq 1$. Hence, we have a pair of points $(i_1, j_1) = (i + m, j)$ and $(i_2, j_2) = (i, j + n)$ in the support of P with $i, j \geq 0$ and $m, n \geq 1$. Since $(i + m, j)$ and $(i, j + n)$ are in the support of P, $P_{i+m,j} > 0$ and $P_{i,j+n} > 0$.

Let $t = \min(P_{i+m,j}, P_{i,j+n})$ and note that $t > 0$. Now consider $\hat{P}$ such that

$$\begin{aligned} \hat{P}_{i,j+n} &= P_{i,j+n} - t, \\ \hat{P}_{i,j} &= P_{i,j} + t, \\ \hat{P}_{i+m,j+n} &= P_{i+m,j+n} + t, \\ \hat{P}_{i+m,j} &= P_{i+m,j} - t, \end{aligned}$$

and $\hat{P}_{i,j} = P_{i,j}$ for all other points where $i, j \geq 0$. Note that $\hat{P}_{i,j} \geq 0$ for all $i, j = 0, 1, 2, \ldots$ and

$\sum_i^\infty \sum_j^\infty \hat{P}_{i,j} = \sum_i^\infty \sum_j^\infty P_{i,j} = 1$. Therefore, $\hat{P}$ is a probability distribution. Moreover,

$$\begin{aligned} h(\hat{P}) &= \sum_{ij} ij\hat{P}_{i,j} \\ &= \sum_{ij} ijP_{i,j} - t \cdot i(j+n) + t \cdot ij + t \cdot (i+m)(j+n) - t \cdot j(i+m) \\ &= \sum_{ij} ijP_{i,j} + tmn \\ &= h(P) + tmn, \end{aligned}$$

thereby contradicting the maximality of P since $tmn > 0$. Therefore, if P solves the optimization problem (2.3) for the case where extremize denotes max, then P must be comonotone. □

Lemma 11. *Suppose that a distribution P solves the optimization problem (2.3) for the case where extremize denotes* min*, then P is antimonotone.*

The proof is similar to Lemma 10 and is omitted.

Theorem 1 (EJD Theorem in two-dimensions [74])**.** *There exists a unique bivariate discrete distribution $\hat{P}^{(1)}$ that satisfies Assumption 1 and solves the problem (2.3) in the maximization case and it satisfies*

$$\hat{P}^{(1)}_{i,j} = [\min(F^{(1)}_i, F^{(2)}_j) - \max(F^{(1)}_{i-1}, F^{(2)}_{j-1})]^+ \quad i,j = 0,1,2,\ldots, \tag{2.4}$$

where $[\,x\,]^+ = \max(x,0)$ and $F^{(k)}$ denotes the marginal cdf corresponding to $Q^{(k)}$, with $F^{(k)}_{-1} = 0$, for $k \in \{1,2\}$.

There exists a unique bivariate discrete distribution $\hat{P}^{(2)}$ that solves the problem (2.3) in the minimization case and it satisfies

$$\hat{P}^{(2)}_{i,j} = [\min(F^{(1)}_i, \bar{F}^{(2)}_{j-1}) - \max(F^{(1)}_{i-1}, \bar{F}^{(2)}_j)]^+ \quad i,j = 0,1,2,\ldots, \tag{2.5}$$

where $\bar{F}^{(2)}_j = 1 - F^{(2)}_j$ and $\bar{F}^{(2)}_{-1} = 1$.

PROOF: Recall that, by Remark 6, a solution to (2.3) exists. Lemmas 12 and 13 below derive equations (2.4) and (2.5) from which uniqueness also follows. Thus, Theorem 1 follows immediately from Lemmas 12 and 13 below. □

The semi-analytic equations (2.4) and (2.5) take on a positive value only for points (i,j) belonging to the support of $\hat{P}^{(k)}, k \in \{1,2\}$; they take the value 0 otherwise. Thus, the EJD theorem provides a method for checking whether a point (i,j) in the domain of an extreme measure (extreme joint distribution) $\hat{P}^{(k)}$ belongs to its support. However, this is an inefficient use of Theorem 1 that is impractical in higher dimensions ($d \geq 3$) where the number of points to check increases exponentially with d.

Remark 24. *Note the generality of Theorem 1: we only assumed, from Assumption 1, that the marginal distributions have finite variance and satisfy $Q^{(k)}_i > 0$ for $i = 0,1,2,\ldots$ and $k \in \{1,2\}$*

Our approach to the proof of the EJD theorem is a probabilistic argument based on Borel's Law of Large Numbers (LLN) [40]. The probabilistic approach also forms the basis for the extension of the

EJD theorem to the general d-dimensional setting (Section 2.4). For the two-dimensional case that we consider in this subsection, consider a sample $\{(\hat{X}_n^{(1)}, \hat{X}_n^{(2)})\}_{n=1}^N$ from the extreme joint distribution $\hat{P}^{(1)}$ maximizing the correlation coefficient. Since $\hat{P}^{(1)}$ is a comonotone distribution, the samples $\{(\hat{X}_n^{(1)}, \hat{X}_n^{(2)})\}_{n=1}^N$ form a comonotonic set (Lemma 9) and, therefore, there exists a permutation π ordering the samples such that both coordinates $X_n^{(1)} = \hat{X}_{\pi(n)}^{(1)}$ and $X_n^{(2)} = \hat{X}_{\pi(n)}^{(2)}$ for $n = 1, 2, \ldots, N$ are monotonically increasing. As $N \to \infty$ we obtain two sequences of increasing length,

$$\begin{aligned} X^{(1)} &: \overbrace{0, 0, \ldots, 0}^{N_{X^{(1)}}(0)}, \overbrace{1, 1, \ldots, 1}^{N_{X^{(1)}}(1)}, \overbrace{2, 2, \ldots, 2}^{N_{X^{(1)}}(2)}, \ldots \overbrace{k, k \ldots, k}^{N_{X^{(1)}}(k)}, \ldots \\ X^{(2)} &: \underbrace{0, \ldots, 0}_{N_{X^{(2)}}(0)}, \underbrace{1, \ldots, 1}_{N_{X^{(2)}}(1)}, \underbrace{2, \ldots, 2}_{N_{X^{(2)}}(2)}, \ldots, \underbrace{k, k, k, \ldots, k}_{N_{X^{(2)}}(k)}, \ldots \end{aligned} \tag{2.6}$$

where $N_{X^{(1)}}(k)$ is the number of times the value k occurs in the sequence $\{X_n^{(1)}\}_{n=1}^N$ and $N_{X^{(2)}}(k)$ is the number of times the value k occurs in the sequence $\{X_n^{(2)}\}_{n=1}^N$. By Borel's LLN,

$$\begin{aligned} \lim_{N\to\infty} \frac{N_{X^{(1)}}(k)}{N} &= Q_k^{(1)} \quad k = 0, 1, 2, \ldots \quad \text{almost surely} \\ \lim_{N\to\infty} \frac{N_{X^{(2)}}(k)}{N} &= Q_k^{(2)} \quad k = 0, 1, 2, \ldots \quad \text{almost surely.} \end{aligned} \tag{2.7}$$

Denote

$$M_{X^{(1)}}(i) = \sum_{l=0}^{i} N_{X^{(1)}}(l) \quad i = 0, 1, 2, \ldots \quad \text{and} \quad M_{X^{(2)}}(j) = \sum_{l=0}^{j} N_{X^{(2)}}(l) \quad j = 0, 1, 2, \ldots \tag{2.8}$$

and let $M_{X^{(k)}}(-1) = 0$ for $k \in \{1, 2\}$.

Let us define the cdfs $F_i^{(1)} = \mathbb{P}(X^{(1)} \leq i)$ and $F_j^{(2)} = \mathbb{P}(X^{(2)} \leq j)$ corresponding to the marginal distributions $Q^{(1)}$ and $Q^{(2)}$, respectively, satisfying

$$F_i^{(1)} = \sum_{l=0}^{i} Q_l^{(1)}, \quad F_j^{(2)} = \sum_{l=0}^{j} Q_l^{(2)} \quad i, j = 0, 1, 2, \ldots . \tag{2.9}$$

In addition, set $F_{-1}^{(k)} = 0$ for $k \in \{1, 2\}$. The following two lemmas prove (2.4) and (2.5).

Lemma 12 ([74])**.** *Let N_{ij} denote the number of times the pair of numbers (i, j) occurs in the sequence $\{(X_n^{(1)}, X_n^{(2)})\}_{n=1}^N$. The limits*

$$\hat{P}_{i,j}^{(1)} = \lim_{N\to\infty} \frac{N_{ij}}{N} \tag{2.10}$$

exist almost surely for all $i, j = 0, 1, 2 \ldots$, as $N \to \infty$. Moreover, $\hat{P}_{i,j}^{(1)}$ satisfies (2.4).

Remark 25. *Note that the while the limit (2.10) exists almost surely, $\hat{P}_{i,j}^{(1)}$ satisfies (2.4) without the almost surely qualifier, as is explained in the proof below.*

PROOF: The first part of the Lemma follows immediately from Borel's Law of Large Numbers. Hence, all that remains is to show that $\hat{P}_{i,j}^{(1)}$ satisfies (2.4). To that end, consider a sample $\{(X_n^{(1)}, X_n^{(2)})\}_{n=1}^N$ from the extreme joint distribution $\hat{P}^{(1)}$ depicted in (2.6). We first show below that N_{ij} satisfies

$$N_{ij} = \left[\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1))\right]^+ \tag{2.11}$$

where $x^+ := \max(0, x)$.

To prove that (2.11) holds, we consider two cases:

(1) (i, j) does not occur in the sequence $\{(X_n^{(1)}, X_n^{(2)})\}_{n=1}^N$, and

(2) (i, j) does occur in the sequence $\{(X_n^{(1)}, X_n^{(2)})\}_{n=1}^N$.

To show that (2.11) holds in Case (1), we further divide Case (1) into three subcases:

(1a) i does not occur in the sequence $\{X_n^{(1)}\}_{n=1}^N$,

(1b) j does not occur in the sequence $\{X_n^{(2)}\}_{n=1}^N$, and

(1c) i occurs in the sequence $\{X_n^{(1)}\}_{n=1}^N$ and j occurs in the sequence $\{X_n^{(2)}\}_{n=1}^N$, but the pair (i, j) does not occur in the sequence $\{(X_n^{(1)}, X_n^{(2)})\}_{n=1}^N$.

Consider Case (1a) first. In this case, since i does not occur in the sequence $\{X_n^{(1)}\}_{n=1}^N$, the pair (i, j) cannot occur in the sequence $\{(X_n^{(1)}, X_n^{(2)})\}_{n=1}^N$. Therefore, $N_{ij} = 0$. So, we need to show that the right side of (2.11) is 0 too. To this end, note that, since i does not occur in the sequence $\{X_n^{(1)}\}_{n=1}^N$, we must have $N_{X^{(1)}}(i) = 0$. Hence, from (2.8), $M_{X^{(1)}}(i-1) = M_{X^{(1)}}(i)$. Therefore,

$$\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1)) \le M_{X^{(1)}}(i) - M_{X^{(1)}}(i-1) = 0,$$

whence

$$\left[\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1))\right]^+ = 0.$$

Hence, we have shown that (2.11) holds in Case (1a).

The proof that (2.11) holds in Case (1b) is very similar to the proof that (2.11) holds in Case (1a). So, for brevity, we omit it.

Now consider Case (1c). Since the pair (i, j) does not occur in the sequence $\{(X_n^{(1)}, X_n^{(2)})\}_{n=1}^N$, we again have that $N_{ij} = 0$. So, we need to show that the right side of (2.11) is 0 in Case (1c) too.

However, in Case (1c), i does occur in the sequence $\{X_n^{(1)}\}_{n=1}^N$. Moreover, since we ordered the $X_n^{(1)}$ from smallest to largest, the $X_n^{(1)}$ that are equal to i occur consecutively in the sequence $\{X_n^{(1)}\}_{n=1}^N$. Therefore, there exist $n_{\text{low}}^{(1)}$ and $n_{\text{high}}^{(1)}$ such that $X_n^{(1)} = i$ if and only if n satisfies $n_{\text{low}}^{(1)} \le n \le n_{\text{high}}^{(1)}$. Similarly, since j does occur in the sequence $\{X_n^{(2)}\}_{n=1}^N$ and the $X_n^{(2)}$ are also ordered from smallest to largest, there exist $n_{\text{low}}^{(2)}$ and $n_{\text{high}}^{(2)}$ such that $X_n^{(2)} = j$ if and only if n satisfies $n_{\text{low}}^{(2)} \le n \le n_{\text{high}}^{(2)}$. However, in Case (1c), the pair (i, j) does not occur in the sequence $\{(X_n^{(1)}, X_n^{(2)})\}_{n=1}^N$. Therefore, it must be that either

(1cα) all the $X_n^{(1)} = i$ occur before any of the $X_n^{(2)} = j$, or

(1cβ) all the $X_n^{(2)} = j$ occur before any of the $X_n^{(1)} = i$.

In subcase (1cα), note that, since all the $X_n^{(1)} = i$ occur before any of the $X_n^{(2)} = j$, we must have that $n_{\text{high}}^{(1)} < n_{\text{low}}^{(2)}$. Similarly, in subcase (1cβ), we must have $n_{\text{high}}^{(2)} < n_{\text{low}}^{(1)}$. Now note that, from (2.6) and

(2.8), it follows that

$$\begin{aligned} n^{(1)}_{\text{low}} &= M_{X^{(1)}}(i-1)+1, \\ n^{(1)}_{\text{high}} &= M_{X^{(1)}}(i), \\ n^{(2)}_{\text{low}} &= M_{X^{(2)}}(j-1)+1, \\ n^{(2)}_{\text{high}} &= M_{X^{(2)}}(j). \end{aligned} \tag{2.12}$$

So, in subcase (1cα), it follows from $n^{(1)}_{\text{high}} < n^{(2)}_{\text{low}}$ that $M_{X^{(1)}}(i) < M_{X^{(2)}}(j-1)+1$, whence $M_{X^{(1)}}(i) \le M_{X^{(2)}}(j-1)$. Therefore,

$$\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1)) \le M_{X^{(1)}}(i) - M_{X^{(2)}}(j-1) \le 0,$$

whence

$$\big[\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1))\big]^+ = 0.$$

That is, we have shown that, in subcase (1cα), the right side of (2.11) is 0 as required. Similarly, in subcase (1cβ), it follows from $n^{(2)}_{\text{high}} < n^{(1)}_{\text{low}}$ that $M_{X^{(2)}}(j) < M_{X^{(1)}}(i-1)+1$, whence $M_{X^{(2)}}(j) \le M_{X^{(1)}}(i-1)$. Therefore,

$$\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1)) \le M_{X^{(2)}}(j) - M_{X^{(1)}}(i-1) \le 0,$$

whence

$$\big[\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1))\big]^+ = 0.$$

That is, we have shown that, in subcase (1cβ), the right side of (2.11) is 0 as required. Hence, we have completed the proof that (2.11) holds in Case (1c).

Now consider Case (2): (i,j) does occur in the sequence $\{(X^{(1)}_n, X^{(2)}_n)\}_{n=1}^N$. We use the notation established in Case (1c) above to show that (2.11) holds in Case (2) too. To this end, note that, if $(X^{(1)}_n, X^{(2)}_n) = (i,j)$, then we must have $X^{(1)}_n = i$, whence $n^{(1)}_{\text{low}} \le n \le n^{(1)}_{\text{high}}$. Moreover, for the same n, we also have $X^{(2)}_n = j$, whence $n^{(2)}_{\text{low}} \le n \le n^{(2)}_{\text{high}}$. Therefore, $(X^{(1)}_n, X^{(2)}_n) = (i,j)$ if and only if $n \in [n^{(1)}_{\text{low}}, n^{(1)}_{\text{high}}] \cap [n^{(2)}_{\text{low}}, n^{(2)}_{\text{high}}]$. Moreover, since there is at least one n for which $(X^{(1)}_n, X^{(2)}_n) = (i,j)$, $[n^{(1)}_{\text{low}}, n^{(1)}_{\text{high}}] \cap [n^{(2)}_{\text{low}}, n^{(2)}_{\text{high}}] \ne \emptyset$.

If $n^{(1)}_{\text{low}} \le n^{(2)}_{\text{low}}$, then we must have $n^{(2)}_{\text{low}} \le n^{(1)}_{\text{high}}$, since otherwise we would have $n^{(1)}_{\text{low}} \le n^{(1)}_{\text{high}} < n^{(2)}_{\text{low}} \le n^{(2)}_{\text{high}}$, whence $[n^{(1)}_{\text{low}}, n^{(1)}_{\text{high}}] \cap [n^{(2)}_{\text{low}}, n^{(2)}_{\text{high}}] = \emptyset$, contradicting our result above that $[n^{(1)}_{\text{low}}, n^{(1)}_{\text{high}}] \cap [n^{(2)}_{\text{low}}, n^{(2)}_{\text{high}}] \ne \emptyset$. Therefore, in this subcase, we must have either

(2a) $n^{(1)}_{\text{low}} \le n^{(2)}_{\text{low}} \le n^{(2)}_{\text{high}} \le n^{(1)}_{\text{high}}$, or

(2b) $n^{(1)}_{\text{low}} \le n^{(2)}_{\text{low}} \le n^{(1)}_{\text{high}} \le n^{(2)}_{\text{high}}$.

On the other hand, if $n^{(2)}_{\text{low}} \le n^{(1)}_{\text{low}}$, then a similar argument shows that we must have either

(2c) $n^{(2)}_{\text{low}} \le n^{(1)}_{\text{low}} \le n^{(1)}_{\text{high}} \le n^{(2)}_{\text{high}}$, or

(2d) $n^{(2)}_{\text{low}} \le n^{(1)}_{\text{low}} \le n^{(2)}_{\text{high}} \le n^{(1)}_{\text{high}}$.

Consider Case (2a) first. Since

$$n^{(1)}_{\text{low}} \le n^{(2)}_{\text{low}} \le n^{(2)}_{\text{high}} \le n^{(1)}_{\text{high}},$$

it follows that

$$[n^{(1)}_{\text{low}}, n^{(1)}_{\text{high}}] \cap [n^{(2)}_{\text{low}}, n^{(2)}_{\text{high}}] = [n^{(2)}_{\text{low}}, n^{(2)}_{\text{high}}].$$

Hence, the number of values of n for which $(X^{(1)}_n, X^{(2)}_n) = (i, j)$ is

$$N_{i,j} = n^{(2)}_{\text{high}} - n^{(2)}_{\text{low}} + 1.$$

On the other hand, it follows in this case from (2.12) that the right side of (2.11) is

$$\begin{aligned}
&\big[\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1))\big]^+ \\
&\quad = \big[\min(n^{(1)}_{\text{high}}, n^{(2)}_{\text{high}}) - \max(n^{(1)}_{\text{low}} - 1, n^{(2)}_{\text{low}} - 1)\big]^+ \\
&\quad = \big[n^{(2)}_{\text{high}} - (n^{(2)}_{\text{low}} - 1)\big]^+ \\
&\quad = n^{(2)}_{\text{high}} - n^{(2)}_{\text{low}} + 1.
\end{aligned}$$

Therefore, (2.11) holds in this case.

A similar argument shows that, in Cases (2b)–(2d), (2.11) also holds. Therefore, we have shown that (2.11) holds in all cases.

To see that $\hat{P}_{i,j}$ satisfies (2.4), divide both sides of (2.11) by N to get

$$\begin{aligned}
\frac{N_{ij}}{N} &= \frac{1}{N}\big[\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1))\big]^+ \\
&= \left[\frac{1}{N}\big(\min(M_{X^{(1)}}(i), M_{X^{(2)}}(j)) - \max(M_{X^{(1)}}(i-1), M_{X^{(2)}}(j-1))\big)\right]^+ \\
&= \left[\min\left(\frac{M_{X^{(1)}}(i)}{N}, \frac{M_{X^{(2)}}(j)}{N}\right) - \max\left(\frac{M_{X^{(1)}}(i-1)}{N}, \frac{M_{X^{(2)}}(j-1)}{N}\right)\right]^+.
\end{aligned} \tag{2.13}$$

Now note that

$$\begin{aligned}
\lim_{N\to\infty} \frac{M_{X^{(1)}}(i)}{N} &= F^{(1)}_i \quad \text{almost surely for } i = 0, 1, 2, \ldots \\
\lim_{N\to\infty} \frac{M_{X^{(2)}}(j)}{N} &= F^{(2)}_j \quad \text{almost surely for } j = 0, 1, 2, \ldots .
\end{aligned} \tag{2.14}$$

Therefore, taking limits in (2.13), using (2.10) and (2.14) and noting that max, min and $[\cdot]^+$ are continuous functions, we get

$$\begin{aligned}
\hat{P}_{i,j} &= \lim_{N\to\infty} \frac{N_{ij}}{N} \\
&= \lim_{N\to\infty} \left[\min\left(\frac{M_{X^{(1)}}(i)}{N}, \frac{M_{X^{(2)}}(j)}{N}\right) - \max\left(\frac{M_{X^{(1)}}(i-1)}{N}, \frac{M_{X^{(2)}}(j-1)}{N}\right)\right]^+ \\
&= \left[\min\left(\lim_{N\to\infty} \frac{M_{X^{(1)}}(i)}{N}, \lim_{N\to\infty} \frac{M_{X^{(2)}}(j)}{N}\right) - \max\left(\lim_{N\to\infty} \frac{M_{X^{(1)}}(i-1)}{N}, \lim_{N\to\infty} \frac{M_{X^{(2)}}(j-1)}{N}\right)\right]^+ \\
&= \big[\min(F^{(1)}_i, F^{(2)}_j) - \max(F^{(1)}_{i-1}, F^{(2)}_{j-1})\big]^+,
\end{aligned} \tag{2.15}$$

almost surely.

Thus, we have shown that $\hat{P}_{i,j}$ satisfies (2.4) almost surely. Taking intersections of the sets of

probability-one events on which (2.10) and (2.15) holds gives a set of full-measure for which both limits (2.10) and (2.15) coincide. Since both the right side of (2.10) and the right side of (2.15) are constants (independent of ω), equality on a set of measure one implies the equality of the constants themselves. This implies that the deterministic identity (2.4) holds for all (i, j) and we can drop the "almost sure" qualifier. □

The case of minimal correlations is similar. Consider a sample $\{(\hat{X}_n^{(1)}, \hat{X}_n^{(2)})\}_{n=1}^N$ from the extreme joint distribution $\hat{P}^{(2)}$ minimizing the correlation coefficient. Since $\hat{P}^{(2)}$ is an antimonotone distribution, the samples $\{(\hat{X}_n^{(1)}, \hat{X}_n^{(2)})\}_{n=1}^N$ also form an antimonotone set (Lemmas 9) and therefore there exists a permutation π ordering the samples such that the first coordinate is monotonically increasing,

$$X_1 \leq X_2 \leq \cdots \leq X_N \quad \text{where} \quad X_n^{(1)} = \hat{X}_{\pi(n)}^{(1)},$$

and the second coordinate is monotonically decreasing,

$$X_1 \geq X_2 \geq \cdots \geq X_N \quad \text{where} \quad X_n^{(2)} = \hat{X}_{\pi(n)}^{(2)}.$$

Taking $N \to \infty$, we obtain two sequences

$$\begin{aligned} X^{(1)} &: \overbrace{0, 0, \ldots, 0}^{N_{X^{(1)}}(0)}, \overbrace{1, 1, \ldots, 1}^{N_{X^{(1)}}(1)}, \overbrace{2, 2, \ldots, 2}^{N_{X^{(1)}}(2)}, \ldots \overbrace{k, k \ldots, k}^{N_{X^{(1)}}(k)}, \ldots \\ X^{(2)} &: \ldots, \underbrace{k, k, k, \ldots, k}_{N_{X^{(2)}}(k)} \ldots \underbrace{k-1, \ldots, k-1}_{N_{X^{(2)}}(k-1)}, \ldots\ldots, \underbrace{0, \ldots, 0, 0}_{N_{X^{(2)}}(0)} \ldots . \end{aligned} \tag{2.16}$$

Lemma 13 ([74]). *Let N_{ij} denote the number of times the pair of numbers (i, j) occurs in the sequence $\{(X_n^{(1)}, X_n^{(2)})\}_{n=1}^N$. The limits*

$$\hat{P}_{i,j}^{(2)} = \lim_{N \to \infty} \frac{N_{ij}}{N} \tag{2.17}$$

exist almost surely for all $i, j = 0, 1, 2 \ldots$, as $N \to \infty$. Moreover, $\hat{P}_{i,j}^{(2)}$ satisfies (2.5).

Remark 26. *Note that the while the limit (2.17) exists almost surely, $\hat{P}_{i,j}^{(1)}$ satisfies (2.5) without the almost surely qualifier, as explained in the proof of (2.4) in Lemma 12.*

PROOF: The proof is similar to that of Lemma 12 and is omitted for brevity. □

2.2.3 Equivalence to the Fréchet-Hoeffding Theorem

In the discrete case, the Fréchet-Hoeffding theorem can be formulated as follows. Consider the space $\Pi(Q^{(1)}, Q^{(2)})$ of discrete bivariate distributions that satisfy the marginal equations (2.3b) and (2.3c), where $Q^{(1)}$ and $Q^{(2)}$ are the marginal distributions in (2.3b) and (2.3c), respectively.

Theorem 2 (Fréchet [49], Hoeffding [64]). *The bivariate cdf $H_{i,j}^{(1)} := \mathbb{P}(X^{(1)} \leq i, X^{(2)} \leq j)$ maximizing the correlation coefficient of $X^{(1)}$ and $X^{(2)}$ is*

$$H_{i,j}^{(1)} = \min(Q_i^{(1)}, Q_j^{(2)}) \qquad i, j = 0, 1, 2, \ldots . \tag{2.18}$$

Similarly, the bivariate cdf $H_{i,j}^{(2)}$ minimizing the correlation coefficient of $X^{(1)}$ and $X^{(2)}$ is

$$H_{i,j}^{(2)} = \max(0, Q_i^{(1)} + Q_j^{(2)} - 1) \qquad i, j = 0, 1, 2, \ldots . \tag{2.19}$$

The EJD theorem in two-dimensions (Theorem 1) was shown to be equivalent to the Fréchet-Hoeffding Theorem in Proposition 9.4 on page 216 of [74]. While the Fréchet-Hoeffding theorem is a two-dimensional result and does not appear to extend easily to the general d-dimensional setting, our EJD theorem, being equivalent in two-dimensions to Fréchet-Hoeffding, does.

2.3 The EJD Algorithm in two-dimensions

While the Extreme Joint Distribution (EJD) theorem (Theorem 1) provides semi-analytical forms—Equations (2.4) and (2.5)—to compute the probabilities of the extreme measure, $\hat{P}^{(j)}$, naive application of this formula to compute the supports of $\hat{P}^{(j)}$ is inefficient in the two-dimensional setting and computationally prohibitive in the general d-dimensional setting (Section 2.5).

An alternative approach to solving the optimization problem (2.3) is the EJD algorithm[6] listed on page 45, introduced in [74] for the two-dimensional case, an efficient algorithm that simultaneously computes each point belonging to the support of the extreme measure and its corresponding probability. The key realization is that, while it is difficult to construct a joint probability distribution such that the marginal distributions are satisfied, it is much easier to construct a joint cumulative distribution function that satisfies the marginal cumulative distribution functions. Moreover, Lemmas 10 and 11 show that any solution to the optimization problem (2.3) must be a comonotone or antimonotone distribution. Hence, the main idea of the algorithm is to exploit our knowledge of the properties of the supports of comonotone and antimonotone distributions and the fact that any solution $\hat{P}^{(j)}$ to (2.3) must satisfy the marginal constraints (2.3b) and (2.3c).

This is accomplished by exploiting our knowledge of the key properties of the $\mathcal{S}$-path that allows us to efficiently compute the $\mathcal{S}$-path and hence the support of $\hat{P}^{(j)}$, $j = 1, 2$. The marginal cdfs $F^{(1)}$ and $F^{(2)}$ play a large role in the determination of the probability $\hat{P}^{(j)}_{s_l}$ corresponding to the l^{th} point of the $\mathcal{S}$-path. The set $\Pi_Z = \Pi_{X^{(1)}} \vee \Pi_{X^{(2)}}$ consisting of the unique ordered values of the marginal cdfs $F^{(1)}$ and $F^{(2)}$ allows us to determine $\hat{P}^{(j)}_{s_l}$ in a manner consistent with (2.4) in the comonotone case and (2.5) in the antimonotone cases. Specifically, the starting point of the $\mathcal{S}$-path is given by Lemma 1 in the comonotone case and by Remark 21 in the antimonotone case. Then, the candidate subsequent points are given by Lemma 2 and Lemma 6 in the comonotone and antimonotone cases, respectively. As the $\mathcal{S}$-path is determined, this also allows us to determine the relative ordering of the values of $F^{(1)}$ and $F^{(2)}$ in Π_Z, which, in turn, allows us to compute $\hat{P}^{(j)}_{s_l}$.

In this section we provide a detailed exposition of the intuition and the mechanics of the EJD algorithm in both the comonotone and antimonotone cases. We restrict our discussion of the algorithm to finite discrete probability distributions. Although there are many ways to truncate an infinite discrete distribution, we do so as follows. Given an infinite discrete probability distribution Q such that $Q_i > 0$ for $i = 0, 1, 2, \ldots$, we denote by $\tilde{Q}$ a finite approximation of Q supported on the set of integers $\{0, 1, \ldots, i_{\max}\}$ such that

$$\tilde{Q}_i = \begin{cases} Q_i & i = 0, 1, \ldots, i_{max} - 1 \\ 1 - \sum_{l=0}^{i_{max}-1} \tilde{Q}_l & i = i_{\max}. \end{cases} \tag{2.20}$$

Note that it is also possible to obtain discretized approximations of continuous probability distributions (see, for example, [15] and [31]). While it is not necessary to truncate probability distributions so that they are finite, we do so for practical purposes. We briefly sketch in Section 2.6 how the EJD algorithm can be modified to operate directly (i.e., without truncation) on infinite discrete probability distributions.

Therefore, we make the following assumption for the rest of this section:

Assumption 2. *The distribution P has marginals $Q^{(1)}$ and $Q^{(2)}$, discrete distributions with finite support and finite variance, satisfying the equations (2.3b) and (2.3c), $Q^{(1)}_i > 0$ for $i = 0, 1, \ldots, i_{\max}$ and $Q^{(2)}_j > 0$ for $j = 0, 1, \ldots, j_{\max}$.*

[6] An open source C++ implementation is available at [16].

It is also useful to expand the equations (2.9) for the cdfs $F^{(1)}$ and $F^{(2)}$ associated with probability distributions $Q^{(1)}$ and $Q^{(2)}$, respectively, for the marginal constraints (2.3b) and (2.3c) as follows

$$\sum_{l=0}^{i}\sum_{j=0}^{j_{\max}} P_{l,j} = \sum_{l=0}^{i} Q_l^{(1)} = F_i^{(1)}, \quad i = 0, 1, \ldots, i_{\max}, \tag{2.21}$$

$$\sum_{l=0}^{j}\sum_{i=0}^{i_{\max}} P_{i,l} = \sum_{l=0}^{j} Q_l^{(2)} = F_j^{(2)}, \quad j = 0, 1, \ldots, j_{\max}. \tag{2.22}$$

Let us denote by s_0 the 1^{st} point on the $\mathcal{S}$-path associated with $\hat{P}^{(j)}$, for j either 1 or 2. For finite-dimensional marginal distributions $Q^{(1)}$ and $Q^{(2)}$, denote by $i_{\max}$ and $j_{\max}$ the greatest integers such that $Q^{(1)}_{i_{\max}} > 0$ and $Q^{(2)}_{j_{\max}} > 0$. Let $s_0 = (0, 0)$ in the comonotone case (Lemma 1) and $s_0 = (0, j_{\max})$ in the antimonotone case (Lemma 5 and Remark 21), respectively, be the 1^{st} point on the $\mathcal{S}$-path and if s_{l-1} for $l = 1, 2, \ldots$, is the l^{th} point on the $\mathcal{S}$-path, then the subsequent point, s_l, to s_{l-1} is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path. Finally, let $l_{\max}$ denote the final iteration of Algorithm 1.

Remark 27 (Terminal points of the $\mathcal{S}$-path)**.** *In the case of finite marginal distributions $Q^{(1)}$ and $Q^{(2)}$ that have support $\{0, 1, \ldots, i_{\max}\}$ and $\{0, 1, \ldots, j_{\max}\}$, respectively, the point $s_{l_{\max}}$ corresponding to iteration $l_{\max}$ of Algorithm 1 is the terminal point on the $\mathcal{S}$-path for $\hat{P}^{(1)}$ and takes the value*

$$(s^{(1)}_{l_{\max}}, s^{(2)}_{l_{\max}}) = (i_{\max}, j_{\max}) \tag{2.23}$$

in the comonotone case (Remark 13) and the terminal point $s_{l_{\max}}$ takes the value

$$(s^{(1)}_{l_{\max}}, s^{(2)}_{l_{\max}}) = (i_{\max}, 0) \tag{2.24}$$

in the antimonotone case (Remark 21) in order for the marginal constraints (2.3b) and (2.3c) to be satisfied.

We show the correctness of Algorithm 1 through proof by induction in the comonotone case. For the antimonotone case, we show that the preprocessing of one of the input marginal distributions and postprocessing of the resultant $\mathcal{S}$-path enables us to transform the problem into a comonotone problem, thus enabling reuse of much of Algorithm 1.

2.3.1 The Comonotone Case

Let $\Pi_{X^{(1)}} = \{F_0^{(1)}, F_1^{(1)}, \ldots, F_{i_{\max}}^{(1)}\}$ and $\Pi_{X^{(2)}} = \{F_0^{(2)}, F_1^{(2)}, \ldots, F_{j_{\max}}^{(2)}\}$ denote the ordered sets of the cdfs of the marginal distributions $Q^{(1)}$ and $Q^{(2)}$ defined in (2.9). Note that the values of the cdfs partition the unit interval and are strictly increasing

$$\begin{aligned} 0 = F_{-1}^{(1)} < F_0^{(1)} < F_1^{(1)} < \cdots < F_{i_{\max}}^{(1)} = 1 \\ 0 = F_{-1}^{(2)} < F_0^{(2)} < F_1^{(2)} < \cdots < F_{j_{\max}}^{(2)} = 1. \end{aligned} \tag{2.25}$$

This follows immediately from (2.9), Assumption 2, the artificial values $F_{-1}^{(1)} = F_{-1}^{(2)} = 0$ and the finiteness of the marginal probability distributions. Let us also define a joint partition of the unit interval as follows.

Definition 21 (Joint partition of the unit interval)**.** *Let $\Pi_Z = \Pi_{X^{(1)}} \vee \Pi_{X^{(2)}} = \{z_0, z_1, \ldots\}$ denote a partition of the unit interval where the points z_i are the unique ordered values of the union of the sets*

$\Pi_{X^{(1)}}$ *and* $\Pi_{X^{(2)}}$*. The point* z_i *is the* $(i+1)^{\text{st}}$ *smallest value in* Π_Z*. That is,* z_0 *is the smallest element in* Π_Z*,* z_1 *is the second smallest element in* Π_Z*,* z_2 *is the third smallest element in* Π_Z*, etc. Moreover, we define* $z_{-1} = 0$*.*

Remark 28. *Note that there cannot be duplicates in* Π_Z*, that is* $z_i \neq z_{i+1}$ *for all* i*, since the joint partition is taken from the* unique *values of the union of the sets* $\Pi_{X^{(1)}}$ *and* $\Pi_{X^{(2)}}$*.*

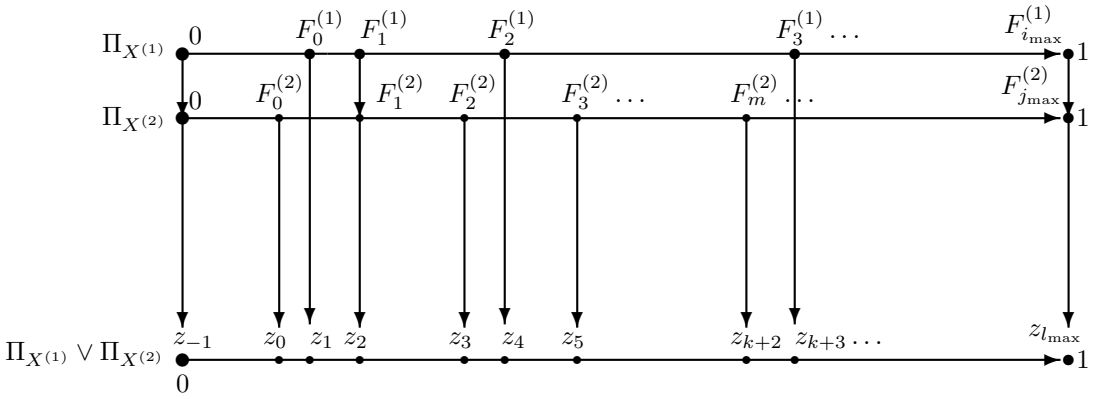

Figure 2.2: Partitions of the unit interval in the comonotone case.

Figure 2.2 illustrates the fact that the joint partition $\Pi_Z = \Pi_{X^{(1)}} \vee \Pi_{X^{(2)}} = \{z_0, z_1, \dots\}$ is constructed by taking the unique ordered values of the union of the sets $\Pi_{X^{(1)}}$ and $\Pi_{X^{(2)}}$. The arrows in the figure represent the partitions of unity corresponding to each marginal distribution. That the two arrows $\Pi_{X^{(1)}}$ and $\Pi_{X^{(2)}}$ are oriented in the same direction indicates the comonotone relationship between the marginal distributions.

The main result of this section is the following theorem which establishes the correctness of Algorithm 1 in the comonotone case. The proof can be found in Appendix A.

Theorem 3. *The probability* $\hat{P}^{(1)}_{s_l^{(1)}, s_l^{(2)}}$ *computed by Algorithm 1 is correct in the sense that it agrees with* (2.4)*.*

Remark 29 (Staircase-like Property of the $\mathcal{S}$-path in the Comonotone Case)**.** *Comonotone* $\mathcal{S}$*-paths are sparse (see Remark 15). That the comonotone* $\mathcal{S}$*-path determined by Algorithm 1 exhibits a staircase-like property can be seen from Lines 8-17 of Algorithm 1, where given the current point* $s_l = (s_l^{(1)}, s_l^{(2)})$ *of the* $\mathcal{S}$*-path, determines the subsequent point* s_{l+1} *which, by Definition 17, must be one of* $(s_l^{(1)} + 1, s_l^{(2)})$*,* $(s_l^{(1)}, s_l^{(2)} + 1)$ *or* $(s_l^{(1)} + 1, s_l^{(2)} + 1)$ *with the smallest Euclidean distance.*

2.3.2 The Antimonotone Case

Algorithm 1 on page 45 in the antimonotone case only differs from the comonotone case in two places:

1. The construction of the input marginal cdf corresponding to $Q^{(2)}$ on Line 1 of Algorithm 1 which calls Algorithm 3 on page 46.
2. Postprocessing[7] of the second coordinate of the $\mathcal{S}$-path on Line 20 of Algorithm 1 which calls Algorithm 2 on page 45.

The main steps of Algorithm 1 (Lines 2-18) remain the same. We begin by first describing the preprocessing and postprocessing steps.

The preprocessing of $Q^{(2)}$ is done in Lines 3-5 of Algorithm 3 on page 46 which is called by Line 1 of Algorithm 1. Algorithm 3 returns the vectors $[F_0^{(1)}, \ldots, F_{i_{\max}}^{(1)}]$ and $[F_0^{(2)}, \ldots, F_{j_{\max}}^{(2)}]$ used in the rest of Algorithm 1. In the antimonotone case of Algorithm 1, the if statement in Line 2 of Algorithm 3 evaluates to true. Then, Lines 3-5 of Algorithm 3 are executed with input vector

$$\mathbf{Q}^{(2)} = [Q_0^{(2)}, Q_1^{(2)}, \ldots, Q_{j_{\max}}^{(2)}] \tag{2.26}$$

and output vector

$$\breve{\mathbf{Q}}^{(2)} = [Q_{j_{\max}}^{(2)}, Q_{j_{\max}-1}^{(2)}, \ldots, Q_0^{(2)}]. \tag{2.27}$$

Algorithm 3 then takes a cumulative sum of the vector $\breve{\mathbf{Q}}^{(2)}$ in Line 7 to obtain

$$\begin{aligned} \breve{\mathbf{F}}^{(2)} &= [Q_{j_{\max}}^{(2)}, Q_{j_{\max}}^{(2)} + Q_{j_{\max}-1}^{(2)}, \ldots, Q_{j_{\max}}^{(2)} + Q_{j_{\max}-1}^{(2)} + \cdots + Q_{j_{\max}-n}^{(2)}, \ldots, 1] \\ &= [\breve{F}_0^{(2)}, \breve{F}_1^{(2)}, \ldots, \breve{F}_n^{(2)}, \ldots, \breve{F}_{j_{\max}}^{(2)}]. \end{aligned} \tag{2.28}$$

Note that the marginal cdf corresponding to $Q^{(1)}$ is constructed normally. That is, given the vector

$$\mathbf{Q}^{(1)} = [Q_0^{(1)}, \ldots, Q_{i_{\max}}^{(1)}]$$

as input, Line 6 of Algorithm 3 constructs

$$\begin{aligned} \mathbf{F}^{(1)} &= [Q_0^{(1)}, Q_0^{(1)} + Q_1^{(1)}, \ldots, 1] \\ &= [F_0^{(1)}, F_1^{(1)}, \ldots, F_{j_{\max}}^{(1)}], \end{aligned}$$

which is the cdf corresponding to $Q^{(1)}$. (See (2.9).)

The postprocessing of the $\mathcal{S}$-path by reversing the second coordinate of s_l is carried out by Algorithm 2 on page 45. For each s_l, Algorithm 2 constructs $\bar{s}_l$ by the following mapping

$$\begin{aligned} \bar{s}_l &= (\bar{s}_l^{(1)}, \bar{s}_l^{(2)}) \\ &= (s_l^{(1)},\ j_{\max} - s_l^{(2)}) \qquad \text{for} \quad l = 0, 1, \ldots, l_{\max}. \end{aligned} \tag{2.29}$$

We use Algorithm 1 to solve (2.3) in the minimization case with marginals $Q^{(1)}$ and $\breve{Q}^{(2)}$, defined above in (2.27), instead of $Q^{(2)}$. This returns a solution consisting of the probabilities $\hat{P}^{(1)}$ with corre-

[7] The postprocessing step is missing in the description of the EJD algorithm in [17]

sponding $\mathcal{S}$-path s_l for $l = 0, 1, \ldots, l_{\max}$. (Recall that the $\mathcal{S}$-path s_l is the support of $\hat{P}^{(1)}$.) We denote by $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$ the probabilities obtained from Algorithm 1 in the maximization case when using marginals $Q^{(1)}$ and $\breve{Q}^{(2)}$, where $\breve{\mathbf{F}}^{(2)}$ is defined above in (2.28). The solution $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$ is related to the solution $\hat{P}^{(2)}$ of (2.3) in the minimization case with marginals $Q^{(1)}$ and $Q^{(2)}$ as follows. The $\mathcal{S}$-path of $\hat{P}^{(2)}$ is $\bar{s}_l = (s_l^{(1)}, j_{\max} - s_l^{(2)})$ for $l = 0, 1, \ldots, l_{\max}$, where $s_l = (s_l^{(1)}, s_l^{(2)})$ for $l = 0, 1, \ldots, l_{\max}$ is the $\mathcal{S}$-path for $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$. Moreover, all the non-zero values of $\hat{P}^{(2)}$ for $l = 0, 1, \ldots, l_{\max}$ are given by

$$\hat{P}^{(2)}_{\bar{s}_l} = \hat{P}^{(1)}_{s_l \,|\, \breve{\mathbf{F}}^{(2)}} \quad \text{for} \quad l = 0, 1, \ldots, l_{\max}. \tag{2.30}$$

Therefore, in this subsection we show that:

1. The probabilities $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$ computed by using $\breve{\mathbf{F}}^{(2)}$ defined in (2.28) in the comonotone case of Algorithm 1 is consistent with (2.5). In particular, we show that (2.30) holds.

2. The $\mathcal{S}$-path computed using the comonotone case of Algorithm 1 with the second components transformed according to (2.29) in Algorithm 2 is the $\mathcal{S}$-path corresponding to $\hat{P}^{(2)}$

We begin by proving the following Lemma from which (2.30) immediately follows.

Lemma 14. *The probabilities $\hat{P}^{(1)}_{i,j|\,\breve{\boldsymbol{F}}^{(2)}}$ that solve (2.3) in the maximization case using the marginal cdfs $F^{(1)}$ and $\breve{F}^{(2)}$ are related to the probabilities $\hat{P}^{(2)}_{i,j}$ that solve (2.3) in the minimization case using the marginal cdfs $F^{(1)}$ and $F^{(2)}$ by the equation*

$$\hat{P}^{(2)}_{i,j_{\max}-j} = \hat{P}^{(1)}_{i,j\,|\,\breve{\boldsymbol{F}}^{(2)}} \quad \textit{for} \quad i = 0, 1, 2, \ldots, i_{\max} \quad \textit{and} \quad j = 0, 1, 2, \ldots, j_{\max}. \tag{2.31}$$

In particular, the point (i, j) is in the support of $\hat{P}^{(1)}_{\breve{\boldsymbol{F}}^{(2)}}$ if and only if the point $(i, j_{\max} - j)$ is in the support of $\hat{P}^{(2)}$.

PROOF: Recall from (2.5) that

$$\hat{P}^{(2)}_{i,j_{\max}-j} = [\min(F_i^{(1)}, \bar{F}^{(2)}_{j_{\max}-j-1}) - \max(F^{(1)}_{i-1}, \bar{F}^{(2)}_{j_{\max}-j})]^+ \tag{2.32}$$

for $i = 0, 1, 2, \ldots, i_{\max}$, $j = 0, 1, 2, \ldots, j_{\max}$ where $F^{(k)}_{-1} = 0$ for $k = 1, 2$,

$$\bar{F}^{(2)}_j = 1 - F^{(2)}_j \tag{2.33}$$

and $\bar{F}^{(2)}_{-1} = 1$. Similarly, recall from (2.4) and the fact that we use $\breve{\mathbf{F}}^{(2)}$ instead of $F^{(2)}$ that

$$\hat{P}^{(1)}_{i,j\,|\,\breve{\mathbf{F}}^{(2)}} = [\min(F_i^{(1)}, \breve{F}^{(2)}_j) - \max(F^{(1)}_{i-1}, \breve{F}^{(2)}_{j-1})]^+. \tag{2.34}$$

Therefore, in order to show that (2.31) is true, we need only to show that

$$\breve{F}^{(2)}_j = \bar{F}^{(2)}_{j_{\max}-j-1} \tag{2.35}$$

and

$$\breve{F}^{(2)}_{j-1} = \bar{F}^{(2)}_{j_{\max}-j} \tag{2.36}$$

for all $j = 0, 1, ..., j_{\max}$.
To this end, let us first show that (2.35) holds:

$$\begin{aligned}
\bar{F}^{(2)}_{j_{\max}-j-1} &= 1 - F^{(2)}_{j_{\max}-j-1} \\
&= 1 - \sum_{i=0}^{j_{\max}-j-1} Q_i^{(2)} \\
&= \sum_{i=0}^{j_{\max}} Q_i^{(2)} - \sum_{i=0}^{j_{\max}-j-1} Q_i^{(2)} \\
&= \sum_{i=j_{\max}-j}^{j_{\max}} Q_i^{(2)} \\
&= \breve{F}^{(2)}_j.
\end{aligned}$$

The proof that (2.36) holds is similar to the proof above that (2.35) holds. □

Next, we show that the $\mathcal{S}$-path $\bar{s}_l$ obtained from transforming the second coordinate of the $\mathcal{S}$-path s_l computed by the comonotone case of Algorithm 1 according to (2.29) in Algorithm 2 is the correct $\mathcal{S}$-path that corresponds to $\hat{P}^{(2)}$. (Recall that in the antimonotone case, s_l is the $\mathcal{S}$-path that corresponds to $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$.)

Proposition 2. *For $l = 0, 1, \ldots, l_{\max}$, if $s_l = (s_l^{(1)}, s_l^{(2)})$ is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\breve{\mathbf{F}}^2}$, then $\bar{s}_l = (s_l^{(1)}, j_{\max} - s_l^{(2)})$ is the $(l+1)^{\text{st}}$ point on the S-path associated with $\hat{P}^{(2)}$.*

The proof of Proposition 2 can be found in Appendix A.

Since we have proved in Appendix A that Algorithm 1 correctly computes the $\mathcal{S}$-path s_l for $l = 0, 1, \ldots, l_{\max}$, associated with $\hat{P}^{(1)}$, it follows immediately from Proposition 2 and Algorithms 2 and 3 that Algorithm 1 correctly computes the $\mathcal{S}$-path $\bar{s}_l$ for $l = 0, 1, \ldots, l_{\max}$, associated with $\hat{P}^{(2)}$. To see that Algorithm 1 also returns the correct probabilities $\hat{P}^{(2)}_{\bar{s}_l}$ for $l = 0, 1, \ldots, l_{\max}$, associated with the $\mathcal{S}$-path $\bar{s}_l$ for $l = 0, 1, \ldots, l_{\max}$, note that Algorithm 1 computes the probabilities as $\hat{P}^{(j)}_l$ for $l = 0, 1, \ldots, l_{\max}$ and actually returns a vector of probabilities $(\hat{P}^{(j)}_0, \hat{P}^{(j)}_1, \ldots, \hat{P}^{(j)}_{l_{\max}})$. In the antimonotone case, these are the probabilities

$$(\hat{P}^{(1)}_{s_0|\breve{\mathbf{F}}^{(2)}}, \hat{P}^{(1)}_{s_1|\breve{\mathbf{F}}^{(2)}}, \ldots, \hat{P}^{(1)}_{s_{l_{\max}}|\breve{\mathbf{F}}^{(2)}}) = (\hat{P}^{(2)}_{\bar{s}_0}, \hat{P}^{(2)}_{\bar{s}_1}, \ldots, \hat{P}^{(2)}_{\bar{s}_{l_{\max}}}).$$

Therefore, Algorithm 1 also returns that correct probabilities $\hat{P}^{(2)}_{\bar{s}_l}$ for $l = 0, 1, \ldots, l_{\max}$, associated with the $\mathcal{S}$-path $\bar{s}_l$ for $l = 0, 1, \ldots, l_{\max}$, for $\hat{P}^{(2)}$.

Remark 30 (Staircase-like Property of the $\mathcal{S}$-path in the Antimonotone Case)**.** *Antimonotone $\mathcal{S}$-paths are sparse (see Remark 23). That the antimonotone $\mathcal{S}$-path determined by Algorithm 1 exhibits a staircase-like property can be seen from Lines 8-17 and Lines 19-20 of Algorithm 1, where, given the current point $s_l = (s_l^{(1)}, s_l^{(2)})$ of the $\mathcal{S}$-path, Algorithm 1 determines the subsequent point s_{l+1}, which, by Definition 19, must be one of $(s_l^{(1)} + 1, s_l^{(2)})$, $(s_l^{(1)}, s_l^{(2)} - 1)$ or $(s_l^{(1)} + 1, s_l^{(2)} - 1)$ with the smallest Euclidean distance.*

Algorithm 1 Extreme Joint Distribution Algorithm in two-dimensions

Require: $[Q_0^{(1)}, \ldots, Q_{i_{\max}}^{(1)}]$
$[Q_0^{(2)}, \ldots, Q_{j_{\max}}^{(2)}]$
$j = 1$ for maximization and $j = 2$ for minimization

Output: Extreme measure $[\hat{P}_0^{(j)}, \ldots, \hat{P}_{l_{\max}}^{(j)}]$ and its corresponding support $[s_0, \ldots, s_{l_{\max}}]$

1: $[F_0^{(1)}, \ldots, F_{i_{\max}}^{(1)}], [F_0^{(2)}, \ldots, F_{j_{\max}}^{(2)}]$
$\leftarrow$ Construct_Marginal_CDFs$([Q_0^{(1)}, \ldots, Q_{i_{\max}}^{(1)}], [Q_0^{(2)}, \ldots, Q_{j_{\max}}^{(2)}], j)$
2: $l \leftarrow 0$
3: $s_0^{(1)} \leftarrow 0$ and $s_0^{(2)} \leftarrow 0$
4: $z_0, \ldots, z_{l_{\max}} \leftarrow$ Partition_Unity$([F_0^{(1)}, \ldots, F_{i_{\max}}^{(1)}], [F_0^{(2)}, \ldots, F_{j_{\max}}^{(2)}])$
5: $\hat{P}_0^{(j)} \leftarrow z_0$
6: **while** $z_l \neq 1$ **do**
7: $\quad l \leftarrow l + 1$
8: $\quad$ **if** $z_{l-1} == F^{(1)}_{s^{(1)}_{l-1}} == F^{(2)}_{s^{(2)}_{l-1}}$ **then**
9: $\quad\quad s_l^{(1)} = s_{l-1}^{(1)} + 1$
10: $\quad\quad s_l^{(2)} = s_{l-1}^{(2)} + 1$
11: $\quad$ **else if** $z_{l-1} == F^{(1)}_{s^{(1)}_{l-1}}$ and $z_{l-1} \neq F^{(2)}_{s^{(2)}_{l-1}}$ **then**
12: $\quad\quad s_l^{(1)} = s_{l-1}^{(1)} + 1$
13: $\quad\quad s_l^{(2)} = s_{(l-1)}^{(2)}$
14: $\quad$ **else if** $z_{l-1} \neq F^{(1)}_{s^{(1)}_{l-1}}$ and $z_{l-1} == F^{(2)}_{s^{(2)}_{l-1}}$ **then**
15: $\quad\quad s_l^{(1)} = s_{(l-1)}^{(1)}$
16: $\quad\quad s_l^{(2)} = s_{l-1}^{(2)} + 1$
17: $\quad$ Save the l-th support point $s_l = (s_l^{(1)}, s_l^{(2)})$
18: $\quad \hat{P}_l^{(j)} \leftarrow z_l - z_{l-1}$
19: **if** $j == 2$ **then** ▷ *Antimonotone Case*
20: $\quad [s_0, \ldots, s_{l_{\max}}] \leftarrow$ Reverse_Support$([s_0, \ldots, s_{l_{\max}}])$
21: **return** $[\hat{P}_0^{(j)}, \ldots, \hat{P}_{l_{\max}}^{(j)}], [s_0, \ldots, s_{l_{\max}}]$

Remark 31. *The* Partition_Unity *subroutine referenced in Line 4 of Algorithm 1, above, is listed in Algorithm 8.*

Remark 32. *Since* Partition_Unity *sorts the unique elements of the union of the marginal cdfs in ascending order and Algorithm 1, in Lines 8-17, increments the coordinate of the support according to the ordering of* $\{z_0, z_1, \ldots, z_{l_{\max}}\}$*, therefore,* $z_l = \min\{F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\}$.

Algorithm 2 Subroutine: Reverse Support

1: **procedure** Reverse_Support$([s_0, \ldots, s_{l_{\max}}])$
2: $\quad$ **for** $l \leftarrow 0, \ldots, l_{\max}$ **do**
3: $\quad\quad \bar{s}_l^{(1)} \leftarrow s_l^{(1)}$
4: $\quad\quad \bar{s}_l^{(2)} \leftarrow j_{\max} - s_l^{(2)}$
5: $\quad$ **return** $[\bar{s}_0, \ldots, \bar{s}_{l_{\max}}]$

Algorithm 3 Subroutine: Construct Marginal CDFs

procedure CONSTRUCT_MARGINAL_CDFS($[Q_0^{(1)}, \ldots, Q_{i_{\max}}^{(1)}], [Q_0^{(2)}, \ldots, Q_{j_{\max}}^{(2)}], j$)
  if $j == 2$ **then** ▷ *Antimonotonic Case*
    $[t_0, \ldots, t_{j_{\max}}] \leftarrow [Q_0^{(2)}, \ldots, Q_{j_{\max}}^{(2)}]$
    for $i = 0, \ldots, j_{\max}$ **do**
      $Q_i^{(2)} \leftarrow t_{j_{\max}-i}$
  $[F_0^{(1)}, \ldots, F_{i_{\max}}^{(1)}] \leftarrow$CUMSUM$([Q_0^{(1)}, \ldots, Q_{i_{\max}}^{(1)}])$
  $[F_0^{(2)}, \ldots, F_{j_{\max}}^{(2)}] \leftarrow$CUMSUM$([Q_0^{(2)}, \ldots, Q_{j_{\max}}^{(2)}])$
  return $[F_0^{(1)}, \ldots, F_{i_{\max}}^{(1)}]$, $[F_0^{(2)}, \ldots, F_{j_{\max}}^{(2)}]$

Remark 33. *Algorithm 1 is very sensitive to rounding errors; implementations of Algorithm 1 in floating-point arithmetic need to account for rounding-errors. For simplicity, we assume in this thesis that all computations are carried out in exact arithmetic.*

2.4 Extreme Measures in d-dimensions

In this section, we extend the Extreme Joint Distribution (EJD) approach to the general d-dimensional setting for $d \geq 3$. We consider random vectors $(X^{(1)}, \ldots, X^{(d)})$ on the positive quadrant of the d-dimensional lattice, $(i_1, \ldots, i_d) \in \mathbb{N}_0^d$, where each $X^{(k)}$, for $k \in \{1, \ldots, d\}$, has the associated discrete probability distribution $Q_i^{(k)}$, where $i = 0, 1, \ldots$, having finite variance supported on $\mathbb{N}_0$. Note that $\mathbb{N}_0$ is the set of non-negative integers and $\mathbb{N}_0^d$ is the d-dimensional integer lattice of non-negative integers. In this section, the $Q_i^{(k)}$, for $k \in \{1, 2, \ldots, d\}$ and $i = 0, 1, \ldots$, are the given marginal distributions, similar to $Q_i^{(1)}$ and $Q_j^{(2)}$ in Definition 14 (the two-dimensional case of Definition 22 below). We also denote by $\mathbf{0}^d = (0, \ldots, 0) \in \mathbb{R}^d$ the d-dimensional 0 vector. Note that we augment the notation in settings where the dimensionality may be unclear. For example, $P^{(d)}$ refers to a general probability measure in d-dimensions, $\hat{P}^{(j,d)}$ refers to the j^{th} extreme measure in d-dimensions, and $Q^{(d;u)}$ and $F^{(d;u)}$ denote the u^{th} marginal distribution and marginal cdf in the d-dimensional case, respectively.

The following definition extends Definition 14 to d-dimensions. Note that $j \in \{1, \ldots, n\}$ in Definition 22 below, where $n = 2^{d-1}$ (see Lemma 15). For the two-dimensional case, $j = 1$ corresponds to the max case and $j = 2$ corresponds to the min case of (2.3). For the general d-dimensional case, with $d \geq 3$, each $e_k^{(j,d)}$ for $k \in \{1, \ldots, d\}$ and $j \in \{1, \ldots, n\}$ is associated with a monotonicity structure that describes the extremal dependency structure, as discussed in Subsection 2.4.1.

Definition 22 (Extreme measures in d-dimensions)**.** *For each $j = \{1, \ldots, n\}$, the extreme measure $\hat{P}^{(j,d)}$ having associated monotone structure $\boldsymbol{e}^{(j,d)}$ is the solution to the following multi-objective infinite-dimensional LP*

$$\begin{aligned}
&\text{extremize} && h_{u,v}^{(j,d)}(P^{(d)}) && 1 \leq u < v \leq d && \text{(2.37a)}\\
&\textit{subject to} && \sum_{v \in \mathcal{I}_k} \sum_{i_v=0}^{\infty} P^{(d)}_{i_1,\ldots i_{k-1}, i_k, i_{k+1}, \ldots, i_d} = Q_{i_k}^{(k)} && \begin{matrix} k = 1, \ldots, d \\ i_k = 0,1,\ldots \end{matrix} && \text{(2.37b)}\\
& && P^{(d)}_{i_1,\ldots,i_d} \geq 0 && && \text{(2.37c)}
\end{aligned}$$

where

$$\text{extremize } h_{u,v}^{(j,d)}(P^{(d)}) = \begin{cases} \max h_{u,v}^{(j,d)}(P^{(d)}) & \text{if } e_u^{(j,d)} = e_v^{(j,d)} \\ \min h_{u,v}^{(j,d)}(P^{(d)}) & \text{if } e_u^{(j,d)} \neq e_v^{(j,d)}, \end{cases} \tag{2.38}$$

$\mathcal{I}_k = \{r : 1 \leq r \leq d, r \neq k\}$, $Q_i^{(k)} \geq 0$, for $i = 0, 1, 2, \ldots$ and $k = 1, 2, \ldots, d$ and $\sum_{i=0}^{\infty} Q_i^{(k)} = 1$ for $k = 1, \ldots, d$. Each objective function takes the form

$$h_{u,v}^{(j,d)}(P^{(d)}) := \mathbb{E}\left[X^{(u)} X^{(v)}\right] = \sum_{i_u=0}^{\infty} \sum_{i_v=0}^{\infty} i_u i_v \, P^{(d;u,v)}_{i_u,i_v} \qquad 1 \leq u < v \leq d \tag{2.39}$$

where

$$P^{(d;u,v)}_{i_u,i_v} = \sum_{j \in \mathcal{I}_{u,v}} \sum_{i_j=0}^{\infty} P^{(d)}_{i_1,\ldots i_{u-1}, i_u, i_{u+1}, \ldots, i_{v-1}, i_v, i_{v+1}, \ldots, i_d} \tag{2.40}$$

with $\mathcal{I}_{u,v} = \{r : 1 \leq r \leq d, r \neq u, r \neq v\}$.

Remark 34. *Analogous to Remark 5, Definition 22 is consistent with finite-dimensional $Q^{(k)}$, since $Q^{(k)}$ can be simply extended to the infinite-dimensional case by setting $Q_i^{(k)} = 0$ for $i = i_{\max}^{(k)} + 1, i_{,\max}^{(k)} + 2, \ldots$.*

Remark 35. *The multi-objective program (2.37) is, in fact, a multi-objective multi-marginal MKP, the solutions of which determine extreme measures. Again, we mention the connection to mass transportation for the sake of completeness; it is not a focus of this thesis.*

Remark 36. *There are $m = d(d-1)/2$ objective functions in (2.37a) where each $h_{u,v}^{(j,d)}(P^{(d)})$ extremizes the dependency between a pair of coordinates.*

Analogous to the bivariate setting, extreme measures in the general setting are defined to be the solutions of the (multi-objective) infinite-dimensional LP (2.37). Naturally, there are differences between extreme measures in the general setting and extreme measures in the bivariate setting. The most significant difference is the multi-objective nature of (2.37). For example, for each $j \in \{1, 2, \ldots, n\}$, the extreme measure $\hat{P}^{(j,d)}$ satisfies the m constrained optimization problems (2.37a). In general, solutions to multi-objective problems lie on a Pareto frontier; it is typical that no unique solution exists that satisfies all the constituent sub-problems [2]. Standard techniques in multi-objective optimization are unable to compute a solution that can satisfy all m constrained optimization problems (2.37a). A surprising property of the structure of the multi-objective problem (2.37) is that the Pareto frontier of solutions to the multi-objective problem is actually a single point (solution). In other words, each extreme measure $\hat{P}^{(j,d)}$ for $j \in \{1, 2, \ldots, d\}$ actually extremizes $h_{u,v}^{(j,d)}$ for each pair u, v satisfying $1 \leq u < v \leq d$ given the constraints (2.37b) and (2.37c). We prove this in the course of showing the correctness of Algorithm 5.

Remark 37. *We show in what follows that the $m = d(d-1)/2$ solutions to the two-dimensional LP (2.3) are all embedded within the solution to the d-dimensional multi-objective LP (2.37). This is significant since, in general, solutions to multi-objective optimization problems are not optimal solutions to all of the constituent sub-problems of a multi-objective problem. Typically, solutions to multi-objective optimization problems lie on a Pareto frontier (a d-dimensional curve) where different points of the frontier correspond to different trade-offs between the optimality of the solution to each sub-problem.*

Lemma 15. *For any given set of marginal distributions $Q^{(1)}, Q^{(2)}, \ldots, Q^{(d)}$ associated with the optimization problem (2.37), the number of extreme measures is $n = 2^{d-1}$.*

Proof: Let us prove this lemma for the case $d \geq 3$, since the case $d = 2$ is obvious. Recall that the set of marginal distributions $Q^{(1)}, Q^{(2)}, \ldots, Q^{(d)}$ corresponds to the d-dimensional random vector $(X^{(1)}, \ldots, X^{(d)})$. For each two-dimensional projection $(X^{(u)}, X^{(v)})$, where $u, v \in \{1, \ldots, d\}$ and $u \neq v$, the corresponding joint distribution is either comonotone or antimonotone by the construction of the optimization problem (2.37). Denote by d_c the number of coordinates that are comonotone to $X^{(1)}$. Then, the number of random variables antimonotone with $X^{(1)}$ satisfies

$$d_a = d - 1 - d_c.$$

The total number of partitions of the number $d-1$ in the additive form, $d - 1 = d_a + d_c$, is $n = 2^{d-1}$. Clearly, n does not depend on the choice of the first random variable. □

In what follows, it is necessary to project random vectors and their corresponding multivariate distributions down to two-dimensions. We define bivariate projections as follows.

Definition 23 (Bivariate Projections)**.** *Let $(m_1, \ldots, m_d)$ be a d-dimensional vector and, for any $u, v \in$*

$\{1, 2, \ldots, d\}$ with $u \neq v$, define its bivariate projection:

$$\text{Proj}_{u,v}(m_1, \ldots, m_d) = (m_u, m_v). \tag{2.41}$$

Similarly, let $\boldsymbol{e}^{(j,d;u,v)}$ denote the projection of the d-dimensional monotone structure $\boldsymbol{e}^{(j,d)}$ onto the u^{th} and v^{th} coordinates:

$$\boldsymbol{e}^{(j,d;u,v)} = \text{Proj}_{u,v}\, \boldsymbol{e}^{(j,d)} = (e_u^{(j,d)}, e_v^{(j,d)}). \tag{2.42}$$

Let $P^{(d)}$ be a d-dimensional probability distribution. The projection of $P^{(d)}$ to the u^{th} and v^{th} coordinates is given by the marginalization of all components l where $l \neq u, v$, that is,

$$P_{i_u,i_v}^{(d;u,v)} = \sum_{l \in \mathcal{I}_{u,v}} \sum_{i_l=0}^{\infty} P_{i_1,\ldots,i_{u-1},i_u,i_{u+1},\ldots,i_{v-1},i_v,i_{v+1},\ldots,i_d}^{(d)}, \tag{2.43}$$

where $\mathcal{I}_{u,v} = \{l : 1 \leq l \leq d, l \neq u, l \neq v\}$. Similarly, define

$$P_{i_v}^{(d;v)} = \sum_{l \in \mathcal{I}_v} \sum_{i_l=0}^{\infty} P_{i_1,\ldots i_{v-1},i_v,i_{v+1},\ldots,i_d}^{(d)}, \tag{2.44}$$

where $\mathcal{I}_v = \{l : 1 \leq l \leq d, l \neq v\}$.

Remark 38. *Note that (2.43) and (2.44) apply straightforwardly to $\hat{P}^{(j,d;u,v)}$ and $\hat{P}^{(j,d;v)}$, respectively. Moreover, $\hat{P}^{(j,d;u,v)}$ is comonotone if $e_u^{(j,d)} = e_v^{(j,d)}$ and antimonotone otherwise.*

2.4.1 Monotonicity Structures

In the two-dimensional case, the components of a bivariate distribution can only exhibit extreme positive and extremal negative dependence in the form of extremal positive and negative correlations, $\hat{C}^{(1)}$ and $\hat{C}^{(2)}$, respectively. Note that both $\hat{C}^{(1)}$ and $\hat{C}^{(2)}$ are scalars. In the general d-dimensional case, each component exhibits either extremal positive or extremal negative dependence relative to another component, resulting in $n = 2^{d-1}$ possible combinations of pairwise dependencies (Lemma 15). Consequently, there are n extremal $d \times d$ correlation matrices, $\hat{C}^{(j,d)}$, for $j = 1, 2, \ldots, n$. We introduce the notion of monotonicity structures to describe the possible extremal dependencies between the components of a random vector $(X^{(1)}, \ldots, X^{(d)})$.

Definition 24 (Monotonicity Structure)**.** *For $j \in \{1, 2, \ldots, n\}$, the j^{th} d-dimensional monotonicity structure $\boldsymbol{e}^{(j,d)}$ is a binary column vector*

$$\boldsymbol{e}^{(j,d)} = \begin{pmatrix} e_1^{(j,d)} \\ \vdots \\ e_d^{(j,d)} \end{pmatrix} \tag{2.45}$$

describing the pairwise extremal dependency structure between the components of a random vector $(X^{(1)}, \ldots, X^{(d)})$. The element $e_k^{(j,d)}$, for $k \in \{2, 3, \ldots, d\}$, of the vector $\boldsymbol{e}^{(j,d)}$, known as a monotonicity

indicator, takes on binary values

$$e_k^{(j,d)} = \begin{cases} 0, & \text{if } X^{(1)} \text{ and } X^{(k)} \text{ are comonotone} \\ 1, & \text{if } X^{(1)} \text{ and } X^{(k)} \text{ are antimonotone} . \end{cases}$$

By convention, we take $e_1^{(j,d)} = 0$.

Figure 2.3 provides a graphical illustration of monotonicity structures. Note that the two monotonicity structures describe the *same* extreme measure even though the arrows are pointing in different directions. The important aspect here is the relative orientation of the arrows with respect to other arrows within the same monotonicity structure.

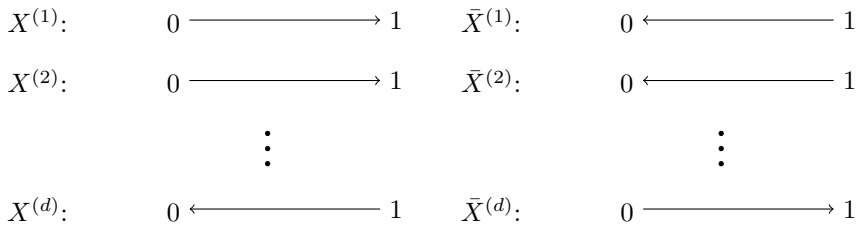

Figure 2.3: Monotonicity structures of extreme measures. The direction of the arrows indicates the direction of monotonically increasing support.

Monotonicity Structures as Binary Numbers

Monotonicity structures are useful not only in describing the extremal dependence structure between the components of a multivariate process, but they are also useful as inputs to the d-dimensional version of the EJD theorem (Subsection 2.4.2) and algorithm (Section 2.5) in order to compute the correct k^{th} extreme measure. Thus, we require a method to easily generate monotone structures. Fortunately, they admit a representation as binary numbers. This is best illustrated through an example. To that end, suppose that $d = 3$, resulting in the following monotone structure:

$$\begin{matrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{matrix} \tag{2.46}$$

The 3×4 matrix in (2.46) indicates that there are 4 extreme measures (the number of columns), each of dimension 3 (the number of rows). The first column corresponds to $\mathbf{e}^{(1,3)}$ representing extreme positive dependence between all marginals (comonotonicity). The remaining columns correspond to extreme measures with two components exhibiting extreme negative dependence and one component exhibiting extreme positive dependence. Moreover, note that for the j^{th} column in (2.46), where $j \in \{1, 2, 3, 4\}$, the top element corresponds to $\mathbf{e}_1^{(j,3)}$, the second from the top element corresponds to $\mathbf{e}_2^{(j,3)}$, and the bottom element corresponds to $\mathbf{e}_3^{(j,3)}$.

We show that (2.46) is easy to construct. The key realization is that the extremal structure represented by each column of the monotonicity structure *is just the index of the column in binary*, using d bits to represent the column indices, where the columns are numbered $0, 1, \ldots, d-1$, instead of being numbered according to 1-indexing $1, 2, \ldots, d$, and the most significant bit of the binary number is at the

top of the column. We rewrite (2.46) below in order to elucidate the relationship between the index of the column and its monotonicity structure

$$\begin{array}{cccc} 0 & 1 & 2 & 3 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{array}. \tag{2.47}$$

Algorithm 4 Generation of monotonicity structures

Require: Dimension of the problem: d

Output: Monotonicity structure matrix: $\mathbf{E} = [\mathbf{e}^{(1,d)}, \ldots, \mathbf{e}^{(n,d)}] \in \mathbb{R}^{d \times n}$

1: $n \leftarrow 2^{(d-1)}$
2: **for** $j = 1, \ldots, n$ **do**
3: $\quad \mathbf{e}^{(j,d)} \leftarrow$ BINARY_REPRESENTATION$(j-1, d)$ ▷ *returns the binary representation, using d bits, of the integer $j-1$ as an integer vector*
4: $\quad \mathbf{e}^{(j,d)} \leftarrow \text{transpose}(\mathbf{e}^{(j,d)})$ ▷ *so we get column vectors with the most significant bit in the first row*
5: **return** $\mathbf{E} = [\mathbf{e}^{(1,d)}, \ldots, \mathbf{e}^{(n,d)}]$ ▷ *concatenate column vectors into a matrix*

2.4.2 Extreme Joint Distributions in d-dimensions

Theorem 4 (EJD Theorem in d-dimensions)**.** *If problem (2.37) has a solution $\hat{P}^{(j,d)}$, then the solution must satisfy*

$$\begin{aligned} \hat{P}^{(j,d)}_{i_1,\ldots,i_d} = [&\min(\bar{F}_1(i_1 - e_1^{(j,d)}; e_1^{(j,d)}), \ldots, \bar{F}_d(i_d - e_d^{(j,d)}; e_d^{(j,d)})) \\ &- \max(\bar{F}_1(i_1 + (e_1^{(j,d)} - 1); e_1^{(j,d)}), \ldots, \bar{F}_d(i_d + (e_d^{(j,d)} - 1); e_d^{(j,d)}))]^+ \end{aligned} \tag{2.48}$$

where $[\cdot]^+ = \max(0, \cdot)$ and, for $k \in \{1, 2, \ldots, d\}$, $\bar{F}_k$ is defined as

$$\bar{F}_k(i_k; e_k^{(j,d)}) = \begin{cases} F^{(k)}_{i_k} & \text{if } \; e_k^{(j,d)} = 0 \\ 1 - F^{(k)}_{i_k} & \text{if } \; e_k^{(j,d)} = 1 \end{cases} \tag{2.49}$$

for $i_k = -1, 0, 1, \ldots$ where $F^{(k)}_{-1} = 0$ and, for $i_k = 0, 1, \ldots$, $F^{(k)}_{i_k}$ denotes the i_k^{th} point of the marginal cdf corresponding to $Q^{(k)}$, assuming that $e_1^{(j,d)} = 0$. Moreover, if the problem (2.37) has a solution $\hat{P}^{(j,d)}$, then $\hat{P}^{(j,d)}$ is unique.

Remark 39. *While we need to assume the existence of a solution to (2.37) in the proof of Theorem 4 below—since the proof relies on sampling from $\hat{P}^{(j,d)}$ and the subsequent use of Borel's Law of Large Numbers to show convergence; see (2.54)–(2.58)—the correctness of Algorithm 5 proves that there is a solution to (2.37). Therefore, the correctness of Algorithm 5 and Theorem 4 together imply that there is a unique solution to the problem (2.37) and it satisfies (2.48).*

PROOF: We begin by defining two sets of integers

$$\mathcal{I}_k = \{j : 1 \leq j \leq d, \quad j \neq k, \}$$

and

$$\mathcal{I}_{k,l} = \{j : 1 \le j \le d, \quad j \ne k, \quad j \ne l\}.$$

Let us first show that, if $d = 2$, then (2.48) is equivalent to (2.4) in the case of comonotonicity and to (2.5) in the case of antimonotonicity. To this end, note that, in the bivariate comonotone case, the monotone structure is $\mathbf{e}^{(1,2)} = (0, 0)$ and thus $\bar{F}_1(i, e_1^{(1,2)}) = F_i^{(1)}$ and $\bar{F}_2(i, e_2^{(1,2)}) = F_i^{(2)}$ for all $i \ge 0$. In the antimonotone case, the monotone structure is $\mathbf{e}^{(2,2)} = (0, 1)$ and thus $\bar{F}_1(i, e_1^{(2;2)}) = F_i^{(1)}$ but $\bar{F}_2(i, e_2^{(2;2)}) = 1 - F_i^{(2)}$ for all $i \ge 0$. Therefore, (2.48) is equivalent to (2.4) and (2.5) in the two-dimensional case. Consequently, Theorem 1 shows that Theorem 4 holds in the case $d = 2$.

Let us now consider the general case $d \ge 3$. There are two subsets of coordinates of $(X^{(1)}, \dots, X^{(d)})$, comonotone coordinates and antimonotone coordinates. Denote their indices by

$$\mathcal{I}_{\mathcal{C}} = \{r : e_r^{(j,d)} = 0\} \quad \text{and} \quad \mathcal{I}_{\mathcal{A}} = \{r : e_r^{(j,d)} = 1\}. \tag{2.50}$$

Assume that a solution $\hat{P}^{(j,d)}$ to (2.37) exists and generate a sample of size N, $\{(\hat{X}_n^{(1)}, \dots, \hat{X}_n^{(d)})\}_{n=1}^N$, from $\hat{P}^{(j,d)}$. An extension of Lemma 9 (see Remark 40) to the d-dimensional case shows that there exists a permutation π ordering the samples such that the first coordinate and the other comonotone coordinates are monotonically increasing

$$X_1^{(k)} \le X_2^{(k)} \le \cdots \le X_N^{(k)} \quad \text{where} \quad X_n^{(k)} = \hat{X}_{\pi(n)}^{(k)} \quad \text{for} \quad k \in \mathcal{I}_C \tag{2.51}$$

and the antimonotone coordinates are monotonically decreasing

$$X_1^{(k)} \ge X_2^{(k)} \ge \cdots \ge X_N^{(k)} \quad \text{where} \quad X_n^{(k)} = \hat{X}_{\pi(n)}^{(k)} \quad \text{for} \quad k \in \mathcal{I}_A. \tag{2.52}$$

Suppose that the indices $1 = k_1 < k_2 < k_3 < \cdots < k_C \le d$ belong to $\mathcal{I}_{\mathcal{C}}$ and the complimentary set of indices is $\mathcal{I}_{\mathcal{A}} = \{l_1, l_2, \dots, l_A\}$. A permuted sample is illustrated in (2.53) below,

$$\begin{aligned}
X^{(1)} &: \overbrace{0, \dots, 0}^{N_1(0)}, \dots, \overbrace{i-1, \dots, i-1}^{N_1(i-1)}, \ \overbrace{i, i, \dots, i}^{N_1(i)}, \dots \overbrace{k, k \dots, k}^{N_1(k)}, \dots \\
&\vdots \\
X^{(k_2)} &: \underbrace{0, 0, \dots, 0}_{N_{k_2}(0)}, \dots, \underbrace{i-1, \dots, i-1}_{N_{k_2}(i-1)}, \ \underbrace{i, \dots, i}_{N_{k_2}(i)}, \ \dots, \\
&\vdots \\
X^{(l_A)} &: \dots \underbrace{i, i, \dots, i}_{N_{l_A}(i)}, \underbrace{i-1, \dots, i-1}_{N_{l_A}(i-1)}, \dots \underbrace{2, 2, 2, \dots 2}_{N_{l_A}(2)}, \dots,
\end{aligned} \tag{2.53}$$

where $N_k(m)$ denotes the number of realizations of m in the sample of the k^{th} coordinate $X^{(k)}$. If $k \in \mathcal{I}_{\mathcal{C}}$, the first position, $I_k^C(m)$, where the number m appears in the sorted sample of the random variable $X^{(k)}$ is

$$I_k^C(m) = 1 + \sum_{i=0}^{m-1} N_k(i).$$

If $k \in \mathcal{I}_{\mathcal{C}}$, the last position, $E_k^C(m)$, where the number m appears in the sorted sample of the random

variable $X^{(k)}$ is

$$E_k^C(m) = \sum_{i=0}^{m} N_k(i).$$

As the sample size $N \to \infty$, we have

$$\lim_{N\to\infty} \frac{N_k(m)}{N} = \hat{P}_m^{(j,d;k)} \quad \textbf{a.s.}\,. \tag{2.54}$$

Since $\hat{P}^{(j,d)}$ satisfies the constraints (2.37b), it follows from (2.44) that $\hat{P}_m^{(j,d;k)} = Q_m^{(k)}$. Therefore, for $k \in \mathcal{I}_{\mathcal{C}}$,

$$\lim_{N\to\infty} \frac{I_k^C(m)}{N} = F_{m-1}^{(k)} \quad \textbf{a.s.} \tag{2.55}$$

and

$$\lim_{N\to\infty} \frac{E_k^C(m)}{N} = F_m^{(k)} \quad \textbf{a.s.}\,. \tag{2.56}$$

In the case of the subset of antimonotone coordinates, $l \in \mathcal{I}_{\mathcal{A}}$, the first index, ${I_l}^A(m)$, where a number m appears in the sorted sample of the r.v. X_l is

$$ {I_l}^A(m) = 1 + N - \sum_{i=0}^{m} N_l(i), \quad l \in \mathcal{I}_{\mathcal{A}}.$$

The last position, ${E_l}^A(m)$, where a number m appears in the sorted sample of the r.v. $X^{(l)}$ is

$$ {E_l}^A(m) = N - \sum_{i=0}^{m-1} N_l(i), \quad l \in \mathcal{I}_{\mathcal{A}}.$$

As $N \to \infty$, we have for $l \in \mathcal{I}_{\mathcal{A}}$

$$\lim_{N\to\infty} \frac{{I_l}^A(m)}{N} = 1 - F_m^{(l)} \quad \textbf{a.s.} \tag{2.57}$$

and

$$\lim_{N\to\infty} \frac{{E_l}^A(m)}{N} = 1 - F_{m-1}^{(l)} \quad \textbf{a.s.}\,. \tag{2.58}$$

Denote by $\mathbf{m_N}(\{(X^{(1)}, \ldots, X^{(d)}) = (i_1, \ldots, i_d)\})$ the number of samples that take the value $(i_1, \ldots, i_d)$. The empirical measure of $\mathbf{m_N}$ is the number of elements in the set

$$\{m : (X_m^{(1)} = i_1, \ldots, X_m^{(d)} = i_d)\} = \Big\{ \bigcap_{k\in\mathcal{I}_{\mathcal{C}}} \{m : X_m^{(k)} = i_k\}\Big\} \bigcap \Big\{ \bigcap_{l\in\mathcal{I}_{\mathcal{A}}} \{m : X_m^{(l)} = i_l\}\Big\}. \tag{2.59}$$

Observe that the right side of (2.59) coincides with the intersection of the intervals and can be written as

$$\Big\{ \bigcap_{k\in\mathcal{I}_{\mathcal{C}}} [I_k^C(i_k), E_k^C(i_k)]\Big\} \bigcap \Big\{ \bigcap_{l\in\mathcal{I}_{\mathcal{A}}} [{I_l}^A(i_l), {E_l}^A(i_l)]\Big\}. \tag{2.60}$$

Now note that (2.60) can be rewritten as follows. The right end of the intersection of the intervals is

$$\mathcal{R} = \min\Big(\min_{k\in\mathcal{I}_{\mathcal{C}}} (E_k^C(i_k)), \min_{l\in\mathcal{I}_{\mathcal{A}}} ({E_l}^A(i_l))\Big) \tag{2.61}$$

and the left end is

$$\mathcal{L} = \max\Big(\max_{k\in\mathcal{I}_C}(I_k^C(i_k)), \max_{l\in\mathcal{I}_A}(I_l{}^A(i_l))\Big). \tag{2.62}$$

Now, let us define

$$\mu_{\mathbf{N}} = \frac{\mathbf{m_N}}{N}. \tag{2.63}$$

Using (2.61) and (2.62), we can rewrite (2.63) as

$$\mu_{\mathbf{N}}(\{(X^{(1)},\dots,X^{(d)}) = (i_1,\dots,i_d)\}) = \frac{(\mathcal{R}-\mathcal{L})^+}{N}.$$

Note that the length of the intersection of intervals is 0 in the case $\mathcal{R} \le \mathcal{L}$. As $N \to \infty$, we obtain from (2.55)–(2.58) that

$$\begin{aligned}\lim_{N\to\infty} \mu_{\mathbf{N}}(\{(X^{(1)},\dots,X^{(d)}) = (i_1,\dots,i_d)\}) = \\ \big[\min(\bar{F}_1(i_1 - e_1^{(j,d)}; e_1^{(j,d)}),\dots,\bar{F}_d(i_d - e_d^{(j,d)}; e_d^{(j,d)})) \\ - \max(\bar{F}_1(i_1 + (e_1^{(j,d)} - 1); e_1^{(j,d)}),\dots,\bar{F}_d(i_d + (e_d^{(j,d)} - 1); e_d^{(j,d)}))\big]^+ \quad \textbf{a.s.}.\end{aligned} \tag{2.64}$$

Finally, note that

$$\lim_{N\to\infty} \mu_{\mathbf{N}}(\{(X^{(1)},\dots,X^{(d)}) = (i_1,\dots,i_d)\}) = \hat{P}^{(j,d)}_{i_1\dots,i_d} \quad \textbf{a.s.}. \tag{2.65}$$

Taking intersections of the sets of probability-one events on which (2.64) and (2.65) holds gives a set of full-measure for which both limits (2.64) and (2.65) coincide. Since both the right side of (2.64) and the right side of (2.65) are constants (independent of ω), equality on a set of measure one implies the equality of the constants themselves. This implies the deterministic identity (2.48) holds for all $(i_1,\dots,i_d)$ and we can drop the "a.s." qualifier.

Thus (2.48) is derived and the first part of the theorem is proved. Moreover, if the problem (2.37) has a solution $\hat{P}^{(j,d)}$, then, since we have shown above that $\hat{P}^{(j,d)}$ must satisfy (2.48), which specifies an analytical form for the probabilities, $\hat{P}^{(j,d)}$ must be unique. □

Remark 40. *As noted in the proof of Theorem 4 above, Lemma 9 can be extended in a straightforward manner to the d-dimensional setting. That is, if $\{\hat{X}_m^{(1)},\dots,\hat{X}_m^{(d)}\}_{m=1}^M$ is a sample of size M from $\hat{P}^{(j,d)}$, then we can construct a permutation π such that*

$$X_1^{(k)} \le X_2^{(k)} \le \cdots \le X_M^{(k)} \quad \textit{where} \quad X_m^{(k)} = \hat{X}_{\pi(m)}^{(k)} \quad \textit{for} \quad k \in \mathcal{I}_C \tag{2.66}$$

and

$$X_1^{(k)} \ge X_2^{(k)} \ge \cdots \ge X_M^{(k)} \quad \textit{where} \quad X_m^{(k)} = \hat{X}_{\pi(m)}^{(k)} \quad \textit{for} \quad k \in \mathcal{I}_A. \tag{2.67}$$

where $\mathcal{I}_C$ and $\mathcal{I}_A$ are defined in (2.50). We sketch a proof of this result below. To that end, let us begin by applying Lemma 9 to the first two components of the d-dimensional comonotonic case, $\boldsymbol{e}^{(1,d)} = (0,\dots,0)$, resulting in $X_1^{(k)},\cdots,X_M^{(k)}$ satisfying (2.66) for $k = \{1,2\}$. We can apply the arguments in Lemma 9 again to the coordinates $k = 1,2,3$. As noted for $k = 2$ in Lemma 9, this may result in a permutation π for which

$$X_{\pi(m)}^{(3)} > X_{\pi(m+1)}^{(3)} \tag{2.68}$$

for some $m = \{0, 1, 2, \ldots, M-1\}$. *However, we can construct a new permutation* $\tilde{\pi}$ *using the same arguments as in Lemma 9 such that*

$$X^{(3)}_{\tilde{\pi}(m)} \leq X^{(3)}_{\tilde{\pi}(m+1)} \tag{2.69}$$

for $m = \{0, 1, \ldots, M-1\}$. *As a result, taking* $\pi = \tilde{\pi}$, *we get that (2.66) is satisfied for* $k \in \{1, 2, 3\}$. *Note that we can repeat this argument iteratively for each coordinate* $k \in \{4, \ldots, d\}$, *ensuring that we can find a permutation* $\tilde{\pi}$ *for which*

$$X^{(k)}_{\tilde{\pi}(m)} \leq X^{(k)}_{\tilde{\pi}(m+1)} \tag{2.70}$$

for $k = \{1, 2, \ldots, d\}$ *and* $m = \{0, 1, \ldots, M-1\}$. *As a result, again taking* $\pi = \tilde{\pi}$, *we get that (2.66) is satisfied for* $k \in \{1, 2, \ldots, d\}$. *A similar argument holds for* $\boldsymbol{e}^{(j,d)}$ *where* $j = \{2, 3, \ldots, n\}$.

Moreover, this implies that, for each $k, l \in \{1, 2, \ldots, n\}$ *with* $k \neq l$, $X^{(k)}_m$ *and* $X^{(l)}_m$ *are also correctly ordered with respect to each other. That is,*

(a) if $\boldsymbol{e}^{(j,d)}_k = \boldsymbol{e}^{(j,d)}_l$, *then either both* $X^{(k)}$ *and* $X^{(l)}$ *satisfy (2.66) if* $\boldsymbol{e}^{(j,d)}_k = \boldsymbol{e}^{(j,d)}_l = 0$ *or they both satisfy (2.67) if* $\boldsymbol{e}^{(j,d)}_k = \boldsymbol{e}^{(j,d)}_l = 1$, *whence in either case they are comonotone, and*

(b) if $\boldsymbol{e}^{(j,d)}_k \neq \boldsymbol{e}^{(j,d)}_l$, *then either* $X^{(k)}_m$ *satisfies (2.66) and* $X^{(l)}_m$ *satisfies (2.67) if* $\boldsymbol{e}^{(j,d)}_k = 0$ *and* $\boldsymbol{e}^{(j,d)}_l = 1$ *or* $X^{(k)}_m$ *satisfies (2.67) and* $X^{(l)}_m$ *satisfies (2.66) if* $\boldsymbol{e}^{(j,d)}_k = 1$ *and* $\boldsymbol{e}^{(j,d)}_l = 0$, *whence in either case they are antimonotone.*

Remark 41. *Since we showed in Remark 40 that, for each* $j \in \{1, 2, \ldots, n\}$, *the permutation* π *not only correctly orders the coordinates* $k = 2, 3, \ldots, d$ *with respect to the first coordinate, but also correctly orders the coordinates with each other. Thus, if one of the* $h^{(j,d)}_{u,v}$ *instead of being maximized is minimized, or vice versa, then there may be no solution to (2.37) that simultaneously solves all of the* $h^{(j,d)}_{u,v}$ *for all* u *and* v *satisfying* $1 \leq u < v \leq d$.

Remark 42 (Embedding of the 2D optimization problem in the multi-objective optimization problem)**.** *For* $j \in \{1, 2, \ldots, n\}$, *the* d*-dimensional multi-objective optimization problem (2.37) with corresponding monotone structure,* $\boldsymbol{e}^{(j,d)}$, *consists of* $m = d(d-1)/2$ *two-dimensional optimization problems of the form (2.3) with the extremization determined by* $\boldsymbol{e}^{(j,d;u,v)}$. *That is, the following two statements hold:*

(1) For any u *and* v *satisfying* $1 \leq u < v \leq d$, $h^{(j,d)}_{u,v}\left(P^{(d)}\right) = h(P^{(d;u,v)})$ *where the extremization of* $h^{(j,d)}_{u,v}$ *on the left side depends on* $\boldsymbol{e}^{(j,d)}$, *the* h *on the right side is the two-dimensional* h *function from (2.3e) and* $P^{(d;u,v)}$ *is the two-dimensional probability measure obtained from the* d*-dimensional probability distribution* $P^{(d)}$ *by (2.40).*

To see this point, note that $h^{(j,d)}_{u,v}(P^{(d)})$ *in problem (2.37) corresponds directly to* $h(P^{(d;u,v)})$ *in (2.3e), as can be seen from the definition of the objective function (2.39).*

(2) For each of the i_k *equations where* $k \in \{u, v\}$, *(2.37b) can be rewritten as*

$$\sum_{i_l=0}^{\infty} P^{(d;k,l)}_{i_k,i_l} = Q^{(k)}_{i_k} \qquad i_k = 0, 1, \ldots \tag{2.71}$$

$$\sum_{i_k=0}^{\infty} P^{(d;k,l)}_{i_k,i_l} = Q^{(l)}_{i_l} \qquad i_l = 0, 1, \ldots \tag{2.72}$$

for all $k \in \{1, \ldots, d\}$ and all $l \in \{1, \ldots, d\}$, excluding $k = l$, respectively. If $k = u$, then (2.71) corresponds to (2.3b). Otherwise, if $k = v$, then (2.72) corresponds to (2.3c). Similarly, (2.37c) reduces to (2.3d).

For a more detailed explanation of why point (2) holds, note that we can rewrite (2.37b) as

$$\sum_{i_l=0}^{\infty} \sum_{v \in \mathcal{I}_{k,l}} \sum_{i_v=0}^{\infty} P^{(d)}_{i_1,\ldots,i_{k-1},i_k,i_{k+1},\ldots,i_d} = Q^{(k)}_{i_k} \qquad k = 1, \ldots, d \text{ and } i_k = 0, 1, \ldots \tag{2.73}$$

and observe that

$$\begin{aligned} &\sum_{v \in \mathcal{I}_{k,l}} \sum_{i_v=0}^{\infty} P^{(d)}_{i_1,\ldots,i_{k-1},i_k,i_{k+1},\ldots,i_d} \\ &\quad = \sum_{v \in \mathcal{I}_{k,l}} \sum_{i_v=0}^{\infty} P^{(d)}_{i_1,\ldots,i_{k-1},i_k,i_{k+1},\ldots,i_{l-1},i_l,i_{l+1},\ldots,i_d} \\ &\quad = P^{(d;k,l)}_{i_k,i_l}. \end{aligned} \tag{2.74}$$

Therefore, substituting (2.74) into (2.73), we get

$$\sum_{i_l=0}^{\infty} P^{(d;k,l)}_{i_k,i_l} = Q^{(k)}_{i_k} \qquad i_k = 0, 1, \ldots . \tag{2.75}$$

Note that (2.75) is the same as (2.71). Hence, we have shown that, if (2.37b) holds, then (2.71) holds too. Similarly, if (2.37b) holds, then (2.72) holds too. Thus, we have shown the first statement in (2), above. For the second statement in (2), a similar argument shows that (2.37c) reduces to (2.3d).

2.4.3 Monotonicity in Higher Dimensions

Comonotonicity and antimonotonicity, introduced in Section 2.2, are properties of bivariate distributions that solve the optimization problem (2.3). Comonotone and antimonotone distributions exhibit extreme positive and extreme negative dependence, respectively, between their components [103]. Comonotonicity generalizes directly to the d-dimensional setting of random vectors $(X^{(1)}, \ldots, X^{(d)})$ and has been well studied in the literature [29, 30]. In contrast, while there are measures of extremal negative dependence that generalize antimonotonicity [126], it is well known that there is no direct extension of antimonotonicity to the general d-dimensional setting [103].

Recall that, in the two-dimensional setting, our motivation for constructing bivariate distributions with extremal dependence between its components was to obtain extreme positive and extreme negative correlations, denoted $\hat{C}^{(1)}$ and $\hat{C}^{(2)}$, respectively. Then, any correlation $C \in [\hat{C}^{(1)}, \hat{C}^{(2)}]$ can be obtained by solving the simple linear equation (2.1). This approach, which extends to the d-dimensional setting and is described in Section 2.8, requires the d-dimensional notion of extremal positive and extremal negative dependence. Since the Pearson correlation is inherently a pairwise concept, we define extremal dependence in the general d-dimensional setting as follows.

Definition 25 (Pairwise Monotonicity)**.** *A random vector $(X^{(1)}, \ldots, X^{(d)})$ where each $X^{(k)} \sim Q^{(k)}$ for $k = 1, \ldots, d$, is considered pairwise monotone if all of its bivariate projections $(X^{(u)}, X^{(v)})$ where $u, v \in \{1, \ldots, d\}$ and $u \neq v$ are either comonotone or antimonotone.*

Pairwise monotonicity is a natural choice for the measure of extremal dependence in the general d-dimensional setting for many reasons. From a modelling perspective, the dependence structure of multivariate distributions with pairwise extremal dependence results in $n = 2^{d-1}$ extreme measures (Lemma 15), each having an associated extreme correlation matrix. The n extreme measures can be used for efficient calibration (Section 2.8) and simulation (Section 2.9) of multivariate discrete distributions. Pairwise monotonicity is also natural from the perspective of optimization since each objective function (2.38) must either be minimized or maximized.

Similar to Subsection 2.2.1, we prove results on the properties of pairwise monotone distributions. The development of these results relies heavily on projecting the multidimensional problem down to the bivariate setting.

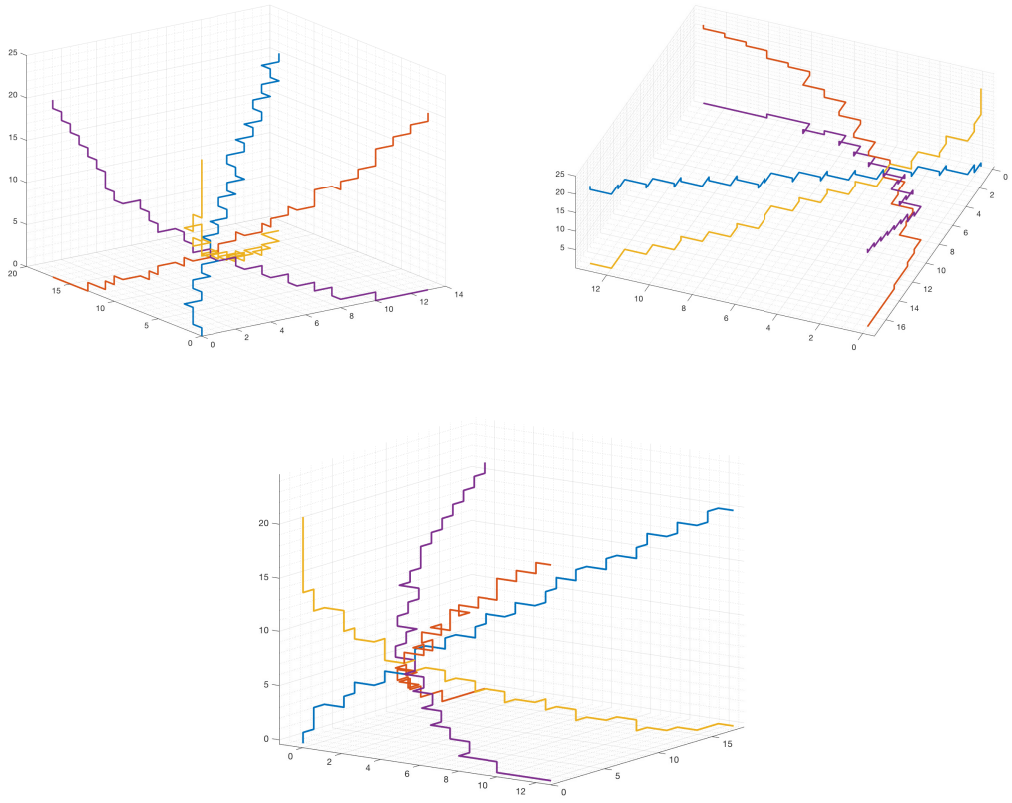

Figure 2.4: Different perspectives of the support of the 4 extreme measures for a three-dimensional joint distribution.

We have defined extreme measures as solutions to the d-dimensional multi-objective optimization problem (2.37) where each of the $j = \{1, \ldots, n\}$ for $n = 2^{d-1}$ (Lemma 15) bivariate expectations is maximized or minimized (2.38).

Figure 2.4 illustrates all of the $n = 2^{3-1} = 4$ extreme measures in the three-dimensional case from different perspectives. Note the sparsity of the supports of each extreme measure. Note also the staircase property seen in the bivariate setting is preserved in the general d-dimensional setting.

2.5 The EJD Algorithm in d-dimensions

The EJD algorithm in d-dimensions, Algorithm 5 listed on page 70, is a generalization of the EJD algorithm in the bivariate setting (Algorithm 1). This extension is non-trivial, not just computationally, but notationally as well: the complexity of coherently representing various quantities across multiple dimensions, due to the various projections, whilst keeping track of their associated extremal dependence structures necessitates additional encumbrance of the notation on the monotone structure $\mathbf{e}^{(j,d)}$ through the extreme point index j. For example, in this chapter, the notation for marginal distributions now highlights its associated monotone structure denoted by j

$$\mathbf{Q}^{(j,d;k)} = \begin{cases} [Q_0^{(j,d;k)}, \ldots, Q_{i_{\max}^{(d;k)}-1}^{(j,d;k)}, Q_{i_{\max}^{(d;k)}}^{(j,d;k)}] & \text{if} \quad e_k^{(j,d)} = 0 \\ [Q_{i_{\max}^{(d;k)}}^{(j,d;k)}, Q_{i_{\max}^{(d;k)}-1}^{(j,d;k)}, \ldots, Q_0^{(j,d;k)}] & \text{if} \quad e_k^{(j,d)} = 1. \end{cases} \tag{2.76}$$

We also define, analogous to the bivariate case, associated random variables consistently: $X^{(j,d;k)} \sim Q^{(j,d;k)}$. The additional encumbrance of the marginal distributions on the monotone structure results in a simpler and more consistent notation for marginal cdfs

$$\mathbf{F}^{(j,d;k)} = [Q_0^{(j,d;k)}, \ldots, \sum_{i=0}^{n} Q_i^{(j,d;k)}, \ldots, 1] \quad \text{where} \quad 0 \le n \le i_{\max}^{(d;k)}. \tag{2.77}$$

We also encumber the iteration counter, $l^{(j,d)}$, on the monotone structure $\mathbf{e}^{(j,d)}$, again through the extreme points index j, to reflect the dependence of the number of iterations, $l_{\max}^{(j,d)}$, of Algorithm 5 on the monotone structure in addition to the marginal distributions. Consequently, we denote the support point $l^{(j,d)}$ computed by Algorithm 5 to be $s_{l^{(j,d)}}^{(j,d)} = (s_{l^{(j,d)}}^{(j,d;1)}, \ldots, s_{l^{(j,d)}}^{(j,d;d)})$. We also index the extreme measures by the index counter $l^{(j,d)}$ which is equivalent to indexing the probabilities by the support points as they are themselves indexed by the iteration counter above. That is,

$$\hat{P}_{l^{(j,d)}}^{(j,d)} = \hat{P}_{s_{l^{(j,d)}}^{(j,d)}}^{(j,d)} = \hat{P}_{s_{l^{(j,d)}}^{(j,d;1)}, \ldots, s_{l^{(j,d)}}^{(j,d;d)}}^{(j,d)}. \tag{2.78}$$

Remark 43. *Note that, for a given $j \in \{1, \ldots, n\}$, for each $k \in \{1, \ldots, d\}$ and each $i_k \in \{0, \ldots, i_{\max}^{(d;k)}\}$,*

$$\begin{aligned} \sum_{v \in \mathcal{I}_k} \sum_{i_v=0}^{i_{\max}^{(d;k)}} \tilde{P}_{i_1, \ldots i_{k-1}, i_k, i_{k+1}, \ldots, i_d}^{(d)} &= \sum_{l \in \mathcal{A}_{i_k}} \tilde{P}_{s_l^{(j,d;1)}, \ldots, i_k, \ldots, s_l^{(j,d;d)}}^{(d)} + \sum_{l \notin \mathcal{A}_{i_k}} \tilde{P}_{s_l^{(j,d;1)}, \ldots, i_k, \ldots, s_l^{(j,d;d)}}^{(d)} \\ &= \sum_{\{l \,:\, s_l^{(j,d;k)} = i_k\}} \tilde{P}_{s_l^{(j,d;1)}, \ldots, s_l^{(j,d;d)}}^{(d)} \end{aligned} \tag{2.79}$$

where

$$\mathcal{A}_{i_k} = \{\, l : \tilde{P}_{s_l^{(j,d;1)}, \ldots, i_k, \ldots, s_l^{(j,d;d)}}^{(d)} > 0 \,\}. \tag{2.80}$$

From an intuitive and high-level perspective, Algorithm 5 exploits the structure of the problem by traversing the d-dimensional $\mathcal{S}$-path to simultaneously determine the support of $\hat{P}^{(j,d)}$ and its corresponding probabilities. This is crucial since, in d-dimensions, the exponential increase in the number of points in the domain as d increases precludes the brute force approach of repeatedly applying the EJD

theorem in d-dimensions (Theorem 4) to each point in the domain to determine the probabilities (and thereby the support) of $\hat{P}^{(j,d)}$.

In the bivariate setting, the two-dimensional version of the EJD algorithm was proved using the properties of the $\mathcal{S}$-path proved in Section 2.2. Recall that the $\mathcal{S}$-path is defined as the directed path of support points of a comonotonic or antimonotonic distribution P. While a similar definition can be made in the d-dimensional setting, we neither prove detailed results on their properties nor on the properties of multivariate discrete distributions that are coordinate-pairwise monotone. Such an approach is exceedingly tedious and unnecessary. Instead, we exploit the structure of the multi-dimensional optimization problem (2.37) and the results from the bivariate setting to prove the correctness of Algorithm 5. A particular problem structure that we exploit is the embedding of the two-dimensional problem within the d-dimensional problem. (See Remark 42.) Indeed, an analogous property holds true for the two-dimensional and the d-dimensional EJD algorithms. (See Remark 49.)

Algorithm 5, similar to Algorithm 1, relies on the joint partition of unity to simultaneously determine the probabilities and support of $\hat{P}^{(j,d)}$. The joint partition of unity extends to the d-dimensional setting in a straightforward manner and is illustrated in Figure 2.5 wherein the lines represent the partitions of unity corresponding to the marginal distributions and the joint partition (bottom-most line).

Definition 26 (Joint partition of the unit interval in d-dimensions)**.** *Let* $\Pi_{Z^{(j,d)}} = \Pi_{X^{(j,d;1)}} \vee \Pi_{X^{(j,d;2)}} \cdots \vee \Pi_{X^{(j,d;d)}} = \{z_0^{(j,d)}, z_1^{(j,d)}, \ldots, z_{l_{\max}^{(j,d)}}^{(j,d)}\}$ *denote a partition of the unit interval corresponding to the monotone structure* $\boldsymbol{e}^{(j,d)}$ *where the points* $z_i^{(j,d)}$ *are the unique ordered values of the union of the sets* $\Pi_{X^{(j,d;1)}}, \cdots, \Pi_{X^{(j,d;d)}}$ *where* $X^{(j,d;k)} \sim Q^{(j,d;k)}$ *for* $k = 1, \ldots, d$*. The point* $z_i^{(j,d)}$ *is the* $(i+1)^{\text{st}}$ *smallest value in* $\Pi_{Z^{(j,d)}}$*. That is,* $z_0^{(j,d)}$ *is the smallest element in* $\Pi_{Z^{(j,d;d)}}$*,* $z_1^{(j,d)}$ *is the second smallest element in* $\Pi_{Z^{(j,d)}}$*,* $z_2^{(j,d)}$ *is the third smallest element in* $\Pi_{Z^{(j,d)}}$*, etc. Moreover, we define* $z_{-1}^{(j,d)} = 0$*. If* $F_i^{(j,d;u)} = F_j^{(j,d;v)}$*, for some* $u \neq v$*, then there is only one* $z_{l^{(j,d)}}^{(j,d)} \in \Pi_{Z^{(j,d)}}$ *that satisfies* $z_{l^{(j,d)}}^{(j,d)} = F_i^{(j,d;u)} = F_j^{(j,d;v)}$*. That is, there are no duplicate values in* $\Pi_{Z^{(j,d)}}$ *(i.e., there is no* $z_k^{(j,d)} = z_l^{(j,d)}$ *for* $k \neq l$*).*

Remark 44. *For* $j \in \{1, \ldots, n\}$ *and* $k \in \{1, 2, \ldots, d\}$*,* $\Pi_{X^{(j,d;k)}} = \{F_0^{(j,d;k)}, F_1^{(j,d;k)}, \ldots, F_{i_{\max}^{(d;k)}}^{(j,d;k)}\}$*. It follows directly from the definition of* $\Pi_{Z^{(j,d)}}$ *that* $\boldsymbol{z}^{(j,d)} = \cup_{k=1}^{d} \{F_i^{(j,d;k)} : i = 0, 1, \ldots, i_{\max}^{(d,k)}\}$*.*

Remark 45. *For every* $z_i^{(j,2)}$ *in* Π_{Z^2} *there is a* $z_{i'}^{(j,d)}$ *in* $\Pi_{Z^{(j,d)}}$ *such that* $z_i^{(j,2)} = z_{i'}^{(j,d)}$*. Moreover, note that, from Remark 28 and Definition 26, there are no duplicate values in either* $\Pi_{Z^{(j,2)}}$ *or* $\Pi_{Z^{(j,d)}}$*.*

Algorithm 5 also works similarly to Algorithm 1 in that it begins with a preprocessing step (Algorithm 7) that constructs the marginal cdfs $\mathbf{F}^{(j,d;1)}, \ldots, \mathbf{F}^{(j,d;d)}$ consistent with the given monotone structure $\mathbf{e}^{(j,d)}$. Next, the main body of Algorithm 5 proceeds as if all the marginal distributions possessed mutual comonotonic dependence. Finally, a postprocessing step (Algorithm 6) is applied to the support $\{s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}\}$ where the k^{th} coordinate is reversed to restore the antimonotone relationship between marginal distributions 1 and k if $e_1^{(j,d)} \neq e_k^{(j,d)}$. A notable difference between the d-dimensional case and the bivariate case is that, if the k^{th} and k'^{th} marginal distribution are both antimonotone in relation to marginal distribution 1, then the k^{th} and k'^{th} marginal distributions are comonotonic with respect to each other. One way to see this is through Algorithm 6. Recall that comonotonic distributions have monotonically increasing supports and that the postprocessing step in Algorithm 6 reverses both the k^{th} and k'^{th} coordinate as they are both antimonotone with respect to the 1^{st} coordinate. After the reversal, the k^{th} and k'^{th} coordinates are monotonically increasing and, therefore, comonotone. This is explained in greater detail in Lemma 24.

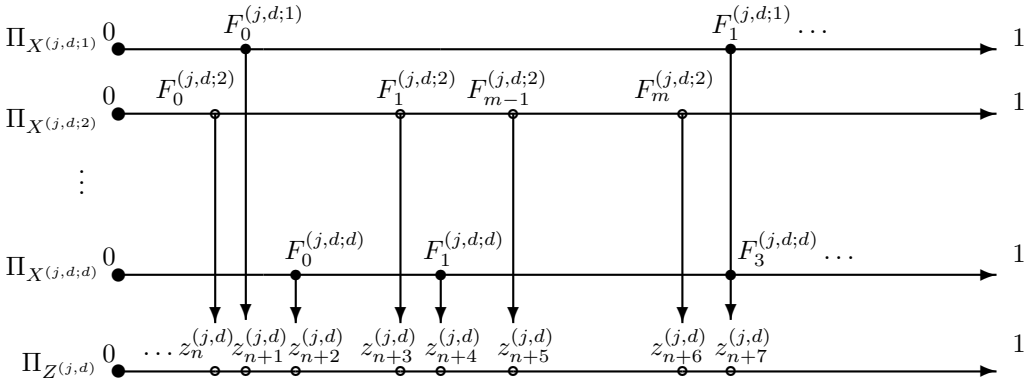

Figure 2.5: Partitions of the unit interval in the multivariate case

Our main result for this section is the correctness of the d-dimensional EJD algorithm (Algorithm 5), encapsulated in Theorem 5. Similar to the discussion of Algorithm 1 in Section 2.3, we restrict our discussion of the algorithm to finite discrete probability distributions; for each $k \in \{1, 2, \ldots, d\}$ there is an $i_{\max}^{(d;k)}$ such that $Q_i^{(j,d;k)} > 0$ for $i = 0, 1, \ldots, i_{\max}^{(d;k)}$ and $Q_i^{(j,d;k)} = 0$ for $i > i_{\max}^{(d;k)}$. The finite-dimensional approach can be extended to the case of infinite probability distributions and is sketched in Section 2.6. In addition, as noted in Remark 8 for bivariate distributions, we believe that the assumption that $Q_i^{(j,d;k)} > 0$ for $i = 0, 1, \ldots, i_{\max}^{(d;k)}$ and for $k = 1, 2, \ldots, d$, can also be relaxed for d-dimensional distributions, whether $i_{\max}^{(d;k)}$ is finite or infinite. However, as noted in Remark 8, we do not pursue this extension in this thesis.

For the remainder of the section, let the integers u and v satisfying $1 \leq u < v \leq d$ refer to the bivariate subset of coordinates from the d-dimensional set $\{1, \ldots, d\}$ that the problem is projected onto and that the u^{th} and v^{th} coordinates in the d-dimensional case correspond to the first and second coordinates in the bivariate case. That is, the marginals $Q^{(j,d;u)}$ and $Q^{(j,d;v)}$ associated with $\hat{P}^{(j,d;u,v)}$ correspond to the marginals $Q^{(j',2;1)}$ and $Q^{(j',2;2)}$ associated with $\hat{P}^{(j',2)}$ where $j' \in \{1, 2\}$ is determined by $\mathbf{e}^{(j;d;u,v)}$.

Definition 27 (Bivariate projections of the d-dimensional support)**.** *Denote by*

$$s_{l^{(j,d)}}^{(j,d;u,v)} = \text{Proj}_{u,v}\ s_{l^{(j,d)}}^{(j,d)} = (s_{l^{(j,d)}}^{(j,d;u)}, s_{l^{(j,d)}}^{(j,d;v)}) = \left(\left(s_{l^{(j,d)}}^{(j,d)}\right)_u, \left(s_{l^{(j,d)}}^{(j,d)}\right)_v\right) \tag{2.81}$$

the bivariate projection of $s_{l^{(j,d)}}^{(j,d)}$ onto its u^{th} and v^{th} coordinates. Moreover, let

$$\boldsymbol{s}^{(j,d;u,v)} = \{s_0^{(j,d;u,v)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d;u,v)}\} \tag{2.82}$$

denote the bivariate projection of the d-dimensional support $\boldsymbol{s}^{(j,d)}$ onto its u^{th} and v^{th} coordinates having duplicates in the resulting set.

Finally, note that, for a given set of marginal distributions $[Q_0^{(j,d;k)}, \ldots, Q_{i_{\max}^{(d;k)}}^{(j,d;k)}]_{k=1}^d$ and for a given monotone structure $\mathbf{e}^{(j,d)}$, Algorithm 5 computes a d-dimensional probability distribution which we

denote by $\tilde{P}^{(j,d)}$ since we have not yet shown that Algorithm 5 computes the solution, $\hat{P}^{(j,d)}$, to (2.37). In a similar vein, the support belonging to $\tilde{P}^{(j,d)}$ should also be denoted as $\tilde{\mathbf{s}}^{(j,d)}$ for consistency. However, due to our encumbered notation which will become increasingly apparent throughout this section, we use instead $\mathbf{s}^{(j,d)}$ to denote the support computed by Algorithm 5 corresponding to $\tilde{P}^{(j,d)}$.

We first sketch the proof Theorem 5: we must show that $\tilde{P}^{(j,d)}$ is the solution to the multi-objective problem (2.37), whence, $\tilde{P}^{(j,d)} = \hat{P}^{(j,d)}$. This can be accomplished by showing that the following statements $\mathbb{S}_i$, for $i = 0, 1, 2, 3, 4$, hold.

$\mathbb{S}_0$: $\tilde{P}^{(j,d)}$ is a probability distribution that satisfies the constraints (2.37b) and (2.37c).

$\mathbb{S}_1$: For every pair of integers u and v satisfying $1 \leq u < v \leq d$,

$$\tilde{P}^{(j,d;u,v)} = \hat{P}^{(j',2)} \tag{2.83}$$

where $\hat{P}^{(j',2)}$ is computed by Algorithm 1 using the marginals $Q^{(j',2;1)} = Q^{(j,d;u)}$ and $Q^{(j',2;2)} = Q^{(j,d;v)}$ and $j' \in \{1, 2\}$ is determined by $\mathbf{e}^{(j,d;u,v)}$.

$\mathbb{S}_2$: $\hat{P}^{(j',2)}$ solves the bivariate optimization problem (2.3) with $Q^{(j',2;1)} = Q^{(j,d;u)}$ and $Q^{(j',2;2)} = Q^{(j,d;v)}$.

$\mathbb{S}_3$: For a given pair of integers u and v satisfying $1 \leq u < v \leq d$, $\tilde{P}^{(j,d;u,v)}$ solves the two-dimensional optimization problem obtained from projecting (2.37) onto the u^{th} and v^{th} coordinates for the given pair of integers u and v (see Remark 42) since $\tilde{P}^{(j,d;u,v)} = \hat{P}^{(j',2)}$ and $\hat{P}^{(j',2)}$ solves the bivariate optimization problem (2.3).

$\mathbb{S}_4$: Since for each pair of integers u and v satisfying $1 \leq u < v \leq d$, $\tilde{P}^{(j,d;u,v)}$ solves the two-dimensional problem obtained by projecting (2.37) onto the u^{th} and v^{th} coordinates for that pair of u and v, $\tilde{P}^{(j,d)}$ satisfies the multi-objective optimization problem (2.37). Moreover, Theorem 4 asserts that, if problem (2.37) has a solution, it is unique, almost surely. Therefore, $\tilde{P}^{(j,d)} = \hat{P}^{(j,d)}$ almost surely.

Remark 46. *Note that this avoids the problems associated with the Pareto frontier, since* $\tilde{P}^{(j,d)} = \hat{P}^{(j,d)}$ *gives the best solution possible to each of the* $m = d(d-1)/2$ *optimization problems extremizing* $h_{u,v}^{(j,d)}(P^{(j,d)})$ *subject to the constraints* (2.37b) *and* (2.37c) *for each* u *and* v *satisfying* $1 \leq u < v \leq d$.

The rest of this section is as follows. We begin with some remarks on the structure of the solution and the corresponding algorithm (Algorithm 5) in d-dimensions and its relation to the two-dimensional setting. Then, we list a series of lemmas that both elucidate some aspect of the problem or some aspect of the projected bivariate problem and are directly used in proving Theorem 5. For the majority of the lemmas, their proofs can be found in Appendix B. These lemmas show that statements $\mathbb{S}_1, \ldots, \mathbb{S}_4$ hold.

We begin with some remarks on the lengths of the solution computed by Algorithm 5.

Remark 47. $l_{\max}^{(j,d)} \geq l_{\max}^{(j,2)}$. *Observe that there is the same number of* $z_{l^{(j,d)}}^{(j,d)}$ *as there are unique values in* $\cup_{k=1}^{d} \{F_i^{(j,d;k)} : i = 0, 1, \ldots, i_{\max}^{(d;k)}\}$ *and, similarly, that there is the same number of* $z_{l^2}^{(j,2)}$ *as there are unique values in* $\cup_{k=1}^{2} \{F_i^{(j,2;k)} : i = 0, 1, \ldots, i_{\max}^{(2;k)}\}$ *where* $F^{(j,d;u)} = F^{(j,2;1)}$ *and* $F^{(j,d;v)} = F^{(j,2;2)}$, *since*

$Q^{(j,d;u)} = Q^{(j,2;1)}$ *and* $Q^{(j,d;v)} = Q^{(j,2;2)}$. *The result follows from the fact that*

$$\cup_{k=1}^{2}\{F_i^{(j,2;k)} : i = 0, 1, \ldots, i_{\max}^{(2;k)}\} = \{F_i^{(j,d;u)} : i = 0, 1 \ldots, i_{\max}^{(d;u)}\} \cup \{F_i^{(j,d;v)} : i = 0, 1 \ldots, i_{\max}^{(d;v)}\}$$
$$\subseteq \cup_{k=1}^{d}\{F_i^{(j,d;k)} : i = 0, 1, \ldots, i_{\max}^{(d;k)}\}.$$

Remark 48. *If* $e^{(j,d;u,v)} = (0,0)$ *and* $e^{(j',d;u,v)} = (1,1)$, *then*

$$l_{\max}^{(j,d;u,v)} = l_{\max}^{(j',d;u,v)}, \tag{2.84}$$

since there is a one-to-one correspondence between $\boldsymbol{z}^{(j,2;u,v)}$ *and* $\boldsymbol{z}^{(j',2;u,v)}$ *by Lemma 16. For this reason, we use the abbreviated notation* $l_{\max}^2$ *instead of* $l_{\max}^{(j,d;u,v)}$ *or* $l_{\max}^{(j',d;u,v)}$ *throughout most of this section when discussing the cases* $e^{(j,d;u,v)} = (0,0)$ *and* $e^{(j',d;u,v)} = (1,1)$ *and we use* $l^2 \in \{0, 1, \ldots, l_{\max}^2\}$ *to index the sets* $\underline{\tilde{\boldsymbol{s}}}^{(j,d;u,v)}$, $\underline{\tilde{\boldsymbol{s}}}^{(j',d;u,v)}$, $\underline{\boldsymbol{s}}^{(j,d;u,v)}$, $\underline{\boldsymbol{s}}^{(j',d;u,v)}$, $\boldsymbol{z}^{(j,2;u,v)}$, *and* $\boldsymbol{z}^{(j',2;u,v)}$. *Moreover, we also use* $l_{\max}^2$ *in place of* $l_{\max}^{(j,d;u,v)}$, *in the case of* $\boldsymbol{e}^{(j,d;u,v)} = (0,0)$, *to simplify the notation whether or not the dual index* j' *is present.*

The sets $\mathbf{z}^{(j,2;u,v)}$, and $\mathbf{z}^{(j',2;u,v)}$ are defined below on page 63. The sets $\underline{\tilde{\boldsymbol{s}}}^{(j,d;u,v)}$, $\underline{\tilde{\boldsymbol{s}}}^{(j',d;u,v)}$, $\underline{\mathbf{s}}^{(j,d;u,v)}$, and $\underline{\mathbf{s}}^{(j',d;u,v)}$ are defined below on page 65.

Remark 49. *Algorithm 5 reduces to Algorithm 1 in the case* $d = 2$.

That Remark 49 holds can been seen as follows. The EJD algorithms 1 and 5 can each be divided into three main phases: preprocessing, main body, and postprocessing. This corresponds to Lines 1-5, Lines 6-18, and Lines 19-21 in Algorithm 1 and Lines 1-5, Lines 6-14, and Lines 15-16 in Algorithm 5. We begin with the preprocessing steps. If $d = 2$, the inputs are equivalent in Algorithm 1 and Algorithm 5 and Lines 1-5 in Algorithm 5 are equivalent to Lines 1-5 in Algorithm 1.

Next, the main bodies of Algorithm 5 and Algorithm 1, while having a different form, are also equivalent. To see this, first note that, if $d = 2$, Lines 13-14 of Algorithm 5 can be clearly seen to be equivalent to Lines 17-18 of Algorithm 1. Consequently, it remains to show that Lines 8-16 of Algorithm 1 are equivalent to Lines 8-12 of Algorithm 5. Note that the conditions in the if statements on Lines 8, 11, and 14 of Algorithm 1 and their corresponding statement bodies can be refactored according to the dimensions. For example, note that in Algorithm 1, if $z_{l-1} == F^{(1)}_{s_{l-1}^{(j,2;1)}}$ in any of the if statement conditions, then $s_l^{(j,2;1)} = s_{l-1}^{(j,2;1)} + 1$ in each of the corresponding statement bodies. Similarly for $F^{(2)}_{s_{l-1}^{(j,2;2)}}$ and $s_l^{(j,2;2)}$. Then, note that, for the case $d = 2$ in Algorithm 5, for each iteration, l, of the algorithm, there are two iterations of the for loop (Line 8). For $k = 1$, the if statement (Line 9) checks that $z_{l-1} == F^{(1)}_{s_{l-1}^{(1)}}$ and if true, increments the corresponding support point, that is, $s_l^{(j,2;1)} = s_{l-1}^{(j,2;1)} + 1$ is executed (Line 10). Otherwise, the corresponding support point is not incremented (Line 12). Similarly for $F^{(2)}_{s_{l-1}^{(j,2;2)}}$ and $s_l^{(j,2;2)}$. Therefore the main bodies of Algorithm 1 and Algorithm 5 are equivalent for $d = 2$.

Finally, to see that the postprocessing stage of Algorithm 1 is equivalent to Algorithm 5, note that, in Algorithm 1, the postprocessing subroutine, Algorithm 2, is only called in the antimonotone case (Line 19). In contrast, Algorithm 5 calls the multidimensional postprocessing subroutine, Algorithm 6, immediately after the main body has executed since, in the general case, each pair of components

can either be comonotone or antimonotone. For each dimension, k, Line 4 of Algorithm 6 checks the monotonic relationship between the pair of components $(1, k)$; if $d = 2$, Line 4 of Algorithm 6 checks whether it is the comonotone or antimonotone case, similar to Line 19 of Algorithm 1. If, for the pair of components $(1, k)$, the antimonotone case is true, Line 5 is executed, reversing the k^{th} component of the support similar to Line 4 of Algorithm 1. Otherwise, Line 7 is executed and the k^{th} component of the support is not reversed, similar to Line 3 of Algorithm 1. Therefore, the postprocessing phases of Algorithm 5, with $d = 2$, is equivalent to Algorithm 1. Since, for $d = 2$, the preprocessing, main body, and post processing phases of Algorithm 5 and Algorithm 1 are equivalent, Algorithm 5 reduces to Algorithm 1 in the case $d = 2$.

Next, we consider an important partition of unity that will be needed later in the chapter: the partition of unity corresponding to the two-dimensional joint probability distribution having monotone structure $\mathbf{e}^{(j,d;u,v)} = (0, 0)$

$$\mathbf{z}^{(j,2;u,v)} = \{z_0^{(j,2;u,v)}, \ldots, z_{l^2_{\max}}^{(j,2;u,v)}\} \tag{2.85}$$

where $l^2_{\max} = l^{(j,d;u,v)}_{\max}$ was introduced in Remark 48 to simplify the notation in this section. Note that $\mathbf{z}^{(j,2;u,v)}$ can be constructed similarly to the set $\{z_{\tilde{l}_0}^{(j,d)}, \ldots, z_{\tilde{l}_{\max}}^{(j,d)}\}$ (which can be found in the setup of the proof of Lemma 24 in Appendix B.7 on page 196) by choosing a set of indices $\{\tilde{l}_0, \ldots, \tilde{l}_{\max}\} \subset \{l_0^{(j,d)}, \ldots, l_{\max}^{(j,d)}\}$ such that the following hold

$$\{z_0^{(j,2;u,v)}, \ldots, z_{l^2_{\max}}^{(j,2;u,v)}\} = \{z_{\tilde{l}_0}^{(j,d)}, \ldots, z_{\tilde{l}_{\max}}^{(j,d)}\} = \cup_{k=1}^{2} \{F_i^{(j,2;k)} : i = 0, 1, \ldots, i_{\max}^{(2;k)}\}. \tag{2.86}$$

Clearly, we can assume that the subscripts on the z's above are chosen so that, for all $l^2 \in \{1, \ldots, l^2_{\max}\}$,

$$z_{l^2-1}^{(j,2;u,v)} < z_{l^2}^{(j,2;u,v)}. \tag{2.87}$$

Remark 50. *Moreover, the set $\{z_0^{(j,2;u,v)}, \ldots, z_{l^2_{\max}}^{(j,2;u,v)}\}$ is equivalent to the two-dimensional partition of unity, $\boldsymbol{z}^{(j,2)}$, constructed from $F^{(j,d;u)}$ and $F^{(j,d;v)}$ since (2.86) corresponds to Line 2 of Algorithm 8 for $d = 2$, $u = 1$, and $v = 2$, and (2.87) corresponds to Line 4 of Algorithm 8. Taking only unique elements of $\{z_0^{(j,d;u,v)}, \ldots, z_{l^{(j,d)}_{\max}}^{(j,d;u,v)}\}$ is equivalent to Line 3 of Algorithm 8.*

Remark 51. *In some instances, it is useful to prepend the two-dimensional partition of unity (2.85) with the element $z_{-1}^{(j,2;u,v)} = 0 = F_{-1}^{(j,2,d;u)} = F_{-1}^{(j,2,d;v)}$. Then, we can write (2.86) as*

$$\boldsymbol{z}^{(j,2;u,v)} = \{F_{-1}^{(j,d;u)}, F_0^{(j,d;u)}, \ldots, F_{i_{max}^{(d;u)}}^{(j,d;u)}\} \cup \{F_{-1}^{(j,d;v)}, F_0^{(j,d;v)}, \ldots, F_{i_{max}^{(d;v)}}^{(j,d;v)}\} \tag{2.88}$$

where $F_{-1}^{(j,d;k)} = 0$ for $k \in \{u, v\}$. Note that this is consistent with how the d-dimensional partition of unity is defined in Definition 26.

Remark 52. *For all $j \in \{1, 2, \ldots, n\}$ and all u, v satisfying $1 \le u < v \le d$, the set $\{z_0^{(j,2;u,v)}, \ldots, z_{l^2_{\max}}^{(j,2;u,v)}\}$ is embedded within the set $\{z_0^{(j,d)}, \ldots, z_{l^{(j,d)}_{\max}}^{(j,d)}\}$. More specifically, for all $l^2 \in \{0, 1, \ldots, l^2_{\max}\}$ there exists a unique $l^{(j,d)} \in \{0, 1, \ldots, l^{(j,d)}_{\max}\}$ such that $z_{l^2}^{(j,2;u,v)} = z_{l^{(j,d)}}^{(j,d)}$. This can be seen easily from the construction of the set $\boldsymbol{z}^{(j,2;u,v)}$ in Remark 50 and from the definition of the set $\boldsymbol{z}^{(j,d)}$ in Definition 26, which also ensures that there are no duplicate values in the set $z^{(j.d)}$.*

Lemma 16. *For any $k \in \{1, 2, \ldots, d\}$ and any $i \in \{-1, 0, 1, \ldots, i_{\max}^{(d;k)}\}$,*

$$1 - F_i^{(j',d;k)} = F_{i_{\max}^{(d;k)} - i - 1}^{(j,d;k)}, \tag{2.89}$$

where $e_k^{(j,d)} = 0$ and $e_k^{(j',d)} = 1$.

The proof of Lemma 16 can be found in Appendix B.1.

Lemma 17. *For any $l \in \{-1, 0, 1, \ldots, l_{\max}^2\}$ and $1 \leq u < v \leq d$,*

$$1 - z_l^{(j',2;u,v)} = z_{l_{\max}^2 - l - 1}^{(j,2;u,v)}, \tag{2.90}$$

where $\boldsymbol{e}^{(j,d;u,v)} = (0, 0)$, $\boldsymbol{e}^{(j',d;u,v)} = (1, 1)$, and we have prepended $\boldsymbol{z}^{(j,2;u,v)}$ with the element $z_{-1}^{(j,2;u,v)} = 0$ and $\boldsymbol{z}^{(j',2;u,v)}$ with the element $z_{-1}^{(j',2;u,v)} = 0$. (Remark 2.88.) Moreover, for each $l \in \{-1, 0, 1, \ldots, l_{\max}^2\}$, there exists a $k \in \{u, v\}$ and an $i \in \{-1, 0, 1, \ldots, i_{\max}^{(d;k)}\}$ such that

$$z_l^{(j',2;u,v)} = F_i^{(j',d;k)} \tag{2.91}$$

and

$$z_{l_{\max}^2 - l - 1}^{(j,2;u,v)} = F_{i_{\max}^{(d;k)} - i - 1}^{(j,d;k)} = 1 - F_i^{(j',d;k)}, \tag{2.92}$$

where the k's and the i's are the same in (2.91) and (2.92) above.

The proof of Lemma 17 can be found in Appendix B.2

Lemma 18. *The order in which $F^{(j,d;u)}$ and $F^{(j,d;v)}$ are matched in Algorithm 5 in the $e^{(j,d;u,v)} = (0, 0)$ case is reversed in Algorithm 5 in the $e^{(j',d;u,v)} = (1, 1)$ case. That is, for each $l^2 \in \{0, \ldots, l_{\max}^2\}$, at least one of*

$$z_{l^2}^{(j,2;u,v)} = F_{i^{(d;u)}}^{(j,d;u)} \tag{2.93}$$

or

$$z_{l^2}^{(j,2;u,v)} = F_{i^{(d;v)}}^{(j,d;v)} \tag{2.94}$$

holds for some $i^{(d;u)} \in \{0, 1, \ldots, i_{\max}^{(d;u)}\}$ or $i^{(d;v)} \in \{0, 1, \ldots, i_{\max}^{(d;v)}\}$. Moreover, (2.93) holds iff

$$z_{l_{\max}^2 - l^2 - 1}^{(j',2;u,v)} = F_{i_{\max}^{(d;u)} - i^{(d;u)} - 1}^{(j',d;u)} \tag{2.95}$$

holds, and (2.94) holds iff

$$z_{l_{\max}^2 - l^2 - 1}^{(j',2;u,v)} = F_{i_{\max}^{(d;v)} - i^{(d;v)} - 1}^{(j',d;v)} \tag{2.96}$$

holds.

PROOF: That at least one of (2.93) or (2.94) holds can be seen from the construction of $\mathbf{z}^{(j,2;u,v)}$ which, by (2.86) and Remark 50, ensures that it satisfies $\mathbf{z}^{(j,2;u,v)} = F^{(j,d;u)} \bigcup F^{(j,d;v)}$.

Next, note that (2.90), which relates $z_{l^2}^{(j,2;u,v)}$ to $z_{l_{\max}^2 - l^2 - 1}^{(j',2;u,v)}$, and (2.89) in Lemma 16, which relates $F_i^{(j,d;k)}$ and $F_{i_{\max}^{(d;k)} - i - 1}^{(j',d;k)}$ where $k \in \{u, v\}$, can both be applied to obtain our desired result. To see this, start by assuming that (2.93) holds for some $l^2 \in \{0, 1, \ldots, l_{\max}^2\}$ and some $i^{(d;u)} \in \{0, 1, \ldots, i_{\max}^{(d;u)}\}$. Subtracting both sides of (2.93) from 1 yields

$$1 - z_{l^2}^{(j,2;u,v)} = 1 - F_{i^{(d;u)}}^{(j,d;u)}. \tag{2.97}$$

Making a suitable change of variable in (2.90), rearranging, and applying the result to the left side of (2.97) and making a similar change of variable in (2.89), rearranging and applying the result to the right side of (2.97) gives

$$z^{(j',2;u,v)}_{l^2_{\max}-l^2-1} = F^{(j',d;u)}_{i^{(d;u)}_{\max}-i^{(d;u)}-1}. \tag{2.98}$$

Thus, we have shown that, if (2.93) holds, then (2.95) holds. A similar argument shows that, if (2.95) holds, then (2.93) holds. Thus, (2.93) holds iff (2.95) holds.

A similar argument shows that (2.94) holds iff (2.96) holds. □

Next, we list some results on the support computed by Algorithm 5. Particularly useful is the counting interpretation of the support.

Lemma 19. *Given $j \in \{1,\ldots,n\}$, for each $k \in \{1,\ldots,d\}$ and each $i \in \{0,\ldots,i^{(d;k)}_{\max}\}$, there exists an integer w such that $s^{(j,d;k)}_w = i$.*

PROOF: Given $j \in \{1,\ldots,n\}$, for each $k \in \{1,\ldots,d\}$, $s^{(j,d;k)}_0 = 0$ by Line 4 of Algorithm 5. Line 10 of Algorithm 5 increments $s^{(j,d;k)}_{l^{(j,d)}}$ by 1 only if Line 9 of Algorithm 5 evaluates to true. For each $k = \{1,\ldots,d\}$, Line 9 of Algorithm 5, evaluates to true $i^{(d;k)}_{\max}+1$ times since the partition of unity, computed by Algorithm 8 on Line 2 of Algorithm 5, is obtained by taking the unique elements of the union of the marginal cdfs. Since, for $l^{(j,d)} \in \{0,1,\ldots,l^{(j,d)}_{\max}\}$, $s^{(j,d;k)}_{l^{(j,d)}}$ must take on the values $0,\ldots,i^{(d;k)}_{\max}$, therefore, there must exist an integer w such that $s^{(j,d;k)}_w = i$. □

Corollary 1. *For each $k \in \{1,\ldots,d\}$ and each $i \in \{0,\ldots,i^{(d;k)}_{\max}\}$, let q be the smallest integer such that $s^{(j,d;k)}_q = i$ and w be the greatest integer such that $s^{(j,d;k)}_w = i$. If $q < w$, then $s^{(j,d;k)}_l = i$ for all l satisfying $q \leq l \leq w$.*

Let us denote by $\tilde{\mathbf{s}}^{(j,d)}$ the support corresponding to $\tilde{P}^{(j,d)}$ before Algorithm 6 is applied. The underline denotes when applied to a support set denotes removing the duplicates in the support set. Thus, the set $\underline{\tilde{\mathbf{s}}}^{(j,d)}$ is equivalent to the set $\tilde{\mathbf{s}}^{(j,d)}$ since there are no duplicates in the $\tilde{\mathbf{s}}^{(j,d)}$. However, $\underline{\mathbf{s}}^{(j,d;u,v)} = \{\underline{s}^{(j,d;u,v)}_0,\ldots,\underline{s}^{(j,d;u,v)}_{l^2_{\max}}\}$ is the support corresponding to $\tilde{P}^{(j,d;u,v)}$ having monotone structure $\mathbf{e}^{(j,d;u,v)} = (0,0)$ obtained by projecting each element of the support $\mathbf{s}^{(j,d)} = \{s^{(j,d)}_0,\ldots,s^{(j,d)}_{l^{(j,d)}_{\max}}\}$ of $\tilde{P}^{(j,d)}$ generated by Algorithm 5 onto the u^{th} and v^{th} coordinates and eliminating duplicates, but retaining the order by which the support points are generated by Algorithm 5. Similarly, let $\underline{\mathbf{s}}^{(j',d;u,v)} = \{\underline{s}^{(j',d;u,v)}_0,\ldots,\underline{s}^{(j',d;u,v)}_{l^2_{\max}}\}$ denote the support corresponding to $\tilde{P}^{(j',d;u,v)}$ having monotone structure $\mathbf{e}^{(j',d;u,v)} = (1,1)$ obtained by projecting each element of the support $\mathbf{s}^{(j',d)} = \{s^{(j',d)}_0,\ldots,s^{(j',d)}_{l^{(j,d)}_{\max}}\}$ of $\tilde{P}^{(j',d)}$ onto the u^{th} and v^{th} coordinates and eliminating duplicates, but retaining the order by which the support points are generated by Algorithm 5.

Remark 53. *For all $j \in \{1,2,\ldots,n\}$, for all u,v satisfying $1 \leq u < v \leq d$, the endmost point in the sets $\underline{\tilde{\boldsymbol{s}}}^{(j,d;u,v)}$ and $\tilde{\boldsymbol{s}}^{(j,d;u,v)}$ must be the same since the set $\underline{\tilde{\boldsymbol{s}}}^{(j,d;u,v)}$ is obtained directly from the set $\tilde{\boldsymbol{s}}^{(j,d;u,v)}$ by eliminating duplicates, but retaining the order by which the support points are generated by Algorithm 5. Whence, $\tilde{s}^{(j,d;u,v)}_{l^{(j,d;u,v)}_{\max}} = \tilde{s}^{(j,d;u,v)}_{l^{(j,d)}_{\max}}$. Moreover, it must also be that $\tilde{s}^{(j,d;u)}_{l^{(j,d;u,v)}_{\max}} = i^{(d;u)}_{\max}$, since the set $\{F^{(j,d;u)}_0,\ldots,F^{(j,d;u)}_{i^{(d;u)}_{\max}}\} \subset \boldsymbol{z}^{(j,d)}$ by Line 2 of Algorithm 5, whence Lines 6-12 of Algorithm 5 imply that $\tilde{s}^{(j,d;u)}_{l^{(j,d)}_{\max}} = i^{(d;u)}_{\max}$. Similarly, $\tilde{s}^{(j,d;v)}_{l^{(j,d;u,v)}_{\max}} = i^{(d;v)}_{\max}$.*

Lemma 20. *For all $j \in \{1,2,\ldots,n\}$, all $k \in \{1,2,\ldots,d\}$, and all $l^{(j,d)} \in \{0,1,\ldots,l^{(j,d)}_{\max}\}$, $\tilde{s}^{(j,d;k)}_{l^{(j,d)}}$ is the number of $i \in \{0,1,\ldots,i^{(d;k)}_{\max}\}$ such that $F^{(j,d;k)}_i = z^{(j,d)}_l$ for some $l \in \{0,1,\ldots,l^{(j,d)}-1\}$.*

PROOF: This is evident from the fact that Line 4 of Algorithm 5 initializes $\tilde{s}_0^{(j,d;k)} = 0$ for all $j \in \{1, 2, \ldots, n\}$ and for all $k \in \{1, 2, \ldots, d\}$ and Line 10 of Algorithm 5 executes if and only if Line 9 of Algorithm 5 evaluates to true. □

A variant of Lemma 20 also applies to the set $\underline{\tilde{\mathbf{s}}}^{(j,d;u,v)}$ as it is obtained directly from the set $\tilde{\mathbf{s}}^{(j,d;u,v)}$.

Corollary 2. *For all $j \in \{1, 2, \ldots, n\}$, all $k \in \{u, v\}$, where u, v are such that $1 \leq u < v \leq d$, and all $l^2 \in \{0, 1, \ldots, l_{\max}^{(j,d;u,v)}\}$, $\tilde{\underline{s}}_{l^2}^{(j,d;k)}$ is the number of $i \in \{0, 1, \ldots, i_{\max}^{(d;k)}\}$ such that $F_i^{(j,d;k)} = z_l^{(j,2;u,v)}$ for some $l \in \{0, 1, \ldots, l^2 - 1\}$.*

The proof of Corollary 2 can be found in Appendix B.4.

Remark 54. *For all $j \in \{1, 2, \ldots, n\}$, for all u, v satisfying $1 \leq u < v \leq d$, and for each $l^2 \in \{0, 1, \ldots, l_{\max}^2\}$, there is at least one $l^{(j,d)} \in \{0, 1, \ldots, l_{\max}^{(j,d)}\}$ such that*

$$\tilde{\underline{s}}_{l^2}^{(j,d;u,v)} = \tilde{s}_{l^{(j,d)}}^{(j,d;u,v)} \tag{2.99}$$

since the set $\underline{\tilde{\boldsymbol{s}}}^{(j,d;u,v)}$ is obtained directly from the set $\tilde{\boldsymbol{s}}^{(j,d;u,v)}$ by eliminating duplicates, but retaining the order by which the support points are generated by Algorithm 5. Note that, in some cases, for a given l^2, (2.99) holds for more than one $l^{(j,d)}$.

Finally, the results that directly lead to the proof of our main result are as follows.

Lemma 21. *If u and v are such that $1 \leq u < v \leq d$ and j and $j' \in \{1, 2, \ldots, n\}$ are such that $e^{(j,d;u,v)} = (0, 0)$ and $e^{(j',d;u,v)} = (1, 1)$, then for all $l^2 \in \{0, 1, \ldots, l_{\max}^2\}$*

$$\underline{s}_{l^2}^{(j,d;u,v)} = \underline{s}_{l_{\max}^2 - l^2}^{(j',d;u,v)}. \tag{2.100}$$

The proof of Lemma 21 can be found in Appendix B.6.

Lemma 22. *For $l^2 \in \{0, 1, \ldots, l_{\max}^2\}$,*

$$\tilde{P}_{l^2}^{(j,d;\,u,v)} = \hat{P}_{l^2}^{(j',2)} \tag{2.101}$$

where $j = 1$ and $j' = 1$ and where $\hat{P}_{l^2}^{(j,d)}$ is computed by Algorithm 5 and $\hat{P}_{l^2}^{(j',2)}$ is computed by Algorithm 1.

The proof of Lemma 22 can be found in Appendix B.5

Lemma 23. *The $\tilde{P}^{(j,d)}$ computed by Algorithm 5 is a probability distribution that satisfies the constraints (2.37b) and (2.37c).*

The proof of Lemma 23 can be found in Appendix B.3

Lemma 24. *The bivariate probability distribution $\tilde{P}^{(j,d;\,u,v)}$ obtained by marginalizing, according to (2.43), the d-dimensional probability distribution $\tilde{P}^{(j,d)}$ having the monotone structure $\boldsymbol{e}^{(j,d)}$ obtained from Algorithm 5 is identical to the bivariate extreme measure $\hat{P}^{(j',2)}$ obtained from Algorithm 1 where j' is determined by $\boldsymbol{e}^{(j,d;u,v)}$. The marginals $Q^{(j,d;u)}$ and $Q^{(j,d;v)}$ associated with $\tilde{P}^{(j,d;u,v)}$ and the marginals $Q^{(j',2;1)}$ and $Q^{(j',2;2)}$ associated with $\hat{P}^{(j',2)}$ satisfy $Q^{(j,d;u)} = Q^{(j',2;1)}$ and $Q^{(j,d;v)} = Q^{(j',2;2)}$.*

The proof of Lemma 24 can be found in Appendix B.7.

The results in the section up to this point prove our main result. We formalize this in Theorem 5 below.

Theorem 5. *The probability $\tilde{P}^{(j,d)}$ computed by Algorithm 5 is correct in the sense that it is the almost sure unique solution to the optimization problem (2.37).*

PROOF: We can now show the correctness of Algorithm 5 by showing that all of the statements, $\mathbb{S}_i$ for $i = 0, 1, 2, 3, 4$, listed on page 61, near the beginning of this section, hold true.

We begin by noting that statements $\mathbb{S}_0$ and $\mathbb{S}_1$ follow directly from Lemma 23 and Lemma 24, respectively.

Statement $\mathbb{S}_2$ states that the two-dimensional probability distribution, $\hat{P}^{(j',2)}$, computed by Algorithm 1 is a solution to the two-dimensional optimization problem (2.3), which was the main result shown in Section 2.2.

Statement $\mathbb{S}_3$ holds true since $\hat{P}^{(j',2)}$ solves the bivariate problem (2.3) and since (2.83) was shown to be true in $\mathbb{S}_1$, $\tilde{P}^{(j,d;u,v)}$ also solves (2.3). Since, by Remark 42, the two-dimensional optimization problem obtained from projecting (2.37) onto the u^{th} and v^{th} coordinates for the given pair of integers u and v is equivalent to the bivariate optimization problem, therefore, $\tilde{P}^{(j,d;u,v)}$ solves the two-dimensional optimization problem obtained from (2.37) by projection.

Finally, statement $\mathbb{S}_4$ states that the solution, $\tilde{P}^{(j,d)}$, computed by Algorithm 5 is a solution to the d-dimensional multi-objective problem (2.37) and that $\tilde{P}^{(j,d)}$ is in fact the unique solution, almost surely. That is, $\tilde{P}^{(j,d)} = \hat{P}^{(j,d)}$. The first part of statement $\mathbb{S}_4$ can be seen as follows. We know from $\mathbb{S}_3$ that, for any pair of integers u and v satisfying $1 \leq u < v \leq d$, $\tilde{P}^{(j,d)}$ projected onto the u^{th} and v^{th} coordinates, $\tilde{P}^{(j,d;u,v)}$, is a solution to the corresponding two-dimensional problem (2.3). By the results in Section 2.3, $\tilde{P}^{(j,d;u,v)}$ is the unique solution. Since the multi-objective problem (2.37) consists of $m = d(d-1)/2$ objective functions, all of which are satisfied, and since by statement $\mathbb{S}_0$, $\tilde{P}^{(j,d)}$ satisfies the constraints (2.37b) and (2.37c), the first part of statement $\mathbb{S}_4$ holds true. Since we have shown that $\tilde{P}^{(j,d)}$ is a solution to the multi-objective problem (2.37) and we know from Theorem 4 that the solution to (2.37) is unique almost surely, the second part of $\mathbb{S}_4$ also holds true almost surely. Therefore, we have shown the correctness of Algorithm 5.

□

2.6 The Infinite Dimensional Case

The focus of the thesis is on developing the EJD approach in the practical setting of finite-dimensional marginal distributions that are approximations, typically via truncations and ensuring the resulting probability mass sums to one, of the true marginal distributions in the infinite (space) dimensional setting. Two natural questions to consider are whether the EJD approach can be extended to the setting of infinite-dimensional marginals and the effect of using the finite approximations (2.20) on the precision of the resulting correlation structure. Since the focus of this thesis is on the finite-dimensional setting, we leave the detailed analysis of the affect of finite approximations on the precision of correlation structures as future work. Instead, we briefly outline below, at a high level, how Algorithm 1 can be extended to the case of infinite marginal distributions. We believe that Algorithm 5 can also be extended to the case of infinite marginal distributions, but we leave that task to future work.

The Comonotone Case

In the comonotone case, the extension of Algorithm 1 is straightforward. Since the initial point of the $\mathcal{S}$-path is $(0,0)$ by Lemma 1, we only have to determine a suitable stopping criteria. One such criteria is to stop the algorithm at iteration l if

$$1 - F_l^{(1)} < \epsilon^{(1)} \quad \text{and/or} \quad 1 - F_l^{(2)} < \epsilon^{(2)} \tag{2.102}$$

for some $\epsilon^{(j)} \in \mathbb{R}_+$ for $j \in \{1,2\}$. In addition, instead of using finite approximations $\hat{Q}^{(1)}$ and $\hat{Q}^{(2)}$ computed in (2.20) we use instead the infinite probability distributions $Q^{(1)}$ and $Q^{(2)}$, without truncating or approximating them. However, since the infinite dimensional marginal distributions $Q^{(1)}$ and $Q^{(2)}$ cannot be passed as arguments into Algorithm 1, Algorithm 1 must be modified such that $Q_i^{(1)}$ and $Q_j^{(2)}$ can be computed as needed in Lines 2-18 of Algorithm 1. Similarly for the infinite dimensional marginal cdfs $F^{(1)}$ and $F^{(2)}$ and the $z's$ in Line 4 of Algorithm 1.

The Antimonotone Case

The antimonotone case differs from the comonotone case in two main ways that complicate a direct extension of Algorithm 1 to the case of infinite-dimensional marginal distributions. The first is the fact that the antimonotone case has a preprocessing step (Lines 2-5 of Algoritm 3 called on Line 1 of Algorithm 1) that reverses the marginal distribution $Q^{(2)}$—it is meaningless to "reverse" the full infinite-dimensional marginal distribution. The second is the fact that the solution $\hat{P}^{(2)}$ to (2.3) has antimonotonic support (Lemma 11) and that, by Lemma 5, there exists points $(0,n)$ and $(m,0)$, for some integers m and n, that belong to the $\mathcal{S}$-path corresponding to an antimonotone distribution. Moreover, we know by Remark 18 that, if $(0,n)$ and $(m,0)$ belong to the $\mathcal{S}$-path, then so must $(0,n+1),(0,n+2),\ldots$ and $(m+1,0),(m+2,0),\ldots$ in the infinite-dimensional case. That is, all but a finite number of support points reside on the axes in the infinite-dimensional antimonotone case. Thus, unlike the infinite-dimensional extension of Algorithm 1 in the comonotone case, the infinite-dimensional extension in the antimonotone case only needs to compute a finite number of points on the $\mathcal{S}$-path similar to the finite-dimensional case since any support point on either axis can be computed by using formula (2.5) of Theorem 1 directly. In essence, there is no infinite-dimensional case for the antimonotone setting.

While it may be possible to modify Algorithm 1 such that it works in the antimonotone case without

having to reverse a full infinite-dimensional (marginal) distribution and use infinitely many $Q^{(2)}$ elements, this extension, as mentioned in the beginning of this section, is not a focus of this thesis. However, as an alternative, the following approach can be used for infinite-dimensional $Q^{(1)}$ and $Q^{(2)}$.

1. Start with a point $(0, n)$ on the $\mathcal{S}$-path. Note that this point can be found easily as follows. Pick any point $(0, n)$ and use formula (2.5) of Theorem 1 to determine if it is a support point. If not, double the value of n, denote it by n' and use Theorem 1 again to check if it is a support point. This can be repeated until we find some point $(0, n')$ where $n' > n$ that is a support point.

2. If the point (m', n') is on the $\mathcal{S}$-path, then we know from the results in Section 2.2 that one of $(m'+1, n')$, $(m', n'-1)$, or $(m'+1, n'-1)$ must be on the $\mathcal{S}$-path. Moreover, $(m'+1, n')$ and $(m', n'-1)$ cannot both be on the $\mathcal{S}$-path. Therefore, the next point on the $\mathcal{S}$-path is one of the $(m'+1, n')$ or $(m', n'-1)$ if either of them is a support point. Otherwise, the next point on the $\mathcal{S}$-path is $(m'+1, n'-1)$.

3. Determine the next point on the $\mathcal{S}$-path and its associated probability, using formula (2.5) of Theorem 1 to compute $\hat{P}^{(2)}_{(m'+1,n')}$, $\hat{P}^{(2)}_{(m',n'-1)}$ and/or $\hat{P}^{(2)}_{(m'+1,n'-1)}$ as needed.

   (a) If $\hat{P}^{(2)}_{(m'+1,n')} > 0$, then $(m'+1, n')$ is the next point on the $\mathcal{S}$-path, and $\hat{P}^{(2)}_{(m'+1,n')}$ is the associated probability.

   (b) Else if $\hat{P}^{(2)}_{(m',n'-1)} > 0$, then $(m', n'-1)$ is the next point on the $\mathcal{S}$-path, and $\hat{P}^{(2)}_{(m',n'-1)}$ is the associated probability.

   (c) Otherwise, $(m'+1, n'-1)$ is the next point on the $\mathcal{S}$-path, and $\hat{P}^{(2)}_{(m'+1,n'-1)}$ is the associated probability.

4. Update the current point on the $\mathcal{S}$-path to the newly determined point and repeat Steps 2–3.

5. Continue this procedure until the point $(m, 0)$ is reached. By the results in Section 2.2, this procedure must terminate at $(m, 0)$ after finitely many steps.

Remark 55. *For the computation of the correlation coefficient corresponding to $\hat{P}^{(2)}$, the probabilities associated with axes support points need not be computed. To see this, note that the bivariate expectation is computed as the sum $\sum_{i=0}^{\infty}\sum_{j=0}^{\infty} ij\hat{P}^{(2)}_{ij}$. Obviously, we only need to include in this sum the terms $ij\hat{P}^{(2)}_{i,j}$ for which i, j and $\hat{P}^{(2)}_{i,j}$ are all nonzero. For anly point on the axes, either $i = 0$ or $j = 0$ and, for any point not on the $\mathcal{S}$-path, $\hat{P}^{(2)}_{i,j} = 0$. Hence, we need to include in the sum only the points associated with the $\mathcal{S}$-path between $(0, n)$ and $(m, 0)$.*

Algorithm 5 Extreme Joint Distribution Algorithm in d-dimensions

Require: Marginal distributions: $[Q_0^{(d;k)}, \dots, Q_{i_{\max}^{(d;k)}}^{(d;k)}]_{k=1}^{d}$

Monotonicity Structures: $\mathbf{e}^{(j,d)}$ where $j \in \{1, \dots, n\}$

Output: Extreme measure $[\hat{P}_0^{(j,d)}, \dots, \hat{P}_{l_{\max}^{(j,d)}}^{(j,d)}]$ and its support $[s_0^{(j,d)}, \dots, s_{l_{\max}^{(j,d)}}^{(j,d)}]$

1: $[F_0^{(j,d;1)}, \dots, F_{i_{max}^{(d;1)}}^{(j,d;1)}], [F_0^{(j,d;2)}, \dots, F_{i_{max}^{(d;2)}}^{(j,d;2)}], \dots,$
$[F_0^{(j,d;d)}, \dots, F_{i_{max}^{(d;d)}}^{(j,d;d)}] \leftarrow$ CONSTRUCT_MARGINAL_CDFS_D($[Q_0^{(d;k)}, \dots, Q_{i_{\max}^{(d;k)}}^{(d;k)}]_{k=1}^{d}, [\mathbf{e}^{(j,d;1)}, \dots, \mathbf{e}^{(j,d;d)}]$)

2: $z_0^{(j,d)}, \dots, z_{l_{\max}^{(j,d)}}^{(j,d)} \leftarrow$ PARTITION_UNITY($[F_0^{(j,d;1)}, \dots, F_{i_{max}^{(d;1)}}^{(j,d;1)}], [F_0^{(j,d;2)}, \dots, F_{i_{max}^{(d;2)}}^{(j,d;2)}], \dots,$
$[F_0^{(j,d;d)}, \dots, F_{i_{max}^{(d;d)}}^{(j,d;d)}]$)

3: $l^{(j,d)} \leftarrow 0$

4: $(s_0^{(j,d;1)}, \dots, s_0^{(j,d;d)}) \leftarrow (0, \dots, 0)$

5: $\hat{P}_0^{(j,d)} \leftarrow z_0^{(j,d)}$

6: **while** $z_{l^{(j,d)}}^{(j,d)} \neq 1$ **do**

7: $\quad l^{(j,d)} \leftarrow l^{(j,d)} + 1$

8: $\quad$ **for** $k = 1, \dots, d$ **do**

9: $\quad\quad$ **if** $z_{l^{(j,d)}-1}^{(j,d)} == F_{s_{(l^{(j,d)}-1)}^{(j,d;k)}}^{(d;k)}$ **then**

10: $\quad\quad\quad s_{l^{(j,d)}}^{(j,d;k)} \leftarrow s_{(l^{(j,d)}-1)}^{(j,d;k)} + 1$

11: $\quad\quad$ **else**

12: $\quad\quad\quad s_{l^{(j,d)}}^{(j,d;k)} \leftarrow s_{(l^{(j,d)}-1)}^{(j,d;k)}$

13: $\quad$ Save the l-th support point $s_{l^{(j,d)}}^{(j,d)} = (s_{l^{(j,d)}}^{(j,d;1)}, \dots, s_{l^{(j,d)}}^{(j,d;d)})$

14: $\quad \hat{P}_{l^{(j,d)}}^{(j,d)} \leftarrow z_{l^{(j,d)}}^{(j,d)} - z_{l^{(j,d)}-1}^{(j,d)}$

15: $[\bar{s}_0^{(j,d)}, \dots, \bar{s}_{l_{\max}^{(j,d)}}^{(j,d)}] \leftarrow$ REVERSE_SUPPORT_D($[s_0^{(j,d)}, \dots, s_{l_{\max}^{(j,d)}}^{(j,d)}]$, $[\mathbf{e}^{(j,d;1)}, \dots, \mathbf{e}^{(j,d;d)}]$)

16: **return** $[\hat{P}_0^{(j,d)}, \dots, \hat{P}_{l_{\max}^{(j,d)}}^{(j,d)}], [\bar{s}_0^{(j,d)}, \dots, \bar{s}_{l_{\max}^{(j,d)}}^{(j,d)}]$

Remark 56. *Similar to Remark 32,* PARTITION_UNITY *sorts the unique elements of the union of the marginal cdfs in ascending order and Line 9 of Algorithm 5 increments the coordinate of the support according to the ordering of* $\{z_0^{(j,d)}, z_1^{(j,d)}, \dots, z_{l_{\max}^{(j,d)}}^{(j,d)}\}$. *Therefore,* $z_{l^{(j,d)}}^{(j,d)} = \min\{F_{s_{l^{(j,d)}}^{(j,d;1)}}^{(j,d;1)}, \dots, F_{s_{l^{(j,d)}}^{(j,d;d)}}^{(j,d;d)}\}$.

Remark 57. *Similar to Remark 33, Algorithm 5 is sensitive to rounding errors; implementations of Algorithm 5 in floating-point arithmetic need to account for rounding-errors. We assume in this thesis that all computations are carried out in exact arithmetic.*

Remark 58. *Note that the choice of* $X^{(1)}$ *does not matter in Algorithm 5. This can be seen from the inner loop in Lines 8-12 of Algorithm 5. All coordinates are treated the same; ordering of the coordinates does not affect accuracy or efficiency of the algorithm.*

Algorithm 6 Subroutine: Reverse support in d-dimensions

procedure REVERSE_SUPPORT_D($[s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}], [e_1^{(j,d)}, \ldots, e_d^{(j,d)}]$)
  for $l^{(j,d)} \leftarrow 0, \ldots, l_{\max}^{(j,d)}$ **do**
    for $k = 1, \ldots, d$ **do**
      if $e_k^{(j,d)} == 1$ **then**
        $\bar{s}_{l^{(j,d)}}^{(j,d;k)} \leftarrow i_{\max}^{(d;k)} - s_{l^{(j,d)}}^{(j,d;k)}$
      else
        $\bar{s}_{l^{(j,d)}}^{(j,d;k)} \leftarrow s_{l^{(j,d)}}^{(j,d;k)}$
  return $[\bar{s}_0^{(j,d)}, \ldots, \bar{s}_{l_{\max}^{(j,d)}}^{(j,d)}]$

Algorithm 7 Subroutine: Construct Marginal CDFs in d-dimensions

procedure CONSTRUCT_MARGINALS_CDFS_D($[Q_0^{(d;k)}, \ldots, Q_{i_{\max}^{(d;k)}}^{(d;k)}]_{k=1}^{d}$, $[e_1^{(j,d)}, \ldots, e_d^{(j,d)}]$)
  for $k = 1, \ldots, d$ **do**
    if $e_k^{(j,d)} == 1$ **then**
      $[t_0, \ldots, t_{i_{\max}^{(d;k)}}] \leftarrow [Q_0^{(d;k)}, \ldots, Q_{i_{\max}^{(d;k)}}^{(d;k)}]$
      for $i = 0, \ldots, i_{\max}^{(d;k)}$ **do**
        $Q_i^{(d;k)} \leftarrow t_{i_{\max}^{(d;k)} - i}$
    $[F_0^{(d;k)}, \ldots, F_{i_{\max}^{(d;k)}}^{(d;k)}] \leftarrow$ CUMSUM($[Q_0^{(d;k)}, \ldots, Q_{i_{\max}^{(d;k)}}^{(d;k)}]$)
  return $[F_0^{(d;1)}, \ldots, F_{i_{\max}^{(d;1)}}^{(d;1)}], \ldots, [F_0^{(d;k)}, \ldots, F_{i_{\max}^{(d;k)}}^{(d;k)}]$

Algorithm 8 Subroutine: Partition of Unity

procedure PARTITION_UNITY($[F_0^{(d;k)}, \ldots, F_{i_{\max}^{(d;k)}}^{(d;k)}]_{k=1}^{d}$)
  $\mathbf{F} \leftarrow$ append($[F_0^{(d;1)}, \ldots, F_{i_{\max}^{(d;1)}}^{(d;1)}], \ldots, [F_0^{(d;d)}, \ldots, F_{i_{\max}^{(d;d)}}^{(d;d)}]$)▷ *combine multiple vectors into a single vector*
  $\mathbf{F} \leftarrow$ unique($\mathbf{F}$)▷ *take only the unique elements of the input vector*
  $\mathbf{F} \leftarrow$ sort($\mathbf{F}$)▷ *sort the vector in ascending order*
  return F

2.7 Numerical Example

In this section, we provide a numerical example illustrating the Extreme Joint Distribution (EJD) approach for $d = 3$[8] by constructing all $n = 2^{3-1} = 4$ extreme measures (extreme joint distributions). The prescribed marginal distributions are Poisson distributions with mean parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$. Each of the marginal distributions are discretized by first truncating the probability distribution at a point of support $i_{\max}$ such that $\mathbb{P}(X \leq i_{\max}) \geq 1 - \epsilon$, where $\epsilon = 0.0001$. Finally, the probability weight, $\mathbb{P}(X = i_{\max})$, corresponding to the last support point, $i_{\max}$, is adjusted such that $\sum_{i=0}^{i_{\max}} \mathbb{P}(X = i) = 1$.

2.7.1 Support

Figure 2.6 illustrates the supports of the $n = 4$ 3-dimensional extreme measures. Each subfigure corresponds to an extreme measure with the extremal dependency structure described by the monotonicity structure at the top of the subfigure. The blue curve in each figure is the support of the multivariate extreme measure. Note the sparsity and the staircase like property of the support (blue curve) of extreme measures. The red, teal, and green curves are the projections of the support onto the xy, xz, and yz coordinate planes, respectively. The projections are the extreme measures in $d = 3$ marginalized to the bivariate setting. The monotonicity of the projections show that the solution to the bivariate case is preserved in higher dimensions.

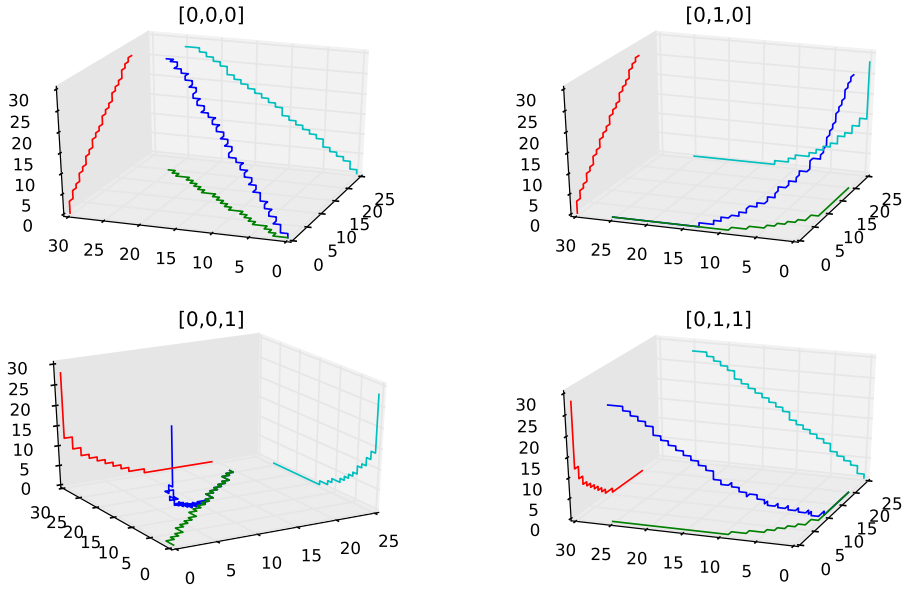

Figure 2.6: Supports of multivariate ($d = 3$) extreme measures with Poisson marginals having parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$ under all possible combinations of extremal dependence between the coordinates. The binary vector in square brackets on top of each figure indicates the corresponding monotonicity structure.

[8] We chose to compute numerical examples in the 3-dimensional case since it is the only dimension greater than the two-dimensional case that can be visualized.

2.7.2 Marginal Distributions and the Joint Partition

Table 2.1 lists the joint partition for the extreme measure having the monotone structure $[0, 0, 0]$ (the comonotone case) having Poisson marginals with parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$. Also listed is the support point corresponding to each point of the joint partition and the corresponding marginal cdf value that the joint partition came from.

l	z_l	s_l	Corresponding F	l	z_l	s_l	Corresponding F
0	0.0009	(0,0,0)	$F_0^{(3)}$	26	0.9467	(0,0,0)	$F_{11}^{(3)}$
1	0.0067	(0,0,1)	$F_0^{(2)}$	27	0.9665	(0,0,1)	$F_6^{(1)}$
2	0.0073	(0,1,1)	$F_1^{(3)}$	28	0.9682	(0,1,1)	$F_9^{(2)}$
3	0.0296	(0,1,2)	$F_2^{(3)}$	29	0.9730	(0,1,2)	$F_{12}^{(3)}$
4	0.0404	(0,1,3)	$F_1^{(2)}$	30	0.9863	(0,1,3)	$F_{10}^{(2)}$
5	0.0498	(0,2,3)	$F_0^{(1)}$	31	0.9872	(0,2,3)	$F_{13}^{(3)}$
6	0.0818	(1,2,3)	$F_3^{(3)}$	32	0.9881	(1,2,3)	$F_7^{(1)}$
7	0.1247	(1,2,4)	$F_2^{(2)}$	33	0.9943	(1,2,4)	$F_{14}^{(3)}$
8	0.1730	(1,3,4)	$F_4^{(3)}$	34	0.9945	(1,3,4)	$F_{11}^{(2)}$
9	0.1991	(1,3,5)	$F_1^{(1)}$	35	0.9962	(1,3,5)	$F_8^{(1)}$
10	0.2650	(2,3,5)	$F_3^{(2)}$	36	0.9976	(2,3,5)	$F_{15}^{(3)}$
11	0.3007	(2,4,5)	$F_5^{(3)}$	37	0.9979	(2,4,5)	$F_{12}^{(2)}$
12	0.4232	(2,4,6)	$F_2^{(1)}$	38	0.9989	(2,4,6)	$F_9^{(1)}$
13	0.4405	(3,4,6)	$F_4^{(2)}$	39	0.9990	(3,4,6)	$F_{16}^{(3)}$
14	0.4497	(3,5,6)	$F_6^{(3)}$	40	0.9993	(3,5,6)	$F_{13}^{(2)}$
15	0.5987	(3,5,7)	$F_7^{(3)}$	41	0.99964	(3,5,7)	$F_{17}^{(3)}$
16	0.6160	(3,5,8)	$F_5^{(2)}$	42	0.99970	(3,5,8)	$F_{10}^{(1)}$
17	0.6472	(3,6,8)	$F_3^{(1)}$	43	0.99977	(3,6,8)	$F_{14}^{(2)}$
18	0.7291	(4,6,8)	$F_8^{(3)}$	44	0.99987	(4,6,8)	$F_{18}^{(3)}$
19	0.7622	(4,6,9)	$F_6^{(2)}$	45	0.99992	(4,6,9)	$F_{11}^{(1)}$
20	0.8153	(4,7,9)	$F_4^{(1)}$	46	0.99993	(4,7,9)	$F_{15}^{(2)}$
21	0.8305	(5,7,9)	$F_9^{(3)}$	47	0.99995	(5,7,9)	$F_{19}^{(3)}$
22	0.8666	(5,7,10)	$F_7^{(2)}$	48	0.99998	(5,7,10)	$F_{16}^{(2)}$
23	0.9015	(5,8,10)	$F_{10}^{(3)}$	49	0.999983	(5,8,10)	$F_{12}^{(1)}$
24	0.9161	(5,8,11)	$F_5^{(1)}$	50	0.999985	(5,8,11)	$F_{20}^{(3)}$
25	0.9319	(6,8,11)	$F_8^{(2)}$	51	1	(6,8,11)	$F_{13}^{(1)}$

Table 2.1: Joint partition of the extreme measure corresponding to comonotone case and the associated support and marginal cdf for each point.

2.7.3 Probability Weights

Tables 2.2 and 2.3 list the supports and the probabilities of all the $n = 2^{d-1} = 4$ extreme measures having Poisson marginals with parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$. Note that the extremal dependency between a pair of coordinates can also be deduced from whether the supports are both increasing or both decreasing (comonotonicity) or whether one has increasing and the other decreasing support (antimonotonicity).

l	$\hat{P}^{(1,3)}_{s_l^{(1,3)}}$	$s_l^{(1,3)}$						
0	0.0009	(0,0,0)	38	0.0009	(9,13,16)	21	0.0265	(3,5,6)
1	0.0058	(0,0,1)	39	0.0001	(10,13,16)	22	0.1098	(3,5,7)
2	0.0006	(0,1,1)	40	0.0003	(10,13,17)	23	0.0392	(3,4,7)
3	0.0223	(0,1,2)	41	0.0003	(10,14,17)	24	0.0485	(3,4,8)
4	0.0108	(0,1,3)	42	6.947e-5	(10,14,18)	25	0.0819	(4,4,8)
5	0.0094	(0,2,3)	43	6.608e-5	(11,14,18)	26	0.0059	(4,4,9)
6	0.0320	(1,2,3)	44	9.640e-5	(11,15,18)	27	0.0803	(4,3,9)
7	0.0429	(1,2,4)	45	5.846e-5	(11,15,19)	28	0.0152	(5,3,9)
8	0.0483	(1,3,4)	46	2.378e-5	(12,15,19)	29	0.0449	(5,3,10)
9	0.0262	(1,3,5)	47	2.460e-5	(12,16,19)	30	0.0261	(5,2,10)
10	0.0659	(2,3,5)	48	2.453e-5	(12,16,20)	31	0.1460	(5,2,11)
11	0.0357	(2,4,5)	49	3.720e-5	(12,17,20)	32	0.0305	(6,2,11)
12	0.1225	(2,4,6)	50	1.654e-5	(13,17,20)	33	0.0129	(6,2,12)
13	0.0173	(3,4,6)	51	1.450e-5	(13,17,21)	34	0.0069	(6,1,12)
14	0.0092	(3,5,6)				35	0.0065	(7,1,12)
15	0.1490	(3,5,7)				36	0.0142	(7,1,13)
16	0.0172	(3,5,8)	l	$\hat{P}^{(3,3)}_{s_l^{(3,3)}}$	$s_l^{(3,3)}$	37	0.0009	(7,1,14)
17	0.0313	(3,6,8)	0	1.987e-5	(0,17,0)	38	0.0052	(8,1,14)
18	0.0819	(4,6,8)	1	4.914e-5	(0,16,0)	39	0.0010	(8,0,14)
19	0.0331	(4,6,9)	2	0.0001	(0,15,0)	40	0.0019	(8,0,15)
20	0.0531	(4,7,9)	3	0.0004	(0,14,0)	41	0.0014	(9,0,15)
21	0.0152	(5,7,9)	4	0.0002	(0,13,0)	42	0.0013	(9,0,16)
22	0.0361	(5,7,10)	5	0.0011	(0,13,1)	43	0.0001	(10,0,16)
23	0.0349	(5,8,10)	6	0.0034	(0,12,1)	44	0.0005	(10,0,17)
24	0.0146	(5,8,11)	7	0.0018	(0,11,1)	45	6.945e-5	(10,0,18)
25	0.0158	(6,8,11)	8	0.0064	(0,11,2)	46	0.0001	(11,0,18)
26	0.0147	(6,9,11)	9	0.0159	(0,10,2)	47	5.845e-5	(11,0,19)
27	0.0198	(6,9,12)	10	0.0022	(0,10,3)	48	2.698e-5	(12,0,19)
28	0.0017	(7,9,12)	11	0.0180	(0,9,3)	49	2.823e-5	(12,0,20)
29	0.0048	(7,10,12)	12	0.0183	(1,9,3)	50	1.654e-5	(13,0,20)
30	0.0133	(7,10,13)	13	0.0137	(1,8,3)	51	1.449e-5	(13,0,21)
31	0.0008	(7,11,13)	14	0.0516	(1,8,4)			
32	0.0009	(7,11,14)	15	0.0396	(1,7,4)			
33	0.0062	(8,11,14)	16	0.0262	(1,7,5)			
34	0.0003	(8,11,15)	17	0.0387	(2,7,5)			
35	0.0017	(8,12,15)	18	0.0629	(2,6,5)			
36	0.0014	(9,12,15)	19	0.0833	(2,6,6)			
37	0.0004	(9,12,16)	20	0.0391	(2,5,6)			

Table 2.2: Support and probabilities of three-dimensional extreme measures having Poisson marginals with parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$ corresponding to the monotone structures $(0, 0, 0)$ and $(0, 1, 0)$.

l	$\hat{P}^{(2,3)}_{s_l^{(2,3)}}$	$s_l^{(2,3)}$						
0	1.445e-5	(0,0,21)	38	0.0046	(8,11,2)	21	0.0348	(1,8,10)
1	2.991e-5	(0,0,20)	39	0.0018	(8,11,1)	22	0.0361	(1,7,10)
2	8.545e-5	(0,0,19)	40	0.0017	(8,12,1)	23	0.0296	(1,7,9)
3	0.0002	(0,0,18)	41	0.0018	(9,12,1)	24	0.0386	(2,7,9)
4	0.0005	(0,0,17)	42	0.0009	(9,13,1)	25	0.0331	(2,6,9)
5	0.0014	(0,0,16)	43	0.0002	(10,13,1)	26	0.1131	(2,6,8)
6	0.0033	(0,0,15)	44	0.0004	(10,13,0)	27	0.0172	(2,5,8)
7	0.0010	(0,0,14)	45	0.0002	(10,14,0)	28	0.0219	(2,5,7)
8	0.0061	(0,1,14)	46	0.0005	(11,14,0)	29	0.1271	(3,5,7)
9	0.0142	(0,1,13)	47	6.608e-5	(11,15,0)	30	0.0092	(3,5,6)
10	0.0134	(0,1,12)	48	1.549-4	(12,15,0)	31	0.0877	(3,4,6)
11	0.0094	(0,2,12)	49	2.378e-6	(12,16,0)	32	0.0521	(4,4,6)
12	0.0036	(1,2,12)	50	4.914e-5	(12,17,0)	33	0.0357	(4,4,5)
13	0.0452	(1,2,11)	51	3.720e-6	(13,17,0)	34	0.0803	(4,3,5)
14	0.0261	(1,2,10)				35	0.0117	(5,3,5)
15	0.0449	(1,3,10)				36	0.0484	(5,3,4)
16	0.0296	(1,3,9)	l	$\hat{P}^{(4,3)}_{s_l^{(4,3)}}$	$s_l^{(4,3)}$	37	0.0407	(5,2,4)
17	0.0659	(2,3,9)	0	1.449e-5	(0,17,21)	38	0.0022	(6,2,4)
18	0.0059	(2,4,9)	1	5.374e-6	(0,17,20)	39	0.0413	(6,2,3)
19	0.1304	(2,4,8)	2	2.453e-5	(0,16,20)	40	0.0069	(6,1,3)
20	0.0219	(2,4,7)	3	2.461e-5	(0,16,19)	41	0.0039	(7,1,3)
21	0.0173	(3,4,7)	4	6.084e-5	(0,15,19)	42	0.0177	(7,1,2)
22	0.1098	(3,5,7)	5	9.640e-5	(0,15,18)	43	0.0046	(8,1,2)
23	0.0657	(3,5,6)	6	0.0001	(0,14,18)	44	0.0006	(8,1,1)
24	0.0313	(3,6,6)	7	0.0003	(0,14,17)	45	0.0029	(8,0,1)
25	0.0521	(4,6,6)	8	0.0003	(0,13,17)	46	0.0027	(9,0,1)
26	0.0629	(4,6,5)	9	0.0011	(0,13,16)	47	0.0002	(10,0,1)
27	0.0531	(4,7,5)	10	0.0004	(0,12,16)	48	0.0006	(10,0,0)
28	0.0117	(5,7,5)	11	0.0030	(0,12,15)	49	0.0002	(11,0,0)
29	0.0396	(5,7,4)	12	0.0003	(0,11,15)	50	5.524e-5	(12,0,0)
30	0.0495	(5,8,4)	13	0.0071	(0,11,14)	51	1.615e-5	(13,0,0)
31	0.0022	(6,8,4)	14	0.0008	(0,11,13)			
32	0.0137	(6,8,3)	15	0.0133	(0,10,13)			
33	0.0035	(6,9,3)	16	0.0048	(0,10,12)			
34	0.0017	(7,9,3)	17	0.0179	(0,9,12)			
35	0.0022	(7,10,3)	18	0.0036	(1,9,12)			
36	0.0159	(7,10,2)	19	0.0147	(1,9,11)			
37	0.0018	(7,11,2)	20	0.0304	(1,8,11)			

Table 2.3: Support and probabilities of three-dimensional extreme measures having Poisson marginals with parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$ corresponding to the monotone structures $(0, 0, 1)$ and $(0, 1, 1)$.

2.7.4 Extreme Correlations

Figure 2.7 below illustrates the extreme correlation matrices, $\hat{C}^{(j,d)}$, corresponding to each extreme measure, $\hat{P}^{(j,d)}$.

$$\hat{C}^{(1,3)} = \begin{pmatrix} 1.0 & 0.93688 & 0.931861 \\ 0.93688 & 1.0 & 0.967188 \\ 0.931861 & 0.967188 & 1.0 \end{pmatrix} \qquad \hat{C}^{(3,3)} = \begin{pmatrix} 1.0 & -0.81193 & 0.931861 \\ -0.81193 & 1.0 & -0.90135 \\ 0.931861 & -0.90135 & 1.0 \end{pmatrix}$$

$$\hat{C}^{(2,3)} = \begin{pmatrix} 1.0 & 0.93688 & -0.84624 \\ 0.93688 & 1.0 & -0.90135 \\ -0.84624 & -0.90135 & 1.0 \end{pmatrix} \qquad \hat{C}^{(4,3)} = \begin{pmatrix} 1.0 & -0.81193 & -0.84624 \\ -0.81193 & 1.0 & 0.967188 \\ -0.84624 & 0.967188 & 1.0 \end{pmatrix}$$

Figure 2.7: Extreme correlation matrices $\hat{C}^{(j,d)}$ corresponding to extreme measures $\hat{P}^{(j,d)}$ with given Poisson marginals having parameters $(\lambda_1, \lambda_2, \lambda_3) = (3, 5, 7)$.

2.8 Calibration

Calibration is useful for many reasons. For example, in a practical context such as the motivating problem from Operational Risk (see Figure **??** for an example of an observed correlation matrix of operational events), calibration of a multivariate Poisson distribution to an observed correlation matrix of operational events within a large financial institution is necessary for scenario generation of operational events that are used in stochastic simulation to compute regulatory risk numbers such as VaR [96]. In a more general context, calibration is useful since failure to calibrate (i.e., the optimization problem (2.107) has no solution using the methodology described herein) means that no multivariate discrete distribution with the specified parameters that can generate the observed data exists. This implies that the parameter assumptions of the input marginal distributions are incorrect. Finally, calibration is necessary for the Backward Simulation of correlated multivariate Poisson processes where it is used to construct joint distributions having a specified correlation structure. (See Chapters 3, 4, and 5.)

Our approach to calibration is to find a representation of $P^{(C)}$ in terms of a *finite mixture* of extreme measures $\hat{P}^{(1)}, \ldots, \hat{P}^{(n)}$, where $P^{(C)}$ is the probability distribution P possessing some desired correlation structure C. Note that since $P^{(C)}$ is a finite mixture of extreme measures, it also satisfies the marginal constraints $Q^{(1)}, \ldots, Q^{(d)}$. Previous work in the literature such as [70] have also used a finite mixture model approach to construct multivariate Poisson distributions that are able to exhibit a limited amount of negative correlation. However, their approach is unable to directly control the correlation of $P^{(C)}$, whereas our linear algebraic approach uses constrained optimization to ensure that, if the correlation structure is admissible—that is, within the convex hull of the extreme correlation matrices $\hat{C}^{(j)}$ corresponding to the extreme measures $\hat{P}^{(j,d)}$—then the resulting multivariate probability distribution $P^{(C)}$ obtained from calibration posses the desired correlation structure C.

The calibration problem is trivial in the two-dimensional setting. There are only two possible extremal dependencies between a pair of marginal distributions $Q^{(1)}$ and $Q^{(2)}$: extreme positive and extreme negative dependence. Using Algorithm 1, we construct the extreme measures $\hat{P}^{(1)}$ and $\hat{P}^{(2)}$ corresponding to probability distributions maximizing and minimizing, respectively, the joint expectation $\mathbb{E}[XY]$, leading to extreme positive and extreme negative values for the correlation coefficients $\hat{C}^{(1)}$ and $\hat{C}^{(2)}$. To calibrate to a desired correlation coefficient, C, within the admissible correlation range $[\hat{C}^{(2)}, \hat{C}^{(1)}]$, we only need to solve the following linear equation, reproduced below, from (2.1)

$$C = w\,\hat{C}^{(1)} + (1-w)\,\hat{C}^{(2)}$$

for the value of $w \in [0,1]$ that, when plugged into (2.2), also reproduced below,

$$P^{(C)} = w\hat{P}^{(1)} + (1-w)\hat{P}^{(2)},$$

gives us the probability distribution $P^{(C)}$ having marginal distributions $Q^{(1)}$ and $Q^{(2)}$ and correlation coefficient C. It is important to note that, while the motivating problem stems from the need to calibrate multivariate Poisson distributions to observed correlation structures, the methods described herein apply to general multivariate discrete extreme measures constructed using the EJD method.

In this section, we discuss the calibration problem in the d-dimensional setting and how it may be

solved. To that end, recall that the d-dimensional analogue of (2.1) is:

$$C = w_1 \hat{C}^{(1,d)} + \cdots + w_n \hat{C}^{(n,d)} \tag{2.103}$$

with the constraints that $w_j \geq 0$ for $j = 1, \ldots, n$ and $\sum_j w_j = 1$. In Section 2.4.1, we showed that there are $n = 2^{d-1}$ extreme measures, $\hat{P}^{(j,d)}$, each with a corresponding monotonicity structure $\mathbf{e}^{(j,d)}$ describing its dependency structure. The objective function (2.38) of each extreme measure, $\hat{P}^{(j,d)}$, takes as input $\mathbf{e}^{(j,d)}$ to determine whether to minimize or maximize each pairwise joint expectation. Therefore, to each extreme measure $\hat{P}^{(j,d)}$ corresponds an extreme correlation matrix $\hat{C}^{(j,d)}$. Thus, analogous to the two-dimensional case, after $w_1, \ldots, w_n$ is determined from solving (2.103), it can be plugged into

$$P^{(C)} = w_1 \hat{P}^{(1,d)} + \cdots + w_n \hat{P}^{(n,d)} \tag{2.104}$$

to obtain a multivariate discrete distribution $P^{(C)}$ with correlation structure C.

2.8.1 A Linear Algebraic Approach

One approach in solving (2.103) is a linear algebraic approach. To this end, (2.103) must first be converted to a constrained system of linear equations by flattening the strictly upper triangular part of each extreme correlation matrix $\hat{C}^{(j,d)}$ into a column vector $A_j \in \mathbb{R}^m$ where $m = d(d-1)/2$. Since C and all $\hat{C}^{(j,d)}$ are symmetric with 1s on their diagonal, this can be done by taking each row in the strictly upper triangular part of each $\hat{C}^{(j,d)}$, appending them into a row vector and taking the transpose to be A_j to obtain $\mathbf{A} = [A_1, \ldots, A_n] \in \mathbb{R}^{m \times n}$, representing the extreme points of our problem in correlation space. Similarly, we can flatten the strictly upper triangular part of the given correlation matrix C on the left side of (2.103) to a vector $b \in \mathbb{R}^m$. Then, (2.103) and the constraints $w_j \geq 0$ for $j = 1, 2, ..., n$ and $\sum_{j=1}^n w_j = 1$ are equivalent to the constrained system of linear equations

$$\mathbf{A}\mathbf{w} = b, \tag{2.105a}$$

$$\mathbf{1}^T \mathbf{w} = 1, \tag{2.105b}$$

$$w_j \geq 0 \quad j = 1, 2, \ldots, n. \tag{2.105c}$$

There are many possible solutions to the constrained system of equations (2.105). One approach is heuristic and relies on the context of the particular application in choosing a suitable objective function[9] and then using (2.105) as the constraints for an optimization problem with that objective function. However, if the goal is just to find *any* solution to (2.105), then a simpler approach is to reformulate (2.105) as

$$\hat{A}w = \hat{b}, \tag{2.106a}$$

$$w_j \geq 0 \qquad j = 1, 2, \ldots, n, \tag{2.106b}$$

where $\hat{A}$ is A with the row $\mathbf{1}^T$ appended to the bottom of it and $\hat{b}$ is b with a 1 appended to the bottom of it. Then, note that (2.106) has the form of the standard constraints for a Linear Programming Problem (LPP). Moreover, the first stage of many LPP codes finds a solution to (2.106). As explained in Section

[9] This is also the subject of future work.

13.5 of [93], a standard approach to finding a solution to (2.106) is to solve the LPP

$$\begin{aligned} \min \quad & \mathbf{1}^T z && \text{(2.107a)} \\ \text{subject to} \quad & \hat{A}w + Ez = \hat{b} && \text{(2.107b)} \\ & (w, z) \geq 0 && \text{(2.107c)} \end{aligned}$$

where $z \in \mathbb{R}^{m+1}$ and E is a $(m+1) \times (m+1)$ diagonal matrix such that $E_{ii} = +1$ if $\hat{b}_i \geq 0$ and $E_{ii} = -1$ if $\hat{b}_i < 0$. Clearly, $w = 0$ and $z = |\, b \,|$ satisfies the constraints (2.107b) and (2.107c). So, we can use $w = 0$ and $z = |\, b \,|$ as a staring point for the simplex method to solve (2.107). It's clear from the constraint $z \geq 0$ that the solution satisfies $\mathbf{1}^T z \geq 0$. Moreover, if $\mathbf{1}^T z = 0$ then $z = 0$. Hence, (2.107) has a solution $\mathbf{1}^T z = 0$ if and only if $\hat{A}w = \hat{b}$, $w \geq 0$ has a solution. Hence, the simplex method applied to (2.107) will find a solution to (2.105), if a solution exists.

Despite the fact that the problem size grows exponentially in d, due to the structure of the problem (2.105) and the fact that the simplex method needs to explicitly access $m + 1$ columns of $\hat{A}$ at a time (assuming you have some clever way to decide which new vector to bring into the active set at at each step of the simplex method without explicitly accessing all the columns of A that are in the inactive set) the LP (2.107) can be solved for a surprisingly large d, e.g., $d = 51$, which corresponds to $n = 2^{50} \approx 10^{15}$; see [86].

Remark 59. *While it is difficult to provide rigorous estimates on the run times and accuracy of solving* (2.107) *numerically due to the nature of both the simplex algorithm and the input (marginal) distributions, we remark that the typical run times of the problem in the six-dimensional case are roughly around thirty minutes on an Apple laptop with an Intel Core i7-9750H processor. The approximation error is bounded by* 10^{-3}.

2.8.2 The Independent Case

It is important to note that our approach, consisting of taking finite convex combinations of the extreme measures (2.103), does not give all distributions $P^{(C)}$ having correlation structure C. This can be seen from the fact that the finite convex combination in (2.103) does not include the zero correlation matrix $\hat{C}^{(0,d)}$. However, incorporating $\hat{C}^{(0,d)}$ into the calibration process brings no benefits and, in fact, completely destroys the sparsity inherited from the extreme measures (Remarks 15 and 23). To see this, consider adding $\hat{C}^{\,(0,d)}$ to the finite convex combination of extreme correlations (2.103) resulting in

$$C = w_0' \hat{C}^{\,(0,d)} + w_1' \hat{C}^{\,(1,d)} + \cdots + w_n' \hat{C}^{\,(n,d)} \tag{2.108}$$

with the constraints $w_j' \geq 0$ for $j = 0, 1, \ldots, n$ and $\sum_j w_j = 1$. Since $\hat{C}^{\,(0,d)}$ is an identity matrix, the flattening process described in Section 2.8.1 necessary to convert (2.108) into a constrained system of equations (2.105) flattens $\hat{C}^{\,(0,d)}$ into a vector of 0s. Since having a column of 0s in $\mathbf{A}$ would make it singular, it must be excluded from the construction of $\mathbf{A}$. This makes intuitive sense, since the zero correlation case does not provide additional information. Therefore, w_0' must be specified so that instead of the constraint (2.105b), we have the modified constraint

$$\mathbf{1}^T \mathbf{w'} = 1 - w_0', \tag{2.109}$$

where $\mathbf{w'} = (w'_1, \ldots, w'_n)$. Note that this has the effect of removing a degree of freedom from the solution since the introduction of w'_0 and the constraint

$$w'_0 + w'_1 + \cdots + w'_n = 1$$

constrains the possible values of $w'_1, \ldots, w'_n$. Clearly, the value of the weights $w'_1, \ldots, w'_n$, differ from the weights $w_1, \ldots, w_n$ obtained from solving (2.103). Plugging $w'_0, \ldots, w'_n$ into

$$\bar{P}^{(C)} = w'_0 \hat{P}^{(0,d)} + w'_1 \hat{P}^{(1,d)} + \cdots + w'_n \hat{P}^{(n,d)}, \tag{2.110}$$

where $\hat{P}^{(j,d)}$ corresponds to $\hat{C}^{(j,d)}$ gives the probability distribution $\bar{P}^{(C)}$ having correlation C. Note that since $\hat{P}^{(0,d)}$ is the (discrete) probability distribution corresponding to the independent case, $\hat{C}^{(0,d)}$, its support must be the larger than the support of the extreme measures, thus, the support of $\hat{P}^{(0,d)}$ cannot be sparse. Hence, the support of $\bar{P}^{(C)}$ is also not sparse. Moreover, since $w'_0 > 0$ in order for the independent case, $\hat{C}^{(0,d)}$ to be included, $\bar{P}^{(C)}$ must also be supported on every point in its domain. The inclusion of the independent case in our EJD methodology is detrimental to the numerical efficiency of our approach, which is inherited from the sparsity of the extreme measures $\hat{P}^{(1,d)}, \cdots, \hat{P}^{(n,d)}$. (See Remarks 15 and 23.)

2.8.3 Admissible Correlations

If no solution to the optimization problem (2.107) exists, then the given correlation matrix, C, cannot be generated by a joint distribution having marginal distributions $Q^{(1)}, \ldots, Q^{(d)}$. We say that C is an *inadmissible correlation matrix*. Similarly, if a solution to (2.107) exists for a given correlation matrix C and a given set of marginal distributions $Q^{(1)}, \ldots, Q^{(d)}$, we say that C is an *admissible correlation matrix*. This is due to the fact that the optimization problem (2.107) generates *all possible* correlation matrices of a joint distribution having a set of given marginals distributions. To see this, we first note that (2.105) and (2.107) are equivalent and that the columns of the matrix A in (2.105) correspond to the flattened correlation matrices associated with the extreme measures $\hat{P}^{(1,d)}, \ldots, \hat{P}^{(n,d)}$. Recall that each extreme measure, $\hat{P}^{(j,d)}$, has an associated monotone structure, $\mathbf{e}^{(j,d)}$, describing the extreme dependence between its components and that the monotonicity structure, $\mathbf{e}^{(j,d)} = (\mathbf{e}_1^{(j,d)}, \ldots, \mathbf{e}_d^{(j,d)})$, describes all possible extreme dependence structures (see Section 2.4.1). That is, the extreme measures $\hat{P}^{(1,d)}, \ldots, \hat{P}^{(n,d)}$ are extreme points. Moreover, since to each extreme measure, $\hat{P}^{(j,d)}$, there is an associated extreme correlation matrix $\hat{C}^{(j,d)}$, the extreme correlation matrices $\hat{C}^{(1,d)}, \ldots, \hat{C}^{(n,d)}$ are also extreme points in the space of correlation matrices. The extreme correlation matrices $\hat{C}^{(1,d)}, \ldots, \hat{C}^{(n,d)}$ form a convex set since the set of all $d \times d$ correlation matrices form a compact convex set [77]. Since, by Carathéodory's theorem [35], any point in the convex hull of the extreme points $\hat{C}^{(1,d)}, \ldots, \hat{C}^{(n,d)}$ can be represented by at most $d+1$ points in the set $\{\hat{C}^{(1,d)}, \ldots, \hat{C}^{(n,d)}\}$, therefore, the constrained system of linear equations (2.105) (and the optimization problem 2.107) generates all *possible* correlations for a multivariate joint distribution satisfying a set of marginal distributions.

In the case of inadmissible correlation matrices, this typically means one of two things: 1) the assumptions on the parameters of the marginal distributions are wrong; or 2) the desired correlation is incorrect—there could be errors in its estimation. It is up to the practitioner to determine a suitable course of action. If the practitioner believes that both the marginal distribution and the desired corre-

lation matrix are correct, one possibility is to solve for the closest admissible matrix under some choice of norm within the admissible set of correlation matrices. That is, solve

$$\begin{aligned} \min_{w\in\mathbb{R}^n} \quad & \|\hat{C}(w) - C\| \\ \text{subject to} \quad & w_l \geq 0, \quad l = 1, \ldots, n, \\ & \sum_{l=1}^{n} w_l = 1, \end{aligned} \tag{2.111}$$

where $\hat{C}(w) = \sum_{l=1}^{n} w_l \hat{C}^{(l,d)}$.

2.8.4 Calibration Algorithm

Algorithm 9 Calibration of Extreme Measures

Require: Marginal distributions: $[Q_0^{(d;k)}, \ldots, Q_{i_{\max}^{(d;k)}}^{(d;k)}]_{k=1}^{d}$
Desired correlation matrix: C

Output: Extreme measures $[\hat{P}_0^{(j,d)}, \ldots, \hat{P}_{l_{\max}^{(j,d)}}^{(j,d)}]$ and their supports $[s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}]$ where $j = 1, \ldots, n$
Weights of the convex combination of extreme measures $(w_1, \ldots, w_n)$

$n \leftarrow 2^{d-1}$ ▷ *number of extreme points*
$m \leftarrow d(d-1)/2$ ▷ *number of elements in the strictly upper triangular part of a $d \times d$ correlation matrix*
$A \leftarrow \mathbf{zeros}(m, n)$ ▷ *construct matrix of extreme points of correlations*
for $j \leftarrow 1, \ldots, n$ **do**
    $[\hat{P}_0^{(j,d)}, \ldots, \hat{P}_{l_{\max}^{(j,d)}}^{(j,d)}], [s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}] \leftarrow$ Algorithm 5 $([Q_0^{(d;k)}, \ldots, Q_{i_{\max}^{(d;k)}}^{(d;k)}]_{k=1}^{d}, \ \mathbf{e}^{(j,d)})$
    $\hat{C}^{(j,d)} \leftarrow$ Algorithm 11$([\hat{P}_0^{(j,d)}, \ldots, \hat{P}_{l_{\max}^{(j,d)}}^{(j,d)}], [s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}])$
    $A_j \leftarrow$ FLATTEN_ABOVE_DIAG$(\hat{C}^{(j,d)})$
$A \leftarrow [A; \mathbf{ones}(1, n)]$ ▷ *append vector of ones to the bottom row*
$b \leftarrow$ FLATTEN_ABOVE_DIAG(C)
$b \leftarrow [b; 1]$
$(w_1, \ldots, w_n)$, success $\leftarrow$ Solve (2.107)
if success **then**
    return $(w_1, \ldots, w_n)$
else
    error: no solution to (2.107) exists for the given marginal distributions and correlation matrix

Algorithm 10 Subroutine: Flatten entries above diagonal of A

procedure FLATTEN_ABOVE_DIAG(A)
    nRows, nCols $\leftarrow$ size(A)
    if nRows $\neq$ nCols **then**
        error: A should be square
    $m \leftarrow$ nRows $\cdot$ (nRows $-$ 1)/2
    $[b_0, \ldots, b_{m-1}] \leftarrow$ **zeros**(1,m) ▷ *row vector*
    $k \leftarrow 0$
    for $i \leftarrow 0, \ldots,$ nRows $- 2$ **do**
        for $j \leftarrow i + 1, \ldots,$ nCols $- 1$ **do**
            $b_k \leftarrow A_{i,j}$
            $k \leftarrow k + 1$
    return $[b_0, \ldots, b_{m-1}]$

Algorithm 11 Compute Correlation Matrix from a d-dimensional Extreme Measures

Require: d-dimensional extreme measure $[\hat{P}_0^{(j,d)}, \ldots, \hat{P}_{l_{\max}^{(j,d)}}^{(j,d)}]$ and its support $[s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}]$

Output: Corresponding extreme correlation matrix $\hat{C}^{(j,d)}$

for $u \leftarrow 1, \ldots, d$ **do**
  $\hat{C}_{u,u}^{(j,d)} \leftarrow 1$
  for $v \leftarrow u+1, \ldots, d$ **do**
    $\hat{P}^{(j,d;u,v)}, \mathbf{s}^{(j,d;u,v)} \leftarrow$ Algorithm 12$([\hat{P}_0^{(j,d)}, \ldots, \hat{P}_{l_{\max}^{(j,d)}}^{(j,d)}], [s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}], (u,v))$
    $\hat{C}_{u,v}^{(j,d)} \leftarrow$ Correlation$(\hat{P}^{(j,d;u,v)}, \mathbf{s}^{(j,d;u,v)})$
    $\hat{C}_{v,u}^{(j,d)} \leftarrow \hat{C}_{u,v}^{(j,d)}$
return $\hat{C}^{(j,d)}$

Algorithm 12 Marginalize d-dimensional Extreme Measure into a d'-dimensional Extreme Measure

Require: d-dimensional extreme measure $[\hat{P}_0^{(j,d)}, \ldots, \hat{P}_{l_{\max}^{(j,d)}}^{(j,d)}]$, its support $\mathbf{s}^{(j,d)} = [s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}]$
Coordinates $(u_1, \ldots, u_{d'})$ to project onto by marginalization ▷ *indices to keep*

Output: d'-dimensional extreme measure $[\hat{P}_0^{(j,d')}, \ldots, \hat{P}_{l_{\max}^{(j,d')}}^{(j,d')}]$ and its support $[s_0^{(j,d')}, \ldots, s_{l_{\max}^{(j,d')}}^{(j,d')}]$

▷ *marginalized support* ◁
$s^{(j,d')} = s^{(j,d)}([u_1, \ldots, u_{d'}], :)$ ▷ *each s_i^d is a column vector; $s^{d'}$ is also a column vector*
$\tilde{s}^{(j,d')} = \text{unique}(s^{(j,d')})$
$l = \text{length}(\tilde{s}^{(j,d')})$
▷ *marginalized weights* ◁
$\hat{P}^{(j,d')} = \text{zeros}(\text{size}(\tilde{s}^{(j,d')}, 2))$
for (count, cols) = enumerate(eachcol($\tilde{s}^{(j,d')}$)) **do** ▷ *iterates through the unique support points*
  indices = findall(x→x==cols, $s^{(j,d')}$) ▷ *returns indices of the duplicated columns*
  $\hat{P}_{\text{count}}^{(j,d')} = \text{sum}(\hat{P}^{(j,d)}[\text{indices}])$
return $([\hat{P}_0^{(j,d')}, \ldots, \hat{P}_l^{(j,d')}], [\tilde{s}_0^{(j,d')}, \ldots, \tilde{s}_l^{(j,d')}])$

- `enumerate` has the behavior of the python function with the same name that alters the for-loop such that at every iteration, the for-loop returns a tuple containing a count corresponding to the loop number and the value from iterating over the (iterable) function argument.

- `eachcol` is a helper function that iterates over the columns of a matrix or an array

- `findall` takes in an anonymous function as its first argument and returns a vector of indices of where the anonymous function applied to the array (second argument) returns true.

Algorithm 13 Subroutine: Correlation Between Components of a Bivariate Extreme Measure

procedure CORRELATION($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}$)
  $e \leftarrow$ BIVARIATE_EXPECTATION($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}$)
  $b_1 \leftarrow$ DOT($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}, 1$)
  $b_2 \leftarrow$ DOT($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}, 2$)
  $v_1 \leftarrow$ VARIANCE($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}, 1$)
  $v_2 \leftarrow$ VARIANCE($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}, 2$)
  return $(e - b_1 \cdot b_2)/\sqrt{v_1 \cdot v_2}$

Algorithm 14 Subroutine: Expectation of a Bivariate Extreme Measure

procedure BIVARIATE_EXPECTATION($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}$)
  $e \leftarrow 0$
  for $i \leftarrow 0, \ldots, l^2_{\max}$ **do**
    $e \leftarrow e + s_i^{(j,2;1)} \cdot s_i^{(j,2;2)} \cdot \hat{P}_i^{(j,2)}$
  return e

Algorithm 15 Subroutine: Variance of a Bivariate Extreme Measure

procedure VARIANCE($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}, k$)
  $v \leftarrow 0$
  $e \leftarrow$ DOT($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}, k$) ▷ *mean of the extreme measure* $\hat{P}^{(j,2)}$
  for $i \leftarrow 0, \ldots, l^2_{\max}$ **do**
    $v \leftarrow v + (s_i^{(j,2;k)})^2 \cdot \hat{P}_i^{(j,2)}$
  return $v - (e)^2$

Algorithm 16 Subroutine: Dot Product Specialized to Birvariate Measures

procedure DOT($\hat{P}_0^{(j,2)}, \ldots, \hat{P}_{l^2_{\max}}^{(j,2)}, s_0^{(j,2)}, \ldots, s_{l^2_{\max}}^{(j,2)}, k$)
  $e \leftarrow 0$
  for $i \leftarrow 0, \ldots, l^2_{\max}$ **do**
    $e \leftarrow e + s_i^{(j,2;k)} \cdot \hat{P}_i^{(j,2)}$
  return e

2.9 Sampling from Multivariate Extreme Measures

One of the main advantages of the EJD approach is that sampling from multivariate extreme measures, $\hat{P}^{(j,d)}$, obtained by Algorithm 5 is equivalent to sampling from univariate probability distribution. This is possible because of the sparsity of the supports which behaves like a univariate distribution. In particular, the support of multivariate extreme measures are directed graphs that exhibit a staircase-like property (Figure 2.4 provides an illustration) thereby enabling the use of the inverse CDF method. (See Remarks 29 and 30.)

Algorithm 17 Sampling from Multivariate Extreme Measures

Require: Extreme Measure: $[\hat{P}_0^{(j,d)}, \ldots, \hat{P}_{l_{\max}}^{(j,d)}]$, $[s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}]$
Number of samples: n
Output: $\hat{P}^{(j,d)}$ distributed random vector $[\tilde{s}_0, \ldots, \tilde{s}_{n-1}]$ of size n

```
for r ← 0, ..., n − 1 do
    Generate a uniform random variable u
    ▷ Inverse cdf through sequential search
    l ← 0
    q ← P̂_l^(j,d)
    while u > q do
        l ← l + 1
        q ← q + P̂_l^(j,d)
    s̃_r ← s_l^(j,d)_(j,d)
return [s̃_0, ..., s̃_{n−1}]
```

2.9.1 Sampling from Calibrated Measures

In Section 2.8, we discussed a linear algebraic approach to the construction of calibrated measures, $P^{(C)}$, that are a convex combination of the extreme measures $\hat{P}^{(1,d)}, \ldots, \hat{P}^{(n,d)}$ having correlation C. The calibration algorithm, listed in Algorithm 9, returns a series of weights $(w_1, \ldots, w_n)$ that satisfies $w_j \geq 0$ for $j = 1, 2, \ldots, n$ and $\sum_{j=1}^{n} w_j = 1$. When substituted into (2.104), the weights give a multivariate probability distribution $P^{(C)}$ having correlation matrix C. Since the weights $(w_1, \ldots, w_n)$ are a convex sum, they can be interpreted as *probabilities* when sampling from a correlated measure $P^{(C)}$. For each draw, the w_j for $j = 1, 2, \ldots, n$, is the probability that the extreme measure $\hat{P}^{(j,d)}$ should be sampled from. Having chosen the particular extreme measure to sample from, we can then apply Algorithm 17 to obtain samples from the extreme measure.

Algorithm 18 Sampling from Calibrated Measures

Require: Extreme measures: $[\hat{P}_0^{(j,d)}, \ldots, \hat{P}_{l_{\max}^{(j,d)}}^{(j,d)}]_{j=1}^n$
Supports: $[s_0^{(j,d)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d)}]_{j=1}^n$
Calibrated weights: $\mathbf{w} = [w_1, \ldots, w_n]$ such that $P^{(C)}$ has correlation C
Number of samples: n

Output: $P^{(C)}$ distributed random vector, $[\tilde{s}_0, \ldots, \tilde{s}_{n-1}]$

for $r \leftarrow 0, \ldots, n-1$ **do**
  Generate a uniform random variable u_1
  $l \leftarrow 1$
  $q_1 \leftarrow w_l$
  while $u_1 > q_1$ **do**
    $l \leftarrow l + 1$
    $q_1 \leftarrow q_1 + w_l$
  Generate a uniform random variable u_2
  $m \leftarrow 0$
  $q_2 \leftarrow \hat{P}_m^{(l,d)}$
  while $u_2 > q_2$ **do**
    $m \leftarrow m + 1$
    $q_2 \leftarrow q_2 + \hat{P}_m^{(l,d)}$
  $\tilde{s}_r \leftarrow s_m^{(l,d)}$
return $[\tilde{s}_0, \ldots, \tilde{s}_{n-1}]$

2.10 Computational Complexity

2.10.1 EJD Construction

The EJD algorithms construct the set of discrete joint distributions $\{\hat{P}^{(j,d)}\}_{j=1}^{2^{d-1}}$ satisfying given marginal distributions according to a monotone structure. Recall that d denotes the problem dimension (i.e., the correlation matrix C is $d \times d$) , $l_{\max}^{(j,d)}$ denotes the number of iterations of Algorithm 5 and $1 + l_{\max}^{(j,d)}$ is the length of the $\mathcal{S}$-path associated with $\hat{P}^{(j,d)}$, $i_{\max}^{(k)}$ denotes the length of the k^{th} marginal distribution, A is the matrix that has the columns that are the flattened strictly-upper-triangular part of the extreme correlation matrix $\hat{C}^{(j,d)}$ associated with the extreme measure $\hat{P}^{(j,d)}$ and has size $m \times n$ where $m = d(d-1)/2$ and $n = 2^{d-1}$, and $\hat{A}$ is the matrix A with the row $\mathbf{1}^T$ appended to the bottom of it.

Time Complexity

Algorithm 5, which generates each $\hat{P}^{(j,d)}$, iterates over d marginal dimensions for every iteration of the algorithm. Therefore, the complexity for each EJD construction is

$$O(l_{\max}^{(j,d)} \cdot d). \tag{2.112}$$

It is easy to see from Algorithm 5 that $l_{\max}^{(j,d)} \leq i_{\max}^{(k)} \cdot d$, where k is the index of the input marginal distribution with the greatest length. Hence, the time complexity becomes $O(i_{\max}^{(k)} \cdot d^2)$. In the worst case where all the 2^{d-1} extreme measures need to be computed, the total cost is $O(2^{d-1} \cdot l_{\max}^{(j,d)} \cdot d)$, although, typically, only a small subset of the 2^{d-1} extreme measures need to be computed.

Memory Complexity

Each extreme measure $\hat{P}^{(j,d)}$ and its associated support points occupy

$$O(l_{\max}^{(j,d)} \cdot d) \tag{2.113}$$

space. Recall again that, typically, only a small subset of the 2^{d-1} extreme measures have to be computed and stored. In the worst case that all 2^{d-1} distributions have to be stored, then the total memory complexity is $O(2^{d-1} \cdot l_{\max}^{(j,d)} \cdot d)$. Note that d is typically small for most applications.

Remark 60. *The EJD construction is highly efficient in high-dimensional settings even though the construction of each extreme measure has quadratic time complexity in d since d is small for most applications. More importantly, the full set of $n = 2^{d-1}$ extreme measures rarely need to be constructed which prevents the complexity from being exponential in practice.*

2.10.2 Extreme Correlation Matrix $C^{(j,d)}$ Construction

For each extreme measure $\hat{P}^{(j,d)}$ produced by Algorithm 5, we construct its associated extreme correlation matrix $\hat{C}^{(j,d)}$ by Algorithm 11. Recall that Algorithm 11 computes all strictly-upper-triangular elements of the extreme correlation matrix $\hat{C}^{(j,d)}$ (and sets the diagonal to 1) via numerical evaluation of expectations and variances of bivariate components of the extreme measure. Let $m = d(d-1)/2$ denote the the number of strictly-upper-triangular entries of a $d \times d$ correlation matrix.

Time Complexity

The correlation, joint expectations, and variances are all arithmetic operations over the support of the extreme measure $\hat{P}^{(j,d)}$. Therefore, the time complexity of constructing $\hat{C}^{(j,d)}$ for a single extreme measure is

$$O(m \cdot l_{\max}^{(j,d)}) = O(d^2 \cdot l_{\max}^{(j,d)}). \tag{2.114}$$

If we are constructing *all* the correlation matrices associated with all the extreme measures, then the time complexity is $O(2^{d-1} \cdot m \cdot l_{\max}^{(j,d)})$ and is exponential in d. However, as noted throughout this section, d is typically small and the full set of extreme measures (and their associated correlations) rarely need to be constructed in full.

Memory Complexity

The correlation matrix $\hat{C}^{(j,d)}$ requires $O(d^2)$ space. Storing only the strictly-upper-triangular part of the matrix still requires $O(m) = O(d^2)$ space.

2.10.3 Calibration

The calibration problem (2.107) relies on the construction of the $\hat{A}$ matrix in (2.107b), the complexity of which is equal to $O(2^{d-1} \cdot m \cdot l_{\max}^{(j,d)})$ if the full matrix is constructed. We note that MacDonald's thesis [86] develops a modified simplex method that does not require the construction of the full $\hat{A}$ matrix. Rather, the method maintains an "active set" consisting of $m+1$ columns of $[\hat{A}, E]$ and swaps columns of $[\hat{A}, E]$ in and out of the active set as needed. This drastically reduces the computational complexity in practical settings. For this reason, we focus on the complexity of the mathematical programming calibration problem itself and omit the cost of setting up the linear program.

Time Complexity

The simplex approach to solving linear programs is well-known to have an exponential worst-case time complexity in n when the A matrix of the linear program is of size $m \times n$. (This means that for our problem, the complexity is double exponential in d.) The worst-case time complexity depends on the solution methods themselves. Note however, that in practice, the simplex algorithm works quite well in practice and simplex-based approaches like that of [86] can prevent the need for constructing the full $\hat{A}$ matrix.

Memory Complexity

If a simplex-like approach like [86] is used that doesn't require constructing the full A matrix, then the worst-case memory complexity is well-known to be $O(m^2)$.

Remark 61. *Similar to Remark 60, in high-dimensional settings, the EJD calibration approach is "usually" efficient due to the fact that simplex-based approaches like [86] do not need the construction of the full set of $n = 2^{d-1}$ extreme measures and associated correlation matrices and also because the simplex algorithm is much faster, in practice, than its theoretical complexity.*

2.11 Summary

In this chapter, we introduced the Extreme Joint Distribution (EJD) approaching consisting of the EJD theorem (Theorem 1) and the EJD algorithm (Algorithm 1). We proved rigorously the foundations of the theory in the bivariate setting. Next, we extended both the theorem and algorithm to the general d-dimensional setting (Theorem 4 and Algorithm 5, respectively). Having a method to construct multivariate extreme measures, we discussed the calibration of a joint distribution satisfying given marginal distributions to a prescribed correlation using the multivariate measures generated by Algorithm 5. Finally, we discussed sampling from extreme measures and calibrated measures.

Our contributions are summarized below:

- Proved, rigorously, various properties of comonotone and antimonotone distributions. (Subsection 2.2.1.)
- We proved the correctness of Algorithm 1. (Section 2.3.)
- Extended the notion of extreme measures to the d-dimensional setting (Section 2.4.)
- Extended the EJD algorithm to d-dimensions and proved its correctness. (Section 2.5.)
- Introduced a method for the calibration of multivariate discrete extreme measures to observed correlation structures. (Section 2.8.)

Chapter 3

Backward Simulation of Poisson Processes

This chapter discusses the core ideas of simulating a Poisson process backwards in time within a simulation interval $[0, T]$. In general, there are two methods for the simulation of Poisson processes: forward and backward simulation. A stochastic process X_t is simulated forwards in time by starting from time $t = 0$ and then subsequently simulating the process at the next grid point, X_{t+h}, where h is the time step, until the terminal time $t = T$ is reached [27]. The prevalence of forward simulation is due to its intuitive simplicity and universal applicability. In contrast, backward simulation of a stochastic process X_t in an interval $[0, T]$ requires the conditional distribution of the process given some terminal value $\mathbb{P}(X_t \,|X_T = n)$ be known or readily obtainable. Backward simulation is becoming increasingly widespread [110] and has been applied to fields such as finance [6, 12, 13, 17, 25, 80], stochastic control [115], stochastic simulation [9, 11, 50, 81], and Monte Carlo statistical inference [83].

Backward simulation of continuous processes is well known and is often referred to as *bridge sampling* in the literature. The class of processes that bridge sampling applies to are known as *bridge processes*, the most well known being the Brownian Bridge [13, 107]. Other bridge processes include the gamma bridge [5, 108], the Ornstein-Uhlenbeck (OU) bridge [20], the Cox-Ingersoll-Ross (CIR) bridge [65], the Bessel bridge processes [87, 101], and many others. Bridge processes for more general diffusions are know as *diffusion bridges* [9, 11, 81]. In contrast, backward simulation for discrete valued stochastic processes, to the best of our knowledge, is not as widely discussed. However, the crucial property that backward simulation of Poisson processes depends is well known:

Proposition 3 (Conditional Uniformity [51])**.** *The joint probability density function, f, of the arrival moments $T_1, T_2, \ldots, T_n$ of the Poisson process, N_t, conditioned on the event $N_t = n$, is given by*

$$f_{(T_1,\ldots,T_n|\, N_t=n)}(x_1, x_2, \ldots, x_n) = \begin{cases} \frac{n!}{t^n} & \text{if} \quad 0 \le x_1 \le x_2 \le \cdots \le x_n \le t \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

Thus, the conditional uniformity property of Poisson processes suggests a simulation method: draw from a Poisson distribution at terminal time to obtain the number of events, n, at terminal time. Then, draw n uniform variates $U_1, \ldots, U_n$. Reordering the uniform variables $U_1, \ldots, U_n$ in increasing order

gives the arrival moments $T_1, T_2, \ldots, T_n$ of the Poisson process. The conditional uniformity property is an instance of the more general *order statistic* property [26, 39, 79, 104]. Briefly, the order statistic property states that:

Definition 28 (Order Statistic Property [39])**.** *For a point process $\{M_t; t \geq 0\}$ with right-continuous paths, unit steps at times $T_1, T_2, \ldots$, and the non decreasing mean value function $m_t = \mathbb{E}\, M_t < \infty$, $t \geq 0$, the successive jump times $\{T_1, \ldots, T_n\}$, conditional on $M_t - M_0 = n$ are distributed as the order statistics of n independent identically distributed random variables with distribution function F supported on $[0, t]$.*

Indeed, the backward simulation method above is effectively a Poisson bridge, analogous to the well-known Brownian bridge. Just as a Brownian bridge connects two endpoints with a Brownian path, the Poisson bridge connects the initial state (typically zero events at $t = 0$) to a specific terminal count at time T using conditional uniformity. However, our contribution extends significantly beyond the standard univariate Poisson bridge. By combining the EJD methodology with backwards simulation, we construct what can be seen as a *multivariate Poisson bridge* that not only results in a set of multivariate Poisson paths, but also reproduces any desired admissible correlation structure at terminal time. This is a substantial generalization of the classical Poisson bridge to the multivariate setting.

Although the backward simulation of multivariate Poisson processes is not discussed much in the literature, the backward simulation of univariate Poisson processes is well known. One reason why the multivariate case has not been discussed much might be that neither the reproducibility of the correlation structure between the components of a multivariate Poisson process at the terminal simulation time nor the time structure of the correlations between the components of a multivariate Poisson process within the simulation interval have been explored—to the best of the author's knowledge—in the literature before [74], upon which our work builds. In contrast, our approach to backward simulation is able to match any desired admissible correlation structure at the terminal simulation time. This relies heavily on the EJD approach, discussed in Chapter 2, to generate extreme joint distributions that are used in the calibration of a multivariate Poisson distribution possessing the desired correlation structure. Then, since the backward simulation approach exploits the conditional uniformity property, the number of terminal events required for the simulation of the process is simply obtained by sampling the calibrated joint distribution. In addition, our analysis below shows that the time structure of correlation under backward simulation is *linear* in time.

We also discuss the Forward Continuation (FC) of the Backward Simulation (BS), introduced in [17], for continuing the process past the original simulation interval $[0, T]$ to an integer multiple of the original simulation length mT for some integer m. We demonstrate that the correlation structure under the Forward Continuation of Backward Simulation reaches asymptotic stationarity.

It is important to note that while much of the exposition in this chapter and in Chapters 4 and 5 is in the bivariate setting, the discussions generalize immediately to the multivariate setting.

3.0.1 Outline

The outline for this chapter is as follows. We introduce Backward Simulation in Section 3.1, where we prove the main result (Theorem 6) enabling the Backward Simulation approach for Poisson processes and provide an accompanying algorithm (Algorithm 19). Section 3.2 analyzes the correlation structure under Backward Simulation. Section 3.3 introduces the Forward Continuation (FC) of the Backward

Simulation (BS) for correlated multivariate Poisson processes. Having introduced Backward Simulation and the Forward Continuation, we then compare the backward and the forward approaches in Section 3.4. Section 3.6 concludes the chapter.

3.0.2 Notation

We make use of the following notation in this chapter.

Symbol	Definition
T	End of the simulation interval
T_i	i^{th} arrival moment
$\Delta T_i := T_i - T_{i-1}$	i^{th} inter-arrival time
$T_i^{(k)}$	i^{th} arrival moment corresponding to the k^{th} component
N	Poisson-distributed random variable
λ	Mean parameter of a Poisson distribution
n	Realization of a Poisson random variable
$\mathbf{N} = (N^{(1)}, \ldots, N^{(d)})$	d-dimensional Poisson-distributed random vector
N_t	Poisson process
$\mathbf{N}_t = (N_t^{(1)}, \ldots, N_t^{(d)})$	d-dimensional Poisson process
X_t	Poisson process obtained from Backward Simulation
$\stackrel{d}{=}$	Equal in distribution
$\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(d)})$	Multivariate Poisson process obtained from Backward Simulation
$\Delta_s X_t$	The increment $X_{t+s} - X_t$ of the process X_t in interval $[t, t+s]$
ζ, ξ	General integer valued random variables
$p_k := \mathbb{P}(\xi = k)$	Probability of the random variable ξ taking on the value k
$\hat{p}(z)$	Generating function of the distribution of the random variable ξ
$q_k(x)$	Probability that depends on $p_{\hat{k}}$ for $\hat{k} = k, k+1, k+2, \ldots$
$\hat{q}(z; x)$	Generating function of $q_k(x)$
$\pi(\mathbf{k}; \mathbf{x})$	Probability distribution on the d-dimensional integer lattice
$\hat{\pi}(\mathbf{z}; \mathbf{x})$	Generating function of $\pi(\mathbf{k}; \mathbf{x})$
$G(z)$	Generating function of a general discrete random variable X
$p_{k,l} = \mathbb{P}(\zeta_1 = k, \zeta_2 = l)$	Bivariate probability distribution of the bivariate random variable (ζ_1, ζ_2)
$\hat{p}(z, w)$	Generating function of $p_{k,l}$
$q_{k,l}(x, y)$	Probability that depends on $p_{\hat{k},\hat{l}}$ for $\hat{k} = k, k+1, \ldots$ and $\hat{l} = l, l+1, \ldots$
$\hat{q}(z, w)$	Generating function of $q_{k,l}$
$\kappa = \lambda_1/\lambda_2$	Ratio of the mean parameters of Poisson processes with intensities $\lambda_1 t$ and $\lambda_2 t$
$\rho(t) = \text{corr}(X_t^{(1)}, X_t^{(2)})$	Correlation at time t between $X_t^{(1)}$ and $X_t^{(2)}$

3.1 Backward Simulation

In this section, we prove the fundamental result, first given in [74], enabling the Backward Simulation of Poisson processes. We consider, for the remainder of this section, a process, X_t for $0 \leq t \leq T$, defined as

$$X_t = \sum_{i=1}^{n} \mathbb{1}(T_i \leq t), \tag{3.2}$$

where n is a realization of the random variable $X_T \sim \text{Pois}(\lambda T)$ and the random variables $\{T_i\}_{i=1}^n$ are iid having a uniform conditional distribution

$$\mathbb{P}(T_i \leq t \,|\, X_T = n) = \frac{t}{T} \quad i = 1, 2, \ldots, n \quad \text{and} \quad 0 \leq t \leq T \tag{3.3}$$

in the interval $[0, T]$. Theorem 6 below is essentially the converse of the well-known conditional uniformity result for Poisson distributions but adapted for Poisson processes, ensuring that discrete non-decreasing processes constructed according to (3.2) and (3.3) are indeed Poisson processes. It is crucial to note that, although Theorem 6 is formulated and proved in the univariate setting, the results extend directly to the multivariate setting. That is, Theorem 6 extends directly to the case where X_t is a correlated multivariate Poisson process. This is because of the fact that, while the dependence structure is specified through the joint distribution, every coordinate (marginal) of a correlated multivariate Poisson process is itself a (univariate) Poisson process[1]. Therefore, given the joint number of events sampled from a suitable joint distribution[2] at time T, the correlated multivariate Poisson process can be constructed within the simulation interval $[0, T]$ by applying Theorem 6 independently to each coordinate.

For the rest of the chapter, unless otherwise stated, N refers to a Poisson distributed random variable and N_t refers to a Poisson process, whereas X_t refers to a Poisson process obtained through Backward Simulation. Moreover, note that a bold typeface denotes the vector or multivariate counterpart. That is, $\mathbf{N} = (N^{(1)}, \ldots, N^{(d)})$ refers to a d-dimensional random vector that is multivariate Poisson distributed, $\mathbf{N}_t = (N_t^{(1)}, \ldots, N_t^{(d)})$ denotes a multivariate Poisson process and $\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(d)})$ denotes a multivariate Poisson process obtained through Backward Simulation.

Finally, we note that the results in this section can be applied directly to the multivariate setting since the conditional uniformity property of Poisson processes ensures that, given the terminal vector of (joint) events, each component's arrival times are just the order statistics of i.i.d Unif$(0, T)$ variables which are conditionally independent across components. Therefore, simulation reduces to generating the joint events (counts) and then simulating each component as a univariate Poisson process conditioned on the number of terminal events. The dependence between components arises entirely through the joint distribution by construction, each component of the multivariate Poisson process is a Poisson process with independent increments in the interval $[0, T]$.

The main result of this chapter is the following theorem.

Theorem 6 ([74]). *For $t \in [0, T]$, define the process X_t by*

$$X_t = \sum_{i=1}^{n} \mathbb{1}(T_i \leq t), \tag{3.4}$$

[1] See Definition 5.

[2] Recall that Chapter 2 is concerned with constructing extreme distributions that exhibit extreme correlations and constructing distributions that exhibit any *admissible* correlation that is a convex combination of extreme correlations.

where n is a realization of the random variable $X_T \sim \text{Pois}(\lambda T)$ and the random variables $\{T_i\}_{i=1}^n$ have the uniform conditional distribution (3.3). Then, X_t is a Poisson process with intensity λ in the interval $[0, T]$.

To prove Theorem 6, we must show that:

1. For any interval $(t, t+s] \subset [0, T]$ of length $s \geq 0$, the increments $\Delta_s X_t = X_{t+s} - X_t$ of the process X_t are Poisson distributed, with $\Delta_s X_t = X_{t+s} - X_t \sim \text{Pois}(\lambda s)$. Moreover, the distribution of the increments $\Delta_s X_t$ does not depend on t.

2. For any $l = 2, 3, \ldots$ disjoint sub-intervals $(t_i, t_i + s_i] \subset [0, T]$ for $i = 1, 2, \ldots, l$, the random variables $\Delta_{s_i} X_{t_i} = X_{t_i+s_i} - X_{t_i}$, for $i = 1, 2, \ldots, l$, are mutually independent.

We can prove points 1 and 2 above by applying Lemma 25 and Lemma 26, respectively, below. Lemma 25 is a statement about discrete generating functions and Lemma 26 is the extension of Lemma 25 to the vector setting. For this reason, we only prove Lemma 26. We state Lemma 25 for clarity of exposition.

Lemma 25 ([74]). *Consider a discrete random variable, ξ, taking non-negative integer values with probabilities $p_k = \mathbb{P}(\xi = k)$, $k = 0, 1, 2, \ldots$, and denote its generating function by*

$$\hat{p}(z) = \sum_{k=0}^{\infty} p_k z^k, \quad |\, z \,| \leq 1.$$

Consider a sequence

$$q_k(x) = \sum_{m=0}^{\infty} p_{k+m} \binom{k+m}{k} x^k (1-x)^m, \quad 0 \leq x \leq 1, \quad k = 0, 1, 2, \ldots . \tag{3.5}$$

Then, for any fixed $x \in [0, 1]$, the sequence $\{q_k(x)\}_{k=0}^{\infty}$ is a probability distribution and its generating function $\hat{q}(z; x)$ takes the form

$$\hat{q}(z; x) = \hat{p}(1 - x + xz). \tag{3.6}$$

To generalize Lemma 25 to the vector setting, we require some additional notation. For a general d-dimensional vector[3], $\mathbf{k} = (k_1, k_2, \ldots, k_d) \in \mathbb{N}_0^d$, with non-negative integer coordinates, $k_j \geq 0$, we denote the norm of the vector by

$$\|\mathbf{k}\| = \sum_{j=1}^{d} k_j.$$

For any d-dimensional vector, $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, with non-negative real coordinates ($x_j \geq 0$ for $j = 1, \ldots, d$) and $\mathbf{k} \in \mathbb{N}_0^d$, we denote

$$\mathbf{x}^{\mathbf{k}} := \prod_{j=1}^{d} {x_j}^{k_j}.$$

[3]We emphasize that the d here and in Lemma 26, below, refers to the dimension of a generic vector and does not refer to the dimensionality of multivariate Poisson processes.

We also introduce the multinomial coefficient

$$\binom{\mathbf{k}+l}{\mathbf{k}} := \frac{\left(l+\sum_{i=1}^{d} k_i\right)!}{l! \cdot \prod_{i=1}^{d} k_i!}.$$

Lemma 26. *Consider a discrete random variable, ξ, taking non-negative integer values with probabilities $p_k = \mathbb{P}(\xi = k), k = 0, 1, 2, \ldots,$ and denote its generating function by $\hat{p}(z) = \sum_{k=0}^{\infty} p_k z^k$, $|z| \leq 1$. Let $\boldsymbol{k} \in \mathbb{N}_0^d$ and consider the function $\pi : \mathbb{N}_0^d \to \mathbb{R}$ defined by*

$$\pi(\boldsymbol{k}; \boldsymbol{x}) = \sum_{l=0}^{\infty} p_{\|\boldsymbol{k}\|+l} \binom{\boldsymbol{k}+l}{\boldsymbol{k}} \cdot \boldsymbol{x}^{\boldsymbol{k}} \cdot y^l, \tag{3.7}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)$, $x_j \geq 0$, $\sum_{j=1}^{d} x_j \leq 1$ and $y = 1 - \sum_{j=1}^{d} x_j$. For any fixed $\boldsymbol{x}$, $\{\pi(\boldsymbol{k}; \boldsymbol{x})\}_{\boldsymbol{k} \in \mathbb{N}_0^d}$ is a probability distribution on the d-dimensional integer lattice. Furthermore, denote its generating function by

$$\hat{\pi}(\boldsymbol{z}; \boldsymbol{x}) := \sum_{\boldsymbol{k} \in \mathbb{N}_0^d} \pi(\boldsymbol{k}; \boldsymbol{x}) \boldsymbol{z}^{\boldsymbol{k}}, \tag{3.8}$$

where $\boldsymbol{z} = (z_1, z_2, \ldots, z_d)$ and $\max\{|z_1|, \ldots, |z_d|\} \leq 1$, then

$$\hat{\pi}(\boldsymbol{z}; \boldsymbol{x}) = \hat{p}\left(1 - \sum_{j=1}^{d} x_j(1 - z_j)\right). \tag{3.9}$$

PROOF: We first show (3.9) holds. To this end, note that the generating function (3.8) can be rewritten, by substituting in (3.7), as

$$\hat{\pi}(\mathbf{z}; \mathbf{x}) = \sum_{j=1}^{d} \sum_{k_j=0}^{\infty} \sum_{l=0}^{\infty} p_{\|\mathbf{k}\|+l} \cdot \mathbf{z}^{\mathbf{k}} \cdot \binom{\mathbf{k}+l}{\mathbf{k}} \cdot \mathbf{x}^{\mathbf{k}} \cdot y^l. \tag{3.10}$$

Let $n = l + \sum_{j=1}^{d} k_j$ and introduce the partial sums

$$K_J = \sum_{j=1}^{J} k_j, \quad J = 1, 2, \ldots, d \tag{3.11}$$

and note that $K_d = K_{d-1} + k_d$.

Then, by expanding the multinomial coefficient, using the partial sum defined above, and substituting $y = 1 - \sum_{j=1}^{d} x_j$ in (3.10), we see that

$$\begin{aligned}\hat{\pi}(\mathbf{z}; \mathbf{x}) = \sum_{n=0}^{\infty} p_n \cdot \sum_{k_1=0}^{n} \binom{n}{k_1} (x_1 z_1)^{k_1} \sum_{k_2=0}^{n-k_1} \binom{n-k_1}{k_2} (x_2 z_2)^{k_2} \sum_{k_3=0}^{n-K_2} \binom{n-K_2}{k_3} (x_3 z_3)^{k_3} \\ \cdots \sum_{k_d=0}^{n-K_{d-1}} \binom{n-K_{d-1}}{k_d} (x_d z_d)^{k_d} \cdot \left(1 - \sum_{j=1}^{d} x_j\right)^{n-K_d}.\end{aligned} \tag{3.12}$$

Denote the last sum in (3.12) by

$$S_d = \sum_{k_d=0}^{n-K_{d-1}} \binom{n-K_{d-1}}{k_d} (x_d z_d)^{k_d} \cdot \left(1 - \sum_{j=1}^{d} x_j\right)^{n-K_d}. \tag{3.13}$$

Recall the binomial theorem:

$$(x+y)^{n'} = \sum_{k=0}^{n'} \binom{n'}{k} x^k y^{n'-k}$$

and note that by making the following substitutions

$$\begin{aligned} n' &= n - K_{d-1}, \\ k &= k_d, \\ n' - k &= n - K_{d-1} - k_d = n - K_d, \\ x &= (x_d z_d), \\ y &= 1 - \sum_{j=1}^{d} x_j, \end{aligned}$$

in (3.13), we can apply the binomial theorem to S_d, obtaining

$$S_d = \left(1 - \sum_{j=1}^{d-1} x_j - x_d(1-z_d)\right)^{n-K_{d-1}}.$$

Applying this transformation recursively to the sums over k_j for $j = d-1, d-2, \ldots, 1$ in (3.12), we get

$$\begin{aligned} \hat{\pi}(\mathbf{z};\mathbf{x}) &= \sum_{n=0}^{\infty} p_n \cdot \left(1 - \sum_{j=1}^{d} x_j(1-z_j)\right)^n \\ &= \hat{p}\left(1 - \sum_{j=1}^{d} x_j(1-z_j)\right) \end{aligned}$$

as required.

Having proved that the generating function $\hat{\pi}(\mathbf{z};\mathbf{x})$ takes the form in (3.9), we can now easily show that, for any fixed $\mathbf{x}$, $\{\pi(\mathbf{k};\mathbf{x})\}_{\mathbf{k}\in\mathbb{N}_0^d}$ is a probability distribution on the d-dimensional integer lattice. Firstly, it is clear from (3.7) that for any $\mathbf{k} \geq 0$, $\pi(\mathbf{k};\mathbf{x}) \geq 0$, since $\mathbf{x} \geq 0$, $y \geq 0$ and all other terms therein are non-negative. Next, we need to show that

$$\sum_{\mathbf{k}\in\mathbb{N}_0^d} \pi(\mathbf{k};\mathbf{x}) = 1.$$

To that end, recall that, for a generating function, $G(z)$, of a discrete random variable X,

$$G(1) = \mathbb{P}(X=0) + \mathbb{P}(X=1) + \cdots.$$

Therefore, it suffices to show that $\hat{\pi}(\mathbf{z};\mathbf{x}) = 1$ when $\mathbf{z} = (1,1,\ldots,1) \in \mathbb{R}^d$ for all admissible choices of $\mathbf{x}$. Indeed, this can be seen directly from (3.9), since $\hat{p}(1) = 1$. Thus, $\{\pi(\mathbf{k};\mathbf{x})\}_{\mathbf{k}\in\mathbb{N}_0^d}$ is a probability

distribution on the d-dimensional integer lattice for any fixed $\mathbf{x} \in [0,1]^d$ satisfying $\sum_{j=1}^{d} x_j \leq 1$. □

Having proved Lemma 26 we can show the main result of this section.

Proof of Theorem 6

PROOF: We begin by proving the first point of the Theorem. To this end, let us denote by $\Delta_s X_t = X_{t+s} - X_t$, the number of events occurring in the interval $[t, t+s]$ of the process X_t. Then, the probability that k events occur in the interval $[t, t+s]$ of the process can be expressed, by the law of total probability, in terms of the conditional probability, as

$$\mathbb{P}(\Delta_s X_t = k) = \sum_{m=0}^{\infty} \mathbb{P}(\Delta_s X_t = k \,|\, X_T = k+m) \cdot \mathbb{P}(X_T = k+m). \tag{3.14}$$

Now note that

$$\mathbb{P}(\Delta_s X_t = k \,|\, X_T = k+m) = \binom{k+m}{k} \left(\frac{s}{T}\right)^k \left(1 - \frac{s}{T}\right)^m, \quad m = 0, 1, \ldots \tag{3.15}$$

since the conditional probability of k events occurring in the sub-interval $[t, t+s] \subset [0, T]$ given a total of $X_T = n = k+m$ independent, uniformly distributed events occurring in the full interval $[0, T]$ is equal to choosing k out of $n = k+m$ total events in the interval $[t, t+s]$ of length s, with the rest of the events being in the remainder of the interval $[0, T]$. Moreover, since X_T is Poisson distributed (and known at terminal simulation time T), its generating function takes the form

$$\hat{p}(z) := \mathbb{E}\left[z^{X_T}\right] = \exp(\lambda T(z-1)). \tag{3.16}$$

We use Lemma 25 to show that the generating function of $\Delta_s X_t$ also takes the form (3.16), but with T replaced by s. Thus, $\Delta_s X_t$ has a Poisson distribution with parameter λs. To this end, let $p_k = \mathbb{P}(X_T = k)$, for any $k \in \mathbb{N}_0$, and $x = s/T$. Then, note that, by the law of total probability and (3.15)

$$\begin{aligned}
\mathbb{P}(\Delta_s X_t = k) &= \sum_{m=0}^{\infty} \mathbb{P}(\Delta_s X_t = k \,|\, X_T = k+m) \cdot \mathbb{P}(X_T = k+m) \\
&= \sum_{m=0}^{\infty} \binom{k+m}{k} \left(\frac{s}{T}\right)^k \left(1 - \frac{s}{T}\right)^m \cdot \mathbb{P}(X_T = k+m) \\
&= \sum_{m=0}^{\infty} p_{k+m} \binom{k+m}{k} x^k (1-x)^m.
\end{aligned}$$

Observe that the last line in the set of equations above has the form of $q_k(x)$ in Lemma 25. So, let

$$q_k(x) = \sum_{m=0}^{\infty} p_{k+m} \binom{k+m}{k} x^k (1-x)^m = \mathbb{P}(\Delta_s X_t = k), \qquad k = 0, 1, 2, \ldots .$$

Since $q_k(x) = \mathbb{P}(\Delta_s X_t = k)$, for $k = 0, 1, 2, \ldots$, the generating function $\mathbb{E}\left[z^{\Delta_s X_t}\right]$ of the random variable $\Delta_s X_t$ must be $\hat{q}(z; x)$, the generating function associated with $q_k(x)$, for $k = 0, 1, 2, \ldots$. From this

observation and Lemma 25, it follows that

$$\begin{aligned}\mathbb{E}\left[z^{\Delta_s X_t}\right] &= \hat{q}(z;x)\\ &= \hat{p}(1-x+xz)\\ &= \hat{p}(1-\frac{s}{T}+\frac{s}{T}z)\\ &= \hat{p}(1+\frac{s}{T}(z-1))\\ &= \exp(\lambda T((1+\frac{s}{T}(z-1))-1))\\ &= \exp(\lambda s(z-1))\end{aligned}$$

where, in the fourth line above, we substituted (3.16). The set of equations above shows that the generating function of $\Delta_s X_t$ is $\exp(\lambda s(z-1))$, which is the generating function of a Poisson random variable with parameter λs. Therefore, the increments $\Delta_s X_t = X_{t+s} - X_t$ of the process X_t are Poisson distributed with $\Delta_s X_t \sim \text{Pois}(\lambda s)$. Moreover, the distribution of the increments $\Delta_s X_t$ does not depend on t.

Next, we prove the second point of the Theorem. To this end, for any $l = 2, 3, \ldots$, consider any l disjoint sub-intervals $(t_i, t_i + s_i]$, for $i = 1, 2, \ldots, l$, where each $(t_i, t_i + s_i] \subset [0, T]$. For $i = 1, 2, \ldots, l$, denote by $\Delta_{s_i} X_{t_i} = X_{t_i+s_i} - X_{t_i}$ the number of events occurring within the sub-interval $(t_i, t_i + s_i]$. For $i = 1, 2, \ldots, l$, let $x_i = (t_i + s_i - t_i)/T = s_i/T \in [0, 1]$. Also, let $\mathbf{x} = (x_1, \ldots, x_l)$ and $y = 1 - \sum_{i=1}^{l} x_i$. Note that $y \in [0, 1]$. Since the intervals $(t_i, t_i + s_i]$, for $i = 1, 2, \ldots, l$, are disjoint and the events have the uniform conditional distribution (3.3), it follows that

$$\mathbb{P}\left(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l \,|\, X_T = m + \sum_{i=1}^{l} k_i\right) = \binom{\mathbf{k}+m}{\mathbf{k}} \cdot \mathbf{x}^{\mathbf{k}} \cdot y^m. \tag{3.17}$$

For any $k \in \mathbb{N}_0$, let $p_k = \mathbb{P}(X_T = k)$. Then, using (4.14) and the law of total probability, we see that

$$\begin{aligned}&\mathbb{P}\big(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l\big)\\ &= \sum_{m=0}^{\infty} \mathbb{P}\left(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l \,|\, X_T = m + \sum_{i=1}^{l} k_i\right) \cdot \mathbb{P}\left(X_T = m + \sum_{i=1}^{l} k_i\right)\\ &= \sum_{m=0}^{\infty} \binom{\mathbf{k}+m}{\mathbf{k}} \cdot \mathbf{x}^{\mathbf{k}} \cdot y^m \cdot \mathbb{P}(X_T = \|\mathbf{k}\| + m)\\ &= \sum_{m=0}^{\infty} p_{\|\mathbf{k}\|+m} \binom{\mathbf{k}+m}{\mathbf{k}} \cdot \mathbf{x}^{\mathbf{k}} \cdot y^m.\end{aligned}$$

Observe that the last line in the set of equations above has the form of $\pi(\mathbf{k}; \mathbf{x})$ in Lemma 26. So, for all $\mathbf{k} = (k_1, \ldots, k_l) \in \mathbb{N}_0^l$, let

$$\pi(\mathbf{k}; \mathbf{x}) = \sum_{m=0}^{\infty} p_{\|\mathbf{k}\|+m} \binom{\mathbf{k}+m}{\mathbf{k}} \cdot \mathbf{x}^{\mathbf{k}} \cdot y^m = \mathbb{P}\big(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l\big).$$

Since, for all $\mathbf{k} = (k_1, \ldots, k_l) \in \mathbb{N}_0^l$, $\pi(\mathbf{k}; \mathbf{x}) = \mathbb{P}\big(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l\big)$, the generating function $\mathbb{E}\left[z_1^{\Delta_{s_1} X_{t_1}} \cdots z_l^{\Delta_{s_l} X_{t_l}}\right]$ of the joint probability distribution of the increments $\Delta_{s_1} X_{t_1}, \ldots, \Delta_{s_l} X_{t_l}$ must be

$\hat{\pi}(\mathbf{z};\mathbf{x})$, the generating function associated with the probabilities $\pi(\mathbf{k};\mathbf{x})$, $\mathbf{k} \in \mathbb{N}_0^l$. From this observation and Lemma 26 with $\hat{p}(z) = \mathbb{E}\left[z^{X_T}\right] = \exp(\lambda T(z-1))$, it follows that

$$\begin{aligned}
\mathbb{E}\left[z_1^{\Delta_{s_1}X_{t_1}} \cdots z_l^{\Delta_{s_l}X_{t_l}}\right] &= \hat{\pi}(\mathbf{z};\mathbf{x}) \\
&= \hat{p}\Big(1 - \sum_{i=1}^{l} x_i(1-z_i)\Big) \\
&= \exp\left(\lambda T\big(1 - \sum_{i=1}^{l} x_i(1-z_i) - 1\big)\right) \\
&= \exp\left(\sum_{i=1}^{l} \lambda T x_i(z_i - 1)\right) \\
&= \prod_{i=1}^{l} \exp(\lambda T x_i(z_i-1)) \\
&= \prod_{i=1}^{l} \exp(\lambda s_i(z_i-1)) \\
&= \prod_{i=1}^{l} \mathbb{E}\left[z_i^{\Delta_{s_i}X_{t_i}}\right]
\end{aligned}$$

where the last line above follows from our proof of the first point of the Theorem, which implies that

$$\mathbb{E}\left[z_i^{\Delta_{s_i}X_{t_i}}\right] = \exp(\lambda s_i(z_i-1)) \qquad \text{for } i = 1, 2, \ldots, l.$$

We see from the set of equations above that the generating function of the joint distribution of the increments $\Delta_{s_1}X_{t_1}, \ldots, \Delta_{s_l}X_{t_l}$ factors multiplicatively into the product of the generating functions of the individual increments $\Delta_{s_i}X_{t_i}$, $i = 1, \ldots, l$. Therefore, the increments $\Delta_{s_i}X_{t_i}$, $i = 1, \ldots, l$, are mutually independent. That is, the process X_t for $0 < t \leq T$ has independent increments.

□

3.1.1 Backward Simulation Algorithm in d-dimensions

Algorithm 19 Backward Simulation of correlated multivariate Poisson processes

Require: Vector of marginal Poisson distributions at terminal time T

$\mathbf{Pois}(\boldsymbol{\lambda} T) = (\text{Pois}(\lambda_1 T), \ldots, \text{Pois}(\lambda_d T))$

Correlation matrix C

Output: Scenarios of the multivariate Poisson process in $[0, T]$

1: Construct $\text{Pois}(\lambda_k T)$ distributed marginals $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]$ for $k = 1, \ldots, d$

2: Generate the calibrated Poisson measure $P^{(C)}$ using Algorithm 9, which takes as input $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]_{k=1}^{d}$ and C

3: Generate samples $(N^{(1)}, \ldots, N^{(d)}) \sim P^{(C)}$ using Algorithm 18 ▷ *Get the number of events at terminal time T*

4: **for** $k = 1, \ldots, d$ **do** ▷ *this can be done in parallel*

5: Generate $N^{(k)}$ iid uniform random variables in the interval $[0, T]$: $\mathbf{T}^{(k)} = (T_1^{(k)}, \ldots, T_{N^{(k)}}^{(k)})'$ ▷ *column vector*

6: Sort $\mathbf{T}^{(k)}$ in ascending order

7: **return** $\mathbf{T} = (\mathbf{T}^{(1)}, \ldots, \mathbf{T}^{(d)})$

Remark 62. *More sophisticated numerical methods, such as Quasi Monte Carlo (QMC) [45, 46], can be used to implement the Backward Simulation of Poisson processes, to achieve a much faster rate of convergence.*

3.2 Correlation Structure

The main result of this section (Theorem 7) describes, analytically, the behavior of the correlation coefficient between a pair of Poisson processes generated using Backward Simulation within the simulation interval $[0, T]$. The exposition in this section is in the bivariate setting since Pearson correlation is inherently a bivariate concept. The results in this section extend directly to the multivariate setting by application of the results to each pair of components of a multivariate Poisson process.

We prove Theorem 7 using generating functions. To that end, let us first prove Lemma 27 below, which can be straightforwardly applied to show Theorem 7. For the remainder of the section, consider a bivariate integer-valued random vector $\zeta = (\zeta_1, \zeta_2)$ and denote its probability distribution by

$$p_{k,l} = \mathbb{P}(\zeta_1 = k, \zeta_2 = l), \quad k, l = 0, 1, 2, \dots \tag{3.18}$$

and its generating function by

$$\hat{p}(z, w) := \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} p_{k,l} \cdot z^k w^l, \quad |z| \leq 1, \quad |w| \leq 1. \tag{3.19}$$

Furthermore, consider another bivariate integer-valued random vector $\xi = (\xi_1, \xi_2)$, such that, for all $k = 0, 1, 2, \dots, k'$, $k' = 0, 1, 2, \dots$, and $l = 0, 1, 2, \dots, l'$, $l' = 0, 1, 2, \dots$, ξ has the conditional probability distribution

$$\mathbb{P}\big(\xi_1 = k, \xi_2 = l \,|\zeta_1 = k', \zeta_2 = l'\big) = \binom{k'}{k} x^k (1-x)^{k'-k} \cdot \binom{l'}{l} y^l (1-y)^{l'-l}, \tag{3.20}$$

where x and y are fixed values satisfying $0 \leq x \leq 1$ and $0 \leq y \leq 1$. By setting $k' = k + m$ and $l' = l + n$ in (3.20) and using (3.20) together with the law of total probability, for any $k, l \in \{0, 1, 2, \dots\}$, we can write the unconditional joint probability of $\xi = (\xi_1, \xi_2)$ as

$$\begin{aligned} \mathbb{P}(\xi_1 = k, \xi_2 = l) &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \mathbb{P}\big(\xi_1 = k, \xi_2 = l \,|\zeta_1 = k + m, \zeta_2 = l + n\big) \cdot \mathbb{P}(\zeta_1 = k + m, \zeta_2 = l + n) \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} p_{k+m,l+n} \binom{k+m}{k} x^k (1-x)^{k+m-k} \cdot \binom{l+n}{l} y^l (1-y)^{l+n-l}, \end{aligned} \tag{3.21}$$

where $p_{k+m,l+n} = \mathbb{P}(\zeta_1 = k + m, \zeta_2 = l + n)$. Note that the last line in the set of equations above has a form analogous to $q_k(x)$ in Lemma 25 but in two-dimensions. Therefore, let

$$\begin{aligned} q_{k,l}\,(x, y) &= \mathbb{P}(\xi_1 = k, \xi_2 = l) \\ &= \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} p_{k+m,l+n} \cdot \binom{k+m}{k} x^k (1-x)^m \cdot \binom{l+n}{l} y^l (1-y)^n \quad k, l = 0, 1, \dots . \end{aligned} \tag{3.22}$$

Lemma 27 ([74]). *Suppose that the variance and the first moment of the random variables ζ_1 and ζ_2 are equal, that is, $\mathbb{E}[\zeta_i] = \sigma^2(\zeta_i)$ for $i \in \{1, 2\}$. Then, the generating function*

$$\hat{q}(z, w) := \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} q_{k,l}(x, y)\, z^k w^l$$

associated with the bivariate random vector ξ satisfies the relation

$$\hat{q}(z,w) = \hat{p}(1-x+xz, 1-y+yw), \quad |z| \leq 1, \quad |w| \leq 1, \tag{3.23}$$

where $\hat{p}(z,w)$ given in (3.19) is the generating function of the bivariate random vector ζ. Moreover, the Pearson correlation coefficient of the random variables ξ_1 and ξ_2 satisfies

$$\rho(\xi_1, \xi_2) = \sqrt{xy} \cdot \rho(\zeta_1, \zeta_2), \tag{3.24}$$

where $\rho(\xi_1, \xi_2)$ and $\rho(\zeta_1, \zeta_2)$ denote the Pearson correlation coefficients between the components of the bivariate random variables ξ and ζ, respectively.

PROOF: The first part of the Lemma is analogous to Lemma 25. Therefore, for brevity, we omit its proof. We focus on the second part of the Lemma and show (3.24).

To begin, since $\hat{q}(z,w)$ is the generating function of bivariate random vector $\xi = (\xi_1, \xi_2)$, it follows from the definition of a generating function that

$$\hat{q}(z,w) = \mathbb{E}\left[z^{\xi_1} w^{\xi_2}\right]. \tag{3.25}$$

Differentiating (3.25), we get

$$\begin{aligned}
\frac{\partial \hat{q}(z,w)}{\partial z} &= \mathbb{E}\left[\xi_1 \, z^{\xi_1 - 1} \, w^{\xi_2}\right], \\
\frac{\partial \hat{q}(z,w)}{\partial w} &= \mathbb{E}\left[z^{\xi_1} \, \xi_2 \, w^{\xi_2 - 1}\right], \\
\frac{\partial^2 \hat{q}(z,w)}{\partial z^2} &= \mathbb{E}\left[\xi_1(\xi_1 - 1) \, z^{\xi_1 - 2} \, w^{\xi_2}\right], \\
\frac{\partial^2 \hat{q}(z,w)}{\partial z \partial w} &= \mathbb{E}\left[\xi_1 \, z^{\xi_1 - 1} \, \xi_2 \, w^{\xi_2 - 1}\right], \\
\frac{\partial^2 \hat{q}(z,w)}{\partial w^2} &= \mathbb{E}\left[z^{\xi_1} \, \xi_2(\xi_2 - 1) \, w^{\xi_2 - 2}\right].
\end{aligned}$$

Evaluating the equations above at $z = w = 1$, we get the well-know results

$$\left.\frac{\partial \hat{q}(z,w)}{\partial z}\right|_{z=w=1} = \mathbb{E}\left[\xi_1\right], \tag{3.26}$$

$$\left.\frac{\partial \hat{q}(z,w)}{\partial w}\right|_{z=w=1} = \mathbb{E}\left[\xi_2\right], \tag{3.27}$$

$$\left.\frac{\partial^2 \hat{q}(z,w)}{\partial z^2}\right|_{z=w=1} = \mathbb{E}\left[\xi_1(\xi_1 - 1)\right] = \mathbb{E}\left[\xi_1^2\right] - \mathbb{E}\left[\xi_1\right], \tag{3.28}$$

$$\left.\frac{\partial^2 \hat{q}(z,w)}{\partial z \partial w}\right|_{z=w=1} = \mathbb{E}\left[\xi_1 \, \xi_2\right], \tag{3.29}$$

$$\left.\frac{\partial^2 \hat{q}(z,w)}{\partial w^2}\right|_{z=w=1} = \mathbb{E}\left[\xi_2(\xi_2 - 1)\right] = \mathbb{E}\left[\xi_2^2\right] - \mathbb{E}\left[\xi_2\right]. \tag{3.30}$$

On the other hand, differentiating (3.23) and using (3.19), we also get

$$\begin{aligned}\frac{\partial \hat{q}(z,w)}{\partial z} &= \frac{\partial}{\partial z}\big(\hat{p}(1-x+xz, 1-y+yw)\big)\\ &= \frac{\partial}{\partial z}\left(\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} p_{k,l}(1-x+xz)^k(1-y+yw)^l\right)\\ &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} p_{k,l}\cdot k\cdot(1-x+xz)^{k-1}\cdot x\cdot(1-y+yw)^l.\end{aligned}$$

Similarly, we get

$$\begin{aligned}\frac{\partial \hat{q}(z,w)}{\partial w} &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} p_{k,l}\cdot(1-x+xz)^k\cdot l\cdot(1-y+yw)^{l-1}\cdot y,\\ \frac{\partial^2 \hat{q}(z,w)}{\partial z^2} &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} p_{k,l}\cdot k(k-1)\cdot(1-x+xz)^{k-2}\cdot x^2\cdot(1-y+yw)^l,\\ \frac{\partial^2 \hat{q}(z,w)}{\partial z\partial w} &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} p_{k,l}\cdot k\cdot(1-x+xz)^{k-1}\cdot x\cdot l\cdot(1-y+yw)^{l-1}\cdot y,\\ \frac{\partial^2 \hat{q}(z,w)}{\partial w^2} &= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} p_{k,l}\cdot(1-x+xz)^k\cdot l(l-1)\cdot(1-y+yw)^{l-2}\cdot y^2.\end{aligned}$$

Evaluating the equations above at $z=w=1$ and using (3.18), we get

$$\left.\frac{\partial \hat{q}(z,w)}{\partial z}\right|_{z=w=1} = x\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} k\cdot p_{k,l} = x\,\mathbb{E}\left[\zeta_1\right], \tag{3.31}$$

$$\left.\frac{\partial \hat{q}(z,w)}{\partial w}\right|_{z=w=1} = y\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} l\cdot p_{k,l} = y\,\mathbb{E}\left[\zeta_2\right], \tag{3.32}$$

$$\left.\frac{\partial^2 \hat{q}(z,w)}{\partial z^2}\right|_{z=w=1} = x^2\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} k(k-1)\cdot p_{k,l} = x^2\,\mathbb{E}\left[\zeta_1^2\right] - x^2\,\mathbb{E}\left[\zeta_1\right], \tag{3.33}$$

$$\left.\frac{\partial^2 \hat{q}(z,w)}{\partial z\partial w}\right|_{z=w=1} = xy\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} k\,l\cdot p_{k,l} = xy\,\mathbb{E}\left[\zeta_1\,\zeta_2\right], \tag{3.34}$$

$$\left.\frac{\partial^2 \hat{q}(z,w)}{\partial w^2}\right|_{z=w=1} = y^2\sum_{k=0}^{\infty}\sum_{l=0}^{\infty} l(l-1)\cdot p_{k,l} = y^2\,\mathbb{E}\left[\zeta_2^2\right] - y^2\,\mathbb{E}\left[\zeta_2\right]. \tag{3.35}$$

It follows from (3.26) and (3.31) that

$$\mathbb{E}\left[\xi_1\right] = x\,\mathbb{E}\left[\zeta_1\right], \tag{3.36}$$

from (3.27) and (3.32) that

$$\mathbb{E}\left[\xi_2\right] = y\,\mathbb{E}\left[\zeta_2\right], \tag{3.37}$$

and from (3.29) and (3.34) that

$$\mathbb{E}\left[\xi_1\,\xi_2\right] = xy\,\mathbb{E}\left[\zeta_1\,\zeta_2\right]. \tag{3.38}$$

Using (3.36), (3.37) and (3.38), we get that

$$\begin{aligned}\operatorname{Cov}(\xi_1, \xi_2) &= \mathbb{E}\left[\xi_1, \xi_2\right] - \mathbb{E}\left[\xi_1\right]\mathbb{E}\left[\xi_2\right] \\ &= \left(xy\,\mathbb{E}\left[\zeta_1\,\zeta_2\right]\right) - \left(x\,\mathbb{E}\left[\zeta_1\right]\right)\,\left(y\,\mathbb{E}\left[\zeta_2\right]\right) \\ &= xy\left(\mathbb{E}\left[\zeta_1\,\zeta_2\right] - \mathbb{E}\left[\zeta_1\right]\mathbb{E}\left[\zeta_2\right]\right) \\ &= xy\,\operatorname{Cov}(\zeta_1, \zeta_2).\end{aligned} \tag{3.39}$$

Using (3.26), (3.28), (3.31), (3.33) and the assumption in the Lemma that $\mathbb{E}\left[\zeta_1\right] = \sigma^2(\zeta_1)$, we get that

$$\begin{aligned}\sigma^2(\xi_1) &= \mathbb{E}\left[\xi_1^2\right] - \left(\mathbb{E}\left[\xi_1\right]\right)^2 \\ &= \left(\mathbb{E}\left[\xi_1^2\right] - \mathbb{E}\left[\xi_1\right]\right) + \mathbb{E}\left[\xi_1\right] - \left(\mathbb{E}\left[\xi_1\right]\right)^2 \\ &= \left.\frac{\partial^2 \hat{q}(z,w)}{\partial z^2}\right|_{z=w=1} + \left.\frac{\partial \hat{q}(z,w)}{\partial z}\right|_{z=w=1} - \left(\left.\frac{\partial \hat{q}(z,w)}{\partial z}\right|_{z=w=1}\right)^2 \\ &= \left(x^2\,\mathbb{E}\left[\zeta_1^2\right] - x^2\,\mathbb{E}\left[\zeta_1\right]\right) + x\,\mathbb{E}\left[\zeta_1\right] - \left(x\,\mathbb{E}\left[\zeta_1\right]\right)^2 \\ &= x^2\,\left(\mathbb{E}\left[\zeta_1^2\right] - \left(\mathbb{E}\left[\zeta_1\right]\right)^2\right) + \left(x - x^2\right)\mathbb{E}\left[\zeta_1\right] \\ &= x^2\,\sigma^2(\zeta_1) + (x - x^2)\sigma^2(\zeta_1) \\ &= x\,\sigma^2(\zeta_1).\end{aligned} \tag{3.40}$$

Using an argument similar to that used to establish (3.40), we get that

$$\sigma^2(\xi_2) = y\,\sigma^2(\zeta_2). \tag{3.41}$$

From (3.39), (3.40) and (3.41), it follows that

$$\begin{aligned}\rho(\xi_1, \xi_2) &= \frac{\operatorname{Cov}(\xi_1, \xi_2)}{\sigma(\xi_1)\,\sigma(\xi_2)} \\ &= \frac{xy\,\operatorname{Cov}(\zeta_1, \zeta_2)}{\sqrt{x}\,\sigma(\zeta_1)\,\sqrt{y}\,\sigma(\zeta_2)} \\ &= \sqrt{xy}\,\frac{\operatorname{Cov}(\zeta_1, \zeta_2)}{\sigma(\zeta_1)\,\sigma(\zeta_2)} \\ &= \sqrt{xy}\,\rho(\zeta_1\,\zeta_2).\end{aligned}$$

Therefore, we have proven that (3.24) holds.

□

Theorem 7 (Time Structure of the Correlation Coefficient [74])**.** *Consider a bivariate Poisson process* $(X_t^{(1)}, X_t^{(2)})$ *such that the sample paths of* $X_t^{(1)}$ *and* $X_t^{(2)}$ *are generated by Backward Simulation in the interval* $[0, T]$*. Let the correlation coefficient at time* T *be* $\rho(T) := \text{corr}(X_T^{(1)}, X_T^{(2)})$*. Then* $\rho(t) = \text{corr}(X_t^{(1)}, X_t^{(2)})$ *satisfies*

$$\rho(t) = \rho(T) \cdot \frac{t}{T}, \quad 0 < t \leq T. \tag{3.42}$$

Proof: Lemma 27 can be used to show Theorem 7 without much further work by setting $\zeta = (N_T^{(1)}, N_T^{(2)})$ and $\xi = (N_t^{(1)}, N_t^{(2)})$. The conditional probabilities $\mathbb{P}(N_t^{(1)} = k, N_t^{(2)} = l \,|\, N_T^{(1)} = k', N_T^{(2)} = l')$ satisfy (3.20) with $x = y = tT^{-1}$, $\rho(\xi_1, \xi_2) = \text{corr}(N_t^{(1)}, N_t^{(2)})$, and $\rho(\zeta_1, \zeta_2) = \text{corr}(N_T^{(1)}, N_T^{(2)})$. Making these substitutions in (3.24) implies (3.42). □

Figure 3.1 depicts the correlation structure obtained from simulating a bivariate Poisson process with intensities 3 and 5, calibrated to correlations of 0.7 and -0.9, respectively, at terminal time $T = 5$, within the simulation interval $[0, 5]$ under Backward Simulation. We used Backward Simulation with 1,000,000 Monte Carlo samples to compute the dashed black lines in Figure 3.1. The blue circles depict the theoretical values according to Theorem 7. Note the good agreement between the theoretical and the empirical results. Also note that the extreme positive and extreme negative correlations in this case are 0.9955 and -0.9897, respectively, and that any correlation, at the terminal time $T = 5$, within the interval $[-0.9897, 0.9955]$ is attainable under our Backward Simulation approach.

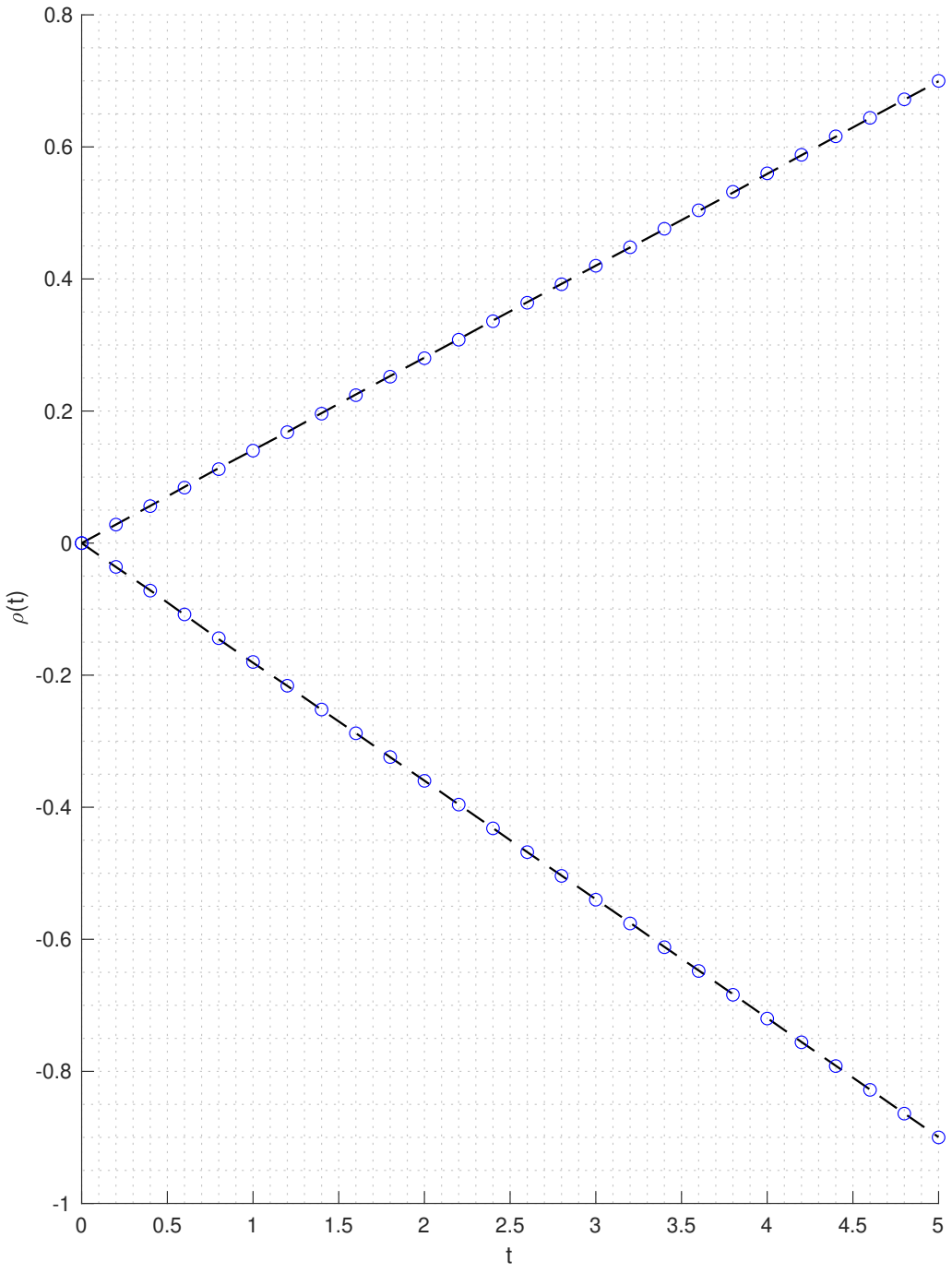

Figure 3.1: The dashed black lines depict the correlation structure for two bivariate Poisson process, each with intensities 3 and 5, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.9$, respectively, computed by Backward Simulation. The blue circles depict the theoretical values according to Theorem 7.

3.3 Forward Continuation of the Backward Simulation

In Forward Simulation, the notion of continuing forward a process past the original simulation interval $[0, T]$ is natural and, for this reason, not given special consideration as this is simply accomplished by continuing the Forward Simulation for another time step. In Backward Simulation, however, continuing forward a process past the original simulation interval is not as natural and requires some consideration.

One approach to continuing forward a process past its original simulation interval, which we denote the Forward Continuation of Backward Simulation[4], is as follows. First, consider the case where a bivariate Poisson process $(X_t^{(1)}, X_t^{(2)})$ has been simulated in the interval $[0, T]$ using Backward Simulation and assume that we wish to continue forward the process to the subsequent interval $[T, 2T]$. To begin, we draw an independent sample from the joint distribution at time T:

$$(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)}) \overset{d}{=} (X_T^{(1)}, X_T^{(2)}). \tag{3.43}$$

The bivariate random variable $(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$ has the same prescribed Poisson marginals and prescribed correlation C between its components as $(X_T^{(1)}, X_T^{(2)})$. Next, we set

$$(X_{2T}^{(1)}, X_{2T}^{(2)}) = (X_T^{(1)}, X_T^{(2)}) + (\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)}). \tag{3.44}$$

As in Backward Simulation, we use $(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$ and the conditional independence property (similar to (3.3)) to generate the uniform arrival moments $\{T_i^{(1)}\}_{i=1}^{\Delta_T X_T^{(1)}}$ and $\{T_i^{(2)}\}_{i=1}^{\Delta_T X_T^{(2)}}$ in the interval $[T, 2T]$. This process can be repeatedly applied to extend a bivariate Poisson process simulated by Backwards Simulation within the interval $[0, T]$ to any subsequent interval $[mT, (m+1)T]$ for any non-negative integer m and is detailed in Algorithm 20 in Section 3.3.2. We call this algorithm Forward-Backward Simulation since it starts with one step of Backward Simulation followed by $m - 1$ steps of Forward Continuation of Backward Simulation.

We emphasize that the Forward Continuation method introduced here *does not* require the underlying multivariate Poisson process to have independent increments. Backward Simulation relies solely on the conditional uniformity of arrival times given the terminal count $N_T = n$, a property that holds irrespective of any increment-independence assumptions. While each component of the multivariate Poisson process has independent increments across time (simulation) intervals, the dependence structure *across* components is introduced explicitly through the calibrated joint distribution of increments rather than imposed by a Lévy-type assumption. Therefore, the method does not implicitly construct a Lévy process, since a Lévy process has independent increments, whereas in our construction the increments between intervals may exhibit arbitrary dependence—including negative correlation—as determined by the EJD based calibration.

Remark 63. *Although the exposition in this section is in the bivariate setting, the results extend straightforwardly to the general d-dimensional setting.*

3.3.1 Forward Correlation Structure

We analyze the behavior of the correlation coefficient under the Forward Continuation of Backward Simulation by deriving an expression for $\rho(mT + \tau) = \text{corr}(X_{mT+\tau}^{(1)}, X_{mT+\tau}^{(2)})$ as a function of $\rho(T) =$

[4]This work was published in [17]

$\text{corr}(X_T^{(1)}, X_T^{(2)})$ for any non-negative integer m and any $\tau \in [0,T]$ in Theorem 8 below. That $\rho(mT+\tau)$ attains asymptotic stationarity follows by Corollary 3 below.

Theorem 8. *The correlation coefficient* $\rho(mT+\tau) = \text{corr}(X_{mT+\tau}^{(1)}, X_{mT+\tau}^{(2)})$ *for any non-negative integer* m *and any* $\tau \in [0,T]$ *as a function of* $\rho(T) = \text{corr}(X_T^{(1)}, X_T^{(2)})$ *takes the form*

$$\rho(mT+\tau) = \rho(T) \cdot \frac{m + \tau^2 \cdot T^{-2}}{m + \tau T^{-1}}. \tag{3.45}$$

PROOF: First note that, for $m = 0$, (3.45) follows from Theorem 7. To show that (3.45) also holds for all $m \geq 1$, we begin by deriving an expression for $\rho(T+\tau)$. To this end, let $\Delta_\tau X_T^{(1)} = X_{T+\tau}^{(1)} - X_T^{(1)}$ and $\Delta_\tau X_T^{(2)} = X_{T+\tau}^{(2)} - X_T^{(2)}$ and note that each of the increments $\Delta_\tau X_T^{(1)}$ and $\Delta_\tau X_T^{(2)}$ is independent of both $X_T^{(1)}$ and $X_T^{(2)}$. Therefore, the covariance of the bivariate Poisson process at time $T+\tau$ satisfies

$$\begin{aligned} \text{Cov}(X_{T+\tau}^{(1)}, X_{T+\tau}^{(2)}) &= \text{Cov}(X_T^{(1)} + \Delta_\tau X_T^{(1)}, X_T^{(2)} + \Delta_\tau X_T^{(2)}) \\ &= \text{Cov}(X_T^{(1)}, X_T^{(2)}) + \text{Cov}(\Delta_\tau X_T^{(1)}, \Delta_\tau X_T^{(2)}) \end{aligned} \tag{3.46}$$

and, for each $i \in \{1,2\}$, the variance satisfies

$$\begin{aligned} \sigma^2(X_{T+\tau}^{(i)}) &= \sigma^2(X_T^{(i)} + \Delta_\tau X_T^{(i)}) \\ &= \sigma^2(X_T^{(i)}) + \sigma^2(\Delta_\tau X_T^{(i)}). \end{aligned} \tag{3.47}$$

Moreover, using an argument similar to that used to prove Lemma 27 and Theorem 7, it can be shown that

$$\text{Cov}(\Delta_\tau X_T^{(1)}, \Delta_\tau X_T^{(2)}) = \text{Cov}(X_T^{(1)}, X_T^{(2)}) \cdot \frac{\tau^2}{T^2} \tag{3.48}$$

and, for each $i \in \{1,2\}$,

$$\sigma^2(\Delta_\tau X_T^{(i)}) = \sigma^2(X_T^{(i)}) \cdot \frac{\tau}{T}. \tag{3.49}$$

It follows from (3.46) and (3.48) that

$$\begin{aligned} \text{Cov}(X_{T+\tau}^{(1)}, X_{T+\tau}^{(2)}) &= \text{Cov}(X_T^{(1)}, X_T^{(2)}) + \text{Cov}(\Delta_\tau X_T^{(1)}, \Delta_\tau X_T^{(2)}) \\ &= \text{Cov}(X_T^{(1)}, X_T^{(2)}) \left(1 + \frac{\tau^2}{T^2}\right) \end{aligned} \tag{3.50}$$

and, from (3.47) and (3.49), that

$$\begin{aligned} \sigma^2(X_{T+\tau}^{(i)}) &= \sigma^2(X_T^{(i)}) + \sigma^2(\Delta_\tau X_T^{(i)}) \\ &= \sigma^2(X_T^{(i)}) \left(1 + \frac{\tau}{T}\right) \end{aligned} \tag{3.51}$$

for $i \in \{1,2\}$. Dividing (3.50) through by $\sigma(X_{T+\tau}^{(1)})\,\sigma(X_{T+\tau}^{(2)})$ and using (3.51), we get that

$$\frac{\text{Cov}(X_{T+\tau}^{(1)}, X_{T+\tau}^{(2)})}{\sigma(X_{T+\tau}^{(1)})\sigma(X_{T+\tau}^{(2)})} = \frac{\text{Cov}(X_T^{(1)}, X_T^{(2)})}{\sigma(X_T^{(1)})\sigma(X_T^{(2)})} \cdot \frac{(1 + \frac{\tau^2}{T^2})}{1 + \frac{\tau}{T}}$$

whence

$$\rho(T+\tau) = \rho(T) \cdot \frac{1+\tau^2 \cdot T^{-2}}{1+\tau T^{-1}}. \tag{3.52}$$

By induction on m and using arguments similar to those used to derive (3.46) and (3.47), we obtain

$$\begin{aligned} \mathrm{Cov}(X^{(1)}_{mT}, X^{(2)}_{mT}) &= m\ \mathrm{Cov}(X^{(1)}_{T}, X^{(2)}_{T}), \\ \sigma^2(X^{(i)}_{mT}) &= m\,\sigma^2(X^{(i)}_{T}) \end{aligned}$$

for $i \in \{1,2\}$. Similar arguments used to show (3.50) and (3.51) can also be used to show that

$$\begin{aligned} \mathrm{Cov}(X^{(1)}_{mT+\tau}, X^{(2)}_{mT+\tau}) &= \mathrm{Cov}(X^{(1)}_{T}, X^{(2)}_{T}) \cdot \left(m + \frac{\tau^2}{T^2}\right), \\ \sigma^2(X^{(i)}_{mT+\tau}) &= \sigma^2(X^{(i)}_{T}) \cdot \left(m + \frac{\tau}{T}\right) \end{aligned}$$

for $i \in \{1,2\}$, whence

$$\rho(mT+\tau) = \rho(T) \cdot \frac{m+\tau^2 \cdot T^{-2}}{m+\tau T^{-1}}. \tag{3.53}$$

□

We show below how (3.45) leads to the asymptotic stationarity of the correlation coefficient.

Corollary 3 (Asymptotic Stationarity of Forward Continuation)**.** *The correlation $\rho(mT+\tau)$ achieves asymptotic stationarity as $m \to \infty$. More specifically,*

$$\lim_{m\to\infty} \rho(mT+\tau) = \rho(T), \quad \textit{for any} \quad \tau \in [0,T] \tag{3.54}$$

Proof: From (3.45) we have that

$$\begin{aligned} \rho(mT+\tau) &= \rho(T)\frac{m+\tau^2 \cdot T^{-2}}{m+\tau \cdot T^{-1}} \\ &= \rho(T)\frac{m(1+\tau^2 \cdot T^{-2} \cdot m^{-1})}{m(1+\tau \cdot T^{-1} \cdot m^{-1})} \\ &= \rho(T)\frac{1+\tau^2 \cdot T^{-2} \cdot m^{-1}}{1+\tau \cdot T^{-1} \cdot m^{-1}}. \end{aligned} \tag{3.55}$$

Passing to the limit as $m \to \infty$ in the standard manner, we obtain

$$\lim_{m\to\infty} \rho(mT+\tau) = \rho(T) \quad \text{for all} \quad \tau \in [0,T],$$

as was to be proved. □

Figure 3.2 depicts the correlation structure obtained from simulating a bivariate Poisson process with intensities 3 and 5, calibrated to correlations of 0.7 and -0.9, respectively, within the simulation interval $[0,5]$ using Backward Simulation on the interval $[0,1]$ and Forward Continuation of Backward Simulation on each of the intervals $[m, m+1]$ for $m = 1,2,3,4$. We refer to this as the Forward-Backward Simulation approach. We used 1,000,000 samples to compute the dashed black lines in Figure 3.2. The blue circles depict the theoretical values according to (3.45). Note the good agreement between the theoretical and

the empirical results. Also note that the first few intervals of the process simulated under the Forward-Backward Simulation can be discarded, similar to a burn-in phase, so that the process exhibits a near constant correlation structure.

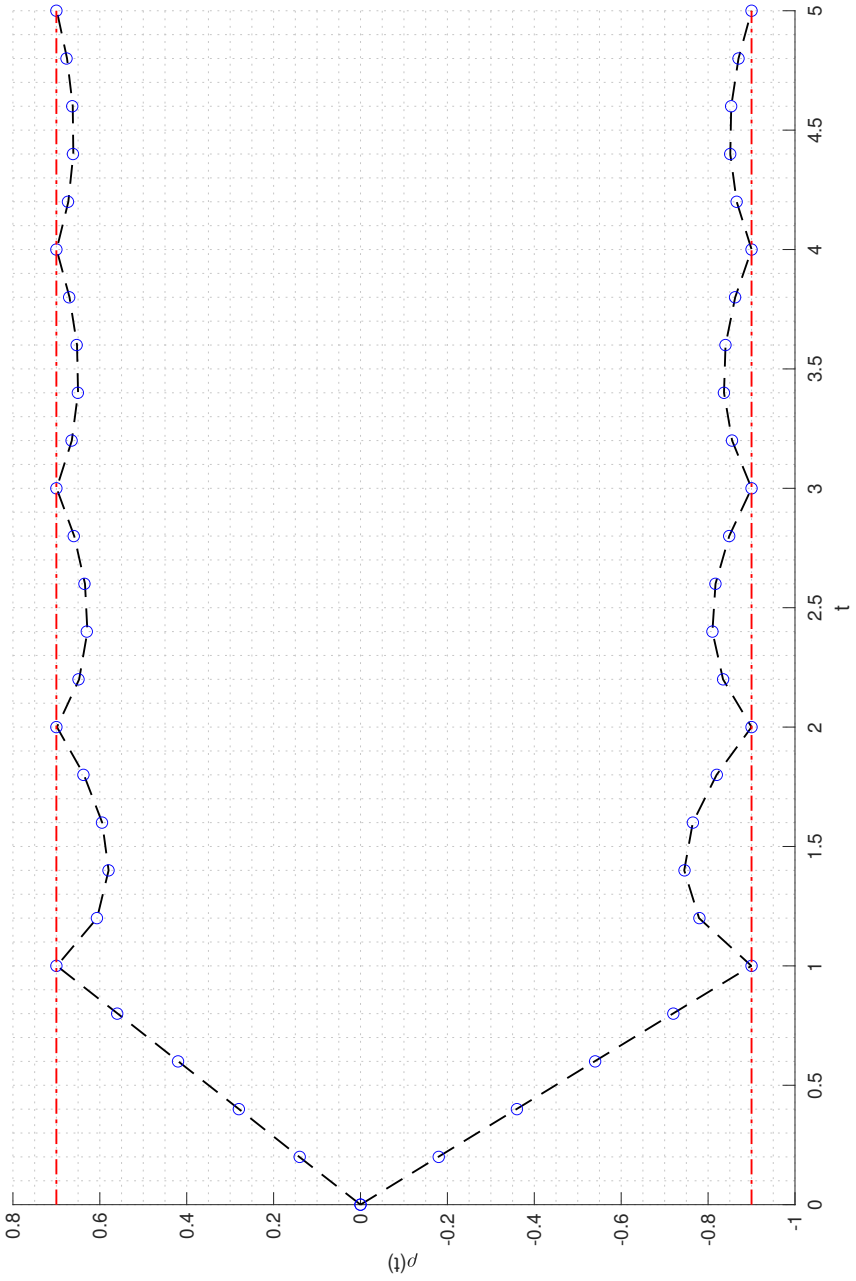

Figure 3.2: The dotted black lines depict the correlation structure for a bivariate Poisson process with intensities 3 and 5, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.9$, respectively, computed by the Forward-Backward Simulation approach. The blue circles depict the theoretical values according to (3.45).

3.3.2 Forward-Backward Simulation Algorithm in d-dimensions

Algorithm 20 Forward-Backward Simulation of correlated multivariate Poisson processes

Require: Vector of marginal Poisson distributions at time T

$\mathbf{Pois(\boldsymbol{\lambda} T)} = (\text{Pois}(\lambda_1 T), \ldots, \text{Pois}(\lambda_d T))$

Correlation matrix C

The number of intervals m (i.e., $[0,T],[T,2T],\cdots,[(m-1)T,mT]$) used in the Forward-Backward Simulation Algorithm

Output: Scenarios of the multivariate Poisson process in the interval $[0, mT]$

1: Construct $\text{Pois}(\lambda_k T)$ distributed marginals $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]$ for $k = 1, \ldots, d$

2: Generate the Poisson calibrated measure $P^{(C)}$ using Algorithm 9, which takes as input $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]_{k=1}^{d}$ and C

3: **for** $i = 1, \ldots, m$ **do**

4: Generate samples $(N^{(1)}, \ldots, N^{(d)}) \sim P^{(C)}$ using Algorithm 18 ▷ *Get the number of events in the interval* $[(i-1)T, iT]$

5: **for** $k = 1, \ldots, d$ **do** ▷ *this can be done in parallel*

6: Generate $N^{(k)}$ iid uniform random variables in the interval $[(i-1)T, iT]$: $\mathbf{T}_i^{(k)} = (T_1^{(k)}, \ldots, T_{N^{(k)}}^{(k)})'$ ▷ *column vector*

7: Sort $\mathbf{T}_i^{(k)}$ in ascending order

8: Append $\mathbf{T}_i^{(k)}$ to $\mathbf{T}^{(k)}$

9: **return** $\mathbf{T} = (\mathbf{T}^{(1)}, \ldots, \mathbf{T}^{(d)})$

Remark 64. *Note that for most applications, especially in risk management, $\rho(nT)$ for $n \in \{1, \ldots, m\}$, where m is the number of forward intervals as defined in Algorithm 20, should vary less than 3%.*

3.4 Forward versus Backward Simulation

We introduced Backward Simulation in Section 3.1, analyzed the correlation structure obtained under Backward Simulation in Section 3.2 and showed that the correlation coefficient between the components of a multivariate Poisson process obtained under Backward Simulation is a linear function of time within the simulation interval $[0, T]$:

$$\rho(t) = \frac{t}{T} \cdot \rho(T)$$

where $\rho(T)$ is any admissible correlation matrix. Recall that an admissible correlation matrix is a correlation matrix that can be the correlation matrix of a discrete multivariate Poisson distribution with the specified intensities; see Section 2.8 for a discussion of admissible correlation matrices. Then, in Section 3.3, we introduced Forward Continuation of Backward Simulation, a method for extending a process constructed using Backward Simulation within $[0, T]$ to an interval $[mT, (m+1)T]$ for any positive integer m. A surprising result is that the correlation structure of a multivariate Poisson process constructed using the Forward-Backward method attains asymptotic stationarity:

$$\lim_{m \to \infty} \rho(mT + \tau) = \rho(T) \quad \text{for any} \quad \tau \in [0, T].$$

In this section, we compare the Forward-Backward method with the well-known Forward Simulation approach and discuss the advantages of the backward approach over the forward approach.

3.4.1 Forward Simulation

We begin by reviewing the forward approach in one dimension. Given a simulation interval $[0, T]$, stochastic processes are typically simulated forwards in time. This is due to the fact that it is both conceptually natural and technically simpler to do so. Forward Simulation (FS) consists of starting at the beginning of the simulation interval, $t = 0$, and advancing the process forward by some time-step, h, by simulating the increments of the process until the end of the simulation interval, $t = T$, is reached. For a univariate Poisson process having arrival moments $T_1, T_2, \ldots$, the inter-arrival times, $\Delta T_k = T_k - T_{k-1}$ for $k = 1, 2, \ldots$, where $T_0 := 0$, are exponentially distributed:

$$\mathbb{P}(\Delta T_k \leq t) = 1 - e^{-\lambda t}. \tag{3.56}$$

Thus, Forward Simulation of the univariate Poisson process

$$N_t = \sum_{i=1}^{\infty} \mathbb{1}(T_i \leq t)$$

within $[0, T]$ consists of repeated sampling of exponentially distributed random variables, ΔT_k, whilst $\sum_k \Delta T_k \leq T$. This remains true for the case of uncorrelated multivariate Poisson processes since the components are mutually independent—Forward Simulation of the multivariate process reduces to the Forward Simulation of each individual component.

However, it is not as straightforward to simulate correlated multivariate Poisson processes. To see why, consider first how a bivariate Poisson process $(N_t^{(1)}, N_t^{(2)})$, where $N_t^{(i)} \sim \text{Pois}(\lambda_i t)$ for $i \in \{1, 2\}$, can be constructed such that it exhibits extreme correlations under Forward Simulation. To accomplish this,

we rely on the well known Fréchet-Hoeffding theorem[5] [49, 64]. When applied to the Poisson setting, the Fréchet-Hoeffding theorem says that the inter-arrival times $\{\Delta T_k^{(i)}\}_{k\geq 1}$ for $i \in \{1,2\}$ must satisfy

$$\lambda_1 \Delta T_k^{(1)} = \lambda_2 \Delta T_k^{(2)}\,, \quad k = 1, 2, \ldots \tag{3.57}$$

in the case of extreme positive correlation and

$$\exp\left(-\lambda_1 \cdot \Delta T_k^{(1)}\right) + \exp\left(-\lambda_2 \cdot \Delta T_k^{(2)}\right) = 1\,, \quad k = 1, 2, \ldots \tag{3.58}$$

in the case of extreme negative correlation. Thus, to forward simulate a bivariate Poisson process with extreme correlations, one must first sample from the inter-arrival distribution (3.56) for one of the components and then solve either (3.57) or (3.58) to obtain the inter-arrival time for the other component.

3.4.2 Range of Correlations Restricted under FS

Extreme correlations in the two-dimensional case obtained by the Backward Simulation approach are more extreme than the extreme correlations attained obtained by Forward Simulation. This can be seen by the Monte Carlo experiment summarized in Figure 9.8 on page 223 of [74], where the extreme correlations generated by Backward Simulation attain values close to 1 and -1, while the extreme correlations generated by Forward Simulation do not attain values close to 1 and -1. Therefore, even if we could find a method based upon Forward Simulation to compute bivariate Poisson processes with correlations between the extreme correlations attainable by Forward Simulation, this method would not be capable of simulating bivariate Poisson processes with as wide a range of correlations as is possible with Backward Simulation coupled with Forward Continuation of Backward Simulation.

Expanding on the experimental result of [74] noted in the paragraph above, we show analytically, below, that the extreme positive correlation for a bivariate Poisson process computed by Forward Simulation is a constant determined by the ratio of the intensities of the processes. Moreover, if one intensity is much larger than the other, this constant extreme positive correlation is much smaller than 1. A similar analysis can be made for the extreme negative case.

Extreme positive correlations via Forward Simulation In the case of extreme positive correlation under Forward Simulation, (3.57) implies that

$$\lambda_1 T_k^{(1)} = \lambda_2 T_k^{(2)}\,, \quad k = 0, 1, 2, \ldots\,. \tag{3.59}$$

Define $\kappa = \lambda_1/\lambda_2$. Obviously, $0 < \kappa < \infty$. To determine the correlation between the process $N_t^{(1)}$associated with the events $T_k^{(1)}$, $k = 1, 2, \ldots$, and the process $N_t^{(2)}$ associated with the events $T_k^{(2)}$, $k = 1, 2, \ldots$, we begin by showing that, for all $t > 0$,

$$N_t^{(1)} = N_{\kappa t}^{(2)}. \tag{3.60}$$

[5] For discrete distributions, Fréchet-Hoeffding is equivalent to the EJD theorem in two-dimensions [74]

To this end, choose any $t > 0$ and let $N_t^{(1)} = n$ where n is a non-negative integer. This implies that the arrival moments associated with $N_t^{(1)}$ satisfies the inequality

$$T_n^{(1)} \leq t < T_{n+1}^{(1)}. \tag{3.61}$$

It follows immediately from (3.59) that the arrival moments for $N_t^{(2)}$ must satisfy

$$T_k^{(2)} = \kappa T_k^{(1)} \quad \text{for all} \quad k = 0, 1, 2, \ldots . \tag{3.62}$$

Multiplying (3.61) through by κ and using (3.62), we see that $T_n^{(2)} \leq \kappa t < T_{n+1}^{(2)}$, which in turns implies that $N_{\kappa t}^{(2)} = n$. Hence, $N_t^{(1)} = n = N_{\kappa t}^{(2)}$. We have shown (3.60) since $t > 0$ is arbitrary.

Now we compute the correlation coefficient of $N_t^{(1)}$ and $N_t^{(2)}$ in the case $\kappa \geq 1$. To this end, write $N_{\kappa t}^{(2)}$ as

$$N_{\kappa t}^{(2)} = N_t^{(2)} + \Delta N_{\kappa t}^{(2)}, \tag{3.63}$$

where $\Delta N_{\kappa t}^{(2)} = N_{\kappa t}^{(2)} - N_t^{(2)}$ represents the increment of $N_t^{(2)}$ in the interval $[t, \kappa t]$ and is independent of $N_t^{(2)}$. Then, we obtain

$$\begin{aligned}
\operatorname{Cov}(N_t^{(1)}, N_t^{(2)}) &= \mathbb{E}\,[N_t^{(1)} N_t^{(2)}] - \mathbb{E}\,[N_t^{(1)}]\,\mathbb{E}\,[N_t^{(2)}] \\
&= \mathbb{E}\,[N_{\kappa t}^{(2)} N_t^{(2)}] - \mathbb{E}\,[N_{\kappa t}^{(2)}]\,\mathbb{E}\,[N_t^{(2)}] \\
&= \mathbb{E}\,[(N_t^{(2)} + \Delta N_{\kappa t}^{(2)}) N_t^{(2)}] - \mathbb{E}\,[N_t^{(2)} + \Delta N_{\kappa t}^{(2)}]\,\mathbb{E}\,[N_t^{(2)}] \\
&= \mathbb{E}\,[N_t^{(2)} N_t^{(2)}] + \mathbb{E}\,[\Delta N_{\kappa t}^{(2)} N_t^{(2)}] - \mathbb{E}\,[N_t^{(2)}]\,\mathbb{E}\,[N_t^{(2)}] - \mathbb{E}\,[\Delta N_{\kappa t}^{(2)}]\,\mathbb{E}\,[N_t^{(2)}] \\
&= \mathbb{E}\,[N_t^{(2)} N_t^{(2)}] + \mathbb{E}\,[\Delta N_{\kappa t}^{(2)}]\,\mathbb{E}\,[N_t^{(2)}] - \mathbb{E}\,[N_t^{(2)}]\,\mathbb{E}\,[N_t^{(2)}] - \mathbb{E}\,[\Delta N_{\kappa t}^{(2)}]\,\mathbb{E}\,[N_t^{(2)}] \\
&= \mathbb{E}\,[(N_t^{(2)})^2] - \mathbb{E}\,[N_t^{(2)}]\,\mathbb{E}\,[N_t^{(2)}] \\
&= \sigma^2(N_t^{(2)}),
\end{aligned}$$

where we used (3.60) in the second line, (3.63) in the third line, and the independence of $\Delta N_{\kappa t}^{(2)}$ and $N_t^{(2)}$ in the fifth line of the set of equations above. Since $\sigma^2(N_t^{(1)}) = \lambda_1 t$, $\sigma^2(N_t^{(2)}) = \lambda_2 t$, and $\kappa = \lambda_1/\lambda_2$, the set of equations above implies that

$$\rho(N_t^{(1)}, N_t^{(2)}) = \frac{1}{\sqrt{\kappa}}, \qquad \text{where } \kappa \geq 1. \tag{3.64}$$

Similar reasoning in the case $0 < \kappa < 1$ leads to

$$\rho(N_t^{(1)}, N_t^{(2)}) = \sqrt{\kappa}. \tag{3.65}$$

Therefore, we can see, from Equations (3.64) and (3.65), that the extreme positive correlation coefficient obtained via Forward Simulation is a function of the ratios of the intensities of the components of a bivariate Poisson process that does not depend on time. For example, if the correlated bivariate Poisson with intensities 3 and 5 from Figure 3.1 was simulated to time $T = 5$ by Forward Simulation, the attainable extreme positive correlation would only be $\sqrt{3/5} \approx 0.7746$, whereas the extreme positive correlation is 0.9955 using Backward Simulation. Finally, consider a more extreme example of a bivariate Poisson process with intensities 1 and 100. The attainable extreme positive correlation would only be 0.1 using Forward Simulation whereas the extreme positive correlation is 0.9193 using Backward Simulation.

3.5 Computational Complexity

The computational complexity of the Backward Simulation algorithms for the more complicated Poisson processes discussed in Chapters 4 and 5 are similar. For this reason, we omit repeating the discussion here in those (more general) settings.

3.5.1 Backwards Simulation

The Backward Simulation method constructs multivariate Poisson sample paths having a dependence structure specified by the joint distribution calibrated from the EJD approach. Each simulation consists of, for each dimension, generation of a number of uniformly distributed random variables equal to the terminal number of events. Let M denote the number of Monte Carlo simulations, T denote the terminal simulation time, and $N_T^{(k)} \sim \text{Pois}(\lambda_k T)$ the number of uniform random variables we have to generate per dimension.

Time Complexity

For a d-dimensional multivariate Poisson process and M Monte Carlo simulations, the pure time complexity of the Backwards Simulation, ignoring the calibration and sampling of the EJD approach, is:

$$O(M \cdot T \cdot \sum_{k=1}^{d} \lambda_k) \tag{3.66}$$

since Backward Simulation *does not* generate a constant-sized vector per dimension. The time complexity of the EJD approach can be found in Section 2.10.

Memory Complexity

Since the Backwards Simulation relies on the joint distribution of the process at terminal time T obtained by calibration via the EJD approach, we refer the memory complexity of the EJD construction in Section 2.10

Remark 65. *The computational complexity per forward interval for the Forward Backward Simulation is the same as that of Backwards Simulation, since each forward interval requires only one additional draw from the calibrated joint distribution and a standard backward simulation step.*

3.6 Summary

In this chapter, we introduced the Backward Simulation method for Poisson processes that, in conjunction with extreme joint distributions constructed using the EJD approach introduced in Chapter 2, enables the simulation of correlated multivariate Poisson processes. In particular, Backward Simulation is able to generate all admissible distributions at the endpoint T of the interval $[0, T]$. Forward Continuation of Backward Simulation enables the process generated by Backward Simulation within the interval $[0, T]$ to be extended to intervals $[mT, (m+1)T]$ for any positive integer m, where the correlation displays the asymptotic behavior

$$\rho(mT + \tau) \approx \rho(T)$$

for all m sufficiently large, all τ in $[0, T]$, and for any admissible correlation matrix $\rho(T)$.

Our contributions are summarized below:

- We introduced the Forward Continuation of Backward Simulation for Poisson processes. This enabled the continuation of correlated multivariate Poisson processes simulated using Backward Simulation within $[0, T]$ to simulation intervals $[mT, (m + 1)T]$ for any non-negative integer m. This was published in [17] for the Poisson case.
- We showed that the correlation structure of a correlated multivariate Poisson process attains asymptotic stationarity.
- We showed formally that the extreme positive correlations attained under Forward Simulation in the bivariate setting is constant and determined by the ratio of the intensities of the bivariate Poisson process.

Chapter 4

Backward Simulation of Mixed Poisson Processes

In Chapter 3, we introduced the Backward Simulation for multivariate Poisson processes that, in conjunction with the Extreme Joint Distribution approach (Chapter 2) for constructing joint distributions with prescribed marginal distributions and a prescribed dependence structure in the form of a correlation matrix, enabled the construction of *correlated* multivariate Poisson processes. In particular, our approach enables the construction of correlated multivariate Poisson processes capable of exhibiting extreme positive and even *extreme negative* correlations between its components. A natural question to ask next is whether the Backward Simulation methodology applies to more general Poisson processes and, if so, to which generalizations does the methodology still apply. One generalization of the Poisson process to consider is the mixed Poisson process, where the intensity of the Poisson process is a random variable instead of a constant scalar value. Consequently, the increments of a mixed Poisson process are no longer independent of each other [58]. However, the increments are *conditionally independent*, a property that we exploit in showing that our Backward Simulation methodology can be applied to mixed Poisson processes. Mixed Poisson processes are widely used in cases where Poisson processes are deemed inadequate; some examples can be found in finance [19, 56], Operational Risk [37], physics [41], and software reliability [62].

Similar to the Poisson setting, we note that while the majority of the exposition in this chapter is in the bivariate setting, the discussions generalize immediately to the multivariate setting.

4.0.1 Outline

In Section 4.1, we review the basics of mixed Poisson processes and mixed Poisson distributions. The Backward Simulation of mixed Poisson processes is discussed in Section 4.2. The correlation structure of mixed Poisson processes simulated using Backward Simulation is discussed in Section 4.3. The Forward Continuation of the Backward Simulation for mixed Poisson processes is discussed in Section 4.4. Section 4.5 closes and summarizes the chapter.

4.0.2 Notation

We make use of the following notation in this chapter.

Symbol	**Definition**
T	End of the simulation interval
T_i	i^{th} arrival moment
$\Delta T_i := T_i - T_{i-1}$	i^{th} inter-arrival time
$T_i^{(k)}$	i^{th} arrival moment corresponding to the k^{th} component
λ	Mean parameter of a Poisson distribution or the intensity parameter of a Poisson process
N_t	Poisson process
$\mathbf{N}_t = (N_t^{(1)}, \ldots, N_t^{(d)})$	d-dimensional Poisson process
$\hat{N}_t$	Mixed Poisson process
$\hat{\mathbf{N}}_t = (\hat{N}_t^{(1)}, \ldots, \hat{N}_t^{(d)})$	d-dimensional mixed Poisson process
NB_t	Negative Binomial process
$\mathbf{NB}_t = (\text{NB}_t^{(1)}, \ldots, \text{NB}_t^{(d)})$	d-dimensional Negative Binomial process
n	Realization of a Poisson, mixed Poisson, or Negative Binomial random variable
U	Structure distribution of a mixed Poisson process
Λ	U-distributed non-negative random variable
$\bar{\lambda}$	Mean of the intensity parameter of a mixed Poisson process
$\text{MP}(U)$	Mixed Poisson distribution with structure distribution U
$\text{MP}(t, U)$	Mixed Poisson distribution with structure distribution U at time t
$\text{MPP}(U)$	Mixed Poisson processes with structure distribution U
ζ, ξ	General integer valued random variables
X_t	Mixed Poisson process obtained from Backward Simulation
$\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(d)})$	Multivariate mixed Poisson process obtained from Backward Simulation
$\Delta_s X_t$	The increment $X_{t+s} - X_t$ of the process X_t in interval $[t, t+s]$
$\hat{g}(t, z)$	Generating function of a mixed Poisson process
$\hat{g}_{NB}(t, z)$	Generating function of a Negative Binomial process

4.1 Mixed Poisson Process

We briefly review some properties of the mixed Poisson process. The main results of the theory of mixed Poisson processes can be found in [58]. Recent results on the characterization of multivariate mixed Poisson processes are in [130]. First consider a Poisson process

$$N_t = \sum_{i=1}^{\infty} \mathbb{1}\left(T_i \leq t\right) \tag{4.1}$$

where the arrival moments of the process, $\{T_i\}_{i=1}^{\infty}$, are iid. The inter-arrival times, $\Delta T_i := T_i - T_{i-1}$ for $i = 1, 2, \ldots$ and $T_0 = 0$, form a sequence of independent identically distributed random variables having an exponential distribution with parameter λ. It is well known that the number of events, N_t, in the interval $[0, t]$ has the Poisson distribution with parameter λt:

$$\mathbb{P}(N_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \ldots \quad t > 0. \tag{4.2}$$

A natural generalization of the Poisson distribution is to randomize the intensity parameter λ in (4.2), leading to the mixed Poisson Distribution (MPD).

Definition 29 (Mixed Poisson Distribution [58])**.** *A discrete random variable N is said to be mixed Poisson distributed, MP(U), with structure distribution U, if*

$$\begin{aligned} p_n :=\ & \mathbb{P}(N = n) = \mathbb{E}\left[\frac{(\Lambda)^n}{n!}e^{-\Lambda}\right] \\ =\ & \int_{0-}^{\infty} \frac{(\lambda)^n}{n!}\, e^{-\lambda} dU(\lambda), \quad n = 0, 1, 2, \ldots \end{aligned} \tag{4.3}$$

where Λ is a random variable distributed according to U.

Remark 66. *The structure distribution U can be viewed as a prior distribution, which allows us to view (4.2) as a conditional distribution, given a realization of the intensity parameter $\Lambda = \lambda$ and (4.3) as an unconditional distribution.*

Remark 67. *Another interpretation of (4.3) is that it is a mixture of Poisson distributions.*

Definition 30 (Mixed Poisson Process [58])**.** *$\hat{N}_t$ is a mixed Poisson process if it is $\mathrm{MP}(t, U)$-distributed for all $t \geq 0$. That is,*

$$\mathbb{P}(\hat{N}_t = n) = \mathbb{E}\left[\frac{(\Lambda t)^n}{n!}e^{-\Lambda t}\right] = \int_{0-}^{\infty} \frac{(\lambda t)^n}{n!}\, e^{-\lambda t} dU(\lambda), \quad n = 0, 1, 2, \ldots . \tag{4.4}$$

The mixed Poisson process is a Poisson process with a non-negative random intensity.

Remark 68. *Note that we use $\mathrm{MP}(t, U)$, instead of $\mathrm{MP}(U)$, when we want to make explicit the dependence on t in an expression such as (4.4).*

Lundberg also showed that there exists a mixed Poisson process for each structure distribution U and that the process is uniquely defined [84]. In what follows, we denote by $\mathrm{MPP}(U)$ the class of mixed Poisson processes with structure distribution U. It is not difficult to see that if $\hat{N}_t \in \mathrm{MPP}(U)$, then the generating function takes the following form.

Definition 31 (Generating function of a mixed Poisson process [58])**.**

$$\hat{g}(t;z) := \mathbb{E}\left[z^{\hat{N}_t}\right] = \int_0^\infty e^{xt(z-1)}\,\mathrm{d}U(x). \tag{4.5}$$

The moments take the following form.

Definition 32 (Moments of the mixed Poisson process [58])**.**

$$\mathbb{E}\left[\hat{N}_t\right] = \bar{\lambda}t, \quad \sigma^2(\hat{N}_t) = \bar{\lambda}t + \sigma^2(\Lambda)t^2, \tag{4.6}$$

where

$$\bar{\lambda} = \mathbb{E}\left[\Lambda\right] = \int_0^\infty \lambda\,\mathrm{d}U(\lambda), \quad \sigma^2(\Lambda) = \int_0^\infty (\lambda-\bar{\lambda})^2\,\mathrm{d}U(\lambda). \tag{4.7}$$

Remark 69. *When we construct multivariate mixed Poisson distributions using the EJD approach, there is no direct dependence between the structure variables. The structure variables indirectly affect the correlation structures since different realizations (draws) of structure variables results in different marginal distributions altogether.*

Crucially, the mixed Poisson process also posses the conditional uniformity property.

Proposition 4 (Conditional uniformity of the mixed Poisson process [89])**.** *Let $\hat{N}_t$ be a mixed Poisson process defined on the interval $[0,T]$ and suppose that the total number of events n is known at time T:*

$$\hat{N}_t = \sum_{i=1}^{n} \mathbb{1}\left(T_i \le t\right) \quad \textit{for} \;\; t \in [0,T].$$

Then, the arrival moments $\{T_i\}_{i=1}^n$ are independent, identically distributed random variables having a uniform conditional distribution

$$\mathbb{P}(T_i \le t \,|\, n) = \frac{t}{T} \quad i = 1,2,\ldots,n \quad \textit{and} \quad 0 \le t \le T. \tag{4.8}$$

4.2 Backward Simulation

In this section, we prove the fundamental result enabling the Backward Simulation for mixed Poisson processes. That is, we consider, for the remainder of this section, a process, X_t for $0 \leq t \leq T$, defined as

$$X_t = \sum_{i=1}^{n} \mathbb{1}(T_i \leq t), \tag{4.9}$$

where n is a realization of the random variable $X_T \sim \text{MP}(T, U)$ and the random variables $\{T_i\}_{i=1}^n$ are iid having the uniform conditional distribution (4.8) in the interval $[0, T]$. We remind the reader that although Theorem 9 below is formulated and proved in the univariate setting, the results extend directly to the multivariate setting. That is, similar to Theorem 6 in Chapter 3, Theorem 9 extends directly to the case where X_t is a correlated multivariate mixed Poisson process. This is because of the fact that, while the dependence structure is specified through the joint distribution, X_T, every coordinate (marginal) of a correlated multivariate mixed Poisson process is itself a (univariate) mixed Poisson process[1]. Therefore, given the joint number of events sampled from a suitable joint distribution[2], X_T, the correlated multivariate mixed Poisson process can be constructed within the simulation interval $[0, T]$ by applying Theorem 9 independently to each coordinate.

We begin by reviewing some notation to keep this chapter as self contained as possible. For a general d-dimensional vector[3], $\mathbf{k} = (k_1, k_2, \ldots, k_d) \in \mathbb{N}_0^d$, with non-negative integer coordinates, $k_j \geq 0$, we denote the norm of the vector by

$$\|\mathbf{k}\| = \sum_{j=1}^{d} k_j.$$

For any d-dimensional vector, $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, with non-negative real coordinates, and $\mathbf{k} \in \mathbb{N}_0^d$, we denote

$$\mathbf{x}^{\mathbf{k}} := \prod_{j=1}^{d} {x_j}^{k_j}$$

and introduce the multinomial coefficient

$$\binom{\mathbf{k}+l}{\mathbf{k}} := \frac{\left(l + \sum_{i=1}^{d} k_i\right)!}{l! \cdot \prod_{i=1}^{d} k_i!}.$$

The main result of the chapter is the following theorem.

Theorem 9. *For $t \in [0, T]$, define the process X_t by*

$$X_t = \sum_{i=1}^{n} \mathbb{1}(T_i \leq t) \tag{4.10}$$

where n is a realization of the random variable $X_T \sim \text{MP}(T, U)$ and $\{T_i\}_{i=1}^n$ are independent, identically

[1] See Definition 9.

[2] Recall that Chapter 2 is concerned with constructing extreme distributions that exhibit extreme correlations and constructing distributions that exhibit any *admissible* correlation that is a convex combination of extreme correlations.

[3] We emphasize that the d here refers to the dimension of a generic vector and does not refer to the dimensionality of multivariate Poisson processes.

distributed random variables having a uniform conditional distribution (4.8). *Then,* X_t *is a mixed Poisson process having distribution* $\mathrm{MP}(t, U)$ *in the interval* $[0, T]$.

Proof: We prove the following two statements.

1. For any interval $(t, t+s] \subset [0, T]$ of length $s \geq 0$, the increments $\Delta_s X_t = X_{t+s} - X_t$ of the process X_t are mixed Poisson distributed, with $\Delta_s X_t = X_{t+s} - X_t \sim \mathrm{MP}(s, U)$. Moreover, the distribution of the increments $\Delta_s X_t$ does not depend on t.

2. For any $l = 2, 3, \ldots$ disjoint sub-intervals $(t_i, t_i + s_i] \subset [0, T]$ for $i = 1, 2, \ldots, l$, the random variables $\Delta_{s_i} X_{t_i} = X_{t_i+s_i} - X_{t_i}$, for $i = 1, 2, \ldots, l$, are mutually independent.

Similar to the proof for the Poisson case in Chapter 3, the theorem follows immediately from applying Lemma 25 and Lemma 26. We begin by proving the first statement above. Note that, since the mixed Poisson process also possesses the conditional uniformity property (see Proposition 4), the proof of the first statement is very similar to the Poisson case. Nevertheless, we present the full details of the proof, since this is the main result of this chapter and we want to keep the chapters as self-contained as possible.

To this end, denote by $\Delta_s X_t = X_{t+s} - X_t$, the number of events occurring in the interval $[t, t+s]$. Then, the probability that k events occur in the interval $[t, t+s]$ of the process can be expressed, by the law of total probability, in terms of the conditional probability, as

$$\mathbb{P}(\Delta_s X_t = k) = \sum_{m=0}^{\infty} \mathbb{P}(\Delta_s X_t = k \,|\, X_T = k + m) \cdot \mathbb{P}(X_T = k + m). \tag{4.11}$$

Since the conditional probability of k events occurring in the sub-interval $[t, t+s] \subset [0, T]$ given a total of $X_T = n = k + m$ independent, uniformly distributed events occurring in the full interval $[0, T]$ is equal to choosing k out of $n = k + m$ total events in the interval $[t, t+s]$ of length s, with the rest of the events being in the remainder of the interval $[0, T]$, it follows that

$$\mathbb{P}(\Delta_s X_t = k \,|\, X_T = k + m) = \binom{k+m}{k} \left(\frac{s}{T}\right)^k \left(1 - \frac{s}{T}\right)^m, \quad m = 0, 1, \ldots . \tag{4.12}$$

Moreover, since X_T is $\mathrm{MP}(T, U)$ distributed, its generating function takes the form

$$\hat{p}(z) := \mathbb{E}\left[z^{X_T}\right] = \int_0^{\infty} e^{\lambda T(z-1)} \, \mathrm{d}U(\lambda). \tag{4.13}$$

We use Lemma 25 to show that the generating function of $\Delta_s X_t$ also takes the form (4.13), but with T replaced by s. Thus, $\Delta_s X_t$ is $\mathrm{MP}(s, U)$ distributed. To this end, let $p_k = \mathbb{P}(X_T = k)$, for any $k \in \mathbb{N}_0$, and $x = s/T$. Then, note that, by (4.11) and (4.12),

$$\begin{aligned}
\mathbb{P}(\Delta_s X_t = k) &= \sum_{m=0}^{\infty} \mathbb{P}(\Delta_s X_t = k \,|\, X_T = k + m) \cdot \mathbb{P}(X_T = k + m) \\
&= \sum_{m=0}^{\infty} \binom{k+m}{k} \left(\frac{s}{T}\right)^k \left(1 - \frac{s}{T}\right)^m \cdot \mathbb{P}(X_T = k + m) \\
&= \sum_{m=0}^{\infty} p_{k+m} \binom{k+m}{k} x^k (1-x)^m.
\end{aligned}$$

Observe that the last line in the set of equations above has the form of $q_k(x)$ in Lemma 25. So, let

$$q_k(x) = \sum_{m=0}^{\infty} p_{k+m} \binom{k+m}{k} x^k (1-x)^m = \mathbb{P}(\Delta_s X_t = k), \qquad k = 0, 1, 2, \ldots .$$

Since $q_k(x) = \mathbb{P}(\Delta_s X_t = k)$, for $k = 0, 1, 2, \ldots$, the generating function $\mathbb{E}\left[z^{\Delta_s X_t}\right]$ of the random variable $\Delta_s X_t$ must be $\hat{q}(z; x)$, the generating function associated with $q_k(x)$, for $k = 0, 1, 2, \ldots$. From this observation and Lemma 25, it follows that

$$\begin{aligned}
\mathbb{E}\left[z^{\Delta_s X_t}\right] &= \hat{q}(z; x) \\
&= \hat{p}(1 - x + xz) \\
&= \hat{p}(1 - \frac{s}{T} + \frac{s}{T} z) \\
&= \hat{p}(1 + \frac{s}{T}(z-1)) \\
&= \int_0^{\infty} \exp(\lambda T((1 + \frac{s}{T}(z-1)) - 1))\, \mathrm{d}U(\lambda) \\
&= \int_0^{\infty} \exp(\lambda s(z-1))\, \mathrm{d}U(\lambda)
\end{aligned}$$

where, in the fourth line above, we substituted (4.13). The set of equations above shows that the generating function of $\Delta_s X_t$ is $\int_0^{\infty} \exp(\lambda s(z-1))\, \mathrm{d}U(\lambda)$, which is the generating function of a mixed Poisson random variable distributed according to $\mathrm{MP}(s, U)$. Therefore, the increments $\Delta_s X_t = X_{t+s} - X_t$ of the process X_t are mixed Poisson distributed with $\Delta_s X_t \sim \mathrm{MP}(s, U)$. Moreover, the distribution of the increments $\Delta_s X_t$ does not depend on t.

Next, we prove the second statement listed at the start of the proof. To this end, for any $l = 2, 3, \ldots$, consider any l disjoint sub-intervals $(t_i, t_i + s_i]$, for $i = 1, 2, \ldots, l$, where each $(t_i, t_i + s_i] \subset [0, T]$. For $i = 1, 2, \ldots, l$, denote by $\Delta_{s_i} X_{t_i} = X_{t_i + s_i} - X_{t_i}$ the number of events occurring within the sub-interval $(t_i, t_i + s_i]$. For $i = 1, 2, \ldots, l$, let $x_i = (t_i + s_i - t_i)/T = s_i/T \in [0, 1]$. Also, let $\mathbf{x} = (x_1, \ldots, x_l)$ and $y = 1 - \sum_{i=1}^{l} x_i$. Note that $y \in [0, 1]$. Since the intervals $(t_i, t_i + s_i]$, for $i = 1, 2, \ldots, l$, are disjoint and the events have the uniform conditional distribution (4.8), it follows that

$$\mathbb{P}\left(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l \,|\, X_T = m + \sum_{i=1}^{l} k_i\right) = \binom{\mathbf{k}+m}{\mathbf{k}} \cdot \mathbf{x}^{\mathbf{k}} \cdot y^m. \tag{4.14}$$

For any $k \in \mathbb{N}_0$, let $p_k = \mathbb{P}(X_T = k)$. Then, using (4.14) and the law of total probability, we see that

$$\begin{aligned}
&\mathbb{P}\big(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l\big) \\
&= \sum_{m=0}^{\infty} \mathbb{P}\left(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l \,|\, X_T = m + \sum_{i=1}^{l} k_i\right) \cdot \mathbb{P}\left(X_T = m + \sum_{i=1}^{l} k_i\right) \\
&= \sum_{m=0}^{\infty} \binom{\mathbf{k}+m}{\mathbf{k}} \cdot \mathbf{x}^{\mathbf{k}} \cdot y^m \cdot \mathbb{P}(X_T = \|\mathbf{k}\| + m) \\
&= \sum_{m=0}^{\infty} p_{\|\mathbf{k}\|+m} \binom{\mathbf{k}+m}{\mathbf{k}} \cdot \mathbf{x}^{\mathbf{k}} \cdot y^m.
\end{aligned}$$

Observe that the last line in the set of equations above has the form of $\pi(\mathbf{k}; \mathbf{x})$ in Lemma 26. So, for all

$\mathbf{k} = (k_1, \ldots, k_l) \in \mathbb{N}_0^l$, let

$$\pi(\mathbf{k}; \mathbf{x}) = \sum_{m=0}^{\infty} p_{\|\mathbf{k}\|+m} \binom{\mathbf{k}+m}{\mathbf{k}} \cdot \mathbf{x}^{\mathbf{k}} \cdot y^m = \mathbb{P}\big(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l\big).$$

Since, for all $\mathbf{k} = (k_1, \ldots, k_l) \in \mathbb{N}_0^l$, $\pi(\mathbf{k}; \mathbf{x}) = \mathbb{P}\big(\Delta_{s_1} X_{t_1} = k_1, \ldots, \Delta_{s_l} X_{t_l} = k_l\big)$, the generating function $\mathbb{E}\left[z_1^{\Delta_{s_1} X_{t_1}} \cdots z_l^{\Delta_{s_l} X_{t_l}}\right]$ of the joint probability distribution of the increments $\Delta_{s_1} X_{t_1}, \ldots, \Delta_{s_l} X_{t_l}$ must be $\hat{\pi}(\mathbf{z}; \mathbf{x})$, the generating function associated with the probabilities $\pi(\mathbf{k}; \mathbf{x})$, $\mathbf{k} \in \mathbb{N}_0^l$. From this observation and Lemma 26 with $\hat{p}(z) = \mathbb{E}\left[z^{X_T}\right] = \int_0^\infty \exp(\lambda T(z-1))\, \mathrm{d}U(\lambda)$, it follows that

$$\begin{aligned}
\mathbb{E}\left[z_1^{\Delta_{s_1} X_{t_1}} \cdots z_l^{\Delta_{s_l} X_{t_l}}\right] &= \hat{\pi}(\mathbf{z}; \mathbf{x}) \\
&= \hat{p}\left(1 - \sum_{i=1}^{l} x_i(1 - z_i)\right) \\
&= \int_0^\infty \exp\left(\lambda T\big(1 - \sum_{i=1}^{l} x_i(1 - z_i) - 1\big)\right) \mathrm{d}U(\lambda) \\
&= \int_0^\infty \exp\left(\sum_{i=1}^{l} \lambda T x_i(z_i - 1)\right) \mathrm{d}U(\lambda) \\
&= \int_0^\infty \prod_{i=1}^{l} \exp(\lambda T x_i(z_i - 1))\, \mathrm{d}U(\lambda) \\
&= \prod_{i=1}^{l} \int_0^\infty \exp(\lambda s_i(z_i - 1))\, \mathrm{d}U(\lambda) \\
&= \prod_{i=1}^{l} \mathbb{E}\left[z_i^{\Delta_{s_i} X_{t_i}}\right].
\end{aligned}$$

where the last line above follows from our proof of the first point of the Theorem, which implies that

$$\mathbb{E}\left[z_i^{\Delta_{s_i} X_{t_i}}\right] = \int_0^\infty \exp(\lambda s_i(z_i - 1))\, \mathrm{d}U(\lambda) \qquad \text{for } i = 1, 2, \ldots, l.$$

We see from the set of equations above that the generating function of the joint distribution of the increments $\Delta_{s_1} X_{t_1}, \ldots, \Delta_{s_l} X_{t_l}$ factors multiplicatively into the product of the generating functions of the individual increments $\Delta_{s_i} X_{t_i}$, $i = 1, \ldots, l$. Therefore, the increments $\Delta_{s_i} X_{t_i}$, $i = 1, \ldots, l$, are mutually independent. That is, the process X_t for $0 < t \leq T$ has conditionally independent increments. □

4.2.1 Backward Simulation Algorithm for mixed Poisson processes in d-dimensions

Algorithm 21 Backward Simulation of correlated multivariate mixed Poisson processes

Require: Vector of mixed Poisson distributions at terminal time T
$\mathbf{MP}(T, \mathbf{U}) = (\mathrm{MP}(T, U^{(1)}), \ldots, \mathrm{MP}(T, U^{(d)}))$
Correlation matrix C

Output: Scenarios of the multivariate mixed Poisson process in $[0, T]$

1: Construct $\mathrm{MP}(T, U^{(k)})$ distributed marginals $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]$ for $k = 1, \ldots, d$
2: Generate the calibrated mixed Poisson measure $P^{(C)}$ using Algorithm 9, which takes as input $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]_{k=1}^{d}$ and C
3: Generate samples $(N^{(1)}, \ldots, N^{(d)}) \sim P^{(C)}$ using Algorithm 18 ▷ *Get the number of events at terminal time T*
4: **for** $k = 1, \ldots, d$ **do** ▷ *this can be done in parallel*
5: Generate $N^{(k)}$ iid uniform random variables in the interval $[0, T]$: $\mathbf{T}^{(k)} = (T_1^{(k)}, \ldots, T_{N^{(k)}}^{(k)})'$ ▷ *column vector*
6: Sort $\mathbf{T}^{(k)}$ in ascending order
7: **return** $\mathbf{T} = (\mathbf{T}^{(1)}, \ldots, \mathbf{T}^{(d)})$

4.3 Correlation Structure

The main result of this section (Theorem 10) describes, analytically, the behavior of the correlation coefficient between a pair of mixed Poisson processes generated using Backward Simulation within the simulation interval $[0, T]$. The exposition in this section is in the bivariate setting since Pearson correlation is inherently a bivariate concept. The results in this section extend directly to the multivariate setting by application of the results to each pair of components of a multivariate mixed Poisson process.

Theorem 10. *Consider a bivariate process $(X_t^{(1)}, X_t^{(2)})$ generated using Backward Simulation in the interval $[0, T]$, whence $X_t^{(1)}$ and $X_t^{(2)}$ possess the conditional uniformity property (4.8). Let the correlation coefficient at time T, $\rho(T) := \operatorname{corr}(X_T^{(1)}, X_T^{(2)})$, be known. Then $\rho(t) = \operatorname{corr}(X_t^{(1)}, X_t^{(2)})$ takes the form*

$$\rho(t) = \rho(T) \cdot \frac{Z(T)}{Z(t)}, \quad 0 < t \leq T, \tag{4.15}$$

where

$$Z(t) = \frac{\sigma(X_t^{(1)})\,\sigma(X_t^{(2)})}{t^2}, \quad t > 0.$$

and $\sigma^2(X_t^{(i)})$ denotes the variance of $X_t^{(i)}$ for $i \in \{1, 2\}$.

Proof: Lemma 27 from Chapter 3 does not apply to the mixed Poisson setting since the mean and the variance of the process are no longer equal. However, the derivation of (3.39) in Lemma 27 does not rely on the assumption that the mean equals the variance and thus applies to the mixed Poisson setting. Recall from (3.39) that

$$\operatorname{Cov}(\xi_1, \xi_2) = xy \operatorname{Cov}(\zeta_1, \zeta_2),$$

where ξ and ζ are bivariate integer-valued random variables. By substituting $\zeta = (X_T^{(1)}, X_T^{(2)})$, $\xi = (X_t^{(1)}, X_t^{(2)})$, and $x = y = t/T$ we obtain

$$\operatorname{Cov}(X_t^{(1)}, X_t^{(2)}) = \frac{t^2}{T^2} \operatorname{Cov}(X_T^{(1)}, X_T^{(2)}).$$

Dividing both sides of the equation above by $\sigma(X_t^{(1)})\,\sigma(X_t^{(2)})$, we arrive at

$$\begin{aligned}
\rho(t) &= \frac{\operatorname{Cov}(X_t^{(1)}, X_t^{(2)})}{\sigma(X_t^{(1)})\sigma(X_t^{(2)})} \\
&= \frac{t^2}{T^2} \cdot \frac{\operatorname{Cov}(X_T^{(1)}, X_T^{(2)})}{\sigma(X_t^{(1)})\sigma(X_t^{(2)})} \\
&= \frac{t^2}{T^2} \cdot \frac{\operatorname{Cov}(X_T^{(1)}, X_T^{(2)})}{\sigma(X_t^{(1)})\sigma(X_t^{(2)})} \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sigma(X_T^{(1)})\sigma(X_T^{(2)})} \\
&= \rho(T) \cdot \frac{t^2}{T^2} \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sigma(X_t^{(1)})\sigma(X_t^{(2)})} \\
&= \rho(T) \cdot \frac{Z(T)}{Z(t)},
\end{aligned} \tag{4.16}$$

as was to be proved. □

Remark 70. *Crucially, the correlation at terminal time is also preserved in the mixed Poisson case. This can be seen by substituting T for t in (4.15), whereby the second term on the right side of the equation cancels out.*

Remark 71. *In the Poisson case, the auxiliary function $Z(T)/Z(t)$ in (4.15) reduces to tT^{-1}, resulting in a correlation structure that is linear in time within the simulation interval $[0, T]$. This is not true in general for mixed Poisson processes. Consider, for example, the Negative Binomial processes, where the auxiliary functions take the form*

$$\begin{aligned}\rho(t) &= \rho(T) \cdot \frac{Z(T)}{Z(t)} \\ &= \rho(T) \cdot \frac{t}{T} \cdot \sqrt{\frac{(\bar{\lambda}^{(1)} + \sigma^2(\lambda_1)T)(\bar{\lambda}^{(2)} + \sigma^2(\lambda_2)T)}{(\bar{\lambda}^{(1)} + \sigma^2(\lambda_1)t)(\bar{\lambda}^{(2)} + \sigma^2(\lambda_2)t)}}. \end{aligned} \tag{4.17}$$

The difference in the correlation structure between the Poisson and Negative Binomial case can be seen in Figure 4.1. The square root term in (4.17) causes the correlation function in the Negative Binomial case to display a convex behavior in comparison to the linear behavior displayed in the Poisson case.

The Negative Binomial (NB) process is a concrete instance of a mixed Poisson process where the structure distribution, U, is gamma distributed [58]. Unlike the Poisson process, the Negative Binomial process does not have the restriction that its mean equals its variance and, for this reason, is widely used to model count data that exhibits overdispersion.

Figure 4.1 depicts the correlation structure obtained from simulating a bivariate Negative Binomial process where the mean of the intensities are 3 and 5 and the variance of the intensities are 5 and 7, respectively, calibrated to correlations of 0.7 and -0.9, respectively, at terminal time $T = 5$, within the simulation interval $[0, 5]$ under Backward Simulation. We used Backward Simulation with 1,000,000 Monte Carlo samples to compute the dashed black lines in Figure 4.1. The blue circles depict the theoretical values according to (4.17). Note the good agreement between the theoretical and the empirical results. Also note that the extreme positive and extreme negative correlations in this case are 0.9972 and -0.96, respectively, and that any correlation, at the terminal time $T = 5$, within the interval $[0.9972, -0.96]$ is attainable under our Backward Simulation approach. The red lines depict the Poisson case from Figure 3.1 for comparison.

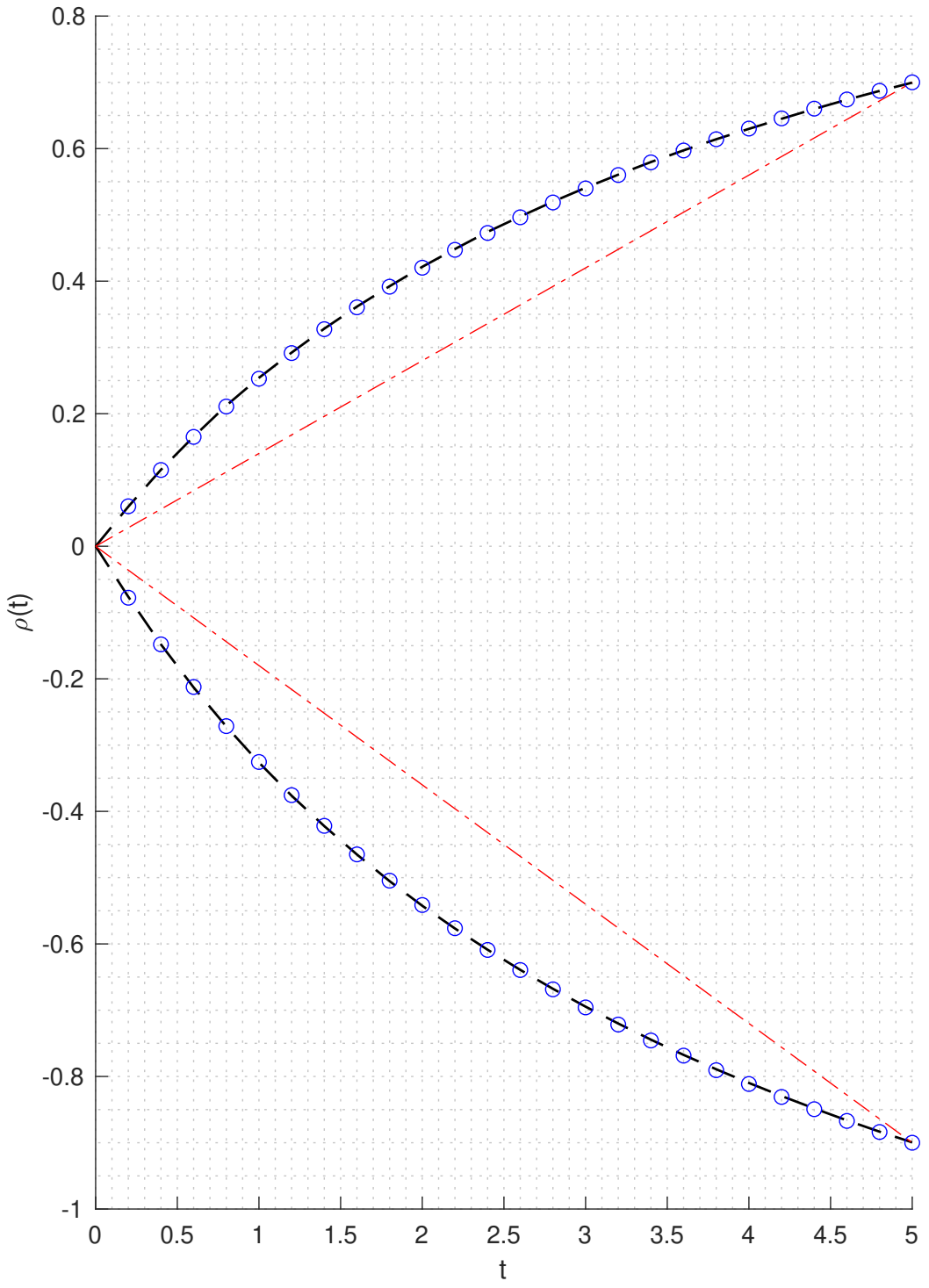

Figure 4.1: The dashed black lines depict the correlation structure for two bivariate Negative Binomial process, with means 3 and 5 and variances 5 and 7, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.9$, respectively, computed by Backward Simulation. The blue circles depict the theoretical values according to (4.17). The dashed red lines depict the Poisson case: two bivariate Poisson processes with intensities 3 and 5, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.9$, respectively, computed by Backward Simulation.

4.4 Forward Continuation of the Backward Simulation for Mixed Poisson Processes

The conditional independence of the increments of a mixed Poisson process, as noted in Section 3.3, presents a challenge in continuing forward a mixed Poisson process already simulated within an interval $[0, T]$. Our particular approach, however, to Forward Continuation of a Poisson process extends naturally to the mixed Poisson process setting. To be more specific, consider a bivariate mixed Poisson process $(X_t^{(1)}, X_t^{(2)})$ already simulated within the interval $[0, T]$ and suppose that we wish to continue forward the process to the subsequent interval $[T, 2T]$. First, we draw an independent sample from the joint distribution at time T:

$$(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)}) \stackrel{d}{=} (X_T^{(1)}, X_T^{(2)}),$$

where $\Delta_T X_T^{(i)} = X_{2T}^{(i)} - X_T^{(i)}$, for $i \in \{1, 2\}$. Note that the bivariate random variable $(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$ has the same prescribed mixed Poisson marginals and prescribed correlation, C, between its components as $(X_T^{(1)}, X_T^{(2)})$. Then, we set

$$(X_{2T}^{(1)}, X_{2T}^{(2)}) = (X_T^{(1)}, X_T^{(2)}) + (\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$$

and, as in Backward Simulation, we use $(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$ and the conditional independence property (similar to (4.8)) to generate the uniform arrival moments $\{T_i^{(1)}\}_{i=1}^{\Delta_T X_T^{(1)}}$ and $\{T_i^{(2)}\}_{i=1}^{\Delta_T X_T^{(2)}}$ in the interval $[T, 2T]$.

Note that the Forward Continuation method remains applicable in the mixed Poisson setting because a mixed Poisson process is defined via a single structural variable Λ that is shared across all times. Conditional on Λ, the process is a time-homogeneous Poisson process with independent and stationary increments.

Forward Continuation, described above, can be repeatedly applied to extend a bivariate mixed Poisson process simulated by Backwards Simulation on an interval $[0, T]$ to any subsequent interval $[mT, (m + 1)T]$ for any non-negative integer m. Algorithm 22 in Section 4.4.2 below combines Backward Simulation and Forward Continuation of Backward Simulation to generate arrival moments for a d-dimensional mixed Poisson process on an interval $[0, mT]$, for any positive integer m. We call this algorithm Forward-Backward Simulation because it starts with one step of Backward Simulation followed by $m - 1$ steps of Forward Continuation of Backward Simulation.

Finally, we note again that, although the exposition in this section is in the bivariate setting, the results extend straightforwardly to the general d-dimensional setting.

4.4.1 Forward Correlation Structure

We analyze the behavior of the correlation coefficient under the Forward Continuation of Backward Simulation by deriving an expression for $\rho(mT + \tau) = \text{corr}(X_{mT+\tau}^{(1)}, X_{mT+\tau}^{(2)})$ as a function of $\rho(T) = \text{corr}(X_T^{(1)}, X_T^{(2)})$ for any non-negative integer m and any $\tau \in [0, T]$ in Theorem 11 below. Similar to the Poisson case, $\rho(mT + \tau)$ attains asymptotic stationarity follows by Corollary 4 below.

Theorem 11. *The correlation coefficient* $\rho(mT + \tau) = \text{corr}(X_{mT+\tau}^{(1)}, X_{mT+\tau}^{(2)})$ *for any non-negative*

integer m and any $\tau \in [0,T]$ as a function of $\rho(T) = \mathrm{corr}(X_T^{(1)}, X_T^{(2)})$ takes the form

$$\rho(mT+\tau) = \rho(T) \cdot (m + \frac{\tau^2}{T^2}) \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sqrt{m\,\sigma^2(X_T^{(1)}) + \sigma^2(X_\tau^{(1)})}\sqrt{m\,\sigma^2(X_T^{(2)}) + \sigma^2(X_\tau^{(2)})}}. \tag{4.18}$$

PROOF: First note that, for $m = 0$, (4.18) follows from Theorem 11. To show that (4.18) also holds for all $m \geq 1$, we begin by deriving an expression for $\rho(T+\tau)$. To this end, let $\Delta_\tau X_T^{(1)} = X_{T+\tau}^{(1)} - X_T^{(1)}$ and $\Delta_\tau X_T^{(2)} = X_{T+\tau}^{(2)} - X_T^{(2)}$ and note that each of the increments $\Delta_\tau X_T^{(1)}$ and $\Delta_\tau X_T^{(2)}$ is independent of both $X_T^{(1)}$ and $X_T^{(2)}$. Therefore, the covariance of the bivariate mixed Poisson process satisfies

$$\begin{aligned} \mathrm{Cov}(X_{T+\tau}^{(1)}, X_{T+\tau}^{(2)}) &= \mathrm{Cov}(X_T^{(1)} + \Delta_\tau X_T^{(1)}, X_T^{(2)} + \Delta_\tau X_T^{(2)}) \\ &= \mathrm{Cov}(X_T^{(1)}, X_T^{(2)}) + \mathrm{Cov}(\Delta_\tau X_T^{(1)}, \Delta_\tau X_T^{(2)}) \end{aligned} \tag{4.19}$$

and, for each $i \in \{1,2\}$, the variance satisfies

$$\begin{aligned} \sigma^2(X_{T+\tau}^{(i)}) &= \sigma^2(X_T^{(i)} + \Delta_\tau X_T^{(i)}) \\ &= \sigma^2(X_T^{(i)}) + \sigma^2(\Delta_\tau X_T^{(i)}). \end{aligned} \tag{4.20}$$

Unlike the Poisson case, Lemma 27 in Chapter 3 cannot be applied since for a mixed Poisson random variable the mean is not equal to the variance. However, (3.39) holds true in the mixed Poisson setting. By substituting $\zeta = (X_T^{(1)}, X_T^{(2)})$, $\xi = (\Delta_\tau X_T^{(1)}, \Delta_\tau X_T^{(2)})$, and $x = y = \tau T^{-1}$ into (3.39), it can be shown that

$$\mathrm{Cov}(X_{T+\tau}^{(1)}, X_{T+\tau}^{(2)}) = (1 + \frac{\tau^2}{T^2}) \cdot \mathrm{Cov}(X_T^{(1)}, X_T^{(2)}). \tag{4.21}$$

Dividing both sides of (4.21) by $\sigma(X_{T+\tau}^{(1)})\,\sigma(X_{T+\tau}^{(2)})$ and using arguments similar to those used to show (4.16), we obtain

$$\rho(T+\tau) = \rho(T) \cdot (1 + \frac{\tau^2}{T^2}) \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sigma(X_{T+\tau}^{(1)})\sigma(X_{T+\tau}^{(2)})}.$$

By using induction on m and arguments similar those used in deriving (4.21), we obtain

$$\mathrm{Cov}(X_{mT+\tau}^{(1)}, X_{mT+\tau}^{(2)}) = (m + \frac{\tau^2}{T^2}) \cdot \mathrm{Cov}(X_T^{(1)}, X_T^{(2)}) \tag{4.22}$$

whence

$$\rho(mT+\tau) = \rho(T) \cdot (m + \frac{\tau^2}{T^2}) \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sigma(X_{mT+\tau}^{(1)})\sigma(X_{mT+\tau}^{(2)})}. \tag{4.23}$$

Induction on m can be used to show that $X_{mT}^{(i)}$ and $\Delta_\tau X_{mT}^{(i)}$ are independent, similar to $\Delta_\tau X_T^{(i)}$ and $X_T^{(i)}$, for $i \in \{1,2\}$. Therefore, we have that

$$\sigma^2(X_{mT}^{(i)} + \Delta_\tau X_{mT}^{(i)}) = \sigma^2(X_{mT}^{(i)}) + \sigma^2(\Delta_\tau X_{mT}^{(i)}) \tag{4.24}$$

for $i \in \{1,2\}$. Furthermore, since our construction of the process at the end of the time interval $[T, 2T]$ consists of drawing another bivariate random variable $(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$ that has the same joint

distribution as $(X_T^{(1)}, X_T^{(2)})$, we have that $\sigma^2(X_T^{(1)}) = \sigma^2(\Delta_T X_T^{(1)})$, whence

$$\begin{aligned}\sigma^2(X_{2T}^{(1)}) &= \sigma^2(X_T^{(1)} + \Delta_T X_T^{(1)}) \\ &= \sigma^2(X_T^{(1)}) + \sigma^2(\Delta_T X_T^{(1)}) \\ &= 2\sigma^2(X_T^{(1)}).\end{aligned} \tag{4.25}$$

By induction on m and an argument similar to the one used in deriving (4.25), we get that

$$\sigma^2(X_{mT}^{(1)}) = m\,\sigma^2(X_T^{(1)}). \tag{4.26}$$

Similarly,

$$\sigma^2(X_{mT}^{(2)}) = m\sigma^2(X_T^{(2)}). \tag{4.27}$$

Then, by using (4.24), (4.26), and (4.27), we can derive from (4.23) that

$$\begin{aligned}\rho(mT+\tau) &= \rho(T)(m + \frac{\tau^2}{T^2}) \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sqrt{\sigma^2(X_{mT}^{(1)} + \Delta_\tau X_{mT}^{(1)})}\sqrt{\sigma^2(X_{mT}^{(2)} + \Delta_\tau X_{mT}^{(2)})}} \\ &= \rho(T)(m + \frac{\tau^2}{T^2}) \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sqrt{\sigma^2(X_{mT}^{(1)}) + \sigma^2(\Delta_\tau X_{mT}^{(1)})}\sqrt{\sigma^2(X_{mT}^{(2)}) + \sigma^2(\Delta_\tau X_{mT}^{(2)})}} \\ &= \rho(T)(m + \frac{\tau^2}{T^2}) \\ &\quad \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sqrt{m\,\sigma^2(X_T^{(1)}) + \sigma^2(X_\tau^{(1)})}\sqrt{m\,\sigma^2(X_T^{(2)}) + \sigma^2(X_\tau^{(2)})}}.\end{aligned} \tag{4.28}$$

□

We show below how (4.28) leads to the asymptotic stationarity of the correlation coefficient.

Corollary 4 (Asymptotic Stationarity of the Forward Continuation)**.** *The correlation $\rho(mT+\tau)$ achieves asymptotic stationarity as $m \to \infty$. More specifically,*

$$\lim_{m\to\infty} \rho(mT+\tau) = \rho(T), \quad \textit{for any} \quad \tau \in [0,T]. \tag{4.29}$$

PROOF: From (4.28) we have that

$$\begin{aligned}\rho(mT+\tau) &= \rho(T)(m + \frac{\tau^2}{T^2}) \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sqrt{m\,\sigma^2(X_T^{(1)}) + \sigma^2(X_\tau^{(1)})}\sqrt{m\,\sigma^2(X_T^{(2)}) + \sigma^2(X_\tau^{(2)})}} \\ &= \rho(T)(m + \frac{\tau^2}{T^2}) \cdot \frac{1}{m} \cdot \frac{\sigma(X_T^{(1)})\sigma(X_T^{(2)})}{\sqrt{\sigma^2(X_T^{(1)}) + (1/m)\sigma^2(X_\tau^{(1)})}\sqrt{\sigma^2(X_T^{(2)}) + (1/m)\sigma^2(X_\tau^{(2)})}}.\end{aligned}$$

Passing to the limit as $m \to \infty$ in the standard manner, we obtain that

$$\lim_{m\to\infty} \rho(mT+\tau) = \rho(T),$$

as was to be proved. □

Figure 4.2 depicts the correlation structure obtained from simulating a bivariate Negative Binomial process where the mean of the intensities are 3 and 5 and the variance of the intensities are 5 and 7, respectively, calibrated to correlations of 0.7 and -0.8, respectively, at terminal time $T = 1$ using Backward Simulation on the interval $[0, 1]$ within the simulation interval $[0, 5]$ where Forward Continuation of the Backward Simulation is used on each of the intervals $[m, m + 1]$ for $m = 1, 2, 3, 4$. We used Backward Simulation with 1,000,000 Monte Carlo samples to compute the dashed black lines in Figure 4.2. The blue circles depict the theoretical values according to (4.28). Note the good agreement between the theoretical and the empirical results. The dashed red lines depict the Poisson case: two bivariate Poisson processes with intensities 3 and 5, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.8$, respectively, computed by the Forward-Backward Simulation approach.

The extreme positive and extreme negative correlations are 0.9972 and -0.8010 in the Negative Binomial case. Note the difference between these values and the extreme positive and negative correlations of 0.9972 and -0.96 from Figure 4.1 in Section 4.2. The difference in extreme negative correlations stems from the fact that the bivariate process is calibrated to the terminal time $T = 5$ in Figure 4.1 but to a terminal time of $T = 1$ in Figure 4.2, below.

4.4.2 Forward-Backward Simulation Algorithm in d-dimensions

Algorithm 22 Forward-Backward Simulation of correlated multivariate mixed Poisson processes

Require: Vector of marginal mixed Poisson distributions at time T
$\mathbf{MP}(T, \mathbf{U}) = (\mathrm{MP}(T, U^{(1)}), \ldots, \mathrm{MP}(T, U^{(d)}))$
Correlation matrix C
The number of intervals m (i.e., $[0, T], [T, 2T], \cdots, [(m-1)T, mT]$) used in the Forward-Backward Simulation Algorithm

Output: Scenarios of the multivariate Poisson process in the interval $[0, mT]$

Construct $\mathrm{MP}(T, U^{(k)})$ distributed marginals $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]$ for $k = 1, \ldots, d$
Generate the calibrated mixed Poisson measure $P^{(C)}$ using Algorithm 9, which takes as input $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]_{k=1}^{d}$ and C
for $i = 1, \ldots, m$ **do**
    Generate samples $(N^{(1)}, \ldots, N^{(d)}) \sim P^{(C)}$ using Algorithm 18 ▷ *Get the number of events in the interval* $[(i-1)T, iT]$
    for $k = 1, \ldots, d$ **do** ▷ *this can be done in parallel*
        Generate $N^{(k)}$ iid uniform random variables in the interval $[(i-1)T, iT]$: $\mathbf{T}_i^{(k)} = (T_1^{(k)}, \ldots, T_{N^{(k)}}^{(k)})'$ ▷ *column vector*
        Sort $\mathbf{T}_i^{(k)}$ in ascending order
        Append $\mathbf{T}_i^{(k)}$ to $\mathbf{T}^{(k)}$
return $\mathbf{T} = (\mathbf{T}^{(1)}, \ldots, \mathbf{T}^{(d)})$

4.5 Summary

In this chapter, we extended the Backward Simulation methodology to mixed Poisson processes. Similar to the Poisson case, the extreme joint distributions, constructed using the EJD approach in Chapter 2, enable the simulation of correlated multivariate mixed Poisson processes. In particular, Backward Simulation is able to generate all admissible distributions at the endpoint T of the interval $[0, T]$. Forward Continuation of Backward Simulation enables the process generated by Backward Simulation within the interval $[0, T]$ to be extended to intervals $[mT, (m+1)T]$ for any positive integer m, where the correlation displays the asymptotic behavior

$$\rho(mT + \tau) \approx \rho(T)$$

for all m sufficiently large, all τ in $[0, T]$, and for any admissible correlation matrix $\rho(T)$.

Our contributions, published in [18], are summarized below:

- We showed rigorously that the Backward Simulation approach applies to mixed Poisson processes.
- We derived an analytic form for the correlation structure between bivariate mixed Poisson processes. Note that this extends directly to multivariate settings since the Pearson correlation is bivariate in nature.
- We extended the Forward Continuation of the Backward Simulation to mixed Poisson processes and showed that the correlation structure under Forward Continuation attains asymptotic stationarity.

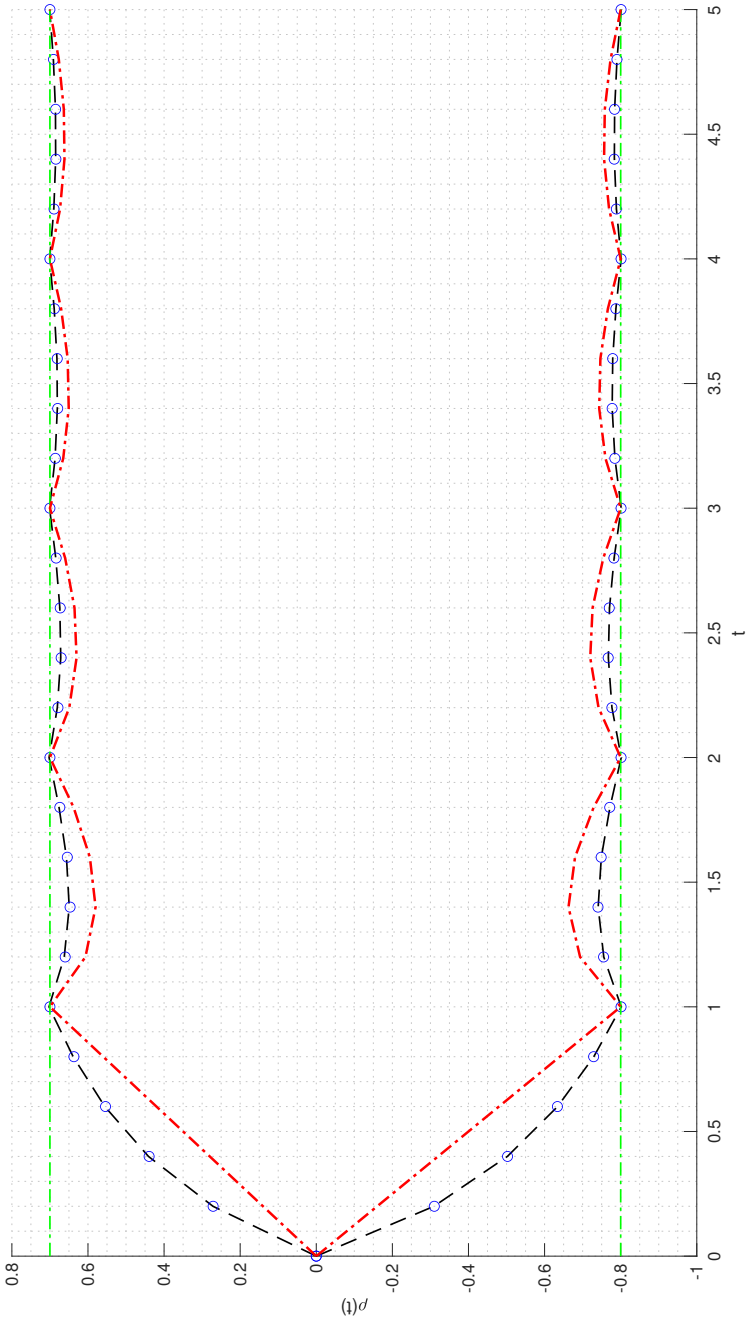

Figure 4.2: The dashed black lines depict the correlation structure for two bivariate Negative Binomial processes with means 3 and 5 and variances 5 and 7, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.8$, respectively, computed by the Forward-Backward Simulation approach. The blue circles depict the theoretical values according to (4.28). The dashed red lines depict the Poisson case: two bivariate Poisson processes with intensities 3 and 5, calibrated to $\rho(T) = 0.7$ and $\rho(T) = -0.8$, respectively, computed by the Forward-Backward Simulation approach.

Chapter 5

Backward Simulation of Compound Poisson Processes

The compound Poisson process is an important generalization of Poisson processes where the jump size is a random variable instead of being unit size. Compound Poisson processes, having jump sizes that are integer-valued, are known as *discrete compound Poisson processes*; they are known simply as *compound Poisson processes* when the jump size is real-valued. Compound Poisson processes are widely used in many areas related to applied probability. Some examples include Operational Risk [96], Insurance [36], Change-point Theory [22] and Algorithmic Trading [52].

Backward Simulation and Forward Continuation of Backward Simulation can be extended to compound Poisson processes. This is due to the special structure of the compound Poisson process which our Backward Simulation approach exploits.

Similar to the Poisson and the mixed Poisson cases, previous work on correlating bivariate compound Poisson distributions had restrictions on the attainable correlations and were unable to achieve the full range of admissible correlations [28, 118]. In particular, negative correlation between the components of a multivariate compound Poisson process, to the best of our knowledge, has been unattainable by previous work in the literature.

Finally, we note that while the majority of the exposition in this chapter is in the bivariate setting, the discussions generalize immediately to the multivariate setting. We also note that, while the results in this chapter are discussed in the setting of compound Poisson processes for simplicity, the results extend easily to compound mixed Poisson processes.

5.0.1 Outline

In Section 5.1, we review the basics of compound Poisson processes. The Backward Simulation of compound Poisson processes is discussed in Section 5.2. The correlation structure of compound Poisson processes generated by Backward Simulation is discussed in Section 5.3. The Forward Continuation of the Backward Simulation for compound Poisson processes and its correlation structure are discussed in Section 5.4. Section 5.5 closes and summarizes the chapter.

5.0.2 Notation

We make use of the following notation in this chapter.

Symbol	Definition
T_i	i^{th} arrival moment
S	Random variable representing a random sum
S_n	Random variable representing a random sum consisting of n elements
n	Realization of the random variable X_T
Z	Generic random variable or the jump variable for a compound Poisson process
F_Z	Secondary distribution of a compound Poisson process
$\boldsymbol{F_Z} = (F_Z^{(1)}, \ldots, F_Z^{(d)})$	d-dimensional vector of secondary distributions
$\mathbf{z}$	Secondary matrix
N_t	Poisson process
$\mathbf{N}_t = (N_t^{(1)}, \ldots, N_t^{(d)})$	d-dimensional Poisson process
X_t	Poisson process obtained from Backward Simulation
$\mathbf{X}_t = (X_t^{(1)}, \ldots, X_t^{(d)})$	Multivariate Poisson process obtained from Backward Simulation
λ	Mean parameter of a Poisson distribution or the intensity parameter of a Poisson process
$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_d)$	d-dimensional vector of intensities corresponding to $\mathbf{N}_t$ or $\mathbf{X}_t$
L_t	Compound Poisson process
Y_t	Compound Poisson process obtained from Backward Simulation
$\mathbf{Y}_t = (Y_t^{(1)}, \ldots, Y_t^{(d)})$	d-dimensional compound Poisson process obtained from Backward Simulation
$\Delta_s Y_t$	The increment $Y_{t+s} - Y_t$ of the process Y_t in interval $[t, t+s]$
$\stackrel{d}{=}$	Equal in distribution
$M_Z(u)$	Moment generating function of the random variable Z
D	Generic random variable
D_i	Random variable identically and independently distributed having the same distribution as D
C	Correlation matrix
$Q^{(k)}$	k^{th} marginal probability distribution
$Q_i^{(k)}$	k^{th} marginal probability at point i of the domain
$\rho(t)$	Correlation at time t between $Y_t^{(1)}$ and $Y_t^{(2)}$
$\hat{\rho}(t)$	Correlation at time t between $X_t^{(1)}$ and $X_t^{(2)}$

5.1 Compound Poisson Processes

We briefly review the basics of compound Poisson processes. Standard results for compound Poisson processes can be found in [40, 58].

Definition 33 (Compound Poisson Distribution [40]). *The random variable*

$$S = Z_1 + Z_2 + \cdots + Z_N$$

is said to be compound Poisson distributed if $N \sim \text{Pois}(\lambda)$ *and the random variables* $\{Z_i\}_{i=1}^{\infty}$ *are identically and independently distributed having the common distribution* F_Z *which is independent of* N *and has finite expectation and finite variance.* N *is known as the primary random variable having, in this case,* $\text{Pois}(\lambda)$ *as the primary distribution and* F_Z *is known as the secondary distribution.*

Remark 72. *For the remainder of this chapter, we assume that the secondary distribution* F_Z *has finite expectation and variance.*

Remark 73. *It is well known that the moment generating function of the sum* $S_n = Z_1 + \cdots + Z_n$*, where* $Z_1, \ldots, Z_n$ *are iid with the common distribution* F_Z *and* n *is fixed, is*

$$\begin{aligned} M_{S_n}(u) &= M_{Z_1}(u) M_{Z_2}(u) \cdots M_{Z_n}(u) \\ &= [M_Z(u)]^n, \end{aligned} \tag{5.1}$$

where $M_Z(u) = \mathbb{E}[\exp(uZ)]$ is the moment generating function of Z.

Definition 34 (Compound Poisson Process [40]). *The process*

$$L_t = \sum_{i=1}^{N_t} Z_i \tag{5.2}$$

is said to be a compound Poisson process if N_t*, known as the primary process, is a Poisson process with intensity* λ *and the random variables* Z_i*, are identically and independently distributed having the common distribution* F_Z *which has finite expectation and finite variance and is independent of the primary process* N_t*.*

Remark 74. *In Operational Risk,* N_t *in (5.2) is known as the frequency process and the secondary distribution* F_Z *is known as the loss distribution.*

The mean and the variance of a compound Poisson process are given as follows.

Proposition 5 (Moments of a Compound Poisson Process [40]). *The mean and variance of a compound Poisson process* L_t *having Poisson primary process* N_t *with intensity* λ *and secondary distribution* F_Z *is given by*

$$\mathbb{E}[L_t] = \lambda t \, \mathbb{E}[Z] \tag{5.3}$$

and

$$\sigma^2(L_t) = \lambda t \, \mathbb{E}(Z^2). \tag{5.4}$$

Proposition 6 (Moment Generating Function of a Compound Poisson Process [40])**.** *The moment generating function for a compound Poisson process L_t is*

$$\begin{aligned} M_{L_t}(u) &= \mathbb{E}\left[\exp(u \cdot \sum_{i=1}^{N_t} Z_i)\right] \\ &= \exp(\lambda t \left[M_Z(u) - 1\right]). \end{aligned} \tag{5.5}$$

Moreover, for an increment $\Delta_s L_t = L_{t+s} - L_t$ of the compound Poisson process L_t, the moment generating function takes the form

$$M_{\Delta_s L_t}(u) = \exp(\lambda s \left[M_Z(u) - 1\right]). \tag{5.6}$$

5.2 Backward Simulation

In this section, we prove the fundamental result, enabling the Backward Simulation of compound Poisson processes. That is, we consider, for the remainder of this section, a process Y_t, for $0 \leq t \leq T$, defined as

$$Y_t = \sum_{i=1}^{n} \mathbb{1}(T_i \leq t) \cdot Z_i \tag{5.7}$$

where n is a realization of the random variable $X_T \sim \text{Pois}(\lambda T)$, $\{T_i\}_{i=1}^n \subset [0,T]$ are independent and identically distributed random variables having a uniform conditional distribution,

$$\mathbb{P}(T_i \leq t \,|\, X_T = n) = \frac{t}{T} \quad i = 1, 2, \ldots, n \quad \text{and} \quad 0 \leq t \leq T \tag{5.8}$$

and the Z_i are independently and identically distributed random variables having the common distribution F_Z and also independent of X_T and T_i for $i = 1, 2, \ldots, n$. We remind the reader that, similar to the Poisson and the mixed Poisson cases, although Theorem 12 below is formulated and proved in the univariate setting, the results extend directly to the case where Y_t is a correlated multivariate compound Poisson process. This is because of the fact that, while the dependence structure is specified through the joint distribution of the multivariate Poisson random variable X_T, every coordinate (marginal) of a correlated multivariate compound Poisson process is itself a (univariate) compound Poisson process[1]. Therefore, given the joint number of events sampled from a suitable joint distribution[2] for X_T, the correlated multivariate compound Poisson process can be constructed within the simulation interval $[0,T]$ by applying Theorem 12 independently to each coordinate. See Algorithm 23 below.

The main result of the chapter is the following theorem.

Theorem 12 (Backward Simulation of Compound Poisson Processes)**.** *For $t \in [0,T]$, define the processes X_t and Y_t by*

$$X_t = \sum_{i=1}^{n} \mathbb{1}(T_i \leq t), \tag{5.9}$$

$$Y_t = \sum_{i=1}^{n} \mathbb{1}(T_i \leq t), \cdot Z_i \tag{5.10}$$

where n is a realization of the random variable $X_T \sim \text{Pois}(\lambda T)$, $\{T_i\}_{i=1}^n$ are identically and independently distributed random variables having the conditional uniform distribution

$$\mathbb{P}\left(T_i \leq t \,|\, X_T = n\right) = \frac{t}{T} \qquad i = 1, 2, \ldots, n, \quad 0 \leq t \leq T \tag{5.11}$$

and $\{Z_i\}_{i=1}^n$ is a sequence of identically and independently distributed random variables having the common distribution F_Z, which is independent of X_T and T_i for $i = 1, 2, \ldots, n$. Then, for $t \in [0,T]$, X_t is a Poisson process with intensity λ and Y_t is a compound Poisson process having the primary process X_t and the secondary distribution F_Z.

Proof: To begin, note that we proved in Theorem 6 that, for $t \in [0,T]$, the random process X_t

[1] See Definition 13.

[2] Recall that Chapter 2 is concerned with constructing extreme distributions that exhibit extreme correlations and constructing distributions that exhibit any *admissible* correlation that is a convex combination of extreme correlations.

defined in (5.9) is a Poisson process with intensity λ. Therefore, for $t \in [0,T]$, $s \geq 0$ and $[t,t+s] \subset [0,T]$, $\Delta_s X_t = X_{t+s} - X_t \sim \text{Pois}(\lambda s)$. Hence, for any non-negative integer k,

$$\mathbb{P}(\Delta_s X_t = k) = \mathrm{e}^{-\lambda s} \frac{(\lambda s)^k}{k!}. \tag{5.12}$$

Also, note that we can renumber the T_i, $i = 1, 2, \ldots, n$, so that $T_1 \leq T_2 \leq \cdots \leq T_n$. Since the T_i are uniformly distributed (5.11), we have that $T_1 < T_2 < \cdots < T_n$, almost surely. This renumbering does not affect the values of either of the sums (5.10) or (5.9) in Theorem 12, but it does allow us to rewrite (5.10) as

$$Y_t = \sum_{i=1}^{X_t} Z_i. \tag{5.13}$$

To use (5.13) in the proof below, assume that the T_i, $i = 1, 2, \ldots, n$, are ordered from smallest to largest for the remainder of this proof.

Since $Y_0 = 0$ almost surely, it follows from Definition 3 and Proposition 6 that, to prove the Y_t defined in (5.10) is a compound Poisson process for $t \in [0,T]$ having the primary process X_t, a Poisson process with intensity λ generated by Backward Simulation on the interval $[0,T]$, and having the secondary distribution F_Z, it is sufficient to prove that the following two statements hold.

1. For any interval $(t, t+s] \subset [0,T]$ of length $s \geq 0$, the increment $\Delta_s Y_t = Y_{t+s} - Y_t$ of the process Y_t has the moment generating function $M_{\Delta_s Y_t}(u) = \exp(\lambda s[M_Z(u) - 1])$, where $M_Z(u)$ is the moment generating function of the secondary distribution F_Z. Moreover, the distribution of the increment $\Delta_s Y_t$ does not depend on t.

2. For any $l = 2, 3, \ldots$ disjoint sub-intervals $(t_i, t_i + s_i] \subset [0,T]$ for $i = 1, 2, \ldots, l$, the random variables $\Delta_{s_i} Y_{t_i} = Y_{t_i+s_i} - Y_{t_i}$, for $i = 1, 2, \ldots, l$, are mutually independent.

We begin by proving the first statement above. To this end, choose any $t \geq 0$ and $s \geq 0$ such that $(t, t+s] \subset [0,T]$. Using (5.13), we get that

$$\begin{aligned} \Delta_s Y_t &= Y_{t+s} - Y_t \\ &= \sum_{i=1}^{X_{t+s}} Z_i - \sum_{i=1}^{X_t} Z_i \\ &= \sum_{i=X_t+1}^{X_{t+s}} Z_i \\ &= \sum_{i=1}^{\Delta_s X_t} Z_{X_t+i}. \end{aligned} \tag{5.14}$$

Using the definition of the moment generating function and (5.14), we get that

$$\begin{aligned} M_{\Delta_s Y_t}(u) &= \mathbb{E}\left[\exp(u \Delta_s Y_t)\right] \\ &= \mathbb{E}\left[\exp\left(u \sum_{i=1}^{\Delta_s X_t} Z_{X_t+i}\right)\right]. \end{aligned} \tag{5.15}$$

Expanding the last line of (5.15) as a conditional expectation and using the property $\mathrm{e}^{x+y} = e^x \, \mathrm{e}^y$ of

the exp function, we get that

$$\begin{aligned} M_{\Delta_s Y_t}(u) &= \sum_{k=0}^{\infty} \mathbb{E}\left[\exp\left(u \sum_{i=1}^{k} Z_{X_t+i}\right) \middle| \Delta_s X_t = k\right] \mathbb{P}(\Delta_s X_t = k) \\ &= \sum_{k=0}^{\infty} \mathbb{E}\left[\prod_{i=1}^{k} \exp\left(u Z_{X_t+i}\right) \middle| \Delta_s X_t = k\right] \mathbb{P}(\Delta_s X_t = k). \end{aligned} \tag{5.16}$$

Since the Z_{X_t+i}, for $i = 1, 2, \ldots, k$, are iid with the same distribution as the secondary distribution F_Z, it follows from (5.16) that

$$\begin{aligned} M_{\Delta_s Y_t}(u) &= \sum_{k=0}^{\infty} \left(\prod_{i=1}^{k} \mathbb{E}\left[\exp\left(u Z_{X_t+i}\right)\right]\right) \mathbb{P}(\Delta_s X_t = k) \\ &= \sum_{k=0}^{\infty} \left(\mathbb{E}\left[\exp\left(u Z\right)\right]\right)^k \mathbb{P}(\Delta_s X_t = k). \end{aligned} \tag{5.17}$$

Using $M_Z(u) = \mathbb{E}\left[\exp\left(uZ\right)\right]$ and (5.12), it follows from (5.17) that

$$\begin{aligned} M_{\Delta_s Y_t}(u) &= \sum_{k=0}^{\infty} \left(M_Z(u)\right)^k \mathbb{P}(\Delta_s X_t = k) \\ &= \sum_{k=0}^{\infty} \left(M_Z(u)\right)^k \mathrm{e}^{-\lambda s} \frac{(\lambda s)^k}{k!} \\ &= \mathrm{e}^{-\lambda s} \sum_{k=0}^{\infty} \frac{\left(\lambda s M_Z(u)\right)^k}{k!} \\ &= \mathrm{e}^{-\lambda s}\, \mathrm{e}^{\lambda s M_Z(u)} \\ &= \exp(\lambda s[M_Z(u) - 1]). \end{aligned} \tag{5.18}$$

Therefore, we have shown that $M_{\Delta_s Y_t}(u) = \exp(\lambda s[M_Z(u) - 1])$, where $M_Z(u)$ is the moment generating function of the secondary distribution F_Z. Moreover, the distribution of the increment $\Delta_s Y_t$ does not depend on t. Hence, we have shown that the first statement listed near the start of the proof holds.

Next, we prove the second statement listed near the start of the proof. To this end, for any $l = 2, 3, \ldots,$ consider any l disjoint sub-intervals $(t_i, t_i + s_i]$, for $i = 1, 2, \ldots, l$, where each $(t_i, t_i + s_i] \subset [0, T]$. Since the intervals are disjoint, there is a permutation σ of the integers $\{1, 2, \ldots, l\}$ such that

$$t_{\sigma(1)} \leq t_{\sigma(1)} + s_{\sigma(1)} \leq t_{\sigma(2)} \leq t_{\sigma(2)} + s_{\sigma(2)} \leq \cdots \leq t_{\sigma(l)} \leq t_{\sigma(l)} + s_{\sigma(l)}. \tag{5.19}$$

To ease the notational burden, we assume for the remainder of the proof that the t_i and s_i have been renumbered in accordance with (5.19). That is,

$$t_1 \leq t_1 + s_1 \leq t_2 \leq t_2 + s_2 \leq \cdots \leq t_l \leq t_l + s_l. \tag{5.20}$$

It follows from (5.20) and from our earlier stated assumption that the arrival moments are ordered that

$$X_{t_1} \leq X_{t_1+s_1} \leq X_{t_2} \leq X_{t_2+s_2} \leq \cdots \leq X_{t_l} \leq X_{t_l+s_l}. \tag{5.21}$$

Denote by $\Delta_{s_i} X_{t_i} = X_{t_i+s_i} - X_{t_i}$ the number of events occurring within the sub-interval $(t_i, t_i + s_i]$. Note that (5.21) implies that the sets

$$\{X_{t_i+j} : j = 1, 2, ..., \Delta_{s_i} X_{t_i}\} \qquad \text{for} \quad i = 1, 2, \ldots, l \tag{5.22}$$

are disjoint since the last element of the m^{th} set is $X_{t_m} + \Delta_{s_m} X_{t_m} = X_{t_m+s_m}$, the first element of the $(m+1)^{st}$ set is $X_{t_{m+1}}+1$ for $m = 1, 2, \ldots, l-1$ and we have from (5.21) that $X_{t_m+s_m} \leq X_{t_{m+1}} < X_{t_{m+1}}+1$.

The joint moment generating function of the l increments, $\Delta_{s_i} Y_{t_i}$, $i = 1, \ldots, l$, takes the form

$$M_{\Delta_{s_1} Y_{t_1} \cdots \Delta_{s_l} Y_{t_l}}(u_1, u_2, \ldots, u_l) = \mathbb{E}\left[\exp(u_1 \Delta_{s_1} Y_{s_1} + \cdots + u_l \Delta_{s_l} Y_{s_l})\right]. \tag{5.23}$$

Using (5.14), we can rewrite the right side of (5.23) as

$$\mathbb{E}\left[\exp\left(u_1 \cdot \sum_{i=1}^{\Delta s_1 X_{t_1}} Z_{X_{t_1+i}} + \cdots + u_l \cdot \sum_{i=1}^{\Delta s_l X_{t_l}} Z_{X_{t_l+i}}\right)\right]. \tag{5.24}$$

By the law of iterated expectations, we can expand (5.24) into a joint conditional expectation:

$$\begin{aligned} &\mathbb{E}\left[\mathbb{E}\left[\exp\left(u_1 \cdot \sum_{i=1}^{\Delta s_1 X_{t_1}} Z_{X_{t_1+i}} + \cdots + u_l \cdot \sum_{i=1}^{\Delta s_l X_{t_l}} Z_{X_{t_l+i}}\right) \middle| \Delta s_1 X_{t_1} = k_1, \ldots, \Delta s_l X_{t_l} = k_l\right]\right] \\ &\quad = \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \mathbb{E}\left[\exp(u_1 \cdot \sum_{i=1}^{k_1} Z_{X_{t_1+i}} + \cdots + u_l \cdot \sum_{i=1}^{k_l} Z_{X_{t_l+i}})\right] \\ &\qquad\qquad \cdot \mathbb{P}(\Delta s_1 X_{t_1} = k_1, \ldots, \Delta s_l X_{t_l} = k_l). \end{aligned} \tag{5.25}$$

Since, by Theorem 6, the increments $\Delta_{s_i} X_{t_i}$, for $i = 1, 2, \ldots, l$, are independent, the joint probability in (5.25) factors multiplicatively:

$$\mathbb{P}(\Delta s_1 X_{t_1} = k_1, \ldots, \Delta s_l X_{t_l} = k_l) = \mathbb{P}(\Delta s_1 X_{t_1} = k_1) \cdots \mathbb{P}(\Delta s_l X_{t_l} = k_l). \tag{5.26}$$

Moreover, using the basic property that $\mathrm{e}^{x+y} = \mathrm{e}^x \mathrm{e}^y$, we can rewrite the expectation term in the second

line of (5.25) as

$$\begin{aligned}&\mathbb{E}\Big[\exp(u_1\cdot\sum_{i=1}^{k_1}Z_{X_{t_1+i}}+\cdots+u_l\cdot\sum_{i=1}^{k_l}Z_{X_{t_l+i}})\Big]\\&\quad=\mathbb{E}\Big[\Big(\prod_{i=1}^{k_1}\exp(u_1Z_{X_{t_1+i}})\Big)\cdots\Big(\prod_{i=1}^{k_l}\exp(u_lZ_{X_{t_l+i}})\Big)\Big].\end{aligned}\tag{5.27}$$

Since, as noted above, the set of indices (5.22) corresponding to the sets $\{Z_{X_{t_i+j}}: j=1,2,...,\Delta_{s_i}X_{t_i}\}$, for $i=1,2,\ldots,l$, are disjoint and non-decreasing (5.21), each $Z_{X_{t_i+j}}$ appears exactly once in the product of exponential terms in the second line of (5.27). In addition, from the hypothesis of Theorem 12, the $Z_{X_{t_i+j}}$ are iid. Therefore, we can factor the expectation of products in the second line of (5.27) into a product of expectations:

$$\begin{aligned}&\mathbb{E}\Big[\Big(\prod_{i=1}^{k_1}\exp(u_1Z_{X_{t_1+i}})\Big)\cdots\Big(\prod_{i=1}^{k_l}\exp(u_lZ_{X_{t_l+i}})\Big)\Big]\\&\quad=\Big(\prod_{i=1}^{k_1}\mathbb{E}\left[\exp(u_1Z_{X_{t_1+i}})\right]\Big)\cdots\Big(\prod_{i=1}^{k_l}\mathbb{E}\left[\exp(u_lZ_{X_{t_l+i}})\right]\Big).\end{aligned}\tag{5.28}$$

Using again the property we used above to factor the expectation of a product of iid random variables into a product of expectations of those random variables, but now in the reverse direction, as well as the basic property that $\mathrm{e}^{x+y}=\mathrm{e}^x\mathrm{e}^y$, we can rewrite the last line of (5.28) as

$$\begin{aligned}&\Big(\prod_{i=1}^{k_1}\mathbb{E}\left[\exp(u_1Z_{X_{t_1+i}})\right]\Big)\cdots\Big(\prod_{i=1}^{k_l}\mathbb{E}\left[\exp(u_lZ_{X_{t_l+i}})\right]\Big)\\&\quad=\mathbb{E}\Big[\prod_{i=1}^{k_1}\exp(u_1Z_{X_{t_1+i}})\Big]\cdots\mathbb{E}\Big[\prod_{i=1}^{k_l}\exp(u_lZ_{X_{t_l+i}})\Big]\\&\quad=\mathbb{E}\Big[\exp\Big(u_1\cdot\sum_{i=1}^{k_1}Z_{X_{t_1+i}}\Big)\Big]\cdots\mathbb{E}\Big[\exp\Big(u_l\cdot\sum_{i=1}^{k_l}Z_{X_{t_l+i}}\Big)\Big].\end{aligned}\tag{5.29}$$

Combining (5.23)–(5.29), we get that

$$
\begin{aligned}
&M_{\Delta_{s_1} Y_{t_1} \cdots \Delta_{s_l} Y_{t_l}}(u_1, u_2, \ldots, u_l) \\
&\quad = \sum_{k_1=0}^{\infty} \cdots \sum_{k_l=0}^{\infty} \mathbb{E}\left[\exp\left(u_1 \cdot \sum_{i=1}^{k_1} Z_{X_{t_1+i}}\right)\right] \cdots \mathbb{E}\left[\exp\left(u_l \cdot \sum_{i=1}^{k_l} Z_{X_{t_l+i}}\right)\right] \\
&\qquad\qquad \cdot \mathbb{P}(\Delta s_1 X_{t_1} = k_1) \cdots \mathbb{P}(\Delta s_l X_{t_l} = k_l) \\
&\quad = \left(\sum_{k_1=0}^{\infty} \mathbb{E}\left[\exp\left(u_1 \cdot \sum_{i=1}^{k_1} Z_{X_{t_1+i}}\right)\right] \cdot \mathbb{P}(\Delta s_1 X_{t_1} = k_1)\right) \\
&\qquad\qquad \cdots \left(\sum_{k_l=0}^{\infty} \mathbb{E}\left[\exp\left(u_l \cdot \sum_{i=1}^{k_l} Z_{X_{t_l+i}}\right)\right] \cdot \mathbb{P}(\Delta s_l X_{t_l} = k_l)\right) \\
&\quad = \left(\sum_{k_1=0}^{\infty} \mathbb{E}\left[\exp\left(u_1 \cdot \sum_{i=1}^{\Delta s_1 X_{t_1}} Z_{X_{t_1+i}}\right) \middle| \Delta s_1 X_{t_1} = k_1\right] \cdot \mathbb{P}(\Delta s_1 X_{t_1} = k_1)\right) \\
&\qquad\qquad \cdots \left(\sum_{k_l=0}^{\infty} \mathbb{E}\left[\exp\left(u_l \cdot \sum_{i=1}^{\Delta s_l X_{t_l}} Z_{X_{t_l+i}}\right) \middle| \Delta s_l X_{t_l} = k_l\right] \cdot \mathbb{P}(\Delta s_l X_{t_l} = k_l)\right) \\
&\quad = \mathbb{E}\left[\exp\left(u_1 \cdot \sum_{i=1}^{\Delta s_1 X_{t_1}} Z_{X_{t_1+i}}\right)\right] \cdots \mathbb{E}\left[\exp\left(u_l \cdot \sum_{i=l}^{\Delta s_l X_{t_l}} Z_{X_{t_l+i}}\right)\right] \\
&\quad = \mathbb{E}\left[\exp(u_1\, \Delta_{s_1} Y_{t_1})\right] \cdots \mathbb{E}\left[\exp(u_l\, \Delta_{s_l} Y_{t_l})\right] \\
&\quad = M_{\Delta_{s_1} Y_{t_1}}(u_1) \cdots M_{\Delta_{s_l} Y_{t_l}}(u_l).
\end{aligned}
$$

Hence, we have shown that

$$M_{\Delta_{s_1} Y_{t_1} \cdots \Delta_{s_l} Y_{t_l}}(u_1, u_2, \ldots, u_l) = M_{\Delta_{s_1} Y_{t_1}}(u_1) \cdots M_{\Delta_{s_l} Y_{t_l}}(u_l).$$

That is, the joint moment generating function for the increments $\Delta_{s_i} Y_{t_i}$, $i = 1, \ldots, l$, factors into a product of the moment generating functions for each increment. Therefore, the random variables $\Delta_{s_i} Y_{t_i}$, $i = 1, \ldots, l$, are mutually independent. Consequently, we have shown that the second statement listed near the start of the proof holds.

□

Remark 75. *Theorem 12 is more general than Theorem 6 and Theorem 9 for the Poisson and mixed Poisson case, respectively, as it applies to both general compound Poisson and discrete compound Poisson processes.*

5.2.1 Backward Simulation Algorithm for compound Poisson processes in d-dimensions

Algorithm 23 Backward Simulation of correlated multivariate compound Poisson Processes

Require: Vector of marginal Poisson distributions at terminal time T

$\mathbf{Pois}(\boldsymbol{\lambda} T) = (\text{Pois}(\lambda_1 T), \ldots, \text{Pois}(\lambda_d T))$

Correlation matrix C

Vector of secondary distributions $\boldsymbol{F_Z} = (F_{Z^{(1)}}, \ldots, F_{Z^{(d)}})$

Output: Scenarios of the correlated multivariate compound Poisson process in $[0, T]$

1: Construct $\text{Pois}(\lambda_k T)$ distributed marginals $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]$ for $k = 1, \ldots, d$

2: Generate the Poisson calibrated measure $P^{(C)}$ using Algorithm 9, which takes as input $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]_{k=1}^{d}$ and C

3: Generate samples $(N^{(1)}, \ldots, N^{(d)}) \sim P^{(C)}$ using Algorithm 18 ▷ *Get the number of events at terminal time T*

4: **for** $k = 1, \ldots, d$ **do** ▷ *this can be done in parallel*

5: Generate $N^{(k)}$ iid uniform random variables in the interval $[0, T]$: $\mathbf{T}^{(k)} = (T_1^{(k)}, \ldots, T_{N^{(k)}}^{(k)})'$ ▷ *column vector*

6: Sort $\mathbf{T}^{(k)}$ in ascending order

7: Draw $N^{(k)}$ iid random variables having the distribution $F_{Z^{(k)}}$: $\mathbf{z}^{(k)} = (z_1^{(k)}, \cdots, z_{N^{(k)}}^{(k)})'$ ▷ *column vector*

8: **return** $\mathbf{T} = (\mathbf{T}^{(1)}, \ldots, \mathbf{T}^{(d)})$ and $\mathbf{z} = (\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(d)})$

5.3 Correlation Structure

Theorem 13. *Consider a bivariate compound Poisson process* $\boldsymbol{Y}_t = (Y_t^{(1)}, Y_t^{(2)})$*, having as its primary process* $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$*, a correlated bivariate Poisson process generated by Backward Simulation, with correlation* $\hat{\rho}(t) = \text{corr}(X_t^{(1)}, X_t^{(2)})$*, and having as its secondary distribution* $\boldsymbol{F_Z} = (F_{Z^{(1)}}, F_{Z^{(2)}})$*, where the jump variables,* $Z^{(1)}$ *and* $Z^{(2)}$*, are mutually independent and also independent to the primary process. Let* $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$ *denote the Pearson correlation coefficient of the bivariate compound Poisson process. Then,* $\rho(t)$ *satisfies*

$$\begin{aligned} \rho(t) &= \frac{t}{T} \cdot \rho(T) \\ &= \frac{t}{T} \cdot \frac{\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}} \cdot \hat{\rho}(T). \end{aligned} \tag{5.30}$$

Proof: Recall the well-known definition of the Pearson correlation coefficient

$$\text{corr}(Y_t^{(1)}, Y_t^{(2)}) = \frac{\text{Cov}(Y_t^{(1)}, Y_t^{(2)})}{\sigma(Y_t^{(1)})\,\sigma(Y_t^{(2)})}. \tag{5.31}$$

The covariance of a bivariate process $(Y_t^{(1)}, Y_t^{(2)})$ can be written as

$$\text{Cov}(Y_t^{(1)}, Y_t^{(2)}) = \mathbb{E}\left[Y_t^{(1)}\, Y_t^{(2)}\right] - \mathbb{E}\left[Y_t^{(1)}\right]\mathbb{E}\left[Y_t^{(2)}\right]. \tag{5.32}$$

As noted in the proof of Theorem 12, we can re-order the $T_i, i = 1, 2, ..., n$, from smallest to largest thereby allowing us to rewrite (5.10) as (5.13). Applying the same approach here, we can rewrite the first term to the right of the equal sign in (5.32) as

$$\mathbb{E}\left[Y_t^{(1)}\, Y_t^{(2)}\right] = \mathbb{E}\left[\sum_{i=1}^{X_t^{(1)}} Z_i^{(1)} \cdot \sum_{j=1}^{X_t^{(2)}} Z_j^{(2)}\right]. \tag{5.33}$$

Then, by Wald's identity [66], (5.33) can be written as

$$\mathbb{E}\left[Y_t^{(1)}\, Y_t^{(2)}\right] = \mathbb{E}\left[X_t^{(1)} X_t^{(2)} Z^{(1)} Z^{(2)}\right]. \tag{5.34}$$

Since the jump variables $(Z^{(1)}, Z^{(2)})$ are mutually independent as well as independent of $X_t^{(1)}$ and $X_t^{(2)}$, we can rewrite (5.34) as

$$\mathbb{E}\left[X_t^{(1)} X_t^{(2)}\right]\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]. \tag{5.35}$$

Using (5.3)[3], (5.34) and $\mathbb{E}\left[X_t^{(i)}\right] = \lambda_i t$, which follows from Theorem 12, the covariance (5.32) can be

[3]Since we showed that Y_t is also a compound Poisson process in Theorem 12.

written in terms of the covariance of the primary process

$$\begin{aligned}\text{Cov}\left(Y_t^{(1)}, Y_t^{(2)}\right) &= \mathbb{E}\left[X_t^{(1)} X_t^{(2)}\right]\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right] - (\lambda_1 t)\,\mathbb{E}\left[Z^{(1)}\right]\cdot(\lambda_2 t)\,\mathbb{E}\left[Z^{(2)}\right] \\ &= \mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]\cdot\left(\mathbb{E}\left[X_t^{(1)} X_t^{(2)}\right] - (\lambda_1 t)\,(\lambda_2 t)\right) \\ &= \mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]\cdot\left(\mathbb{E}\left[X_t^{(1)} X_t^{(2)}\right] - \mathbb{E}\left[X_t^{(1)}\right]\mathbb{E}\left[X_t^{(2)}\right]\right) \\ &= \mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]\cdot\text{Cov}(X_t^{(1)}, X_t^{(2)}).\end{aligned} \tag{5.36}$$

Then, by using (5.36), (5.4)[4], (5.31) and Theorem 7 in Chapter 3, we can derive

$$\begin{aligned}\text{corr}(Y_t^{(1)}, Y_t^{(2)}) &= \frac{\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]\cdot\text{Cov}(X_t^{(1)}, X_t^{(2)})}{\sqrt{\lambda_1 t\,\mathbb{E}\left[(Z^{(1)})^2\right]}\sqrt{\lambda_2 t\,\mathbb{E}\left[(Z^{(2)})^2\right]}} \\ &= \frac{\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}}\cdot\frac{\text{Cov}(X_t^{(1)}, X_t^{(2)})}{\sqrt{\lambda_1 t}\sqrt{\lambda_2 t}} \\ &= \frac{\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}}\cdot\text{corr}(X_t^{(1)}, X_t^{(2)}) \\ &= \frac{\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}}\cdot\frac{t}{T}\cdot\text{corr}(X_T^{(1)}, X_T^{(2)}) \\ &= \frac{\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}}\cdot\frac{t}{T}\cdot\hat{\rho}(T).\end{aligned} \tag{5.37}$$

In going from the second to third line in (5.37), above, we used the fact that $\sigma^2(X_t^{(k)}) = \lambda_k\, t$, for $k \in \{1, 2\}$, since, as proven in Theorem 12, $X_t^{(k)}$, for $k \in \{1, 2\}$, is a Poisson process. It follows immediately from (5.37) that

$$\rho(t) = \frac{\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}}\cdot\frac{t}{T}\cdot\hat{\rho}(T), \tag{5.38}$$

which proves that $\rho(t)$ is equal to the second line of (5.30). To complete the proof of (5.30), note that (5.38) implies that

$$\rho(T) = \frac{\mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}}\cdot\hat{\rho}(T). \tag{5.39}$$

It follows immediately from (5.38) and (5.39) that

$$\rho(t) = \frac{t}{T}\cdot\rho(T),$$

which completes the proof of (5.30). □

Remark 76. *The compound Poisson process, $\boldsymbol{Y}_t$, inherits the correlation structure of the underlying primary process, $\boldsymbol{X}_t$ as can be seen from (5.37).*

Figure 5.1 depicts the correlation structure obtained from simulating a bivariate compound Poisson process, $\boldsymbol{Y}_t = (Y_t^{(1)}, Y_t^{(2)})$, using Backward Simulation. The primary process, $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$, is a bivariate Poisson process having intensities 3 and 5, calibrated to correlations of 0.7 and -0.9, respectively,

[4]Since we showed that Y_t is also a compound Poisson process in Theorem 12.

at terminal time $T = 5$, within the simulation interval $[0, 5]$ computed by Backward Simulation. The correlation of the primary processes, $\hat{\rho}(t) = \text{corr}(X_t^{(1)}, X_t^{(2)})$, is depicted by the dashed red lines in the figure. The jump sizes, $Z^{(1)}$ and $Z^{(2)}$, are lognormally distributed with parameters $(2.1235, 0.5)$ and $(1.9449, 0.75)$, respectively; the corresponding correction factor in (5.37) takes a value of 0.6661. We used Backward Simulation with 1,000,000 Monte Carlo samples to compute $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$, depicted by the dashed black lines in Figure 5.1. The blue circles depict the theoretical values for $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$, according to Theorem 13. Note the good agreement between the theoretical and the empirical results.

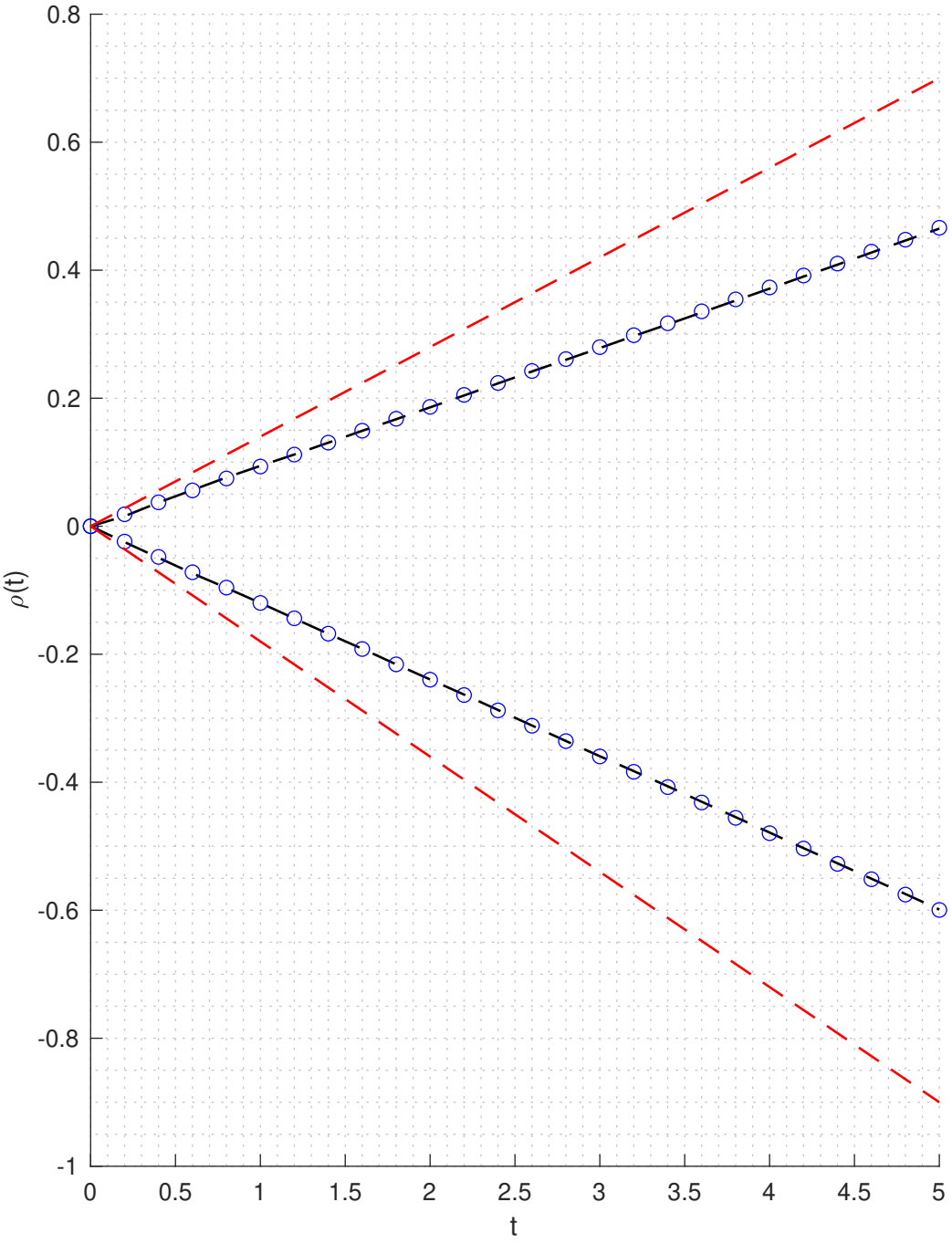

Figure 5.1: The dashed black lines depict the correlation structure, $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$, for a bivariate compound Poisson process, $\boldsymbol{Y}_t = (Y_t^{(1)}, Y_t^{(2)})$, having a primary processes, $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$, a bivariate Poisson process with intensities 3 and 5, calibrated to $\hat{\rho}(5) = 0.7$ and $\hat{\rho}(5) = -0.9$, respectively, computed by Backward Simulation. The jump sizes $F_{Z^{(1)}}$ and $F_{Z^{(2)}}$ are lognormally distributed with parameters $(2.1235, 0.5)$ and $(1.9449, 0.75)$, respectively. The parameters for the jump distributions were selected for illustrative purposes. The blue circles depict the theoretical values for $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$ according to Theorem 13. The dashed red lines depict the correlation structure, $\hat{\rho}(t) = \text{corr}(X_t^{(1)}, X_t^{(2)})$, of the primary process, i.e., the bivariate Poisson process $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$.

5.4 Forward Continuation of the Backward Simulation

Forward Continuation also applies to the compound Poisson setting. Suppose that a bivariate compound Poisson process $\boldsymbol{Y}_t = (Y_t^{(1)}, Y_t^{(2)})$ has already been simulated in the interval $[0, T]$ by Backward Simulation and that we wish to continue forward the process to the subsequent interval $[T, 2T]$. First, we draw an independent sample of the joint frequency distribution at time T

$$(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)}) \stackrel{d}{=} (X_T^{(1)}, X_T^{(2)}).$$

Note that the bivariate random variable $(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$ has the same prescribed marginal Poisson distributions and prescribed correlation, C, between its components as $(X_T^{(1)}, X_T^{(2)})$. Then, we set

$$(X_{2T}^{(1)}, X_{2T}^{(2)}) = (X_T^{(1)}, X_T^{(2)}) + (\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$$

and, as in Backward Simulation, we use $(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$ and the conditional independence property (similar to (3.3)) to generate the uniform arrival moments $\{T_i^{(1)}\}_{i=1}^{\Delta_T X_T^{(1)}}$ and $\{T_i^{(2)}\}_{i=1}^{\Delta_T X_T^{(2)}}$ in the interval $[T, 2T]$. Unlike the Poisson case, since the jumps of the compound Poisson process are no longer of unit size, we must also generate $\Delta_T X_T^{(i)}$ number of $Z^{(i)}$ distributed variables for $i \in \{1, 2\}$.

The process described above can be repeatedly applied to extend a bivariate compound Poisson process simulated within an interval $[0, T]$ to any subsequent interval $[mT, (m+1)T]$ for any non-negative integer m. Moreover, similar to the Poisson and the mixed Poisson cases, although the exposition here is in the bivariate setting, the results extend straightforwardly to the general d-dimensional setting.

Algorithm 24 in Section 5.4.2 below combines Backward Simulation and Forward Continuation of Backward Simulation to generate arrival moments for a d-dimensional compound Poisson process on an interval $[0, mT]$, for any positive integer m. We call this algorithm Forward-Backward Simulation because it starts with one step of Backward Simulation followed by $m-1$ steps of Forward Continuation of Backward Simulation.

5.4.1 Forward Correlation Structure

We analyze the behavior of the correlation coefficient under the Forward Continuation of Backward Simulation by deriving an expression for $\rho(mT + \tau) = \text{corr}(Y_{mT+\tau}^{(1)}, Y_{mT+\tau}^{(2)})$ as a function of $\rho(T) = \text{corr}(Y_T^{(1)}, Y_T^{(2)})$ for any non-negative integer m and any $\tau \in [0, T]$ in Theorem 14 below. That $\rho(mT+\tau)$ attains asymptotic stationarity follows by Corollary 5 below.

Theorem 14. *The correlation coefficient* $\rho(mT+\tau) = \text{corr}(Y_{mT+\tau}^{(1)}, Y_{mT+\tau}^{(2)})$ *for any non-negative integer* m *and any* $\tau \in [0, T]$ *as a function of* $\rho(T) = \text{corr}(Y_T^{(1)}, Y_T^{(2)})$ *takes the form*

$$\rho(mT + \tau) = \rho(T) \frac{m + \tau^2 \cdot T^{-2}}{m + \tau T^{-1}}. \tag{5.40}$$

Proof: First note that, for $m = 0$, (5.40) follows from Theorem 13. To show that (5.40) also holds for all $m \geq 1$, we begin by deriving an expression for $\rho(T + \tau)$. To that end, let $\Delta_\tau Y_T^{(1)} = Y_{T+\tau}^{(1)} - Y_T^{(1)}$ and $\Delta_\tau Y_T^{(2)} = Y_{T+\tau}^{(2)} - Y_T^{(2)}$ and note that each of the increments $\Delta_\tau Y_T^{(1)}$ and $\Delta_\tau Y_T^{(2)}$ is independent of both $Y_T^{(1)}$ and $Y_T^{(2)}$. Therefore the covariance of the bivariate compound Poisson process $(Y_t^{(1)}, Y_t^{(2)})$

satisfies

$$\mathrm{Cov}(Y^{(1)}_{T+\tau}, Y^{(2)}_{T+\tau}) = \mathrm{Cov}(Y^{(1)}_T, Y^{(2)}_T) + \mathrm{Cov}(\Delta_\tau Y^{(1)}_T, \Delta_\tau Y^{(2)}_T). \tag{5.41}$$

Similar to the proof of Theorem 13, let us rewrite the covariance of the bivariate compound Poisson process $\boldsymbol{Y}_t = (Y^{(1)}_t, Y^{(2)}_t)$ in terms of its primary processes $\boldsymbol{X}_t = (X^{(1)}_t, X^{(2)}_t)$ and its secondary distributions $(F_{Z^{(1)}}, F_{Z^{(2)}})$. Towards that end, we begin by considering the second term on the right side of (5.41). Using an analysis similar to that used to derive (5.36), we can rewrite the second term on the right side of (5.41) as a function of the covariance of its frequency process:

$$\mathrm{Cov}(\Delta_\tau Y^{(1)}_T, \Delta_\tau Y^{(2)}_T) = \mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right] \cdot \mathrm{Cov}(\Delta_\tau X^{(1)}_T, \Delta_\tau X^{(2)}_T). \tag{5.42}$$

Since the primary process, $\boldsymbol{X}_t = (X^{(1)}_t, X^{(2)}_t)$, is a Poisson process, we can use (3.48) directly in (5.42) to obtain

$$\mathrm{Cov}(\Delta_\tau Y^{(1)}_T, \Delta_\tau Y^{(2)}_T) = \mathbb{E}\left[Z^{(1)}\right]\mathbb{E}\left[Z^{(2)}\right] \cdot \mathrm{Cov}(X^{(1)}_T, X^{(2)}_T) \cdot \frac{\tau^2}{T^2}. \tag{5.43}$$

Using (5.36) and (5.43), we can derive from (5.41) that

$$\mathrm{Cov}(Y^{(1)}_{T+\tau}, Y^{(2)}_{T+\tau}) = \mathbb{E}\left[Z^{(1)}Z^{(2)}\right] \cdot \frac{T^2+\tau^2}{T^2} \cdot \mathrm{Cov}(X^{(1)}_T, X^{(2)}_T). \tag{5.44}$$

Dividing both sides of (5.44) by $\sigma(Y^{(1)}_{T+\tau})\,\sigma(Y^{(2)}_{T+\tau})$ and using (5.4)[5], we obtain

$$\begin{aligned}
\mathrm{corr}(Y^{(1)}_{T+\tau}, Y^{(2)}_{T+\tau}) &= \frac{\mathbb{E}\left[Z^{(1)}Z^{(2)}\right]}{\sigma(Y^{(1)}_{T+\tau})\,\sigma(Y^{(2)}_{T+\tau})} \cdot \frac{T^2+\tau^2}{T^2} \cdot \mathrm{Cov}(X^{(1)}_T, X^{(2)}_T) \\
&= \frac{\mathbb{E}\left[Z^{(1)}Z^{(2)}\right]}{\sqrt{\lambda_1(T+\tau)\mathbb{E}\left[(Z^{(1)})^2\right]}\sqrt{\lambda_2(T+\tau)\mathbb{E}\left[(Z^{(2)})^2\right]}} \cdot \frac{T^2+\tau^2}{T^2} \cdot \mathrm{Cov}(X^{(1)}_T, X^{(2)}_T) \\
&= \frac{\mathbb{E}\left[Z^{(1)}Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}} \cdot \frac{1}{\sqrt{\lambda_1(T+\tau)\lambda_2(T+\tau)}} \cdot \frac{T^2+\tau^2}{T^2} \cdot \mathrm{Cov}(X^{(1)}_T, X^{(2)}_T) \\
&= \frac{\mathbb{E}\left[Z^{(1)}Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}} \cdot \frac{1}{\sqrt{\lambda_1\lambda_2}} \cdot \frac{T^2+\tau^2}{(T+\tau)T^2} \cdot \mathrm{Cov}(X^{(1)}_T, X^{(2)}_T) \\
&= \frac{T^2+\tau^2}{(\tau+T)T} \cdot \frac{\mathbb{E}\left[Z^{(1)}Z^{(2)}\right]}{\sqrt{\mathbb{E}\left[(Z^{(1)})^2\right]\mathbb{E}\left[(Z^{(2)})^2\right]}} \cdot \mathrm{corr}(X^{(1)}_T, X^{(2)}_T) \\
&= \frac{T^2+\tau^2}{(\tau+T)T}\ \mathrm{corr}(Y^{(1)}_T, Y^{(2)}_T),
\end{aligned} \tag{5.45}$$

whence

$$\rho(T+\tau) = \rho(T) \cdot \frac{1+\tau^2 \cdot T^{-2}}{1+\tau T^{-1}}. \tag{5.46}$$

Using an argument similar to that used to derive (3.45) from (3.52) in Chapter 3, it follows from (5.46) that

$$\rho(mT+\tau) = \rho(T)\frac{m+\tau^2 \cdot T^{-2}}{m+\tau T^{-1}} \tag{5.47}$$

for any integer $m \geq 1$ and any $\tau \in [0, T]$. □

[5] Since we showed that Y_t is also a compound Poisson process in Theorem 12.

Corollary 5 (Asymptotic Stationarity of the Forward Continuation)**.** *The correlation* $\rho(mT+\tau) = \text{corr}(Y^{(1)}_{mT+\tau}, Y^{(2)}_{mT+\tau})$ *achieves asymptotic stationarity as* $m \to \infty$*. More specifically,*

$$\lim_{m\to\infty} \rho(mT+\tau) = \rho(T) \qquad \textit{for any } \tau \in [0,T]. \tag{5.48}$$

PROOF: Note that (5.40) has the exact same form as (3.45) in Corollary 3 in Chapter 3. Therefore the proof of Corollary 5 is the same as the proof in Corollary 3. □

Figure 5.2 depicts the correlation structure obtained from simulating a bivariate compound Poisson process, $\boldsymbol{Y}_t = (Y^{(1)}_t, Y^{(2)}_t)$, within the simulation interval $[0,5]$ using Backward Simulation on the interval $[0,1]$ and Forward Continuation of the Backward Simulation on each of the intervals $[m, m+1]$ for $m = 1,2,3,4$. The primary process, $\boldsymbol{X}_t = (X^{(1)}_t, X^{(2)}_t)$, is a bivariate Poisson process having intensities 3 and 5, calibrated to correlations of 0.7 and -0.9, respectively, at terminal time $T = 1$. The correlation of the primary processes, $\hat{\rho}(t) = \text{corr}(X^{(1)}_t, X^{(2)}_t)$, is depicted by the dashed red lines in the figure. The secondary distributions, $F_{Z^{(1)}}$ and $F_{Z^{(2)}}$, are lognormal distributions with parameters $(2.1235, 0.5)$ and $(1.9449, 0.75)$, respectively; the corresponding correction factor in (5.45) takes a value of 0.6661. We used Backward Simulation with 1,000,000 Monte Carlo samples to compute $\rho(t) = \text{corr}(Y^{(1)}_t, Y^{(2)}_t)$, the dashed black lines in Figure 5.2. The blue circles depict the theoretical values for $\rho(t) = \text{corr}(Y^{(1)}_t, Y^{(2)}_t)$, according to (5.40). Note the good agreement between the theoretical and the empirical results.

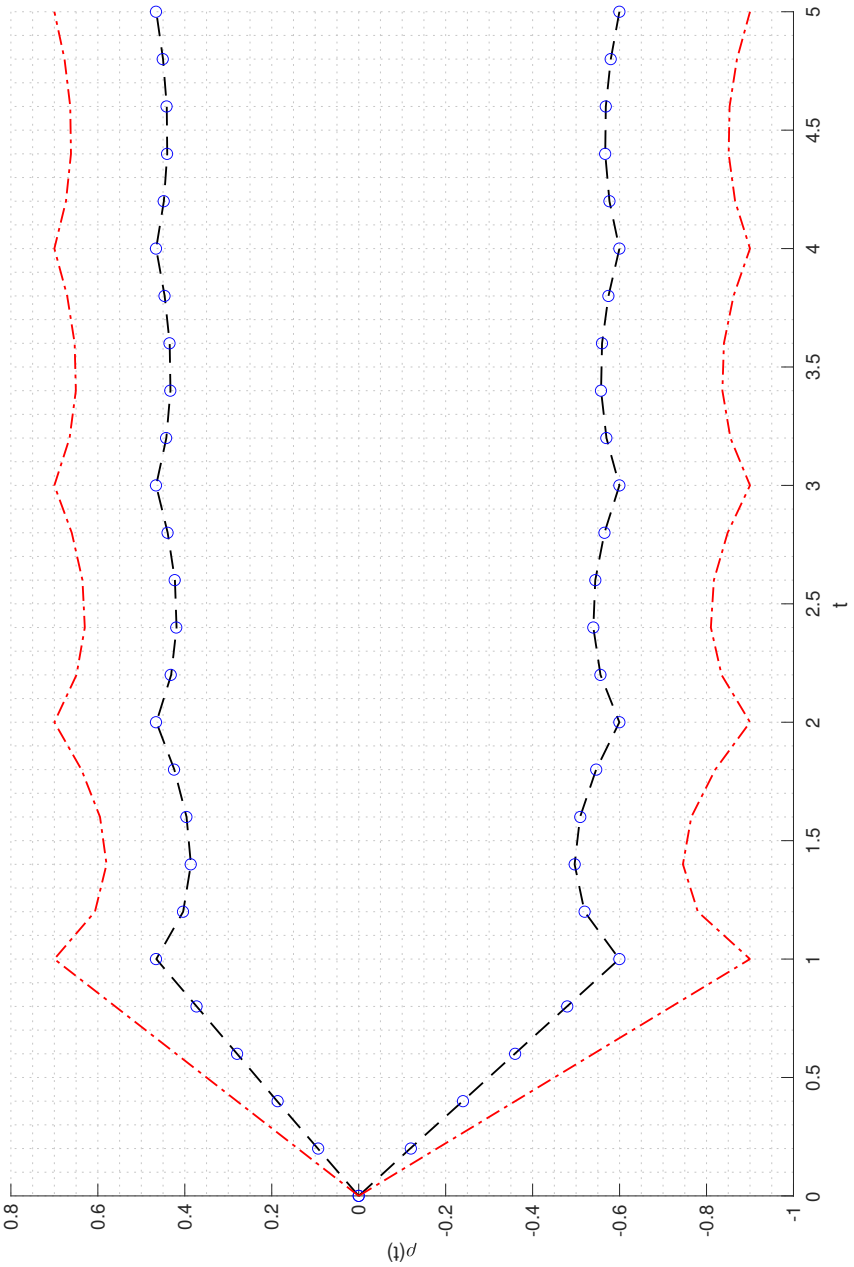

Figure 5.2: The dashed black lines depict the correlation structure, $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$, for a bivariate compound Poisson process, $\boldsymbol{Y}_t = (Y_t^{(1)}, Y_t^{(2)})$, computed by the Forward-Backward Simulation approach. The primary process $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$, is a bivariate Poisson process with intensities 3 and 5, calibrated to $\hat{\rho}(1) = 0.7$ and $\hat{\rho}(1) = -0.9$, respectively. The secondary distributions $F_{Z^{(1)}}$ and $F_{Z^{(2)}}$ are lognormal distributions with parameters $(2.1235, 0.5)$ and $(1.9449, 0.75)$, respectively. The blue circles depict the theoretical values for $\rho(t) = \text{corr}(Y_t^{(1)}, Y_t^{(2)})$ according to (5.40). The dashed red line depicts the correlation structure, $\hat{\rho}(t) = \text{corr}(X_t^{(1)}, X_t^{(2)})$, of the primary process, $\boldsymbol{X}_t = (X_t^{(1)}, X_t^{(2)})$.

5.4.2 Forward-Backward Simulation Algorithm in d-dimensions

Algorithm 24 Forward-Backward Simulation of correlated multivariate compound Poisson processes

Require: Vector of marginal Poisson distributions at terminal time T
$\mathbf{Pois(\boldsymbol{\lambda} T)} = (\text{Pois}(\lambda_1 T), \ldots, \text{Pois}(\lambda_d T))$
Correlation matrix C
Vector of secondary distributions $\boldsymbol{F_Z} = (F_{Z^{(1)}}, \ldots, F_{Z^{(d)}})$
The number of intervals m (i.e., $[0,T], [T,2T], \cdots, [(m-1)T, mT]$) used in the Forward-Backward Simulation Algorithm

Output: Scenarios of the multivariate compound Poisson process in the interval $[0, mT]$

Construct $\text{Pois}(\lambda_k T)$ distributed marginals $\{[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]\}$ for $k = 1, \ldots, d$
Generate the Poisson calibrated measure $P^{(C)}$ using Algorithm 9, which takes as input $[Q_0^{(k)}, \ldots, Q_{i_{\max}^{(k)}}^{(k)}]_{k=1}^{d}$ and C
for $i = 1, \ldots, m$ **do**
    Generate samples $(N^{(1)}, \ldots, N^{(d)}) \sim P^{(C)}$ using Algorithm 18 ▷ *Get the number of events in the interval* $[(i-1)T, iT]$
    for $k = 1, \ldots, d$ **do** ▷ *this can be done in parallel*
        Generate $N^{(k)}$ iid uniform random variables in the interval $[(i-1)T, iT]$: $\mathbf{T}_i^{(k)} = (T_1^{(k)}, \ldots, T_{N^{(k)}}^{(k)})'$ ▷ *column vector*
        Sort $\mathbf{T}_i^{(k)}$ in ascending order
        Draw $N^{(k)}$ iid uniform random variables having the distribution $F_{Z^{(k)}}$: $\mathbf{z}_i^{(k)} = (z_1^{(k)}, \cdots, z_{N^{(k)}}^{(k)})'$
        Append $\mathbf{T}_i^{(k)}$ to $\mathbf{T}^{(k)}$
        Append $\mathbf{z}_i^{(k)}$ to $\mathbf{z}^{(k)}$
return $\mathbf{T} = (\mathbf{T}^{(1)}, \ldots, \mathbf{T}^{(d)})$ and $\mathbf{z} = (\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(d)})$

5.5 Summary

In this chapter, we extended the Backward Simulation methodology to compound Poisson processes that, in conjunction with the extreme joint distributions constructed using the EJD approach introduced in Chapter 2, enables the simulation of correlated multivariate compound Poisson processes. In particular, Backward Simulation can generate all admissible distributions at the endpoint T of the interval $[0, T]$. Forward Continuation of Backward Simulation also extends to the compound Poisson case and enables the process generated by Backward Simulation within the interval $[0, T]$ to be extended to intervals $[mT, (m+1)T]$ for any positive integer m, where the correlation also displays the asymptotic behavior

$$\rho(mT + \tau) \approx \rho(T)$$

for all m sufficiently large, all τ in $[0, T]$, and for any admissible correlation matrix $\rho(T)$.

Our contributions are summarized below:

- We showed rigorously that the Backward Simulation approach can be applied to the compound Poisson process.
- We derived an analytic form for the correlation structure between bivariate compound Poisson processes computed by Backward Simulation.
- We extended the Forward-Continuation of the Backward Simulation to the compound Poisson process and derived an analytic form for the correlation structure between bivariate compound Poisson processes computed by Forward-Continuation of the Backward Simulation.

Chapter 6

Conclusions and Future Work

6.1 Summary and Conclusions

In this thesis, we develop a methodology for constructing multivariate Poisson processes that exhibit negative correlations between their components and satisfy given marginal constraints. Our methodology relies on two pillars:

- The Extreme Joint Distribution (EJD) approach for constructing joint distributions having given marginal distributions and a specified correlation structure.
- Backward Simulation of multivariate Poisson processes within a simulation interval $[0, T]$.

First introduced in the bivariate setting [74], the EJD methodology is a pure probabilistic based approach for constructing joint distributions having extreme correlation structures, referred to simply as extreme joint distributions (or extreme measures) in this thesis. In the bivariate setting, which we review in Chapter 2, there are only extreme positive and extreme negative correlations; we denote the associated extreme joint distributions by $\hat{P}^{(1)}$ and $\hat{P}^{(2)}$ respectively. An extreme measure can be characterized as the solution to an infinite-dimensional linear program where the objective function maximizes or minimizes the joint expectation and where the constraints are the given marginal distributions that the joint distribution must satisfy (Definition 14). The theoretical underpinnings for our approach is given in Theorem 1 and in Algorithm 1. The importance of constructing extreme joint distributions is due to the fact that, for each extreme joint distribution $\hat{P}^{(i)}$, there corresponds an extreme correlation $\hat{C}^{(i)}$ for $i \in \{1, 2\}$ in the bivariate setting and that any correlation $C \in [\hat{C}^{(1)}, \hat{C}^{(2)}]$ can be obtained via the solution of the linear equation

$$C = w\hat{C}^{(1)} + (1 - w)\hat{C}^{(2)} \tag{6.1}$$

where $0 \leq w \leq 1$. By itself, the solution w to (6.1) is not significant. However, the w that satisfies (6.1) can be substituted into

$$P = w\hat{P}^{(1)} + (1 - w)\hat{P}^{(2)} \tag{6.2}$$

to construct the probability distribution P having the given marginals and the associated correlation C. Any correlation, C, that can be obtained from (6.1) with $0 \leq w \leq 1$ is called *admissible*.

Chapter 2 also extends the EJD methodology to the general d-dimensional setting for $d \geq 3$. We show that both the EJD theorem and algorithm extend to d-dimensions. In higher dimensions, instead

of having only two extreme measures, with extreme positive and extreme negative dependence, there are $n = 2^{(d-1)}$ extreme measures with varying extreme dependencies between their components. To describe and keep track of the possible extreme dependence between pairs of components of a multivariate distribution, we introduced the concept of monotone structures (Section 2.4.1). In d-dimensions, an extreme measure can be characterized as the solution to a multi-objective infinite-dimensional linear programming problem, where the (bivariate) objective functions, specified by the associated monotone structure, minimize or maximize the joint expectation of a pair of components of the multivariate distribution and where the constraints are the marginal distributions that the joint distribution must satisfy (Definition 22). In addition, the extreme correlations $\hat{C}^{(1)}, \ldots, \hat{C}^{(n)}$ are matrices, in this case.

In addition, our approach associated with (6.1) extends to d-dimensions with some modification: since there are n extreme measures, the linear equation (6.1) becomes

$$C = w_1 \hat{C}^{(1)} + \cdots + w_n \hat{C}^{(n)} \tag{6.3}$$

where $0 \leq w_i \leq 1$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} w_i = 1$. Similar to the bivariate setting, by solving (6.3) for the weights $(w_1, \ldots, w_n)$ and substituting them into

$$P^{(C)} = w_1 \hat{P}^{(1)} + \cdots + w_n \hat{P}^{(n)} \tag{6.4}$$

we obtain a probability measure $P^{(C)}$ having the given marginals and the associated correlation matrix C from (6.3). Note that (6.3) is a linear system of equations with constraints on the coefficients, $\{w_i : i = 1, 2, \ldots, n\}$, that can be solved using techniques from Mathematical Programming, as explained in Section 2.8.1. If there does not exist a solution to (6.3), then the desired correlation matrix C is said to be *inadmissible*. That is, there is no joint distribution with the given marginals that has the correlation matrix C (see Section 2.8.3). Otherwise, C is said to be *admissible*. The EJD methodology is a crucial pillar of this thesis since it enables the construction of joint distributions with a prescribed dependence structure satisfying given marginals, which are necessary inputs to the Backward Simulation methodology.

The second pillar of this thesis is the Backward Simulation of correlated multivariate Poisson processes. Backward Simulation relies on the conditional uniformity property of Poisson processes (Proposition 3): given the number of events, n, at terminal simulation time T, the arrival moments of the Poisson process are uniformly distributed. The conditional uniformity property suggests a (stochastic) simulation method for a Poisson process within the simulation interval $[0, T]$: 1) compute a pseudo-random variable that gives the number of events, n, at terminal time; 2) draw n iid uniform variables; 3) sort the n uniform variables. Note that the importance of the EJD method in this context is now clear: the joint distribution obtained from the EJD algorithm is necessary for sampling the number of events of the Poisson process at terminal time, T. A major advantage of Backward Simulation is that the aforementioned procedure extends directly to the multivariate setting. Since a vector of terminal events can be obtained from a suitable joint distribution, Backward Simulation of a multivariate Poisson process simply consists of applying Backward Simulation to each univariate component using the corresponding number of terminal events; the dependency information is contained within the joint distribution. A key property of Backward Simulation is that the correlation structure is a linear function of time (Theorem 3.42). That is,

$$\rho(t) = \frac{t}{T} \cdot \rho(T)$$

We also introduced a method for extending a Poisson process simulated using Backward Simulation in $[0, T]$ to any subsequent interval $[mT, (m+1)T]$ for any integer $m \geq 1$, known as the Forward Continuation of Backward Simulation. We refer to the combined Backward Simulation and Forward Continuation of Backward Simulation as the Forward-Backward Simulation Algorithm (or the Forward-Backward approach). Consider extending a bivariate bivariate correlated Poisson process simulated using Backward Simulation on $[0, T]$ to $[T, 2T]$. Independently draw a bivariate random variable $(\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$ having the same prescribed marginal Poisson distributions and prescribed correlation, C, between its components as $(X_T^{(1)}, X_T^{(2)})$ from the joint distribution and define the number of terminal events at $t = 2T$ as follows:

$$(X_{2T}^{(1)}, X_{2T}^{(2)}) = (X_T^{(1)}, X_T^{(2)}) + (\Delta_T X_T^{(1)}, \Delta_T X_T^{(2)})$$

The process is filled in using Backward Simulation in $[T, 2T)$. This Forward Continuation of Backward Simulation can be extended to any interval $[mT, (m+1)T]$ for any integer $m \geq 1$. Surprisingly, the correlation structure of such a process that is simulated by the Forward-Backward approach attains asymptotic stationarity (Corollary 3). We extend Backward Simulation and Forward Continuation of Backward Simulation to the mixed Poisson processes (Chapter 4) and to the compound Poisson process in Chapter 5. We also analyze their correlation structure in Chapters 4 and 5.

Before our work in this area, it was not possible to construct multivariate Poisson process having extreme positive and negative correlation between their components. Our EJD approach, by constructing joint distributions with extreme dependence between their components, in conjunction with the Forward-Backward Simulation Algorithm is able to construct correlated multivariate Poisson processes with extreme dependence between their components and simulate them within a simulation interval $[0, mT]$, for any integer $m \geq 1$. Moreover, this approach works for any admissible correlation matrix. In particular, this thesis enables the accurate modeling of counting processes that exhibit negative correlation between their components.

6.1.1 Our Contributions

The work in this thesis builds on earlier work of Kreinin [74]. I list below the contributions that I made to our research project.

1. Proved various properties of comonotone and antimonotone distributions. (Subsection 2.2.1.)
2. Proved the correctness of Algorithm 1. (Section 2.3.)
3. Extended the notion of extreme measures to the d-dimensional setting and proved that the EJD theorem can be extended to the d-dimensional setting. (Section 2.4.)
4. Extended the EJD algorithm to d-dimensions and proved its correctness. (Section 2.5.)
5. Introduced a method for the calibration of multivariate discrete extreme measures to observed correlation structures. (Section 2.8.)
6. Introduced a method for sampling extreme measures and calibrated measures. (Section 2.9.)
7. Introduced the Forward Continuation of Backward Simulation for Poisson processes. This enabled the continuation of correlated multivariate Poisson processes simulated using Backward Simulation

within $[0, T]$ to simulation intervals $[mT, (m+1)T]$ for any non-negative integer m. This was published in [17] for the Poisson case. (Section 3.3.)

8. Showed that the correlation structure of a correlated multivariate Poisson process computed by Forward Continuation of Backward Simulation attains asymptotic stationarity. (Section 3.3.)

9. Showed formally that the extreme positive correlations attained under Forward Simulation in the bivariate setting is constant and determined by the ratio of the intensities of the bivariate Poisson process. (Section 3.4.)

10. Extended the Backward Simulation approach to mixed Poisson processes (Section 4.2) and showed that the correlation structure of a multivariate mixed Poisson process constructed by Backward Simulation depends on the correlation at terminal time adjusted by a correction factor (Section 4.3).

11. Extended the Forward-Backward approach to mixed Poisson processes and showed that a multivariate mixed Poisson process constructed by the Forward-Backward approach attains asymptotic stationarity (Section 4.4).

12. Extended the Backward Simulation approach to compound Poisson processes (Section 5.2) and showed that the correlation structure of a multivariate compound Poisson process constructed by Backward Simulation is a linear function of the correlation at time T, similar to the Poisson case (Section 5.3).

13. Extended the Forward-Backward approach to compound Poisson processes and showed that a multivariate compound Poisson process constructed by the Forward-Backward approach attains asymptotic stationarity (Section 5.4).

6.2 Future Work

The problem studied in this thesis touches many fields of applied probability. We list some directions for future work stemming from this thesis.

- Extending the Backward Simulation methodology to compound mixed-Poisson processes.
- A natural extension of our work that is worth investigating is whether Backward Simulation can be applied to inhomogeneous Poisson processes[1]. Upon preliminary investigation, inhomogeneous Poisson processes do satisfy an order statistic property [76]. Thus, it is very likely that Backward Simulation can be applied to multivariate inhomogeneous Poisson processes. Investigating their correlation structures should also be fruitful. We hope to investigate this in our future work.
- Antithetic variates, a popular variance reduction technique, is similar, in spirit, to the EJD approach. Both seek to generate objects with extreme dependence. It is worth studying whether the EJD method of generating joint distributions can be used as a variance reduction technique.
- Studying whether or not the EJD algorithm can be extended to the continuous case.
- Section 2.8.1 proposes one approach to solving the calibration problem (2.105). It may be worthwhile investigating other approaches.

[1] This was also asked by some individuals from industry expressing interest in our work.

- The optimization formulation of our problem, given by equations (2.3)-(2.3e) in the bivariate case and equations (2.37)-(2.39) in the general d-dimensional case, is broadly known as the "marginal problem" or the problem of finding "distributions with given marginals and correlations" in the literature. There has been much previous work in the dual problem associated with the optimization problems mentioned above [72]. Dual problems often reveal the sensitivities of a problem and allow for more efficient numerical methods. It may be worthwhile to see if there is a dual version of the EJD algorithm based on the dual problems to (2.3)-(2.3e) or (2.37)-(2.39).

- Quasi Monte Carlo (QMC) has been applied to both the simulation of Poisson processes [45] and in Backward Simulation methodologies [13], where it has been shown to be superior to crude Monte Carlo, especially for problems in Finance [127]. It is worthwhile exploring how QMC can be applied to the Backward Simulation of correlated multivariate Poisson and mixed Poisson processes.

Bibliography

[1] Operational Riskdata eXchange Association ORX. https://managingrisktogether.orx.org/orx-membership/loss-data.

[2] In J. Branke, K. Deb, K. Miettinen, and R. Slowiński, editors, *Multiobjective optimization: Interactive and evolutionary approaches*, volume 5252 of *Lecture Notes in Computer Science*. Springer Science & Business Media, 2008.

[3] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja. *A First Course in Order Statistics*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1992.

[4] F. Aue and M. Kalkbrener. LDA at work: Deutsche Bank's approach to quantifying operational risk. *Journal of Operational Risk*, 1(4):49–93, 2006.

[5] A. N. Avramidis, P. L'Ecuyer, and P.-A. Tremblay. Efficient simulation of gamma and variance-gamma processes. In S. Chick, P.J. Sánchez, D. Ferrin, and D.J. Morrice, editors, *Proceedings of the 2003 Winter Simulation Conference*, pages 319–326, 2003.

[6] A. N. Avramidis and P. L'Ecuyer. Efficient Monte Carlo and quasi–Monte Carlo option pricing under the variance gamma model. *Management Science*, 52(12):1930–1944, 2006.

[7] T. Bae and A. Kreinin. A backward construction and simulation of correlated Poisson processes. *Journal of Statistical Computation and Simulation*, 87(8):1593–1607, 2017.

[8] T. Bae and M. Mazjini. Backward simulation of correlated negative binomial Lévy processes. *Mathematics and Statistics*, 7(5):191–196, 2019.

[9] A. Beskos, G. Roberts, A. Stuart, and J. Voss. MCMC methods for diffusion bridges. *Stochastics and Dynamics*, 8(03):319–350, 2008.

[10] M. Bladt, S. Finch, and M. Sørensen. Simulation of multivariate diffusion bridges. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(2):343–369, 2016.

[11] M. Bladt and M. Sørensen. Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli*, 20(2):645–675, 2014.

[12] G. Bormetti, G. Callegaro, G. Livieri, and A. Pallavicini. A backward Monte Carlo approach to exotic option pricing. *European Journal of Applied Mathematics*, 29(1):146–187, 2018.

[13] R. E. Caflisch, W. Morokoff, and A. B. Owen. Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension. *Journal of Computational Finance*, 1(1):27–46, 1997.

[14] M. C. Cario and B. L. Nelson. Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix. Technical report, Department of Industrial Engineering and Management Sciences, Northwester University, Illinois, 1997.

[15] S. Chakraborty. Generating discrete analogues of continuous probability distributions-A survey of methods and constructions. *Journal of Statistical Distributions and Applications*, 2(6):1–30, 2015.

[16] M. Chiu. Implementation of the extreme joint distribution algorithm in C++. https://github.com/chiumichael/ejd, 2019.

[17] M. Chiu, K. R. Jackson, and A. Kreinin. Correlated multivariate Poisson processes and extreme measures. *Model Assisted Statistics and Applications*, 12(4):369–385, 2017.

[18] M. Chiu, K. R. Jackson, and A. Kreinin. Backward simulation of multivariate mixed Poisson processes. *Journal of Statistical Computation and Simulation*, pages 3549–3572, 2021.

[19] R. Cont and P. Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC Press, 2004.

[20] S. Corlay. Properties of the Ornstein-Uhlenbeck bridge. *arXiv preprint arXiv:1310.5617*, 2013.

[21] M. Cruz, G. W. Peters, and P. V. Shevchenko. *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. Wiley Handbooks in Financial Engineering and Econometrics. John Wiley & Sons, New Jersey, 2015.

[22] S. Dachian and I. Negri. On compound Poisson processes arising in change-point type statistical models as limiting likelihood ratios. *Statistical Inference for Stochastic Processes*, 14(3):255–271, 2011.

[23] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume I: Elementary Theory and Methods*. Probability and Its Applications. Springer Science & Business Media, New York, 2003.

[24] D. J. Daley and D. Vere-Jones. *An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure*. Probability and Its Applications. Springer Science & Business Media, New York, 2007.

[25] M. Davis and M. Pistorius. Quantification of counterparty risk via Bessel bridges. Working paper, Department of Mathematics, Imperial College, London, 2010.

[26] A. Deffner and E. Haeusler. A characterization of order statistic point processes that are mixed Poisson processes and mixed sample processes simultaneously. *Journal of Applied Probability*, 22(2):314–323, 1985.

[27] L. Devroye. *Non-Uniform Random Variate Generation*. Springer, 1986.

[28] D. K. Dey and Y. Chung. Compound Poisson distributions: Properties and estimation. *Communications in Statistics-Theory and Methods*, 21(11):3097–3121, 1992.

[29] J. Dhaene and M. Denuit. The safest dependence structure among risks. *Insurance: Mathematics and Economics*, 25(1):11–21, 1999.

[30] J. Dhaene, M. Denuit, M. J. Goovaerts, R. Kaas, and D. Vyncke. The concept of comonotonicity in actuarial science and finance: theory. *Insurance: Mathematics and Economics*, 31(1):3–33, 2002.

[31] Z. Drezner and D. Zerom. A simple and effective discretization of a continuous random variable. *Communications in Statistics-Simulation and Computation*, 45(10):3798–3810, 2016.

[32] K. Duch, Y. Jiang, and A. Kreinin. New approaches to operational risk modeling. *IBM Journal of Research and Development*, 58(4):3–1, 2014.

[33] F. Durante and J. F. Sánchez. On the approximation of copulas via shuffles of Min. *Statistics & Probability Letters*, 82(10):1761–1767, 2012.

[34] M. Dwass and H. Teicher. On infinitely divisible random vectors. *The Annals of Mathematical Statistics*, 28(2):461–470, 1957.

[35] H. G. Eggleston. *Convexity*. Cambridge Tracts in Mathematics. Cambridge University Press, 1958.

[36] P. Embrechts, R. Frey, and A. McNeil. *Quantitative Risk Management: Concepts, Techniques, and Tools*. Princeton University Press, 2015.

[37] P. Embrechts, H. Furrer, and R. Kaufmann. Quantifying regulatory capital for operational risk. *Derivatives Use, Trading and Regulation*, 9(3):217–233, 2003.

[38] P. Embrechts and M. Hofert. A note on generalized inverses. *Mathematical Methods of Operations Research*, 77(3):423–432, 2013.

[39] P. D. Feigin. On the characterization of point processes with the order statistic property. *Journal of Applied Probability*, 16(2):297–304, 1979.

[40] W. Feller. *An Introduction to Probability Theory and its Applications, vol 1 & 2*. John Wiley & Sons, 2008.

[41] A. Ferrari, G. Letac, and J.-Y. Tourneret. Multivariate mixed Poisson distributions. In *2004 12th European Signal Processing Conference*, pages 1067–1070. IEEE, 2004.

[42] G. S. Fishman. Variance reduction in simulation studies. *Journal of Statistical Computation and Simulation*, 1(2):173–182, 1972.

[43] G. S. Fishman and L. R. Moore III. Sampling from a discrete distribution while preserving monotonicity. *The American Statistician*, 38(3):219–223, 1984.

[44] T. R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2013.

[45] B. L. Fox. Generating Poisson processes by quasi-Monte Carlo. Technical report, Boulder CO: SIM-OPT Consulting, 1996.

[46] B. L. Fox and P. W. Glynn. Computing Poisson probabilities. *Communications of the ACM*, 31(4):440–445, 1988.

[47] A. Frachot and T. Roncalli. Mixing internal and external data for managing operational risk. Working paper, Group de Recherche Opérationnelle, Crédit Lyonnais, France, 2002.

[48] A. Frachot, T. Roncalli, and E. Salomon. The correlation problem in operational risk. MPRA Paper 28052, University Library of Munich, Germany, 2004.

[49] M. Fréchet. Sur les tableaux dont les marges et des bornes sont données. *Revue de l'Institut international de statistique*, A. 14:53–77, 1951.

[50] M. C. Fu. Variance-gamma and Monte Carlo. In M. C. Fu, R. A. Jarrow, J.-Y. J. Yen, and R. J. Elliot, editors, *Advances in Mathematical Finance*, pages 21–34. Springer, 2007.

[51] R. G. Gallager. *Stochastic Processes: Theory for Applications*. Cambridge University Press, 2013.

[52] X. Gao and Y. Wang. Optimal market making in the presence of latency. *Quantitative Finance*, 20(9):1495–1512, 2020.

[53] S. Ghosh and S. G. Henderson. Chessboard distributions and random vectors with specified marginals and covariance matrix. *Operations Research*, 50(5):820–834, 2002.

[54] S. Ghosh and S. G. Henderson. Properties of the NORTA method in higher dimensions. In E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, editors, *Proceedings of the 2002 Winter Simulation Conference*, 2002.

[55] S. Ghosh and S. G. Henderson. Behavior of the NORTA method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation*, 13(3):1–19, 2003.

[56] P. Glasserman and J. Li. Importance sampling for a mixed Poisson model of portfolio credit risk. In S. Chick, P.J. Sánchez, D. Ferrin, and D.J. Morrice, editors, *Proceedings of the 2003 Winter Simulation Conference*, pages 267–275, 2003.

[57] P. Glasserman and D. D. Yao. Optimal couplings are totally positive and more. *Journal of Applied Probability*, 41:321–332, 2004.

[58] J. Grandell. *Mixed Poisson Processes*. CRC Press, 1997.

[59] R. C. Griffiths, R. K. Milne, and R. Wood. Aspects of correlation in bivariate Poisson distributions and processes. *Australian Journal of Statistics*, 21(3):238–255, 1979.

[60] J. M. Hammersley and J. G. Mauldon. General principles of antithetic variates. *Mathematical proceedings of the Cambridge Philosophical Society*, 52(3):476–481, 1956.

[61] G. Hardy, J. E. Littlewood, and G. Pólya. *Inequalities*. Cambridge University Press, 1952.

[62] Y. Hayakawa and G. Telfar. Mixed Poisson-type processes with application in software reliability. *Mathematical and Computer Modelling*, 31(10-12):151–156, 2000.

[63] R. R. Hill and C. H. Reilly. Composition for multivariate random variables. In J.D. Tew, S. Manivannan, D.A. Sadowski, and A. F. Seila, editors, *Proceedings of the 1994 Winter Simulation Conference*, pages 332–339, 1994.

[64] W. Hoeffding. Masstabinvariante korrelations-theorie. *Schriften Math. Inst. Univ. Berlin.*, 2:181–233, 1940.

[65] W. Hu and J. Zhou. Backward simulation methods for pricing American options under the CIR process. *Quantitative Finance*, 17(11):1683–1695, 2017.

[66] J. Janssen and R. Manca. *Applied Semi-Markov Processes*. Springer Science & Business Media, 2006.

[67] M. E. Johnson. *Multivariate Statistical Simulation: A Guide to Selecting and Generating Continuous Multivariate Distributions*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 2013.

[68] N. L. Johnson, A. W. Kemp, and S. Kotz. *Univariate Discrete Distributions*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2005.

[69] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Discrete Multivariate Distributions*. Wiley Series in Probability and Statistics. John Wiley & Sons, 1996.

[70] D. Karlis and L. Meligkotsidou. Finite mixtures of multivariate Poisson distributions with application. *Journal of Statistical Planning and Inference*, 137(6):1942–1960, 2007.

[71] A. Karr. *Point Processes and Their Statistical Inference*. Routledge, 2017.

[72] H. G. Kellerer. Duality theorems for marginal problems. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 67(4):399–432, 1984.

[73] A. Kolesárová, R. Mesiar, and C. Sempi. Measure-preserving transformations, copulæ and compatibility. *Mediterranean Journal of Mathematics*, 5(3):325–339, 2008.

[74] A. Kreinin. Correlated Poisson processes and their applications in financial modeling. In A. N Akansu, S. R Kulkarni, and D. M Malioutov, editors, *Financial Signal Processing and Machine Learning*, chapter 9, pages 191–230. John Wiley & Sons, 2016.

[75] G. Last. Stochastic analysis for Poisson processes. In *Stochastic analysis for Poisson point processes*, pages 1–36. Springer, 2016.

[76] P. A. W. Lewis and G. S. Shedler. Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413, 1979.

[77] C.-K. Li and B.-S. Tam. A note on extreme correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, 15(3):903–908, 1994.

[78] S. T. Li and J. L. Hammond. Generation of pseudorandom numbers with specified univariate distributions and correlation coefficients. *IEEE Transactions on Systems, Man, and Cybernetics*, (5):557–561, 1975.

[79] U. Liberman. An order statistic characterization of the Poisson renewal process. *Journal of Applied Probability*, 2(3):717–722, 1985.

[80] J. Lin and X. Wang. New Brownian bridge construction in quasi-Monte Carlo methods for computational finance. *Journal of Complexity*, 24(2):109–133, 2008.

[81] M. Lin, R. Chen, and P. Mykland. On generating Monte Carlo samples of continuous diffusion bridges. *Journal of the American Statistical Association*, 105(490):820–838, 2010.

[82] F. Lindskog and A. J. McNeil. Common Poisson shock models: Applications to insurance and credit risk modelling. *ASTIN Bulletin: The Journal of the IAA*, 33(2):209–238, 2003.

[83] F. Lindsten and T. B. Schön. Backward simulation methods for Monte Carlo statistical inference. *Foundations and Trends in Machine Learning*, 6(1):1–143, 2013.

[84] O. Lundberg. *On Random Processes and their Application to Sickness and Accident Statistics*. PhD thesis, Almqvist & Wiksell, Uppsala., 1964.

[85] P. M. Lurie and M. S. Goldberg. An approximate method for sampling correlated random variables from partially-specified distributions. *Management Science*, 44(2):149–283, 1998.

[86] Z. MacDonald. A modified simplex method for solving Ax = b, $x \geq 0$, for very large A arising from a calibration problem. MSc, University of Toronto, 2020.

[87] R. N. Makarov and D. Glew. Exact simulation of Bessel diffusions. *Monte Carlo Methods and Applications*, 16(3-4):283–306, 2010.

[88] R. J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Mathematical Journal*, 80(2):309–324, 1995.

[89] J. A. McFadden. The mixed Poisson process. *Sankhyā: The Indian Journal of Statistics, Series A*, 25(1):83–92, 1965.

[90] G. J. Minty. On the monotonicity of the gradient of a convex function. *Pacific Journal of Mathematics*, 14(1):243–247, 1964.

[91] M. Moscadelli. The modelling of operational risk: experience with the analysis of the data collected by the Basel committee. *Available at SSRN 557214*, 2004.

[92] L. Nenna. *Numerical Methods for Multi-Marginal Optimal Transportation*. PhD thesis, PSL Research University, 2016.

[93] G. Nocedal and S. Wright. *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer, 2006.

[94] Basel Committee on Banking Supervision. Basel II: International convergence of capital measurement and capital standards: A revised framework - comprehensive version. https://www.bis.org/publ/bcbs128.pdf, June 2006.

[95] E. S. Page. On Monte Carlo methods in congestion problems: II. simulation of queuing systems. *Operations Research*, 13(2):300–305, 1965.

[96] H. H. Panjer. *Operational Risk: Modeling Analytics*, volume 620 of *Wiley Series in Probability and Statistics*. John Wiley & Sons, 2006.

[97] B. Pass. Multi-marginal optimal transport: Theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1771–1790, 2015.

[98] R. Pasupathy. Generating homogeneous Poisson processes. In J. J. Cochran, L. A. Cox, P. Keskinocak, J.P. Kharoufeh, and J. C. Smith, editors, *Wiley Encyclopedia of Operations Research and Management Science*, pages 1–11. John Wiley & Sons, 2011.

[99] H. Penikas. History of the Basel internal-ratings-based (irb) credit risk regulation. *Model Assisted Statistics and Applications*, 15(1):81–98, 2020.

[100] V. V. Petrov. *Sums of Independent Random Variables*. De Gruyter, 1975.

[101] J. Pitman and M. Yor. A decomposition of Bessel bridges. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 59(4):425–457, 1982.

[102] M. Powojowski, D. Reynolds, and H. Tuenter. Dependent events and operational risk. *Algo Research Quarterly*, 5(2):65–73, 2002.

[103] G. Puccetti and R. Wang. Extremal dependence concepts. *Statistical Science*, 30(4):485–517, 2015.

[104] P. S. Puri. On the characterization of point processes with the order statistic property without the moment condition. *Journal of Applied Probability*, 19(1):39–51, 1982.

[105] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume I: Theory*. Probability and its Applications. Springer, 1998.

[106] S. T. Rachev and L. Rüschendorf. *Mass Transportation Problems: Volume II: Applications*. Probability and its Applications. Springer, 1998.

[107] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der mathematischen Wissenchaften*. Springer, 2013.

[108] C. Ribeiro and N. Webber. Valuing path-dependent options in the variance-gamma model by Monte Carlo with a gamma bridge. *Journal of Computational Finance*, 7(2):81–100, 2004.

[109] L. Rüschendorf. Construction of multivariate distributions with given marginals. *Annals of the Institute of Statistical Mathematics*, 37:225–233, 1985.

[110] P. Sabino. Forward or backward simulation? a comparative study. *Quantitative Finance*, 20(7):1213–1226, 2020.

[111] S. Santini. Analysis of traffic flow in urban areas using web cameras. In *Proceedings Fifth IEEE Workshop on Applications of Computer Vision*, pages 140–145. IEEE, 2000.

[112] Juliana Schulz and Christian Genest. Comonotonicity and counter-monotonicity: Review and implications for likelihood-based estimation. *Communications in Statistics-Theory and Methods*, 54(8):2482–2505, 2025.

[113] D. W. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2015.

[114] T. C. Sharkey. Infinite linear programs. In J. J. Cochran, L. A. Cox, P. Keskinocak, J.P. Kharoufeh, and J. C. Smith, editors, *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, 2010.

[115] Z. Shen and C. Weng. A backward simulation method for stochastic optimal control problems. *Available at SSRN 3319160*, 2019.

[116] P. V. Shevchenko. *Modelling Operational Risk Using Bayesian Inference*. Springer, 2011.

[117] P. Shi and E. A. Valdez. Multivariate negative binomial models for insurance claim counts. *Insurance: Mathematics and Economics*, 55:18–29, 2014.

[118] D. Z. Stein and J. M. Juritz. Bivariate compound Poisson distributions. *Communications in Statistics-Theory and Methods*, 16(12):3591–3607, 1987.

[119] V. Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

[120] H. Teicher. On the multivariate Poisson distribution. *Scandinavian Actuarial Journal*, 1:1–9, 1954.

[121] E.-M. Tiit. Random vectors with given arbitrary marginal and given correlation matrix. *Acta et Commentationes Universitatis Tartuensis*, 733:14–39, 1986.

[122] E.-M. Tiit. Mixture of extremal distributions - a model alternative to the log-linear model. *Prob. Theory and Math. Stat*, pages 703–714, 1994.

[123] E.-M. Tiit. Mixtures of multivariate quasi-extremal distributions having given marginals. In *Distributions with Fixed Marginals and Related Topics*, IMS Lecture Notes - Monograph Series Vol. 28, pages 337–357. Institute of Mathematical Statistics, 1996.

[124] E.-M. Tiit. Existence of multivariate distributions with given marginals. In *Distributions With Given Marginals and Statistical Modelling*, pages 229–241. Springer, 2002.

[125] C. Villani. *Optimal Transport: Old and New*, volume 338 of *Grundlehren der mathematischen Wissenschaften*. Springer, 2008.

[126] B. Wang and R. Wang. Joint mixability. *Mathematics of Operations Research*, 41(3):808–826, 2016.

[127] X. Wang and K.-T. Fang. The effective dimension and quasi-Monte Carlo integration. *Journal of Complexity*, 19(2):101–124, 2003.

[128] W. Whitt. Bivariate distributions with given marginals. *The Annals of Statistics*, 4(6):1280–1289, 1976.

[129] K. C. Yuen, J. Guo, and X. Wu. On a correlated aggregate claims model with Poisson and Erlang risk processes. *Insurance: Mathematics and Economics*, 31(2):205–214, 2002.

[130] M. Zocher. *Multivariate Mixed Poisson Processes*. PhD thesis, Technische Universitat Dresden, 2005.

Appendix A

Chapter 2.3 Proofs

A.1 Proof of Theorem 3

PROOF: We prove Theorem 3 by induction on l.

We use $\mathbb{S}_l^{(i)}$ to stand for the i^{th} statement below and refer to all three statements together as $\mathbb{S}_l$ for the l^{th} iteration of Algorithm 1.

$\mathbb{S}_l^{(1)}$: The point $s_l = (s_l^{(1)}, s_l^{(2)})$ computed by Algorithm 1 is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$.

$\mathbb{S}_l^{(2)}$: The value z_l computed by Algorithm 1 satisfies

$$\begin{aligned} F^{(1)}_{s_l^{(1)}-1} \le z_{l-1} < z_l \le F^{(1)}_{s_l^{(1)}}, \\ F^{(2)}_{s_l^{(2)}-1} \le z_{l-1} < z_l \le F^{(2)}_{s_l^{(2)}}. \end{aligned} \tag{A.1}$$

In addition, at least one of

$$\begin{aligned} F^{(1)}_{s_l^{(1)}-1} = z_{l-1}, \\ F^{(2)}_{s_l^{(2)}-1} = z_{l-1} \end{aligned} \tag{A.2}$$

holds and at least one of

$$\begin{aligned} z_l = F^{(1)}_{s_l^{(1)}}, \\ z_l = F^{(2)}_{s_l^{(2)}} \end{aligned} \tag{A.3}$$

holds. Moreover, z_l computed by Algorithm 1 is the $(l+1)^{\text{st}}$ smallest element in $\Pi_Z = \Pi_{X^{(1)}} \vee \Pi_{X^{(2)}}$.

$\mathbb{S}_l^{(3)}$: The probability $\hat{P}^{(1)}_{s_l^{(1)}, s_l^{(2)}}$ computed by Algorithm 1 is correct in the sense that it agrees with (2.7).

We begin by using induction on l to prove that $\mathbb{S}_l$ holds for $l = 0, 1, 2, \ldots, l_{\max}$.

For the base case of the induction we show that $\mathbb{S}_0$ holds true. To this end, note that Line 3 of Algorithm 1 gives

$$s_0 = (0, 0).$$

From Lemma 1, the origin is always a support point for bivariate comonotonic distributions. Therefore, $s_0 = (0, 0)$ is the 1^{st} point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$. Thus, we have verified that $\mathbb{S}_0^{(1)}$ holds.

Next, we show that $\mathbb{S}_0^{(2)}$ holds. On Line 4, Algorithm 1 computes the partition of unity by calling Algorithm 8. Therefore, we have that

$$z_0 = \min\left(F_0^{(1)}, F_0^{(2)}\right).$$

From the discussion above, the fact that the cdfs are increasingly ordered (2.25) and $F_0^{(1)} > 0$ and $F_0^{(2)} > 0$, whence $z_0 > 0$, we have that

$$\begin{aligned} 0 = F_{-1}^{(1)} = z_{-1} < z_0 \leq F_0^{(1)}, \\ 0 = F_{-1}^{(2)} = z_{-1} < z_0 \leq F_0^{(2)}. \end{aligned}$$

Hence, (A.1) holds for $l = 0$. In addition, for $l = 0$, both equations in (A.2) hold and at least one of the equations in (A.3) holds. Moreover, from (2.25), it follows that the smallest element in $\Pi_Z = \Pi_{X^{(1)}} \vee \Pi_{X^{(2)}}$ is the smaller of $F_0^{(1)}$ and $F_0^{(2)}$. Hence, $z_0 = \min\left(F_0^{(1)}, F_0^{(2)}\right)$ computed by Algorithm 1 on Line 4 is the smallest element in Π_Z. That is, the value z_0 computed by Algorithm 1 is consistent with Definition 21. Therefore, we have verified that $\mathbb{S}_0^{(2)}$ holds.

Finally, we show that $\mathbb{S}_0^{(3)}$ holds. From Line 5 of Algorithm 1:

$$\hat{P}^{(1)}_{s_0^{(1)}, s_0^{(2)}} = z_0.$$

Moreover, we have from Line 4 of Algorithm 1 that

$$z_0 = \min\left(F_0^{(1)}, F_0^{(2)}\right).$$

Therefore, Algorithm 1 gives

$$\hat{P}^{(1)}_{s_0^{(1)}, s_0^{(2)}} = \min\left(F_0^{(1)}, F_0^{(2)}\right).$$

In addition, since $s_0 = (0, 0)$, $\min\left(F_0^{(1)}, F_0^{(2)}\right) > 0$, and $F_{-1}^{(1)} = F_{-1}^{(2)} = 0$, we see that

$$\begin{aligned} \hat{P}^{(1)}_{s_0^{(1)}, s_0^{(2)}} &= z_0 \\ &= \min(F_0^{(1)}, F_0^{(2)}) - 0 \\ &= \min(F_0^{(1)}, F_0^{(2)}) - \max(F_{-1}^{(1)}, F_{-1}^{(2)}) \\ &= \left[\min\left(F_0^{(1)}, F_0^{(2)}\right) - \max\left(F_{-1}^{(1)}, F_{-1}^{(2)}\right)\right]^+. \end{aligned}$$

Thus, the value of $\hat{P}_{0,0}^{(1)}$ computed on Line 5 of Algorithm 1 is consistent with the $\hat{P}_{0,0}^{(1)}$ given by (2.4). Therefore, we have verified that $\mathbb{S}_0^{(3)}$ holds.

Consequently, the base case of the induction proof must be true since we have verified that all three statements in $\mathbb{S}_0$ hold.

For the induction step, we choose any $l \geq 1$ and assume that all three statements in $\mathbb{S}_{l-1}$ hold. Our goal is to prove that all three statements in $\mathbb{S}_l$ hold.

We break the proof into three cases depending on whether

(1) $z_{l-1} = F^{(1)}_{s^{(1)}_{l-1}} = F^{(2)}_{s^{(2)}_{l-1}}$ (Line 8),

(2) $z_{l-1} = F^{(1)}_{s^{(1)}_{l-1}}$ and $z_{l-1} \neq F^{(2)}_{s^{(2)}_{l-1}}$ (Line 11),

(3) $z_{l-1} \neq F^{(1)}_{s^{(1)}_{l-1}}$ and $z_{l-1} = F^{(2)}_{s^{(2)}_{l-1}}$ (Line 14).

Note that exactly one of the cases above must be true, since from the construction of the partition of unity in Line 4 of Algorithm 1, we must have that

$$z_{l-1} = \min\left(F^{(1)}_{s^{(1)}_{l-1}}, F^{(2)}_{s^{(2)}_{l-1}}\right).$$

In what follows, it is helpful in the proof of the induction step to rewrite z_{l-1} using Lines 5 and 18 as follows

$$\begin{aligned} z_{l-1} &= \hat{P}^{(j)}_{s_{l-1}} + z_{l-2}, \\ z_{l-1} &= \hat{P}^{(j)}_{s_{l-1}} + \hat{P}^{(j)}_{s_{l-2}} + \cdots + \hat{P}^{(j)}_{s_0}. \end{aligned} \tag{A.4}$$

If we define the sets $\mathcal{A}(s_{l-1}) = \{(i,j) : 0 \leq i \leq s^{(1)}_{l-1},\ 0 \leq j \leq s^{(2)}_{l-1},\ \hat{P}^{(1)}_{i,j} > 0\}$ and $\mathcal{A}^c(s_{l-1}) = \{(i,j) : 0 \leq i \leq s^{(1)}_{l-1},\ 0 \leq j \leq s^{(1)}_{l-1},\ \hat{P}^{(1)}_{i,j} = 0\}$, then we can see that z_{l-1} can also be written in terms of the probabilities $\hat{P}^{(1)}_{i,j}$ associated with the points on the domain in the rectangle with one corner at the origin, $(0,0)$, and the other corner at the point $s_{l-1} = (s^{(1)}_{l-1}, s^{(2)}_{l-1})$

$$\begin{aligned} z_{l-1} &= \sum_{\mathcal{A}(s_{l-1})} \hat{P}^{(1)}_{i,j} + \sum_{\mathcal{A}^c(s_{l-1})} \hat{P}^{(1)}_{i,j} \\ &= \sum_{i=0}^{s^{(1)}_{l-1}} \sum_{j=0}^{s^{(2)}_{l-1}} \hat{P}^{(1)}_{i,j}. \end{aligned} \tag{A.5}$$

Case 1: $z_{l-1} = F^{(1)}_{s^{(1)}_{l-1}} = F^{(2)}_{s^{(2)}_{l-1}}$ (Line 8)

In this case, the condition in the if statement on Line 8 of Algorithm 1 is true and so Lines 9 and 10 of Algorithm 1 are executed. Hence, Algorithm 1 computes

$$\begin{aligned} s^{(1)}_l &= s^{(1)}_{l-1} + 1, \\ s^{(2)}_l &= s^{(2)}_{l-1} + 1. \end{aligned} \tag{A.6}$$

To verify that $\mathbb{S}^{(1)}_l$ holds, we need to show that the point s_l computed by Algorithm 1 on Lines 9 and 10 and specified above in (A.6) is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$.

To this end, note that the induction hypothesis $\mathbb{S}^{(1)}_{l-1}$ gives that the point s_{l-1} in (A.6) is the l^{th} point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$. So, if we can show that the point s_l specified in (A.6) is the subsequent point to s_{l-1}, then it follows immediately that the point s_l specified in (A.6) is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$ and we are done.

To see that the point s_l specified in (A.6) is the subsequent point to s_{l-1}, note that, from Definition 17 and Remark 12, one of the three points $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$, $(s_{l-1}^{(1)}, s_{l-1}^{(2)}+1)$ or $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)}+1)$ must be the subsequent point to s_{l-1}. So, if we can show that $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$ and $(s_{l-1}^{(1)}, s_{l-1}^{(2)}+1)$ are not support points of $\hat{P}^{(1)}$, then it follows that the point s_l specified in (A.6) is the subsequent point to s_{l-1} on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$.

We show by proof by contradiction that the subsequent point to s_{l-1} on the $\mathcal{S}$-path cannot be $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$. To that end, assume that the subsequent point to s_{l-1} is $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$, which implies that $\hat{P}^{(1)}_{s_{l-1}^{(1)}+1, s_{l-1}^{(2)}} > 0$. Now, since, in Case 1, $z_{l-1} = F^{(2)}_{s_{l-1}^{(2)}}$ we have from (A.5) that

$$\sum_{i=0}^{s_{l-1}^{(1)}} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} = F^{(2)}_{s_{l-1}^{(2)}}.$$

However, as noted above, $\hat{P}^{(1)}_{s_{l-1}^{(1)}+1, s_{l-1}^{(2)}} > 0$. Therefore, (2.22) and the discussion above leads to the contradiction

$$\begin{aligned}
F^{(2)}_{s_{l-1}^{(2)}} = \sum_{i=0}^{\infty} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} &= \sum_{i=0}^{s_{l-1}^{(1)}} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} + \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{s_{l-1}^{(1)}+1,j} + \sum_{i=s_{l-1}^{(1)}+2}^{\infty} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} \\
&\geq \sum_{i=0}^{s_{l-1}^{(1)}} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} + \hat{P}^{(1)}_{s_{l-1}^{(1)}+1, s_{l-1}^{(2)}} \\
&= F^{(2)}_{s_{l-1}^{(2)}} + \hat{P}^{(1)}_{s_{l-1}^{(1)}+1, s_{l-1}^{(2)}} \\
&> F^{(2)}_{s_{l-1}^{(2)}}.
\end{aligned}$$

Hence, the subsequent point to s_{l-1} cannot be $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$.

A similar argument shows that the subsequent point to s_{l-1} cannot be $(s_{l-1}^{(1)}, s_{l-1}^{(2)}+1)$. Therefore, by Lemma 2, the subsequent point must be $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)}+1)$. Hence, the point s_l computed by Algorithm 1 on Lines 9 and 10 and specified above in (A.6) is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$. Therefore, we have verified that $\mathbb{S}_l^{(1)}$ holds.

Next, we verify that $\mathbb{S}_l^{(2)}$ holds. To this end, recall that, in Case 1, $z_{l-1} = F^{(1)}_{s_{l-1}^{(1)}}$ and $z_{l-1} = F^{(2)}_{s_{l-1}^{(2)}}$. Also, in Case 1, Lines 9 and 10 of Algorithm 1 are executed. Hence, s_l satisfies

$$\begin{aligned}
s_l^{(1)} &= s_{l-1}^{(1)} + 1, \\
s_l^{(2)} &= s_{l-1}^{(2)} + 1.
\end{aligned}$$

Therefore,

$$F^{(1)}_{s_l^{(1)}-1} = F^{(1)}_{s_{l-1}^{(1)}} = z_{l-1} = F^{(2)}_{s_{l-1}^{(2)}} = F^{(2)}_{s_l^{(2)}-1},$$

whence

$$\begin{aligned}
F^{(1)}_{s_l^{(1)}-1} &= z_{l-1}, \\
F^{(2)}_{s_l^{(2)}-1} &= z_{l-1}.
\end{aligned} \tag{A.7}$$

Line 4 of Algorithm 1 calls Algorithm 8 to compute the partition of unity. Therefore,

$$z_l = \min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right).$$

Since

$$z_{l-1} = F^{(1)}_{s^{(1)}_{l-1}} = F^{(2)}_{s^{(2)}_{l-1}}$$

and

$$\begin{aligned} F^{(1)}_{s^{(1)}_{l-1}} &= F^{(1)}_{s^{(1)}_l - 1} < F^{(1)}_{s^{(1)}_l}, \\ F^{(2)}_{s^{(2)}_{l-1}} &= F^{(2)}_{s^{(2)}_l - 1} < F^{(2)}_{s^{(2)}_l} \end{aligned}$$

since $F^{(1)}_i$ and $F^{(2)}_j$ are strictly increasing (see (2.25)), it follows that

$$z_{l-1} = F^{(1)}_{s^{(1)}_{l-1}} = F^{(2)}_{s^{(2)}_{l-1}} < \min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right) = z_l. \tag{A.8}$$

Putting (A.7) and (A.8) together with

$$\begin{aligned} z_l &= \min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right) \le F^{(1)}_{s^{(1)}_l}, \\ z_l &= \min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right) \le F^{(2)}_{s^{(2)}_l} \end{aligned}$$

we get

$$\begin{aligned} F^{(1)}_{s^{(1)}_l - 1} &= z_{l-1} < z_l \le F^{(1)}_{s^{(1)}_l}, \\ F^{(2)}_{s^{(2)}_l - 1} &= z_{l-1} < z_l \le F^{(2)}_{s^{(2)}_l}. \end{aligned} \tag{A.9}$$

That is, we have shown that (A.1) holds. Moreover, from (A.9), we see that both the equations in (A.2) hold. In addition, since $z_l = \min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right)$, at least one of the equations in (A.3) must hold. Furthermore, since we know from the induction hypothesis $\mathbb{S}^{(2)}_{l-1}$ that z_{l-1} is the l^{th} smallest element in $\Pi_Z = \Pi_{X^{(1)}} \vee \Pi_{X^{(2)}}$, it follows from (A.9) that the smaller of $F^{(1)}_{s^{(1)}_l}$ and $F^{(2)}_{s^{(2)}_l}$ must be the $(l+1)^{\text{st}}$ smallest element in Π_Z. Hence, $z_l = \min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right)$ is the $(l+1)^{\text{st}}$ smallest element in Π_Z. That is, the value z_l computed by Algorithm 1 is consistent with Definition 21. Therefore, we have verified that $\mathbb{S}^{(2)}_l$ holds.

Finally, we verify that $\mathbb{S}^{(3)}_l$ holds. The value of $\hat{P}^{(1)}_{s^{(1)}_l, s^{(2)}_l}$ computed on Line 18 of Algorithm 1 is

$$\hat{P}^{(1)}_{s^{(1)}_l, s^{(2)}_l} = z_l - z_{l-1}. \tag{A.10}$$

We know from Line 4 of Algorithm 1 which calls Algorithm 8 to compute the partition of unity that

$$z_l = \min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right). \tag{A.11}$$

Substituting $l-1$ for l in (A.11), we see that, if $l \geq 2$, on the previous iteration

$$z_{l-1} = \min\left(F^{(1)}_{s^{(1)}_{l-1}}, F^{(2)}_{s^{(2)}_{l-1}}\right). \tag{A.12}$$

If $l = 1$, we see that Line 4 of Algorithm 1 also implies that (A.12) holds. Now, recall again that, in Case 1, $z_{l-1} = F^{(1)}_{s^{(1)}_{l-1}}$ and $z_{l-1} = F^{(2)}_{s^{(2)}_{l-1}}$, whence $F^{(1)}_{s^{(1)}_{l-1}} = F^{(2)}_{s^{(2)}_{l-1}}$. Therefore,

$$\begin{aligned} z_{l-1} &= \min\left(F^{(1)}_{s^{(1)}_{l-1}}, F^{(2)}_{s^{(2)}_{l-1}}\right) \\ &= \max\left(F^{(1)}_{s^{(1)}_{l-1}}, F^{(2)}_{s^{(2)}_{l-1}}\right). \end{aligned} \tag{A.13}$$

From (A.8), it follows that

$$z_l - z_{l-1} > 0. \tag{A.14}$$

Combining (A.10), (A.11), (A.13) and (A.14) together with

$$\begin{aligned} s^{(1)}_l &= s^{(1)}_{l-1} + 1, \\ s^{(2)}_l &= s^{(2)}_{l-1} + 1, \end{aligned}$$

which follows from Lines 9 and 10 of Algorithm 1 in Case 1, we get

$$\begin{aligned} \hat{P}^{(1)}_{s^{(1)}_l, s^{(2)}_l} &= z_l - z_{l-1} \\ &= \min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right) - \max\left(F^{(1)}_{s^{(1)}_{l-1}}, F^{(2)}_{s^{(2)}_{l-1}}\right) \\ &= \min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right) - \max\left(F^{(1)}_{s^{(1)}_l - 1}, F^{(2)}_{s^{(2)}_l - 1}\right) \\ &= \left[\min\left(F^{(1)}_{s^{(1)}_l}, F^{(2)}_{s^{(2)}_l}\right) - \max\left(F^{(1)}_{s^{(1)}_l - 1}, F^{(2)}_{s^{(2)}_l - 1}\right)\right]^+. \end{aligned}$$

Therefore, the probability $\hat{P}^{(1)}_{s^{(1)}_l, s^{(2)}_l}$ computed on Line 18 of Algorithm 1 is correct in the sense that it agrees with (2.7) and we have verified that $\mathbb{S}^{(3)}_l$ holds.

Thus, we have shown that all three statements in $\mathbb{S}_l$ hold in Case 1.

Case 2: $z_{l-1} = F^{(1)}_{s^{(1)}_{l-1}}$ and $z_{l-1} \neq F^{(2)}_{s^{(2)}_{l-1}}$ (Line 11)

In this case, the condition in the if statement in Line 11 is true and so Lines 12 and 13 of Algorithm 1 are executed. Hence, Algorithm 1 computes

$$\begin{aligned} s^{(1)}_l &= s^{(1)}_{l-1} + 1, \\ s^{(2)}_l &= s^{(2)}_{l-1}. \end{aligned} \tag{A.15}$$

We begin by showing that $\mathbb{S}^{(1)}_l$ is true. Thus, our goal is to show that the s_l specified in (A.15) is the subsequent point of s_{l-1} on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$. It follows from this that s_l is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path, since, by the induction hypothesis, $\mathbb{S}^{(1)}_{l-1}$, s_{l-1} is the l^{th} point on the $\mathcal{S}$-path.

To see that the s_l specified in (A.15) is the subsequent point to s_{l-1} on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$, first note that, by Lemma 2, the subsequent point to s_{l-1} must be one of $(s^{(1)}_{l-1}+1, s^{(2)}_{l-1})$, $(s^{(1)}_{l-1}, s^{(2)}_{l-1}+$

1) or $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)}+1)$. We show below that the subsequent point cannot be either $(s_{l-1}^{(1)}, s_{l-1}^{(2)}+1)$ or $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)}+1)$. Therefore, the subsequent point must be $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$ computed by Algorithm 1 in Lines 12 and 13.

We first show by proof by contradiction that the subsequent point to s_{l-1} cannot be $(s_{l-1}^{(1)}, s_{l-1}^{(2)}+1)$. To that end, suppose that the subsequent point of s_{l-1} is $(s_{l-1}^{(1)}, s_{l-1}^{(2)}+1)$, then, $\hat{P}^{(1)}_{s_{l-1}^{(1)}, s_{l-1}^{(2)}+1} > 0$. Since, in Case 2, $z_{l-1} = F^{(1)}_{s_{l-1}^{(1)}}$, this together with (A.5) implies that:

$$\sum_{i=0}^{s_{l-1}^{(1)}} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} = z_{l-1} = F^{(1)}_{s_{l-1}^{(1)}}. \tag{A.16}$$

However, as noted above, $\hat{P}^{(1)}_{s_{l-1}^{(1)}, s_{l-1}^{(2)}+1} > 0$. Therefore, (2.21) and the discussion above leads to the contradiction

$$\begin{aligned} F^{(1)}_{s_{l-1}^{(1)}} = \sum_{i=0}^{s_{l-1}^{(1)}} \sum_{j=0}^{\infty} \hat{P}^{(1)}_{i,j} &= \sum_{i=0}^{s_{l-1}^{(1)}} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} + \sum_{i=0}^{s_{l-1}^{(1)}} \hat{P}^{(1)}_{i, s_{l-1}^{(2)}+1} + \sum_{i=0}^{s_{l-1}^{(1)}} \sum_{j=s_{l-1}^{(2)}+2}^{\infty} \hat{P}^{(1)}_{i,j} \\ &\geq \sum_{i=0}^{s_{l-1}^{(1)}} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} + \hat{P}^{(1)}_{s_{l-1}^{(1)}, s_{l-1}^{(2)}+1} \\ &= F^{(1)}_{s_{l-1}^{(1)}} + \hat{P}^{(1)}_{s_{l-1}^{(1)}, s_{l-1}^{(2)}+1} \\ &> F^{(1)}_{s_{l-1}^{(1)}}. \end{aligned} \tag{A.17}$$

Therefore, the subsequent point of s_{l-1} on the $\mathcal{S}$-path cannot be $(s_{l-1}^{(1)}, s_{l-1}^{(2)}+1)$.

Next, we show, again by proof by contradiction, that the subsequent point of s_{l-1} on the $\mathcal{S}$-path cannot be $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)}+1)$. To that end, suppose that the subsequent point of s_{l-1} on the $\mathcal{S}$-path is $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)}+1)$, which implies that $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$ cannot also be on the $\mathcal{S}$-path, since if both $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)}+1)$ and $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$ are on the $\mathcal{S}$-path then, by Definition 17, $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$ is the subsequent point of $(s_{l-1}^{(1)}, s_{l-1}^{(2)})$. Moreover, since $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$ is not on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$, by Lemma 4, $(s_{l-1}^{(1)}+1, s_{l-1}^{(2)})$ is not a support point of $\hat{P}^{(1)}$, whence $\hat{P}^{(1)}_{s_{l-1}^{(1)}+1, s_{l-1}^{(2)}} = 0$. Now note that (2.22) and (A.16) imply that

$$\begin{aligned} F^{(2)}_{s_{l-1}^{(2)}} = \sum_{i=0}^{\infty} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} &= \sum_{i=0}^{s_{l-1}^{(1)}} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} + \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{s_{l-1}^{(1)}+1, j} + \sum_{i=s_{l-1}^{(1)}+2}^{\infty} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j} \\ &= F^{(1)}_{s_{l-1}^{(1)}} + \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{s_{l-1}^{(1)}+1, j} + \sum_{i=s_{l-1}^{(1)}+2}^{\infty} \sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{i,j}. \end{aligned} \tag{A.18}$$

We first show that

$$\sum_{j=0}^{s_{l-1}^{(2)}} \hat{P}^{(1)}_{s_{l-1}^{(1)}+1, j} = 0. \tag{A.19}$$

To this end, note that, in our proof by contradiction, we assumed above that $\hat{P}^{(1)}_{s^{(1)}_{l-1}+1,s^{(2)}_{l-1}} = 0$. In addition, note that, from the induction hypothesis $\mathbb{S}^{(1)}_{l-1}$, $s_{l-1} = (s^{(1)}_{l-1}, s^{(2)}_{l-1})$ is on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$, whence $s_{l-1} = (s^{(1)}_{l-1}, s^{(2)}_{l-1})$ is a support point of $\hat{P}^{(1)}$. Therefore, none of the points $(s^{(1)}_{l-1}+1, j)$ for $0 \leq j < s^{(2)}_{l-1}$ can be a support point of $\hat{P}^{(1)}$, since, if they were, this would violate the comonotonicity of $\hat{P}^{(1)}$. Consequently, $\hat{P}^{(1)}_{s^{(1)}_{l-1}+1,j} = 0$ for all $j = 0, 1, ..., s^{(2)}_{l-1} - 1$. Hence, (A.19) must hold.

Next we show that

$$\sum_{i=s^{(1)}_{l-1}+2}^{\infty} \sum_{j=0}^{s^{(2)}_{l-1}} \hat{P}^{(1)}_{i,j} = 0. \tag{A.20}$$

To this end, note that in our proof by contradiction, we assumed above that $(s^{(1)}_{l-1} + 1, s^{(2)}_{l-1} + 1)$ is the subsequent point to $s_{l-1} = (s^{(1)}_{l-1}, s^{(2)}_{l-1})$ on the S-path associated with $\hat{P}^{(1)}$, whence $(s^{(1)}_{l-1} + 1, s^{(2)}_{l-1} + 1)$ is a support point of $\hat{P}^{(1)}$. Therefore, none of the points (i, j) for $i \geq s^{(1)}_{l-1} + 2$ and $j \in \{0, 1, ..., s^{(2)}_{l-1}\}$ can be a support point of $\hat{P}^{(1)}$, since, if they were, this would violate the comonotonicity of $\hat{P}^{(1)}$. Hence, (A.20) must hold.

Combining (A.18) with (A.19) and (A.20), we get that $F^{(2)}_{s^{(2)}_{l-1}} = F^{(1)}_{s^{(1)}_{l-1}}$. However, in Case 2, $z_{l-1} = F^{(1)}_{s^{(1)}_{l-1}}$ and $z_{l-1} \neq F^{(2)}_{s^{(2)}_{l-1}}$, whence $F^{(2)}_{s^{(2)}_{l-1}} \neq F^{(1)}_{s^{(1)}_{l-1}}$. Hence, we have arrived at a contradiction. Therefore, $(s^{(1)}_{l-1} + 1, s^{(2)}_{l-1} + 1)$ cannot be the subsequent point to $s_{l-1} = (s^{(1)}_{l-1}, s^{(2)}_{l-1})$ on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$. Ergo, the subsequent point to s_{l-1} on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$ must be $s_l = (s^{(1)}_{l-1}+1, s^{(2)}_{l-1})$ computed on Lines 12 and 13 of Algorithm 1 and specified by (A.15). From the induction hypothesis $\mathbb{S}^{(1)}_{l-1}$, it follows that the point s_{l-1} computed by Algorithm 1 is the l^{th} point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$. Therefore, s_l, the subsequent point to s_{l-1} on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$, is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}$. Therefore, we have verified that $\mathbb{S}^{(1)}_l$ holds.

Next, we verify that $\mathbb{S}^{(2)}_l$ holds. To this end, note that, from (A.15), $s^{(1)}_{l-1} = s^{(1)}_l - 1$ and, in Case 2, $z_{l-1} = F^{(1)}_{s^{(1)}_{l-1}}$, whence

$$F^{(1)}_{s^{(1)}_l - 1} = F^{(1)}_{s^{(1)}_{l-1}} = z_{l-1}.$$

Also, $F^{(1)}_{s^{(1)}_l - 1} < F^{(1)}_{s^{(1)}_l}$, since, as noted above, the marginal cdfs $F^{(1)}$ and $F^{(2)}$ are strictly increasing (see (2.25)). Therefore,

$$F^{(1)}_{s^{(1)}_l - 1} = z_{l-1} < F^{(1)}_{s^{(1)}_l}. \tag{A.21}$$

Since the induction hypothesis $\mathbb{S}^{(2)}_{l-1}$ holds, we can replace l by $l-1$ in the second line of (A.1) to get

$$F^{(2)}_{s^{(2)}_{l-1}-1} \leq z_{l-2} < z_{l-1} \leq F^{(2)}_{s^{(2)}_{l-1}}.$$

However, in Case 2, $z_{l-1} \neq F^{(2)}_{s^{(2)}_{l-1}}$. Therefore, we must have

$$F^{(2)}_{s^{(2)}_{l-1}-1} \leq z_{l-2} < z_{l-1} < F^{(2)}_{s^{(2)}_{l-1}}.$$

Dropping the z_{l-2} from the line above, we have

$$F^{(2)}_{s^{(2)}_{l-1}-1} < z_{l-1} < F^{(2)}_{s^{(2)}_{l-1}}.$$

From (A.15), $s_l^{(2)} = s_{l-1}^{(2)}$. Therefore,

$$F^{(2)}_{s_l^{(2)}-1} < z_{l-1} < F^{(2)}_{s_l^{(2)}}. \tag{A.22}$$

Now note that Line 18 of Algorithm 1 computes

$$z_l = \min\left(F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\right).$$

Since, from (A.21) and (A.22),

$$z_{l-1} < F^{(1)}_{s_l^{(1)}} \qquad \text{and} \qquad z_{l-1} < F^{(2)}_{s_l^{(2)}},$$

it follows that

$$z_{l-1} < \min\left(F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\right) = z_l. \tag{A.23}$$

Moreover,

$$\begin{aligned} z_l &= \min\left(F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\right) \le F^{(1)}_{s_l^{(1)}}, \\ z_l &= \min\left(F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\right) \le F^{(2)}_{s_l^{(2)}}. \end{aligned} \tag{A.24}$$

Therefore, it follows from (A.21), (A.22) (A.23) and (A.24) that

$$\begin{aligned} F^{(1)}_{s_l^{(1)}-1} &= z_{l-1} < z_l \le F^{(1)}_{s_l^{(1)}}, \\ F^{(2)}_{s_l^{(2)}-1} &< z_{l-1} < z_l \le F^{(2)}_{s_l^{(2)}}. \end{aligned} \tag{A.25}$$

Therefore, (A.1) holds. Moreover, from (A.25), it follows that the first equation in (A.2) holds. In addition, since

$$z_l = \min\left(F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\right),$$

z_l must be equal to at least one of $F^{(1)}_{s_l^{(1)}}$ or $F^{(2)}_{s_l^{(2)}}$. Hence, at least one of the equations in (A.3) holds. Furthermore, since, from the induction hypothesis $\mathbb{S}_{l-1}^{(2)}$, we know that z_{l-1} is the l^{th} smallest element in $\Pi_Z = \Pi_{X^{(1)}} \vee \Pi_{X^{(2)}}$, it follows from (A.25) that the smaller of $F^{(1)}_{s_l^{(1)}}$ and $F^{(2)}_{s_l^{(2)}}$ must be the $(l+1)^{\text{st}}$ smallest element in Π_Z. Hence, $z_l = \min\left(F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\right)$ is the $(l+1)^{\text{st}}$ smallest element in Π_Z. Therefore, we have verified that $\mathbb{S}_l^{(2)}$ holds.

Now we verify that $\mathbb{S}_l^{(3)}$ holds. The value of $\hat{P}^{(1)}_{s_l^{(1)}, s_l^{(2)}}$ computed on Line 18 of Algorithm 1 is

$$\hat{P}^{(1)}_{s_l^{(1)}, s_l^{(2)}} = z_l - z_{l-1}. \tag{A.26}$$

We know from Line 4 of Algorithm 1 which calls Algorithm 2 to compute the partition of unity that

$$z_l = \min\left(F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\right). \tag{A.27}$$

In addition, we proved above that $\mathbb{S}_l^{(2)}$ holds. Therefore, we have from (A.1) that

$$F^{(1)}_{s_l^{(1)}-1} \leq z_{l-1} \quad \text{and} \quad F^{(2)}_{s_l^{(2)}-1} \leq z_{l-1}$$

and from (A.2) that at least one of

$$F^{(1)}_{s_l^{(1)}-1} = z_{l-1} \quad \text{and} \quad F^{(2)}_{s_l^{(2)}-1} = z_{l-1}$$

holds. Therefore,

$$z_{l-1} = \max\left(F^{(1)}_{s_l^{(1)}-1}, F^{(2)}_{s_l^{(2)}-1}\right). \tag{A.28}$$

Since we verified above that $\mathbb{S}_l^{(2)}$ holds, we also have $z_{l-1} < z_l$, whence $z_l - z_{l-1} > 0$. Combining $z_l - z_{l-1} > 0$ with (A.26), (A.27) and (A.28), we get

$$\begin{aligned}
\hat{P}^{(1)}_{s_l^{(1)},s_l^{(2)}} &= z_l - z_{l-1} \\
&= \min\left(F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\right) - \max\left(F^{(1)}_{s_l^{(1)}-1}, F^{(2)}_{s_l^{(2)}-1}\right) \\
&= \left[\min\left(F^{(1)}_{s_l^{(1)}}, F^{(2)}_{s_l^{(2)}}\right) - \max\left(F^{(1)}_{s_l^{(1)}-1}, F^{(2)}_{s_l^{(2)}-1}\right)\right]^+.
\end{aligned} \tag{A.29}$$

Therefore, the probability $\hat{P}^{(1)}_{s_l^{(1)},s_l^{(2)}}$ computed on Line 18 of Algorithm 1 is correct in the sense that it agrees with (2.7). Therefore, we have verified that $\mathbb{S}_l^{(3)}$ holds.

Thus, we have verified that all three statements in $\mathbb{S}_l$ hold in Case 2.

Case 3: $z_{l-1} \neq F^{(1)}_{s_{l-1}^{(1)}}$ and $z_{l-1} = F^{(2)}_{s_{l-1}^{(2)}}$ (Line 14)

The proof that Algorithm 1 is correct in this case is very similar to the proof in Case 2. Therefore, for the sake of brevity, we omit the proof in this case.

Since we have shown that for each $l = 0, 1, 2, \ldots, l_{\max}$, $\mathbb{S}_l$ holds true for all three cases, we have shown that Theorem 3 holds true.

□

A.2 Proof of Proposition 2

PROOF: We prove Proposition 2 by induction on l.

To begin, we prove the base case of the induction. We know from Lemma 1 that $s_0 = (s_0^{(1)}, s_0^{(2)}) = (0, 0)$ is the 1^{st} point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\vec{\mathbf{F}}^2}$ and we know from Remark 21 that $\bar{s}_0 = (0, j_{\max})$, is the 1^{st} point on the $\mathcal{S}$-path associated with $\hat{P}^{(2)}$. Therefore, we have that $\bar{s}_0 = (s_0^{(1)}, j_{\max} - s_0^{(2)})$, as required.

For the induction step, choose any $l \in \{1, 2, \ldots, l_{\max}\}$ and, for the induction hypothesis, assume that Proposition 2 holds for $l-1$. That is, for the induction hypothesis, assume that, if $s_{l-1} = (s_{l-1}^{(1)}, s_{l-1}^{(2)})$ is the l^{th} point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\vec{\mathbf{F}}^{(2)}}$, then $\bar{s}_{l-1} = (s_{l-1}^{(1)}, j_{\max} - s_{l-1}^{(2)})$ is the l^{th} point on the $\mathcal{S}$-path associated with $\hat{P}^{(2)}$. Our goal is to show that, if $s_l = (s_l^{(1)}, s_l^{(2)})$ is the $(l+1)^{\text{st}}$ point on the

$\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$, then $\bar{s}_l = (s_l^{(1)}, j_{\max} - s_l^{(2)})$ is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(2)}$.

To this end, suppose s_l is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$. By the induction hypothesis, s_{l-1} is the l^{th} point on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$. Therefore, s_l must be the subsequent point to s_{l-1} on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$. Hence, by Definition 17 and Remark 12, s_l must be one of the three points

$$(s_{l-1}^{(1)} + 1, s_{l-1}^{(2)}), \quad (s_{l-1}^{(1)}, s_{l-1}^{(2)} + 1), \quad \text{or} \quad (s_{l-1}^{(1)} + 1, s_{l-1}^{(2)} + 1). \tag{A.30}$$

To complete the proof, consider three cases depending on which of the three points in (A.30) is equal to s_l.

Case 1: $s_l = (s_{l-1}^{(1)} + 1, s_{l-1}^{(2)})$

Since $s_l = (s_{l-1}^{(1)} + 1, s_{l-1}^{(2)})$ is the subsequent point to $s_{l-1} = (s_{l-1}^{(1)}, s_{l-1}^{(2)})$ on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$, it follows that

$$\hat{P}^{(1)}_{s_{l-1}^{(1)}+1, s_{l-1}^{(2)} | \breve{\mathbf{F}}^{(2)}} > 0.$$

Consequently, it follows from Lemma 14 that

$$\hat{P}^{(2)}_{s_{l-1}^{(1)}+1, j_{\max} - s_{l-1}^{(2)}} > 0.$$

Therefore, $\bar{s}_l = (s_{l-1}^{(1)} + 1, j_{\max} - s_{l-1}^{(2)})$ is the subsequent point to $\bar{s}_{l-1} = (s_{l-1}^{(1)}, j_{\max} - s_{l-1}^{(2)})$ on the $\mathcal{S}$-path associated with $\hat{P}^{(2)}$. From the induction hypothesis, $\bar{s}_{l-1} = (s_{l-1}^{(1)}, j_{\max} - s_{l-1}^{(2)})$ is the l^{th} point on the $\mathcal{S}$-path associated with $\hat{P}^{(2)}$. Hence, $\bar{s}_l = (s_{l-1}^{(1)} + 1, j_{\max} - s_{l-1}^{(2)})$ is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(2)}$. To complete the proof in this case, note that, since $s_l = (s_{l-1}^{(1)} + 1, s_{l-1}^{(2)})$, $\bar{s}_l = (s_l^{(1)}, j_{\max} - s_l^{(2)})$, as required.

Case 2: $s_l = (s_{l-1}^{(1)}, s_{l-1}^{(2)} + 1)$

The proof in this case is very similar to the proof in Case 1. Therefore, for brevity, we omit it.

Case 3: $s_l = (s_{l-1}^{(1)} + 1, s_{l-1}^{(2)} + 1)$

Since $s_l = (s_{l-1}^{(1)} + 1, s_{l-1}^{(2)} + 1)$ is the subsequent point to $s_{l-1} = (s_{l-1}^{(1)}, s_{l-1}^{(2)})$ on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$, it follows that

$$\hat{P}^{(1)}_{s_{l-1}^{(1)}+1, s_{l-1}^{(2)}+1 | \breve{\mathbf{F}}^{(2)}} > 0.$$

However, in this case, we must also have

$$\hat{P}^{(1)}_{s_{l-1}^{(1)}+1, s_{l-1}^{(2)} | \breve{\mathbf{F}}^{(2)}} = 0,$$
$$\hat{P}^{(1)}_{s_{l-1}^{(1)}, s_{l-1}^{(2)}+1 | \breve{\mathbf{F}}^{(2)}} = 0$$

since otherwise one of the points $(s_{l-1}^{(1)} + 1, s_{l-1}^{(2)})$ or $(s_{l-1}^{(1)}, s_{l-1}^{(2)} + 1)$ would be the subsequent point to $s_{l-1} = (s_{l-1}^{(1)}, s_{l-1}^{(2)})$ on the $\mathcal{S}$-path associated with $\hat{P}^{(1)}_{\breve{\mathbf{F}}^{(2)}}$. Consequently, it follows from Lemma 14

that

$$\begin{aligned}\hat{P}^{(2)}_{s^{(1)}_{l-1}+1,j_{\max}-s^{(2)}_{l-1}-1} &> 0,\\ \hat{P}^{(2)}_{s^{(1)}_{l-1}+1,j_{\max}-s^{(2)}_{l-1}} &= 0,\\ \hat{P}^{(2)}_{s^{(1)}_{l-1},j_{\max}-s^{(2)}_{l-1}-1} &= 0.\end{aligned}$$

Therefore, $\bar{s}_l = (s^{(1)}_{l-1}+1, j_{\max}-s^{(2)}_{l-1}-1)$ is the subsequent point to $\bar{s}_{l-1} = (s^{(1)}_{l-1}, j_{\max}-s^{(2)}_{l-1})$ on the $\mathcal{S}$-path associated with $\hat{P}^{(2)}$. From the induction hypothesis, $\bar{s}_{l-1} = (s^{(1)}_{l-1}, j_{\max}-s^{(2)}_{l-1})$ is the l^{th} point on the $\mathcal{S}$-path associated with $\hat{P}^{(2)}$. Hence, $\bar{s}_l = (s^{(1)}_{l-1}+1, j_{\max}-s^{(2)}_{l-1}-1)$ is the $(l+1)^{\text{st}}$ point on the $\mathcal{S}$-path associated with $\hat{P}^{(2)}$. To complete the proof in this case, note that, since $s_l = (s^{(1)}_{l-1}+1, s^{(2)}_{l-1}+1)$, $\bar{s}_l = (s^{(1)}_l+1, j_{\max}-s^{(2)}_l-1)$, as required.

□

Appendix B

Chapter 2.5 Proofs

B.1 Proof of Lemma 16

PROOF: To begin, we consider the two special cases: $i = -1$ and $i = i_{\max}^{(d;k)}$. For any $k \in \{1, 2, \ldots, d\}$ and $i = -1$, (2.89) reduces to

$$1 - F_{-1}^{(j',d;k)} = F_{i_{\max}^{(d;k)}}^{(j,d;k)} \tag{B.1}$$

which is clearly true, since $F_{-1}^{(j',d;k)} = 0$ and $F_{i_{\max}}^{(j,d;k)} = 1$.

Similarly, for any $k \in \{1, 2, \ldots, d\}$ and $i = i_{\max}^{(d;k)}$, (2.89) reduces to

$$1 - F_{i_{\max}^{(d;k)}}^{(j',d;k)} = F_{-1}^{(j,d;k)} \tag{B.2}$$

which is clearly true, since $F_{-1}^{(j,d;k)} = 0$ and $F_{i_{\max}^{(d;k)}}^{(j',d;k)} = 1$.

Now consider the remaining general case $i \in \{0, 1, \ldots, i_{\max}^{(d;k)} - 1\}$. From (2.76) and (2.77), it follows that, for any $k \in \{1, 2, \ldots, d\}$ and any $i \in \{0, 1, \ldots, i_{\max}^{(d;k)} - 1\}$,

$$\begin{aligned}
1 - F_i^{(j',d;k)} &= 1 - \sum_{n=0}^{i} Q_n^{(j',d;k)} \\
&= 1 - \sum_{n=i_{\max}^{(d;k)}-i}^{i_{\max}^{(d;k)}} Q_n^{(j,d;k)} \\
&= 1 - (Q_{i_{\max}^{(d;k)}}^{(j,d;k)} + \cdots + Q_{i_{\max}^{(d;k)}-i}^{(j,d;k)}) \\
&= (Q_{i_{\max}^{(d;k)}}^{(j,d;k)} + \cdots + Q_{i_{\max}^{(d;k)}-i}^{(j,d;k)} + Q_{i_{\max}^{(d;k)}-i-1}^{(j,d;k)} + \cdots + Q_0^{(j,d;k)}) - (Q_{i_{\max}^{(d;k)}}^{(j,d;k)} + \cdots + Q_{i_{\max}^{(d;k)}-i}^{(j,d;k)}) \\
&= Q_0^{(j,d;k)} + \cdots + Q_{i_{\max}^{(d;k)}-i-1}^{(j,d;k)} \\
&= F_{i_{\max}^{(d;k)}-i-1}^{(j,d;k)}.
\end{aligned}$$

□

B.2 Proof of Lemma 17

PROOF: Before we begin the proof, note that $z_{-1}^{(j,2;u,v)} = 0$ is the smallest element in $\mathbf{z}^{(j,2;u,v)}$ and $z_0^{(j,2;u,v)}$ is the second smallest element in $\mathbf{z}^{(j,2;u,v)}$, and so on. Similarly, $z_{-1}^{(j',2;u,v)} = 0$ and $z_0^{(j',2;u,v)}$ are the smallest element and second smallest element, respectively, in $\mathbf{z}^{(j',2;u,v)}$.

We begin by showing that (2.90) is true. To this end, choose any $l \in \{-1, 0, 1, \ldots, l_{\max}^2 - 1, l_{\max}^2\}$. By the construction of the partition of unity, $z_l^{(j',2;u,v)}$ is the $(l+2)^{\text{th}}$ element from the left in the sequence $\boldsymbol{z}^{(j',2;u,v)}$ and must equal $F_i^{(j',d;k)}$ for some $k \in \{u, v\}$ and some $i \in \{-1, 0, 1, \ldots, i_{\max}^{(d;k)}\}$. That is, $z_l^{(j',2;u,v)} = F_i^{(j',d;k)}$ for some $k \in \{u, v\}$ and some $i \in \{-1, 0, 1, \ldots, i_{\max}^{(d;k)}\}$. Thus, there must be $(l+1)$ values in the set $\cup_{k\in\{u,v\}}\{F_i^{(j',d;k)} : i = -1, 0, \ldots, i_{\max}^{(d;k)}\}$ smaller than $F_i^{(j',d;k)}$ and $l_{\max}^2 - l$ values in the set $\cup_{k\in\{u,v\}}\{F_i^{(j',d;k)} : i = -1, 0, \ldots, i_{\max}^{(d;k)}\}$ larger than $F_i^{(j',d;k)}$. Moreover, $1 - F_i^{(j',d;k)}$ must be the $(l+2)^{\text{th}}$ element from the right in the set $\cup_{k\in\{u,v\}}\{1 - F_i^{(j',d;k)} : i = -1, 0, \ldots, i_{\max}^{(d;k)}\}$. That is, there must be $l_{\max}^2 - l$ values in the set $\cup_{k\in\{u,v\}}\{1 - F_i^{(j',d;k)} : i = -1, 0, \ldots, i_{\max}^{(d;k)}\}$ smaller than $1 - F_i^{(j',d;k)}$ and $l+1$ items greater than $1 - F_i^{(j',d;k)}$. By Lemma 16, we must have that $1 - F_i^{(j',d;k)} = F_{i_{\max}^{(d;k)}-i-1}^{(j,d;k)}$, whence, $F_{i_{\max}^{(d;k)}-i-1}^{(j,d;k)}$ must also be the $(l+2)^{\text{th}}$ element from the right in the set $\cup_{k\in\{u,v\}}\{F_i^{(j,d;k)} : i = -1, 0, \ldots, i_{\max}^{(d;k)}\}$. Therefore, it must be that $z_{l_{\max}^2-l-1}^{(j,2;u,v)} = F_{i_{\max}^{(d;k)}-i-1}^{(j,d;k)} = 1 - F_i^{(j',d;k)} = 1 - z_l^{(j',2;u,v)}$, since $z_{l_{\max}-l-1}^{(j,2;u,v)}$ is the $(l+2)^{\text{th}}$ element from the right in the set $z^{(j',2;u,v)} = \cup_{k\in\{u,v\}}\{F_i^{(j,d;k)} : i = -1, 0, \ldots, i_{\max}^{(d;k)}\}$. Therefore, (2.90) holds true.

Finally, note that, in the argument in the paragraph above, the same dimension index, k, and the same iteration index, i, are used in the proof of both $z_l^{(j',2;u,v)} = F_i^{(j',d;k)}$ and $z_{l_{\max}^2-l-1}^{(j,2;u,v)} = F_{i_{\max}^{(d;k)}-i-1}^{(j,d;k)} = 1 - F_i^{(j',d;k)}$ Therefore, both (2.91) and (2.92) above hold true, where the k's and i's are the same in (2.91) and (2.92).

□

B.3 Proof of Lemma 23

PROOF: We begin by showing that $\tilde{P}^{(j,d)}$ is a probability distribution. That is,

$$\tilde{P}_{i_1,\ldots,i_d}^{(j,d)} \geq 0 \tag{B.3}$$

for all $k \in \{1, \ldots, d\}$ and all $i_k \in \{0, 1, \ldots, i_{\max}^{(d;k)}\}$, and that

$$\sum_{i_1=0}^{i_{\max}^{(d;1)}} \cdots \sum_{i_d=0}^{i_{\max}^{(d;d)}} \tilde{P}_{i_1,\ldots,i_d}^{(j,d)} = 1. \tag{B.4}$$

We start with (B.4) by noting that the left side can be written as

$$\sum_{i_1=0}^{i_{\max}^{(d;1)}} \cdots \sum_{i_d=0}^{i_{\max}^{(d;d)}} \tilde{P}_{i_1,\ldots,i_d}^{(j,d)} = \sum_{l=0}^{l_{\max}^{(j,d)}} \tilde{P}_{s_{l^{(j,d)}}^{(j,d;1)},\ldots,s_{l^{(j,d)}}^{(j,d;d)}}^{(j,d)} = \sum_{l=0}^{l_{\max}^{(j,d)}} \tilde{P}_{l^{(j,d)}}^{(j,d)} \tag{B.5}$$

since Algorithm 5 computes all the points, $\{s_{l^{(j,d)}}^{(j,d;1)}, \ldots, s_{l^{(j,d)}}^{(j,d;d)}\}_{l^{(j,d)}=0}^{l_{\max}^{(j,d)}}$, belonging to the support of $\tilde{P}^{(j,d)}$ and all other points in the domain are defined to be 0 since they do not belong to the support. Since we have that $\tilde{P}_0^{(j,d)} = z_0^{(j,d)}$ by Line 8 of Algorithm 5 and $\tilde{P}_{l^{(j,d)}}^{(j,d)} = z_{l^{(j,d)}}^{(j,d)} - z_{l^{(j,d)}-1}^{(j,d)}$ for $l^{(j,d)} \geq 1$ from

Line 14 of Algorithm 5, the right-most sum in (B.5) satisfies

$$\sum_{l^{(j,d)}=0}^{l_{\max}^{(j,d)}} \tilde{P}_l^{(j,d)} = z_0^{(j,d)} + (z_1^{(j,d)} - z_0^{(j,d)}) + \cdots + (z_{l_{\max}^{(j,d)}}^{(j,d)} - z_{l_{\max}^{(j,d)}-1}^{(j,d)}) = z_{l_{\max}^{(j,d)}}^{(j,d)}. \tag{B.6}$$

Therefore, by (B.5), (B.6), and the fact that $z_{l_{\max}^{(j,d)}}^{(j,d)} = 1$, (B.4) holds true.

Next, we show that (B.3) is valid. To this end, choose any point $(i_1, ..., i_d)$ in the domain of $\tilde{P}^{(j,d)}$ and note that $(i_1, ..., i_d)$ must satisfy either

(a) $(i_1, ..., i_d) = s_{l^{(j,d)}}^{(j,d)}$ for some $l^{(j,d)} \in \{0, 1, ..., l_{\max}^{(j,d)}\}$, or

(b) $(i_1, ..., i_d) \neq s_{l^{(j,d)}}^{(j,d)}$ for any $l^{(j,d)} \in \{0, 1, ..., l_{\max}^{(j,d)}\}$.

In case (a), $\tilde{P}_{i_1,...,i_d}^{(j,d)} = \tilde{P}_{s_{l^{(j,d)}}^{(j,d)}}^{(j,d)} = \tilde{P}_{l^{(j,d)}}^{(j,d)}$. Also, by Algorithm 5, $\tilde{P}_0^{(j,d)} = z_0^{(j,d)}$ and $\tilde{P}_{l^{(j,d)}}^{(j,d)} = z_{l^{(j,d)}}^{(j,d)} - z_{l^{(j,d)}-1}^{(j,d)}$ for $l^{(j,d)} > 0$. In addition, by Algorithm 8, $\{z_0^{(j,d)}, z_1^{(j,d)}, ..., z_{l_{\max}^{(j,d)}}^{(j,d)}\}$ is a strictly increasing sequence of positive values. Hence, $\tilde{P}_0^{(j,d)} = z_0^{(j,d)} > 0$ and $\tilde{P}_{l^{(j,d)}}^{(j,d)} = z_{l^{(j,d)}}^{(j,d)} - z_{l^{(j,d)}-1}^{(j,d)} > 0$ for $l^{(j,d)} > 0$. Therefore, $\tilde{P}_{i_1,...,i_d}^{(j,d)} = \tilde{P}_{s_{l^{(j,d)}}^{(j,d)}}^{(j,d)} = \tilde{P}_{l^{(j,d)}}^{(j,d)} > 0$. On the other hand, in case (b), $(i_1, ..., i_d)$ is not in the support of $\tilde{P}^{(j,d)}$, since Algorithm 5 computes all the support points, $s_{l^{(j,d)}}^{(j,d)}$, of $\tilde{P}^{(j,d)}$ and $(i_1, ..., i_d) \neq s_{l^{(j,d)}}^{(j,d)}$ for any $l^{(j,d)} \in \{0, 1, ..., l_{\max}^{(j,d)}\}$. Hence, $\tilde{P}_{i_1,...,i_d}^{(j,d)} = 0$. Therefore, whether case (a) or case (b) holds, $\tilde{P}_{i_1,...,i_d}^{(j,d)} \geq 0$, whence (B.3) is valid.

Since (B.3) and (B.4) hold true, we have shown that $\tilde{P}^{(j,d)}$ is a probability distribution and that it satisfies (2.37c).

All that remains is to show that $\tilde{P}^{(j,d)}$ satisfies (2.37b). To this end, we note that, by Remark 43, for each $k \in \{1, \ldots, d\}$ and each $i_k \in \{0, 1, \ldots, i_{\max}^{(d;k)}\}$, we can rewrite the equation on the left of (2.37b) as

$$\sum_{v \in \mathcal{I}_k} \sum_{i_v=0}^{i_{\max}^{(d;v)}} \tilde{P}_{i_1,...i_{k-1},i_k,i_{k+1},...,i_d}^{(j,d)} = \sum_{\{l\,:\,s_{l^{(j,d)}}^{(j,d;k)} = i_k\}} \tilde{P}_{s_{l^{(j,d)}}^{(j,d;1)},...,s_{l^{(j,d)}}^{(j,d;d)}}^{(j,d)} \tag{B.7}$$

because, while (2.37b) is written assuming that $\tilde{P}^{(j,d)}$ is an infinite-dimensional distribution, it applies equally well, with minor modifications, to finite dimensional distributions. Lemma 19 and Corollary 1 guarantee the existence of the integers q and w, where q is the smallest integer such that $s_q^{(j,d;k)} = i_k$ and w is the largest integer such that $s_w^{(j,d;k)} = i_k$, enabling us to write (B.7) as

$$\sum_{v \in \mathcal{I}_k} \sum_{i_v=0}^{i_{\max}^{(d;v)}} \tilde{P}_{i_1,...i_{k-1},i_k,i_{k+1},...,i_d}^{(j,d)} = \sum_{l^{(j,d)}=q}^{w} \tilde{P}_{s_{l^{(j,d)}}^{(j,d;1)},...,s_{l^{(j,d)}}^{(j,d;d)}}^{(j,d)} = \sum_{l^{(j,d)}=q}^{w} \tilde{P}_{l^{(j,d)}}^{(j,d)} = \tilde{P}_q^{(j,d)} + \cdots + \tilde{P}_w^{(j,d)}. \tag{B.8}$$

Since we have that $\tilde{P}_{l^{(j,d)}}^{(j,d)} = z_{l^{(j,d)}}^{(j,d)} - z_{l^{(j,d)}-1}^{(j,d)}$ from Line 14 of Algorithm 5, we can rewrite (B.8) as

$$\tilde{P}_q^{(j,d)} + \cdots + \tilde{P}_w^{(j,d)} = (z_q^{(j,d)} - z_{q-1}^{(j,d)}) + \cdots + (z_w^{(j,d)} - z_{w-1}^{(j,d)}) = z_w^{(j,d)} - z_{q-1}^{(j,d)}. \tag{B.9}$$

Note that (B.9) holds true in the edge case $q = 0$, since $\tilde{P}_0^{(j,d)} = z_0^{(j,d)}$ by Line 5 of Algorithm 5 and

$z_0^{(j,d)} = z_0^{(j,d)} - z_{-1}^{(j,d)}$, since $z_{-1}^{(j,d)} = 0$ by Definition 26, whence $\tilde{P}_0^{(j,d)} = z_0^{(j,d)} - z_{-1}^{(j,d)}$. It follows from Lines 8-12 of Algorithm 5 that either $s_{w+1}^{(j,d;k)} = s_w^{(j,d;k)} + 1$ or $s_{w+1}^{(j,d;k)} = s_w^{(j,d;k)}$. Therefore, if w is the greatest integer such that $s_w^{(j,d;k)} = i_k$, then it must be that $s_{w+1}^{(j,d;k)} = s_w^{(j,d;k)} + 1 = i_k + 1$. Consequently, Line 10 must get executed in the subsequent iteration of the algorithm. In order for Line 10 to get executed, the condition in Line 9 must evaluate to true, from which we can deduce that

$$z_w^{(j,d)} = F_{s_w^{(j,d;k)}}^{(j,d;k)} = F_{i_k}^{(j,d;k)}. \tag{B.10}$$

Similarly, since q is the smallest integer such that $s_q^{(j,d;k)} = i_k$, we can also deduce that

$$z_{q-1}^{(j,d)} = F_{s_q^{(j,d;k)}-1}^{(j,d;k)} = F_{i_k-1}^{(j,d;k)}. \tag{B.11}$$

Note that (B.11) holds for the edge case $q = 0$ since $s_0^{(j,d;k)} = 0$ by Line 4 of Algorithm 5 and by definition $F_{-1}^{(j,d;k)} = 0$. (See the beginning of this section.) Therefore, we have from (B.8), (B.9), (B.10), and (B.11) that

$$\sum_{v \in \mathcal{I}_k} \sum_{i_v=0}^{i_{\max}^{(d;v)}} \tilde{P}_{i_1,\ldots i_{k-1},i_k,i_{k+1},\ldots,i_d}^{(j,d)} = \sum_{l^{(j,d)}=q}^{w} \tilde{P}_{s_{l^{(j,d)}}^{(j,d;1)},\ldots,s_{l^{(j,d)}}^{(j,d;d)}}^{(j,d)} = F_{i_k}^{(d;k)} - F_{i_k-1}^{(d;k)} = Q_{i_k}^{(d;k)}. \tag{B.12}$$

Note that (B.12) holds true for the edge case $i_k = 0$ since $F_{-1}^{(j,d;k)} = 0$ and, by definition, $Q_0^{(j,d;k)} = F_0^{(j,d;k)}$. Therefore, the probability distribution $\tilde{P}^{(j,d)}$ satisfies the constraint (2.37b). □

B.4 Proof of Corollary 2

PROOF: Choose any $j \in \{1, 2, \ldots, n\}$, u, v such that $1 \leq u < v \leq d$, any $k \in \{u, v\}$, and any $l^2 \in \{0, 1, \ldots, l_{\max}^{(j,d;u,v)}\}$. By Remark 52, there exists a unique $l^{(j,d)} \in \{0, 1, \ldots, l_{\max}^{(j,d)}\}$ such that

$$z_{l^2}^{(j,2;u,v)} = z_{l^{(j,d)}}^{(j,d)}. \tag{B.13}$$

By Remark 54, there is at least one $l^{(j,d)} \in \{0, 1, \ldots, l_{\max}^{(j,d)}\}$ such that

$$\underline{\tilde{s}}_{l^2}^{(j,d;u,v)} = \tilde{s}_{l^{(j,d)}}^{(j,d;u,v)}. \tag{B.14}$$

For a given l^2, the unique $l^{(j,d)}$ that satisfies (B.13) also satisfies (B.14). We prove this preliminary result by induction on $l^2 \in \{0, 1, \ldots, l_{\max}^{(j,d;u,v)}\}$ and use it in the proof of the corollary.

We begin the induction proof with the base case $l^2 = 0$. First note that $\underline{\tilde{s}}_0^{(j,d;u,v)} = (0, 0)$ from the construction of the set $\underline{\tilde{\mathbf{s}}}^{(j,d;u,v)}$ from the set $\tilde{\mathbf{s}}^{(j,d)}$ and the initialization of $\tilde{s}_0^{(j,d)}$ to $(0, 0, \ldots, 0)$ on Line 4 of Algorithm 5. Next note that, for $l^2 = 0$, there exists, by Remark 52, a unique $l^{(j,d)}$ satisfying (B.13). Denote this $l^{(j,d)}$ by $l_0^{(j,d)}$. From the construction of the set $\mathbf{z}^{(j,2;u,v)}$, $z_0^{(j,2;u,v)}$ must be an element of either $\mathbf{F}^{(j,d;u)}$ or $\mathbf{F}^{(j,d;v)}$. Hence, (B.13) implies that $z_{l_0^{(j,d)}}^{(j,d)}$ must also belong to either $\mathbf{F}^{(j,d;u)}$ or $\mathbf{F}^{(j,d;v)}$. However, since $\mathbf{z}^{(j,2;u,v)} \subset \mathbf{z}^{(j,d)}$, $l_0^{(j,d)}$ may not necessarily be 0. If $l_0^{(j,d)} = 0$, it follows from Line 4 of Algorithm 5 that $\tilde{s}_{l_0^{(j,d)}}^{(j,d;u,v)} = \tilde{s}_0^{(j,d;u,v)} = (0, 0)$. Thus, $l_0^{(j,d)}$ satisfies both (B.13) and (B.14). On the other hand, if $l_0^{(j,d)} \neq 0$, (B.13) implies that $z_{l_0^{(j,d)}}^{(j,d)}$ is the first element of the set $\mathbf{z}^{(j,d)}$ that belongs to

either $\mathbf{F}^{(j,d;u)}$ or $\mathbf{F}^{(j,d;v)}$. Then, Lines 9-12 of Algorithm 5 further imply that $\tilde{s}^{(j,d)}_{l_0^{(j,d)}+1}$ is the first element of the set $\tilde{\mathbf{s}}^{(j,d)}$ where either the u^{th} or the v^{th} coordinate is incremented. Therefore, it follows that $\tilde{s}^{(j,d;u,v)}_{l_0^{(j,d)}} = (0,0)$. Thus, $l_0^{(j,d)}$ satisfies both (B.13) and (B.14).

Next, for the induction step, choose any $l^2 \in \{0,1,\dots,l_{\max}^{(j,d;u,v)}-1\}$ and, for the induction hypothesis, assume that there exists an $l^{(j,d)} \in \{0,1,\dots,l_{\max}^{(j,d)}\}$ that satisfies both (B.13) and (B.14). Denote this $l^{(j,d)}$ by $l_{l^2}^{(j,d)}$. Then, note that, by Remark 52, there exists a unique $l^{(j,d)} \in \{0,1,\dots,l_{\max}^{(j,d)}\}$ corresponding to l^2+1, denoted $l_{l^2+1}^{(j,d)}$, such that

$$z_{l^2+1}^{(j,2;u,v)} = z_{l_{l^2+1}^{(j,d)}}^{(j,d)}. \tag{B.15}$$

To complete the induction proof, we show below that the same $l_{l^2+1}^{(j,d)}$ must also satisfy

$$\underline{\tilde{s}}_{l^2+1}^{(j,d;u,v)} = \tilde{s}_{l_{l^2+1}^{(j,d)}}^{(j,d;u,v)}. \tag{B.16}$$

Towards that end, note that, by the generation of the set $\tilde{\mathbf{s}}^{(j,d)}$ in Algorithm 5 and by the construction of the set $\underline{\tilde{\mathbf{s}}}^{(j,d;u,v)}$ from the set $\tilde{\mathbf{s}}^{(j,d)}$, $\underline{\tilde{s}}_{l^2+1}^{(j,d;u,v)}$ must be one of $(\underline{\tilde{s}}_{l^2}^{(j,d;u)}+1, \underline{\tilde{s}}_{l^2}^{(j,d;v)})$, $(\underline{\tilde{s}}_{l^2}^{(j,d;u)}, \underline{\tilde{s}}_{l^2}^{(j,d;v)}+1)$, or $(\underline{\tilde{s}}_{l^2}^{(j,d;u)}+1, \underline{\tilde{s}}_{l^2}^{(j,d;v)}+1)$. Moreover, note that $z_{l^2}^{(j,2;u,v)} = z_{l_{l^2}^{(j,d)}}^{(j,d)}$ implies that $z_{l_{l^2}^{(j,d)}}^{(j,d)}$ must be an element of either $\mathbf{F}^{(j,d;u)}$ or $\mathbf{F}^{(j,d;v)}$ which, along with Lines 8-12 of of Algorithm 5, implies that $\tilde{s}_{l_{l^2}^{(j,d)}+1}^{(j,d;u,v)}$ must be one of $(\tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;u)}+1, \tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;v)})$, $(\tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;u)}, \tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;v)}+1)$, or $(\tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;u)}+1, \tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;v)}+1)$.

Case 1: $\tilde{s}_{l_{l^2}^{(j,d)}+1}^{(j,d;u,v)} = (\tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;u)}+1, \tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;v)})$

The induction hypothesis, $z_{l^2}^{(j,2;u,v)} = z_{l_{l^2}^{(j,d)}}^{(j,d)}$, along with our assertion (B.15) above that $z_{l^2+1}^{(j,2;u,v)} = z_{l_{l^2+1}^{(j,d)}}^{(j,d)}$ imply that there is no $z_{\hat{l}^{(j,d)}}^{(j,d)}$ with $\hat{l}^{(j,d)}$ satisfying $l_{l^2}^{(j,d)} < \hat{l}^{(j,d)} < l_{l^2+1}^{(j,d)}$ that belongs to either $\mathbf{F}^{(j,d;u)}$ or $\mathbf{F}^{(j,d;v)}$. Hence, for iterations $l_{l^2}^{(j,d)}+2$ to $l_{l^2+1}^{(j,d)}$, Line 9 of Algorithm 5 evaluates to false for $k \in \{u,v\}$, which leads to Line 12 of Algorithm 5 being executed. It then follows that $\tilde{s}_{l_{l^2}^{(j,d)}+1}^{(j,d;u,v)} = \cdots = \tilde{s}_{l_{l^2+1}^{(j,d)}}^{(j,d;u,v)}$, which enables us to rewrite the left side of the equation in our assumption for this case as

$$\tilde{s}_{l_{l^2+1}^{(j,d)}}^{(j,d;u,v)} = (\tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;u)}+1, \tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;v)}). \tag{B.17}$$

Then, since (B.14) holds for the same $l_{l^2}^{(j,d)}$ that satisfies (B.13) by the induction hypothesis, substituting (B.14) into the right side of (B.17) gives

$$\tilde{s}_{l_{l^2+1}^{(j,d)}}^{(j,d;u,v)} = (\underline{\tilde{s}}_{l^2}^{(j,d;u)}+1, \underline{\tilde{s}}_{l^2}^{(j,d;v)}). \tag{B.18}$$

Finally, since in this case we have that $\tilde{s}_{l_{l^2}^{(j,d)}+1}^{(j,d;u,v)} = (\tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;u)}+1, \tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;v)}) = \tilde{s}_{l_{l^2+1}^{(j,d)}}^{(j,d;u,v)}$ and the set $\underline{\tilde{\mathbf{s}}}^{(j,d;u,v)}$ is constructed from the set $\tilde{\mathbf{s}}^{(j,d)}$ by removing duplicates and retaining the order in which the support points were generated, it must then also be that $\tilde{s}_{l_{l^2+1}^{(j,d)}}^{(j,d;u,v)} = (\underline{\tilde{s}}_{l^2}^{(j,d;u)}+1, \underline{\tilde{s}}_{l^2}^{(j,d;v)}) = \underline{\tilde{s}}_{l^2+1}^{(j,d;u,v)}$, as required.

Case 2: $\tilde{s}_{l_{l^2}^{(j,d)}+1}^{(j,d;u,v)} = (\tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;u)}, \tilde{s}_{l_{l^2}^{(j,d)}}^{(j,d;v)}+1)$

In this case, it can be shown that $\tilde{s}^{(j,d;u,v)}_{l^{(j,d)}_{l^2+1}} = (\underline{\tilde{s}}^{(j,d;u)}_{l^2}, \underline{\tilde{s}}^{(j,d;v)}_{l^2} + 1) = \underline{\tilde{s}}^{(j,d;u,v)}_{l^2+1}$. Since the proof is very similar to the proof in Case 1, we omit it for the sake of brevity.

Case 3: $\tilde{s}^{(j,d;u,v)}_{l^{(j,d)}_{l^2}+1} = (\tilde{s}^{(j,d;u)}_{l^{(j,d)}_{l^2}} + 1, \tilde{s}^{(j,d;v)}_{l^{(j,d)}_{l^2}} + 1)$

In this case, it can be shown that $\tilde{s}^{(j.d;u,v)}_{l^{(j,d)}_{l^2+1}} = (\underline{\tilde{s}}^{(j,d;u)}_{l^2} + 1, \underline{\tilde{s}}^{(j,d;v)}_{l^2} + 1) = \underline{\tilde{s}}^{(j,d;u,v)}_{l^2+1}$. Since the proof is very similar to the proof in Case 1, we omit it for the sake of brevity.

Therefore, we have shown that (B.16) holds in all three cases above, which completes the induction step. Hence, we have shown by induction that, for any $l^2 \in \{0, 1, \dots, l^{(j,d;u,v)}_{\max}\}$, the unique $l^{(j,d)} \in \{0, 1, \dots, l^{(j,d)}_{\max}\}$ that satisfies (B.13) also satisfies (B.14).

Next, recall that, by Lemma 20, for every $l^{(j,d)} \in \{0, 1, \dots, l^{(j,d)}_{\max}\}$, $\tilde{s}^{(j,d;k)}_{l^{(j,d)}}$ is the number of $i \in \{0, 1, \dots, i^{(d;k)}_{\max}\}$ such that $F^{(j,d;k)}_i = z^{(j,d)}_l$ for some $l \in \{0, 1, \dots, l^{(j,d)} - 1\}$. Moreover, given the fact that we have shown above that, for every l^2 there exists a corresponding $l^{(j,d)}$ such that both (B.13) and (B.14) hold and the fact that we have fixed an l^2 at the beginning of this proof, (B.14) implies that $\underline{\tilde{s}}^{(j,d;k)}_{l^2}$ must also be the number of $i \in \{0, 1, \dots, i^{(d;k)}_{\max}\}$ such that $F^{(j,d;k)}_i = z^{(j,d)}_l$ for some $l \in \{0, 1, \dots, l^{(j,d)} - 1\}$. Thus, it remains to be shown that the number of $i \in \{0, 1, \dots, i^{(d;k)}_{\max}\}$ such that $F^{(j,d;k)}_i = z^{(j,d)}_l$ for some $l \in \{0, 1, \dots, l^{(j,d)} - 1\}$ must be equal to the number of $i \in \{0, 1, \dots, i^{(d;k)}_{\max}\}$ such that $F^{(j,d;k)}_i = z^{(j,2;u,v)}_l$ for some $l \in \{0, 1, \dots, l^2 - 1\}$, where l^2 was chosen in the first line of the proof and $l^{(j,d)}$ is such that (B.13) and (B.14) hold.

Towards that end, we first show that for any $i \in \{0, 1, \dots, i^{(d;k)}_{\max}\}$, all $j \in \{1, 2, \dots, n\}$, all $k \in \{u, v\}$, where u, v are such that $1 \le u < v \le d$,

$$F^{(j,d;k)}_i = z^{(j,d)}_l \tag{B.19}$$

for some $l \in \{0, 1, \dots, l^{(j,d)} - 1\}$ iff

$$F^{(j,d;k)}_i = z^{(j,2;u,v)}_l \tag{B.20}$$

for some $l \in \{0, 1, \dots, l^2 - 1\}$.

We begin by showing the $\Longleftarrow$ case. Towards this end, fix some $i \in \{0, 1, \dots, i^{(d;k)}_{\max}\}$ and assume that (B.20) holds for some $l \in \{0, 1, \dots, l^2 - 1\}$. Note that there can only be one such $l \in \{0, 1, \dots, l^2 - 1\}$ that satisfies (B.20) since there are no duplicate elements in the set $\mathbf{z}^{(j,2;u,v)}$. Further note that the $z^{(j,2;u,v)}_l$ in (B.20) must also be in the set $\mathbf{z}^{(j,d)}$ since $\mathbf{z}^{(j,2;u,v)} \subset \mathbf{z}^{(j,d)}$ by the construction of $\mathbf{z}^{(j,2;u,v)}$ in Remark 50. That is, there exists an $\hat{l} \in \{0, 1, \dots, l^{(j,d)}_{\max}\}$ such that

$$z^{(j,d)}_{\hat{l}} = z^{(j,2;u,v)}_l. \tag{B.21}$$

Then, (B.20) and (B.21) imply that $F^{(j,d;k)}_i = z^{(j,d)}_{\hat{l}}$. It remains to show that $\hat{l} \in \{0, 1, \dots, l^{(j,d)} - 1\}$. To see this, note that

$$z^{(j,2;u,v)}_l < z^{(j,2;u,v)}_{l^2} \tag{B.22}$$

since $l < l^2$ and the set $\mathbf{z}^{(j,2;u,v)}$ is strictly increasing. Next, substituting (B.13) in the right side of (B.22) and substituting (B.21) in the left side of (B.22) gives

$$z^{(j,d)}_{\hat{l}} < z^{(j,d)}_{l^{(j,d)}}. \tag{B.23}$$

Since $\mathbf{z}^{(j,d)}$ is also a strictly increasing set, (B.23) implies that $\hat{l} < l^{(j,d)}$. Therefore, (B.19) holds for $l = \hat{l}$ and $\hat{l} \in \{0, 1, \ldots, l^{(j,d)} - 1\}$, as required.

The $\Longrightarrow$ case is similar and for this reason we omit its proof. Finally, since we have shown that, for any $i \in \{0, 1, \ldots, i_{\max}^{(d;k)}\}$, (B.19) holds if and only if (B.20) holds, it must then be that the number of $i \in \{0, 1, \ldots, i_{\max}^{(d;k)}\}$ such that $F_i^{(j,d;k)} = z_l^{(j,d)}$ for some $l \in \{0, 1, \ldots, l^{(j,d)} - 1\}$ is equal to the number of $i \in \{0, 1, \ldots, i_{\max}^{(d;k)}\}$ such that $F_i^{(j,d;k)} = z_l^{(j,2;u,v)}$ for some $l \in \{0, 1, \ldots, l^2 - 1\}$. □

B.5 Proof of Lemma 22

PROOF: We begin by showing that (2.101) holds true for $l^2 = 0$. Recall from Lines 3-5 of Algorithm 1 that

$$s_0^{(j',2)} = (0, 0), \quad z_0^{(j',2)} = \min(F_0^{(j',2;1)}, F_0^{(j',2;2)}), \quad \text{and} \quad \hat{P}_0^{(j',2)} = z_0^{(j',2)}. \tag{B.24}$$

Similarly, from Lines 2-5 of Algorithm 5 and from Algorithm 8, we must also have that

$$s_0^{(j,d)} = (0, \ldots, 0), \quad z_0^{(j,d)} = \min(F_0^{(j,d;1)}, \ldots, F_0^{(j,d;d)}), \quad \text{and} \quad \tilde{P}_0^{(j,d)} = z_0^{(j,d)}. \tag{B.25}$$

Case 1: $F_0^{(j',2;1)} \leq F_0^{(j',2;2)}$

In this case, (B.24) implies that

$$\hat{P}^{(j',2)} = F_0^{(j',2;1)}. \tag{B.26}$$

From our assumption in this case that $F_0^{(j',2;1)} \leq F_0^{(j',2;2)}$, the fact that $F^{(j,d;u)} = F^{(j',2;1)}$ and $F^{(j,d;v)} = F^{(j',2;2)}$, we also have that

$$F_0^{(j,d;u)} \leq F_0^{(j,d;v)}. \tag{B.27}$$

By Lemma 19, let m and n be the smallest integers such that

$$s_m^{(j,d;u)} = s_n^{(j,d;v)} = 1. \tag{B.28}$$

Note that we assume here that $i_{\max}^{(d;u)} > 0$ and $i_{\max}^{(d;v)} > 0$ and treat various edge cases at the end of the proof for $l^2 = 0$. We must have that $m \leq n$ since (B.27) implies, by Lines 9-12 of Algorithm 5, that $s_m^{(j,d;u)}$ is incremented before $s_n^{(j,d;v)}$. Then, either

$$z_{m-1}^{(j,d)} = F_0^{(j,d;u)} \quad \text{or} \quad z_{m-1}^{(j,d)} = F_0^{(j,d;v)} \tag{B.29}$$

since for (B.28) to hold in Algorithm 5, each coordinate is incremented if and only if the previous z value is equal to the corresponding cdf with the same coordinate. In fact, we must have that

$$z_{m-1}^{(j,d)} = \min\{F_0^{(j,d;u)}, F_0^{(j,d;v)}\} \tag{B.30}$$

since the partition of unity is constructed in ascending order by Algorithm 8. Using (B.27) in (B.30), we have that

$$z_{m-1}^{(j,d)} = F_0^{(j,d;u)}. \tag{B.31}$$

We also have, from (B.28), $m \leq n$, and Algorithm 5, that

$$s_l^{(j,d;u,v)} = (0,0) \qquad \text{for} \quad l \in \{0,\dots,m-1\}. \tag{B.32}$$

Using (2.43), we compute

$$\begin{aligned}\tilde{P}_0^{(j,d;u,v)} = \tilde{P}_{(0,0)}^{(j,d;u,v)} &= \sum_{j \in \mathcal{I}_{u,v}} \sum_{i_j=0}^{\infty} \tilde{P}^{(j,d)}_{i_1,\dots i_{u-1},0,i_{u+1},\dots,i_{v-1},0,i_{v+1},\dots,i_d} \\ &= \sum_{l^{(j,d)}=0}^{m-1} \tilde{P}^{(j,d)}_{l^{(j,d)}}. \end{aligned} \tag{B.33}$$

Note that the second line in (B.33) follows from the first line by Remark 43. Then, recall from Line 5 of Algorithm 5 that

$$\tilde{P}_0^{(j,d)} = z_0^{(j,d)} \tag{B.34}$$

and from Line 14 of Algorithm 5 that

$$\tilde{P}^{(j,d)}_{l^{(j,d)}} = z^{(j,d)}_{l^{(j,d)}} - z^{(j,d)}_{l^{(j,d)}-1} \tag{B.35}$$

for $l^{(j,d)} > 0$. Using (B.31), (B.34) and (B.35) to rewrite (B.33), we have that

$$\begin{aligned}\tilde{P}_0^{(j,d;\,u,v)} &= z_0^{(j,d)} + (z_1^{(j,d)} - z_0^{(j,d)}) + \cdots + (z_{m-1}^{(j,d)} - z_{m-2}^{(j,d)}) \\ &= z_{m-1}^{(j,d)} = F_0^{(j,d;u)}. \end{aligned} \tag{B.36}$$

Therefore, (2.101) holds true for the case $l^2 = 0$ by (B.26) and (B.36), since $F_0^{(j,d;u)} = F_0^{(j',2;1)}$.

Case 2: $F_0^{(j',2;1)} \geq F_0^{(j',2;2)}$

The proof in this case is very similar to Case 1. We omit the proof in this case for the sake of brevity.

Note that in the edge case $i_{\max}^{(d;u)} = i_{\max}^{(2;1)} > 0$ and $i_{\max}^{(d;v)} = i_{\max}^{(2;2)} = 0$, we have that $F_0^{(j,d;v)} = 1 = F_0^{(j',2;2)}$. Since, by Line 6 of Algorithm 5, the v^{th} coordinate of the support does not get incremented, therefore, the n in (B.28) cannot exist. However, the m in (B.28) does exist. Consequently, we must have, from (B.30), that $z_{m-1}^{(j,d)} = F_0^{(j,d;u)}$. Similar arguments to those in (B.31)-(B.36) allow us to conclude that, for the edge case $i_{\max}^{(d;u)} = i_{\max}^{(2;1)} > 0$ and $i_{\max}^{(d;v)} = i_{\max}^{(2;2)} = 0$, (2.101) also holds true. A similar argument holds for the edge case $i_{\max}^{(d;u)} = i_{\max}^{(2;1)} = 0$ and $i_{\max}^{(d;v)} = i_{\max}^{(2;2)} > 0$.

In the edge case $i_{\max}^{(d;u)} = i_{\max}^{(2;1)} = 0$ and $i_{\max}^{(d;v)} = i_{\max}^{(2;2)} = 0$, however, neither m nor n in (B.28) exist. Since $F_0^{(j,d;u)} = F_0^{(j,d;v)} = 1$, we must have that $z^{(j,d)}_{l_{\max}^{(j,d)}} = 1 = F_0^{(j,d;u)} = F_0^{(j,d;v)}$ which implies that in order to compute $\tilde{P}_{(0,0)}^{(j,d;u,v)}$, the upper limit of summation (B.33) must be $l_{\max}^{(j,d)}$. Therefore, we have that

$$\tilde{P}_{(0,0)}^{(j,d;u,v)} = z_0^{(j,d)} + (z_1^{(j,d)} - z_0^{(j,d)}) + \cdots + (z^{(j,d)}_{l_{\max}^{(j,d)}} - z^{(j,d)}_{l_{\max}^{(j,d)}-1}) = z^{(j,d)}_{l_{\max}^{(j,d)}} = 1. \tag{B.37}$$

Since $\hat{P}_0^{(j',2)} = 1$, (2.101) also holds true for the case $l^2 = 0$, in the edge case $i_{\max}^{(d;u)} = i_{\max}^{(2;1)} = 0$ and $i_{\max}^{(d;v)} = i_{\max}^{(2;2)} = 0$.

Therefore, (2.101) holds true for the case $l^2 = 0$.

Next, we show that (2.101) holds true for any $l^2 \in \{1, 2, \ldots, l^2_{\max}\}$. By Line 18 of Algorithm 1,

$$\begin{aligned}\hat{P}^{(j',2)}_{l^2} &= z^{(j',2)}_{l^2} - z^{(j',2)}_{l^2-1} \\ &= \min\{F^{(j',2;1)}_{s^{(j',2;1)}_{l^2}}, F^{(j',2;2)}_{s^{(j',2;2)}_{l^2}}\} - \min\{F^{(j',2;1)}_{s^{(j',2;1)}_{l^2-1}}, F^{(j',2;2)}_{s^{(j',2;2)}_{l^2-1}}\}. \end{aligned} \tag{B.38}$$

since, by Remark 32, we have that $z^{(j',2)}_{l^2} = \min\{F^{(j',2;1)}_{s^{(2;1)}_{l^2}}, F^{(j',2;2)}_{s^{(2;2)}_{l^2}}\}$.

Let $l^{(j,d)}_-$ be the smallest iteration index in Algorithm 5 corresponding to the index l^2 such that

$$(s^{(j,d;u)}_{l^{(j,d)}_-}, s^{(j,d;v)}_{l^{(j,d)}_-}) = (s^{(j',2;1)}_{l^2}, s^{(j',2;2)}_{l^2}). \tag{B.39}$$

Note that $l^{(j,d)}_-$ must exist due to (B.61), which we have shown above. Then, either

$$z^{(j,d)}_{l^{(j,d)}_- - 1} = F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}_- - 1}} \quad \text{or} \quad z^{(j,d)}_{l^{(j,d)}_- - 1} = F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}_- - 1}} \tag{B.40}$$

since $l^{(j,d)}_-$ is the smallest index such that (B.39) holds and for $(s^{(j,d;u)}_{l^{(j,d)}_-}, s^{(j,d;v)}_{l^{(j,d)}_-})$ to be the current support point, either the u^{th} or the v^{th} coordinate must be incremented. Within Algorithm 5, a coordinate is incremented if and only if the previous z value is equal to the corresponding cdf from the same coordinate. In fact, we must have that

$$z^{(j,d)}_{l^{(j,d)}_- - 1} = \min\{F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}_- - 1}}, F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}_- - 1}}\} \tag{B.41}$$

since the partition of unity is constructed in ascending order by Algorithm 8.

Now, let $l^{(j,d)}_+$ be the greatest iteration index such that

$$(s^{(j,d;u)}_{l^{(j,d)}_+}, s^{(j,d;v)}_{l^{(j,d)}_+} = (s^{(j',2;1)}_{l^2}, s^{(j',2;2)}_{l^2}). \tag{B.42}$$

We assume here that $l^{(j,d)}_+ > l^{(j,d)}_-$, $l^{(j,d)}_- \neq l^{(j,d)}_{\max}$, and that $l^{(j,d)}_+ \neq l^{(j,d)}_{\max}$. We handle the edge cases at the end of the proof. Then, either

$$s^{(j,d;u)}_{l^{(j,d)}_+ + 1} = s^{(j,d;u)}_{l^{(j,d)}_+} + 1 \quad \text{or} \quad s^{(j,d;v)}_{l^{(j,d)}_+ + 1} = s^{(j,d;v)}_{l^{(j,d)}_+} + 1, \tag{B.43}$$

which in turn implies that

$$z^{(j,d)}_{l^{(j,d)}_+} = \min\{F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}_+}}, F^{(j,d;u)}_{s^{(j,d;v)}_{l^{(j,d)}_+}}\} \tag{B.44}$$

from similar arguments to those used above for $l^{(j,d)}_-$. We then use (2.43) to compute

$$\begin{aligned}\tilde{P}^{(j,d;u,v)}_{l^2} = \tilde{P}^{(j,d;u,v)}_{(s^{(j',2;1)}_{l^2}, s^{(j',2;2)}_{l^2})} &= \sum_{j \in \mathcal{I}_{u,v}} \sum_{i_j=0}^{\infty} \tilde{P}^{(j,d)}_{i_1,\ldots i_{u-1}, s^{(j',2;1)}_{l^2}, i_{u+1},\ldots,i_{v-1}, s^{(j',2;2)}_{l^2}, i_{v+1},\ldots,i_d} \\ &= \sum_{l=l^{(j,d)}_-}^{l^{(j,d)}_+} \tilde{P}^{(j,d)}_l = \tilde{P}^{(j,d)}_{l^{(j,d)}_-} + \cdots + \tilde{P}^{(j,d)}_{l^{(j,d)}_+}. \end{aligned} \tag{B.45}$$

Note that the second line in (B.45) follows from the first line by Remark 43. We can rewrite (B.45)

using (B.35) as

$$\begin{aligned}\tilde{P}^{(j,d;u,v)}_{l^2} &= (z^{(j,d)}_{l^{(j,d)}_-} - z^{(j,d)}_{l^{(j,d)}_- - 1}) + \cdots + (z^{(j,d)}_{l^{(j,d)}_+} - z^{(j,d)}_{l^{(j,d)}_+ - 1}) \\ &= z^{(j,d)}_{l^{(j,d)}_+} - z^{(j,d)}_{l^{(j,d)}_- - 1}. \end{aligned} \tag{B.46}$$

Using (B.41) and (B.44) to rewrite (B.46), we arrive at

$$\tilde{P}^{(j,d;u,v)}_{l^2} = \min\{F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}_+}}, F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}_+}}\} - \min\{F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}_- - 1}}, F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}_- - 1}}\} \tag{B.47}$$

$$= \min\{F^{(j',2;1)}_{s^{(j',2;1)}_{l^2}}, F^{(j',2;2)}_{s^{(j',2;2)}_{l^2}}\} - \min\{F^{(j',2;1)}_{s^{(j',2;1)}_{l^2-1}}, F^{(j',2;2)}_{s^{(j',2;2)}_{l^2-1}}\}. \tag{B.48}$$

The first min in (B.48) follows from the first min in (B.47) since $(s^{(j,d;u)}_{l^{(j,d)}_+}, s^{(j,d;v)}_{l^{(j,d)}_+}) = (s^{(j',2;1)}_{l^2}, s^{(j',2;2)}_{l^2})$ by the definition of $l^{(j,d)}_+$ and the fact that $F^{(j,d;u)} = F^{(j',2;1)}$ and $F^{(j,d;v)} = F^{(j',2;2)}$. To see that the second min in (B.48) follows from the second min in (B.47), first note that, from (B.39), we have that

$$(s^{(j,d;u)}_{l^{(j,d)}_-}, s^{(j,d;v)}_{l^{(j,d)}_-}) = (s^{(j',2;1)}_{l^2}, s^{(j',2;2)}_{l^2})$$

and $l^{(j,d)}_-$ is the smallest value for which this holds. Since we know from Statement 1 at the start of the proof of Lemma 24 (which has already been proved above) that the support of $\tilde{P}^{(j,d;u,v)}$ is identical to the support of $\hat{P}^{(j',2)}$, it follows that

$$(s^{(j,d;u)}_{l^{(j,d)}_- - 1}, s^{(j,d;v)}_{l^{(j,d)}_- - 1}) = (s^{(j',2;1)}_{l^2-1}, s^{(j',2;2)}_{l^2-1}).$$

Finally, recall that $F^{(j,d;u)} = F^{(j',2;1)}$ and $F^{(j,d;v)} = F^{(j',2;2)}$. Therefore, (B.38) and (B.48) imply that (2.101) holds in the general case.

Let us now handle the edge cases. In the edge case $l^{(j,d)}_+ = l^{(j,d)}_-$, (B.45) reduces to a single term

$$\tilde{P}^{(j,d;u,v)}_{l^2} = \tilde{P}^{(j,d)}_{l^{(j,d)}_-} = z^{(j,d)}_{l^{(j,d)}_-} - z^{(j,d)}_{l^{(j,d)}_- - 1}. \tag{B.49}$$

Using an argument similar to that used to derive (B.44) and (B.41), we can rewrite (B.49) as

$$\tilde{P}^{(j,d;u,v)}_{l^2} = \min\{F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}_-}}, F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}_-}}\} - \min\{F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}_- - 1}}, F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}_- - 1}}\}. \tag{B.50}$$

Similar reasoning to that used above in showing the equivalence between (B.47) and (B.38) can be used to show the equivalence of (B.50) and (B.38), thereby showing that (2.101) holds in this case too.

In the edge case $l^{(j,d)}_- = l^{(j,d)}_{\max}$, it is clear that we must also have that $l^{(j,d)}_+ = l^{(j,d)}_- = l^{(j,d)}_{\max}$. We omit the proof of this case since the arguments are similar to the edge case $l^{(j,d)}_+ = l^{(j,d)}_-$ considered in the paragraph above.

In the edge case $l^{(j,d)}_- < l^{(j,d)}_+ = l^{(j,d)}_{\max}$, it is clear that

$$z^{(j,d)}_{l^{(j,d)}_+} = z^{(j,d)}_{l^{(j,d)}_{\max}} = 1 \tag{B.51}$$

and that (B.46) takes the form

$$\tilde{P}_{l^2}^{(j,d;u,v)} = z_{l_+^{(j,d)}}^{(j,d)} - z_{l_-^{(j,d)}-1}^{(j,d)} = 1 - z_{l_-^{(j,d)}-1}^{(j,d)} = 1 - \min\{F^{(j,d;u)}_{s^{(j,d;u)}_{l_-^{(j,d)}-1}}, F^{(j,d;v)}_{s^{(j,d;v)}_{l_-^{(j,d)}-1}}\}. \tag{B.52}$$

Note that if $l_+^{(j,d)} = l_{\max}^{(j,d)}$, we must also have that $l^2 = l^2_{\max}$, otherwise, we would have, by (B.39), that $(s^{(j,d;u)}_{l^{(j,d)}_{\max}}, s^{(j,d;v)}_{l^{(j,d)}_{\max}}) = (s^{(j',2;1)}_{\hat{l}^2}, s^{(j',2;2)}_{\hat{l}^2})$, where $\hat{l}^2 \neq l^2_{\max}$, which would contradict the first statement of Lemma 24 which we have already shown above.

Therefore, we have that

$$\hat{P}_{l^2_{\max}}^{(j',2)} = z_{l^2_{\max}}^{(j',2)} - z_{l^2_{\max}-1}^{(j',2)} = 1 - z_{l^2_{\max}-1}^{(j',2)} = 1 - \min\{F^{(j',2;1)}_{s^{(2;1)}_{l^2_{\max}-1}}, F^{(j',2;2)}_{s^{(2;2)}_{l^2_{\max}-1}}\}. \tag{B.53}$$

It only remains to show that the two min terms in (B.52) and (B.53) are equal. Since $F^{(j,d;u)} = F^{(j',2;1)}$ and $F^{(j,d;v)} = F^{(j',2;2)}$, we only need to show that $(s^{(j,d;u)}_{l_-^{(j,d)}-1}, s^{(j,d;v)}_{l_-^{(j,d)}-1}) = (s^{(j',2;1)}_{l^2_{\max}-1}, s^{(j',2;2)}_{l^2_{\max}-1})$. The arguments to show this is very similar to those, above, used in showing that the second min in (B.47) is equal to (B.48); we omit the proof for the sake of brevity.

Thus, we have shown that (2.101) holds true for $l^2 = 0, 1, \ldots, l^2_{\max}$. □

B.6 Proof of Lemma 21

PROOF: Since $l_{\max}^{(j,d;u,v)} = l_{\max}^{(j',d;u,v)} = l^2_{\max}$ by Remark 48, we can alleviate the notational burden by simply using throughout this proof $l^2_{\max}$ instead of $l_{\max}^{(j,d;u,v)}$ and $l_{\max}^{(j',d;u,v)}$ and $l^2 \in \{0, 1, \ldots, l^2_{\max}\}$ to index the sets $\underline{\mathbf{s}}^{(j,d;u,v)}$, $\underline{\mathbf{s}}^{(j',d;u,v)}$, $\tilde{\underline{\mathbf{s}}}^{(j,d;u,v)}$, $\tilde{\underline{\mathbf{s}}}^{(j',d;u,v)}$, $\mathbf{z}^{(j,2;u,v)}$, and $\mathbf{z}^{(j',2;u,v)}$.

Consider first the u component of the left side of (2.100). By Corollary 2, $\tilde{\underline{s}}_{l^2}^{(j,d;u)}$ is the number of $i \in \{0, 1, \ldots, i_{\max}^{(d;u)}\}$ such that $F_i^{(j,d;u)} = z_l^{(j,2;u,v)}$ for some $l \in \{0, 1, \ldots, l^2 - 1\}$. Since Algorithm 6 is not invoked in the $e^{(j,d;u,v)} = (0, 0)$ case, it must be that $\tilde{\underline{s}}_{l^2}^{(j,d;u)} = \underline{s}_{l^2}^{(j,d;u)}$. Whence, $\underline{s}_{l^2}^{(j,d;u)}$ must be the number of $i \in \{0, 1, \ldots, i_{\max}^{(d;u)}\}$ such that $F_i^{(j,d;u)} = z_l^{(j,2;u,v)}$ for some $l \in \{0, 1, \ldots, l^2 - 1\}$.

Next, consider the right side of (2.100). For any $l^2 \in \{0, 1, \ldots, l^2_{\max}\}$, there exists, by arguments similar to those made in Remark 54, at least one $l^{(j',d)} \in \{0, 1, \ldots, l_{\max}^{(j',d)}\}$ such that

$$\underline{s}_{l^2_{\max}-l^2}^{(j',d;u,v)} = s^{(j',d;u,v)}_{l_{\max}^{(j',d)}-l^{(j',d)}}. \tag{B.54}$$

Since Line 5 of Algorithm 6 is invoked in the $e^{(j',d;u,v)} = (1, 1)$ case, we can rewrite the u component of the right side of (B.54) as

$$s^{(j',d;u)}_{l_{\max}^{(j',d)}-l^{(j',d)}} = i_{\max}^{(d;u)} - \tilde{s}^{(j',d;u)}_{l_{\max}^{(j',d)}-l^{(j',d)}}. \tag{B.55}$$

Applying Remark 54 to the right side of (B.55) and substituting that into the right side of (B.54), considering only the u component, gives

$$\underline{s}_{l^2_{\max}-l^2}^{(j',d;u)} = i_{\max}^{(d;u)} - \tilde{\underline{s}}_{l^2_{\max}-l^2}^{(j',d;u)}. \tag{B.56}$$

Note that, since the $l^{(j',d)}$ in both sides of (B.55) are the same, the l^2 in both sides of (B.56) must also

be the same. Next, we show that $\tilde{\underline{s}}^{(j',d;u)}_{l^2_{\max}-l^2}$ in the right side of (B.56) can be written as

$$\tilde{\underline{s}}^{(j',d;u)}_{l^2_{\max}-l^2} = i^{(d;u)}_{\max} - \underline{s}^{(j,d;u)}_{l^2}. \tag{B.57}$$

By Corollary 2, $\tilde{\underline{s}}^{(j',d;u)}_{l^2_{\max}-l^2}$ in the left side of (B.57) is the total number of $i \in \{0,1,\dots,i^{(d;u)}_{\max}\}$ such that $F^{(j',d;u)}_i = z^{(j',2;u,v)}_l$ for some $l \in \{0,1,\dots,l^2_{\max}-l^2-1\}$.

Next, we can use the fact that $i^{(d;u)}_{\max} = \tilde{\underline{s}}^{(j,d;u)}_{l^2_{\max}}$ by Remark 53 and the fact that $\tilde{\underline{s}}^{(j,d;u,v)}_{l^2} = \underline{s}^{(j,d;u,v)}_{l^2}$ in the $e^{(j,d;u,v)} = (0,0)$ case to rewrite the right side of (B.57) as

$$i^{(d;u)}_{\max} - \underline{s}^{(j,d;u)}_{l^2} = \tilde{\underline{s}}^{(j,d;u)}_{l^2_{\max}} - \tilde{\underline{s}}^{(j,d;u)}_{l^2}. \tag{B.58}$$

Then, it is clear by Corollary 2 that the right side of (B.58) is the total number of $i \in \{0,1,\dots,i^{(d;u)}_{\max}\}$ such that $F^{(j,d;u)}_i = z^{(j,2;u,v)}_l$ for some $l \in \{l^2,\dots,l^2_{\max}-1\}$. Since, by Lemma 18, the order in which the $F^{(j,d;u)}$ is matched in Algorithm 5 for the $e^{(j,d;u,v)} = (0,0)$ case is reversed in Algorithm 5 for the $e^{(j',d;u,v)} = (1,1)$ case, the number of $i \in \{0,1,\dots,i^{(d;u)}_{\max}\}$ such that $F^{(j,d;u)}_i = z^{(j,2;u,v)}_l$ for some $l \in \{l^2, l^2+1,\dots,l^2_{\max}-1\}$ must be equal to the number of $i \in \{0,1,\dots,i^{(d;u)}_{\max}\}$ such that $F^{(j',d;u)}_i = z^{(j',2;u,v)}_l$ for some $l \in \{0,1,\dots,l^2_{\max}-l^2-1\}$. Therefore, we have shown that both the left side and right side of (B.57) equal the total number of $i \in \{0,1,\dots,i^{(d;u)}_{\max}\}$ such that $F^{(j',d;u)}_i = z^{(j',2;u,v)}_l$ for some $l \in \{0,1,\dots,l^2_{\max}-l^2-1\}$, whence, (B.57) holds true. Finally, substituting (B.57) into (B.56) gives

$$\begin{aligned}\underline{s}^{(j',d;u)}_{l^2_{\max}-l^2} &= i^{(d;u)}_{\max} - (i^{(d;u)}_{\max} - \underline{s}^{(j,d;u)}_{l^2}) \\ &= \underline{s}^{(j,d;u)}_{l^2}.\end{aligned} \tag{B.59}$$

as required.

Therefore, we have shown that (2.100) holds true for the u component. Since the proof that (2.100) holds true for the v component is similar, we omit it for brevity. Because (2.100) holds true for both u and v components, we must have that (2.100) holds true. □

B.7 Lemma 24

Before we prove Lemma 24, we show some auxiliary results that will make the exposition in the proof easier to follow. We begin by introducing the sets $\{\hat{s}^{(j',2)}_0,\dots,\hat{s}^{(j',2)}_{l^{(j,d)}_{\max}}\}$, $\{\hat{z}^{(j',2)}_0,\dots,\hat{z}^{(j',2)}_{l^{(j,d)}_{\max}}\}$, and $\{\hat{z}^{(j,d;u,v)}_0,\dots,\hat{z}^{(j,d;u,v)}_{l^{(j,d)}_{\max}}\}$. The sets $\{\hat{s}^{(j',2)}_0,\dots,\hat{s}^{(j',2)}_{l^{(j,d)}_{\max}}\}$ and $\{\hat{z}^{(j',2)}_0,\dots,\hat{z}^{(j',2)}_{l^{(j,d)}_{\max}}\}$ are constructed by duplicating (where appropriate) points in the set of support points $\{s^{(j,2)}_0,\dots,s^{(j,2)}_{l^2_{\max}}\}$ and in the partition of unity $\{z^{(j',2)}_0,\dots,z^{(j',2)}_{l^2_{\max}}\}$, respectively, according to Algorithm 25 below.

Algorithm 25 Subroutine: Extend sets constructed in the $l^2_{\max}$ setting to $l^{(j,d)}_{\max}$

$\hat{s}^{(j',2)}_0 \leftarrow s^{(j',2)}_0$, $\hat{z}^{(j',2)}_0 = z^{(j',2)}_0$
$i \leftarrow 0$
for $l^{(j,d)} \leftarrow 1 : l^{(j,d)}_{\max}$ **do**
  if $\text{Proj}_{u,v}\ s^{(j,d)}_{l^{(j,d)}} \neq \text{Proj}_{u,v}\ s^{(j,d)}_{l^{(j,d)}-1}$ **then**
    $i \leftarrow i+1$
  $\hat{s}^{(j',2)}_{l^{(j,d)}} = s^{(j',2)}_i$, $\hat{z}^{(j',2)}_{l^{(j,d)}} = z^{(j',2)}_i$

Clearly, from Algorithm 25, the sets $\{s_0^{(j,d)},\dots,s_{l^{(j,d)}_{\max}}^{(j,d)}\}$, $\{\hat{s}_0^{(j',2)},\dots,\hat{s}_{l^{(j,d)}_{\max}}^{(j',2)}\}$, and $\{\hat{z}_0^{(j',2)},\dots,\hat{z}_{l^{(j,d)}_{\max}}^{(j',2)}\}$ have the same length. Note that we extend the set $\{s_0^{(j',2)},\dots,s_{l^2_{\max}}^{(j',2)}\}$ by duplicating points $s_{l^2}^{(j',2)}$ corresponding to $s_{l^{(j,d)}}^{(j,d)}$ if the u^{th} and v^{th} coordinates of $s_{l^{(j,d)}}^{(j,d)}$ are unchanged from $s_{l^{(j,d)}-1}^{(j,d)}$; similarly for $\{z_0^{(j',2)},\dots,z_{l^2_{\max}}^{(j',2)}\}$.

Note, however, that the d-dimensional partition of unity, $\{z_0^{(j,d)},\dots,z_{l^{(j,d)}_{\max}}^{(j,d)}\}$, is not directly comparable to neither $\{z_0^{(j',2)},\dots,z_{l^2_{\max}}^{(j',2)}\}$ nor $\{\hat{z}_0^{(j',2)},\dots,\hat{z}_{l^{(j,d)}_{\max}}^{(j',2)}\}$ since it may contain values from $\{F_i^{(j,d;k)} : k=1,\dots,d \text{ and } i=0,1,\dots,i_{\max}^{(d;k)}\}$ that do not exist in the two-dimensional setting. Therefore, we must construct a set comparable to $\{\hat{z}_0^{(j',2)},\dots,\hat{z}_{l^{(j,d)}_{\max}}^{(j',2)}\}$ for the induction statement. To this end, let $\{\bar{l}_0,\dots,\bar{l}_{\max}\}\subseteq\{l_0^{(j,d)},\dots,l_{\max}^{(j,d)}\}$ be a set of indices such that $\bar{l}_i<\bar{l}_{i+1}$ and

$$\{z_{\bar{l}_0}^{(j,d)},\dots,z_{\bar{l}_{\max}}^{(j,d)}\}=\cup_{k=1}^{2}\{F_i^{(j',2;k)} : i=0,1,\dots,i_{\max}^{(2;k)}\}, \tag{B.60a}$$

$$z_{\bar{l}_i}^{(j,d)}<z_{\bar{l}_{i+1}}^{(j,d)}. \tag{B.60b}$$

Note that $\bar{l}_{\max}=l_{\max}^{(j,d)}$, since 1 is in the set on the right side of (B.60a) and $z_{l^{(j,d)}_{max}}^{(j,d)}$ is the unique value in $\{z_0^{(j,d)},\dots,z_{l^{(j,d)}_{\max}}^{(j,d)}\}$ that is equal to 1.

Now Algorithm 26 below uses the sets $\{\bar{l}_0,\dots,\bar{l}_{\max}\}$ and $\{z_{\bar{l}_0}^{(j,d)},\dots,z_{\bar{l}_{\max}}^{(j,d)}\}$ to construct the set $\{\hat{z}_0^{(j,d;u,v)},\dots,\hat{z}_{l^{(j,d)}_{\max}}^{(j,d;u,v)}\}$.

Algorithm 26 Subroutine: Construct $\hat{z}^{(j,d;uv)}$ from $z^{(j,d)}$

1: Set each $\hat{z}_0^{(j,d;u,v)},\dots,\hat{z}_{\bar{l}_0}^{(j,d;u,v)}\leftarrow z_{\bar{l}_0}^{(j,d)}$
2: $l_r\leftarrow\bar{l}_0$
3: **for** $\bar{l}\leftarrow\bar{l}_1:\bar{l}_{\max}$ **do**
4: Set each $\hat{z}_{l_r+1}^{(j,d;u,v)},\dots,\hat{z}_{\bar{l}}^{(j,d;u,v)}\leftarrow z_{\bar{l}}^{(j,d)}$
5: $l_r\leftarrow\bar{l}$

The set $\{\hat{z}_0^{(j,d;u,v)},\dots,\hat{z}_{l^{(j,d)}_{\max}}^{(j,d;u,v)}\}$ has length $l_{\max}^{(j,d)}$ and only contains values from the partition of unity $\cup_{k=1}^{2}\{F_i^{(j',2;k)} : i=0,1,\dots,i_{\max}^{(2;k)}\}$, where $F^{(j',2;1)}=F^{(j,d;u)}$ and $F^{(j',2;2)}=F^{(j,d;v)}$.

We prove the following key result concerning the sequences $\{s_0^{(j,d)},\dots,s_{l^{(j,d)}_{\max}}^{(j,d)}\}$ and $\{\hat{s}_0^{(j',2)},\dots,\hat{s}_{l^{(j,d)}_{\max}}^{(j',2)}\}$ and the sequences $\{\hat{z}_0^{(j,d;u,v)},\dots,\hat{z}_{l^{(j,d)}_{\max}}^{(j,d;u,v)}\}$ and $\{\hat{z}_0^{(j',2)},\dots,\hat{z}_{l^{(j,d)}_{\max}}^{(j',2)}\}$.

Lemma 28. *For $l^{(j,d)}\in\{0,1,\dots,l_{\max}^{(j,d)}\}$*

$$\text{Proj}_{u,v}\, s_{l^{(j,d)}}^{(j,d)}=s_{l^{(j,d)}}^{(j,d;u,v)}=\hat{s}_{l^{(j,d)}}^{(j',2)}, \tag{B.61a}$$

$$\hat{z}_{l^{(j,d)}}^{(j,d;u,v)}=\hat{z}_{l^{(j,d)}}^{(j',2)}. \tag{B.61b}$$

PROOF: We prove that (B.61) holds for all $l^{(j,d)}\in\{0,1,\dots,l_{\max}^{(j,d)}\}$ by induction on $l^{(j,d)}$, beginning with the base case, $l^{(j,d)}=0$. We have that $s_0^{(j,d)}=(0,\dots,0)$ from Line 4 of Algorithm 5 and $s_0^{(j',2)}=(0,0)$ from Line 3 of Algorithm 1. Moreover, since $\hat{s}_0^{(j',2)}=s_0^{(j',2)}$ from Line 1 of Algorithm 25, (B.61a) holds true for $l^{(j,d)}=0$.

To see that (B.61b) also holds true for $l^{(j,d)} = 0$, note that the two-dimensional partition of unity, $\{z_0^{(j',2)}, \ldots, z_{l_{\max}^2}^{(j',2)}\}$, is computed on Line 4 of Algorithm 1 and therefore that $z_0^{(j',2)} = \min\{F_0^{(j',2;1)}, F_0^{(j',2;2)}\}$. From Line 1 of Algorithm 26, we have that $\hat{z}_0^{(j,d;u,v)} = z_{\bar{l}_0}^{(j,d)}$. From (B.60b), we can conclude that $z_{\bar{l}_0}^{(j,d)}$ is the smallest element in the set on the right side of (B.60a), whence $\hat{z}_0^{(j,d;u,v)} = z_{\bar{l}_0}^{(j,d)} = \min\{F_0^{(j',2;1)}, F_0^{(j',2;2)}\}$. Since $\hat{z}_0^{(j',2)} = z_0^{(j',2)} = \min\{F_0^{(j',2,1)}, F_0^{(j',2,2)}\}$ from Algorithm 25, $\hat{z}_0^{(j,d;u,v)} = \hat{z}_0^{(j',2)}$. Hence, (B.61b) also holds true for $l^{(j,d)} = 0$.

For the induction step, choose any $l^{(j,d)} \in \{1, 2, \ldots, l_{\max}^{(j,d)}\}$ and, for the induction hypothesis, assume that (B.61) holds for $l^{(j,d)} - 1$. That is, that $\text{Proj}_{u,v}\, s_{l^{(j,d)}-1}^{(j,d)} = \hat{s}_{l^{(j,d)}-1}^{(j',2)}$ and $\hat{z}_{l^{(j,d)}-1}^{(j,d;u,v)} = \hat{z}_{l^{(j,d)}-1}^{(j',2)}$. Given the current support point $s_{l^{(j,d)}-1}^{(j,d)}$, Lines 6-13 of Algorithm 5 increments the u^{th} and/or the v^{th} coordinates of $s_{l^{(j,d)}-1}^{(j,d)}$ if and only if

$$z_{l^{(j,d)}-1}^{(j,d)} = F_{s_{l^{(j,d)}-1}^{(j,d;u)}}^{(j,d;u)} \quad \text{and/or} \quad z_{l^{(j,d)}-1}^{(j,d)} = F_{s_{l^{(j,d)}-1}^{(j,d;v)}}^{(j,d;v)}.$$

Consequently, there are four cases to consider:

Case 1: $z_{l^{(j,d)}-1}^{(j,d)} == F_{s_{l^{(j,d)}-1}^{(j,d;u)}}^{(j,d;u)}$ and $z_{l^{(j,d)}-1}^{(j,d)} == F_{s_{l^{(j,d)}-1}^{(j,d;v)}}^{(j,d;v)}$

In this case, Line 9 of Algorithm 5 evaluates to true for both the u^{th} and the v^{th} coordinates, leading to Line 10 being executed for both $k = u$ and $k = v$. Therefore,

$$s_{l^{(j,d)}}^{(j,d;u)} = s_{l^{(j,d)}-1}^{(j,d;u)} + 1, \tag{B.62a}$$
$$s_{l^{(j,d)}}^{(j,d;v)} = s_{l^{(j,d)}-1}^{(j,d;v)} + 1. \tag{B.62b}$$

By the induction hypothesis (B.61), Algorithm 25, and Algorithm 26, we have that

$$s_{l^{(j,d)}-1}^{(j,d;u,v)} = \hat{s}_{l^{(j,d)}-1}^{(j',2)} = s_{i-1}^{(j',2)}, \tag{B.63a}$$
$$z_{l^{(j,d)}-1}^{(j,d)} = \hat{z}_{l^{(j,d)}-1}^{(j,d;u,v)} = \hat{z}_{l^{(j,d)}-1}^{(j',2)} = z_{i-1}^{(j',2)} \tag{B.63b}$$

where $i-1$ is the most recent iteration for Algorithm 1. In the case of (B.63b), note also that $z_{l^{(j,d)}-1}^{(j,d)} = \hat{z}_{l^{(j,d)}-1}^{(j,d;u,v)}$ from the assumptions in Case 1, (B.60a), and Definition 26.

From the assumption for Case 1 that $z_{l^{(j,d)}-1}^{(j,d)} = F_{s_{l^{(j,d)}-1}^{(j,d;u)}}^{(j,d;u)} = F_{s_{l^{(j,d)}-1}^{(j,d;v)}}^{(j,d;v)}$, the fact that $F_{s_{l^{(j,d)}-1}^{(j,d;u)}}^{(j,d;u)} = F_{\hat{s}_{l^{(j,d)}-1}^{(j',2;1)}}^{(j',2;1)} = F_{s_{i-1}^{(j',2;1)}}^{(j',2;1)}$ and $F_{s_{l^{(j,d)}-1}^{(j,d;v)}}^{(j,d;v)} = F_{\hat{s}_{l^{(j,d)}-1}^{(j',2;2)}}^{(j',2;2)} = F_{s_{i-1}^{(j',2;2)}}^{(j',2;2)}$, the induction hypothesis (B.61), and (B.63b), we must also have that

$$z_{i-1}^{(j',2)} = F_{s_{i-1}^{(j',2;1)}}^{(j',2;1)} \quad \text{and} \quad z_{i-1}^{(j',2)} = F_{s_{i-1}^{(j',2;2)}}^{(j',2;2)}.$$

Consequently, Line 8 of Algorithm 1 must evaluate to true, leading to the execution of Lines 9 and 10 of Algorithm 1. Therefore,

$$s_i^{(j',2)} = (s_{i-1}^{(j',2;1)} + 1, s_{i-1}^{(j',2;2)} + 1). \tag{B.64}$$

In this case, since Line 4 of Algorithm 25 evaluates to true, Line 5 gets executed and we have that

$$\hat{s}^{(j',2)}_{l^{(j,d)}} = s^{(j',2)}_{i-1+1} = s^{(j',2)}_{i}, \tag{B.65a}$$
$$\hat{z}^{(j',2)}_{l^{(j,d)}} = z^{(j',2)}_{i-1+1} = z^{(j',2)}_{i}. \tag{B.65b}$$

Therefore, by (B.62), (B.63), (B.64), and (B.65), (B.61) holds true for $l^{(j,d)}$ as well.

Case 2: $z^{(j,d)}_{l^{(j,d)}-1} == F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}-1}}$ and $z^{(j,d)}_{l^{(j,d)}-1} \neq F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}-1}}$

In this case, Line 9 of Algorithm 5 evaluates to true for the u^{th} coordinate, but not for the v^{th} coordinate, leading to Line 10 being executed for the u^{th} coordinate, but Line 13 being executed for the v^{th} coordinate. Therefore,

$$s^{(j,d;u)}_{l^{(j,d)}} = s^{(j,d;u)}_{l^{(j,d)}-1} + 1, \tag{B.66a}$$
$$s^{(j,d;v)}_{l^{(j,d)}} = s^{(j,d;v)}_{l^{(j,d)}-1}. \tag{B.66b}$$

By the induction hypothesis, Algorithm 25 and Algorithm 26, we have that

$$s^{(j,d;u,v)}_{l^{(j,d)}-1} = \hat{s}^{(j',2)}_{l^{(j,d)}-1} = s^{(j',2)}_{i-1}, \tag{B.67a}$$
$$z^{(j,d)}_{l^{(j,d)}-1} = \hat{z}^{(j,d;u,v)}_{l^{(j,d)}-1} = \hat{z}^{(j',2)}_{l^{(j,d)}-1} = z^{(j',2)}_{i-1} \tag{B.67b}$$

where $i-1$ is the most recent iteration for Algorithm 1. In the case of (B.67b), note also that $z^{(j,d)}_{l^{(j,d)}-1} = z^{(d;u,v)}_{l^{(j,d)}-1}$ from the assumptions in Case 2, (B.60a), and Definition 26.

From the assumption for Case 2 that $z^{(j,d)}_{l^{(j,d)}-1} == F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}-1}}$ and $z^{(j,d)}_{l^{(j,d)}-1} \neq F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}-1}}$, the fact that $F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}-1}} = F^{(j',2;1)}_{\hat{s}^{(j',2;1)}_{l^{(j,d)}-1}} = F^{(j',2;1)}_{s^{(j',2;1)}_{i-1}}$ and $F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}-1}} = F^{(j',2;2)}_{\hat{s}^{(j',2;2)}_{l^{(j,d)}-1}} = F^{(j',2;2)}_{s^{(j',2;2)}_{i-1}}$, the induction hypothesis (B.61), and (B.67b), we must also have that

$$z^{(j',2)}_{i-1} = F^{(j',2;1)}_{\hat{s}^{(j',2;1)}_{i-1}} \quad \text{and} \quad z^{(j',2)}_{i-1} \neq F^{(j',2;2)}_{\hat{s}^{(j',2;2)}_{i-1}}.$$

Consequently, Line 11 of Algorithm 1 must hold true, leading to the execution of Lines 12 and 13 of Algorithm 1. Therefore,

$$s^{(j',2)}_{i} = (s^{(j',2;1)}_{i-1} + 1, s^{(j',2;2)}_{i-1}). \tag{B.68}$$

In this case, since Line 4 of Algorithm 25 evaluates to true, Line 5 gets executed and we have that

$$\hat{s}^{(j',2)}_{l^{(j,d)}} = s^{(j',2)}_{i-1+1} = s^{(j',2)}_{i}, \tag{B.69a}$$
$$\hat{z}^{(j',2)}_{l^{(j,d)}} = z^{(j',2)}_{i-1+1} = z^{(j',2)}_{i}. \tag{B.69b}$$

Therefore, by (B.66), (B.67), (B.68), and (B.69), (B.61) holds true for $l^{(j,d)}$ as well.

Case 3: $z^{(j,d)}_{l^{(j,d)}-1} \neq F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}-1}}$ and $z^{(j,d)}_{l^{(j,d)}-1} == F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}-1}}$

The proof that (B.61) holds true in this case is very similar to the proof in Case 2. For the sake of brevity, we omit the proof in this case.

Case 4: $z^{(j,d)}_{l^{(j,d)}-1} \neq F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}-1}}$ and $z^{(j,d)}_{l^{(j,d)}-1} \neq F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}-1}}$

In this case, Line 9 of Algorithm 5 does not hold true for neither the u^{th} nor the v^{th} coordinate. Hence,

$$s^{(j,d;u)}_{l^{(j,d)}} = s^{(j,d;u)}_{l^{(j,d)}-1}, \tag{B.70a}$$

$$s^{(j,d;v)}_{l^{(j,d)}} = s^{(j,d;v)}_{l^{(j,d)}-1}. \tag{B.70b}$$

By the induction hypothesis, Algorithm 25, and Algorithm 26, we have that

$$s^{(j,d;u,v)}_{l^{(j,d)}-1} = \hat{s}^{(j',2)}_{l^{(j,d)}-1} = s^{(j',2)}_{i-1}, \tag{B.71a}$$

$$\hat{z}^{(j,d;u,v)}_{l^{(j,d)}-1} = \hat{z}^{(j',2)}_{l^{(j,d)}-1} = z^{(j',2)}_{i-1}. \tag{B.71b}$$

where $i-1$ is the most recent iteration for Algorithm 1.

From the assumption for Case 4 that $z^{(j,d)}_{l^{(j,d)}-1} \neq F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}-1}}$ and $z^{(j,d)}_{l^{(j,d)}-1} \neq F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}-1}}$, the fact that $F^{(j,d;u)}_{s^{(j,d;u)}_{l^{(j,d)}-1}} = F^{(j',2;1)}_{\hat{s}^{(j',2;1)}_{l^{(j,d)}-1}} = F^{(j',2;1)}_{s^{(j',2;1)}_{i-1}}$ and $F^{(j,d;v)}_{s^{(j,d;v)}_{l^{(j,d)}-1}} = F^{(j',2;2)}_{\hat{s}^{(j',2;2)}_{l^{(j,d)}-1}} = F^{(j',2;2)}_{s^{(j',2;2)}_{i-1}}$, the induction hypothesis (B.61), and (B.71), we must also have that

$$z^{(j',2)}_{i-1} \neq F^{(j',2;1)}_{s^{(j',2;1)}_{i-1}} \quad \text{and} \quad z^{(j',2)}_{i-1} \neq F^{(j',2;2)}_{s^{(j',2;2)}_{i-1}}. \tag{B.72}$$

Note that there is no corresponding condition in Algorithm 1 to (B.72). However, in this case, the projection condition in Line 4 of Algorithm 25 evaluates to false and Line 5 is not executed. Therefore, we have that

$$\hat{s}^{(j',2)}_{l^{(j,d)}} = s^{(j',2)}_{i-1}, \tag{B.73a}$$

$$\hat{z}^{(j',2)}_{l^{(j,d)}} = z^{(j',2)}_{i-1}. \tag{B.73b}$$

Therefore, by (B.70), (B.71), and (B.73), (B.61) holds true for $l^{(j,d)}$ as well.

We have shown by induction on $l^{(j,d)}$ that (B.61) holds true for all $l^{(j,d)} \in \{0, \ldots, l^{(j,d)}_{\max}\}$. □

B.7.1 Proof of Lemma 24

PROOF: To prove the lemma, we show the following two statements:

1. The support of $\tilde{P}^{(j,d;\,u,v)}$ is identical to the support of $\hat{P}^{(j',2)}$.
2. The probabilities at each point of support are equal. That is, for $l^2 \in \{0, 1, \ldots, l^2_{\max}\}$,

$$\tilde{P}^{(j,d;\,u,v)}_{l^2} = \hat{P}^{(j',2)}_{l^2}.$$

We note that, for the remainder of the proof of Lemma 24, we assume that $d > 2$ since we have already shown the two-dimensional case in Section 2.3 and since, for $d = 2$, Algorithm 5 reduces to Algorithm 1. (See Remark 49.) There are four cases to consider.

The $\mathbf{e}^{(1,d)} = (0, \ldots, 0)$ and $\mathbf{e}^{(1,2)} = (0, 0)$ case

We begin proving Lemma 24 for the d-dimensional case $j = 1$ and the corresponding two-dimensional case $j' = 1$. The case $j = 1$ is the d-dimensional analogue of the comonotone case in the bivariate setting with corresponding monotone structure $\mathbf{e}^{(1,d)} = (0, \ldots, 0)$ describing the dependence structure where all coordinates exhibit extreme positive dependence to one another. It follows easily from $\mathbf{e}^{(1,d)}$ that the case $j = 1$ in d-dimensions corresponds to the case $j' = 1$ in two-dimensions.

We begin by proving the first statement listed at the start of the proof of Lemma 24 by induction. Note that Lemma 28 shows that the support computed by Algorithm 5 projected onto (u, v), $\{s_0^{(j,d;u,v)}, \ldots, s_{l_{\max}^{(j,d)}}^{(j,d;u,v)}\}$ (having duplicate points within the set), is equal to the augmented support set $\hat{\mathbf{s}}^{(j',2)} = \{\hat{s}_0^{(j',2)}, \ldots, \hat{s}_{l_{\max}^{(j,d)}}^{(j',2)}\}$. Moreover, since the augmented support, $\hat{\mathbf{s}}^{(j',2)}$, with the repeated points removed is the set $\{s_0^{(j',2)}, \ldots, s_{l_{\max}^2}^{(j',2)}\}$, the set $\{s_0^{(j,d;u,v)}, \ldots, s_{l_{\max}^2}^{(j,d;u,v)}\}$ with the repeated points removed is equal to the set of support points $\{s_0^{(j',2)}, \ldots, s_{l_{\max}^2}^{(j',2)}\}$ obtained from Algorithm 1. Therefore, the first statement at the start of the proof of Lemma 24 follows from Lemma 28 and holds true for the case $j = 1$ and $j' = 1$.

The second statement of the proof of the lemma was shown directly in Lemma 22.

Therefore, Lemma 24 follows from Lemma 28 and Lemma 22 for the case $\mathbf{e}^{(1,d)} = (0, \ldots, 0)$ and $\mathbf{e}^{(1,2)} = (0, 0)$ case.

The $\mathbf{e}^{(j,d;u,v)} = (0, 0)$ and $\mathbf{e}^{(j',d;u,v)} = (1, 1)$ case

In the two-dimensional case, there are two extreme joint distributions: the comonotone case $\mathbf{e}^{(1,2)} = (0, 0)$ and the antimonotone case $\mathbf{e}^{(2,2)} = (0, 1)$. In the general d-dimensional case, however, there may be monotone structures such that their bivariate projections take the form $\mathbf{e}^{(j,d;u,v)} = (1, 1)$ or $\mathbf{e}^{(j,d;u,v)} = (1, 0)$. We show in this subsection that the bivariate joint probability distribution $\tilde{P}^{(j,d;u,v)}$ having the monotone structure $\mathbf{e}^{(j,d;u,v)} = (0, 0)$ is equivalent to the bivariate joint probability distribution $\tilde{P}^{(j',d;u,v)}$ having the monotone structure $\mathbf{e}^{(j',d;u,v)} = (1, 1)$ by showing again that the two statements hold for this case. Later, on page 201, we discuss the $\mathbf{e}^{(j,d;u,v)} = (0, 1)$ and $\mathbf{e}^{(j,d;u,v)} = (1, 0)$ case.

Lemma 21 directly shows the first statement of the proof of Lemma 24 in the $\mathbf{e}^{(j,d;u,v)} = (0, 0)$ and $\mathbf{e}^{(j',d;u,v)} = (1, 1)$ case.

Next, we show the second statement of the proof also holds true in this special case. That is, for all $l^2 \in \{0, 1, \ldots, l_{\max}^2\}$,

$$\tilde{P}_{l^2}^{(j,d;u,v)} = \tilde{P}_{l_{\max}^2 - l^2}^{(j',d;u,v)}. \tag{B.74}$$

We rewrite the left side of (B.74) using Line 18 of Algorithm 1 as

$$\tilde{P}_{l^2}^{(j,d;u,v)} = z_{l^2}^{(j,2)} - z_{l^2-1}^{(j,2)}. \tag{B.75}$$

Since there is a correspondence, by Remark 50, between the sets $\boldsymbol{z}^{(j,2;u,v)}$ and $\boldsymbol{z}^{(j,2)}$, we can rewrite (B.75) as

$$\tilde{P}_{l^2}^{(j,d;u,v)} = z_{l^2}^{(j,2)} - z_{l^2-1}^{(j,2)} = z_{l^2}^{(2;u,v)} - z_{l^2-1}^{(2;u,v)}. \tag{B.76}$$

Using (2.90) from Lemma 17, we can rewrite (B.76) as

$$\begin{aligned}\tilde{P}_{l^2}^{(j,d;u,v)} &= z_{l^2}^{(j,2;u,v)} - z_{l^2-1}^{(j,2;u,v)} \\ &= (1 - z_{l^2_{\max}-l^2-1}^{(j',2;u,v)}) - (1 - z_{l^2_{\max}-l^2}^{(j',2;u,v)}) \\ &= z_{l^2_{\max}-l^2}^{(j',2;u,v)} - z_{l^2_{\max}-l^2-1}^{(j',2;u,v)} \\ &= \tilde{P}_{l^2_{\max}-l^2}^{(j',d;u,v)}.\end{aligned}$$

Therefore, we have shown that (B.74) holds for all $l^2 \in \{0, 1, \ldots, l^2_{\max}\}$.

Since $\underline{s}_{l^2}^{(j,d;u,v)}$ is the support point corresponding to $\tilde{P}_{l^2}^{(j,d;u,v)}$ and $\underline{s}_{l^2_{max}-l^2}^{(j',d;u,v)}$ is the support point corresponding to $\tilde{P}_{l^2_{max}-l^2}^{(j',d;u,v)}$, it follows from (B.74) that $\tilde{P}_{\underline{s}_{l^2}^{(j,d;u,v)}}^{(j,d;u,v)} = \tilde{P}_{\underline{s}_{l^2_{max}-l^2}^{(j',d;u,v)}}^{(j',d;u,v)}$ for all $l^2 \in \{0, 1, 2, \ldots, l^2_{\max}\}$. This together with Lemma 21 shows that the probability distributions $\tilde{P}^{(j,d;u,v)}$ and $\tilde{P}^{(j',d;u,v)}$ are equal. Therefore, we have shown that Lemma 24 holds true for the special case $\mathbf{e}^{(j,d;u,v)} = (0, 0)$ and $\mathbf{e}^{(j',d;u,v)} = (1, 1)$.

The $\mathbf{e}^{(j,d;u,v)} = (0, 1)$ and $\mathbf{e}^{(j',d;u,v)} = (1, 0)$ case

This case can be proved using arguments similar to the $\mathbf{e}^{(j,d;u,v)} = (0, 0)$ and $\mathbf{e}^{(j',d;u,v)} = (1, 1)$ case shown above. For this reason, we omit its proof.

The general $1 < j \leq n$ setting

Pick a $j \in \{2, \ldots, n\}$ and choose a bivariate set of antimonotone coordinates (u, v) where $1 \leq u < v \leq d$ from the corresponding monotone structure $\mathbf{e}^{(j,d)}$. Recall that to construct the antimonotone distribution $\hat{P}^{(2,2)}$ in Algorithm 1, a preprocessing step is applied to the marginal distribution $Q^{(2;2)}$. The modified $Q^{(2;2)}$ is then used to construct $F^{(2;2)}$ (Lines 3-5 of Algorithm 3). After the execution of the main loop (Lines 6-18), a postprocessing step is applied to the support of $\hat{P}^{(2,2)}$ (Lines 19-20). As we have already shown that (B.61) and (2.101) hold true for the comonotone case $j = j' = 1$, we only need to show that the preprocessing of the marginal distributions and the postprocessing of the antimonotone supports in Algorithm 5 is the same as the postprocessing in Algorithm 1. This can be clearly seen by comparing Algorithm 3 and Algorithm 7 for the preprocessing steps and Algorithm 2 and Algorithm 6 for the postprocessing steps. Therefore, Lemma 24 holds true for $j \in \{2, \ldots, d\}$.

□