Chapter 3

RENEWAL PROCESSES

3.1 Introduction

Recall that a renewal process is an arrival process in which the interarrival intervals are positive,[1] independent and identically distributed (IID) random variables (rv's). Renewal processes (since they are arrival processes) can be specified in three standard ways, first, by the joint distributions of the arrival epochs $S_1, S_2, \dots$, second, by the joint distributions of the interarrival times $X_1, X_2, \dots$, and third, by the joint distributions of the counting rv's, $N(t)$ for $t > 0$. Recall that $N(t)$ represents the number of arrivals to the system in the interval $(0, t]$.

The simplest characterization is through the interarrival times X_i, since they are IID. Each arrival epoch S_n is simply the sum $X_1 + X_2 + \cdots + X_n$ of n IID rv's. The characterization of greatest interest in this chapter is the renewal counting process, $\{N(t);\ t > 0\}$. Recall from (2.2) and (2.3) that the arrival epochs and the counting rv's are related in each of the following equivalent ways.

$$\{S_n \le t\} = \{N(t) \ge n\}; \qquad \{S_n > t\} = \{N(t) < n\}. \tag{3.1}$$

The reason for calling these processes *renewal processes* is that the process probabilistically starts over at each arrival epoch, S_n. That is, if the nth arrival occurs at $S_n = \tau$, then, counting from $S_n = \tau$, the j^{th} subsequent arrival epoch is at $S_{n+j} - S_n = X_{n+1} + \cdots + X_{n+j}$. Thus, given $S_n = \tau$, $\{N(\tau + t) - N(\tau); t \ge 0\}$ is a renewal counting process with IID interarrival intervals of the same distribution as the original renewal process. This interpretation of arrivals as renewals will be discussed in more detail later.

The major reason for studying renewal processes is that many complicated processes have randomly occurring instants at which the system returns to a state probabilistically equiva-

[1] Renewal processes are often defined in a slightly more general way, allowing the interarrival intervals X_i to include the possibility $1 > \Pr\{X_i = 0\} > 0$. All of the theorems in this chapter are valid under this more general assumption, as can be verified by complicating the proofs somewhat. Allowing $\Pr\{X_i = 0\} > 0$ allows multiple arrivals at the same instant, which makes it necessary to allow $N(0)$ to take on positive values, and appears to inhibit intuition about renewals. Exercise 3.3 shows how to view these more general renewal processes while using the definition here, thus showing that the added generality is not worth much.

lent to the starting state. These embedded renewal epochs allow us to separate the long term behavior of the process (which can be studied through renewal theory) from the behavior within each renewal period.

Example 3.1.1. Consider a G/G/m queue. The customer arrivals to a G/G/m queue form a renewal counting process, $\{N(t); t > 0\}$. Each arriving customer waits in the queue until one of m identical servers is free to serve it. The service time required by each customer is a rv, IID over customers, and independent of arrival times and servers. We define a new renewal counting process, $\{N'(t);\, t \geq 0\}$, for which the renewal epochs are those particular arrival epochs in the original process $\{N(t);\; t > 0\}$ at which an arriving customer sees an empty system[2] (i.e., no customer in queue and none in service). To make the time origin probabilistically identical to the renewal epochs in this new renewal process, we assume that a customer arrives to an empty queue at time 0, but does not count as an arrival in $(0, t]$. There are two renewal counting processes, $\{N(t); t > 0\}$ and $\{N'(t);\, t \geq 0\}$ in this example, and the renewals in the second process are those particular customer arrivals that arrive to an empty system. In most situations, we use the words *arrivals* and *renewals* interchangably. For this type of example, the word *arrival* is used for the counting process $\{N(t);\; t > 0\}$ and the word *renewal* is used for $\{N'(t);\; t > 0\}$.

Throughout our study of renewal processes, we use $\overline{X}$ and $\mathsf{E}[X]$ interchangeably to denote the mean inter-renewal interval, and use σ^2 to denote the variance of the inter-renewal interval. We will usually assume that $\overline{X}$ is finite, but, except where explicitly stated, we need not assume that σ^2 is finite. This means, first, that σ^2 need not be calculated (which is often difficult if renewals are embedded into a more complex process), and second, since modeling errors on the far tails of the inter-renewal distribution typically affect σ^2 more than $\overline{X}$, the results are relatively robust to these kinds of modeling errors.

Much of this chapter will be devoted to understanding the behavior of $N(t)$ and $N(t)/t$ as t becomes large. As might appear to be intuitively obvious, and as is proven in Exercise 3.1, $N(t)$ is a rv (*i.e.*, not defective). Also, as proven in Exercise 3.2 $\mathsf{E}[N(t)] < \infty$ for all $t > 0$. It is then also clear that $N(t)/t$, which is interpreted as the time-average renewal rate over (0,t], is also a rv with finite expectation.

One of the major results about renewal theory, which we establish shortly, is that, with probability 1, the family of random variables, $\{N(t)/t;\; t > 0\}$, has a limiting value, $\lim_{t\to\infty} N(t)/t$, equal to $1/\overline{X}$. This result is called the strong law of large numbers for renewal processes. We shall often refer to it by the less precise statement that the time-average renewal rate is $1/\overline{X}$. This result is an analog (and direct consequence) of the strong law of large numbers, Theorem 1.6.

Another important result is the elementary renewal theorem, which states that $\mathsf{E}[N(t)/t]$ also approaches $1/\overline{X}$ as $t \to \infty$. It seems surprising that this does not follow easily from the strong law for renewal processes, but in fact it doesn't, and we shall develop several widely useful results such as Wald's equality, in establishing this theorem. The final major result

[2]Readers who accept without question that $\{N'(t)\; t > 0\}$ is a renewal process here should feel happy about their probabilistic intuition, but should be concerned about a certain degree of carelessness. We return later to provide a sound argument why $\{N'(t)\; t > 0\}$ is indeed a renewal process.

is Blackwell's theorem, which shows that, for appropriate values of δ, the expected number of renewals in an interval $(t, t+\delta]$ approaches $\delta/\overline{X}$ as $t \to \infty$. We shall thus interpret $1/\overline{X}$ as an ensemble-average renewal rate. This rate is the same as the above time-average renewal rate. We shall see the benefits of being able to work with both time-averages and ensemble-averages.

3.2 Strong law of large numbers for renewal processes

To get an intuitive idea why $N(t)/t$ should approach $1/\overline{X}$ for large t, consider Figure 3.1. For any given sample function of $\{N(t);\ t > 0\}$, note that, for any given t, $N(t)/t$ is the slope of a straight line from the origin to the point $(t, N(t))$. As t increases, this slope decreases in the interval between each adjacent pair of arrival epochs and then jumps up at the next arrival epoch. In order to express this as an equation, note that t lies between the $N(t)$th arrival (which occurs at $S_{N(t)}$) and the $(N(t)+1)$th arrival (which occurs at $S_{N(t)+1}$). Thus

$$\frac{N(t)}{S_{N(t)}} \geq \frac{N(t)}{t} > \frac{N(t)}{S_{N(t)+1}}. \tag{3.2}$$

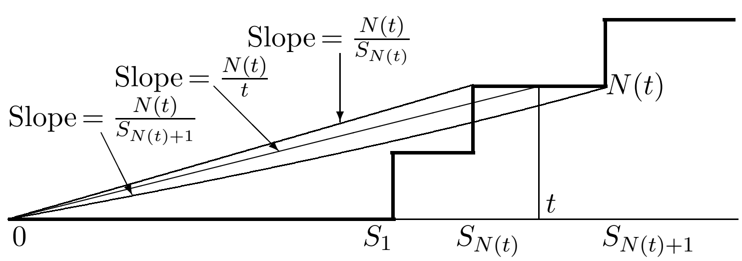


Figure 3.1: Comparison of $N(t)/t$ with $\frac{N(t)}{S_{N(t)}}$ and $\frac{N(t)}{S_{N(t)+1}}$.

We want to show intuitively why the slope $N(t)/t$ in the figure approaches $1/\overline{X}$ as $t \to \infty$. As t increases, we would guess that $N(t)$ increases without bound. Assuming this, the left side of (3.2) increases with increasing t as $1/S_1, 2/S_2, \dots, n/S_n, \dots$, where $n = N(t)$. Since S_n/n converges to $\overline{X}$ WP1 from the strong law of large numbers, we might be brave enough or insightful enough to guess that n/S_n converges to $1/\overline{X}$.

We are now ready to state the strong law for renewal processes as a theorem. Before proving the theorem, we state the above two guesses carefully and prove their validity.

Theorem 3.1 (Strong Law for Renewal Processes). *For a renewal process with mean inter-renewal interval* $\overline{X} < \infty$, $\lim_{t\to\infty} N(t)/t = 1/\overline{X}$ *with probability* 1.

Lemma 3.1. *Let* $\{N(t);\, t > 0\}$ *be a renewal counting process with inter-renewal rv's* $\{X_n;\ n \geq 1\}$. *Then (whether or not* $\overline{X} < \infty$*),* $\lim_{t\to\infty} N(t) = \infty$ *with probability* 1 *and* $\lim_{t\to\infty} \mathsf{E}[N(t)] = \infty$.

Proof of Lemma 3.1: Note that for each sample point $\omega \in \Omega$, $N(t,\omega)$ is a nondecreasing function of t and thus either has a finite limit or an infinite limit. Using (3.1), the probability that this limit is finite with value less than any given n is

$$\lim_{t\to\infty} \Pr\{N(t) < n\} \;=\; \lim_{t\to\infty} \Pr\{S_n > t\} \;=\; 1 - \lim_{t\to\infty} \Pr\{S_n \le t\}$$

Since the X_i are rv's, the sums S_n are also rv's (*i.e.*, nondefective) for each n (see Section 1.3.7), and thus $\lim_{t\to\infty} \Pr\{S_n \le t\} = 1$ for each n. Thus $\lim_{t\to\infty} \Pr\{N(t) < n\} = 0$ for each n. This shows that the set of sample points ω for which $\lim_{t\to\infty} N(t(\omega)) < n$ has probability 0 for all n. Thus the set of sample points for which $\lim_{t\to\infty} N(t,\omega)$ is finite has probability 0 and $\lim_{t\to\infty} N(t) = \infty$ WP1.

Next, $\mathsf{E}[N(t)]$ is nondecreasing in t, and thus has either a finite or infinite limit as $t \to \infty$. For each n, $\Pr\{N(t) \ge n\} \ge 1/2$ for large enough t, and therefore $\mathsf{E}[N(t)] \ge n/2$ for all such t. Thus $\mathsf{E}[N(t)]$ can have no finite limit as $t \to \infty$, and $\lim_{t\to\infty} \mathsf{E}[N(t)] = \infty$. □

The following theorem is quite a bit more general than the second guess above, but it will be useful elsewhere and helps show why the strong law of large numbers is so useful.

Theorem 3.2. *Let $\overline{X}$ be the mean for a sequence of IID rv's $\{X_i;\, i \ge 1\}$ and assume that $\mathsf{E}[|X|] < \infty$. Let $f(x)$ be continuous at $x = \overline{X}$ where f is a finite real valued function of a real variable. Then the relative frequencies S_n/n where $S_n = \sum_{i=1}^{n} X_i$ satisfy*

$$\lim_{n\to\infty} f(S_n/n) = f(\overline{X}) \qquad \text{WP1} \tag{3.3}$$

Proof of Theorem 3.2: First let $a_1, a_2, \dots,$ be a sequence of real numbers such that $\lim_{n\to\infty} a_n = \overline{X}$. From the definition of continuity, we know that for every $\epsilon > 0$, there is a $\delta > 0$ such that $|f(x) - f(\overline{X})| < \epsilon$ for all x such that $|x - \overline{X}| < \delta$. Also, since $\lim_{n\to\infty} a_n = \overline{X}$, we know that for every $\delta > 0$, there is an m such that $|a_n - \overline{X}| \le \delta$ for all $n \ge m$. Putting these two statements together, we know that for every $\epsilon > 0$, there is an m such that $|f(a_n) - f(\overline{X})| < \epsilon$ for all $n \ge m$. Thus $\lim_{n\to\infty} f(a_n) = f(\overline{X})$.

If ω is any sample point such that $\lim_{n\to\infty} S_n(\omega)/n = \overline{X}$, we then have $\lim_{n\to\infty} f(S_n(\omega)/n) = f(\overline{X})$. Since this set of sample points has probability 1, (3.3) follows. □

Proof of Theorem 3.1, Strong law for renewal processes: Since $\Pr\{X > 0\} = 1$ for a renewal process, we see that $\overline{X} > 0$. Choosing $f(x) = 1/x$, we see that $f(x)$ is continuous at $x = \overline{X}$. It follows from Theorem 3.2 that

$$\lim_{n\to\infty} \frac{n}{S_n} = \frac{1}{\overline{X}} \qquad \text{WP1}$$

From Lemma 3.1, we know that $\lim_{t\to\infty} N(t) = \infty$ with probability 1, so, with probability 1, $N(t)$ increases through all the nonnegative integers as t increases from 0 to ∞. Thus

$$\lim_{t\to\infty} \frac{N(t)}{S_{N(t)}} = \lim_{n\to\infty} \frac{n}{S_n} = \frac{1}{\overline{X}} \qquad \text{WP1}$$

Recall that $N(t)/t$ is sandwiched between $N(t)/S_{N(t)}$ and $N(t)/S_{N(t)+1}$, so we can complete the proof by showing that $\lim_{t\to\infty} N(t)/S_{N(t)+1} = 1/\overline{X}$. To show this,

$$\lim_{t\to\infty} \frac{N(t)}{S_{N(t)+1}} = \lim_{n\to\infty} \frac{n}{S_{n+1}} \;=\; \lim_{n\to\infty} \frac{n{+}1}{S_{n+1}}\,\frac{n}{n{+}1} \;=\; \frac{1}{\overline{X}} \qquad \text{WP1.}$$

□

We have gone through the proof of this theorem in great detail, since a number of the techniques are probably unfamiliar to many readers. If one reads the proof again, after becoming familiar with the details, the simplicity of the result will be quite striking. The theorem is also true if the mean inter-renewal interval is infinite; this can be seen by a truncation argument (see Exercise 3.5).

As explained in Section 1.5.5, Theorem 3.1 also implies the corresponding weak law of large numbers for $N(t)$, *i.e.*, for any $\epsilon > 0$, $\lim_{t\to\infty} \Pr\{|N(t)/t - 1/\overline{X}| \geq \epsilon\} = 0$). This weak law could also be derived from the weak law of large numbers for S_n (Theorem 1.4). We do not pursue that here, since the derivation is tedious and uninstructive. As we will see, it is the strong law that is most useful for renewal processes.

Figure 3.2 helps give some appreciation of what the strong law for $N(t)$ says and doesn't say. The strong law deals with time-averages, $\lim_{t\to\infty} N(t,\omega)/t$, for individual sample points ω; these are indicated in the figure as horizontal averages, one for each ω. It is also of interest to look at time and ensemble-averages, $\mathsf{E}[N(t)/t]$, shown in the figure as vertical averages. Note that $N(t,\omega)/t$ is the time-average number of renewals from 0 to t, whereas $\mathsf{E}[N(t)/t]$ averages also over the ensemble. Finally, to focus on arrivals in the vicinity of a particular time t, it is of interest to look at the ensemble-average $\mathsf{E}[N(t+\delta) - N(t)]/\delta$.

Given the strong law for $N(t)$, one would hypothesize that $\mathsf{E}[N(t)/t]$ approaches $1/\overline{X}$ as $t \to \infty$. One might also hypothesize that $\lim_{t\to\infty} \mathsf{E}[N(t+\delta) - N(t)]/\delta = 1/\overline{X}$, subject to some minor restrictions on δ. These hypotheses are correct and are discussed in detail in what follows. This equality of time-averages and limiting ensemble-averages for renewal processes carries over to a large number of stochastic processes, and forms the basis of *ergodic theory*. These results are important for both theoretical and practical purposes. It is sometimes easy to find time averages (just like it was easy to find the time-average $N(t,\omega)/t$ from the strong law of large numbers), and it is sometimes easy to find limiting ensemble-averages. Being able to equate the two then allows us to alternate at will between time and ensemble-averages.

Note that in order to equate time-averages and limiting ensemble-averages, quite a few conditions are required. First, the time-average must exist in the limit $t \to \infty$ with probability one and have a fixed value with probability one; second, the ensemble-average must approach a limit as $t \to \infty$; and third, the limits must be the same. The following example, for a stochastic process very different from a renewal process, shows that equality between time and ensemble averages is not always satisfied for arbitrary processes.

Example 3.2.1. Let $\{X_i; i \geq 1\}$ be a sequence of binary IID random variables, each taking the value 0 with probability 1/2 and 2 with probability 1/2. Let $\{M_n; n \geq 1\}$ be the product process in which $M_n = X_1X_2\cdots X_n$. Since $M_n = 2^n$ if X_1 to X_n each take the value 2 (an event of probability 2^{-n}) and $M_n = 0$ otherwise, we see that $\lim_{n\to\infty} M_n = 0$ with probability 1. Also $\mathsf{E}[M_n] = 1$ for all $n \geq 1$. Thus the time-average exists and equals 0 with probability 1 and the ensemble-average exists and equals 1 for all n, but the two are different. The problem is that as n increases, the atypical event in which $M_n = 2^n$ has a probability approaching 0, but still has a significant effect on the ensemble-average.

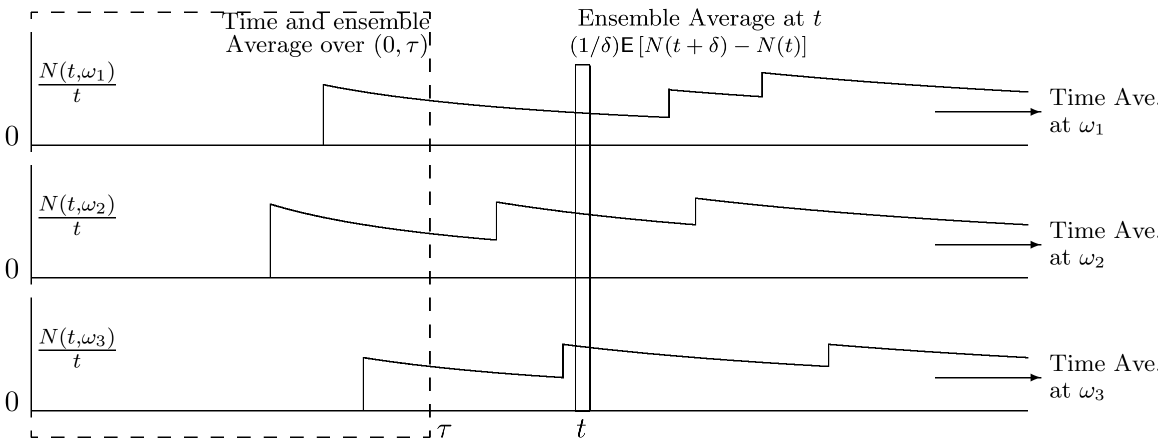


Figure 3.2: The time average at a sample point ω, the time and ensemble average from 0 to a given τ, and the ensemble-average in an interval $(t, t+\delta]$.

Before establishing the results about ensemble-averages, we state and briefly discuss the central limit theorem for renewal processes.

Theorem 3.3 (Central Limit Theorem (CLT) for $N(t)$). *Assume that the inter-renewal intervals for a renewal counting process $\{N(t);\, t > 0\}$ have finite standard deviation $\sigma > 0$. Then*

$$\lim_{t\to\infty} \Pr\left\{ \frac{N(t) - t/\overline{X}}{\sigma \overline{X}^{-3/2}\sqrt{t}} < \alpha \right\} = \Phi(\alpha). \tag{3.4}$$

where $\Phi(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx$.

This says that the distribution function of $N(t)$ tends to the Gaussian distribution with mean $t/\overline{X}$ and standard deviation $\sigma\overline{X}^{-3/2}\sqrt{t}$.

The theorem can be proved by applying Theorem 1.3 (the CLT for a sum of IID rv's) to S_n and then using the identity $\{S_n \le t\} = \{N(t) \ge n\}$. The general idea is illustrated in Figure 3.3, but the details are somewhat tedious, and can be found, for example, in [16]. We simply outline the argument here. For any real α, the CLT states that

$$\Pr\left\{S_n \le n\overline{X} + \alpha\sqrt{n}\sigma\right\} \approx \Phi(\alpha)$$

where $\Phi(\alpha) = \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)\, dx$ and where the approximation becomes exact in the limit $n \to \infty$. Letting

$$t = n\overline{X} + \alpha\sqrt{n}\sigma,$$

and using $\{S_n \le t\} = \{N(t) \ge n\}$,

$$\Pr\{N(t) \ge n\} \approx \Phi(\alpha). \tag{3.5}$$

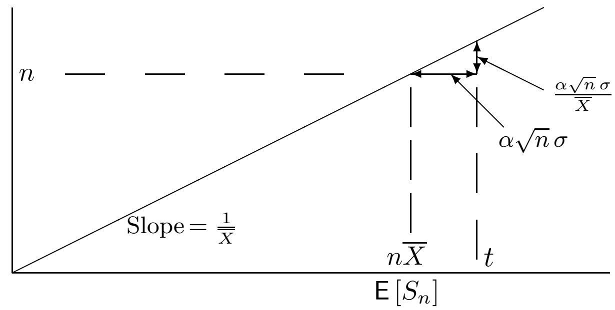


Figure 3.3: Illustration of the central limit theorem (CLT) for renewal processes. A given integer n is shown on the vertical axis, and the corresponding mean, $\mathsf{E}[S_n] = n\overline{X}$ is shown on the horizontal axis. The horizontal line with arrows at height n indicates α standard deviations from $\mathsf{E}[S_n]$, and the vertical line with arrows indicates the distance below $(t/\overline{X})$.

Since t is monotonic in n for fixed α, we can express n in terms of t, getting

$$n = \frac{t}{\overline{X}} - \frac{\alpha\sigma\sqrt{n}}{\overline{X}} \approx \frac{t}{\overline{X}} - \alpha\sigma t^{1/2}(\overline{X})^{-3/2}.$$

Substituting this into (3.5) establishes the theorem for $-\alpha$, which establishes the theorem since α is arbitrary. The omitted details involve handling the approximations carefully.

3.3 Renewal-reward processes; time-averages

There are many situations in which, along with a renewal counting process $\{N(t);\, t > 0\}$, there is another randomly varying function of time, called a *reward function* $\{R(t); t > 0\}$. $R(t)$ models a rate at which the process is accumulating a reward. We shall illustrate many examples of such processes and see that a "reward" could also be a cost or any randomly varying quantity of interest. The important restriction on these *reward functions* is that $R(t)$ at a given t depends only on the particular inter-renewal interval containing t. We start with several examples to illustrate the kinds of questions addressed by this type of process.

Example 3.3.1. (Time-average residual life) For a renewal counting process $\{N(t), t \geq 0\}$, let $Y(t)$ be the residual life at time t. The *residual life* is defined as the interval from t until the next renewal epoch, i.e., as $S_{N(t)+1} - t$. For example, if we arrive at a bus stop at time t and buses arrive according to a renewal process, $Y(t)$ is the time we have to wait for a bus to arrive (see Figure 3.4). We interpret $\{Y(t);\, t \geq 0\}$ as a reward function. The time-average of $Y(t)$, over the interval $(0, t]$, is given by[3] $(1/t)\int_0^t Y(\tau)d\tau$. We are interested in the limit of this average as $t \to \infty$ (assuming that it exists in some sense). Figure 3.4 illustrates

[3] $\int_0^t Y(\tau)d\tau$ is a rv just like any other function of a set of rv's. It has a sample value for each sample function of $\{N(t);\, t > 0\}$, and its distribution function could be calculated in a straightforward but tedious way. For arbitrary stochastic processes, integration and differentiation can require great mathematical sophistication, but none of those subtleties occur here.

a sample function of a renewal counting process $\{N(t);\, t > 0\}$ and shows the residual life $Y(t)$ for that sample function. Note that, for a given sample function $\{Y(t) = y(t)\}$, the integral $\int_0^t y(\tau)\, d\tau$ is simply a sum of isosceles right triangles, with part of a final triangle at the end. Thus it can be expressed as

$$\int_0^t y(\tau)d\tau = \frac{1}{2}\sum_{i=1}^{n(t)} x_i^2 + \int_{\tau=s_{n(t)}}^{t} y(\tau)d\tau$$

where $\{x_i;\, 0 < i < \infty\}$ is the set of sample values for the inter-renewal intervals.

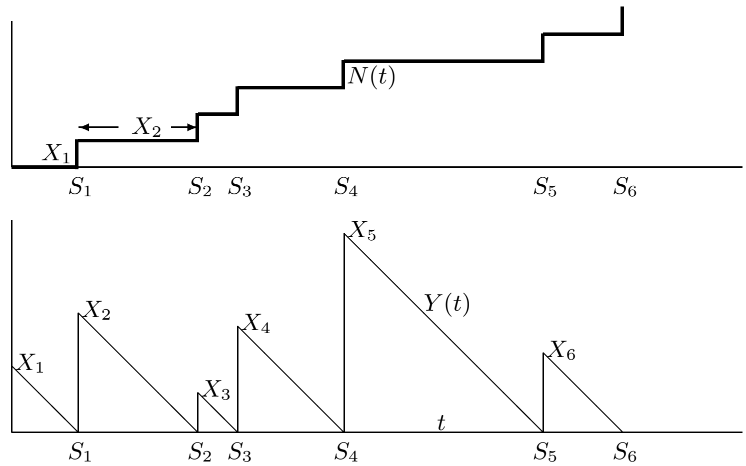


Figure 3.4: Residual life at time t. For any given sample function of the renewal process, the sample function of residual life decreases linearly with a slope of -1 from the beginning to the end of each inter-renewal interval.

Since this relationship holds for every sample point, we see that the random variable $\int_0^t Y(\tau)d\tau$ can be expressed in terms of the inter-renewal random variables X_n as

$$\int_{\tau=0}^{t} Y(\tau)d\tau = \frac{1}{2}\sum_{n=1}^{N(t)} X_n^2 + \int_{\tau=S_{N(t)}}^{t} Y(\tau)d\tau.$$

Although the final term above can be easily evaluated for a given $S_{N(t)}(t)$, it is more convenient to use the following bound:

$$\frac{1}{2t}\sum_{n=1}^{N(t)} X_n^2 \le \frac{1}{t}\int_{\tau=0}^{t} Y(\tau)d\tau \le \frac{1}{2t}\sum_{n=1}^{N(t)+1} X_n^2. \tag{3.6}$$

The term on the left can now be evaluated in the limit $t \to \infty$ (for all sample functions except a set of probability zero) as follows:

$$\lim_{t\to\infty} \frac{\sum_{n=1}^{N(t)} X_n^2}{2t} = \lim_{t\to\infty} \frac{\sum_{n=1}^{N(t)} X_n^2}{N(t)}\, \frac{N(t)}{2t}. \tag{3.7}$$

Consider each term on the right side of (3.7) separately. For the first term, recall that $\lim_{t\to 0} N(t) = \infty$ with probability 1. Thus as $t \to \infty$, $\sum_{n=1}^{N(t)} X_n^2/N(t)$ goes through the same set of values as $\sum_{n=1}^{k} X_n^2/k$ as $k \to \infty$. Thus, using the SLLN,

$$\lim_{t\to\infty} \frac{\sum_{n=1}^{N(t)} X_n^2}{N(t)} = \lim_{k\to\infty} \frac{\sum_{n=1}^{k} X_n^2}{k} = \mathsf{E}\left[X^2\right] \qquad \text{WP1}$$

The second term on the right side of (3.7) is simply $N(t)/2t$. By the strong law for renewal processes, $\lim_{t\to\infty} N(t)/2t = 1/(2\mathsf{E}[X])$ WP1. Thus both limits exist WP1 and

$$\lim_{t\to\infty} \frac{\sum_{n=1}^{N(t)} X_n^2}{2t} = \frac{\mathsf{E}\left[X^2\right]}{2\mathsf{E}[X]} \qquad \text{WP1} \tag{3.8}$$

The right hand term of (3.6) is handled almost the same way:

$$\lim_{t\to\infty} \frac{\sum_{n=1}^{N(t)+1} X_n^2}{2t} = \lim_{t\to\infty} \frac{\sum_{n=1}^{N(t)+1} X_n^2}{N(t)+1} \, \frac{N(t)+1}{N(t)} \, \frac{N(t)}{2t} = \frac{\mathsf{E}\left[X^2\right]}{2\mathsf{E}[X]}. \tag{3.9}$$

Combining these two results, we see that, with probability 1, the time-average residual life is given by

$$\lim_{t\to\infty} \frac{\int_{\tau=0}^{t} Y(\tau)\, d\tau}{t} = \frac{\mathsf{E}\left[X^2\right]}{2\mathsf{E}[X]}. \tag{3.10}$$

Note that this time-average depends on the second moment of X; this is $\overline{X}^2 + \sigma^2 \geq \overline{X}^2$, so the time-average residual life is at least half the expected inter-renewal interval (which is not surprising). On the other hand, the second moment of X can be arbitrarily large (even infinite) for any given value of $\mathsf{E}[X]$, so that the time-average residual life can be arbitrarily large relative to $\mathsf{E}[X]$. This can be explained intuitively by observing that large inter-renewal intervals are weighted more heavily in this time-average than small inter-renewal intervals.

Example 3.3.2. As an example of the effect of improbable but large inter-renewal intervals, let X take on the value ϵ with probability $1-\epsilon$ and value $1/\epsilon$ with probability ϵ. Then, for small ϵ, $\mathsf{E}[X] \sim 1, \mathsf{E}\left[X^2\right] \sim 1/\epsilon$, and the time average residual life is approximately $1/(2\epsilon)$ (see Figure 3.5).

Example 3.3.3. (time-average Age) Let $Z(t)$ be the age of a renewal process at time t where *age* is defined as the interval from the most recent arrival before (or at) t until t, i.e., $Z(t) = t - S_{N(t)}$ (see Figure 3.6). We notice that the age process, for a given sample function of the renewal process, is almost the same as the residual life process—the isosceles right triangles are simply turned around. Thus the same analysis as before can be used to show that the time average of $Z(t)$ is the same as the time-average of the residual life,

$$\lim_{t\to\infty} \frac{\int_{\tau=0}^{t} Z(\tau)\, d\tau}{t} = \frac{\mathsf{E}\left[X^2\right]}{2\mathsf{E}[X]} \qquad \text{WP1.} \tag{3.11}$$

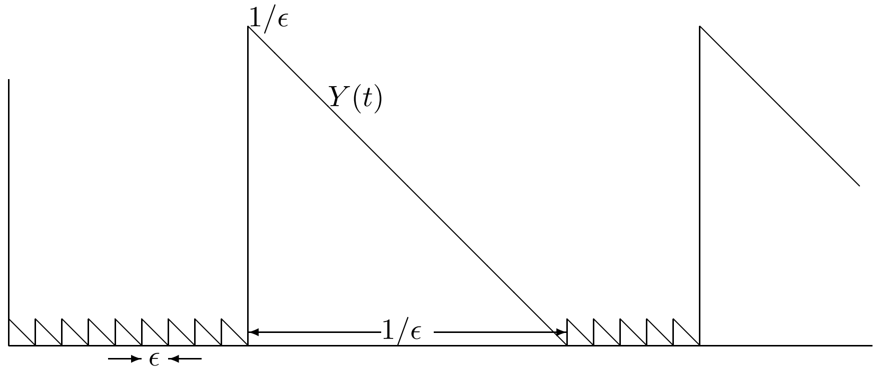


Figure 3.5: Average Residual life is dominated by large interarrival intervals. Each large interval has duration $1/\epsilon$, and the expected aggregate duration between successive large intervals is $1-\epsilon$

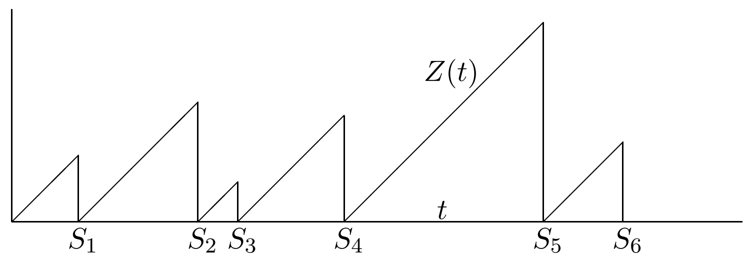


Figure 3.6: Age at time t: For any given sample function of the renewal process, the sample function of age increases linearly with a slope of 1 from the beginning to the end of each inter-renewal interval.

Example 3.3.4. (time-average Duration) Let $X(t)$ be the duration of the inter-renewal interval containing time t, i.e., $X(t) = X_{N(t)+1} = S_{N(t)+1} - S_{N(t)}$ (see Figure 3.7). It is clear that $X(t) = Z(t) + Y(t)$, and thus the time-average of the duration is given by

$$\lim_{t\to\infty} \frac{\int_{\tau=0}^{t} X(\tau)\,d\tau}{t} = \frac{\mathsf{E}\left[X^2\right]}{\mathsf{E}\left[X\right]} \qquad \text{WP1.} \tag{3.12}$$

Again, long intervals are heavily weighted in this average, so that the time-average duration is at least as large as the mean inter-renewal interval and often much larger.

3.3.1 General renewal-reward processes

In each of these examples, and in many other situations, we have a random function of time (i.e., $Y(t)$, $Z(t)$, or $X(t)$) whose value at time t depends only on where t is in the current inter-renewal interval (i.e., on the age $Z(t)$ and the duration $X(t)$ of the current inter-renewal interval). We now investigate the general class of reward functions for which the reward at time t depends at most on the age and the duration at t, *i.e.*, the reward

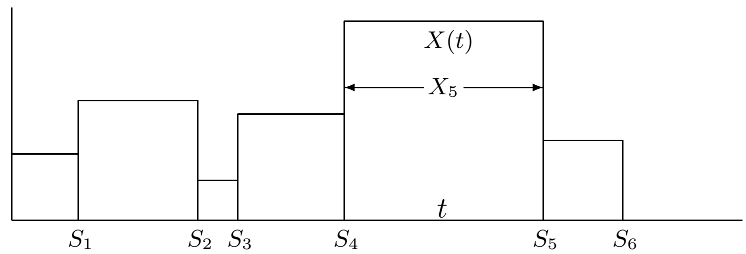


Figure 3.7: Duration $X(t) = X_{N(t)}$ of the inter-renewal interval containing t.

$R(t)$ at time t is given explicitly as a function[4] $\mathcal{R}(Z(t), X(t))$ of the age and duration at t. For the three examples above, the function $\mathcal{R}$ is trivial. That is, the residual life, $Y(t)$, is given by $X(t) - Z(t)$ and the age and duration are given directly.

We now find the time-average value of $R(t)$, namely, $\lim_{t\to\infty} \frac{1}{t}\int_0^t R(\tau)\,d\tau$. As in examples 3.3.1 to 3.3.4 above, we first want to look at the accumulated reward over each inter-renewal period separately. Define R_n as the accumulated reward in the nth renewal interval,

$$\mathrm{R}_n = \int_{S_{n-1}}^{S_n} R(\tau)\,d(\tau) = \int_{S_{n-1}}^{S_n} \mathcal{R}[Z(\tau), X(\tau)]\,d\tau. \tag{3.13}$$

For residual life (see Example 3.3.1), R_n is the area of the nth isosceles right triangle in Figure 3.4. In general, since $Z(\tau) = \tau - S_{n-1}$,

$$\mathrm{R}_n = \int_{S_{n-1}}^{S_n} \mathcal{R}(\tau - S_{n-1}, X_n)\,d\tau = \int_{z=0}^{X_n} \mathcal{R}(z, X_n)\,dz. \tag{3.14}$$

Note that R_n is a function only of X_n, where the form of the function is determined by $\mathcal{R}(Z, X)$. From this, it is clear that $\{\mathrm{R}_n; n \ge 1\}$ is essentially[5] a set of IID random variables. For residual life, $\mathcal{R}(z, X_n) = X_n - z$, so the integral in (3.14) is $X_n^2/2$, as calculated by inspection before. In general, from (3.14), the expected value of R_n is given by

$$\mathsf{E}[\mathrm{R}_n] = \int_{x=0}^{\infty}\int_{z=0}^{x} \mathcal{R}(z, x)\,dz\,d\mathsf{F}_X(x). \tag{3.15}$$

Breaking $\int_0^t R(\tau)\,d\tau$ into the reward over the successive renewal periods, we get

$$\begin{aligned}\int_0^t R(\tau)\,d\tau &= \int_0^{S_1} R(\tau)\,d\tau + \int_{S_1}^{S_2} R(\tau)\,d\tau + \cdots + \int_{S_{N(t)-1}}^{S_{N(t)}} R(\tau)\,d\tau + \int_{S_{N(t)}}^{t} R(\tau)d\tau \\ &= \sum_{n=1}^{N(t)} \mathrm{R}_n + \int_{S_{N(t)}}^{t} R(\tau)\,d\tau. \end{aligned} \tag{3.16}$$

[4]This means that $R(t)$ can be determined at any t from knowing $Z(t)$ and $X(t)$. It does not mean that $R(t)$ must vary as either of those quantities are changed. Thus, for example, $R(t)$ could depend on only one of the two or could even be a constant.

[5]One can certainly define functions $\mathcal{R}(Z, X)$ for which the integral in (3.14) is infinite or undefined for some values of X_n, and thus R_n becomes a defective rv. It seems better to handle this type of situation when it arises rather than handling it in general.

The following theorem now generalizes the results of Examples 3.3.1, 3.3.3, and 3.3.4 to general renewal-reward functions.

Theorem 3.4. *Let $\{R(t); t > 0\} \geq 0$ be a nonnegative renewal-reward function for a renewal process with expected inter-renewal time $\mathsf{E}[X] = \overline{X} < \infty$. If $\mathsf{E}[\mathrm{R}_n] < \infty$, then with probability 1*

$$\lim_{t\to\infty} \frac{1}{t} \int_{\tau=0}^{t} R(\tau)\, d\tau = \frac{\mathsf{E}[\mathrm{R}_n]}{\overline{X}}. \tag{3.17}$$

Proof: Using (3.16), the accumulated reward up to time t can be bounded between the accumulated reward up to the renewal before t and that to the next renewal after t,

$$\frac{\sum_{n=1}^{N(t)} \mathrm{R}_n}{t} \leq \frac{\int_{\tau=0}^{t} R(\tau)\, d\tau}{t} \leq \frac{\sum_{n=1}^{N(t)+1} \mathrm{R}_n}{t}. \tag{3.18}$$

The left hand side of (3.18) can now be broken into

$$\frac{\sum_{n=1}^{N(t)} \mathrm{R}_n}{t} = \frac{\sum_{n=1}^{N(t)} \mathrm{R}_n}{N(t)} \, \frac{N(t)}{t}. \tag{3.19}$$

Each R_n is a given function of X_n, so the R_n are IID. As $t \to \infty$, $N(t) \to \infty$, and, thus, as we have seen before, the strong law of large numbers can be used on the first term on the right side of (3.19), getting $\mathsf{E}[\mathrm{R}_n]$ with probability 1. Also the second term approaches $1/\overline{X}$ by the strong law for renewal processes. Since $0 < \overline{X} < \infty$ and $\mathsf{E}[\mathrm{R}_n]$ is finite, the product of the two terms approaches the limit $\mathsf{E}[\mathrm{R}_n]/\overline{X}$. The right-hand inequality of (3.18) is handled in almost the same way,

$$\frac{\sum_{n=1}^{N(t)+1} \mathrm{R}_n}{t} = \frac{\sum_{n=1}^{N(t)+1} \mathrm{R}_n}{N(t)+1} \, \frac{N(t)+1}{N(t)} \, \frac{N(t)}{t}. \tag{3.20}$$

It is seen that the terms on the right side of (3.20) approach limits as before and thus the term on the left approaches $\mathsf{E}[\mathrm{R}_n]/\overline{X}$ with probability 1. Since the upper and lower bound in (3.18) approach the same limit, $(1/t)\int_0^t R(\tau)\, d\tau$ approaches the same limit and the theorem is proved. □

The restriction to nonnegative renewal-reward functions in Theorem 3.4 is slightly artificial. The same result holds for non-positive reward functions simply by changing the directions of the inequalities in (3.18). Assuming that $\mathsf{E}[\mathrm{R}_n]$ exists (i.e., that both its positive and negative parts are finite), the same result applies in general by splitting an arbitrary reward function into a positive and negative part. This gives us the corollary:

Corollary 3.1. *Let $\{R(t); t > 0\}$ be a renewal-reward function for a renewal process with expected inter-renewal time $\mathsf{E}[X] = \overline{X} < \infty$. If $\mathsf{E}[\mathrm{R}_n]$ exists, then with probability 1*

$$\lim_{t\to\infty} \frac{1}{t} \int_{\tau=0}^{t} R(\tau)\, d\tau = \frac{\mathsf{E}[\mathrm{R}_n]}{\overline{X}}. \tag{3.21}$$

Example 3.3.5. (Distribution of Residual Life) Example 3.3.1 treated the time-average value of the residual life $Y(t)$. Suppose, however, that we would like to find the time-average distribution function of $Y(t)$, i.e., the fraction of time that $Y(t) \le y$ as a function of y. The approach, which applies to a wide variety of applications, is to use an indicator function (for a given value of y) as a reward function. That is, define $R(t)$ to have the value 1 for all t such that $Y(t) \le y$ and to have the value 0 otherwise. Figure 3.8 illustrates this function for a given sample path. Expressing this reward function in terms of $Z(t)$ and $X(t)$, we have

$$R(t) = \mathcal{R}(Z(t), X(t)) = \begin{cases} 1\,; & X(t) - Z(t) \le y \\ 0\,; & \text{otherwise} \end{cases}.$$

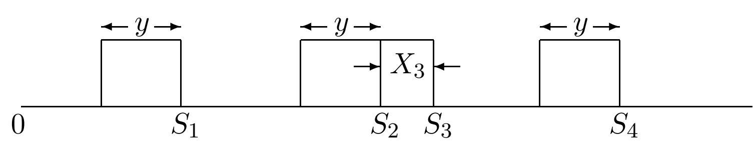


Figure 3.8: Reward function to find the time-average fraction of time that $\{Y(t) \le y\}$. For the sample function in the figure, $X_1 > y$, $X_2 > y$, and $X_4 > y$, but $X_3 < y$

Note that if an inter-renewal interval is smaller than y (such as the third interval in Figure 3.8), then $R(t)$ has the value one over the entire interval, whereas if the interval is greater than y, then $R(t)$ has the value one only over the final y units of the interval. Thus $\mathrm{R}_n = \min[y, X_n]$. Note that the random variable $\min[y, X_n]$ is equal to X_n for $X_n \le y$, and thus has the same distribution function as X_n in the range 0 to y. Figure 3.9 illustrates this in terms of the complementary distribution function. From the figure, we see that

$$\mathsf{E}[\mathrm{R}_n] = \mathsf{E}[\min(X, y)] = \int_{x=0}^{\infty} \Pr\{\min(X, y) > x\}\, dx = \int_{x=0}^{y} \Pr\{X > x\}\, dx. \tag{3.22}$$

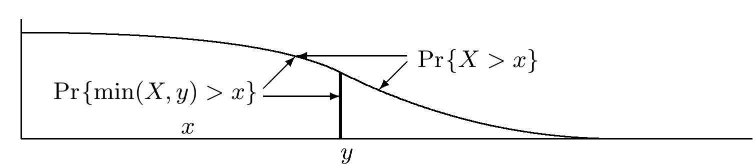


Figure 3.9: R_n for distribution of residual life.

Let $\mathsf{F}_Y(y) = \lim_{t\to\infty}(1/t)\int_0^t R(\tau)\, d\tau$ denote the time-average fraction of time that the residual life is less than or equal to y. From Theorem 3.4 and Eq.(3.22), we then have

$$\mathsf{F}_Y(y) = \frac{\mathsf{E}[\mathrm{R}_n]}{\overline{X}} = \frac{1}{\overline{X}} \int_{x=0}^{y} \Pr\{X > x\}\, dx \qquad \text{WP1.} \tag{3.23}$$

As a check, note that this integral is increasing in y and approaches 1 as $y \to \infty$. Note also that the expected value of Y, calculated from (3.23), is given by $\mathsf{E}\left[X^2\right]/2\overline{X}$, in agreement with (3.10).

The same argument can be applied to the time-average distribution of age (see Exercise 3.9). The time-average fraction of time, $\mathsf{F}_Z(z)$, that the age is at most z is given by

$$\mathsf{F}_Z(z) = \frac{1}{\overline{X}} \int_{x=0}^{z} \Pr\{X > x\}\, dx \qquad \text{WP1.} \tag{3.24}$$

In the development so far, the reward function $R(t)$ has been a function solely of the age and duration intervals, and the aggregate reward over the nth inter-renewal interval is a function only of X_n. In more general situations, where the renewal process is embedded in some more complex process, it is often desirable to define $R(t)$ to depend on other aspects of the process as well. The important thing here is for $\{\mathrm{R}_n;\ n \geq 1\}$ to be an IID sequence. How to achieve this, and how it is related to queueing systems, is described in Section 3.4.3. Theorem 3.4 clearly remains valid if $\{\mathrm{R}_n;\ n \geq 1\}$ is IID. This more general type of renewal-reward function will be required and further discussed in Section 3.7 where we discuss Little's theorem and the M/G/1 expected queueing delay, both of which use this more general structure.

Limiting time-averages are sometimes visualized by the following type of experiment. For some given large time t, let T be a uniformly distributed random variable over $(0, t]$; T is independent of the renewal-reward process under consideration. Then $(1/t)\int_0^t R(\tau)\, d\tau$ is the expected value (over T) of $R(T)$ for a given sample path of $\{R(\tau);\ \tau{>}0\}$. Theorem 3.4 states that in the limit $t \to \infty$, all sample paths (except a set of probability 0) yield the same expected value over T. This approach of viewing a time-average as a random choice of time is referred to as *random incidence*. Random incidence is awkward mathematically, since the random variable T changes with the overall time t and has no reasonable limit. It also blurs the distinction between time and ensemble-averages, so it will not be used in what follows.

3.4 Random stopping times

Visualize performing an experiment repeatedly, observing independent successive sample outputs of a given random variable (i.e., observing a sample outcome of $X_1, X_2, \ldots$ where the X_i are IID). The experiment is stopped when enough data has been accumulated for the purposes at hand.

This type of situation occurs frequently in applications. For example, we might be required to make a choice from several hypotheses, and might repeat an experiment until the hypotheses are sufficiently discriminated. If the number of trials is allowed to depend on the outcome, the mean number of trials required to achieve a given error probability is typically a small fraction of the number of trials required when the number is chosen in advance. Another example occurs in tree searches where a path is explored until further extensions of the path appear to be unprofitable.

The first careful study of experimental situations where the number of trials depends on the data was made by the statistician Abraham Wald and led to the field of sequential analysis (see [21]). We study these situations now since one of the major results, Wald's equality, will be useful in studying $\mathsf{E}[N(t)]$ in the next section. These stopping times are also important in understanding queues and will be used in Section 3.6. Wald's equality will be used again, along with a generating function equality known as Wald's identity, when we study random walks.

An important part of experiments that stop after a random number of trials is the rule for stopping. Such a rule must specify, for each sample path, the trial at which the experiment stops, *i.e.*, the final trial after which no more trials are performed. Thus the rule for stopping must specify a positive, integer valued, random variable J, called the *stopping time*, or *stopping trial*, mapping sample paths to this final trial at which the experiment stops.

We still view the sample space as including the set of sample value sequences for the never-ending sequence of random variables $X_1, X_2, \ldots$. That is, even if the experiment is stopped at the end of the second trial, we still visualize the 3rd, 4th, ... random variables as having sample values as part of the sample function. In other words, we visualize that the experiment continues forever, but that the observer stops watching at the end of the stopping point. From the standpoint of applications, the experiment might or might not continue after the observer stops watching. From a mathematical standpoint, however, it is far preferable to view the experiment as continuing so as to avoid confusion and ambiguity about what it means for the rv's $X_1, X_2, \ldots$ to be IID when the very existence of later variables depends on earlier sample values.

The intuitive notion of stopping a sequential experiment should involve stopping based on the data (*i.e.*, the sample values) gathered up to and including the stopping point. For example, if $X_1, X_2, \ldots$, represent the succesive changes in our fortune when gambling, we might want to stop when our cumulative gain exceeds some fixed value. The stopping time n then depends on the sample values of $X_1, X_2, \ldots, X_n$. At the same time, we want to exclude from stopping times those rules that allow the experimenter to peek at subsequent values before making the decision to stop or not.[6] This leads to the following definition.

Definition 3.1. *A stopping time*[7] *or stopping trial J for a sequence of rv's $X_1, X_2, \ldots,$ is a positive integer-valued rv such that for each $n \geq 1$, the indicator rv $\mathbb{I}_{\{J=n\}}$ is a function of $X_1, X_2, \ldots, X_n$.*

The last clause of the definition means that any given sample value $x_1, \ldots, x_n$ for $X_1, \ldots, X_n$ uniquely determines whether the corresponding sample value of J is n or not. Note that since the stopping time J is defined to be a positive integer-valued rv, the events $\{J = n\}$ and $\{J = m\}$ for $m < n$ are disjoint events, so stopping at trial m makes it impossible to also stop at n for a given sample path. Also the union of the events $\{J = n\}$ over $n \geq 1$ has probability 1. Aside from this final restriction, the definition does not depend on the probability measure and depends solely on the set of events $\{J = n\}$ for each n. In some situations, it is useful to relax the definition further to allow J to be a possibly-defective rv.

[6]For example, poker players do not take kindly to a player who attempts to withdraw his bet when someone else wins the hand. Similarly, a responsible statistician gathering data on product failures is forbidden to observe a failure and then neglect recording it by recording an earlier stopping time.

[7]Stopping times are sometimes called optional stopping times.

In this case the question of whether stopping occurs with probability 1 can be postponed until after specifying the disjoint events $\{J = n\}$.

Example 3.4.1. Consider a Bernoulli process $\{X_n; n \geq 1\}$. A very simple stopping time for this process is to stop at the first occurrence of the string $(0, 1)$. Figure 3.10 illustrates this stopping time by viewing it as a truncation of the tree of possible binary sequences.

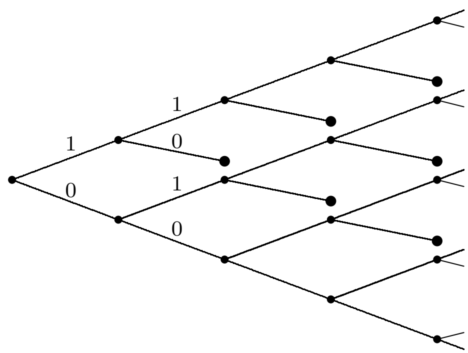


Figure 3.10: A tree representing the set of binary sequences, with a stopping rule viewed as a pruning of the tree. The particular stopping rule here is to stop on the first occurrence of the string $(1, 0)$. The leaves of the tree (*i.e.*, the nodes at which stopping occurs) are marked with large dots and the intermediate nodes (the other nodes) with small dots. Note that each leaf in the tree has a one-to-one correspondence with an initial segment of the tree, so the stopping nodes can be unambiguously viewed either as leaves of the tree or initial segments of the sample sequences.

As can be seen from the figure, the event $\{J = 2\}$, *i.e.*, the event that stopping occurs at time 2, is the event $\{X_1{=}1,\ X_2{=}0\}$. Similarly, the event $\{J = 3\}$ is $\{X_1{=}1,\ X_2{=}1,\ X_3{=}0\} \bigcup \{X_1{=}0,\ X_2{=}1,\ X_3{=}0\}$. The disjointness of $\{J = n\}$ and $\{J = m\}$ for $n \neq m$ is represented in the figure by terminating the tree at each stopping node. It can be seen that the tree never dies out completely, and in fact the number of stopping nodes at time n is $n - 1$. However, the probability that stopping has not occurred by time n goes to zero exponentially with n, which ensures that J is a random variable.

The same approach of representing a stopping time by a pruned tree can be used for any discrete random sequence, although the tree becomes quite unwieldy in all but trivial cases. Visualizing a stopping time in terms of a pruned tree is useful conceptually, but stopping rules are usually stated in other terms. For example, we often consider a stopping trial for the interarrival intervals of a renewal process as the first n for which the arrival epoch S_n satisfies $S_n \geq \alpha$ for some given $\alpha > 0$. The event $\{J = n\}$ is then the event $\{X_1 + \cdots + X_n \geq \alpha\}$ intersected with the events $\{X_1 + \cdots + X_j < \alpha\}$ for each $j < n$. This is a stopping rule whether or not the variables X_i are discrete.

3.4.1 Wald's equality

An important question that arises with stopping times is to evaluate the sum S_J of the random variables up to the stopping time, *i.e.*, $S_J = \sum_{n=1}^{J} X_n$. Many gambling strategies and investing strategies involve some sort of rule for when to stop, and it is important to understand the rv S_J (which can model the overall gain or loss up to that time). Wald's equality is very useful in helping to find $\mathsf{E}[S_J]$.

Theorem 3.5 (Wald's equality). *Let $\{X_n; n \geq 1\}$ be a sequence of IID rv's, each of mean $\overline{X}$. If J is a stopping time for $\{X_n; n \geq 1\}$ and if $\mathsf{E}[J] < \infty$, then the sum $S_J = X_1 + X_2 + \cdots + X_J$ at the stopping time J satisfies*

$$\mathsf{E}[S_J] = \overline{X}\mathsf{E}[J]. \tag{3.25}$$

Proof: Note that X_n is included in $S_J = \sum_{n=1}^{J} X_n$ whenever $n \leq J$, *i.e.*, whenever the indicator function $\mathbb{I}_{J \geq n} = 1$. Thus

$$S_J = \sum_{n=1}^{\infty} X_n \mathbb{I}_{J \geq n} . \tag{3.26}$$

This includes X_n as part of the sum if stopping has not occurred before time n. The event $\{J \geq n\}$ is the complement of $\{J < n\} = \{J = 1\} \bigcup \cdots \bigcup \{J = n-1\}$. All of these latter events are determined by $X_1, \ldots, X_{n-1}$ and are thus independent of X_n. It follows that X_n and $\{J < n\}$ are independent and thus X_n and $\{J \geq n\}$ are also independent.[8] Thus

$$\mathsf{E}[X_n \mathbb{I}_{J \geq n}] = \overline{X}\mathsf{E}[\mathbb{I}_{J \geq n}].$$

We then have

$$\begin{aligned}
\mathsf{E}[S_J] &= \mathsf{E}\left[\sum\nolimits_{n=1}^{\infty} X_n \mathbb{I}_{J \geq n}\right] \\
&= \sum\nolimits_{n=1}^{\infty} \mathsf{E}[X_n \mathbb{I}_{J \geq n}] && (3.27) \\
&= \sum\nolimits_{n=1}^{\infty} \overline{X}\mathsf{E}[\mathbb{I}_{J \geq n}] \\
&= \overline{X}\mathsf{E}[J]. && (3.28)
\end{aligned}$$

The interchange of expectation and infinite sum in (3.27) is obviously valid for a finite sum, and is shown in Exercise 3.15 to be valid for an infinite sum if $\mathsf{E}[J] < \infty$. The example below shows that Wald's equality can be invalid when $\mathsf{E}[J] = \infty$. The final step above comes from the observation that $\mathsf{E}[\mathbb{I}_{J \geq n}] = \Pr\{J \geq n\}$. Since J is a positive integer rv, $\mathsf{E}[J] = \sum_{n=1}^{\infty} \Pr\{J \geq n\}$. One can also obtain the last step by using $J = \sum_{n=1}^{\infty} \mathbb{I}_{J \geq n}$ (see Exercise 3.10). □

What this result essentially says in terms of gambling is that strategies for when to stop betting are not really effective as far as the mean is concerned. This sometimes appears obvious and sometimes appears very surprising, depending on the application.

[8] This can be quite confusing initially, since (as seen in the example of Figure 3.10) X_n is not necessarily independent of the event $\{J = n\}$, nor of $\{J = n + 1\}$, etc. In other words, *given that* stopping has not occurred before trial n, then X_n can have a great deal to do with *the trial* at which stopping occurs. However, as shown above, X_n has nothing to do with *whether* $\{J < n\}$ or $\{J \geq n\}$.

Example 3.4.2 (Stop when you're ahead in coin tossing). We can model a (biased) coin tossing game as a sequence of IID rv's $X_1, X_2, \dots$ where each X is 1 with probability p and -1 with probability $1-p$. Consider the possibly-defective stopping time J where J is the first n for which $S_n = X_1 + \cdots + X_n = 1$, *i.e.*, the first time the gambler is ahead.

We first want to see if J is a rv, *i.e.*, if the probability of eventual stopping, say $\theta = \Pr\{J < \infty\}$, is 1. We solve this by a frequently useful trick, but will use other more systematic approaches in Chapters 5 and 7 when we look at this same example as a birth-death Markov chain and then as a simple random walk. Note that $\Pr\{J = 1\} = p$, *i.e.*, $S_1 = 1$ with probability p and stopping occurs at trial 1. With probability $1-p$, $S_1 = -1$. Following $S_1 = -1$, the only way to become one ahead is to first return to $S_n = 0$ for some $n > 1$, and, after the first such return, go on to $S_m = 1$ at some later time m. The probability of eventually going from -1 to 0 is the same as that of going from 0 to 1, *i.e.*, θ. Also, given a first return to 0 from -1, the probability of reaching 1 from 0 is θ. Thus,

$$\theta = p + (1-p)\theta^2.$$

This is a quadratic equation in θ with two solutions, $\theta = 1$ and $\theta = p/(1-p)$. For $p > 1/2$, the second solution is impossible since θ is a probability. Thus we conclude that J is a rv. For $p = 1/2$ (and this is the most interesting case), both solutions are the same, $\theta = 1$, and again J is a rv. For $p < 1/2$, the correct solution[9] is $\theta = p/(1-p)$. Thus $\theta < 1$ so J is a defective rv.

For the cases where $p \geq 1/2$, *i.e.*, where J is a rv, we can use the same trick to evaluate $\mathsf{E}[J]$,

$$\mathsf{E}[J] = p + (1-p)(1 + 2\mathsf{E}[J]).$$

The solution to this is

$$\mathsf{E}[J] \;=\; \frac{1}{2(1-p)} \;=\; \frac{1}{2p-1}.$$

We see that $\mathsf{E}[J]$ is finite for $p > 1/2$ and infinite for $p = 1/2$.

For $p > 1/2$, we can check that these results agree with Wald's equality. In particular, since S_J is 1 with probability 1, we also have $\mathsf{E}[S_J] = 1$. Since $\overline{X} = 2p-1$ and $\mathsf{E}[J] = 1/(2p-1)$, Wald's equality is satisfied (which of course it has to be).

For $p = 1/2$, we still have $S_J = 1$ with probability 1 and thus $\mathsf{E}[S_J] = 1$. However $\overline{X} = 0$ so $\overline{X}\mathsf{E}[J]$ has no meaning and Wald's equality breaks down. Thus we see that the restriction $\mathsf{E}[J] < \infty$ in Wald's equality is indeed needed. This is an important example, but we will understand it better when we study birth-death Markov chains and random walks.

3.4.2 Using Wald's equality to find $m(t) = \mathsf{E}[N(t)]$

We next use Wald's equality in evaluating $m(t) = \mathsf{E}[N(t)]$ for a renewal counting process. Consider an experiment in which successive interarrival intervals are observed until the sum

[9] This will be shown when we view this example as a birth-death Markov chain in Chapter 5.

first exceeds t. From Figure 3.11, note that $S_{N(t)+1}$ is the epoch of the first arrival after t, and thus $N(t)+1$ is the number of intervals observed until the sum first exceeds t. We now show that $N(t)+1$ is a stopping trial[10] for the interarrival sequence $\{X_n; n \geq 1\}$. Informally, the decision to stop when the sum exceeds t depends only on the interarrival intervals already observed. More formally, for any given t, the indicator rv $\mathbb{I}_{N(t)+1=n}$ for each given n has the value 1 for the event

$$\{N(t) = n-1\} = \{S_{n-1} \leq t\} \bigcap \{S_n > t\}.$$

This is a function only of $X_1, \dots, X_n$, verifying that $N(t)+1$ is a stopping trial for $X_1, X_2 \dots$.

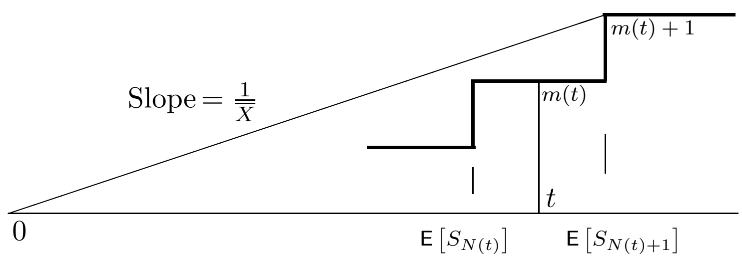


Figure 3.11: Illustration of Wald's equality, 3.29, applied to $N(t)+1$.

Note that $N(t)$ is not a stopping trial for $X_1, X_2, \dots$. For any given n, observation of $X_1, \dots, X_n$ does not specify whether or not $N(t) = n$ (except in the very special case where $S_n = t$). One would have to peek ahead at X_{n+1} to verify that S_{n+1} exceeds t.

Since $N(t)+1$ is a stopping trial, and $\mathsf{E}[N(t)+1]$ exists (see Exercise 3.2), Wald's equality yields

$$\begin{aligned} \mathsf{E}\left[S_{N(t)+1}\right] &= \overline{X}\mathsf{E}[N(t)+1] = \overline{X}[m(t)+1]. \\ m(t) &= \frac{\mathsf{E}\left[S_{N(t)+1}\right]}{\overline{X}} - 1. \end{aligned} \tag{3.29}$$

Since $\mathsf{E}\left[S_{N(t)+1}\right] \geq t$, we have $m(t) \geq t/\overline{X} - 1$, and

$$\frac{m(t)}{t} \geq \frac{1}{\overline{X}} - \frac{1}{t}. \tag{3.30}$$

Section 3.5 combines this with an upper bound on $m(t)/t$ to establish the elementary renewal theorem.

3.4.3 Using stopping times to understand embedded renewals

The above definition of stopping time is quite restrictive in that it refers only to a single sequence of rv's . In many queueing situations, for example, there is a sequence of interarrival

[10]It is important to recognize here that stopping times or trials are defined relative to a particular sequence of rv's , in this case $X_1, X_2, \dots$. The conceptual observer does not observe $N(\tau)$ for $\tau \leq t$, but rather observes only the inter-renewal intervals. Thus stopping does not occur at time t, but rather on the trial $N(t)+1$.

times, say $X_1, X_2, \dots$, for incoming customers, and another sequence of required service times, say $Z_1, Z_2, \dots$, for each of those customers. Consider again the G/G/m queue described in Example 3.1.1. This of course includes G/G/1, M/G/1, G/D/1, etc. as special cases. Assume that the customers are served in First-Come-First-Served (FCFS) order.[11] The sample sequence of interarrival intervals and service times then uniquely specifies when each customer enters and leaves the system. The first trial J at which an arriving customer sees an empty queue (*i.e.*, a renewal as described in Example 3.1.1) is a function of the interarrival and required service times up to and including trial J. This can be viewed as a stopping time with the following small generalization of the stopping time definition.

Definition 3.2 (Generalized stopping times). *A generalized stopping time J for a sequence of pairs of rv's $(X_1, Z_1), (X_2, Z_2) \dots$, is a positive integer-valued rv such that, for each $n \geq 1$, the indicator rv $\mathbb{I}_{\{J=n\}}$ is a function of $X_1, Z_1, X_2, Z_2, \dots, X_n, Z_n$.*

Wald's equality can be trivially generalized for these generalized stopping times.

Theorem 3.6 (Generalized Wald's equality). *Let $\{(X_n, Z_n);\ n \geq 1\}$ be a sequence of pairs of rv's, where each pair is independent and identically distributed (IID) to all other pairs. Assume that each X_i has finite mean $\overline{X}$. If J is a stopping time for $\{(X_n, Z_n);\ n \geq 1\}$ and if $\mathsf{E}[J] < \infty$, then the sum $S_J = X_1 + X_2 + \cdots + X_J$ satisfies*

$$\mathsf{E}[S_J] = \overline{X}\mathsf{E}[J]. \tag{3.31}$$

The proof of this will be omitted, since it is exactly the same as the proof of Theorem 3.5. In fact, the definition of stopping times could be further generalized by replacing the rv's Z_i by vector rv's or by a random number of rv's , and Wald's equality would still hold.[12]

For the example of the G/G/m queue, we can take the X_i as the customer interarrival times, Z_i as the service times, and J as the number of the first arrival to find an empty queue. In this case, $\mathsf{E}[S_J]$ is the expected time until the first arrival to see an empty queue, and $\mathsf{E}[J]$ is the expected number of trials until an empty queue is seen.

It is important here to avoid the confusion created by viewing J as a stopping *time*. The important thing about the sequence $(X_1, Z_1), (X_2, Z_2), \dots$ is the implied *ordering*, not any notion of time, which has an entirely different interpretation in most queueing situations. There is a further possible timing confusion about whether the custormer service times are determined when they arrive or when they complete service. This makes no difference since the ordered sequence of pairs is well-defined and satisfies the appropriate IID condition for using the Wald equality. For all examples like this, it is preferable to refer to stopping trials rather than stopping times.

As is often the case with Wald's equality, it is not obvious how to compute either quantity in (3.31), but it is nice to know that they are so simply related. It is also interesting to see that the service times and the interarrival times need not be independent for (3.31) to hold. This lack of independence implies a more complex type of queue than G/G/m, however.

[11]For single server queues, this is sometimes referred to as First-In-First-Out (FIFO) service.

[12]In fact, some people use an even more general definition of stopping time, stipulating that $\mathbb{I}_{J \geq n}$ is independent of $X_n, X_{n+1}, \dots$ for each n. This makes it easy to prove Wald's equality, but quite hard to see when the definition holds, especially since $\mathbb{I}_{J=n}$, for example, is typically dependent on X_n (see footnote 7).

Perhaps a more important aspect of viewing the first renewal for the G/G/m queue as a stopping time is the ability to show that successive renewals are in fact IID. Let $X_{2,1}, X_{2,2}, \dots$ be the interarrival times following the first arrival to see an empty queue. Conditioning on $J = j$, we have $X_{2,1} = X_{j+1}, X_{2,2} = X_{j+2}, \dots$. Thus $\{X_{2,k};\ k \geq 1\}$ is an IID sequence with the original inter-arrival distribution. Similarly $\{(X_{2,k}, Z_{2,k});\ k \geq 1\}$ is a sequence of IID pairs with the original distribution. This remains the same no matter what sample value j is assumed for the condition $J = j$. Thus $\{(X_{2,k}, Z_{2,k});\ k \geq 1\}$ is statistically independent of J and $(X_i, Z_i);\ 1 \leq i \leq J$.

The argument above can be repeated for subsequent arrivals to an empty system, so we have shown that successive arrivals to an empty system actually form a renewal process.[13] One can define many different stopping times for queues, such as the first time that n or more customers are in the queue. Most of these do not correspond to renewals. The implicit thing in the above argument that made arrivals to an empty system serve as renewals is that the state of the queue, upon a renewal, must be the same as the initial state.

Finally, nothing in the argument above for the G/G/m queue made any use of the FCFS service discipline. One can use any service discipline that is defined solely by the sample sequence of arrivals and departures, and the same results follow, although it is possible for the specific renewal epochs to change.

3.5 Expected number of renewals

The purpose of this section is to evaluate $\mathsf{E}[N(t)]$, denoted $m(t)$, as a function of $t > 0$ for arbitrary renewal processes. We first find an exact expression, in the form of an integral equation, for $m(t)$. This can be easily solved by Laplace transform methods in special cases. For the general case, however, $m(t)$ becomes increasingly messy for large t, so we then find the asymptotic behavior of $m(t)$. Since $N(t)/t$ approaches $1/\overline{X}$ with probability 1, we might expect $m(t)$ to grow with a derivative $m'(t)$ that asymptotically approaches $1/\overline{X}$. This is not true in general. Two somewhat weaker results, however, are true. The first, called the elementary renewal theorem (Theorem 3.7), states that $\lim_{t\to\infty} m(t)/t = 1/\overline{X}$. The second result, called Blackwell's theorem (Theorem 3.8), states that, subject to some limitations on $\delta > 0$, $\lim_{t\to\infty}[m(t+\delta) - m(t)] = \delta/\overline{X}$. This says essentially that the expected renewal rate approaches steady state as $t \to \infty$. We will find a large number of applications of Blackwell's theorem throughout the remainder of the text.

The exact calculation of $m(t)$ makes use of the fact that the expectation of a nonnegative random variable is defined as the integral of its complementary distribution function,

$$m(t) = \mathsf{E}[N(t)] = \sum_{n=1}^{\infty} \Pr\{N(t) \geq n\}.$$

[13] Confession by author: For about 15 years, I mistakenly believed that it was obvious that arrivals to an empty system in a G/G/m queue formed a renewal process. Thus I can not expect readers to be excited about the above proof. However, it is a nice example of how to use stopping times to see otherwise obscure points clearly.

Since the event $\{N(t) \geq n\}$ is the same as $\{S_n \leq t\}$, $m(t)$ is expressed in terms of the distribution functions of S_n, $n \geq 1$, as follows.

$$m(t) = \sum_{n=1}^{\infty} \Pr\{S_n \leq t\}. \tag{3.32}$$

Although this expression looks fairly simple, it becomes increasingly complex with increasing t. As t increases, there is an increasing set of values of n for which $\Pr\{S_n \leq t\}$ is significant, and $\Pr\{S_n \leq t\}$ itself is not that easy to calculate if the interarrival distribution $\mathsf{F}_X(x)$ is complicated. The main utility of (3.32) comes from the fact that it leads to an integral equation for $m(t)$. Since $S_n = S_{n-1} + X_n$ for each $n \geq 1$ (interpreting S_0 as 0), and since X_n and S_{n-1} are independent, we can use the convolution equation (1.11) to get

$$\Pr\{S_n \leq t\} = \int_{x=0}^{t} \Pr\{S_{n-1} \leq t-x\}\ d\mathsf{F}_X(x) \quad \text{for } n \geq 2.$$

Substituting this in (3.32) for $n \geq 2$ and using the fact that $\Pr\{S_1 \leq t\} = \mathsf{F}_X(t)$, we can interchange the order of integration and summation to get

$$\begin{aligned} m(t) &= \mathsf{F}_X(t) + \int_{x=0}^{t} \sum_{n=2}^{\infty} \Pr\{S_{n-1} \leq t-x\}\ d\mathsf{F}_X(x) \\ &= \mathsf{F}_X(t) + \int_{x=0}^{t} \sum_{n=1}^{\infty} \Pr\{S_n \leq t-x\}\ d\mathsf{F}_X(x) \\ &= \mathsf{F}_X(t) + \int_{x=0}^{t} m(t-x) d\mathsf{F}_X(x)\,; \qquad t \geq 0. \end{aligned} \tag{3.33}$$

An alternative derivation is given in Exercise 3.18. This integral equation is called the *renewal equation*, although that somewhat overstates its importance.

3.5.1 Laplace transform approach

If we assume that X has a density $\mathsf{f}_X(x)$, and that this density has a Laplace transform $L_X(s) = \int_0^{\infty} \mathsf{f}_X(x) e^{-sx} dx$, then we can take the Laplace transform of both sides of (3.33). Note that the final term in (3.33) is the convolution of m with f_X, so that the Laplace transform of $m(t)$ satisfies

$$L_m(s) = \frac{L_X(s)}{s} + L_m(s) L_X(s).$$

Solving for $L_m(s)$,

$$L_m(s) = \frac{L_X(s)}{s[1 - L_X(s)]}. \tag{3.34}$$

Example 3.5.1. As a simple example of how this can be used to calculate $m(t)$, suppose $\mathsf{f}_X(x) = (1/2)e^{-x} + e^{-2x}$ for $x \geq 0$. The Laplace transform is given by

$$L_X(s) = \frac{1}{2(s+1)} + \frac{1}{s+2} = \frac{(3/2)s + 2}{(s+1)(s+2)}.$$

Substituting this into (3.34) yields

$$L_m(s) = \frac{(3/2)s + 2}{s^2(s + 3/2)} = \frac{4}{3s^2} + \frac{1}{9s} - \frac{1}{9(s + 3/2)}.$$

Taking the inverse Laplace transform, we then have

$$m(t) = \frac{4t}{3} + \frac{1 - \exp[-(3/2)t]}{9}.$$

The procedure in this example can be used for any inter-renewal density $\mathsf{f}_X(x)$ for which the Laplace transform is a rational function, i.e., a ratio of polynomials. In such cases, $L_m(s)$ will also be a rational function. The Heaviside inversion formula (i.e., factoring the denominator and expressing $L_m(s)$ as a sum of individual poles as done above) can then be used to calculate $m(t)$. In the example above, there was a second order pole at $s = 0$ leading to the linear term $4t/3$ in $m(t)$, there was a first order pole at $s = 0$ leading to the constant $1/9$, and there was a pole at $s = -3/2$ leading to the exponentially decaying term.

We now show that a second order pole at $s = 0$ always occurs when $L_X(s)$ is a rational function. To see this, note that $L_X(0)$ is just the integral of $\mathsf{f}_X(x)$, which is 1; thus $1 - L_X(s)$ has a zero at $s = 0$ and $L_m(s)$ has a second order pole at $s = 0$. To evaluate the residue for this second order pole, we recall that the first and second derivatives of $L_X(s)$ at $s = 0$ are $-\mathsf{E}[X]$ and $\mathsf{E}\left[X^2\right]$ respectively. Expanding $L_X(s)$ in a power series around $s = 0$ then yields $L_X(s) = 1 - s\mathsf{E}[X] + (s^2/2)\mathsf{E}\left[X^2\right]$ plus terms of order s^3 or higher. This gives us

$$L_m(s) = \frac{1 - s\overline{X} + (s^2/2)\mathsf{E}\left[X^2\right] + \cdots}{s^2\left[\overline{X} - (s/2)\mathsf{E}[X^2] + \cdots\right]} = \frac{1}{s^2\overline{X}} + \frac{1}{s}\left(\frac{\mathsf{E}\left[X^2\right]}{2\overline{X}^2} - 1\right) + \cdots. \tag{3.35}$$

The remaining terms are the other poles of $L_m(s)$ with their residues. For values of s with $\Re(s) \geq 0$, we have $|L_X(s)| = |\int \mathsf{f}_X(x)e^{-sx}dx| \leq \int \mathsf{f}_X(x)|e^{-sx}|dx \leq \int \mathsf{f}_X(x)dx = 1$ with strict inequality except for $s = 0$. Thus $L_X(s)$ cannot have any poles on the imaginary axis or the right half plane, and $1 - L_X(s)$ cannot have any zeros there other than the one at $s = 0$. It follows that all the remaining poles of $L_m(s)$ are strictly in the left half plane. This means that the inverse transforms for all these remaining poles die out as $t \to \infty$. Thus the inverse Laplace transform of $L_m(s)$ is

$$m(t) = \frac{t}{\overline{X}} + \frac{\mathsf{E}\left[X^2\right]}{2\overline{X}^2} - 1 + \epsilon(t) \quad \text{for } t \geq 0. \tag{3.36}$$

where $\lim_{t\to\infty} \epsilon(t) = 0$.

We have derived (3.36) only for the special case in which $\mathsf{f}_X(x)$ has a rational Laplace transform. For this case, (3.36) implies both the elementary renewal theorem ($\lim_{t\to\infty} m(t)/t = 1/\overline{X}$) and also Blackwell's theorem ($\lim_{t\to\infty}[m(t+\delta) - m(t)] = \delta/\overline{X}$). We will interpret the meaning of the constant term $\mathsf{E}\left[X^2\right]/(2\overline{X}^2) - 1$ in Section 3.8.

3.5.2 The elementary renewal theorem

Theorem 3.7 (The elementary renewal theorem). *Let $\{N(t);\, t > 0\}$ be a renewal counting process with mean inter-renewal interval $\overline{X}$. Then $\lim_{t\to\infty} \mathsf{E}[N(t)]/t = 1/\overline{X}$.*

Discussion: The following proof is not difficult, but is also not particularly insightful, so readers might want to postpone it. We have already seen that $m(t) = \mathsf{E}[N(t)]$ is finite for all $t > 0$ (see Exercise 3.2. We will prove the theorem by establishing lower and upper bounds to $M(t)/t$ and showing that each approach $1/\mathsf{E}[X]$ as $t \to \infty$. The key element for both bounds is (3.29), which comes from the Wald equality and is repeated below.

$$m(t) = \frac{\mathsf{E}\left[S_{N(t)+1}\right]}{\overline{X}} - 1. \tag{3.37}$$

Proof: The lower bound to $m(t)/t$ comes by recognizing that $S_{N(t)+1}$ is the epoch of the first arrival after t. Thus $\mathsf{E}\left[S_{N(t)+1}\right] > t$. Substituting this into (3.37),

$$\frac{m(t)}{t} > \frac{1}{\mathsf{E}[X]} - \frac{1}{t}$$

Clearly this lower bound approaches $1/\mathsf{E}[X]$ as $t \to \infty$. The upper bound is more difficult,[14] and is established by first truncating $X(t)$ and then applying (3.37) to the truncated process.

For an arbitrary constant $b > 0$, let $\breve{X}_i = \min(b, X_i)$. Since these truncated random variables are IID, they form a related renewal counting process $\{\breve{N}(t); t > 0\}$ with $\breve{m}(t) = \mathsf{E}\left[\breve{N}(t)\right]$ and $\breve{S}_n = \breve{X}_1 + \cdots + \breve{X}_n$. Since $\breve{X}_i \le X_i$ for all i, we see that $\breve{S}_n \le S_n$ for all n. Since $\{S_n \le t\} = \{N(t) \ge n\}$, it follows that $\breve{N}(t) \ge N(t)$ and thus $\breve{m}(t) \ge m(t)$. Finally, in the truncated process, $\breve{S}_{\breve{N}(t)+1} \le t + b$ and thus $\mathsf{E}\left[\breve{S}_{\breve{N}(t)+1}\right] \le t + b$. Thus, applying (3.37) to the truncated process,

$$\frac{m(t)}{t} \le \frac{\breve{m}(t)}{t} = \frac{\mathsf{E}\left[S_{\breve{N}(t)+1}\right]}{t\mathsf{E}\left[\breve{X}\right]} - \frac{1}{t} \;\le\; \frac{t+b}{t\mathsf{E}\left[\breve{X}\right]}$$

Finally, choose $b = \sqrt{t}$. Then

$$\frac{m(t)}{t} \le \frac{1}{\mathsf{E}\left[\breve{X}\right]} + \frac{1}{\sqrt{t}\,\mathsf{E}\left[\breve{X}\right]}$$

Note finally that $\mathsf{E}\left[\breve{X}\right] = \int_0^b [1 - \mathsf{F}_X(x)]\, dx$. Since $b = \sqrt{t}$, we have $\lim_{t\to\infty} \mathsf{E}\left[\breve{X}\right] = \mathsf{E}[X]$, completing the proof. □

Note that this theorem (and its proof) have not assumed finite variance. It can also be seen that the theorem holds when $\mathsf{E}[X]$ is infinite, since $\lim_{t\to\infty} \mathsf{E}\left[\breve{X}\right] = \infty$ in this case.

[14] The difficulty here, and the reason for using a truncation argument, comes from the fact that the residual life, $S_{N(t)+1} - t$ at t might be arbitrarily large. We saw in Section 3.3 that the time-average residual life is infinite if $\mathsf{E}\left[X^2\right]$ is infinite. Figure 3.5 also illustrates why residual life can be so large.

Recall that $N[t, \omega]/t$ is the average number of renewals from 0 to t for a sample function ω, and $m(t)/t$ is the average of this over ω. Combining with Theorem 3.1, we see that the limiting time and ensemble-average equals the time-average renewal rate for each sample function except for a set of probability 0.

Another interesting question is to determine the expected renewal rate in the limit of large t without averaging from 0 to t. That is, are there some values of t at which renewals are more likely than others for large t? If the inter-renewal intervals have an integer distribution function (i.e., each inter-renewal interval must last for an integer number of time units), then each renewal epoch S_n must also be an integer. This means that $N(t)$ can increase only at integer times and the expected rate of renewals is zero at all non-integer times.

An obvious generalization of integer valued inter-renewal intervals is that of inter-renewals that occur only at integer multiples of some real number $d > 0$. Such a distribution is called an *arithmetic distribution*. The *span* of an arithmetic distribution is the largest number d such that this property holds. Thus, for example if X takes on only the values 0, 2, and 6, its distribution is arithmetic with span $d = 2$. Similarly, if X takes on only the values 1/3 and 1/5, then the span is $d = 1/15$. The remarkable thing, for our purposes, is that any inter-renewal distribution that is not an arithmetic distribution leads to a uniform expected rate of renewals in the limit of large t. This result is contained in Blackwell's renewal theorem, which we state without proof.[15]. Recall, however, that for the special case of an inter-renewal density with a rational Laplace transform, Blackwell's renewal theorem is a simple consequence of (3.36).

Theorem 3.8 (Blackwell). *If a renewal process has an inter-renewal distribution that is non-arithmetic, then for each $\delta > 0$,*

$$\lim_{t\to\infty}[m(t+\delta) - m(t)] = \frac{\delta}{\mathsf{E}[X]}. \tag{3.38}$$

If the inter-renewal distribution is arithmetic with span d, then for any integer $n \geq 1$

$$\lim_{t\to\infty}[m(t+nd) - m(t)] = \frac{nd}{\mathsf{E}[X]}. \tag{3.39}$$

Eq. (3.38) says that for non-arithmetic distributions, the expected number of arrivals in the interval $(t, t+\delta]$ is equal to $\delta/\mathsf{E}[X]$ in the limit $t \to \infty$. Since the theorem is true for arbitrarily small δ, the theorem almost seems to be saying that $m(t)$ has a derivative for large t, but this is not true. One can see the reason by looking at an example where X can take on only the values 1 and π. Then no matter how large t is, $N(t)$ can only increase at discrete points of time of the form $k + j\pi$ where k and j are nonnegative integers. Thus $dm(t)/dt$ is either 0 or ∞ for all t. As t gets larger, the jumps in $m(t)$ become both smaller in magnitude and more closely spaced from one to the next. Thus $[m(t+\delta) - m(t)]/\delta$ can approach $1/\mathsf{E}[X]$ as $t \to \infty$ for any fixed δ (as the theorem says), but as δ gets smaller, the convergence in t gets slower. For the above example (and for all discrete non-arithmetic distributions), $[m(t+\delta) - m(t)]/\delta$ does not approach[16] $1/\mathsf{E}[X]$ for any t as $\delta \to 0$.

[15] See Theorem 1 of Section 11.1, of [8]) for a proof

[16] This must seem like mathematical nitpicking to many readers. However, $m(t)$, the expected number of renewals in $(0, t]$, and how $m(t)$ varies with t, is central to this chapter and keeps reappearing.

For an arithmetic renewal process with span d, the asymptotic behavior of $m(t)$ as $t \to \infty$ is much simpler. Renewals can only occur at multiples of d, and since simultaneous renewals are not allowed, either 0 or 1 renewal occurs at each time kd. Thus $m(kd) - m((k{-}1)d)$ is the probability of a renewal at time kd, and (3.39) says that

$$\lim_{k\to\infty} \Pr\{\text{Renewal at } kd\} = \lim_{k\to\infty} m(kd) - m(kd{-}d) = \frac{d}{\overline{X}} \tag{3.40}$$

Section 3.6 contains a more extensive discussion about the limiting behavior of $m(t)$.

3.6 Renewal-reward processes; ensemble-averages

Section 3.3 showed that the time-average of a reward function $R(t)$ is a constant with probability 1. In this section, we explore the ensemble average, $\mathsf{E}[R(t)]$, as a function of time t. It is easy to see that $\mathsf{E}[R(t)]$ typically changes with t, especially for small t, but the question of interest here is whether $\mathsf{E}[R(t)]$ approaches a constant as $t \to \infty$.

In more concrete terms, if the arrival times of busses at a particular place forms a renewal process, then the waiting time for the next bus, *i.e.*, the residual life, starting at time t, can be represented as a reward function $R(t)$. We would like to know if the expected waiting time depends critically on t, where t is the time since the renewal process starts, *i.e.*, the time since bus number 0 arrived. if $\mathsf{E}[R(t)]$ varies significantly with t, even for large t, it means, for example, that our expected waiting time for a bus depends strongly on our arrival time. In other words, it says that the effect of the original arrival at $t = 0$ never dies out as $t \to \infty$.

Blackwell's renewal theorem (and also common sense) tell us that there is a large difference between arithmetic inter-renewal times and non-arithmetic inter-renewal times. For the arithmetic case, all renewals occur at multiples of the span, so the expected waiting time, for example, continues forever to decrease at rate 1 from each multiple of the span time to the next, and it increases discontinuously at each multiple of the span time. For the non-arithmetic case, on the other hand, the expected number of renewals in any small interval of length δ becomes independent of t as $t \to \infty$, so we might guess that $\mathsf{E}[R(t)]$ approaches a limit as $t \to \infty$.

The bottom line for this section is that under very broad conditions, the guess above is correct. This means that for non-arithmetic inter-renewal distibutions, $\lim_{t\to\infty} \mathsf{E}[R(t)]$ exists, and is equal to the time-average of $\lim_{t\to\infty}(1/t)\int_0^t R(\tau)d\tau$ with probability 1. Thus the limiting ensemble average can be computed simply by finding the time-average. The proof of this fact is interesting, the techniques used are interesting, and the intuitive interpretions are somewhat different from those used for the time-average. However, the proof of these results is based on a theorem called the key renewal theorem which we do not prove.

We start by deriving the distribution function of the age for a renewal process at any given time t. The time-average of this distribution function served as an example of the time-average of a particular reward function in Section 3.3. Here it plays a much more fundamental role, since finding $\mathsf{E}[R(t)]$ for an arbitrary reward function depends on knowing the distribution of how long the current inter-renewal interval has been running at time t, *i.e.*, on knowing the distribution function of the age at time t.

3.6.1 The distribution function of age at a given time

Let $\{N(t); t > 0\}$ be a renewal counting process with inter-renewal intervals $X_1, X_2, \dots$. The age and duration rv's are denoted by $Z(t)$ and $X(t)$, $t > 0$. Our first objective is to find (and understand) the distribution function of $Z(t)$ as a function of t. Initially we assume that each X_i is non-arithmetic and show that $Z(t)$ then converges in distribution to a limiting rv Z.

Recall that $Z(t) = t - S_{N(t)}$ is the nonnegative interval from the last renewal before t, (*i.e.*, $S_{N(t)}$) to t itself. It is possible that $S_{N(t)} = t$. Thus for $Z(t)$ to be equal to a given z, $0 \le z < t$ there must be one and only one $n \ge 1$ such that $S_n = t - z$ and $S_{n+1} > t$. We conclude from this that for $0 < z \le t$,

$$\begin{aligned}\Pr\{Z(t) \le z\} &= \sum_{n=1}^{\infty} \Pr\{t{-}z \le S_n \le t,\ S_{n+1} > t\} \\ &= \sum_{n=1}^{\infty} \Pr\{t{-}z \le S_n \le t,\ X_{n+1} > t - S_n\} \qquad (3.41)\end{aligned}$$

Figure 3.12 illustrates the region in the S_n, X_{n+1} plane over which this probability is taken. Note that X_{n+1} and S_n are independent and that the distribution of X_{n+1} does not depend on n. For given S_n, $\Pr\{X_{n+1} > t - S_n \mid S_n = s\} = \mathsf{F}^{\mathsf{c}}_X(t-s)$. Thus, as illustrated in the figure, we can evaluate $\Pr\{t{-}z \le S_n \le t,\ X_{n+1} > t - S_n\}$ as a Riemann-Stieltjes integral over S_n, *i.e.*, essentially as the limit below

$$\Pr\{t{-}z \le S_n \le t,\ X_{n+1} > t{-}S_n\} = \lim_{\delta\to 0} \sum_{k=\lfloor (t-z)/\delta\rfloor}^{\lfloor t/\delta\rfloor} \mathsf{F}^{\mathsf{c}}_X(t-k\delta)\,[\mathsf{F}_{S_n}(k\delta) - F_{S_n}(k\delta{-}\delta)] \qquad (3.42)$$

Recall from (3.32) that $m(t) = \sum_{n=1}^{\infty} \mathsf{F}_{S_n}(t)$. Thus if we substitute (3.42) into (3.41) and interchange the order of summation over n with the limit and sum over k, we get

$$\begin{aligned}\Pr\{Z(t) \le z\} &= \lim_{\delta\to 0} \sum_{k=\lfloor (t-z)/\delta\rfloor}^{\lfloor t/\delta\rfloor} \mathsf{F}^{\mathsf{c}}_X(t-k\delta)\,[m(k\delta) - m(k\delta{-}\delta)] \qquad (3.43) \\ &= \int_{t-z}^{t} \mathsf{F}^{\mathsf{c}}_X(t-\tau)\,dm(\tau) \qquad (3.44)\end{aligned}$$

This integral exists if $\mathsf{F}^{\mathsf{c}}_X$ is Riemann integrable.[17]

We next want to investigate whether this approaches a limit as $t \to \infty$. This certainly looks plausible from (3.43), since Blackwell's renewal theorem says that $m(k\delta) - m(k\delta - \delta) =$

[17] Readers who wish to verify this, and to better understand Riemann-Stiltjes integration, should read Rudin's very clear and elementary treatment in Chapter 6 of [18].

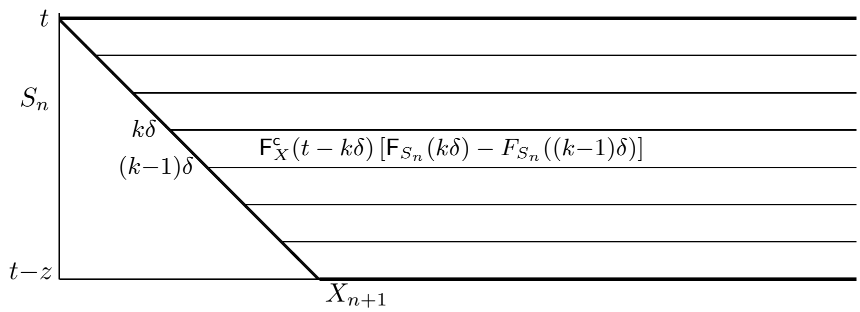


Figure 3.12: Illustration of the region (bounded by thick lines) where $t - z \le S_n \le t$ and $X_{n+1} > t - S_n$. The probability of this region can be found as a Riemann-Stieltjes integral $\int_0^{t-z} \mathsf{F}^c_X(t-s))d\mathsf{F}_{S_n}(s)$. This is interpreted in the figure as the sum of the probabilities of incrementally small horizontal slices of the region.

$\delta/\overline{X} + o(\delta)$. Since we haven't proven Blackwell's theorem anyway, we might as well state (without proof) another equivalent theorem, called the key renewal theorem, which makes taking this type of limit easy. Essentially Blackwell's theorem is easier to interpret, but the key renewal theorem is easier to use.

Theorem 3.9 (Key renewal theorem). *Let $r(z) \ge 0$ be a directly Riemann integrable function, and let $m(t) = \mathsf{E}[N(t)]$ for a non-arithmetic renewal process. Then*

$$\lim_{t\to\infty} \int_{\tau=0}^{t} r(t-\tau)dm(\tau) = \frac{1}{\overline{X}} \int_{x=0}^{\infty} r(x)dx. \tag{3.45}$$

We first explain what directly Riemann integrable means. If $r(z)$ is nonzero only over finite limits, say $[0, b]$, then direct Riemann integration means the same thing as ordinary Riemann integration (as taught in elementary calculus courses). However, if $r(z)$ is nonzero over $[0, \infty)$, then ordinary Riemann integration means to integrate from 0 to b and then take the limit as $b \to \infty$. Direct Riemann integration means to take a Riemann sum over the entire half line, $[0, \infty)$ and then take the limit as the grid becomes finer. Exercise 3.21 gives an example of a simple but bizarre function that is Riemann integrable but not directly Riemann integrable. If $r(z) \ge 0$ can be upper bounded by a decreasing Riemann integrable function, however, then $r(z)$ must be directly Riemann integrable. The bottom line is that restricting $r(z)$ to be directly Riemann integrable is not a major restriction.

Next we interpret the theorem. Note that $m(t) \approx t/\overline{X}$. As t becomes large, $m(t)$ can fluctuate very fast, but $|m(t) - t/\overline{X}|$ fluctuates between decreasing limits. In a sense, $r(t)$ acts like a smoothing filter which, as $t \to \infty$, eliminates the small but rapid fluctuations in $m(t)$. The theorem says that the required smoothing occurs whenever $r(t)$ is directly Riemann integrable.

We can now apply the key renewal theorem to (3.44) by letting $r(x) = \mathsf{F}^c_X(x)$ for $0 \le x \le z$ and $r(x) = 0$ elsewhere.

$$\lim_{t\to\infty} \Pr\{Z(t) \le z\} = \frac{1}{\overline{X}} \int_0^z \mathsf{F}^c_X(\tau)d\tau \tag{3.46}$$

Note that this agrees with the time-average result in (3.24). Also, since $r(x)$ in this case is positive and decreasing toward 0, it must be directly Riemann integrable.

3.6.2 Expected reward at a given time

We next use the distribution function of age at asymptotically large t to derive the expected reward at asymptotically large t for an arbitrary reward function $R(t) = \mathcal{R}(Z(t), X(t))$.

We continue to assume non-arithmetic inter-renewal intervals in this section. The previous subsection showed how to derive the distribution of age in a mathematically rigorous way. It will be clearer at this point if we are less rigorous. We will take the distribution function of $Z(t)$ to be equal to its limiting value, denoted here as F_{Z_t}. Note that this is not a function of t, but the subscript t is used as a reminder that we are talking about the asymptotic distribution of $Z(t)$ (rather than, for example the time-average distribution of $Z(t)$). We then use this limiting distribution F_{Z_t} to derive any desired reward function in the limit $R_t = \lim_{t\to\infty} R(t)$. We will also assume that the limiting age distribution has a probability density, which from (3.46) is given by

$$\mathsf{f}_{Z_t}(z) = \frac{1}{\overline{X}}\mathsf{F}^{\mathsf{c}}_X(z)$$

The exercises guide the interested reader in doing this rigorously. Conditional on $Z(t) = z$, the duration, $X(t)$ is constrained by $X(t) \geq z$, but otherwise has the distribution of an inter-renewal interval X. In other words, assuming that X has a density f_X, the conditional density of $X(t)$ is given by

$$\mathsf{f}_{X(t)|Z(t)}(x \mid z) = \frac{\mathsf{f}_X(x)}{\mathsf{F}^{\mathsf{c}}_X(z)} \qquad \text{for } x > z; \qquad 0 \text{ elsewhere}$$

Thus the asymptotic joint density, denoted by $\mathsf{f}_{Z_t,X_t}(z,x)$, is given by

$$\mathsf{f}_{Z_t,X_t}(z,x) = \frac{\mathsf{f}_X(x)}{\overline{X}}, \quad x > z\,; \qquad \mathsf{f}_{Z_t,X_t}(z,x) = 0 \text{ elsewhere.} \tag{3.47}$$

This joint density is illustrated in Figure 3.13. Note that the argument z does not appear

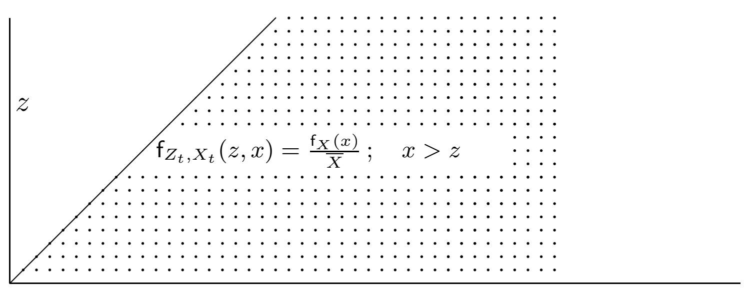


Figure 3.13: Joint density of age and duration.

except in the condition $x > z \geq 0$, but this condition is very important.

This joint density is intuitively satisfying, because it says that conditional on $X(t) = x$, the conditional density of $Z(t)$ is uniform from 0 to x, which says that for a given size of interval containing t, t is equally likely to be anywhere in that interval. The marginal density $X(t)$ in the limit $t \to \infty$ can be found by integrating (3.47) over the constraint region, getting

$$\mathsf{f}_{X_t}(x) = \int_{z=0}^{x} \frac{\mathsf{f}_X(x)\,dz}{\overline{X}} = \frac{x\mathsf{f}_X(x)}{\overline{X}}. \tag{3.48}$$

The factor x serves to weight the interarrival density by the size of the interval, which is intuitively plausible with a random incidence viewpoint. Unfortunately, intuition about these densities is somewhat slippery and acquiring intuition from the time-averages is more satisfying. The limiting marginal density for $Z(t)$ can be found in the same way, getting

$$\mathsf{f}_{Z_t}(z) = \int_{x=z}^{\infty} \frac{\mathsf{f}_X(x)\,dx}{\overline{X}} = \frac{\mathsf{F}^{\mathsf{c}}_X(z)}{\overline{X}}. \tag{3.49}$$

Note that this limiting density is the density for the time-average distribution function of age given in (3.24). The asymptotic mean duration and mean age can be calculated from (3.48) and (3.49) by integration by parts, yielding

$$\lim_{t\to\infty} \mathsf{E}[X(t)] = \frac{\mathsf{E}\left[X^2\right]}{\overline{X}}; \qquad \lim_{t\to\infty} \mathsf{E}[Z(t)] = \frac{\mathsf{E}\left[X^2\right]}{2\overline{X}};. \tag{3.50}$$

These ensemble-averages agree with the time-averages found in (3.11) and (3.12). In calculating time-averages, the somewhat paradoxical result that the time-average duration is greater than $\mathsf{E}[X]$ was explained by the large inter-renewal intervals being weighted more heavily in the time-average. Here the same effect occurs, but it can be given a different interpretation: the joint density, at z and x, for age and duration, is proportional to the inter-renewal density $\mathsf{f}_X(x)$, but the marginal density for duration is weighted by x since the range for age is proportional to x.

Using the joint probability density of $Z(t)$ and $X(t)$ to evaluate the expectation of an arbitrary reward function $R(t) = \mathcal{R}(Z(t), X(t))$ in the limit $t \to \infty$, we get

$$\lim_{t\to\infty} \mathsf{E}[R(t)] = \int_{x=0}^{\infty} \int_{z=0}^{x} \mathcal{R}(z,x)\,dz\, \frac{\mathsf{f}_X(x)\,dx}{\overline{X}} = \frac{\mathsf{E}[\mathrm{R}_n]}{\overline{X}}. \tag{3.51}$$

where $\mathsf{E}[\mathrm{R}_n]$ is the aggregate expected reward over an inter-renewal interval, as defined in (3.15). Thus the limiting ensemble-average is the same as the time-average. This result should not be surprising. Since we are dealing with non-arithmetic renewals, the probability of a renewal in a small interval becomes independent of where the interval is, so long as the interval is far enough removed from 0 for the process to be in "steady state". Since the reward function depends on the process only through the current renewal interval, the reward function must also become independent of time.

The above development assumed a non-arithmetic renewal process. For an arithmetic process, the situation is somewhat simpler mathematically, but in general $\mathsf{E}[R(t)]$ depends on the remainder when t is divided by the span d. Usually with such processes, one is interested only in reward functions that remain constant over intervals of length d, so we can consider $\mathsf{E}[R(t)]$ only for t equal to multiples of d, and thus we assume $t = nd$ here. Thus the function $\mathcal{R}(z,x)$ is of interest only when z and x are multiples of d, and in particular, only for $x = d, 2d, \dots$ and for $z = d, 2d, \dots, x$. We follow the convention that an inter-renewal interval is open on the left and closed on the right, thus including the renewal that ends the interval.

$$\mathsf{E}[R(nd)] = \sum_{i=1}^{\infty}\sum_{j=1}^{i} \mathcal{R}(jd, id)\,\Pr\{\text{renewal at } (n-j)d, \text{next renewal at } (n-j+i)d\}.$$

Let P_i be the probability that an inter-renewal interval has size id. Using (3.40) for the limiting probability of a renewal at $(n-j)d$, this becomes

$$\lim_{n\to\infty} \mathsf{E}[R(nd)] = \sum_{i=1}^{\infty}\sum_{j=1}^{i} \mathcal{R}(jd, id)\,\frac{d}{\overline{X}}\,P_i = \frac{\mathsf{E}[\mathrm{R}_n]}{\overline{X}}, \tag{3.52}$$

where $\mathsf{E}[\mathrm{R}_n]$ is the expected reward over a renewal period. In using this formula, remember that $R(t)$ is piecewise constant, so that the aggregate reward over an interval of size d around nd is $dR(nd)$.

It has been important, and theoretically assuring, to be able to find ensemble-averages for renewal-reward functions in the limit of large t and to show (not surprisingly) that they are the same as the time-average results. The ensemble-average results are quite tricky, though, and it is wise to check results achieved that way with the corresponding time-average results.

3.7 Applications of renewal-reward theory

3.7.1 Little's theorem

Little's theorem is an important queueing result stating that the expected number of customers in a queueing system is equal to the expected time each customer waits in the system times the arrival rate. This result is true under very general conditions; we use the G/G/1 queue with FCFS service as a specific example, but the reason for the greater generality will be clear as we proceed. Note that the theorem does not tell us how to find either the expected number or expected wait; it only says that if one can be found, the other can also be found.

Figure 3.14 illustrates the setting for Little's theorem. It is assumed that an arrival occurs at time 0, and that the subsequent interarrival intervals are IID. $A(t)$ is the number of arrivals from time 0 to t, including the arrival at 0, so $\{A(t) - 1;\ t \geq 0\}$ is a renewal counting process. The departure process $\{D(t); t \geq 0\}$ is the number of departures from 0 to t, and thus increases by one each time a customer leaves the system. The difference, $L(t) = A(t) - D(t)$, is the number in the system at time t.

Assuming FCFS, the system time of customer n, i.e., the time customer n spends in the system, is the interval from the nth arrival to the nth departure. Finally, the figure shows the renewal points $S_1, S_2, \dots$ at which arriving customers find an empty system. As explained in Section 3.4.3, the system probabilistically restarts at each of these renewal instants. The

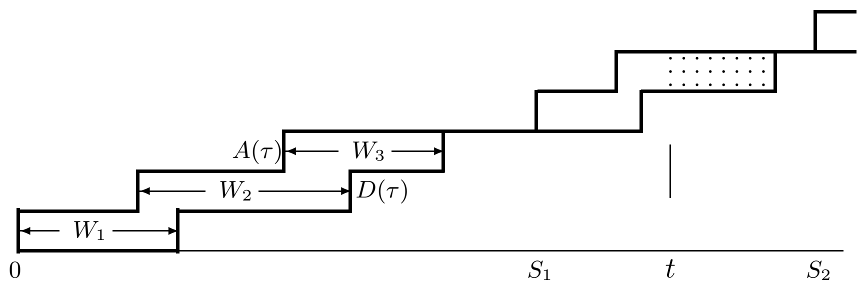


Figure 3.14: Arrival process, departure process, and waiting times for a queue. Renewals occur at S_1 and S_2, i.e., when an arrival sees an empty system. The area between $A(\tau)$ and $D(\tau)$ up to time t is $\int_0^t L(\tau)\,d\tau$ where $L(\tau) = A(\tau) - D(\tau)$. The sum $W_1 + \cdots + W_{A(t)}$ also includes the shaded area to the right of t.

rv's such as $A(t), D(t), W_i$ etc. within an inter-renewal interval are functions of the customer interarrival times and service times for the customers entering during that interarrival interval. Thus the behavior of the system in one inter-renewal interval is independent of that in each other inter-renewal interval.

It is important here to distinguish between two different renewal processes. The arrival process, or more precisely, $\{A(t) - 1; t \geq 0\}$ is one renewal counting process, and the renewal epochs $S_1, S_2, \ldots$ in the figure generate another renewal process. In what follows, $\{A(t);\ t \geq 0\}$ is referred to as the *arrival process* and $\{N(t);\ t > 0\}$, with renewal epochs $S_1, S_2, \ldots$ is referred to as the *renewal process*. The entire system can be viewed as starting anew at each renewal epoch, but not at each arrival epoch.

We now regard $L(t)$, the number of customers in the system at time t, as a reward function over the renewal process. This is slightly more general than the reward functions of Sections 3.3 and 3.6, since $L(t)$ depends on the interarrival intervals and service times within the corresponding busy period (i.e., inter-renewal interval). However, as explained in Section 3.4.3, both the number of these arrivals and the set of interarrival intervals and service times within an inter-renewal interval is independent of those in any other renewal interval.

Conditional on the age $Z(t)$ and duration $X(t)$ of the inter-renewal interval at time t, one could, in principle, calculate the expected value $\mathcal{R}(Z(t), X(t))$ over the parameters other than $Z(t)$ and $X(t)$. Fortunately, this is not necessary and we can use the sample functions of the combined arrival and departure processes directly, which specify $L(t)$ as $A(t) - D(t)$. Assuming that the expected inter-renewal interval is finite, Theorem 3.4 asserts that the time-average number of customers in the system (with probability 1) is equal to $\mathsf{E}[\mathrm{L}_n]/\mathsf{E}[X]$, where $\mathsf{E}[X]$ is the expected inter-renewal interval and $\mathsf{E}[\mathrm{L}_n]$ is the expected area between $A(t)$ and $D(t)$ (i.e., the expected integral of $L(t)$) over an inter-renewal interval. Note that an inter-renewal interval is a busy period followed by an idle period, so $\mathsf{E}[\mathrm{L}_n]$ is also the expected area over a busy period.

From Figure 3.14, we observe that $W_1 + W_2 + W_3$ is the area of the region between $A(t)$ and $D(t)$ in the first inter-renewal interval for the particular sample path in the figure. This is the aggregate reward over the first inter-renewal interval for the reward function $L(t)$. More generally, for any time t, $W_1 + W_2 + \cdots + W_{A(t)}$ is the area between $A(t)$ and $D(t)$

up to a height of $A(t)$. It is equal to $\int_0^t L(\tau)d\tau$ plus the remaining waiting time of each of the customers in the system at time t (see Figure 3.14). Since this remaining waiting time is at most the area between $A(t)$ and $D(t)$ from t until the next time when the system is empty, we have

$$\sum_{n=1}^{N(t)} \mathrm{L}_n \le \int_{\tau=0}^{t} L(\tau)\, d\tau \le \sum_{i=1}^{A(t)} W_i \le \sum_{n=1}^{N(t)+1} \mathrm{L}_n. \tag{3.53}$$

Assuming that the expected inter-renewal interval, $\mathsf{E}[X]$, is finite, we can divide both sides of (3.53) by t and go to the limit $t \to \infty$. From the same argument as in Theorem 3.4, we get

$$\lim_{t\to\infty} \frac{\sum_{i=1}^{A(t)} W_i}{t} = \lim_{t\to\infty} \frac{\int_{\tau=0}^{t} L(\tau)\, d\tau}{t} = \frac{\mathsf{E}[\mathrm{L}_n]}{\mathsf{E}[X]} \qquad \text{with probability 1.} \tag{3.54}$$

We denote $\lim_{t\to\infty}(1/t)\int_0^t L(\tau)d\tau$ as $\overline{L}$. The quantity on the left of (3.54) can now be broken up as waiting time per customer multiplied by number of customers per unit time, i.e.,

$$\lim_{t\to\infty} \frac{\sum_{i=1}^{A(t)} W_i}{t} = \lim_{t\to\infty} \frac{\sum_{i=1}^{A(t)} W_i}{A(t)} \lim_{t\to\infty} \frac{A(t)}{t}. \tag{3.55}$$

From (3.54), the limit on the left side of (3.55) exists (and equals $\overline{L}$) with probability 1. The second limit on the right also exists with probability 1 by the strong law for renewal processes, applied to $\{A(t) - 1; t \ge 0\}$. This limit is called the *arrival rate* λ, and is equal to the reciprocal of the mean interarrival interval for $\{A(t)\}$. Since these two limits exist with probability 1, the first limit on the right, which is the sample-path-average waiting time per customer, denoted $\overline{W}$, also exists with probability 1. We have thus proved Little's theorem,

Theorem 3.10 (Little). *For a FCFS G/G/1 queue in which the expected inter-renewal interval is finite, the time-average number of customers in the system is equal, with probability* 1*, to the sample-path-average waiting time per customer multiplied by the customer arrival rate, i.e.,* $\overline{L} = \lambda\overline{W}$.

The mathematics we have brought to bear here is quite formidable considering the simplicity of the idea. At any time t within an idle period, the sum of customer waiting periods up to time t is precisely equal to t times the time-average number of customers in the system up to t (see Figure 3.14). Renewal theory informs us that the limits exist and that the edge effects (i.e., the customers in the system at an arbitrary time t) do not have any effect in the limit.

Recall that we assumed earlier that customers departed from the queue in the same order in which they arrived. From Figure 3.15, however, it is clear that FCFS order is not required for the argument. Thus the theorem generalizes to systems with multiple servers and arbitrary service disciplines in which customers do not follow FCFS order. In fact, all that the argument requires is that the system has renewals (which are IID by definition of a renewal) and that the inter-renewal interval is finite with probability 1.

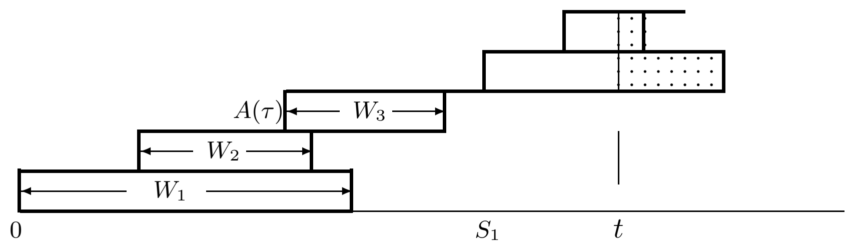


Figure 3.15: Arrivals and departures in non-FCFS systems. The aggregate reward (integral of number of customers in system) up to time t is the enclosed area to the left of t; the sum of waits of customers arriving by t includes the additional shaded area to the right of t.

Finally, suppose the inter-renewal distribution is non-arithmetic; this occurs if and only if the interarrival distribution is non-arithmetic. Then $\overline{L}$, the time-average number of customers in the system, is also equal to $\lim_{t\to\infty} \mathsf{E}[L(t)]$. It is also possible (see Exercise 3.33) to replace the sample-path-average waiting time $\overline{W}$ with $\lim_{n\to\infty} \mathsf{E}[W_n]$. This gives us the following variant of Little's theorem:

$$\lim_{t\to\infty} \mathsf{E}[L(t)] = \lambda\overline{W} = \lim_{n\to\infty} \lambda\mathsf{E}[W_n]\,. \tag{3.56}$$

The same argument as in Little's theorem can be used to relate the average number of customers in a single server queue (not counting service) to the average wait in the queue (not counting service). Renewals still occur on arrivals to an empty system, and the integral of customers in queue over a busy period is still equal to the sum of the queue waiting times. Let $L_q(t)$ be the number in the queue at time t and let $\overline{L}_q = \lim_{t\to\infty}(1/t)\int_0^t L_q(\tau)d\tau$ be the time-average queue wait. Letting $\overline{W}_q$ be the sample-path-average waiting time in queue,

$$\overline{L}_q = \lambda\overline{W}_q. \tag{3.57}$$

If the inter-renewal distribution is non-arithmetic, then

$$\lim_{t\to\infty} \mathsf{E}[L_q(t)] = \lambda\overline{W}_q. \tag{3.58}$$

The same argument can also be applied to the service facility of a single server queue. The time-average of the number of customers in the server is just the fraction of time that the server is busy. Denoting this fraction by ρ and the expected service time by $\overline{Z}$, we get

$$\rho = \lambda\overline{Z}. \tag{3.59}$$

3.7.2 Expected queueing time for an M/G/1 queue

For our last example of the use of renewal-reward processes, we consider the expected queueing time in an M/G/1 queue. We again assume that an arrival to an empty system occurs at time 0 and renewals occur on subsequent arrivals to an empty system. At any given time t, let $L_q(t)$ be the number of customers in the queue (not counting the customer in service, if any) and let $R(t)$ be the residual life of the customer in service. If no customer is in service, $R(t) = 0$, and otherwise $R(t)$ is the remaining time until the current service is completed. Let $U(t)$ be the waiting time in queue that would be experienced by a customer arriving at time t. This is often called the unfinished work in the queueing literature and represents the delay until all the customers currently in the system complete service. Thus the rv $U(t)$ is equal to $R(t)$, the residual life of the customer in service, plus the service times of each of the $L_q(t)$ customers currently waiting in the queue.

$$U(t) = \sum_{i=1}^{L_q(t)} Z_i + R(t). \tag{3.60}$$

where Z_i is the required service time of the ith customer in the queue at time t. Since the service times are independent of the arrival times and of the earlier service times, $L_q(t)$ is independent of $Z_1, Z_2, \dots, Z_{L_q(t)}$, so, taking expected values,

$$\mathsf{E}[U(t)] = \mathsf{E}[L_q(t)]\,\mathsf{E}[Z] + \mathsf{E}[R(t)]. \tag{3.61}$$

Figure 3.16 illustrates how to find the time-average of $R(t)$. Viewing $R(t)$ as a reward

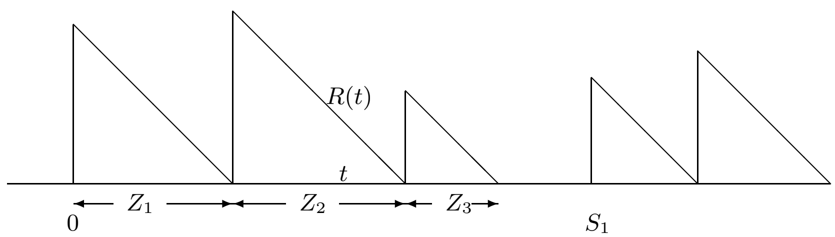


Figure 3.16: Sample value of the residual life function of customers in service.

function, we can find the accumulated reward up to time t as the sum of triangular areas. First, consider $\int R(\tau)d\tau$ from 0 to $S_{N(t)}$, i.e., the accumulated reward up to the last renewal epoch before t. $S_{N(t)}$ is not only a renewal epoch for the renewal process, but also an arrival epoch for the arrival process; in particular, it is the $A(S_{N(t)})$th arrival epoch, and the $A(S_{N(t)}) - 1$ earlier arrivals are the customers that have received service up to time $S_{N(t)}$. Thus,

$$\int_{\tau=0}^{S_{N(t)}} R(\tau)\,d\tau \;=\; \sum_{i=1}^{A(S_{N(t)})-1} \frac{Z_i^2}{2} \;\le\; \sum_{i=1}^{A(t)} \frac{Z_i^2}{2}.$$

We can similarly upper bound the term on the right above by $\int_{\tau=0}^{S_{N(t)+1}} R(\tau)\,d\tau$. We also know (from going through virtually the same argument many times) that $(1/t)\int_{\tau=0}^{t} R(\tau)d\tau$ will approach a limit with probability 1 as $t \to \infty$, and that the limit will be unchanged if t is replaced with $S_{N(t)}$ or $S_{N(t)+1}$. Thus, taking λ as the arrival rate,

$$\lim_{t\to\infty} \frac{\int_0^t R(\tau)\,d\tau}{t} \;=\; \lim_{t\to\infty} \frac{\sum_{i=1}^{A(t)} Z_i^2}{2A(t)}\,\frac{A(t)}{t} \;=\; \frac{\lambda \mathsf{E}\left[Z^2\right]}{2} \qquad \text{WP1.}$$

From (3.51), we can replace the time average above with the limiting ensemble-average, so that

$$\lim_{t\to\infty} \mathsf{E}[R(t)] = \frac{\lambda \mathsf{E}[Z^2]}{2}. \tag{3.62}$$

Finally, we can use Little's theorem, in the limiting ensemble-average form of (3.58), to assert that $\lim_{t\to\infty} \mathsf{E}[L_q(t)] = \lambda \overline{W}_q$. Substituting this plus (3.62) into (3.61), we get

$$\lim_{t\to\infty} \mathsf{E}[U(t)] = \lambda \mathsf{E}[Z]\, \overline{W}_q + \frac{\lambda \mathsf{E}[Z^2]}{2}. \tag{3.63}$$

This shows that $\lim_{t\to\infty} \mathsf{E}[U(t)]$ exists, so that $\mathsf{E}[U(t)]$ is asymptotically independent of t. It is now important to distinguish between $\mathsf{E}[U(t)]$ and $\overline{W}_q$. The first is the expected unfinished work at time t, which is the queueing delay that a customer would incur by arriving at t; the second is the sample-path-average expected queueing delay. For Poisson arrivals, the probability of an arrival in $(t, t+\delta]$ is independent of $U(t)$[18]. Thus, in the limit $t \to \infty$, each arrival faces an expected delay $\lim_{t\to\infty} \mathsf{E}[U(t)]$, so $\lim_{t\to\infty} \mathsf{E}[U(t)]$ must be equal to $\overline{W}_q$. Substituting this into (3.63), we obtain the celebrated *Pollaczek-Khinchin* formula,

$$\overline{W}_q = \frac{\lambda \mathsf{E}[Z^2]}{2(1 - \lambda \mathsf{E}[Z])}. \tag{3.64}$$

This queueing delay has some of the peculiar features of residual life, and in particular, if $\mathsf{E}[Z^2] = \infty$, the limiting expected queueing delay is infinite even though the expected service time is less than the expected interarrival interval.

In trying to visualize why the queueing delay is so large when $\mathsf{E}[Z^2]$ is large, note that while a particularly long service is taking place, numerous arrivals are coming into the system, and all are being delayed by this single long service. In other words, the number of new customers held up by a long service is proportional to the length of the service, and the amount each of them are held up is also proportional to the length of the service. This visualization is rather crude, but does serve to explain the second moment of Z in (3.64). This phenomenon is sometimes called the "slow truck effect" because of the pile up of cars behind a slow truck on a single lane road.

For a G/G/1 queue, (3.63) is still valid, but arrival times are no longer independent of $U(t)$, so that typically $\mathsf{E}[U(t)] \neq \overline{W}_q$. As an example, suppose that the service time is uniformly distributed between $1-\epsilon$ and $1+\epsilon$ and that the interarrival interval is uniformly distributed between $2-\epsilon$ and $2+\epsilon$. Assuming that $\epsilon < 1/2$, the system has no queueing and $\overline{W}_q = 0$. On the other hand, for small ϵ, $\lim_{t\to\infty} \mathsf{E}[U(t)] \sim 1/4$ (i.e., the server is busy half the time with unfinished work ranging from 0 to 1).

[18] This is often called the *PASTA* property, standing for Poisson arrivals see time-averages. This holds with great generality, requiring only that time-averages exist and that the state of the system at a given time t is independent of future arrivals.

3.8 Delayed renewal processes

We have seen a certain awkwardness in our discussion of Little's theorem and the M/G/1 delay result because an arrival was assumed, but not counted, at time 0; this was necessary for the first interarrival interval to be statistically identical to the others. In this section, we correct that defect by allowing the epoch at which the first renewal occurs to be arbitrarily distributed. The resulting type of process is a generalization of the class of renewal processes known as *delayed renewal processes*. The word *delayed* does not necessarily imply that the first renewal epoch is in any sense larger than the other inter-renewal intervals. Rather, it means that the usual renewal process, with IID inter-renewal times, is delayed until after the epoch of the first renewal. What we shall discover is intuitively satisfying — both the time-average behavior and, in essence, the limiting ensemble behavior are not affected by the distribution of the first renewal epoch. It might be somewhat surprising, however, to find that this irrelevance of the distribution of the first renewal epoch holds even when the mean of the first renewal epoch is infinite.

To be more precise, we let $\{X_i;\ i{\geq}1\}$ be a set of independent nonnegative random variables. X_1 has a given distribution function $G(x)$, whereas $\{X_i;\ i \geq 2\}$ are identically distributed with a given distribution function $F(x)$. Typically, $G(x) \neq F(x)$, since if equality held, we would have an ordinary renewal process. Let $S_n = \sum_{i=1}^{n} X_i$ be the nth renewal epoch. We first show that the SLLN still holds despite the deviant behavior of X_1.

Lemma 3.2. *Let $\{X_i;\ i \geq 2$ be IID with a mean $\overline{X}$ satisfying $\mathsf{E}[|X|] < \infty$ and let X_1 be a rv, independent of $\{X_i;\ i \geq 2\}$. Let $S_n = \sum_{i=1}^{n} X_i$. Then $\lim S_n/n = \overline{X}$ WP1.*

This generalizes the SLLN, but it is an almost trivial generalization. Note that

$$\frac{S_n}{n} = \frac{X_1}{n} + \frac{\sum_{i=2}^{n} X_i}{n}$$

Since X_1 is finite WP1, the first term above goes to 0 WP1. as $n \to \infty$. The second term goes to $\overline{X}$, proving the lemma and showing that it is not a big deal.

Now, for the given delayed renewal process, let $N(t)$ be the number of renewal epochs up to and including time t. This is still determined by the fact that $\{N(t) \geq n\}$ if and only if $\{S_n \leq t\}$). $\{N(t);\, t > 0\}$ is then called a delayed renewal counting process. The following simple lemma follows from lemma 3.1.

Lemma 3.3. *Let $\{N(t);\, t > 0\}$ be a delayed renewal counting process. Then $\lim_{t\to\infty} N(t) = \infty$ with probability 1 and $\lim_{t\to\infty} \mathsf{E}[N(t)] = \infty$.*

Proof: Conditioning on $X_1 = x$, we can write $N(t) = 1 + N'(t-x)$ where $N'\{t; t \geq 0\}$ is the ordinary renewal counting process with inter-renewal intervals $X_2, X_3, \dots$. From Lemma 3.1, $\lim_{t\to\infty} N'(t-x) = \infty$ with probability 1, and $\lim_{t\to\infty} \mathsf{E}[N'(t-x)] = \infty$. Since this is true for every finite $x > 0$, and X_1 is finite with probability 1, the lemma is proven.

Theorem 3.11 (Strong Law for Delayed Renewal Processes). *Let $N(t);\ t > 0$ be the renewal counting process for a delayed renewal process where the inter-renewal intervals*

$X_2, X_3, \dots,$ have distribution function F *and finite mean* $\overline{X} = \int_{x=0}^{\infty}[1 - F(x)]\,dx$. *Then*

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\overline{X}} \qquad \textit{WP1.} \tag{3.65}$$

Proof: Using Lemma 3.2, the conditions for Theorem 3.2 are fulfilled, so the proof follows exactly as the proof of Theorem 3.1. □.

Next we look at the elementary renewal theorem and Blackwell's theorem for delayed renewal processes. To do this, we view a delayed renewal counting process $\{N(t);\, t > 0\}$ as an ordinary renewal counting process that starts at a random nonnegative epoch X_1 with some distribution function $G(t)$. Define $N_o(t - X_1)$ as the number of renewals that occur in the interval $(X_1, t]$. Conditional on any given sample value x for X_1, $\{N_o(t-x);\, t-x > 0\}$ is an ordinary renewal counting process and thus, given $X_1 = x$, $\lim_{t\to\infty} \mathsf{E}[N_o(t-x)]/(t-x) = 1/\overline{X}$. Since $N(t) = 1 + N_o(t - X_1)$ for $t > X_1$, we see that, conditional on $X_1 = x$,

$$\lim_{t\to\infty} \frac{\mathsf{E}[N(t) \mid X_1{=}x]}{t} = \lim_{t\to\infty} \frac{\mathsf{E}[N_o(t-x)]}{t-x}\,\frac{t-x}{t} = \frac{1}{\overline{X}}. \tag{3.66}$$

Since this is true for every finite sample value x for X_1, we we have established the following theorem:

Theorem 3.12 (Elementary Delayed Renewal Theorem). *For a delayed renewal process with* $\mathsf{E}[X_i] = \overline{X}$ *for* $i \geq 2$*,*

$$\lim_{t\to\infty} \frac{\mathsf{E}[N(t)]}{t} = \frac{1}{\overline{X}}. \tag{3.67}$$

The same approach gives us Blackwell's theorem. Specifically, if $\{X_i;\, i{\geq}2\}$ is a sequence of IID non-arithmetic rv's, then, for any $\delta > 0$, Blackwell's theorem for ordinary renewal processes implies that

$$\lim_{t\to\infty} \frac{\mathsf{E}[N_o(t - x + \delta) - N_o(t-x)]}{\delta} = \frac{1}{\overline{X}}. \tag{3.68}$$

Thus, conditional on any sample value $X_1 = x$, $\lim_{t\to\infty} \mathsf{E}[N(t{+}\delta) - N(t) \mid X_1{=}x] = \delta/\overline{X}$. Taking the expected value over X_1 gives us $\lim_{t\to\infty} \mathsf{E}[N(t+\delta) - N(t)] = \delta/\overline{X}$. The case in which $\{X_i;\, i{\geq}2\}$ are arithmetic with span d is somewhat more complicated. If X_1 is arithmetic with span d (or a multiple of d), then the first renewal epoch must be at some multiple of d and $d/\overline{X}$ gives the expected number of arrivals at time id in the limit as $i \to \infty$. If X_1 is non-arithmetic or arithmetic with a span other than a multiple of d, then the effect of the first renewal epoch never dies out, since all subsequent renewals occur at multiples of d from this first epoch. We ignore this rather ugly case and state the following theorem for the nice situations.

Theorem 3.13 (Blackwell for Delayed Renewal). *If* $\{X_i;\, i{\geq}2\}$ *are non-arithmetic, then, for all* $\delta > 0$*,*

$$\lim_{t\to\infty} \frac{\mathsf{E}[N(t+\delta) - N(t)]}{\delta} = \frac{1}{\overline{X}}. \tag{3.69}$$

If $\{X_i; i \geq 2\}$ are arithmetic with span d and mean $\overline{X}$ and X_1 is arithmetic with span md for some positive integer m, then

$$\lim_{i\to\infty} \Pr\{\textit{renewal at } t = id\} = \frac{d}{\overline{X}}. \tag{3.70}$$

3.8.1 Delayed renewal-reward processes

We have seen that the distribution of the first renewal epoch has no effect on the time or ensemble-average behavior of a renewal process (other than the ensemble dependence on time for an arithmetic process). This carries over to reward functions with almost no change. In particular, the generalized version of Theorem 3.4 is as follows:

Theorem 3.14. *Let $\{N(t); t > 0\}$ be a delayed renewal counting process where the inter-renewal intervals $X_2, X_3, \dots$ have the distribution function F. Let $Z(t) = t - S_{N(t)}$, let $X(t) = S_{N(t)+1} - S_{N(t)}$, and let $R(t) = \mathcal{R}(Z(t), X(t))$ be a reward function. Assume that*

$$\mathsf{E}[\mathrm{R}_n] = \int_{x=0}^{\infty}\int_{z=0}^{x} \mathcal{R}(z,x)\,dz\,d\mathsf{F}(x) < \infty\,.$$

Then, with probability one,

$$\lim_{t\to\infty} \frac{1}{t}\int_{\tau=0}^{t} R(\tau)d\tau = \frac{\mathsf{E}[\mathrm{R}_n]}{\overline{X}_2} \textit{ for } n \geq 2. \tag{3.71}$$

We omit the proof of this since it is a minor variation of that of theorem 3.4. Finally, the equivalence of time and limiting ensemble averages holds as before, yielding

$$\lim_{t\to\infty} \mathsf{E}[R(t)] = \frac{\mathsf{E}[\mathrm{R}_n]}{\overline{X}_2}. \tag{3.72}$$

3.8.2 Transient behavior of delayed renewal processes

Let $m(t) = \mathsf{E}[N(t)]$ for a delayed renewal process. As in (3.32), we have

$$m(t) = \sum_{n=1}^{\infty} \Pr\{N(t) \geq n\} = \sum_{n=1}^{\infty} \Pr\{S_n \leq t\}\,. \tag{3.73}$$

For $n \geq 2$, $S_n = S_{n-1} + X_n$ where X_n and S_{n-1} are independent. From the convolution equation (1.12),

$$\Pr\{S_n \leq t\} = \int_{x=0}^{t} \Pr\{S_{n-1} \leq t - x\}\; dF(x) \quad \text{for}\, n \geq 2. \tag{3.74}$$

For $n = 1$, $\Pr\{S_n \le t\} = G(t)$. Substituting this in (3.73) and interchanging the order of integration and summation,

$$\begin{aligned} m(t) &= G(t) + \int_{x=0}^{t} \sum_{n=2}^{\infty} \Pr\{S_{n-1} \le t - x\}\, dF(x) \\ &= G(t) + \int_{x=0}^{t} \sum_{n=1}^{\infty} \Pr\{S_n \le t - x\}\, dF(x) \\ &= G(t) + \int_{x=0}^{t} m(t-x) dF(x)\,; \qquad t \ge 0. \end{aligned} \tag{3.75}$$

This is the *renewal equation* for delayed renewal processes and is a generalization of (3.33). It is shown to have a unique solution in [8], Section 11.1.

There is another useful integral equation very similar to (3.75) that arises from breaking up S_n as the sum of X_1 and $\widehat{S}_{n-1}$ where $\widehat{S}_{n-1} = X_2 + \cdots + X_n$. Letting $\widehat{m}(t)$ be the expected number of renewals in time t for an ordinary renewal process with interarrival distribution F, a similar argument to that above, starting with $\Pr\{S_n \le t\} = \int_0^t \Pr\left\{\widehat{S}_{n-1} \le t - x\right\}\, dG(x)$ yields

$$m(t) = G(t) + \int_{x=0}^{t} \widehat{m}(t-x) dG(x). \tag{3.76}$$

This equation brings out the effect of the initial renewal interval clearly, and is useful in computation if one already knows $\widehat{m}(t)$.

Frequently, the most convenient way of dealing with $m(t)$ is through transforms. Following the same argument as that in (3.34), we get $L_m(r) = (1/r)L_G(r) + L_m(r)L_F(r)$. Solving, we get

$$L_m(r) = \frac{L_G(r)}{r[1 - L_F(r)]}. \tag{3.77}$$

We can find $m(t)$ from (3.77) by finding the inverse Laplace transform, using the same procedure as in Example 3.5.1. There is a second order pole at $r = 0$ again, and, evaluating the residue, it is $1/L_F'(0) = 1/\overline{X}_2$, which is not surprising in terms of Blackwell's theorem. We can also expand numerator and denominator of (3.77) in a power series, as in (3.35). The inverse transform, corresponding to (3.36), is

$$m(t) = \frac{t}{\overline{X}} + \frac{\mathsf{E}\left[X_2^2\right]}{2\overline{X}} - \frac{\overline{X}_1}{\overline{X}} + \epsilon(t) \quad \text{for}\, t \to 0. \tag{3.78}$$

where $\lim_{t\to\infty} \epsilon(t) = 0$.

3.8.3 The equilibrium process

Consider an ordinary non-arithmetic renewal process with an inter-renewal interval X of distribution $F(x)$. We have seen that the distribution of the interval from t to the next renewal approaches $\mathsf{F}_Y(y) = (1/\mathsf{E}[X]) \int_0^y [1 - F(x)]dx$ as $t \to \infty$. This suggests that if we look at this renewal process starting at some very large t, we should see a delayed renewal process for which the distribution $G(x)$ of the first renewal is equal to the residual life distribution $\mathsf{F}_Y(x)$ above and subsequent inter-renewal intervals should have the original distribution $F(x)$ above. Thus it appears that such a delayed renewal process is the same as the original ordinary renewal process, except that it starts in "steady state." To verify this, we show that $m(t) = t/\overline{X}$ is a solution to (3.75) if $G(t) = \mathsf{F}_Y(t)$. Substituting $(t-x)/\overline{X}$ for $m(t-x)$, the right hand side of (3.75) is

$$\frac{\int_0^t [1 - F(x)]dx}{\overline{X}_2} + \frac{\int_0^t (t-x)dF(x)}{\overline{X}} = \frac{\int_0^t [1 - F(x)]dx}{\overline{X}} + \frac{\int_0^t F(x)dx}{\overline{X}} = \frac{t}{\overline{X}}.$$

where we have used integration by parts for the first equality. This particular delayed renewal process is called the *equilibrium process*, since it starts off in steady state, and thus has no transients.

3.9 Summary

Sections 3.1 to 3.6 developed the central results about renewal processes that frequently appear in subsequent chapters. The chapter starts with the strong law for renewal processes, showing that the time average rate of renewals, $N(t)/t$, approaches $1/\overline{X}$ with probability 1 as $t \to \infty$. This, combined with the strong law of large numbers in Chapter 1, is the basis for most subsequent results about time-averages. Section 3.3 adds a reward function $R(t)$ to the underlying renewal process. These reward functions are defined to depend only on the inter-renewal interval containing t, and are used to study many surprising aspects of renewal processes such as residual life, age, and duration. For all sample paths of a renewal process (except a subset of probabiity 0), the time-average reward for a given $R(t)$ is a constant, and that constant is the expected aggregate reward over an inter-renewal interval divided by the expect length of an inter-renewal interval.

The next topic, in Section 3.4 is that of stopping times or stopping trials. These have obvious applications to situations where an experiment or game is played until some desired (or undesired) outcome (based on the results up to and including the given trial) occurs. This is a basic and important topic in its right, but is also needed to understand both how the expected renewal rate $\mathsf{E}[N(t)]/t$ varies with time t and how renewal theory can be applied to queueing situations.

This is followed, in Section 3.5, by an analysis of how $\mathsf{E}[N(t)]/t$ varies with t. This starts by using Laplace transforms to get a complete solution of the ensemble-average, $\mathsf{E}[N(t)]/t$, as a function of t, when the distribution of the inter-renewal interval has a rational Laplace transform. For the general case (where the Laplace transform is irrational or non-existant), the elementary renewal theorem shows that $\lim_{t\to\infty} \mathsf{E}[N(t)]/t = 1/\overline{X}$. The fact that the time-average (WP1) and the limiting ensemble-average are the same is not surprising, and the fact that the ensemble-average has a limit is not surprising. These results are so fundamental to other results in probability, however, that they deserve to be understood.

Another fundamental result in Section 3.5 is Blackwell's renewal theorem, showing that the distribution of renewal epochs reach a steady state as $t \to \infty$. The form of that steady state depends on whether the inter-renewal distribution is arithmetic (see (3.39)) or non-arithmetic (see (3.38)).

Section 3.6 ties together the results on rewards in 3.3 to those on ensemble averages in 3.5. Under some very minor restrictions imposed by the key renewal theorem, we found that, for non-arithmetic inter-renewal distributions, $\lim_{t\to\infty} \mathsf{E}[R(t)]$ is the same as the time-average value of reward. These general results were used in Section 3.7 to derive and understand Little's theorem and the Pollaczek-Khinchin expression for the expected delay in an M/G/1 queue.

Finally, all the results above were shown to apply to delayed renewal processes.

For further reading on renewal processes, see Feller,[8], Ross, [16], or Wolff, [22]. Feller still appears to be the best source for deep understanding of renewal processes, but Ross and Wolff are somewhat more accessible.

3.10 Exercises

Exercise 3.1. The purpose of this exercise is to show that for an arbitrary renewal process, $N(t)$, the number of renewals in $(0, t]$ is a (non-defective) random variable.

a) Let $X_1, X_2, \dots$, be a sequence of IID inter-renewal rv's . Let $S_n = X_1 + \cdots + X_n$ be the corresponding renewal epochs for each $n \geq 1$. Assume that each X_i has a finite expectation $\overline{X} > 0$ and, for any given $t > 0$, use the weak law of large numbers to show that $\lim_{n\to\infty} \Pr\{S_n \leq t\} = 0$.

b) Use part a) to show that $\lim_{n\to\infty} \Pr\{N \geq n\} = 0$ and explain why this means that $N(t)$ is a rv, *i.e.*, is not defective.

c) Now suppose that the X_i do not have a finite mean. Consider truncating each X_i to $\breve{X}_i$, where for any given $b > 0$, $\breve{X}_i = \min(X_i, b)$. Let $\breve{N}(t)$ be the renewal counting process for the inter-renewal intervals $\breve{X}_i$. Show that $\breve{N}(t)$ is non-defective for each $t > 0$. Show that $N(t) \leq \breve{N}(t)$ and thus that $N(t)$ is non-defective. Note: Large inter-renewal intervals create small values of $N(t)$, and thus $\mathsf{E}[X] = \infty$ has nothing to do with potentially large values of $N(t)$, so the argument here was purely technical.

Exercise 3.2. The purpose of this exercise is to show that, for an arbitrary renewal process, $N(t)$, the number of renewals in $(0, t]$, has finite expectation.

a) Let the inter-renewal intervals have the distribution $\mathsf{F}_X(x)$, with, as usual, $\mathsf{F}_X(0) = 0$. Using whatever combination of mathematics and common sense is comfortable for you, show that numbers $\epsilon > 0$ and $\delta > 0$ must exist such that $\mathsf{F}_X(\delta) \leq 1-\epsilon$. In other words, you are to show that a positive rv must take on some range of of positive values with positive probability.

b) Show that $\Pr\{S_n \leq \delta\} \leq (1-\epsilon)^n$.

c) Show that $\mathsf{E}[N(\delta)] \leq 1/\epsilon$. Hint: Write (1.30) as a sum.

d) Show that for every integer k, $\mathsf{E}[N(k\delta)] \leq k/\epsilon$ and thus that $\mathsf{E}[N(t)] \leq \frac{t+\delta}{\epsilon\delta}$ for any $t > 0$.

e) Use your result here to show that $N(t)$ is non-defective.

Exercise 3.3. Let $\{X_i; i \geq 1\}$ be the inter-renewal intervals of a renewal process generalized to allow for inter-renewal intervals of size 0 and let $\Pr\{X_i = 0\} = \alpha$, $0 < \alpha < 1$. Let $\{Y_i;\ i \geq 1\}$ be the sequence of non-zero interarrival intervals. For example, if $X_1 = x_1 > 0,\ X_2 = 0,\ X_3 = x_3 > 0, \dots$, then $Y_1 = x_1$, $Y_2 = x_3$, $\dots$, .

a) Find the distribution function of each Y_i in terms of that of the X_i.

b) Find the PMF of the number of arrivals of the generalized renewal process at each epoch at which arrivals occur.

c) Explain how to view the generalized renewal process as an ordinary renewal process with inter-renewal intervals $\{Y_i;\ i \geq 1\}$ and bulk arrivals at each renewal epoch.

d) When a generalized renewal process is viewed as an ordinary renewal process with bulk arrivals, what is the distribution of the bulk arrivals? (The point of this part is to illustrate that bulk arrivals on an ordinary renewal process are considerably more general than generalized renewal processes.)

Exercise 3.4. Is it true for a renewal process that:

a) $N(t) < n$ if and only if $S_n > t$?

b) $N(t) \leq n$ if and only if $S_n \geq t$?

c) $N(t) > n$ if and only if $S_n < t$?

Exercise 3.5. Let $\{X_i;\ i \geq 1\}$ be the inter-renewal intervals of a renewal process and assume that $\mathsf{E}[X_i] = \infty$. Let $b > 0$ be an arbitrary number and $\breve{X}_i$ be a truncated random variable defined by $\breve{X}_i = X_i$ if $X_i \leq b$ and $\breve{X}_i = b$ otherwise.

a) Show that for any constant $M > 0$, there is a b sufficiently large so that $\mathsf{E}\left[\breve{X}_i\right] \geq M$.

b) Let $\{\breve{N}(t);\ t \geq 0\}$ be the renewal counting process with inter-renewal intervals $\{\breve{X}_i; i \geq 1\}$ and show that for all $t > 0$, $\breve{N}(t) \geq N(t)$.

c) Show that for all sample functions $N(t, \omega)$, except a set of probability 0, $N(t, \omega)/t < 2/M$ for all sufficiently large t. Note: Since M is arbitrary, this means that $\lim N(t)/t = 0$ with probability 1.

Exercise 3.6. Let $Y(t) = S_{N(t)+1} - t$ be the residual life at time t of a renewal process. First consider a renewal process in which the interarrival time has density $\mathsf{f}_X(x) = e^{-x}$; $x \geq 0$, and next consider a renewal process with density

$$\mathsf{f}_X(x) = \frac{3}{(x+1)^4}; \quad x \geq 0.$$

For each of the above densities, use renewal-reward theory to find:

i) the time-average of $Y(t)$

ii) the second moment in time of $Y(t)$ (i.e., $\lim_{T\to\infty} \frac{1}{T}\int_0^T Y^2(t)dt$)

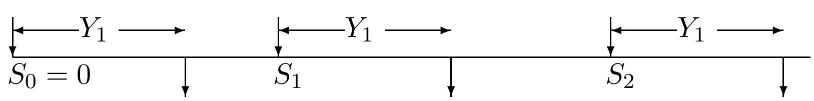
For the exponential density, verify your answers by finding $\mathsf{E}[Y(t)]$ and $\mathsf{E}\left[Y^2(t)\right]$ directly.

Exercise 3.7. Consider a variation of an M/G/1 queueing system in which there is no facility to save waiting customers. Assume customers arrive according to a Poisson process of rate λ. If the server is busy, the customer departs and is lost forever; if the server is not busy, the customer enters service with a service time distribution function denoted by $\mathsf{F}_Y(y)$.

Successive service times (for those customers that are served) are IID and independent of arrival times. Assume that customer number 0 arrives and enters service at time $t = 0$.

a) Show that the sequence of times $S_1, S_2, \ldots$ at which successive customers enter service are the renewal times of a renewal process. Show that each inter-renewal interval $X_i = S_i - S_{i-1}$ (where $S_0 = 0$) is the sum of two independent random variables, $Y_i + U_i$ where Y_i is the ith service time; find the probability density of U_i.

b) Assume that a reward (actually a cost in this case) of one unit is incurred for each customer turned away. Sketch the expected reward function as a function of time for the sample function of inter-renewal intervals and service intervals shown below; the expectation is to be taken over those (unshown) arrivals of customers that must be turned away.



c) Let $\int_0^t R(\tau)d\tau$ denote the accumulated reward (i.e., cost) from 0 to t and find the limit as $t \to \infty$ of $(1/t)\int_0^t R(\tau)d\tau$. Explain (without any attempt to be rigorous or formal) why this limit exists with probability 1.

d) In the limit of large t, find the expected reward from time t until the next renewal. Hint: Sketch this expected reward as a function of t for a given sample of inter-renewal intervals and service intervals; then find the time-average.

e) Now assume that the arrivals are deterministic, with the first arrival at time 0 and the nth arrival at time $n - 1$. Does the sequence of times $S_1, S_2, \ldots$ at which subsequent customers start service still constitute the renewal times of a renewal process? Draw a sketch of arrivals, departures, and service time intervals. Again find $\lim_{t\to\infty}\left(\int_0^t R(\tau)\,d\tau\right)/t$.

Exercise 3.8. Let $Z(t) = t - S_{N(t)}$ be the age of a renewal process and $Y(t) = S_{N(t)+1} - t$ be the residual life. Let $\mathsf{F}_X(x)$ be the distribution function of the inter-renewal interval and find the following as a function of $\mathsf{F}_X(x)$:

a) $\Pr\{Y(t){>}x \mid Z(t){=}s\}$

b) $\Pr\{Y(t){>}x \mid Z(t{+}x/2){=}s\}$

c) $\Pr\{Y(t){>}x \mid Z(t{+}x){>}s\}$ for a Poisson process.

Exercise 3.9. Let $\mathsf{F}_Z(z)$ be the fraction of time (over the limiting interval $(0, \infty)$ that the age of a renewal process is at most z. Show that $\mathsf{F}_Z(z)$ satisfies

$$\mathsf{F}_Z(z) = \frac{1}{\overline{X}} \int_{x=0}^{y} \Pr\{X > x\}\, dx \qquad \text{WP1.}$$

Hint: Follow the argument in Example 3.3.5.

Exercise 3.10. a) Let J be a stopping rule and $\mathbb{I}_{J \geq n}$ be the indicator random variable of the event $\{J \geq n\}$. Show that $J = \sum_{n \geq 1} \mathbb{I}_{J \geq n}$.

b) Show that $\mathbb{I}_{J \geq 1} \geq \mathbb{I}_{J \geq 2} \geq \dots$, *i.e.*, show that for each $n > 1$, $\mathbb{I}_{J \geq n}(\omega) \geq \mathbb{I}_{J \geq n+1}(\omega)$ for each $\omega \in \Omega$ (except perhaps for a set of probability 0).

Exercise 3.11. a) Use Wald's equality to compute the expected number of trials of a Bernoulli process up to and including the kth success.

b) Use elementary means to find the expected number of trials up to and including the first success. Use this to find the expected number of trials up to and including the kth success. Compare with part a).

Exercise 3.12. A gambler with an initial finite capital of $d > 0$ dollars starts to play a dollar slot machine. At each play, either his dollar is lost or is returned with some additional number of dollars. Let X_i be his change of capital on the ith play. Assume that $\{X_i;\, i{=}1, 2, \dots\}$ is a set of IID random variables taking on integer values $\{-1, 0, 1, \dots\}$. Assume that $\mathsf{E}[X_i] < 0$. The gambler plays until losing all his money (i.e., the initial d dollars plus subsequent winnings).

a) Let J be the number of plays until the gambler loses all his money. Is the weak law of large numbers sufficient to argue that $\lim_{n\to\infty} \Pr\{J > n\} = 0$ (i.e., that J is a random variable) or is the strong law necessary?

b) Find $\mathsf{E}[J]$.

Exercise 3.13. Let $\{X_i; i \geq 1\}$ be IID binary random variables with $P_X(0) = P_X(1) = 1/2$. Let J be a positive integer-valued random variable defined on the above sample space of binary sequences and let $S_J = \sum_{i=1}^{J} X_i$. Find the simplest example you can in which J is not a stopping time for $\{X_i; i \geq 1\}$ and where $\mathsf{E}[X]\,\mathsf{E}[J] \neq \mathsf{E}[S_J]$. Hint: Try letting J take on only the values 1 and 2.

Exercise 3.14. Let $J = \min\{n \mid S_n{\leq}b \text{ or } S_n{\geq}a\}$, where a is a positive integer, b is a negative integer, and $S_n = X_1 + X_2 + \cdots + X_n$. Assume that $\{X_i;\, i{\geq}1\}$ is a set of zero mean IID rv's that can take on only the set of values $\{-1, 0, +1\}$, each with positive probability.

a) Is J a stopping rule? Why or why not? Hint: The more difficult part of this is to argue that J is a random variable (*i.e.*, non-defective); you do not need to construct a proof of this, but try to argue why it must be true.

b) What are the possible values of S_J?

c) Find an expression for $\mathsf{E}[S_J]$ in terms of p, a, and b, where $p = \Pr\{S_J \geq a\}$.

d) Find an expression for $\mathsf{E}[S_J]$ from Wald's equality. Use this to solve for p.

Exercise 3.15. Show that the interchange of expectation and sum in (3.27) is valid if $\mathsf{E}[J] < \infty$. Hint: First express the sum as $\sum_{n=1}^{k-1} X_n \mathbb{I}_{J\geq n} + \sum_{n=k}^{\infty}(X_n^+ + X_n^-)\mathbb{I}_{J\geq n}$ and then consider the limit as $k \to \infty$.

Exercise 3.16. Consider a miner trapped in a room that contains three doors. Door 1 leads him to freedom after two-day's travel; door 2 returns him to his room after four-day's travel; and door 3 returns him to his room after eight-day's travel. Suppose each door is equally likely to be chosen whenever he is in the room, and let T denote the time it takes the miner to become free.

a) Define a sequence of independent and identically distributed random variables $X_1, X_2, \ldots$ and a stopping rule J such that

$$T = \sum_{i=1}^{J} X_i.$$

b) Use Wald's equality to find $\mathsf{E}[T]$.

c) Compute $\mathsf{E}\left[\sum_{i=1}^{J} X_i \mid J{=}n\right]$ and show that it is not equal to $\mathsf{E}[\sum_{i=1}^{n} X_i]$.

d) Use part c) for a second derivation of $\mathsf{E}[T]$.

Exercise 3.17. Let $\{N(t);\ t > 0\}$ be a renewal counting process generalized to allow for inter-renewal intervals $\{X_i\}$ of duration 0. Let each X_i have the PMF $\Pr\{X_i = 0\} = 1 - \epsilon$; $\Pr\{X_i = 1/\epsilon\} = \epsilon$.

a) Sketch a typical sample function of $\{N(t);\ t > 0\}$. Note that $N(0)$ can be non-zero (i.e., $N(0)$ is the number of zero interarrival times that occur before the first non-zero interarrival time).

b) Evaluate $\mathsf{E}[N(t)]$ as a function of t.

c) Sketch $\mathsf{E}[N(t)]/t$ as a function of t.

d) Evaluate $\mathsf{E}\left[S_{N(t)+1}\right]$ as a function of t (do this directly, and then use Wald's equality as a check on your work).

e) Sketch the lower bound $\mathsf{E}[N(t)]/t \geq 1/\mathsf{E}[X] - 1/t$ on the same graph with part c).

f) Sketch $\mathsf{E}\left[S_{N(t)+1} - t\right]$ as a function of t and find the time average of this quantity.

g) Evaluate $\mathsf{E}\left[S_{N(t)}\right]$ as a function of t; verify that $\mathsf{E}\left[S_{N(t)}\right] \neq \mathsf{E}[X]\,\mathsf{E}[N(t)]$.

Exercise 3.18. Let $\{N(t);\, t > 0\}$ be a renewal counting process and let $m(t) = \mathsf{E}[N(t)]$ be the expected number of arrivals up to and including time t. Let $\{X_i; i \geq 1\}$ be the inter-renewal times and assume that $\mathsf{F}_X(0) = 0$.

a) For all $x > 0$ and $t > x$ show that $\mathsf{E}[N(t)|X_1{=}x] = \mathsf{E}[N(t-x)] + 1$.

b) Use part a) to show that $m(t) = \mathsf{F}_X(t) + \int_0^t m(t-x)d\mathsf{F}_X(x)$ for $t > 0$. This equation is the renewal equation derived differently in (3.33).

c) Suppose that X is an exponential random variable of parameter λ. Evaluate $L_m(s)$ from (3.34); verify that the inverse Laplace transform is $\lambda t;\ t{\geq}0$.

Exercise 3.19. a) Let the inter-renewal interval of a renewal process have a second order Erlang density, $\mathsf{f}_X(x) = \lambda^2 x \exp(-\lambda x)$. Evaluate the Laplace transform of $m(t) = \mathsf{E}[N(t)]$.

b) Use this to evaluate $m(t)$ for $t \geq 0$. Verify that your answer agrees with (3.36).

c) Evaluate the slope of $m(t)$ at $t = 0$ and explain why that slope is not surprising.

d) View the renewals here as being the even numbered arrivals in a Poisson process of rate λ. Sketch $m(t)$ for the process here and show one half the expected number of arrivals for the Poisson process on the same sketch. Explain the difference between the two.

Exercise 3.20. a) Let $N(t)$ be the number of arrivals in the interval $(0, t]$ for a Poisson process of rate λ. Show that the probability that $N(t)$ is even is $[1 + \exp(-2\lambda t)]/2$. Hint: Look at the power series expansion of $\exp(-\lambda t)$ and that of $\exp(\lambda t)$, and look at the sum of the two. Compare this with $\sum_{n\,\text{even}} \Pr\{N(t) = n\}$.

b) Let $\widetilde{N}(t)$ be the number of even numbered arrivals in $(0, t]$. Show that $\widetilde{N}(t) = N(t)/2 - \mathbb{I}_{\text{odd}}(t)/2$ where $\mathbb{I}_{\text{odd}}(t)$ is a random variable that is 1 if $N(t)$ is odd and 0 otherwise.

c) Use parts a) and b) to find $\mathsf{E}\left[\widetilde{N}(t)\right]$. Note that this is $m(t)$ for a renewal process with 2nd order Erlang inter-renewal intervals.

Exercise 3.21. a) Consider a function $r(z)$ defined for $0 \leq z < \infty$ as follows: For each integer $n \geq 1$ and each integer $k, 1 \leq k < n$, $r(z) = 1$ for $n + k/n \leq z \leq n + k/n + 2^{-n}$. For all other z, $r(z) = 0$. Sketch this function and show that $r(z)$ is not directly Riemann integrable.

b) Evaluate the Riemann integral $\int_0^\infty r(z)dz$.

Exercise 3.22. Let $Z(t), Y(t), X(t)$ denote the age, residual life, and duration at time t for a renewal counting process $\{N(t);\, t > 0\}$ in which the interarrival time has a density given by $f(x)$. Find the following probability densities; assume steady state.

a) $\mathsf{f}_{Y(t)}(y \,|\, Z(t{+}s/2){=}s)$ for given $s > 0$.

b) $\mathsf{f}_{Y(t),Z(t)}(y, z)$.

c) $\mathsf{f}_{Y(t)}(y \,|\, X(t){=}x)$.

d) $\mathsf{f}_{Z(t)}(z \mid Y(t{-}s/2){=}s)$ for given $s > 0$.

e) $\mathsf{f}_{Y(t)}(y \mid Z(t{+}s/2){\geq}s)$ for given $s > 0$.

Exercise 3.23. **a)** Find $\lim_{t\to\infty}\{\mathsf{E}[N(t)] - t/\overline{X}\}$ for a renewal counting process $\{N(t); t > 0\}$ with inter-renewal times $\{X_i; i \geq 1\}$. Hint: use Wald's equation.

b) Evaluate your result for the case in which X is an exponential random variable (you already know what the result should be in this case).

c) Evaluate your result for a case in which $\mathsf{E}[X] < \infty$ and $\mathsf{E}\left[X^2\right] = \infty$. Explain (very briefly) why this does not contradict the elementary renewal theorem.

Exercise 3.24. Customers arrive at a bus stop according to a Poisson process of rate λ. Independently, buses arrive according to a renewal process with the inter-renewal interval distribution $\mathsf{F}_X(x)$. At the epoch of a bus arrival, all waiting passengers enter the bus and the bus leaves immediately. Let $R(t)$ be the number of customers waiting at time t.

a) Draw a sketch of a sample function of $R(t)$.

b) Given that the first bus arrives at time $X_1 = x$, find the expected number of customers picked up; then find $\mathsf{E}\left[\int_0^x R(t)dt\right]$, again given the first bus arrival at $X_1 = x$.

c) Find $\lim_{t\to\infty} \frac{1}{t}\int_0^t R(\tau)d\tau$ (with probability 1). Assuming that F_X is a non-arithmetic distribution, find $\lim_{t\to\infty}\mathsf{E}[R(t)]$. Interpret what these quantities mean.

d) Use part c) to find the time-average expected wait per customer.

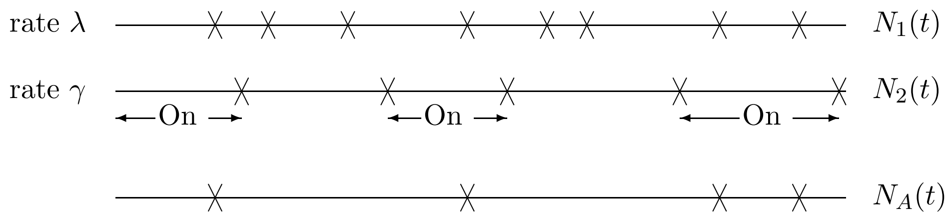
e) Find the fraction of time that there are no customers at the bus stop. (Hint: this part is independent of a), b), and c); check your answer for $\mathsf{E}[X] \ll 1/\lambda$).

Exercise 3.25. Consider the same setup as in Exercise 3.24 except that now customers arrive according to a non-arithmetic renewal process independent of the bus arrival process. Let $1/\lambda$ be the expected inter-renewal interval for the customer renewal process. Assume that both renewal processes are in steady state (i.e., either we look only at $t \gg 0$, or we assume that they are equilibrium processes). Given that the nth customer arrives at time t, find the expected wait for customer n. Find the expected wait for customer n without conditioning on the arrival time.

Exercise 3.26. Let $\{N_1(t); t \geq 0\}$ be a Poisson counting process of rate λ. Assume that the arrivals from this process are switched on and off by arrivals from a non-arithmetic renewal counting process $\{N_2(t); t \geq 0\}$ (see figure below). The two processes are independent.

Let $\{N_A(t); t \geq 0\}$ be the switched process; that is $N_A(t)$ includes arrivals from $\{N_1(t); t \geq 0\}$ while $N_2(t)$ is even and excludes arrivals from $\{N_1(t); t \geq 0\}$ while $N_2(t)$ is odd.

a) Is $N_A(t)$ a renewal counting process? Explain your answer and if you are not sure, look at several examples for $N_2(t)$.



b) Find $\lim_{t\to\infty} \frac{1}{t} N_A(t)$ and explain why the limit exists with probability 1. Hint: Use symmetry—that is, look at $N_1(t) - N_A(t)$. To show why the limit exists, use the renewal-reward theorem. What is the appropriate renewal process to use here?

c) Now suppose that $\{N_1(t); t\geq 0\}$ is a non-arithmetic renewal counting process but not a Poisson process and let the expected inter-renewal interval be $1/\lambda$. For any given δ, find $\lim_{t\to\infty} \mathsf{E}[N_A(t+\delta) - N_A(t)]$ and explain your reasoning. Why does your argument in (b) fail to demonstrate a time-average for this case?

Exercise 3.27. An M/G/1 queue has arrivals at rate λ and a service time distribution given by $\mathsf{F}_Y(y)$. Assume that $\lambda < 1/\mathsf{E}[Y]$. Epochs at which the system becomes empty define a renewal process. Let $\mathsf{F}_Z(z)$ be the distribution of the inter-renewal intervals and let $\mathsf{E}[Z]$ be the mean inter-renewal interval.

a) Find the fraction of time that the system is empty as a function of λ and $\mathsf{E}[Z]$. State carefully what you mean by such a fraction.

b) Apply Little's theorem, not to the system as a whole, but to the number of customers in the server (i.e., 0 or 1). Use this to find the fraction of time that the server is busy.

c) Combine your results in a) and b) to find $\mathsf{E}[Z]$ in terms of λ and $\mathsf{E}[Y]$; give the fraction of time that the system is idle in terms of λ and $\mathsf{E}[Y]$.

d) Find the expected duration of a busy period.

Exercise 3.28. Consider a sequence $X_1, X_2, \dots$ of IID binary random variables. Let p and $1-p$ denote $\Pr\{X_m = 1\}$ and $\Pr\{X_m = 0\}$ respectively. A *renewal* is said to occur at time m if $X_{m-1} = 0$ and $X_m = 1$.

a) Show that $\{N(m); m \geq 0\}$ is a renewal counting process where $N(m)$ is the number of renewals up to and including time m and $N(0)$ and $N(1)$ are taken to be 0.

b) What is the probability that a renewal occurs at time m, $m \geq 2$?

c) Find the expected inter-renewal interval; use Blackwell's theorem here.

d) Now change the definition of renewal; a renewal now occurs at time m if $X_{m-1} = 1$ and $X_m = 1$. Show that $\{N_m^*; m \geq 0\}$ is a delayed renewal counting process where N_m^* is the number of renewals up to and including m for this new definition of renewal ($N_0^* = N_1^* = 0$).

e) Find the expected inter-renewal interval for the renewals of part d).

f) Given that a renewal (according to the definition in (d)) occurs at time m, find the expected time until the next renewal, conditional, first, on $X_{m+1} = 1$ and, next, on $X_{m+1} = 0$. Hint: use the result in e) plus the result for $X_{m+1} = 1$ for the conditioning on $X_{m+1} = 0$.

g) Use your result in f) to find the expected interval from time 0 to the first renewal according to the renewal definition in d).

h) Which pattern requires a larger expected time to occur: 0011 or 0101

i) What is the expected time until the first occurrence of 0111111?

Exercise 3.29. A large system is controlled by n identical computers. Each computer independently alternates between an operational state and a repair state. The duration of the operational state, from completion of one repair until the next need for repair, is a random variable X with finite expected duration $\mathsf{E}[X]$. The time required to repair a computer is an exponentially distributed random variable with density $\lambda e^{-\lambda t}$. All operating durations and repair durations are independent. Assume that all computers are in the repair state at time 0.

a) For a single computer, say the ith, do the epochs at which the computer enters the repair state form a renewal process? If so, find the expected inter-renewal interval.

b) Do the epochs at which it enters the operational state form a renewal process?

c) Find the fraction of time over which the ith computer is operational and explain what you mean by fraction of time.

d) Let $Q_i(t)$ be the probability that the ith computer is operational at time t and find $\lim_{t\to\infty} Q_i(t)$.

e) The system is in failure mode at a given time if all computers are in the repair state at that time. Do the epochs at which system failure modes begin form a renewal process?

f) Let $\Pr\{t\}$ be the probability that the the system is in failure mode at time t. Find $\lim_{t\to\infty} \Pr\{t\}$. Hint: look at part d).

g) For δ small, find the probability that the system enters failure mode in the interval $(t, t+\delta]$ in the limit as $t \to \infty$.

h) Find the expected time between successive entries into failure mode.

i) Next assume that the repair time of each computer has an arbitrary density rather than exponential, but has a mean repair time of $1/\lambda$. Do the epochs at which system failure modes begin form a renewal process?

j) Repeat part f) for the assumption in (i).

Exercise 3.30. Let $\{N_1(t);\, t{\ge}0\}$ and $\{N_2(t);\, t{\ge}0\}$ be independent renewal counting processes. Assume that each has the same distribution function $F(x)$ for interarrival intervals and assume that a density $f(x)$ exists for the interarrival intervals.

a) Is the counting process $\{N_1(t) + N_2(t); t \ge 0\}$ a renewal counting process? Explain.

b) Let $Y(t)$ be the interval from t until the first arrival (from either process) after t. Find an expression for the distribution function of $Y(t)$ in the limit $t \to \infty$ (you may assume that time averages and ensemble-averages are the same).

c) Assume that a reward R of rate 1 unit per second starts to be earned whenever an arrival from process 1 occurs and ceases to be earned whenever an arrival from process 2 occurs.

Assume that $\lim_{t\to\infty}(1/t)\int_0^t R(\tau)\,d\tau$ exists with probability 1 and find its numerical value.

d) Let $Z(t)$ be the interval from t until the first time after t that $R(t)$ (as in part c) changes value. Find an expression for $\mathsf{E}[Z(t)]$ in the limit $t\to\infty$. Hint: Make sure you understand why $Z(t)$ is not the same as $Y(t)$ in part b). You might find it easiest to first find the expectation of $Z(t)$ conditional on both the duration of the $\{N_1(t); t\geq 0\}$ interarrival interval containing t and the duration of the $\{N_2(t); t\geq 0\}$ interarrival interval containing t; draw pictures!

Exercise 3.31. This problem provides another way of treating ensemble-averages for renewal-reward problems. Assume for notational simplicity that X is a continuous random variable.

a) Show that $\Pr\{\text{one or more arrivals in } (\tau,\tau+\delta)\} = m(\tau+\delta)-m(\tau)-o(\delta)$ where $o(\delta)\geq 0$ and $\lim_{\delta\to 0} o(\delta)/\delta = 0$.

b) Show that $\Pr\{Z(t)\in[z,z+\delta),\ X(t)\in(x,x+\delta)\}$ is equal to $[m(t-z)-m(t-z-\delta)-o(\delta)][\mathsf{F}_X(x+\delta)-\mathsf{F}_X(x)]$ for $x\geq z+\delta$.

c) Assuming that $m'(\tau)=dm(\tau)/d\tau$ exists for all τ, show that the joint density of $Z(t)$, $X(t)$ is $\mathsf{f}_{Z(t),X(t)}(z,x) = m'(t-z)\mathsf{f}_X(x)$ for $x>z$.

d) Show that $\mathsf{E}[R(t)] = \int_{z=0}^{t}\int_{x=z}^{\infty}\mathcal{R}(z,x)\mathsf{f}_X(x)dx\, m'(t-z)dz$

Exercise 3.32. This problem is designed to give you an alternate way of looking at ensemble-averages for renewal-reward problems. First we find an exact expression for $\Pr\{S_{N(t)}>s\}$. We find this for arbitrary s and t, $0<s<t$.

a) By breaking the event $\{S_{N(t)}>s\}$into subevents $\{S_{N(t)}>s, N(t)=n\}$, explain each of the following steps:

$$\begin{aligned}
\Pr\{S_{N(t)}>s\} &= \sum_{n=1}^{\infty}\Pr\{t\geq S_n>s, S_{n+1}>t\}\\
&= \sum_{n=1}^{\infty}\int_{y=s}^{t}\Pr\{S_{n+1}{>}t\mid S_n{=}y\}\ d\mathsf{F}_{S_n}(y)\\
&= \int_{y=s}^{t}\mathsf{F}^{\mathsf{c}}_X(t{-}y)\,d\sum_{n=1}^{\infty}\mathsf{F}_{S_n}(y)\\
&= \int_{y=s}^{t}\mathsf{F}^{\mathsf{c}}_X(t{-}y)\,dm(y)\quad \text{where } m(y)=\mathsf{E}[N(y)].
\end{aligned}$$

b) Show that for $0<s<t<u$,

$$\Pr\{S_{N(t)}>s, S_{N(t)+1}>u\} = \int_{y=s}^{t}\mathsf{F}^{\mathsf{c}}_X(u-y)\,dm(y).$$

c) Draw a two dimensional sketch, with age and duration as the axes, and show the region of (age, duration) values corresponding to the event $\{S_N(t) > s, \ S_{N(t)+1} > u\}$.

d) Assume that for large t, $dm(y)$ can be approximated (according to Blackwell) as $(1/\overline{X})dy$, where $\overline{X} = \mathsf{E}[X]$. Assuming that X also has a density, use the result in parts b) and c) to find the joint density of age and duration.

Exercise 3.33. In this problem, we show how to calculate the residual life distribution $Y(t)$ as a transient in t. Let $\mu(t) = dm(t)/dt$ where $m(t) = \mathsf{E}[N(t)]$, and let the interarrival distribution have the density $\mathsf{f}_X(x)$. Let $Y(t)$ have the density $\mathsf{f}_{Y(t)}(y)$.

a) Show that these densities are related by the integral equation

$$\mu(t+y) = \mathsf{f}_{Y(t)}(y) + \int_{u=0}^{y} \mu(t+u)\mathsf{f}_X(y-u)du.$$

b) Let $L_{\mu,t}(r) = \int_{y\geq 0} \mu(t+y)e^{-ry}dy$ and let $L_{Y(t)}(r)$ and $L_X(r)$ be the Laplace transforms of $\mathsf{f}_{Y(t)}(y)$ and $\mathsf{f}_X(x)$ respectively. Find $L_{Y(t)}(r)$ as a function of $L_{\mu,t}$ and L_X.

c) Consider the inter-renewal density $\mathsf{f}_X(x) = (1/2)e^{-x} + e^{-2x}$ for $x \geq 0$ (as in Example 3.5.1). Find $L_{\mu,t}(r)$ and $L_{Y(t)}(r)$ for this example.

d) Find $\mathsf{f}_{Y(t)}(y)$. Show that your answer reduces to that of (3.23) in the limit as $t \to \infty$.

e) Explain how to go about finding $\mathsf{f}_{Y(t)}(y)$ in general, assuming that f_X has a rational Laplace transform.

Exercise 3.34. Show that for a G/G/1 queue, the time-average wait in the system is the same as $\lim_{n\to\infty} \mathsf{E}[W_n]$. Hint: Consider an integer renewal counting process $\{M(n); n \geq 0\}$ where $M(n)$ is the number of renewals in the G/G/1 process of Section 3.7 that have occurred by the nth arrival. Show that this renewal process has a span of 1. Then consider $\{W_n; n \geq 1\}$ as a reward within this renewal process.

Exercise 3.35. If one extends the definition of renewal processes to include inter-renewal intervals of duration 0, with $\Pr\{X{=}0\} = \alpha$, show that the expected number of simultaneous renewals at a renewal epoch is $1/(1-\alpha)$, and that, for a non-arithmetic process, the probability of 1 or more renewals in the interval $(t, t+\delta]$ tends to $(1-\alpha)\delta/\mathsf{E}[X] + o(\delta)$ as $t \to \infty$.

Exercise 3.36. The purpose of this exercise is to show why the interchange of expectation and sum in the proof of Wald's equality is justified when $\mathsf{E}[J] < \infty$ but not otherwise. Let $X_1, X_2, \dots,$ be a sequence of IID rv's, each with the distribution F_X. Assume that $\mathsf{E}[|X|] < \infty$.

a) Show that $S_n = X_1 + \cdots + X_n$ is a rv for each integer $n > 0$. Note: S_n is obviously a mapping from the sample space to the real numbers, but you must show that it is finite with probability 1. Hint: Recall the additivity axiom for the real numbers.

b) Let J be a stopping time for $X_1, X_2, \ldots$. Show that $S_J = X_1 + \cdots X_J$ is a rv. Hint: Represent $\Pr\{S_J\}$ as $\sum_{n=1}^{\infty} \Pr\{J = n\} S_n$.

c) For the stopping time J above, let $J^{(k)} = \min(J, k)$ be the stopping time J truncated to integer k. Explain why the interchange of sum and expectation in the proof of Wald's equality is justified in this case, so $\mathsf{E}[S_{J^{(k)}}] = \overline{X}\mathsf{E}\left[J^{(k)}\right]$.

d) In parts d), e), and f), assume, in addition to the assumptions above, that $\mathsf{F}_X(0) = 0$, *i.e.*, that the X_i are positive rv's. Show that $\lim_{k\to\infty} \mathsf{E}[S_{J^{(k)}}] < \infty$ if $\mathsf{E}[J] < \infty$ and $\lim_{k\to\infty} \mathsf{E}[S_{J^{(k)}}] = \infty$ if $\mathsf{E}[J] = \infty$.

e) Show that

$$\Pr\{S_{J^{(k)}} > x\} \leq \Pr\{S_J > x\}$$

for all k, x.

f) Show that $\mathsf{E}[S_J] = \overline{X}\mathsf{E}[J]$ if $\mathsf{E}[J] < \infty$ and $\mathsf{E}[S_J] = \infty$ if $\mathsf{E}[J] = \infty$.

g) Now assume that X has both negative and positive values with nonzero probability and let $X^+ = \max(0, X)$ and $X^- = \min(X, 0)$. Express S_J as $S_J^+ + S_J^-$ where $S_J^+ = \sum_{i=1}^{J} X_i^+$ and $S_J^- = \sum_{i=1}^{J} X_i^-$. Show that $\mathsf{E}[S_J] = \overline{X}\mathsf{E}[J]$ if $\mathsf{E}[J] < \infty$ and that $\mathsf{E}[S_j]$ is undefined otherwise.

Exercise 3.37. This is a very simple exercise designed to clarify confusion about the roles of past, present, and future in stopping rules. Let $\{X_n;\, n \geq 1\}$ be a sequence of IID binary rv's , each with the pmf $\mathsf{p}_X(1) = 1/2$, $\mathsf{p}_X(0) = 1/2$. Let J be a positive integer-valued rv that takes on the sample value n of the first trial for which $X_n = 1$. That is, for each $n \geq 1$,

$$\{J = n\} = \{X_1{=}0,\ X_2{=}0, \ldots, X_{n-1}{=}0,\ X_n{=}1\}$$

a) Use the definition of stopping time, Definition 3.1 in the text, to show that J is a stopping time for $\{X_n;\, n \geq 1\}$.

b) Show that for any given n, the rv's X_n and $\mathbb{I}_{J=n}$ are *statistically dependent*.

c) Show that for every $m > n$, X_n and $\mathbb{I}_{J=m}$ are *statistically dependent*.

d) Show that for every $m < n$, X_n and $\mathbb{I}_{J=m}$ are *statistically independent*.

e) Show that X_n and $\mathbb{I}_{J\geq n}$ are *statistically independent*. Give the simplest characterization you can of the event $\{J \geq n\}$.

f) Show that X_n and $\mathbb{I}_{J>n}$ are *statistically dependent*.

Note: The results here are characteristic of most sequences of IID rv's. For most people, this requires some realignment of intuition, since $\{J{\geq}n\}$ is the union of $\{J = m\}$ for all $m \geq n$, and all of these events are highly dependent on X_n. The right way to think of this is that $\{J{\geq}n\}$ is the complement of $\{J{<}n\}$, which is determined by $X_1, \ldots, X_{n-1}$. Thus $\{J{\geq}n\}$ is also determined by $X_1, \ldots, X_{n-1}$ and is thus independent of X_n. The moral of the story is that thinking of stopping rules as rv's independent of the future is very tricky, even in totally obvious cases such as this.

Exercise 3.38. Consider a ferry that carries cars across a river. The ferry holds an integer number k of cars and departs the dock when full. At that time, a new ferry immediately appears and begins loading newly arriving cars ad infinitum. The ferry business has been good, but customers complain about the long wait for the ferry to fill up.

a) Assume that cars arrive according to a renewal process. The IID interarrival times have mean $\overline{X}$, variance σ^2 and moment generating function $g_X(r)$. Does the sequence of departure times of the ferries form a renewal process? Explain carefully.

b) Find the expected time that a customer waits, starting from its arrival at the ferry terminal and ending at the departure of its ferry. Note 1: Part of the problem here is to give a reasonable definition of the expected customer waiting time. Note 2: It might be useful to consider $k = 1$ and $k = 2$ first.

c Is there a 'slow truck' phenomenon (a dependence on $\mathsf{E}\left[X^2\right]$) here? Give an intuitive explanation. Hint: Look at $k = 1$ and $k = 2$ again.

d) In an effort to decrease waiting, the ferry managers institute a policy where no customer ever has to wait more than one hour. Thus, the first customer to arrive after a ferry departure waits for either one hour or the time at which the ferry is full, whichever comes first, and then the ferry leaves and a new ferry starts to accumulate new customers. Does the sequence of ferry departures form a renewal process under this new system? Does the sequence of times at which each successive empty ferry is entered by its first customer form a renewal process? You can assume here that $t = 0$ is the time of the first arrival to the first ferry. Explain carefully.

e) Give an expression for the expected waiting time of the first new customer to enter an empty ferry under this new strategy.