1 Basic Probability Theory

A probabilistic model consists of a **sample space** Ω, which is the set of all possible outcomes; and a **probability law** that assigns a probability $P(A)$ to an event A (a set of possible outcomes). The following properties hold:

- *Non-negativity*: $P(A) \geq 0$ for every event A.
- *Additivity*: If A_i's are all disjoint events, $P(A_1 \bigcup A_2 \bigcup ...) = \sum P(A_i)$.
- *Normalization*: The probability of the union of all possible events is 1: $P(\Omega) = 1$.

1.1 Conditional Probability

The probability of an event A occurring, given that we know that an event B has occurred, is denoted and computed as

$$P(A|B) = \frac{P(A \bigcap B)}{P(B)}$$

Under the assumption of course that $P(B) > 0$ (since we know that B has occurred.)

1.1.1 Total Probability Theorem

Let $A_0, ..., A_n$ be disjoint events that form a partition of the sample space (each possible outcome is included in one and only one of the events $A_1, ..., A_n$) and assume that $P(A_i) > 0$ for all $i = 1, .., n$, then for any event B, we have

$$\begin{aligned} P(B) &= P(A_1 \bigcap B) + ... + P(A_n \bigcap B) \\ &= P(A_1)P(B|A_1) + ... + P(A_n)P(B|A_n) \end{aligned}$$

1.1.2 Independence

Events A and B are independent events if and only if,

$$\begin{aligned} P(A \bigcap B) &= P(A)P(B) \\ P(A|B) &= P(A) \end{aligned}$$

Otherwise, the two events are said to be dependent. This can, of course, be generalized and holds for any set of n events.

2 Random Variables

2.1 Discrete Random Variables

For a discrete random variable X, the probability mass function (PMF) gives the probability that X will take on a particular value in its domain and we denote this probability by P_X. In other words:

$$P_X(x) = P(\{X = x\})$$

2.1.1 Expectation

We define the expected value of a discrete random variable X by

$$E[X] = \sum xP_X(x)$$

Let $g(X)$ be a real-valued function of X, we can get the expected value of $g(X)$ by

$$E[g(X)] = \sum g(x)P_X(x)$$

2.2 Continuous Random Variables

For a continuous random variable X, a probability density function (PDF) f_X is a non-negative function such that

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

and

$$\int_{-\infty}^{+\infty} f_X(x)dx = P(-\infty < X < +\infty) = 1$$

The latter is called the *normalizing condition* and must hold for all probability density functions. Note that since

$$P(X = a) = \int_a^a f(x)dx = 0$$

we get that

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b)$$

so we don't need to be too careful when we are writing the inequalities.

2.2.1 Expectation

The expectation of a continuous random variable is defined similarly as

$$E[X] = \int_{-\infty}^{+\infty} xf_X(x)dx$$

And for a similar non-negative real function $g(X)$ we have

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f_X(x)dx$$

2.3 Cumulative Distribution Functions

The cumulative distribution function (CDF) of a random variable X is the probability $P(X \leq x)$, denoted by $F_X(x)$.

If X is a discrete random variable, then we get

$$F_X(x) = P(X \leq x) = \sum_{k \leq x} P_X(k)$$

And similarly if X is a continuous random variable we get

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt$$

In this case we can therefore define f_X in terms of F_X:

$$f_X(x) = \frac{dF_X(x)}{dx}$$

So the probability density function of a continuous random variable is the derivative of its cumulative distribution function.

2.4 Conditional PDF

The conditional PDF $f_{X|A}$ of a continuous random variable X given an event A with $P(A) > 0$, satisfies

$$P(X \epsilon B|A) = \int_B f_{X|A}(x)dx$$

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(X\epsilon A)} \text{if X}\epsilon\text{A} \\ \text{0 otherwise} \end{cases}$$

And results are obtained as before:

$$E[X|A] = \int_{-\infty}^{+\infty} x f_{X|A}(x)dx$$

A version of the total probability theorem:

$$f_X(x) = \sum_{i=0}^{n} P(A_i) f_{X|A}(x)$$

And the total expectation theorem:

$$E[X] = \sum_{i=1}^{n} P(A_i) E[X|A_i]$$

2.5 Exponential Random Variable

An exponential random variable has a PDF of the form

$$f_X(x) \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

where $\lambda > 0$. The CDF is given by

$$F_X(x) = 1 - e^{-\lambda x}, \qquad x \geq 0$$

Verify that for an exponential random variable, we have

$$E[X] = \frac{1}{\lambda}.$$

An important property of an exponential random variable is the memoryless property. Let x be an exponential random variable, then we have that

$$P\{X > r + t \mid X > t\} = P\{X > r\}, \qquad r, t \geq 0.$$

This property can be verified as follows,

$$P\{X > r + t \mid X > t\} = \frac{P\{X > r + t\}}{P\{X > t\}} = \frac{e^{-\lambda(r+t)}}{e^{-\lambda t}} = e^{-\lambda r} = P\{X > r\}.$$

3 Stochastic Processes

Consider a sequence of events that can be thought of as "arrivals". In the Bernoulli process, arrivals occur in discrete time and the interarrival times are geometrically distributed; in the Poisson process, arrivals occur in continuous time and the interarrival times are exponentially distributed. We will talk about the Poisson process in a later section. First, we give a quick overview of the Bernoulli process.

3.1 Bernoulli Process

A Bernoulli process is a sequence $X_1, X_2, \ldots$ of independent Bernoulli random variables X_i with

$$P(X_i = 1) = P(success) = p$$

$$P(X_i = 0) = P(failure) = 1 - p$$

3.1.1 Binomial Random Variable with parameters p and n

This is the number S of successes (or arrivals) in n independent trials.

The PMF is given by

$$P_S(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, .., n$. And the expected number of successes (or arrivals) is given by

$$E[S] = np$$

3.1.2 Geometric Random Variable with parameter p

This is the number, T, of trials up to (and including) the first success.

The PMF is given by

$$P_T(t) = (1-p)^{t-1}p$$

for $t = 1, 2, ...$ And the expectation is

$$E[T] = \frac{1}{p}$$

4 The Poisson Process

Problems modeled by the Poisson process can generally be described in terms of event arrivals (or occurrences). For example, and event occurrence might be the arrival of a new customer, or the pittance of a new particle and so on. In the context of computer network performance Madeline, the arrivals are generally those of messages, or packets. We are interested in how these packets arrive under certain conditions, how many we can expect in a certain amount of time and so on. We denote $P(k,t)$ the probability that we get exactly k arrivals during an interval of length t, where t is any (positive) real number. We also associate with Poisson process an **arrival rate** $\lambda > 0$.

4.1 The Assumption

Every mathematical model imposes restrictions on the phenomena it wishes to model. The fewer the restrictions, the more general the model. The Poisson process has three basic conditions: independence, individuality and uniformity (or homogeneity). Since most of the events in networking meet these conditions, we apply the Poisson process to model and understand the systems in which we are interested.

- *Independence*: The occurrences of an event are independent between two disjoint intervals. That is, the probability of one event occurring in one time interval is independent of the probability of the event occurring during another time interval. The time intervals are assumed to be non-overlapping (disjoint).

- *Individuality (or Small Interval Probabilities)*: The probability of two events occurring during a single time interval is very close to 0. More importantly, as the time interval goes to 0, the probability of two events occurring during one time interval goes to 0 faster. (For the purposes of performance modeling we assume that the probability of two packets arriving in one timer interval to be 0.) More precisely, we have

  $$P(0,t) = 1 - \lambda t + o(t),$$
  $$P(1,t) = \lambda t + o_1(t),$$

  where $o(t)$ and $o_1(t)$ are functions with the property that

  $$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0 \qquad \lim_{\Delta t \to 0} \frac{o_1(\Delta t)}{\Delta t} = 0$$

- *Uniformity (or Time-Homogeneity)*: The occurrence of events is uniform over time. That is, the probability of an event occurring during a time interval must be proportional to the length of the time interval. (We will see how this applies in section 3).

The Poisson process assumes that the times between arrivals are independent and equally distributed random variables. The assumption means that the probability of an arrival during one time period is identical to, and independent of, the probability of an arrival during any other time period (of the same length of course).

An important property that arises from the Poisson process is the so-called *memoryless* property. This property means, simply, that what happens in the past does not affect the probability of events occurring in the future. For example, if we know that for a certain amount of time no packet has arrived, then we cannot use this information to infer probabilities of inter-arrival times after this time period. Using the notation of conditional probability we illustrate this property as follows: if the inter-arrival time, described by a random variable X, and $s \geq 0$, $t \geq 0$, then the memoryless property means that:

$$P(X > t + s \mid X > s) = P(X > t)$$

4.2 The Poisson Distribution

We wish to model the arrival of packets within a certain time period T. We start with the following approximation. We break up this time period into N equal slots of length $\Delta t = \frac{T}{N}$. (Omitting the $o(t)$ terms) The probability of exactly one packet arriving in this time slot is $\lambda \Delta t$ (because of our *uniformity* condition) and no packets arriving with probability $1 - \lambda \Delta t$. Our *individuality* condition tells us that the probability of two or more packets arriving during a single time interval vanishes to 0 (given that each time period is sufficiently small).

We are interested in the probability that n packets arrive within this time period $[0, T]$. Since there can be at most one packet arriving in each time slot, and we have a total of N time slots, there can be at most N packets arriving during this time period, and so n must range from 0 to N.

Now, if we had n packets arriving, then $N - n$ packets did not arrive, and therefore we have, for one instance, the probability $(\lambda \Delta t)^n (1 - \lambda \Delta t)^{N-n}$ that n packets arrive during that time period. However, there are many ways in which we can choose the n times slots that have an arrival. Specifically, there are N choose n different configurations in which n packets can arrive. Therefore, in total we have the following probability, P_n, that n packets arrive in the time period $[0, T]$:

$$P_n = \binom{N}{n} (\lambda \Delta t)^n (1 - \lambda \Delta t)^{N-n} \tag{1}$$

The above model is an approximation: the Poisson process is a continuous-time process which we approximated through a discrete-time process by dividing time into slots. This approximation becomes more and more accurate, as we let the duration of time slot become smaller and smaller. Or equivalently, the approximation becomes accurate as number of time slots N goes to infinity. First, we rewrite (1) to make it easier to take this limit.

$$\begin{aligned} P_n &= \binom{N}{n} (\lambda \Delta t)^n (1 - \lambda \Delta t)^{N-n} \\ &= \frac{N!}{(N-n)!n!} \left(\lambda \frac{T}{N}\right)^n (1 - \lambda \Delta t)^{\frac{T}{\Delta t}} (1 - \lambda \Delta t)^{-n} \end{aligned}$$

$$= \frac{N(N-1)...(N-n+1)}{N^n}\frac{(\lambda T)^n}{n!}(1-\lambda\Delta t)^{\frac{T}{\Delta t}}(1-\lambda\Delta t)^{-n}$$

Now as we let N go to infinity, $\Delta t = \frac{T}{N}$ goes to 0. Using

$$\lim_{\Delta t \to 0}(1-\lambda\Delta t)^{\frac{T}{\Delta t}} = e^{-\lambda T}$$

we get the following result for P_n:

$$P_n = (1)\frac{(\lambda T)^n}{n!}e^{-\lambda T}(1)^{-n}$$

$$= \frac{(\lambda T)^n}{n!}e^{-\lambda T}$$

This is the Poisson distribution, T is the time period and λ is the arrival rate.
Since the probabilities must add up to 1, we need to verify that:

$$\sum_{n=0}^{\infty} P_n = 1$$

Here's how we do it:

$$\begin{aligned}
\sum_{n=0}^{\infty} P_n &= \sum_{n=0}^{\infty}\frac{(\lambda T)^n}{n!}e^{-\lambda T} \\
&= e^{-\lambda T}\sum_{n=0}^{\infty}\frac{(\lambda T)^n}{n!} \\
&= e^{-\lambda T}e^{\lambda T} \\
&= 1
\end{aligned}$$

Since $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ...$, and therefore we have our required result.

4.3 Expectation

We showed that the probability of n packets arriving during a time period is dependent on λ. The parameter λ is the average rate of arrivals per unit of time, so λT is the average number of arrivals during the time period T. We now prove this result.

$$\begin{aligned}
E[n] &= \sum_{n=0}^{\infty} n\frac{(\lambda T)^n}{n!}e^{-\lambda T} \\
&= (\lambda T)e^{-\lambda T}\sum_{n=1}^{\infty}\frac{(\lambda T)^{n-1}}{(n-1)!} \\
&= (\lambda T)e^{-\lambda T}e^{\lambda T} \\
&= \lambda T
\end{aligned}$$

4.4 The Exponential Distribution of Interrarival Times

So far we've talked about the probability of arrivals, but we haven't really looked at the time between these arrivals. Interarrival times in a Poisson process are exponentially distributed.

Let t be the time between two successive arrivals and let $P(t \leq s)$ be the probability of t being less than some non-negative, real value s. We have the following relation

$$P(t \leq s) = 1 - P(t\ s),$$

where $P(t\ s)$ is the probability that the interarrival time is larger than s. This probability is equal to the probability that we have no arrival in the interval $[0, s]$. Using the Poisson distribution, this probability is equal to

$$P(t\ s) = e^{-\lambda s}.$$

Therefore, the probability of t being less than some non-negative, real value s is equal to

$$P(t \leq s) = 1 - e^{-\lambda s}$$

Which implies that its probability density function (PDF) is

$$p(t) = \lambda e^{-\lambda t}$$

The expected value of t is

$$E[t] = \frac{1}{\lambda}$$

4.5 Modeling Packet Arrivals as a Poisson Process

We will often model arrivals of data packets at a node (router) as a Poisson process. If we let $A(t)$ represent the total number of arrivals in our system at time t with arrival rate λ, we know the following is true:

$$\begin{aligned}
P\{A(t + \Delta t) - A(t) = 0\} &= 1 - \lambda\Delta t + o(\Delta t) \\
P\{A(t + \Delta t) - A(t) = 1\} &= \lambda\Delta t + o(\Delta t) \\
P\{A(t + \Delta t) - A(t) \geq 2\} &= o(\Delta t)
\end{aligned}$$

where

$$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

The first probability is that of no arrivals during a small time interval Δt, while the second gives the probability of exactly one arrival during this time interval.

4.6 Other Properties of a Poisson Process

If we have two or more independent Poisson processes $A_1, ..., A_n$, with corresponding arrival rates $\lambda_1, ..., \lambda_n$, are merged into a single Poisson process, then the resulting process is Poisson, and its parameter is the sum of the parameters of the original processes. That is, the resulting Poisson process, A, has an arrival rate

$$\lambda = \sum_{i=0}^{n} \lambda_i$$

.

This is important for us because we can consider many applications, for example, that are each delivering packages according to a Poisson process to a server, we can sum up the arrival times and consider the entire thing as a single Poisson process.

The dual of this property is splitting up of a single Poisson process. Let A be a Poisson process with arrival rate λ, and let A_i, $i = 1, ..., n$, be processes that are given the arrivals for A independently with probability p_i respectively (such that $\sum_{i=1}^{n} p_i = 1$), then each process A_i is a Poisson process with arrival rate λp_i.

5 Useful Results

These results are useful for manipulating many of the equations that come up when studying probability.

5.1 Geometric Series

For $x \neq 1$,

$$\sum_{k=0}^{n} x^k = \frac{x^{n+1} - 1}{x - 1}$$

When $|x| < 1$,

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1 - x}$$

We can get another result if we differentiate both sides:

$$\sum_{k=0}^{\infty} kx^k = \frac{x}{(1 - x)^2}$$

5.2 Exponentials

Taylor series expansion of e^x is:

$$e^x = 1 + x + \frac{x^2}{2!} + ... = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

Another result is:

$$\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x$$