Correlated Multivariate Poisson Processes and Extreme Measures

Michael Chiu[1], Kenneth R. Jackson[1], and Alexander Kreinin[2]

[1]Department of Computer Science, University of Toronto
[2]Quantitative Research, Risk Analytics, IBM Canada

June 23, 2017

Abstract

Multivariate Poisson processes have many important applications in Insurance, Finance, and many other areas of Applied Probability. In this paper we study the backward simulation approach to modelling multivariate Poisson processes and analyze the connection to the extreme measures describing the joint distribution of the processes at the terminal simulation time.

1 Introduction

Analysis and simulation of dependent Poisson processes is an important problem having many applications in Insurance, Finance, Operational Risk modelling and many other areas (see [1], [3], [4], [6], [9], [13], [5] and references therein). In the modelling of multivariate Poisson processes, the specification of the dependence structure is an intriguing problem. In the literature, several different bivariate processes with Poisson marginals are available for applications in actuarial science and quantitative risk management. One of the most popular models is the common shock model where a common Poisson process drives the dependence between independent Poisson marginal distribution. The resulting correlation structure is time invariant and cannot exhibit negative correlations.

An alternative, more flexible approach to this problem is based on the Backward Simulation (BS) introduced in [8]. The BS of correlated Poisson processes and an approach to the calibration problem using transformations of Gaussians variables was proposed in [5]. In [8], the idea of BS was extended to the class of multivariate processes containing both Poisson and Wiener components. It was also proved that the linear time structure of correlations is observed both in the Poisson and the Poisson-Wiener model. Further steps in the bivariate case were proposed in [2] where the BS was combined with copula functions. This method allows one to extend the correlation pattern by using the Marshall-Olkin type copula functions that are simple to simulate.

In the present paper, we continue the analysis and development of the BS method for the class of multivariate Poisson processes. By the multivariate Poisson process, we understand any vector-valued process such that all its components are (single-dimensional) Poisson processes. The idea of our approach is to use the relationship between the extreme measures describing the joint distribution with maximal or minimal correlation coefficient of the components of the multivariate process at the terminal simulation time and the time structure of correlations. We describe the class of admissible correlation structures given parameters of the marginal Poisson processes and exploit convex combinations of the extreme measures to represent the multivariate Poisson process with given correlations of the components. We believe that our approach can simplify the solution to the calibration problem and extend the variety of the correlation patterns of the multivariate Poisson processes.

There is a connection of our problem to the Optimal Transport literature (see [14] for a general overview of the area and [11, 12] for a more probabilistic focus). Our computation of the extreme measures at the terminal simulation time can be viewed as a solution to a special multi-objective Monge-Kantorovich Mass Transportation (MKP) Problem with the quadratic cost functions. However, this connection is not discussed in the present paper. [1]

The rest of the paper is organized as follows. In Section 2 we begin by discussing the background and motivation for the 2-dimensional problem. We introduce extreme measures and generalize the results of the bivariate problem to higher dimensions in Section 3. In Section 4 we describe a general algorithm for the computation of the joint distribution of the extreme measures. Section 5 is concerned with the calibration problem. We discuss the simulation problem in Section 6 and propose a Forward-Backward extension of the BS method. The paper is concluded with some directions for future research in Section 7.

2 Extreme Measures and Monotonicity of the Joint Distributions

We begin with a description of the Common Shock Model (CSM) and the motivation of the approach proposed in [5]. Afterwards, we discuss the results obtained in [8] for the case of two Poisson processes and describe the computation of the extreme measures in the case $J = 2$.

The CSM has become very popular within actuarial applications as well as in Operational Risk modeling [10]. This model is based on the following idea. Suppose we want to construct two dependent Poisson processes. Consider three independent Poisson processes $\nu_t^{(1)}$, $\nu_t^{(2)}$, $\nu_t^{(3)}$ with the intensities λ_1, λ_2, λ_3. Let $N_t^{(1)} = \nu_t^{(1)} + \nu_t^{(2)}$ and $N_t^{(2)} = \nu_t^{(3)} + \nu_t^{(2)}$, which are also Poisson processes, formed by the superposition operation. Then, the Poisson processes $N^{(1)}(t)$ and $N^{(2)}(t)$ are dependent with the Pearson correlation coefficient

$$\rho(N_t^{(1)}, N_t^{(2)}) = \frac{\lambda_2}{\sqrt{(\lambda_1 + \lambda_2)(\lambda_2 + \lambda_3)}}.$$

Clearly, the correlation coefficient can only be positive.

A more advanced approach to the construction of negatively correlated Poisson processes is based on the idea of the backward simulation of the Poisson processes [8]. The conditional distribution of the arrival moments of a Poisson process, conditional on the value of the process at the terminal simulation time, T, is uniform. Then, using a joint distribution maximizing or minimizing correlation between the components at time, T, one can construct a Poisson process with a linear time structure of correlations in the interval $t \in [0, T]$. Thus, the problem of constructing the 2-dimensional Poisson process with the extreme correlation of the components at time T is reduced to that of random variables having Poisson distributions with the parameters λT and μT, where λ and μ are parameters of the processes. It is not difficult to see that maximization (minimization) of the correlation coefficient of two random variables (r.v.), X and Y, given their marginal distributions, is equivalent to maximization (minimization) of $\mathbb{E}[XY]$, if the r.v. have finite first and second moments and positive variances.

The admissible range of the correlation coefficients can be computed using the Extreme Joint Distributions (EJD) Theorem in [8] (see Theorem 2.2 in this section). The key statement, the characterization of the EJDs, is equivalent to the Frechet-Hoeffding theorem [7] for distributions on the positive quadrant of the two-dimensional lattice, $\mathbb{Z}_{++}^{(2)} = \{(i, j) : i, j = 0, 1, 2, \dots\}$. However, taking into account the numerical aspect of the problem, we prefer to use equations, derived in [8], written in terms of the probability density function, not in terms of the cumulative distribution function. Given marginal distributions of the non-negative, integer-valued random variables X_1 and X_2, with finite first and second moments, there exist two joint distributions, $F^*(i, j)$ and $F^{**}(i, j)$ minimizing and maximizing the correlation, $\rho = \mathrm{corr}(X_1, X_2)$, respectively.

[1] We are mainly concerned with the construction of the multivariate Poisson processes.

Definition 2.1. *The probability measures corresponding to the joint distributions F^* and F^{**} are called extreme probability measures.*

The EJD theorem in [8] allows one to construct the extreme measures p^* and p^{**}, given marginal distributions of X_1 and X_2, with the minimal negative correlation ρ^* and maximal positive correlation ρ^{**}, respectively. The extreme correlation coefficient uniquely defines the extreme measure.

Given a probability measure, p, corresponding to the joint distribution of the vector (X_1, X_2) on $\mathbb{Z}^{(2)}_{++}$ we define a functional $f_\rho(p) = \text{corr}(X_1, X_2)$. Then we have $\rho^* = f_\rho(p^*)$, and $\rho^{**} = f_\rho(p^{**})$. This functional f_ρ preserves the convex combination property. Indeed, taking a convex combination of the extreme measures, $p = \theta p^* + (1-\theta)p^{**}$, $(0 \le \theta \le 1)$, we obtain

$$f_\rho(p) = \theta f_\rho(p^*) + (1-\theta) f_\rho(p^{**}). \tag{1}$$

Connection to Optimization Problem

Computation of the extreme measures in the case $J = 2$ was accomplished in [8] using a very efficient EJD algorithm having linear complexity with respect to the number of points in the support of the marginal distributions. It is interesting to note that this algorithm is applicable to a more general class of linear optimization problems on a lattice. In the case $J > 2$, the corresponding optimization problem becomes multi-objective with $M = J(J-1)/2$ objective functions. Let us first recall the case $J = 2$.

Let (X_1, X_2) be a random vector with support $\mathbb{Z}^{(2)}_{++}$ and given marginal probabilities $\mathbb{P}(X_1 = i) = P_1(i)$ and $\mathbb{P}(X_2 = j) = P_2(j)$. Denote

$$f(p) := \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} ij\, p(i,j)$$

where $p(i,j) = \mathbb{P}(X_1 = i, X_2 = j)$. The objective function $f(p) = \mathbb{E}[X_1 X_2]$. The measure p^{**} is the solution to the problem $f(p) \to \max$ with the constraints on the marginal distributions of p^{**}. Similarly, the extreme measure p^* is the solution to the optimization problem $f(p) \to \min$ with the same constraints [8]. For the sake of brevity, we write these two problems as

$$\begin{aligned} & f(p) \to \text{extr} \qquad\qquad (2) \\ \text{s. t.} \quad & \\ & \sum_{j=0}^{\infty} p(i,j) = P_1(i), \quad i = 0, 1, \dots \\ & \sum_{i=0}^{\infty} p(i,j) = P_2(j), \quad j = 0, 1, \dots \\ & p(i,j) \ge 0 \quad i, j = 0, 1, 2, \dots \end{aligned}$$

where $\sum\limits_{i=0}^{\infty} P_1(i) = \sum\limits_{j=0}^{\infty} P_2(j) = 1$. The symbol extr denotes max in the case of measure p^{**} and min in the case of p^*. It is not difficult to see that Problem (2) is infinite dimensional; its numerical solution requires construction of the compact subset of the lattice for the computation of the approximate solution [8].

A solution to the infinite dimensional optimization problem (2) is the joint distribution describing one of the extreme measures, given the marginal distributions of the random variables. The EJD algorithm discussed in [8] allows one to find a unique solution to the problem with required accuracy. Taking the marginal distributions to be Poissonian, we find the extreme measures, p^* and p^{**}, describing the joint distribution of the processes, $N_T = (N_T^{(1)}, N_T^{(2)})$ with the extreme correlation of the components at time T.

The convex combination of these measures can be calibrated to the desired value of the correlation coefficient, ρ. Then, applying the BS we obtain the sample paths of the processes. Note that the EJD algorithm is applicable to a more general class of linear optimization problems: there is no need to assume normalization conditions as long as $P_1(i) \geq 0$ and $P_2(j) \geq 0$ for all $i \geq 0$ and $j \geq 0$ and these functions are integrable: $\sum_{i=0}^{\infty} P_1(i) < \infty$ and $\sum_{j=0}^{\infty} P_2(j) < \infty$.

Monotone Distributions

Extreme measures are closely connected to the monotone distributions in the case $J = 2$. It was proved in [8] that the joint distribution is comonotone in the case of maximal correlation and antimonotone in the case of minimal (negative) correlation. Let us review the properties of extreme measures used in what follows.

Consider a set $\mathcal{S} = \{s_n\}_{n\geq 0}$, where $s_n = (x_n, y_n) \in \mathbb{R}^2$. Define the two subsets $\mathcal{R}_+ = \{(x, y) \in \mathbb{R}^2 : x \cdot y \geq 0\}$ and $\mathcal{R}_- = \{(x, y) \in \mathbb{R}^2 : x \cdot y \leq 0\}$.

Definition 2.2. *A set $\mathcal{S} = \{s_n\}_{n\geq 0} \subset \mathbb{R}^2$ is comonotone if $\forall\, i, j,\ s_i - s_j \in \mathcal{R}_+$. Similarly, $\mathcal{S}$ is antimonotone if $\forall\, i, j,\ s_i - s_j \in \mathcal{R}_-$.*

Definition 2.3 (Monotone distributions)**.** *We say that a distribution P is comonotone (antimonotone) if its support is a comonotone (antimonotone) set.*

It is also useful to recall the following classical statement on monotone sequences of real numbers, usually attributed to Hardy.

Consider two vectors $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^N$. Their inner product

$$\langle x, y\rangle := \sum_{k=1}^{N} x_k y_k$$

Denote by $\mathfrak{S}_N$ the set of all permutations of N elements.

Lemma 2.1. *For any monotonically increasing sequence, $x_1 \leq x_2 \leq \cdots \leq x_N$ and a vector $y \in \mathbb{R}^N$, there exists permutations π_+ and π_- solving the optimization problems*

$$\langle x, \pi_+ y\rangle = \max_{\pi\in\mathfrak{S}_N} \langle x, \pi y\rangle$$

and

$$\langle x, \pi_- y\rangle = \min_{\pi\in\mathfrak{S}_N} \langle x, \pi y\rangle$$

The permutations π_+ and π_- sort vectors in ascending and descending order, respectively.

Lemma 2.1 motivates the introduction of monotone distributions in the 2-dimesional case.

Theorem 2.2 ([8])**.** *The joint distribution p^{**} for X_1 and X_2 having maximal positive correlation coefficient ρ^{**}, given marginal distributions $F_1(i)$ and $F_2(j)$, is comonotone. The probabilities $p^{**}(i, j) = \mathbb{P}(X_1 = i, X_2 = j)$ satify the equation*

$$p^{**}(i, j) = [\min(F_1(i), F_2(j)) - \max(F_1(i-1), F_2(j-1))]^+ \quad i, j = 0, 1, 2, \ldots \tag{3}$$

where $[\,x\,]^+ = \max(x, 0)$ and $F_i(\cdot)$ denote the marginal CDFs, with $F_i(-1) = 0$.

The joint distribution p^* for X_1 and X_2 having minimal negative correlation coefficient ρ^* is antimonotone. In this case

$$p^*(i, j) = [\min(F_1(i), \bar{F}_2(j-1)) - \max(F_1(i-1), \bar{F}_2(j))]^+ \quad i, j = 0, 1, 2, \ldots \tag{4}$$

where $\bar{F}_i(j) = 1 - F_i(j)$ and $\bar{F}_i(-1) = 1$.

Theorem 2.2 is equivalent to the Frechet theorem in the case the marginal distributions are discrete.

The case of the Poisson marginal distributions is a particular case of Theorem 2.2. This result is applicable to much more general classes of distributions. In particular, one can describe the joint probabilities corresponding to p^* and p^{**} in the case the components of the vector have a negative binomial distribution. The EJD algorithm for computation of the joint probabilities is also applicable to more general cases. If both marginal distributions have finite second moments, the joint distribution can be approximated with a desired accuracy.

3 Extreme Measures in Higher Dimensions

Let us now generalize the main result, Theorem 2.2, discussed in Section 2. We consider a random vector $\vec{X} = (X_1, \dots, X_J)$ on a positive quadrant of the J-dimensional lattice, $\mathbb{Z}_+^{(J)}$. Each coordinate of $\vec{X}$ has a discrete distribution with the support $\mathbb{Z}_+$. We also assume that the random variable X_k has finite second moment and that the variance is positive. In this case, the correlation coefficients, $\rho_{k,l} = \text{corr}(X_k, X_l)$, are defined for all $1 \le k \le l \le J$. We denote the marginal distribution of the r.v. X_k by F_k:

$$F_k(i) = \mathbb{P}(X_k \le i), \quad i \in \mathbb{Z}_+; \quad k = 1, 2, \dots, J.$$

Let us now define the extreme measures on the J-dimensional lattice. If $J = 2$, the extreme measures are described by the joint distribution maximizing and minimizing the correlation coefficient of X_1 and X_2; the corresponding probability density functions satisfy Theorem 2.2. If the number of components $J \ge 3$, the definition of the extreme measure is less obvious.

Denote the (joint) distribution function of $\vec{X}$ by $F(\vec{i})$: $F(i_1, i_2, \dots, i_J) = \mathbb{P}(X_1 \le i_1, X_2 \le i_2, \dots, X_J \le i_J)$ and denote the corresponding probability density function by $p(\vec{i})$. By $p_{k,l}(i_k, i_l)$ we denote the probability density function of the 2-dimensional projection, (X_k, X_l) of $\vec{X}$, $(1 \le k < l \le J)$.

Definition 3.1. *We say that the density* $p(\mathbf{i})$*,*

$$p(i_1, \dots, i_J) = \mathbb{P}(X_1 = i_1, \dots, X_J = i_J), \quad i_k \in \mathbb{Z}_+, \quad k = 1, 2, \dots, J$$

determines an extreme measure on the J*-dimensional lattice if and only if for all* k *and* l*,* $(1 \le k \le l \le J)$*, the associated density* $p_{k,l}$ *determines an extreme measure on* $\mathbb{Z}_{++}^{(2)}$ *in the sense of Definition 2.1.*

Our goal is to describe the extreme measures given the marginal distributions, F_k, and compute the correlation matrices, $\rho = [\rho_{k,l}]$. Let us first find the number of extreme measures.

Lemma 3.1. *For any given set of marginal distributions,* F_k*, on* $\mathbb{Z}_+$ $(k = 1, 2, \dots, J)$ *the number of extreme measures* $N = 2^{J-1}$*.*

Proof. The proof of Lemma 3.1 for $J = 2$ is obvious. Let us prove it for $J \ge 3$. For each 2-dimensional projection (X_k, X_l), the corresponding joint distribution should be either comonotone or antimonotone. Take the first r.v, X_1, and form the first group of random variables from the set X_2, X_3, $\dots, X_J$, that are comonotone with X_1. Denote the number of comonotone r.v. by J_c. The number of r.v. antimonotone with X_1, satisfies

$$J_a = J - 1 - J_c.$$

The total number of partitions of the number $J-1$ in the additive form, $J-1 = J_a + J_c$, is $N = 2^{J-1}$. Clearly, N does not depend on the choice of the first r.v. □

Let us now introduce the monotonicity structure of the extreme measures. Take the first r.v., X_1 and consider the r.v. X_2, X_3, $\dots X_J$. Define the vector of binary variables $\mathbf{e} = (e_1, e_2, \dots, e_J)$ such that $e_1 = 0$, and for $j = 2, 3, \dots, J$

$$e_j = \begin{cases} 1, & \text{if } X_1 \text{ and } X_j \text{ are antimonotone,} \\ 0, & \text{if } X_1 \text{ and } X_j \text{ are comonotone.} \end{cases}$$

We call $\mathbf{e}$ the monotonicity vector; its components are called monotonicity indicators. Figure 1 illustrates this concept. In this example, all coordinates but the last are comonotone with the first r.v., X_1. The last coordinate, X_J is antimonotone. The monotonicity indicators in this case are $e_k = 0$, for $k = 1, 2, \dots, J-1$, and $e_J = 1$.

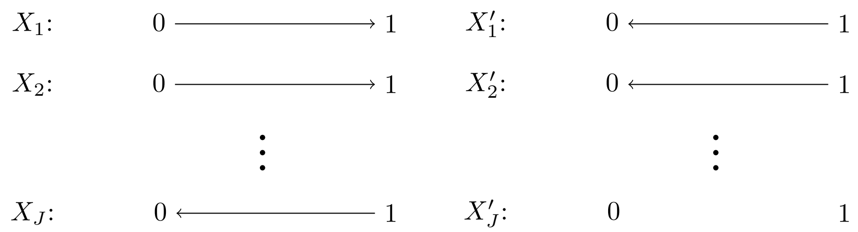

Figure 1: Illustration of the monotonicity structure of an extreme measure. Each distribution is represented by an arrow having unit length. All arrows associated with $X_2, \ldots, X_{J-1}$ are oriented in the same direction as the arrow representing X_1. The last arrow pointing in the opposite direction indicates antimonotonicity of the random variables X_1 and X_J. The monotonicity structure on the right has all of its arrows reversed compared to that on the left. However, note that they both represent the same monotonicity structure.

Optimization Problem: $J \geq 3$.

Since each 2-dimensional projection of the random vector $\vec{X}$ is generated by an extreme measure, the optimization problem in this case is multiobjective. The number of optimization criteria is $M = J(J-1)/2$. The number of constraints is equal to the number of marginal distributions, J. The variables in this problem are the probabilities

$$p(\mathbf{i}) = \mathbb{P}(X_1 = i_1, X_2 = i_2, \ldots, X_J = i_J), \quad i_j \in \mathbb{Z}_+,$$

and, therefore, must satisfy the inequalities $0 \leq p(\mathbf{i}) \leq 1$.

Let us define the set of integers

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

and

$$\mathcal{I}_{k,l} = \{j : 1 \leq j \leq J, \quad j \neq k, \quad j \neq l\}.$$

Then the marginal probabilities, $P_k(i_k)$, can be written as

$$P_k(i_k) = \sum_{j \in \mathcal{I}_k} \sum_{i_j=0}^{\infty} p(i_1, \ldots, i_J), \quad i_k \in \mathbb{Z}_+.$$

The probabilities of the 2-dimensional projections

$$p_{k,l}(i_k, i_l) = \mathbb{P}(X_k = i_k, X_l = i_l), \quad k, l = 1, 2, \ldots, J, \ k \neq l, \ k, l \in \mathbb{Z}_+,$$

are computed as

$$p_{k,l}(i_k, i_l) = \sum_{j \in \mathcal{I}_{k,l}} \sum_{i_j=0}^{\infty} p(i_1, \ldots, i_J).$$

Similarly, the objective functions, $f_{k,l}(p) = \mathbb{E}[X_k X_l]$, take the form

$$f_{k,l}(p) = \sum_{i_k=1}^{\infty} \sum_{i_l=1}^{\infty} i_k i_l p_{k,l}(i_k, i_l), \quad 1 \leq k < l \leq J.$$

The optimization problem can then be written as

$$f_{k,l}(p) \to \ \text{extr} \quad 1 \leq k < l \leq J, \tag{5}$$

s. t.

$$\sum_{j \in \mathcal{I}_k} \sum_{i_j=0}^{\infty} p(i_1, \ldots, i_J), = P_k(i_k) \quad i_k \in \mathbb{Z}_+, \quad k = 1, \ldots, J$$

$$p\,(i_1, \ldots, i_J) \geq 0$$

where $P_j(\cdot)$ are given marginal probabilities ($j = 1, 2, \ldots, J$).

The main theorem

Let us now formulate the main result of the paper. It is convenient to introduce the following notation.

$$\tilde{F}_j(i_j, e) = \begin{cases} F_j(i_j) & \text{if } e_j = 0 \\ 1 - F_j(i_j) & \text{if } e_j = 1 \end{cases} \tag{6}$$

where the marginal distributions, $F_j(\cdot)$, satisfy

$$F_j(i_l) = \sum_{i_k=0}^{i_l} P_j(i_k)$$

Theorem 3.2 (Extreme Joint Distributions in Higher Dimensions)**.** *Given marginal distributions F_1, F_2, $\ldots F_J$ on $\mathbb{Z}_+$ and a binary vector $\vec{e}$, the extreme measure with the monotonicity structure $\vec{e}$ is defined by the probabilities*

$$\begin{aligned} p^{\vec{e}}(\boldsymbol{i}) = \big[&\min(\tilde{F}_1(i_1 - e_1; e_1), \ldots, \tilde{F}_J(i_J - e_J; e_J)) \\ &- \max(\tilde{F}_1(i_1 + (e_1 - 1); e_1), \ldots, \tilde{F}_J(i_J + (e_J - 1); e_J))\big]^+ \end{aligned} \tag{7}$$

Proof. We give a sketch of the proof here for the general case $J \geq 2$. A more complete proof for the case $J = 2$ is given in [8].

Let us first show that, if $J = 2$, then Equation (7) is equivalent to (3), in the case of maximal correlation, and to (4), in the case of minimal correlation. Indeed, in the first case, the distributions of X_1 and X_2 must be comonotone. Hence, $e_1 = e_2 = 0$ and $\tilde{F}_k(i, \vec{e}) = F_k(i)$ for $k = 1$ and 2 and all $i \geq 0$. In the antimonotone case, $e_1 = 0$, but $e_2 = 1$. Thus, $\tilde{F}_1(i, \vec{e}) = F_1(i)$, but $\tilde{F}_2(i, \vec{e}) = 1 - F_2(i - 1)$ for all $i \geq 0$. Therefore, Equation (7) is equivalent to (3) and (4).

Let us now consider the general case, $J \geq 3$. There are two groups of the coordinates of $\vec{X}$: comonotone and antimonotone. Denote their indices by

$$\mathcal{I}_{\mathcal{C}} = \{j : e_j = 0\} \text{ and } \mathcal{I}_{\mathcal{A}} = \{j : e_j = 1\}.$$

Let us now generate a large sample from the distribution $p^{\vec{e}}$ and sort them in the ascending order with respect to the first coordinate. It was shown in [8] that, after sorting, the comonotone coordinates of $\vec{X}$ will be permuted in the ascending order while the antimonotone will be permuted in the descending order.

Suppose that the indices $1 = k_1 < k_2 < k_3 < \cdots < k_C$ belong to $\mathcal{I}_{\mathcal{C}}$ and the complimentary set of indices is $\mathcal{I}_{\mathcal{A}} = \{l_1, l_2, \ldots, l_A\}$. A permuted sample is represented in (8).

$$\begin{aligned} X_1 &: \overbrace{0, \ldots, 0}^{N_1(0)}, \ldots, \overbrace{i-1, \ldots, i-1}^{N_1(i-1)},\ \overbrace{i, i, \ldots, i}^{N_1(i)}, \ldots \overbrace{k, k \ldots, k}^{N_1(k)}, \ldots \\ X_{k_2} &: \underbrace{0, 0, \ldots, 0}_{N_{k_2}(0)}, \ldots, \underbrace{i-1, \ldots, i-1}_{N_{k_2}(i-1)},\ \underbrace{i, \ldots, i}_{N_{k_2}(i)},\ \ldots, \\ &\qquad\qquad \vdots \\ X_{l_A} &: \ldots \underbrace{k, k, \ldots, k}_{N_{l_A}(k)}, \underbrace{k-1, \ldots, k-1}_{N_{l_A}(k-1)}, \ldots \underbrace{2, 2, 2, \ldots 2}_{N_{l_A}(2)}, \ldots \end{aligned} \tag{8}$$

where $N_k(m)$ denotes the number of realizations of m in the kth coordinate, X_k, of $\vec{X}$. The first position, $I_k^C(m)$, where the number m appears in the sorted sample of the r.v. X_k is

$$I_k^C(m) = 1 + \sum_{i=0}^{m-1} N_k(i), \quad k \in \mathcal{I}_{\mathcal{C}}.$$

The last position, $E_k^C(m)$, where the number m appears in the sorted sample of the r.v. X_k is

$$E_k^C(m) = \sum_{i=0}^{m} N_k(i), \quad k \in \mathcal{I}_{\mathcal{C}}.$$

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

$$\lim_{N_S\to\infty} \frac{N_k(m)}{N_S} = p_k(m) \quad \textbf{a.s.} \tag{9}$$

Therefore, for $k \in \mathcal{I_C}$

$$\lim_{N_S\to\infty} \frac{I_k^C(m)}{N_S} = F_k(m-1) \quad \textbf{a.s.}. \tag{10}$$

and

$$\lim_{N_S\to\infty} \frac{E_k^C(m)}{N_S} = F_k(m) \quad \textbf{a.s.}. \tag{11}$$

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

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

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

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

As $N_S \to \infty$, we have for $l \in \mathcal{I_A}$

$$\lim_{N_S\to\infty} \frac{I_l^A(m)}{N_S} = 1 - F_l(m) \quad \textbf{a.s.}. \tag{12}$$

and

$$\lim_{N_S\to\infty} \frac{E_l^A(m)}{N_S} = 1 - F_l(m-1) \quad \textbf{a.s.}. \tag{13}$$

The empirical measure of the event

$$\{\vec{X} = \vec{i}\} = \Big\{ \bigcap_{k\in\mathcal{I_C}} \{X_k = i_k\} \Big\} \bigcap \Big\{ \bigcap_{l\in\mathcal{I_A}} \{X_l = i_l\} \Big\}$$

is $\mathbf{m_{N_S}}(\{\vec{X} = \vec{i}\})$, which coincides with that of the intersection of the intervals

$$\Big\{ \bigcap_{k\in\mathcal{I_C}} [I_k^C(i_k), E_k^C(i_k)] \Big\} \bigcap \Big\{ \bigcap_{l\in\mathcal{I_A}} [I_l^A(i_l), E_l^A(i_l)] \Big\}$$

The latter can be written as follows. The right end of the intersection of the intervals is

$$\mathcal{R} = \min\Big(\min_{k\in\mathcal{I_C}} (E_k^C(i_k)), \min_{l\in\mathcal{I_A}} (E_l^C(i_l)) \Big)$$

and the left end is

$$\mathcal{L} = \max\Big(\max_{k\in\mathcal{I_C}} (I_k^C(i_k)), \max_{l\in\mathcal{I_A}} (I_l^C(i_l)) \Big)$$

Then we obtain

$$\mu_{\mathbf{N_S}}(\{\vec{X} = \vec{i}\} = \frac{(\mathcal{R} - \mathcal{L})^+}{N_S}$$

Note that the length of the intersection of intervals is 0 in the case $\mathcal{R} \le \mathcal{L}$. As $N_S \to \infty$, we obtain from Equations (10)–(13)

$$\lim_{N_S\to\infty} \mu_{\mathbf{N_S}}(\{\vec{X} = \vec{i}\} = \big[\min(\tilde{F}_1(i_1 - e_1; e_1), \ldots, \tilde{F}_J(i_J - e_J; e_J)) \\ - \max(\tilde{F}_1(i_1 + (e_1 - 1); e_1), \ldots, \tilde{F}_J(i_J + (e_J - 1)\big]^+.$$

Finally, note

$$\lim_{N_S\to\infty} \mu_{\mathbf{N_S}}(\{\vec{X} = \vec{i}\} = p^{\vec{e}}(\vec{X} = \vec{i}) \quad \textbf{a.s.}$$

Thus (7) is derived and the theorem is proved. □

As in the case $J = 2$, the support of an extreme measure looks like a staircase. Figure 2 displays the supports of all four extreme measures of a 3-dimensional multivariate Poisson process with intensities $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3) = (3, 5, 7)$. For each extreme measure, it also displays the three associated 2-dimensional projections.

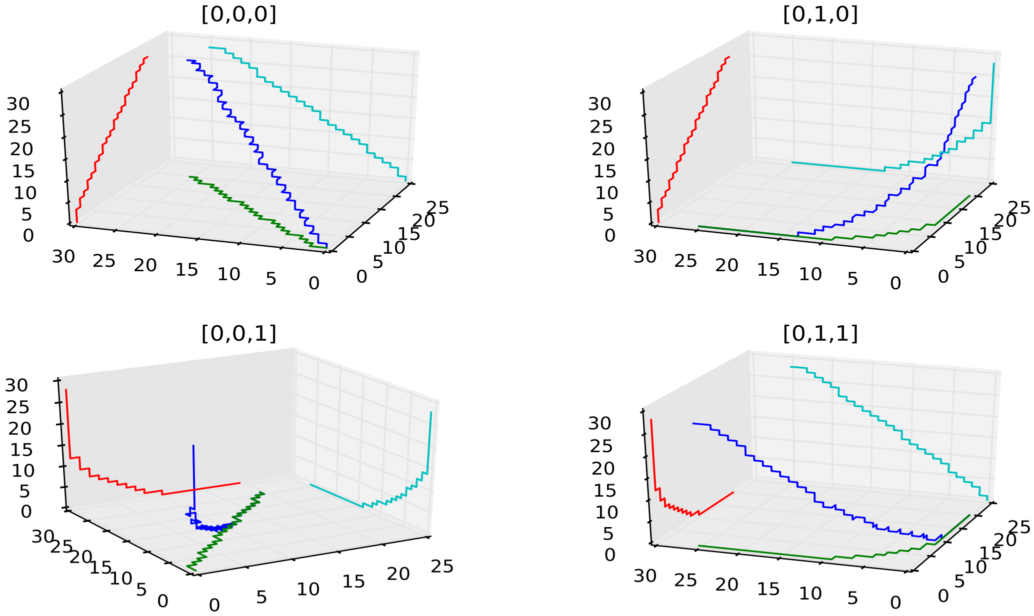

Figure 2: The blue curve in each graph is the support of an extreme measure in the case $J = 3$. The red, teal, and green curves represent the projection of the 3D support onto 2 dimensions. These four graphs completely describe the support of the extreme measures in the case $J = 3$.

4 EJD Algorithm in Higher Dimensions

Let us now describe the algorithm for the computation of the probabilities $p^{\vec{e}}(\vec{x})$. The preliminary step, approximation of marginal distributions by the distributions with finite support, is identical to that in [8]. The main step is the recursive computation of the probabilities $p^{\vec{e}}(\vec{x})$, which can be done as described in the algorithm below. Note that, in that algorithm, $p^{\vec{e}}(\vec{x})$ is assigned a value (in Step 5) only if $\vec{x}$ is in the support of $p^{\vec{e}}$ and the support point $\vec{x}$ is saved (in Step 3). If $\vec{x}$ is not a saved support point (i.e., not saved in Step 3), then $p^{\vec{e}}(\vec{x}) = 0$. To simplify the descrption of the algorithm below, we assume that all the marginal probabilities are positive.

Step 0a. Set $k = 0$
Step 0b. For each $j = 1 : J$
    If $e_j = 1$,
        Set $F_j(i) = 1 - F_j(i)$
        Set Δ_j = -1 and $x_j^0 = \max\{i : P_j(i) \geq 0\}$
    else
        Set $\Delta_j = 1$ and $x_j^0 = 0$
Step 0c. Set $z_0 = \min(F_1(0), \ldots, F_J(0))$ and $p^{\vec{e}}(x_1, \ldots, x_J) = z_0$
Step 1. Set $k = k + 1$
Step 2. For each $j = 1 : J$
    If $z_{k-1} = F_j(i_j)$ for some i_j,
        Set $x_j^k = i_j + \Delta_j$
    else
        Set $x_j^k = x_j^{k-1}$
Step 3. Save the k-th support point $\vec{x}_k = (x_1^k, \ldots, x_J^k)$

Step 4. Set $z_k = \min(F_1(x_1^k), \ldots, F_J(x_J^k))$
Step 5. Set $p^{\vec{e}}(x_1^k, \ldots, x_J^k) = z_k - z_{k-1}$
Step 6. Go to Step 1

5 Calibration of Correlations

In the case $J = 2$, given a correlation coefficient ρ, we want to find a convex combination of extreme measures such that, for any desired correlation ρ within a given admissible correlation range $[\rho^*, \rho^{**}]$, by solving for $w \in [0, 1]$ and setting

$$\rho = w\rho^* + (1-w)\rho^{**}$$

we obtain a probability $p = wp^* + (1-w)p^{**}$ with associated correlation ρ. If $J > 2$, the same idea is applicable. However, we have instead, a system of equations with N_w weights to solve for a given correlation matrix

$$\mathcal{C}_g = w_1\,\mathcal{C}_1 + \cdots + w_{N_w}\,\mathcal{C}_{N_w}, \tag{14}$$

where the $\mathcal{C}_n$ are correlation matrices associated with the extreme distributions, $w_n \geq 0$ and $\sum_{n=1}^{N_w} w_n = 1$. Taking the extreme measures with the same set of marginal distributions, we construct the convex combination

$$p^w = w_1 p^{e_1} + \cdots + w_{N_w} p^{e_{N_w}} \tag{15}$$

where p^w has correlation matrix $\mathcal{C}_g$. The calibration problem is now reduced to finding a minimal N_w to form a convex combination of extreme measures. Indeed, the number of extreme measures is 2^{J-1} and the number of correlation coefficients is $M = J(J-1)/2$. In matrix form (14) can be written as

$$Aw = \hat{\mathcal{C}}_g \tag{16}$$

where A is of dimension M-by-N, the i^{th} column of A is a vectorized version of the upper triangular part of the extreme correlation matrix $\mathcal{C}_i$ and $\hat{\mathcal{C}}_g$ is a vectorized version of the matrix $\mathcal{C}_g$. As the dimensionality of the multivariate Poisson process J increases, A becomes increasingly underdetermined. To find the weights, w_j, one can solve the following linear programming (LP)

$$\begin{aligned} \min \quad & \mathbf{1}^T w \\ \text{s. t.} \quad & Aw = \hat{\mathcal{C}}_g \\ & \mathbf{1}^T w = 1 \\ & w_n \geq 0 \qquad n = 1, 2, \ldots, N. \end{aligned} \tag{17}$$

If (17) does not have a solution, this implies that $\mathcal{C}_g$ is not the correlation matrix of any multivariate Poisson process with the prescribed marginals. Note that solving the LP (17) is not necessary: any feasible solution to (17) suffices. Also, note that once we have found a w satisfting the constraints in (17), we can reduce the number of nonzero components in w to $N_w \leq M + 1$ using, for exmaple, a techinque similiar to that often used in the proof of Carathéodory's theorem, to obtain a vector of N_w nonzero weights satisfying (15) and the assoicated convexity constraints on w.

A matrix $\mathcal{C}$ is called admissible if it is a symmetric, positive semi-definite (PSD) matrix with ones on the diagonal and each entry satisfying $\rho_{ij}^* \leq c_{ij} \leq \rho_{ij}^{**}$, where ρ_{ij}^* and ρ_{ij}^{**} are extreme correlations for the 2-dimensional problem for (X_i, X_j). Notice that the correlation matrices corresponding to the extreme measures are admissible.

Theorem 5.1. *A convex combination of admissible correlation matrices is also an admissible correlation matrix.*

Proof. This fact readily follows from the observation that a convex combination of positive semi-definite quadratic forms is also a PSD form. □

The probabilities $p^{\vec{e}}(\vec{i})$ describing the extreme measures and their supports are very different from the case of independent r.v. X_j. In particular, if $\rho = 0$, the support of the measure p^w is the union of the supports of $p^{\vec{e}}$. By adding an additional edge point p^0 corresponding to the case of independent components of $\vec{X}$, one can obtain a more general solution. We do not discuss this problem further in this paper.

6 Simulation

Up until this point, we have discussed the computation of the multivariate Poisson distribution at some terminal time T via the EJD. That allows us to achieve extreme correlations between the components of the multivariate Poisson process at time T. This also gives us bounds on the admissible correlation between the components. Furthermore, given the characterization of the extreme measures we are able to construct any multivarite Poisson process with correlations between the components satisfying the admissible correlation bounds. With the BS approach, which we will briefly review below, we are then able to construct a correlated multivariate Poisson on $[0, T]$ having desired admissible correlation at time T. Finally, we introduce the Forward-Backward continuation which allows us to construct forward the process in subsequent intervals $[mT, (m+1)T]$ for some m.

The BS simulation technique includes the following steps.

1. Given a finite vector of weights, w_n, $(n = 1, 2, \ldots, N_w)$, satisfying the conditions $w_n \geq 0$, $\sum_{n=1}^{N_w} w_n = 1$, generate an index, n by sampling from the probability distribution defined by w to choose an extreme measure, $p^{\vec{e_n}}$.
2. Generate a random vector $N_T = (N_T(1), \ldots N_T(J))$ from the extreme measure p^{e_n}.
3. Generate arrival moments of the multivariate process $\mathbf{N_t}, (0 \leq t \leq T)$.

Forward Continuation of the Backward Simulation

The BS technique allows one to construct the sample paths of the multivariate Poisson process in a bounded time interval, $[0, T]$. In this section we consider a generalization of the technique, which we call Forward-Backward simulation. We explain the idea of this approach for $J = 2$.

Consider a sequence of time intervals $[0, T)$, $[T, 2T)$, $\ldots$, $[mT, (m+1)T]$. Suppose that a bivariate Poisson process, (X_t, Y_t), has already been simulated in the interval $[0, T)$ using the BS technique. For any τ, $0 \leq \tau < T$, the increments $X_{T+\tau} - X_T$ are independent of X_T and $Y_{T+\tau} - Y_T$ are independent of Y_T. Let us define the joint distribution of the increments as

$$(X_{T+\tau} - X_T, Y_{T+\tau} - Y_T) \stackrel{\mathcal{D}}{=} (\hat{X}_\tau, \hat{Y}_\tau), \quad 0 < \tau \leq T,$$

where $\hat{X}_\tau$ and $\hat{Y}_\tau$ are independent versions of X_t and Y_t, respectively, $\hat{X}_\tau \stackrel{\mathcal{D}}{=} X_t$ and $\hat{Y}_\tau \stackrel{\mathcal{D}}{=} Y_t$. Then we find

$$\mathrm{Cov}(X_{T+\tau}, Y_{T+\tau}) = \mathrm{Cov}(X_T, Y_T) + \mathrm{Cov}(X_\tau, Y_\tau).$$

Taking into account that

$$\mathrm{Cov}(X_\tau, Y_\tau) = \mathrm{Cov}(X_T, Y_T) \cdot \frac{\tau^2}{T^2},$$

we obtain

$$\rho(T + \tau) = \rho(T) \frac{T^2 + \tau^2}{T(T + \tau)}.$$

In particular, we have $\rho(2T) = \rho(T)$ and $\mathrm{Cov}(X_{2T}, Y_{2T}) = 2\,\mathrm{Cov}(X_T, Y_T)$. Suppose now that $\rho(t)$ is defined for all $t \leq nT$.

Consider now the case $t = nT + \tau \in [nT, (n+1)T)$. We have

$$\mathrm{Cov}(X_{nT}, Y_{nT}) = n\,\mathrm{Cov}(X_T, Y_T)$$

and

$$\mathrm{Cov}(X_{nT+\tau}, Y_{nT+\tau}) = \mathrm{Cov}(X_T, Y_T) \cdot \Big(n + \frac{\tau^2}{T^2}\Big).$$

The latter relation implies

$$\rho(nT+\tau) = \rho(T)\frac{n + \tau^2 \cdot T^{-2}}{n + \tau T^{-1}}.$$

Thus, the processes X_t and Y_t have asymptotically stationary correlations as $t \to \infty$. An illustration of this is shown in Figure 3, where maximal (red line) and minimal (blue line) values of the correlation coefficient, $\mathrm{corr}(X_t, Y_t)$ are depicted.

It would be interesting to generalize this result for the class of mixed Poisson processes. The main difficulty is that the increments of the mixed Poisson processes are not independent.

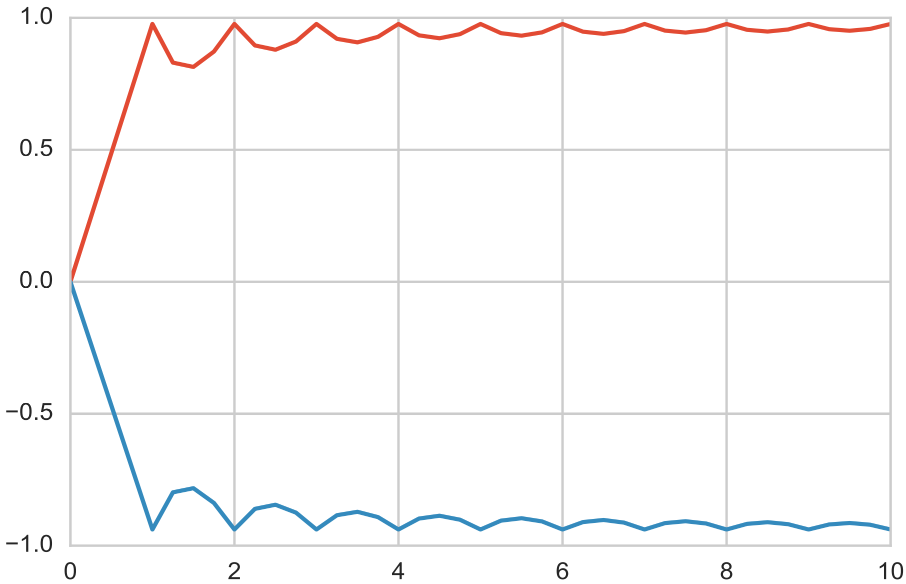

Figure 3: Forward Continuation of Backward Simulation: $\mathrm{corr}(X_t, Y_t)$, $\mu_1 = 3$, $\mu_2 = 5$.

7 Final remarks

We presented the solution to the problem of simulation of multivariate Poisson processes in the case the dimension of the problem is $J > 2$ and we described the admissible parameters for the calibration problem. We also extended the BS approach with the introduction of the Forward Continuation of BS.

There are several directions for future research. One is to extend the EJD approach to more general processes such as the Mixed Poisson processes. Another direction relates to the calibration problem. Another direction is to study the interplay between the optimization problem and the EJD algorithm for computing the probabilities of the extreme measures. Exploring the synthesis of Forward and Backward simulation for more general processes is also worthwhile.

References

[1] Aue, F., Kalkbrener, M.: Lda at work: Deutsche banks approach to quantify operational risk. Journal of Operational Risk **1(4)**, 49–95 (2006)

[2] Bae, T., Kreinin, A.: A backward construction and simulation of correlated poisson processes. Journal of Statistical Computation and Simulation, DOI: 10.1080/00949655.2016.1277428 (2017)

[3] Böcker, K., Klüppelberg, C.: Multivariate models for operational risk. Quantitative Finance **10(8)**, 855–869 (2010)

[4] Chavez-Demoulin, V., Embrechts, P., Nešlehovà, J.: Quantitative models for operational risk: extremes, dependence and aggregation. Journal of Banking and Finance **30(9)**, 2635–2658 (2006)

[5] Duch, K., Kreinin, A., Jiang, Y.: New approaches to operational risk modeling. IBM Journal of Research and Development **3**, 31–45 (2014)

[6] Embrechts, P., Puccetti, G.: Aggregating risk capital with an application to operational risk. The Geneva Risk and Insurance Review **31(2)**, 71–90 (2006)

[7] Fréchet, M.: Sur les tableaux dont les marges et des bornes sont données. Revue de l'Institut international de statistique **A. 14**, 53–77 (1951)

[8] Kreinin, A.: Correlated poisson processes and their applications in financial modeling. Financial Signal Processing and Machine Learning: Correlated Poisson Processes and Their Applications in Financial Modeling (2016)

[9] Panjer, H.H.: Operational Risks: Modeling Analytics. Wiley, New York, NY (2006)

[10] Powojovsky, M., Reynolds, D., Tuenter, H.: Dependent events and operational risk. Algorithmics Research Quarterly **5(2)**, 65–74 (2002)

[11] Rachev, Svetlozar T and Rüschendorf, Ludger.: Mass Transportation Problems: Volume I: Theory Springer Science & Business Media (1998)

[12] Rachev, Svetlozar T and Rüschendorf, Ludger.: Mass Transportation Problems: Volume II: Applications Springer Science & Business Media (1998)

[13] Shevchenko, P.: Modelling Operational Risk Using Bayesian Inference. Springer-Verlag (2011)

[14] Villani, Cédric.: Optimal transport, old and new. Springer Science & Business Media (2008)