5

Stochastic Processes

Contents

A stochastic process is a mathematical model of a probabilistic experiment that evolves in time and generates a sequence of numerical values. For example, a stochastic process can be used to model:

(a) the sequence of daily prices of a stock;

(b) the sequence of scores in a football game;

(c) the sequence of failure times of a machine;

(d) the sequence of hourly traffic loads at a node of a communication network;

(e) the sequence of radar measurements of the position of an airplane.

Each numerical value in the sequence is modeled by a random variable, so a stochastic process is simply a (finite or infinite) sequence of random variables and does not represent a major conceptual departure from our basic framework. We are still dealing with a single basic experiment that involves outcomes governed by a probability law, and random variables that inherit their probabilistic properties from that law.† However, stochastic processes involve some change in emphasis over our earlier models. In particular:

(a) We tend to focus on the **dependencies** in the sequence of values generated by the process. For example, how do future prices of a stock depend on past values?

(b) We are often interested in **long-term averages**, involving the entire sequence of generated values. For example, what is the fraction of time that a machine is idle?

(c) We sometimes wish to characterize the likelihood or frequency of certain **boundary events.** For example, what is the probability that within a given hour all circuits of some telephone system become simultaneously busy, or what is the frequency with which some buffer in a computer network overflows with data?

In this book, we will discuss two major categories of stochastic processes.

(a) *Arrival-Type Processes*: Here, we are interested in occurrences that have the character of an "arrival," such as message receptions at a receiver, job completions in a manufacturing cell, customer purchases at a store, etc. We will focus on models in which the interarrival times (the times between successive arrivals) are independent random variables. In Section 5.1, we consider the case where arrivals occur in discrete time and the interarrival times are geometrically distributed – this is the *Bernoulli process*. In Section 5.2, we consider the case where arrivals occur in continuous time and

† Let us emphasize that all of the random variables arising in a stochastic process refer to a single and common experiment, and are therefore defined on a common sample space. The corresponding probability law can be specified directly or indirectly (by assuming some of its properties), as long as it unambiguously determines the joint CDF of any subset of the random variables involved.

the interarrival times are exponentially distributed – this is the *Poisson process*.

(b) *Markov Processes*: Here, we are looking at experiments that evolve in time and in which the future evolution exhibits a probabilistic dependence on the past. As an example, the future daily prices of a stock are typically dependent on past prices. However, in a Markov process, we assume a very special type of dependence: the next value depends on past values only through the current value. There is a rich methodology that applies to such processes, and which will be developed in Chapter 6.

5.1 THE BERNOULLI PROCESS

The Bernoulli process can be visualized as a sequence of independent coin tosses, where the probability of heads in each toss is a fixed number p in the range $0 < p < 1$. In general, the Bernoulli process consists of a sequence of Bernoulli trials, where each trial produces a 1 (a success) with probability p, and a 0 (a failure) with probability $1 - p$, independently of what happens in other trials.

Of course, coin tossing is just a paradigm for a broad range of contexts involving a sequence of independent binary outcomes. For example, a Bernoulli process is often used to model systems involving arrivals of customers or jobs at service centers. Here, time is discretized into periods, and a "success" at the kth trial is associated with the arrival of at least one customer at the service center during the kth period. In fact, we will often use the term "arrival" in place of "success" when this is justified by the context.

In a more formal description, we define the Bernoulli process as a sequence $X_1, X_2, \ldots$ of **independent** Bernoulli random variables X_i with

$$\mathbf{P}(X_i = 1) = \mathbf{P}(\text{success at the } i\text{th trial}) = p,$$
$$\mathbf{P}(X_i = 0) = \mathbf{P}(\text{failure at the } i\text{th trial}) = 1 - p,$$

for each i.†

Given an arrival process, one is often interested in random variables such as the number of arrivals within a certain time period, or the time until the first arrival. For the case of a Bernoulli process, some answers are already available from earlier chapters. Here is a summary of the main facts.

† Generalizing from the case of a finite number of random variables, the independence of an *infinite* sequence of random variables X_i is defined by the requirement that the random variables $X_1, \ldots, X_n$ be independent for any finite n. Intuitively, knowing the experimental values of any finite subset of the random variables does not provide any new probabilistic information on the remaining random variables, and the conditional distribution of the latter stays the same as the unconditional one.

Some Random Variables Associated with the Bernoulli Process and their Properties

- **The binomial with parameters p and n.** This is the number S of successes in n independent trials. Its PMF, mean, and variance are

$$p_S(k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n,$$

$$\mathbf{E}[S] = np, \qquad \mathrm{var}(S) = np(1-p).$$

- **The geometric with parameter p.** This is the number T of trials up to (and including) the first success. Its PMF, mean, and variance are

$$p_T(t) = (1-p)^{t-1} p, \qquad t = 1, 2, \ldots,$$

$$\mathbf{E}[T] = \frac{1}{p}, \qquad \mathrm{var}(T) = \frac{1-p}{p^2}.$$

Independence and Memorylessness

The independence assumption underlying the Bernoulli process has important implications, including a memorylessness property (whatever has happened in past trials provides no information on the outcomes of future trials). An appreciation and intuitive understanding of such properties is very useful, and allows for the quick solution of many problems that would be difficult with a more formal approach. In this subsection, we aim at developing the necessary intuition.

Let us start by considering random variables that are defined in terms of what happened in a certain set of trials. For example, the random variable $Z = (X_1 + X_3)X_6X_7$ is defined in terms of the first, third, sixth, and seventh trial. If we have two random variables of this type and if the two sets of trials that define them have no common element, then these random variables are independent. This is a generalization of a fact first seen in Chapter 2: if two random variables U and V are independent, then any two functions of them, $g(U)$ and $h(V)$, are also independent.

Example 5.1.

(a) Let U be the number of successes in trials 1 to 5. Let V be the number of successes in trials 6 to 10. Then, U and V are independent. This is because $U = X_1 + \cdots + X_5$, $V = X_6 + \cdots + X_{10}$, and the two collections $\{X_1, \ldots, X_5\}$, $\{X_6, \ldots, X_{10}\}$ have no common elements.

(b) Let U (respectively, V) be the first odd (respectively, even) time i in which we have a success. Then, U is determined by the odd-time sequence $X_1, X_3, \ldots$, whereas V is determined by the even-time sequence $X_2, X_4, \ldots$. Since these two sequences have no common elements, U and V are independent.

Suppose now that a Bernoulli process has been running for n time steps, and that we have observed the experimental values of $X_1, X_2, \ldots, X_n$. We notice that the sequence of future trials $X_{n+1}, X_{n+2}, \ldots$ are independent Bernoulli trials and therefore form a Bernoulli process. In addition, these future trials are independent from the past ones. We conclude that starting from any given point in time, the future is also modeled by a Bernoulli process, which is independent of the past. We refer to this as the **fresh-start** property of the Bernoulli process.

Let us now recall that the time T until the first success is a geometric random variable. Suppose that we have been watching the process for n time steps and no success has been recorded. What can we say about the number $T-n$ of remaining trials until the first success? Since the future of the process (after time n) is independent of the past and constitutes a fresh-starting Bernoulli process, the number of future trials until the first success is described by the same geometric PMF. Mathematically, we have

$$\mathbf{P}(T-n=t \,|\, T>n) = (1-p)^{t-1}p = \mathbf{P}(T=t), \qquad t=1,2,\ldots.$$

This **memorylessness** property can also be derived algebraically, using the definition of conditional probabilities, but the argument given here is certainly more intuitive.

Memorylessness and the Fresh-Start Property of the Bernoulli Process

- The number $T-n$ of trials until the first success after time n has a geometric distribution with parameter p, and is independent of the past.
- For any given time n, the sequence of random variables $X_{n+1}, X_{n+2}, \ldots$ (the future of the process) is also a Bernoulli process, and is independent from $X_1, \ldots, X_n$ (the past of the process).

The next example deals with an extension of the fresh-start property, in which we start looking at the process at a *random* time, determined by the past history of the process.

Example 5.2. Let N be the first time in which we have a success immediately following a previous success. (That is, N is the first i for which $X_{i-1} = X_i = 1$.) What is the probability $\mathbf{P}(X_{N+1} = X_{N+2} = 0)$ that there are no successes in the two trials that follow?

Intuitively, once the condition $X_{N-1} = X_N = 1$ is satisfied, from then on, the future of the process still consists of independent Bernoulli trials. Therefore the probability of an event that refers to the future of the process is the same as in a fresh-starting Bernoulli process, so that $\mathbf{P}(X_{N+1} = X_{N+2} = 0) = (1-p)^2$.

To make this argument precise, we argue that the time N is a random variable, and by conditioning on the possible values of N, we have

$$\begin{aligned}\mathbf{P}(X_{N+1} = X_{N+2} = 0) &= \sum_n \mathbf{P}(N = n)\mathbf{P}(X_{N+1} = X_{N+2} = 0 \mid N = n) \\ &= \sum_n \mathbf{P}(N = n)\mathbf{P}(X_{n+1} = X_{n+2} = 0 \mid N = n)\end{aligned}$$

Because of the way that N was defined, the event $\{N = n\}$ occurs if and only if the experimental values of $X_1, \ldots, X_n$ satisfy a certain condition. But the latter random variables are independent of X_{n+1} and X_{n+2}. Therefore,

$$\mathbf{P}(X_{n+1} = X_{n+2} = 0 \mid N = n) = \mathbf{P}(X_{n+1} = X_{n+2} = 0) = (1-p)^2,$$

which leads to

$$\mathbf{P}(X_{N+1} = X_{N+2} = 0) = \sum_n \mathbf{P}(N = n)(1-p)^2 = (1-p)^2.$$

Interarrival Times

An important random variable associated with the Bernoulli process is the time of the kth success, which we denote by Y_k. A related random variable is the kth interarrival time, denoted by T_k. It is defined by

$$T_1 = Y_1, \qquad T_k = Y_k - Y_{k-1}, \qquad k = 2, 3, \ldots$$

and represents the number of trials following the $k-1$st success until the next success. See Fig. 5.1 for an illustration, and also note that

$$Y_k = T_1 + T_2 + \cdots + T_k.$$

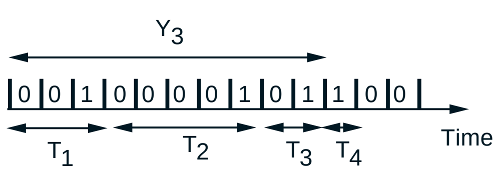


Figure 5.1: Illustration of interarrival times. In this example, $T_1 = 3$, $T_2 = 5$, $T_3 = 2$, $T_4 = 1$. Furthermore, $Y_1 = 3$, $Y_2 = 8$, $Y_3 = 10$, $Y_4 = 11$.

We have already seen that the time T_1 until the first success is a geometric random variable with parameter p. Having had a success at time T_1, the future is a fresh-starting Bernoulli process. Thus, the number of trials T_2 until the next success has the same geometric PMF. Furthermore, past trials (up to and including time T_1) are independent of future trials (from time $T_1 + 1$ onward). Since T_2 is determined exclusively by what happens in these future trials, we see that T_2 is independent of T_1. Continuing similarly, we conclude that the random variables $T_1, T_2, T_3, \ldots$ are independent and all have the same geometric distribution.

This important observation leads to an alternative, but equivalent way of describing the Bernoulli process, which is sometimes more convenient to work with.

Alternative Description of the Bernoulli Process

1. Start with a sequence of independent geometric random variables T_1, $T_2, \ldots$, with common parameter p, and let these stand for the interarrival times.
2. Record a success (or arrival) at times T_1, $T_1 + T_2$, $T_1 + T_2 + T_3$, etc.

Example 5.3. A computer executes two types of tasks, priority and nonpriority, and operates in discrete time units (*slots*). A priority task arises with probability p at the beginning of each slot, independently of other slots, and requires one full slot to complete. A nonpriority task is executed at a given slot only if no priority task is available. In this context, it may be important to know the probabilistic properties of the time intervals available for nonpriority tasks.

With this in mind, let us call a slot *busy* if within this slot, the computer executes a priority task, and otherwise let us call it *idle*. We call a string of idle (or busy) slots, flanked by busy (or idle, respectively) slots, an *idle period* (or *busy period*, respectively). Let us derive the PMF, mean, and variance of the following random variables (cf. Fig. 5.2):

(a) $T =$ the time index of the first idle slot;

(b) $B =$ the length (number of slots) of the first busy period;

(c) $I =$ the length of the first idle period.

We recognize T as a geometrically distributed random variable with parameter $1 - p$. Its PMF is

$$p_T(k) = p^{k-1}(1-p), \qquad k = 1, 2, \ldots.$$

Its mean and variance are

$$\mathbf{E}[T] = \frac{1}{1-p}, \qquad \mathrm{var}(T) = \frac{p}{(1-p)^2}.$$

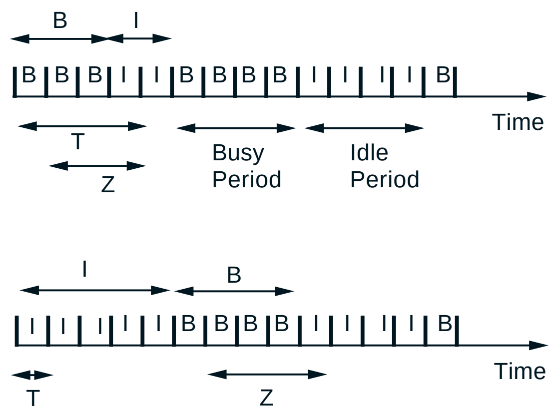


Figure 5.2: Illustration of busy (B) and idle (I) periods in Example 5.3. In the top diagram, $T = 4$, $B = 3$, and $I = 2$. In the bottom diagram, $T = 1$, $I = 5$, and $B = 4$.

Let us now consider the first busy period. It starts with the first busy slot, call it slot L. (In the top diagram in Fig. 5.2, $L = 1$; in the bottom diagram, $L = 6$.) The number Z of subsequent slots until (and including) the first subsequent idle slot has the same distribution as T, because the Bernoulli process starts fresh at time $L + 1$. We then notice that $Z = B$ and conclude that B has the same PMF as T.

If we reverse the roles of idle and busy slots, and interchange p with $1-p$, we see that the length I of the first idle period has the same PMF as the time index of the first busy slot, so that

$$p_I(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots, \qquad \mathbf{E}[I] = \frac{1}{p}, \quad \mathrm{var}(I) = \frac{1-p}{p^2}.$$

We finally note that the argument given here also works for the second, third, etc. busy (or idle) period. Thus the PMFs calculated above apply to the ith busy and idle period, for any i.

The kth Arrival Time

The time Y_k of the kth success is equal to the sum $Y_k = T_1 + T_2 + \cdots + T_k$ of k independent identically distributed geometric random variables. This allows us to derive formulas for the mean, variance, and PMF of Y_k, which are given in the table that follows.

Properties of the kth Arrival Time

- The kth arrival time is equal to the sum of the first k interarrival times

$$Y_k = T_1 + T_2 + \cdots + T_k,$$

  and the latter are independent geometric random variables with common parameter p.

- The mean and variance of Y_k are given by

$$\mathbf{E}[Y_k] = \mathbf{E}[T_1] + \cdots + \mathbf{E}[T_k] = \frac{k}{p},$$

$$\mathrm{var}(Y_k) = \mathrm{var}(T_1) + \cdots + \mathrm{var}(T_k) = \frac{k(1-p)}{p^2}.$$

- The PMF of Y_k is given by

$$p_{Y_k}(t) = \binom{t-1}{k-1} p^k (1-p)^{t-k}, \qquad t = k, k+1, \ldots,$$

  and is known as the **Pascal PMF of order** k.

To verify the formula for the PMF of Y_k, we first note that Y_k cannot be smaller than k. For $t \geq k$, we observe that the event $\{Y_k = t\}$ (the kth success comes at time t) will occur if and only if both of the following two events A and B occur:

(a) event A: trial t is a success;

(b) event B: exactly $k-1$ successes occur in the first $t-1$ trials.

The probabilities of these two events are

$$\mathbf{P}(A) = p$$

and

$$\mathbf{P}(B) = \binom{t-1}{k-1} p^{k-1} (1-p)^{t-k},$$

respectively. In addition, these two events are independent (whether trial t is a success or not is independent of what happened in the first $t-1$ trials). Therefore,

$$p_{Y_k}(t) = \mathbf{P}(Y_k = t) = \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B) = \binom{t-1}{k-1} p^k (1-p)^{t-k},$$

as claimed.

Example 5.4. In each minute of basketball play, Alice commits a single foul with probability p and no foul with probability $1-p$. The number of fouls in different minutes are assumed to be independent. Alice will foul out of the game once she commits her sixth foul, and will play 30 minutes if she does not foul out. What is the PMF of Alice's playing time?

We model fouls as a Bernoulli process with parameter p. Alice's playing time Z is equal to Y_6, the time until the sixth foul, except if Y_6 is larger than 30, in which case, her playing time is 30, the duration of the game; that is, $Z = \min\{Y_6, 30\}$. The random variable Y_6 has a Pascal PMF of order 6, which is given by

$$p_{Y_6}(t) = \binom{t-1}{5} p^6 (1-p)^{t-6}, \qquad t = 6, 7, \ldots$$

To determine the PMF $p_Z(z)$ of Z, we first consider the case where z is between 6 and 29. For z in this range, we have

$$p_Z(z) = \mathbf{P}(Z = z) = \mathbf{P}(Y_6 = z) = \binom{z-1}{5} p^6 (1-p)^{z-6}, \qquad z = 6, 7, \ldots, 29.$$

The probability that $Z = 30$ is then determined from

$$p_Z(30) = 1 - \sum_{z=6}^{29} p_Z(z).$$

Splitting and Merging of Bernoulli Processes

Starting with a Bernoulli process in which there is a probability p of an arrival at each time, consider **splitting** it as follows. Whenever there is an arrival, we choose to either keep it (with probability q), or to discard it (with probability $1-q$); see Fig. 5.3. Assume that the decisions to keep or discard are independent for different arrivals. If we focus on the process of arrivals that are kept, we see that it is a Bernoulli process: in each time slot, there is a probability pq of a kept arrival, independently of what happens in other slots. For the same reason, the process of discarded arrivals is also a Bernoulli process, with a probability of a discarded arrival at each time slot equal to $p(1-q)$.

In a reverse situation, we start with two *independent* Bernoulli processes (with parameters p and q, respectively) and **merge** them into a single process, as follows. An arrival is recorded in the merged process if and only if there is an arrival in at least one of the two original processes, which happens with probability $p+q-pq$ [one minus the probability $(1-p)(1-q)$ of no arrival in either process.] Since different time slots in either of the original processes are independent, different slots in the merged process are also independent. Thus, the merged process is Bernoulli, with success probability $p+q-pq$ at each time step; see Fig. 5.4.

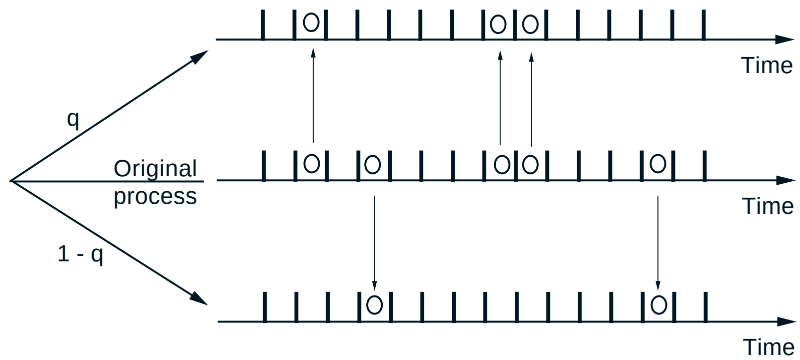


Figure 5.3: Splitting of a Bernoulli process.

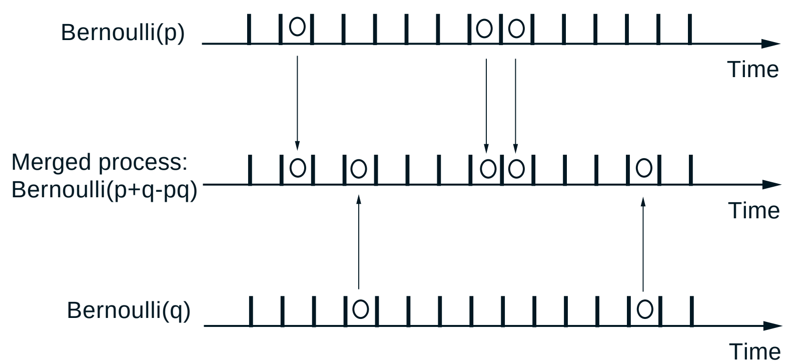


Figure 5.4: Merging of independent Bernoulli process.

Splitting and merging of Bernoulli (or other) arrival processes arises in many contexts. For example, a two-machine work center may see a stream of arriving parts to be processed and split them by sending each part to a randomly chosen machine. Conversely, a machine may be faced with arrivals of different types that can be merged into a single arrival stream.

The Poisson Approximation to the Binomial

The number of successes in n independent Bernoulli trials is a binomial random variable with parameters n and p, and its mean is np. In this subsection, we concentrate on the special case where n is large but p is small, so that the mean np has a moderate value. A situation of this type arises when one passes from discrete to continuous time, a theme to be picked up in the next section. For some more examples, think of the number of airplane accidents on any given day:

there is a large number of trials (airplane flights), but each one has a very small probability of being involved in an accident. Or think of counting the number of typos in a book: there is a large number n of words, but a very small probability of misspelling each one.

Mathematically, we can address situations of this kind, by letting n grow while simultaneously decreasing p, in a manner that keeps the product np at a constant value λ. In the limit, it turns out that the formula for the binomial PMF simplifies to the Poisson PMF. A precise statement is provided next, together with a reminder of some of the properties of the Poisson PMF that were derived in earlier chapters.

Poisson Approximation to the Binomial

- A Poisson random variable Z with parameter λ takes nonnegative integer values and is described by the PMF

$$p_Z(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots.$$

  Its mean and variance are given by

$$\mathbf{E}[Z] = \lambda, \qquad \mathrm{var}(Z) = \lambda.$$

- For any fixed nonnegative integer k, the binomial probability

$$p_S(k) = \frac{n!}{(n-k)!\,k!}p^k(1-p)^{n-k}$$

  converges to $p_Z(k)$, when we take the limit as $n \to \infty$ and $p = \lambda/n$, while keeping λ constant.

- In general, the Poisson PMF is a good approximation to the binomial as long as $\lambda = np$, n is very large, and p is very small.

The verification of the limiting behavior of the binomial probabilities was given in Chapter 2 as as an end-of-chapter problem, and is replicated here for convenience. We let $p = \lambda/n$ and note that

$$\begin{aligned} p_S(k) &= \frac{n!}{(n-k)!\,k!}p^k(1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!}\cdot\frac{\lambda^k}{n^k}\cdot\left(1-\frac{\lambda}{n}\right)^{n-k}. \end{aligned}$$

$$= \frac{n}{n} \cdot \frac{(n-1)}{n} \cdots \frac{(n-k+1)}{n} \cdot \frac{\lambda^k}{k!} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Let us focus on a fixed k and let $n \to \infty$. Each one of the ratios $(n-1)/n$, $(n-2)/n, \ldots, (n-k+1)/n$ converges to 1. Furthermore,†

$$\left(1 - \frac{\lambda}{n}\right)^{-k} \to 1, \qquad \left(1 - \frac{\lambda}{n}\right)^{n} \to e^{-\lambda}.$$

We conclude that for each fixed k, and as $n \to \infty$, we have

$$p_S(k) \to e^{-\lambda} \frac{\lambda^k}{k!}.$$

Example 5.5. As a rule of thumb, the Poisson/binomial approximation

$$e^{-\lambda} \frac{\lambda^k}{k!} \approx \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n,$$

is valid to several decimal places if $n \geq 100$, $p \leq 0.01$, and $\lambda = np$. To check this, consider the following.

Gary Kasparov, the world chess champion (as of 1999) plays against 100 amateurs in a large simultaneous exhibition. It has been estimated from past experience that Kasparov wins in such exhibitions 99% of his games on the average (in precise probabilistic terms, we assume that he wins each game with probability 0.99, independently of other games). What are the probabilities that he will win 100 games, 98 games, 95 games, and 90 games?

We model the number of games X that Kasparov does *not* win as a binomial random variable with parameters $n = 100$ and $p = 0.01$. Thus the probabilities that he will win 100 games, 98, 95 games, and 90 games are

$$\begin{aligned}
p_X(0) &= (1 - 0.01)^{100} = 0.366, \\
p_X(2) &= \frac{100!}{98!2!} 0.01^2 (1 - 0.01)^{98} = 0.185, \\
p_X(5) &= \frac{100!}{95!5!} 0.01^5 (1 - 0.01)^{95} = 0.00290, \\
p_X(10) &= \frac{100!}{90!10!} 0.01^{10} (1 - 0.01)^{90} = 7.006 \times 10^{-8},
\end{aligned}$$

† We are using here, the well known formula $\lim_{x\to\infty}(1 - \frac{1}{x})^x = e^{-1}$. Letting $x = n/\lambda$, we have $\lim_{n\to\infty}(1 - \frac{\lambda}{n})^{n/\lambda} = e^{-1}$, from which it follows that $\lim_{n\to\infty}(1 - \frac{\lambda}{n})^n = e^{-\lambda}$.

respectively. Now let us check the corresponding Poisson approximations with $\lambda = 100 \cdot 0.01 = 1$. They are:

$$p_Z(0) = e^{-1}\frac{1}{0!} = 0.368,$$
$$p_Z(2) = e^{-1}\frac{1}{2!} = 0.184,$$
$$p_Z(5) = e^{-1}\frac{1}{5!} = 0.00306,$$
$$p_Z(10) = e^{-1}\frac{1}{10!} = 1.001 \times 10^{-8}.$$

By comparing the binomial PMF values $p_X(k)$ with their Poisson approximations $p_Z(k)$, we see that there is close agreement.

Suppose now that Kasparov plays simultaneously just 5 opponents, who are, however, stronger so that his probability of a win per game is 0.9. Here are the binomial probabilities $p_X(k)$ for $n = 5$ and $p = 0.1$, and the corresponding Poisson approximations $p_Z(k)$ for $\lambda = np = 0.5$,

$$p_X(0) = 0.590, \qquad p_Z(0) = 0.605,$$
$$p_X(1) = 0.328, \qquad p_Z(1) = 0.303,$$
$$p_X(2) = 0.0729, \qquad p_Z(2) = 0.0758,$$
$$p_X(3) = 0.0081, \qquad p_Z(3) = 0.0126,$$
$$p_X(4) = 0.00045, \qquad p_Z(4) = 0.0016,$$
$$p_X(5) = 0.00001, \qquad p_Z(5) = 0.00016.$$

We see that the approximation, while not poor, is considerably less accurate than in the case where $n = 100$ and $p = 0.01$.

Example 5.6. A packet consisting of a string of n symbols is transmitted over a noisy channel. Each symbol has probability $p = 0.0001$ of being transmitted in error, independently of errors in the other symbols. How small should n be in order for the probability of incorrect transmission (at least one symbol in error) to be less than 0.001?

Each symbol transmission is viewed as an independent Bernoulli trial. Thus, the probability of a positive number S of errors in the packet is

$$1 - \mathbf{P}(S = 0) = 1 - (1 - p)^n.$$

For this probability to be less than 0.001, we must have $1 - (1 - 0.0001)^n < 0.001$ or

$$n < \frac{\ln 0.999}{\ln 0.9999} = 10.0045.$$

We can also use the Poisson approximation for $\mathbf{P}(S = 0)$, which is $e^{-\lambda}$ with $\lambda = np = 0.0001 \cdot n$, and obtain the condition $1 - e^{-0.0001 \cdot n} < 0.001$, which leads to

$$n < \frac{-\ln 0.999}{0.0001} = 10.005.$$

Given that n must be integer, both methods lead to the same conclusion that n can be at most 10.

5.2 THE POISSON PROCESS

The Poisson process can be viewed as a continuous-time analog of the Bernoulli process and applies to situations where there is no natural way of dividing time into discrete periods.

To see the need for a continuous-time version of the Bernoulli process, let us consider a possible model of traffic accidents within a city. We can start by discretizing time into one-minute periods and record a "success" during every minute in which there is at least one traffic accident. Assuming the traffic intensity to be constant over time, the probability of an accident should be the same during each period. Under the additional (and quite plausible) assumption that different time periods are independent, the sequence of successes becomes a Bernoulli process. Note that in real life, two or more accidents during the same one-minute interval are certainly possible, but the Bernoulli process model does not keep track of the exact number of accidents. In particular, it does not allow us to calculate the expected number of accidents within a given period.

One way around this difficulty is to choose the length of a time period to be very small, so that the probability of two or more accidents becomes negligible. But how small should it be? A second? A millisecond? Instead of answering this question, it is preferable to consider a limiting situation where the length of the time period becomes zero, and work with a continuous time model.

We consider an arrival process that evolves in continuous time, in the sense that any real number t is a possible arrival time. We define

$$P(k,\tau) = \mathbf{P}(\text{there are exactly } k \text{ arrivals during an interval of length } \tau),$$

and assume that this probability is the same for all intervals of the same length τ. We also introduce a positive parameter λ to be referred to as the **arrival rate** or **intensity** of the process, for reasons that will soon be apparent.

Definition of the Poisson Process

An arrival process is called a Poisson process with rate λ if it has the following properties:

(a) **(Time-homogeneity.)** The probability $P(k,\tau)$ of k arrivals is the same for all intervals of the same length τ.

(b) **(Independence.)** The number of arrivals during a particular interval is independent of the history of arrivals outside this interval.

(c) **(Small interval probabilities.)** The probabilities $P(k,\tau)$ satisfy

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

Here, $o(\tau)$ and $o_1(\tau)$ are functions of τ that satisfy

$$\lim_{\tau \to 0} \frac{o(\tau)}{\tau} = 0, \qquad \lim_{\tau \to 0} \frac{o_1(\tau)}{\tau} = 0.$$

The first property states that arrivals are "equally likely" at all times. The arrivals during any time interval of length τ are statistically the same, in the sense that they obey the same probability law. This is a counterpart of the assumption that the success probability p in a Bernoulli process is constant over time.

To interpret the second property, consider a particular interval $[t, t']$, of length $t' - t$. The unconditional probability of k arrivals during that interval is $P(k, t' - t)$. Suppose now that we are given complete or partial information on the arrivals outside this interval. Property (b) states that this information is irrelevant: the conditional probability of k arrivals during $[t, t']$ remains equal to the unconditional probability $P(k, t' - t)$. This property is analogous to the independence of trials in a Bernoulli process.

The third property is critical. The $o(\tau)$ and $o_1(\tau)$ terms are meant to be negligible in comparison to τ, when the interval length τ is very small. They can be thought of as the $O(\tau^2)$ terms in a Taylor series expansion of $P(k, \tau)$. Thus, for small τ, the probability of a single arrival is roughly $\lambda\tau$, plus a negligible term. Similarly, for small τ, the probability of zero arrivals is roughly $1 - \lambda\tau$. Note that the probability of two or more arrivals is

$$1 - P(0, \tau) - P(1, \tau) = -o(\tau) - o_1(\tau),$$

and is negligible in comparison to $P(1, \tau)$ as τ gets smaller and smaller.

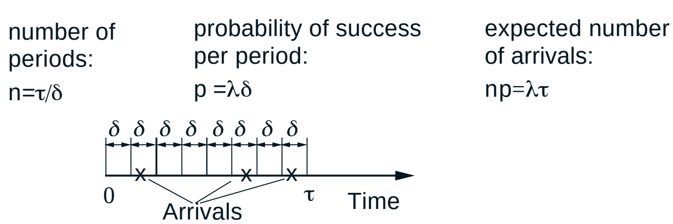


Figure 5.5: Bernoulli approximation of the Poisson process.

Let us now start with a fixed time interval of length τ and partition it into τ/δ periods of length δ, where δ is a very small number; see Fig. 5.5. The probability of more than two arrivals during any period can be neglected, because of property (c) and the preceding discussion. Different periods are independent, by property (b). Furthermore, each period has one arrival with probability approximately equal to $\lambda\delta$, or zero arrivals with probability approximately equal to $1 - \lambda\delta$. Therefore, the process being studied can be approximated by a Bernoulli process, with the approximation becoming more and more accurate the smaller δ is chosen. Thus the probability $P(k, \tau)$ of k arrivals in time τ, is approximately the same as the (binomial) probability of k successes in $n = \tau/\delta$ independent Bernoulli trials with success probability $p = \lambda\delta$ at each trial. While keeping the length τ of the interval fixed, we let the period length δ decrease to zero. We then note that the number n of periods goes to infinity, while the product np remains constant and equal to $\lambda\tau$. Under these circumstances, we saw in the previous section that the binomial PMF converges to a Poisson PMF with parameter $\lambda\tau$. We are then led to the important conclusion that

$$P(k, \tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}, \quad k = 0, 1, \ldots.$$

Note that a Taylor series expansion of $e^{-\lambda\tau}$, yields

$$\begin{aligned} P(0, \tau) &= e^{-\lambda\tau} = 1 - \lambda\tau + O(\tau^2) \\ P(1, \tau) &= \lambda\tau e^{-\lambda\tau} = \lambda\tau - \lambda^2\tau^2 + O(\tau^3) = \lambda\tau + O(\tau^2), \end{aligned}$$

consistent with property (c).

Using our earlier formulas for the mean and variance of the Poisson PMF, we obtain

$$\mathbf{E}[N_\tau] = \lambda\tau, \qquad \mathrm{var}(N_\tau) = \lambda\tau,$$

where N_τ stands for the number of arrivals during a time interval of length τ. These formulas are hardly surprising, since we are dealing with the limit of a binomial PMF with parameters $n = \tau/\delta$, $p = \lambda\delta$, mean $np = \lambda\tau$, and variance $np(1 - p) \approx np = \lambda\tau$.

Let us now derive the probability law for the time T of the first arrival, assuming that the process starts at time zero. Note that we have $T > t$ if and only if there are no arrivals during the interval $[0, t]$. Therefore,

$$F_T(t) = \mathbf{P}(T \le t) = 1 - \mathbf{P}(T > t) = 1 - P(0, t) = 1 - e^{-\lambda t}, \qquad t \ge 0.$$

We then differentiate the CDF $F_T(t)$ of T, and obtain the PDF formula

$$f_T(t) = \lambda e^{-\lambda t}, \qquad t \ge 0,$$

which shows that the time until the first arrival is exponentially distributed with parameter λ. We summarize this discussion in the table that follows. See also Fig. 5.6.

Random Variables Associated with the Poisson Process and their Properties

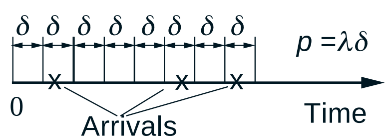
- **The Poisson with parameter $\lambda\tau$.** This is the number N_τ of arrivals in a Poisson process with rate λ, over an interval of length τ. Its PMF, mean, and variance are

$$p_{N_\tau}(k) = P(k,\tau) = \frac{(\lambda\tau)^k e^{-\lambda\tau}}{k!}, \quad k = 0, 1, \ldots.$$

$$\mathbf{E}[N_\tau] = \lambda\tau, \qquad \mathrm{var}(N_\tau) = \lambda\tau.$$

- **The exponential with parameter λ.** This is the time T until the first arrival. Its PDF, mean, and variance are

$$f_T(t) = \lambda e^{-\lambda t}, \quad t \geq 0, \qquad \mathbf{E}[T] = \frac{1}{\lambda}, \qquad \mathrm{var}(T) = \frac{1}{\lambda^2}.$$



	POISSON	**BERNOULLI**
Times of Arrival	Continuous	Discrete
PMF of # of Arrivals	Poisson	Binomial
Interarrival Time CDF	Exponential	Geometric
Arrival Rate	λ/unit time	p/per trial

Figure 5.6: View of the Bernoulli process as the discrete-time version of the Poisson. We discretize time in small intervals δ and associate each interval with a Bernoulli trial whose parameter is $p = \lambda\delta$. The table summarizes some of the basic correspondences.

Example 5.7. You get email according to a Poisson process at a rate of $\lambda = 0.2$ messages per hour. You check your email every hour. What is the probability of finding 0 and 1 new messages?

These probabilities can be found using the Poisson PMF $(\lambda\tau)^k e^{-\lambda\tau}/k!$, with $\tau = 1$, and $k = 0$ or $k = 1$:

$$\mathbf{P}(0,1) = e^{-0.2} = 0.819, \qquad \mathbf{P}(1,1) = 0.2 \cdot e^{-0.2} = 0.164$$

Suppose that you have not checked your email for a whole day. What is the probability of finding no new messages? We use again the Poisson PMF and obtain

$$\mathbf{P}(0,24) = e^{-0.2\cdot 24} = 0.008294.$$

Alternatively, we can argue that the event of no messages in a 24-hour period is the intersection of the events of no messages during each of 24 hours. These latter events are independent and the probability of each is $\mathbf{P}(0,1) = e^{-0.2}$, so

$$\mathbf{P}(0,24) = \big(\mathbf{P}(0,1)\big)^{24} = \big(e^{-0.2}\big)^{24} = 0.008294,$$

which is consistent with the preceding calculation method.

Example 5.8. Sum of Independent Poisson Random Variables. Arrivals of customers at the local supermarket are modeled by a Poisson process with a rate of $\lambda = 10$ customers per minute. Let M be the number of customers arriving between 9:00 and 9:10. Also, let N be the number of customers arriving between 9:30 and 9:35. What is the distribution of $M + N$?

We notice that M is Poisson with parameter $\mu = 10\cdot 10 = 100$ and N is Poisson with parameter $\nu = 10\cdot 5 = 50$. Furthermore, M and N are independent. As shown in Section 4.1, using transforms, $M+N$ is Poisson with parameter $\mu+\nu = 150$. We will now proceed to derive the same result in a more direct and intuitive manner.

Let $\tilde{N}$ be the number of customers that arrive between 9:10 and 9:15. Note that $\tilde{N}$ has the same distribution as N (Poisson with parameter 50). Furthermore, $\tilde{N}$ is also independent of N. Thus, the distribution of $M + N$ is the same as the distribution of $M + \tilde{N}$. But $M + \tilde{N}$ is the number of arrivals during an interval of length 15, and has therefore a Poisson distribution with parameter $10 \cdot 15 = 150$.

This example makes a point that is valid in general. The probability of k arrivals during a *set* of times of total length τ is always given by $P(k, \tau)$, even if that set is not an interval. (In this example, we dealt with the set $[9:00, 9:10] \cup [9:30, 9:35]$, of total length 15.)

Example 5.9. During rush hour, from 8 am to 9 am, traffic accidents occur according to a Poisson process with a rate μ of 5 accidents per hour. Between 9 am and 11 am, they occur as an independent Poisson process with a rate ν of 3 accidents per hour. What is the PMF of the total number of accidents between 8 am and 11 am?

This is the sum of two independent Poisson random variables with parameters 5 and $3 \cdot 2 = 6$, respectively. Since the sum of independent Poisson random variables is also Poisson, the total number of accidents has a Poisson PMF with parameter 5+6=11.

Independence and Memorylessness

The Poisson process has several properties that parallel those of the Bernoulli process, including the independence of nonoverlapping time sets, a fresh-start property, and the memorylessness of the interarrival time distribution. Given that the Poisson process can be viewed as a limiting case of a Bernoulli process, the fact that it inherits the qualitative properties of the latter should be hardly surprising.

(a) **Independence of nonoverlapping sets of times.** Consider two disjoint sets of times A and B, such as $A = [0, 1] \cup [4, \infty)$ and $B = [1.5, 3.6]$, for example. If U and V are random variables that are completely determined by what happens during A (respectively, B), then U and V are independent. This is a consequence of the second defining property of the Poisson process.

(b) **Fresh-start property.** As a special case of the preceding observation, we notice that the history of the process until a particular time t is independent from the future of the process. Furthermore, if we focus on that portion of the Poisson process that starts at time t, we observe that it inherits the defining properties of the original process. For this reason, *the portion of the Poisson process that starts at any particular time $t > 0$ is a probabilistic replica of the Poisson process starting at time 0, and is independent of the portion of the process prior to time t.* Thus, we can say that the Poisson process *starts afresh* at each time instant.

(c) **Memoryless interarrival time distribution.** We have already seen that the geometric PMF (interarrival time in the Bernoulli process) is memoryless: the number of *remaining trials* until the first future arrival does not depend on the past. The exponential PDF (interarrival time in the Poisson process) has a similar property: given the current time t and the past history, the future is a fresh-starting Poisson process, hence the *remaining time* until the next arrival has the same exponential distribution. In particular, if T is the time of the first arrival and if we are told that $T > t$, then the remaining time $T - t$ is exponentially distributed, with the same parameter λ. For an algebraic derivation of this latter fact, we first use the exponential CDF to obtain $\mathbf{P}(T > t) = e^{-\lambda t}$. We then note that for all positive scalars s and t, we have

$$\begin{aligned}\mathbf{P}(T > t + s \,|\, T > t) &= \frac{\mathbf{P}(T > t + s,\, T > t)}{\mathbf{P}(T > t)} \\ &= \frac{\mathbf{P}(T > t + s)}{\mathbf{P}(T > t)} \\ &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ &= e^{-\lambda s}.\end{aligned}$$

Here are some examples of reasoning based on the memoryless property.

Example 5.10. You and your partner go to a tennis court, and have to wait until the players occupying the court finish playing. Assume (somewhat unrealistically) that their playing time has an exponential PDF. Then the PDF of your waiting time (equivalently, their remaining playing time) also has the same exponential PDF, regardless of when they started playing.

Example 5.11. When you enter the bank, you find that all three tellers are busy serving other customers, and there are no other customers in queue. Assume that the service times for you and for each of the customers being served are independent identically distributed exponential random variables. What is the probability that you will be the last to leave?

The answer is 1/3. To see this, focus at the moment when you start service with one of the tellers. Then, the remaining time of each of the other two customers being served, as well as your own remaining time, have the same PDF. Therefore, you and the other two customers have equal probability 1/3 of being the last to leave.

Interarrival Times

An important random variable associated with a Poisson process that starts at time 0, is the time of the kth arrival, which we denote by Y_k. A related random variable is the kth interarrival time, denoted by T_k. It is defined by

$$T_1 = Y_1, \qquad T_k = Y_k - Y_{k-1}, \qquad k = 2, 3, \ldots$$

and represents the amount of time between the $k-1$st and the kth arrival. Note that

$$Y_k = T_1 + T_2 + \cdots + T_k.$$

We have already seen that the time T_1 until the first arrival is an exponential random variable with parameter λ. Starting from the time T_1 of the first arrival, the future is a fresh-starting Poisson process. Thus, the time until the next arrival has the same exponential PDF. Furthermore, the past of the process (up to time T_1) is independent of the future (after time T_1). Since T_2 is determined exclusively by what happens in the future, we see that T_2 is independent of T_1. Continuing similarly, we conclude that the random variables $T_1, T_2, T_3, \ldots$ are independent and all have the same exponential distribution.

This important observation leads to an alternative, but equivalent, way of describing the Poisson process.†

Alternative Description of the Poisson Process

1. Start with a sequence of independent exponential random variables $T_1, T_2, \ldots$, with common parameter λ, and let these stand for the interarrival times.
2. Record an arrival at times T_1, $T_1 + T_2$, $T_1 + T_2 + T_3$, etc.

The kth Arrival Time

The time Y_k of the kth arrival is equal to the sum $Y_k = T_1 + T_2 + \cdots + T_k$ of k independent identically distributed exponential random variables. This allows us to derive formulas for the mean, variance, and PMF of Y_k, which are given in the table that follows.

Properties of the kth Arrival Time

- The kth arrival time is equal to the sum of the first k interarrival times

$$Y_k = T_1 + T_2 + \cdots + T_k,$$

and the latter are independent exponential random variables with common parameter λ.

† In our original definition, a process was called Poisson if it possessed certain properties. However, the astute reader may have noticed that we have not so far established that there exists a process with the required properties. In an alternative line of development, we could have defined the Poisson process by the alternative description given here, and such a process is clearly well-defined: we start with a sequence of independent interarrival times, from which the arrival times are completely determined. Starting with this definition, it is then possible to establish that the process satisfies all of the properties that were postulated in our original definition.

- The mean and variance of Y_k are given by

$$\mathbf{E}[Y_k] = \mathbf{E}[T_1] + \cdots + \mathbf{E}[T_k] = \frac{k}{\lambda},$$

$$\mathrm{var}(Y_k) = \mathrm{var}(T_1) + \cdots + \mathrm{var}(T_k) = \frac{k}{\lambda^2}.$$

- The PDF of Y_k is given by

$$f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}$$

and is known as the **Erlang PDF of order** k.

To evaluate the PDF f_{Y_k} of Y_k, we can argue that for a small δ, the product $\delta \cdot f_{Y_k}(y)$ is the probability that the kth arrival occurs between times y and $y+\delta$.† When δ is very small, the probability of more than one arrival during the interval $[y, y+\delta]$ is negligible. Thus, the kth arrival occurs between y and $y+\delta$ if and only if the following two events A and B occur:

(a) event A: there is an arrival during the interval $[y, y+\delta]$;

(b) event B: there are exactly $k-1$ arrivals before time y.

The probabilities of these two events are

$$\mathbf{P}(A) \approx \lambda\delta, \qquad \text{and} \qquad \mathbf{P}(B) = P(k-1, y) = \frac{\lambda^{k-1} y^{k-1} e^{-\lambda y}}{(k-1)!}.$$

† For an alternative derivation that does not rely on approximation arguments, note that for a given $y \geq 0$, the event $\{Y_k \leq y\}$ is the same as the event

$$\{\text{number of arrivals in the interval } [0, y] \geq k\}.$$

Thus the CDF of Y_k is given by

$$F_{Y_k}(y) = \mathbf{P}(Y_k \leq y) = \sum_{n=k}^{\infty} P(n, y) = 1 - \sum_{n=0}^{k-1} P(n, y) = 1 - \sum_{n=0}^{k-1} \frac{(\lambda y)^n e^{-\lambda y}}{n!}.$$

The PDF of Y_k can be obtained by differentiating the above expression, which by straightforward calculation yields the Erlang PDF formula

$$f_{Y_k}(y) = \frac{d}{dy} F_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}.$$

Since A and B are independent, we have

$$\delta f_{Y_k}(y) \approx \mathbf{P}(y \leq Y_k \leq y+\delta) \approx \mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B) \approx \lambda\delta \frac{\lambda^{k-1} y^{k-1} e^{-\lambda y}}{(k-1)!},$$

from which we obtain

$$f_{Y_k}(y) = \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}, \qquad y \geq 0.$$

Example 5.12. You call the IRS hotline and you are told that you are the 56th person in line, excluding the person currently being served. Callers depart according to a Poisson process with a rate of $\lambda = 2$ per minute. How long will you have to wait on the average until your service starts, and what is the probability you will have to wait for more than an hour?

By the memoryless property, the remaining service time of the person currently being served is exponentially distributed with parameter 2. The service times of the 55 persons ahead of you are also exponential with the same parameter, and all of these random variables are independent. Thus, your waiting time Y is Erlang of order 56, and

$$\mathbf{E}[Y] = \frac{56}{\lambda} = 28.$$

The probability that you have to wait for more than an hour is given by the formula

$$\mathbf{P}(Y \geq 60) = \int_{60}^{\infty} \frac{\lambda^{56} y^{55} e^{-\lambda y}}{55!}\, dy.$$

Computing this probability is quite tedious. In Chapter 7, we will discuss a much easier way to compute approximately this probability. This is done using the central limit theorem, which allows us to approximate the CDF of the sum of a large number of random variables with a normal CDF and then to calculate various probabilities of interest by using the normal tables.

Splitting and Merging of Poisson Processes

Similar to the case of a Bernoulli process, we can start with a Poisson process with rate λ and **s**plit it, as follows: each arrival is kept with probability p and discarded with probability $1-p$, independently of what happens to other arrivals. In the Bernoulli case, we saw that the result of the splitting was also a Bernoulli process. In the present context, the result of the splitting turns out to be a Poisson process with rate λp.

Alternatively, we can start with two independent Poisson processes, with rates λ_1 and λ_2, and merge them by recording an arrival whenever an arrival occurs in either process. It turns out that the merged process is also Poisson with rate $\lambda_1 + \lambda_2$. Furthermore, any particular arrival of the merged process has probability $\lambda_1/(\lambda_1+\lambda_2)$ of originating from the first process and probability $\lambda_2/(\lambda_1+\lambda_2)$ of originating from the second, independently of all other arrivals and their origins.

We discuss these properties in the context of some examples, and at the same time provide a few different arguments to establish their validity.

Example 5.13. Splitting of Poisson Processes. A packet that arrives at a node of a data network is either a local packet which is destined for that node (this happens with probability p), or else it is a transit packet that must be relayed to another node (this happens with probability $1-p$). Packets arrive according to a Poisson process with rate λ, and each one is a local or transit packet independently of other packets and of the arrival times. As stated above, the process of *local* packet arrivals is Poisson with rate λp. Let us see why.

We verify that the process of local packet arrivals satisfies the defining properties of a Poisson process. Since λ and p are constant (do not change with time), the first property (time homogeneity) clearly holds. Furthermore, there is no dependence between what happens in disjoint time intervals, verifying the second property. Finally, if we focus on an interval of small length δ, the probability of a local arrival is approximately the probability that there is a packet arrival, and that this turns out to be a local one, i.e., $\lambda\delta \cdot p$. In addition, the probability of two or more local arrivals is negligible in comparison to δ, and this verifies the third property. We conclude that local packet arrivals form a Poisson process and, in particular, the number L_τ of such arrivals during an interval of length τ has a Poisson PMF with parameter $p\lambda\tau$.

Let us now rederive the Poisson PMF of L_τ using transforms. The total number of packets N_τ during an interval of length τ is Poisson with parameter $\lambda\tau$. For $i = 1, \ldots, N_\tau$, let X_i be a Bernoulli random variable which is 1 if the ith packet is local, and 0 if not. Then, the random variables $X_1, X_2, \ldots$ form a Bernoulli process with success probability p. The number of local packets is the number of "successes," i.e.,

$$L_\tau = X_1 + \cdots + X_{N_\tau}.$$

We are dealing here with the sum of a random number of independent random variables. As discussed in Section 4.4, the transform associated with L_τ is found by starting with the transform associated with N_τ, which is

$$M_{N_\tau}(s) = e^{\lambda\tau(e^s-1)},$$

and replacing each occurrence of e^s by the transform associated with X_i, which is

$$M_X(s) = 1 - p + pe^s.$$

We obtain

$$M_{L_\tau}(s) = e^{\lambda\tau(1-p+pe^s-1)} = e^{\lambda\tau p(e^s-1)}.$$

We observe that this is the transform of a Poisson random variable with parameter $\lambda\tau p$, thus verifying our earlier statement for the PMF of L_τ.

We conclude with yet another method for establishing that the local packet process is Poisson. Let $T_1, T_2, \dots$ be the interarrival times of packets of any type; these are independent exponential random variables with parameter λ. Let K be the total number of arrivals up to and including the first local packet arrival. In particular, the time S of the first local packet arrival is given by

$$S = T_1 + T_2 + \cdots + T_K.$$

Since each packet is a local one with probability p, independently of the others, and by viewing each packet as a trial which is successful with probability p, we recognize K as a geometric random variable with parameter p. Since the nature of the packets is independent of the arrival times, K is independent from the interarrival times. We are therefore dealing with a sum of a random (geometrically distributed) number of exponential random variables. We have seen in Chapter 4 (cf. Example 4.21) that such a sum is exponentially distributed with parameter λp. Since the interarrival times between successive local packets are clearly independent, it follows that the local packet arrival process is Poisson with rate λp.

Example 5.14. Merging of Poisson Processes. People with letters to mail arrive at the post office according to a Poisson process with rate λ_1, while people with packages to mail arrive according to an independent Poisson process with rate λ_2. As stated earlier the merged process, which includes arrivals of both types, is Poisson with rate $\lambda_1 + \lambda_2$. Let us see why.

First, it should be clear that the merged process satisfies the time-homogeneity property. Furthermore, since different intervals in each of the two arrival processes are independent, the same property holds for the merged process. Let us now focus on a small interval of length δ. Ignoring terms that are negligible compared to δ, we have

$$\mathbf{P}(0 \text{ arrivals in the merged process}) \approx (1-\lambda_1\delta)(1-\lambda_2\delta) \approx 1 - (\lambda_1+\lambda_2)\delta,$$

$$\mathbf{P}(1 \text{ arrival in the merged process}) \approx \lambda_1\delta(1-\lambda_2\delta) + (1-\lambda_1\delta)\lambda_2\delta \approx (\lambda_1+\lambda_2)\delta,$$

and the third property has been verified.

Given that an arrival has just been recorded, what is the probability that it is an arrival of a person with a letter to mail? We focus again on a small interval of length δ around the current time, and we seek the probability

$$\mathbf{P}(1 \text{ arrival of person with a letter} \,|\, 1 \text{ arrival}).$$

Using the definition of conditional probabilities, and ignoring the negligible probability of more than one arrival, this is

$$\frac{\mathbf{P}(1 \text{ arrival of person with a letter})}{\mathbf{P}(1 \text{ arrival})} \approx \frac{\lambda_1\delta}{(\lambda_1+\lambda_2)\delta} = \frac{\lambda_1}{\lambda_1+\lambda_2}.$$

Example 5.15. Competing Exponentials. Two light bulbs have independent and exponentially distributed lifetimes $T^{(1)}$ and $T^{(2)}$, with parameters λ_1 and λ_2, respectively. What is the distribution of the first time $Z = \min\{T^{(1)}, T^{(2)}\}$ at which a bulb burns out?

We can treat this as an exercise in derived distributions. For all $z \geq 0$, we have,

$$\begin{aligned} F_Z(z) &= \mathbf{P}\big(\min\{T^{(1)}, T^{(2)}\} \leq z\big) \\ &= 1 - \mathbf{P}\big(\min\{T^{(1)}, T^{(2)}\} > z\big) \\ &= 1 - \mathbf{P}(T^{(1)} > z,\, T^{(2)} > z) \\ &= 1 - \mathbf{P}(T^{(1)} > z)\mathbf{P}(T^{(2)} > z) \\ &= 1 - e^{-\lambda_1 z} e^{-\lambda_2 z} \\ &= 1 - e^{-(\lambda_1+\lambda_2)z}. \end{aligned}$$

This is recognized as the exponential CDF with parameter $\lambda_1 + \lambda_2$. Thus, the minimum of two independent exponentials with parameters λ_1 and λ_2 is an exponential with parameter $\lambda_1 + \lambda_2$.

For a more intuitive explanation of this fact, let us think of $T^{(1)}$ (respectively, $T^{(2)}$) as the times of the first arrival in two independent Poisson process with rate λ_1 (respectively, $T^{(2)}$). If we merge these two Poisson processes, the first arrival time will be $\min\{T^{(1)}, T^{(2)}\}$. But we already know that the merged process is Poisson with rate $\lambda_1 + \lambda_2$, and it follows that the first arrival time, $\min\{T^{(1)}, T^{(2)}\}$, is exponential with parameter $\lambda_1 + \lambda_2$.

The preceding discussion can be generalized to the case of more than two processes. Thus, the total arrival process obtained by merging the arrivals of n independent Poisson processes with arrival rates $\lambda_1, \ldots, \lambda_n$ is Poisson with arrival rate equal to the sum $\lambda_1 + \cdots + \lambda_n$.

Example 5.16. More on Competing Exponentials. Three light bulbs have independent exponentially distributed lifetimes with a common parameter λ. What is the expectation of the time until the last bulb burns out?

We think of the times when each bulb burns out as the first arrival times in independent Poisson processes. In the beginning, we have three bulbs, and the merged process has rate 3λ. Thus, the time T_1 of the first burnout is exponential with parameter 3λ, and mean $1/3\lambda$. Once a bulb burns out, and because of the memorylessness property of the exponential distribution, the remaining lifetimes of the other two lightbulbs are again independent exponential random variables with parameter λ. We thus have *two* Poisson processes running in parallel, and the remaining time T_2 until the first arrival in one of these two processes is now exponential with parameter 2λ and mean $1/2\lambda$. Finally, once a second bulb burns out, we are left with a single one. Using memorylessness once more, the remaining time T_3 until the last bulb burns out is exponential with parameter λ and mean $1/\lambda$. Thus, the expectation of the total time is

$$\mathbf{E}[T_1 + T_2 + T_3] = \frac{1}{3\lambda} + \frac{1}{2\lambda} + \frac{1}{\lambda}.$$

Note that the random variables T_1, T_2, T_3 are independent, because of memorylessness. This also allows us to compute the variance of the total time:

$$\mathrm{var}(T_1 + T_2 + T_3) = \mathrm{var}(T_1) + \mathrm{var}(T_2) + \mathrm{var}(T_3) = \frac{1}{9\lambda^2} + \frac{1}{4\lambda^2} + \frac{1}{\lambda^2}.$$

We close by noting a related and quite deep fact, namely that the sum of a *large* number of (*not* necessarily Poisson) independent arrival processes, can be approximated by a Poisson process with arrival rate equal to the sum of the individual arrival rates. The component processes must have a small rate relative to the total (so that none of them imposes its probabilistic character on the total arrival process) and they must also satisfy some technical mathematical assumptions. Further discussion of this fact is beyond our scope, but we note that it is in large measure responsible for the abundance of Poisson-like processes in practice. For example, the telephone traffic originating in a city consists of many component processes, each of which characterizes the phone calls placed by individual residents. The component processes need not be Poisson; some people for example tend to make calls in batches, and (usually) while in the process of talking, cannot initiate or receive a second call. However, the total telephone traffic is well-modeled by a Poisson process. For the same reasons, the process of auto accidents in a city, customer arrivals at a store, particle emissions from radioactive material, etc., tend to have the character of the Poisson process.

The Random Incidence Paradox

The arrivals of a Poisson process partition the time axis into a sequence of interarrival intervals; each interarrival interval starts with an arrival and ends at the time of the next arrival. We have seen that the lengths of these interarrival intervals are independent exponential random variables with parameter λ and mean $1/\lambda$, where λ is the rate of the process. More precisely, for every k, the length of the kth interarrival interval has this exponential distribution. In this subsection, we look at these interarrival intervals from a different perspective.

Let us fix a time instant t^* and consider the length L of the interarrival interval to which it belongs. For a concrete context, think of a person who shows up at the bus station at some arbitrary time t^* and measures the time from the previous bus arrival until the next bus arrival. The arrival of this person is often referred to as a "random incidence," but the reader should be aware that the term is misleading: t^* is just a particular time instance, not a random variable.

We assume that t^* is much larger than the starting time of the Poisson process so that we can be fairly certain that there has been an arrival prior to time t^*. To avoid the issue of determining how large a t^* is large enough, we can actually assume that the Poisson process has been running forever, so that we can be fully certain that there has been a prior arrival, and that L is well-defined. One might superficially argue that L is the length of a "typical" interarrival interval, and is exponentially distributed, but this turns out to be false. Instead, we will establish that L has an Erlang PDF of order two.

This is known as the *random incidence phenomenon or paradox*, and it can be explained with the help of Fig. 5.7. Let $[U, V]$ be the interarrival interval to which t^* belongs, so that $L = V - U$. In particular, U is the time of the first arrival prior to t^* and V is the time of the first arrival after t^*. We split L into two parts,

$$L = (t^* - U) + (V - t^*),$$

where $t^* - U$ is the elapsed time since the last arrival, and $V - t^*$ is the remaining time until the next arrival. Note that $t^* - U$ is determined by the past history of the process (before t^*), while $V - t^*$ is determined by the future of the process (after time t^*). By the independence properties of the Poisson process, the random variables $t^* - U$ and $V - t^*$ are independent. By the memorylessness property, the Poisson process starts fresh at time t^*, and therefore $V - t^*$ is exponential with parameter λ. The random variable $t^* - U$ is also exponential with parameter λ. The easiest way of seeing this is to realize that if we run a Poisson process backwards in time it remains Poisson; this is because the defining properties of a Poisson process make no reference to whether time moves forward or backward. A more formal argument is obtained by noting that

$$\mathbf{P}(t^* - U > x) = \mathbf{P}\big(\text{no arrivals during } [t^* - x, t^*]\big) = P(0, x) = e^{-\lambda x}, \qquad x \geq 0.$$

We have therefore established that L is the sum of two independent exponential random variables with parameter λ, i.e., Erlang of order two, with mean $2/\lambda$.

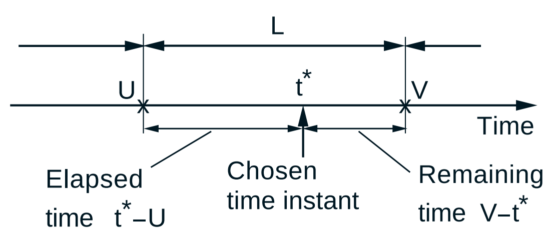


Figure 5.7: Illustration of the random incidence phenomenon. For a fixed time instant t^*, the corresponding interarrival interval $[U, V]$ consists of the elapsed time $t^* - U$ and the remaining time $V - t^*$. These two times are independent and are exponentially distributed with parameter λ, so the PDF of their sum is Erlang of order two.

Random incidence phenomena are often the source of misconceptions and errors, but these can be avoided with careful probabilistic modeling. The key issue is that even though interarrival intervals have length $1/\lambda$ on the average, an observer who arrives at an arbitrary time is more likely to fall in a large rather than a small interarrival interval. As a consequence the expected length seen by the observer is higher, $2/\lambda$ in this case. This point is amplified by the example that follows.

Example 5.17. Random incidence in a non-Poisson arrival process. Buses arrive at a station deterministically, on the hour, and fifteen minutes after the hour. Thus, the interarrival times alternate between 15 and 45 minutes. The average interarrival time is 30 minutes. A person shows up at the bus station at a "random" time. We interpret "random" to mean a time which is uniformly distributed within a particular hour. Such a person falls into an interarrival interval of length 15 with probability 1/4, and an interarrival interval of length 45 with probability 3/4. The expected value of the length of the chosen interarrival interval is

$$15 \cdot \frac{1}{4} + 45 \cdot \frac{3}{4} = 37.5,$$

which is considerably larger than 30, the average interarrival time.