Chapter 9

RANDOM WALKS, LARGE DEVIATIONS, AND MARTINGALES

9.1 Introduction

Definition 9.1.1. *Let* $\{X_i;\, i \geq 1\}$ *be a sequence of IID random variables (rv's), and let* $S_n = X_1 + X_2 + \cdots + X_n$*. The integer-time stochastic process* $\{S_n;\, n \geq 1\}$ *is called a random walk, or, more precisely, the one-dimensional random walk based on* $\{X_i;\, i \geq 1\}$*.*

For any given n, S_n is simply a sum of IID rv's, but here the behavior of the entire random walk *process*, $\{S_n;\, n \geq 1\}$, is of interest. Thus, for a given real number $\alpha > 0$, we might want to find the probability that the sequence $\{S_n;\, n \geq 1\}$ contains any term for which $S_n \geq \alpha$ (*i.e.*, that a threshold at α is ever crossed) or to find the distribution of the smallest n for which $S_n \geq \alpha$.

We know that S_n/n essentially tends to $\mathsf{E}[X] = \overline{X}$ as $n \to \infty$. Thus if $\overline{X} < 0$, S_n will tend to drift downward and if $\overline{X} > 0$, S_n will tend to drift upward. This means that the results to be obtained will depend critically on whether $\overline{X} < 0$, $\overline{X} > 0$, or $\overline{X} = 0$. Since results for $\overline{X} > 0$ can easily be found from results for $\overline{X} < 0$ by considering $\{-S_n;\, n \geq 1\}$, we usually focus either on the case $\overline{X} < 0$ or $\overline{X} = 0$. To avoid trivialities we shall always assume in addition that X is non-deterministic.

In the case $\overline{X} < 0$, the random walk drifts downward, with random excursions around the mean, and those random excursions might result in crossing a threshold at some positive value α. If such a positive threshold is crossed, it usually occurs before n becomes very large.

As one might expect, both the results and the techniques have a very different flavor for a zero-mean random walk, *i.e.*, a walk with $\overline{X} = 0$. In this case S_n/n essentially tends to 0 and we will see that the random walk typically wanders around in a rather aimless fashion. With increasing n, σ_{S_n} increases as $\sqrt{n}$; this behavior is often called diffusion.

We can visualize a zero-mean random walk as a zero-mean gambling game where S_n represents our winnings at time n. A common fallacy in such a game is to imagine that if we have a run of bad luck, then, since $\mathsf{E}[S_n] = 0$ for all n, a run of good luck should occur very soon to return our winnings to 0. In fact, given a loss, $S_k = s_k < 0$ at time k, then $\mathsf{E}[S_n \mid S_k = s_k] = s_k$ for all $n \geq k$. The fallacy arises because $\mathsf{E}[S_n/n \mid S_k = s_k] = s_k/n$, which approaches 0 as $n \to \infty$. One's intuition often confuses the sample average S_n/n with the sample value S_n.

The following three subsections discuss three special cases of random walks. The first two, simple random walks and integer random walks, will be useful throughout as examples, since they can be easily visualized and analyzed. The third special case is that of renewal processes, which we have already studied, and which will provide additional insight into the general study of random walks.

After this, Section 9.2 shows how a major application area, G/G/1 queues, can be viewed in terms of random walks. This section also illustrates why questions related to threshold crossings are so important in random walks.

Section 9.3 then returns to study arbitrary random walks. Many of the interesting problems concerning threshold crossings involve large deviation behavior of the underlying sums, S_n, of IID rv's. This section develops some important large deviation results about S_n. Section 9.4 then combines these results with a renewed study of stopping rules. This leads to a powerful generalization of Wald's equality known as Wald's identity; this will help us to view a random walk as an actual process rather than a collection of individual rv's.

The remainder of the chapter is devoted to a rather general type of stochastic process called martingales. The topic of martingales is both a subject of interest in its own right and also a tool that provides additional insight into random walks, laws of large numbers, and other basic topics in probability and stochastic processes.

9.1.1 Simple random walks

Suppose $X_1, X_2, \dots$ are IID binary rvs, each taking on the value 1 with probability p and -1 with probability $q = 1 - p$. Letting $S_n = X_1 + \cdots + X_n$, the sequence of sums $\{S_n;\, n \geq 1\}$, is called a *simple random walk*. Note that S_n is the difference between positive and negative occurrences in the first n trials, and thus a simple random walk is little more than a notational variation on a Bernoulli process. For the Bernoulli process, X takes on the values 1 and 0, whereas for a simple random walk X takes on the values 1 and -1. For the random walk, if $X_m = 1$ for m out of n trials, then $S_n = 2m - n$, and

$$\Pr\{S_n = 2m - n\} = \frac{n!}{m!(n-m)!}\, p^m (1-p)^{n-m}. \tag{9.1}$$

This distribution allows us to answer questions about S_n for any given n, but it is not very helpful in answering such questions as the following: for any given integer $k > 0$, what is the probability that the sequence $S_1, S_2, \dots$ ever reaches or exceeds k? This probability can be expressed as[1] $\Pr\{\bigcup_{n=1}^{\infty}\{S_n \geq k\}\}$ and is referred to as the probability that the random

[1] This same probability is sometimes expressed as $\Pr\{\sup_{n\geq 1} S_n \geq k\}$, but there is a slight difference, since the event $\{\sup_{n\geq 1} S_n \geq k\}$ can include sample sequences that approach k but never quite meet it. It

walk *crosses a threshold* at k. Exercise 9.1 demonstrates the surprisingly simple result that for a simple random walk with $p \leq 1/2$, this threshold crossing probability is

$$\Pr\left\{\bigcup_{n=1}^{\infty}\{S_n \geq k\}\right\} = \left(\frac{p}{1-p}\right)^k. \tag{9.2}$$

Sections 9.3 and 9.4 treat this same question for general random walks, but the results are less simple. They also treat questions such as the overshoot given a threshold crossing, the time at which the threshold is crossed given that it is crossed, and the probability of crossing such a positive threshold before crossing any given negative threshold.

9.1.2 Integer-valued random walks

Suppose next that $X_1, X_2, \dots$ are arbitrary IID integer-valued random variables. We can again ask for the probability that such an integer-valued random walk crosses a threshold at k, *i.e.*, that the event $\bigcup_{n=1}^{\infty}\{S_n \geq k\}$ occurs, but the question is considerably harder than for the special case of simple random walks. Since this random walk takes on only integer values, it can be represented as a Markov chain with the set of integers forming the state space. A simple random walk can also be represented as a Markov chain, but has the special property of being a birth-death chain.

For the Markov chain representation of the general integer-valued random walk, threshold crossing problems become first passage-time problems. These problems can be attacked by the Markov chain tools we already know, but the special structure of the random walk provides new approaches and simplifications that will be explained in Sections 9.3 and 9.4.

9.1.3 Renewal processes as special cases of random walks

If $X_1, X_2, \dots$ are IID positive random variables, then $\{S_n;\, n \geq 1\}$ is both a special case of a random walk and also the sequence of arrival epochs of a renewal counting process, $\{N(t);\, t > 0\}$. In this special case, the sequence $\{S_n;\, n \geq 1\}$ must eventually cross a threshold at any given positive value α. Thus the question of interest here is not *whether* a threshold is crossed but rather *when* it is crossed and with what overshoot. These are familiar questions from the study of renewal theory, where $N(\alpha)$ was effectively defined as the largest n for which $S_n \leq \alpha$ and thus $N(\alpha)+1$ is the smallest n for which $S_n > \alpha$, *i.e.*, the smallest n for which the threshold at α is strictly exceeded. When there is a positive overshoot, it is given by $S_{N(\alpha)+1} - \alpha$, and is familiar as the residual life at α.

Figure 9.1 illustrates the difference between positive random walks, *i.e.*, renewal processes, and arbitrary random walks. Note that the renewal process in part (b) is illustrated with the axes reversed from the conventional renewal process representation. We usually view each renewal epoch as a time (epoch) and view $N(\alpha)$ as the number of renewals up to time α, whereas with random walks, we usually view the number of trials as a discrete-time variable and view the sum of rv's as some kind of amplitude or cost. There is no mathematical difference between these viewpoints, and each viewpoint is often helpful.

---

is simpler to avoid this unimportant issue by not using the sup notation to refer to threshold crossings.

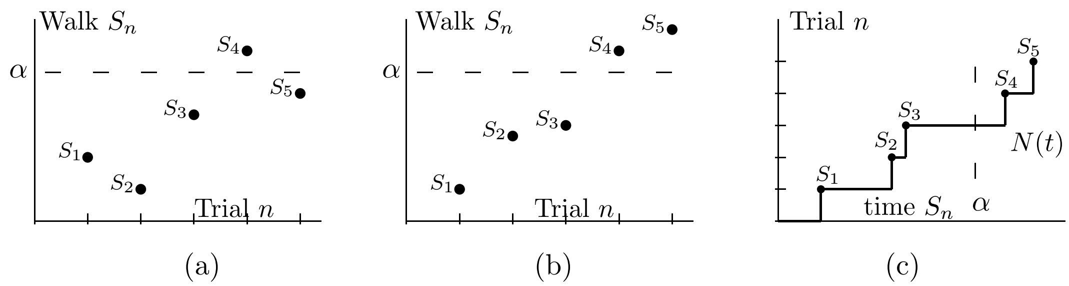


Figure 9.1: The sample function above in part (a) illustrates a random walk $S_1, S_2, \ldots$ with arbitrary (positive and negative) step sizes $\{X_i;\, i \geq 1\}$. The sample function in (b) illustrates a random walk, $S_1, S_2, \ldots$ with only positive step sizes $\{X_i > 0;\, i \geq 1\}$. Thus, $S_1, S_2, \ldots$ in (b) are renewal points in a renewal process. Note that the axes in (b) are reversed from the usual depiction of a renewal process. The usual depiction, illustrated in (c) for the same sample points, also shows the corresponding counting process. The step sizes in part (b) become the time increments between arrivals in (c). The random walks in parts a) and b) each illustrate a threshold at α, which in each case is crossed on trial 4 with an overshoot $S_4 - \alpha$.

9.2 The queueing delay in a G/G/1 queue:

Before analyzing random walks in general, we introduce an important problem area that is often best viewed in terms of random walks. This section will show how to represent the queueing delay in a G/G/1 queue as a threshold crossing problem in a random walk.

Consider a G/G/1 queue with first-come-first-serve (FCFS) service. We shall associate the probability that a customer must wait more than some given time α in the queue with the probability that a certain random walk crosses a threshold at α. Let $X_1, X_2, \ldots$ be the interarrival times of a G/G/1 queueing system; thus these variables are IID with an arbitrary CDF $\mathsf{F}_X(x) = \Pr\{X_i \leq x\}$. Assume that arrival 0 enters an empty system at time 0, and thus $S_n = X_1 + X_2 + \cdots + X_n$ is the epoch of the n^{th} arrival after time 0. Let $Y_0, Y_1, \ldots$ be the service times of the successive customers. These are independent of $\{X_i;\, i \geq 1\}$ and are IID with some given CDF $\mathsf{F}_Y(y)$. Figure 9.2 shows the arrivals and departures for an illustrative sample path of the process and illustrates the queueing delay for each arrival.

Let W_n be the queueing delay for the nth customer, $n \geq 1$. The system time for customer n is then defined as the queueing delay W_n plus the service time Y_n. As illustrated in Figure 9.2, customer $n \geq 1$ arrives X_n time units after the beginning of customer $n-1$'s system time. If $X_n < W_{n-1} + Y_{n-1}$, *i.e.*, if customer n arrives before the end of customer $n-1$'s system time, then customer n must wait in the queue until n finishes service (in the figure, for example, customer 2 arrives while customer 1 is still in the queue). Thus

$$W_n = W_{n-1} + Y_{n-1} - X_n \qquad \text{if } X_n \leq W_{n-1} + Y_{n-1}. \tag{9.3}$$

On the other hand, if $X_n > W_{n-1} + Y_{n-1}$, then customer $n-1$ (and all earlier customers) have departed when n arrives. Thus n starts service immediately and $W_n = 0$. This is the

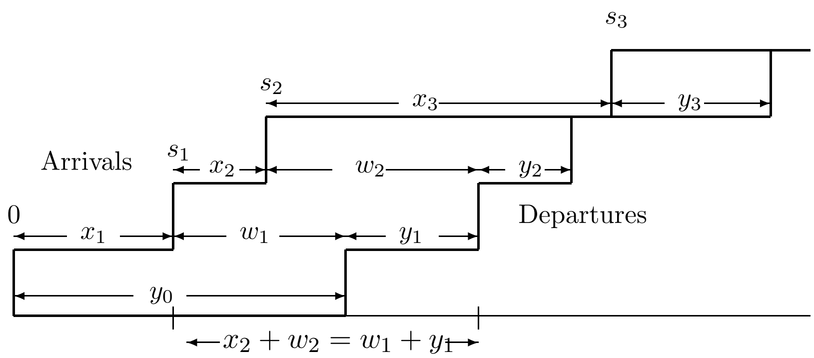


Figure 9.2: Sample path of arrivals and departures from a G/G/1 queue. Customer 0 arrives at time 0 and enters service immediately. Customer 1 arrives at time $s_1 = x_1$. For the case shown above, customer 0 has not yet departed, *i.e.*, $x_1 < y_0$, so customer 1's time in queue is $w_1 = y_0 - x_1$. As illustrated, customer 1's system time (queueing time plus service time) is $w_1 + y_1$.

Customer 2 arrives at $s_2 = x_1 + x_2$. For the case shown above, this is before customer 1 departs at $y_0 + y_1$. Thus, customer 2's wait in queue is $w_2 = y_0 + y_1 - x_1 - x_2$. As illustrated above, x_2+w_2 is also equal to customer 1's system time, so $w_2 = w_1+y_1-x_2$. Customer 3 arrives when the system is empty, so it enters service immediately with no wait in queue, *i.e.*, $w_3 = 0$.

case for customer 3 in the figure. These two cases can be combined in the single equation

$$W_n = \max[W_{n-1} + Y_{n-1} - X_n,\ 0]; \quad \text{for } n \geq 1; \quad W_0 = 0. \tag{9.4}$$

Since Y_{n-1} and X_n are coupled together in this equation for each n, it is convenient to define $U_n = Y_{n-1} - X_n$. Note that $\{U_n; n \geq 1\}$ is a sequence of IID random variables. From (9.4), $W_n = \max[W_{n-1} + U_n, 0]$, and iterating on this equation,

$$\begin{aligned} W_n &= \max[\max[W_{n-2}+U_{n-1},\ 0]+U_n,\ 0] \\ &= \max[(W_{n-2} + U_{n-1} + U_n),\ U_n,\ 0] \\ &= \max[(W_{n-3}+U_{n-2}+U_{n-1}+U_n),\ (U_{n-1}+U_n),\ U_n,\ 0] \\ &= \cdots \quad \cdots \\ &= \max[(U_1+U_2+\dots+U_n),\ (U_2+U_3+\dots+U_n), \dots, (U_{n-1}+U_n),\ U_n,\ 0]. \end{aligned} \tag{9.5}$$

If we look at the terms in the maximization in (9.5) working back from the end, we see that they can be viewed as the terms in a random walk starting at time n and working backwards. To be more specific, the first term in that backward walk is denoted as $Z_1^n = U_n$. The next term is denoted $Z_2^n = U_n + U_{n-1}$. Similarly, the ith term is

$$Z_i^n = \sum_{j=1}^{i} U_{n-j}$$

With this terminology, (9.5) becomes

$$W_n = \max[0, Z_1^n, Z_2^n, \dots, Z_n^n]. \tag{9.6}$$

Note that the terms in $\{Z_i^n;\, 1 \le i \le n\}$ are the first n terms of a random walk, but it is not the random walk based on $U_1, U_2, \dots$, but rather the random walk going backward, starting with U_n. Note also that W_{n+1}, for example, is the maximum of a different set of variables, *i.e.*, it is the walk going backward from U_{n+1}. Fortunately, this doesn't matter for the analysis since the ordered variables $(U_n, U_{n-1} \dots, U_1)$ are statistically identical to $(U_1, \dots, U_n)$. The probability that the wait is greater than or equal to a given value α is

$$\Pr\{W_n \ge \alpha\} = \Pr\{\max[0, Z_1^n, Z_2^n, \dots, Z_n^n] \ge \alpha\}. \tag{9.7}$$

This says that, for the n^{th} customer, $\Pr\{W_n \ge \alpha\}$ is equal to the probability that a random walk $\{Z_i^n;\, 1 \le i \le n\}$ based on the IID rv's U_i crosses a threshold at α by the n^{th} trial.

In the same way, $\Pr\{W_{n+1} \ge \alpha\}$ is the probability that a random walk based on the IID rv's U_i crosses a threshold at α by trial $n+1$. We see from this that

$$\cdots \le \Pr\{W_n \ge \alpha\} \le \Pr\{W_{n+1} \ge \alpha\} \le \cdots. \tag{9.8}$$

Since this sequence of probabilities is non-decreasing and bounded by 1, it must have a limit as $n \to \infty$, and this limit is denoted $\Pr\{W \ge \alpha\}$. Mathematically,[2] this limit is the probability that a random walk based on $\{U_i;\, i \ge 1\}$ ever crosses a threshold at α. Physically, this limit is the probability that the queueing delay is at least α for any given very large-numbered customer (*i.e.*, for customer n when the influence of a busy period starting n customers earlier has died out). These results are summarized in the following theorem.

Theorem 9.2.1. *Let $\{X_i;\, i \ge 1\}$ be the IID interarrival intervals of a G/G/1 queue, let $\{Y_i;\, i \ge 0\}$ be the IID service times, and assume that the system is empty at time 0 when customer 0 arrives. Let W_n be the queueing delay for the n^{th} customer. Let $U_n = Y_{n-1} - X_n$ for $n \ge 1$ and let $Z_i^n = U_n + U_{n-1} + \cdots + U_{n-i+1}$ for $1 \le i \le n$. Then for every $n \ge 1$, $W_n = \max[0, Z_1^n, Z_2^n, \dots, Z_n^n]$. Also, for each $\alpha > 0$, $\Pr\{W_n \ge \alpha\}$ is the probability that the random walk based on $\{U_i;\, i \ge 1\}$ crosses a threshold at α on or before the n^{th} trial. Finally, $\Pr\{W \ge \alpha\} = \lim_{n\to\infty} \Pr\{W_n \ge \alpha\}$ is equal to the probability that the random walk based on $\{U_i;\, i \ge 1\}$ ever crosses a threshold at α.*

It was not required in establishing the above theorem, but we can understand this maximization better by realizing that if the maximization is achieved at $U_i + U_{i+1} + \cdots + U_n$, then a busy period must start with the arrival of customer $i-1$ and continue at least through the service of customer n. To see this intuitively, note that the analysis above starts with the arrival of customer 0 to an empty system at time 0, but the choice of 0 time and customer number 0 has nothing to do with the analysis, and thus the analysis is valid for any arrival to an empty system. Choosing the largest customer number before n that starts a busy period must then give the correct queueing delay, and thus maximize (9.5). Exercise 9.2 provides further insight into this maximization.

Note that the theorem specifies the CDF of W_n for each n, but says nothing about the joint distribution of successive queueing delays. These are not the same as the distribution of successive terms in a random walk because of the reversal of terms above.

[2]More precisely, the sequence of queueing delays $W_1, W_2 \dots$ converge in distribution to W, *i.e.*, $\lim_n \mathsf{F}_{W_n}(w) = \mathsf{F}_W(w)$ for each w. We refer to W as the queueing delay in steady state.

We shall find relatively simple bounds and approximations to the probability that a random walk crosses a positive threshold in Section 9.3 and 9.4. These can be applied, via Theorem 9.2.1, to the distribution of queueing delay for the G/G/1 queue (and thus also for the M/G/1 and M/M/1 queues).

9.3 Threshold crossing probabilities in random walks

Let $\{X_i;\, i \geq 1\}$ be a sequence of IID random variables (rv's), each with the CDF $\mathsf{F}_X(x)$, and let $\{S_n;\, n \geq 1\}$ be a random walk with $S_n = X_1 + \cdots + X_n$. The major objective of this section and the next is to develop results about $\Pr\{\bigcup_{n=1}^{\infty}\{S_n \geq \alpha\}\}$, *i.e.*, the probability that the random walk $\{S_n;\; n \geq 1\}$ ever crosses a given threshold at some $\alpha > 0$. We assume throughout that $\mathsf{E}[X] < 0$ and that $\Pr\{X > 0\} > 0$. We focus on events known as large deviations where α is sufficiently large that $\Pr\{\bigcup_{n=1}^{\infty}\{S_n \geq \alpha\}\}$ is too small for the central limit theorem to provide a good approximation. In this section, we expand on the treatment of the Chernoff bound in Section 1.6.3 to understand the behavior of $\Pr\{S_n \geq \alpha\}$. Then in Section 9.4, we develop a result called the Wald identity and use it along with the Chernoff bound to further study $\Pr\{\bigcup_{n=1}^{\infty}\{S_n \geq \alpha\}\}$.

Let $I(X)$ be the interval of r over which the moment generating function (MGF) $\mathsf{g}(r) = \mathsf{E}\left[e^{rX}\right]$ is finite. As seen by example in Section 1.5.5, $I(X)$ might be finite or infinite at each end, and if finite, might be open or closed at that end. Let r_- and r_+ be the lower and upper end respectively of $I(X)$. Thus r_- and r_+ each might be finite or infinite and $g(r_-)$ and $g(r_+)$ each might be finite or infinite. We assume throughout that $r_- < 0 < r_+$, which implies that X has finite moments of all orders. Under the conditions above (including $\overline{X} < 0$) we will see that $\Pr\{\bigcup_{n=1}^{\infty}\{S_n \geq \alpha\}\}$ is bounded by an exponentially decreasing function of α for $\alpha > 0$. This will also be generalized to a bound on the probability that a threshold at $\alpha > 0$ is crossed before a threshold at some $\beta < 0$ is crossed. We will show that these bounds are exponentially tight in a sense to be specified.

9.3.1 The Chernoff bound

The Chernoff bound was derived and discussed in Section 1.6.3. It was shown in (1.64) that for any $r \geq 0$, $r \in I(X)$,

$$\Pr\{S_n \geq na\} \leq \exp\left(n[\gamma(r) - ra]\right). \tag{9.9}$$

where $\gamma(r) = \ln \mathsf{g}(r)$ is the semi-invariant MGF of X. The tightest bound of this form is found by optimizing over r, *i.e.*, by

$$\Pr\{S_n \geq na\} \leq \exp[n\mu(a)] \qquad \text{where } \mu(a) = \inf_{r\geq 0;\, r\in I(X)} \gamma(r) - ra. \tag{9.10}$$

Lemma 1.6.1 showed that $\mu(a) < 0$ for all $a > \overline{X}$. This in turn implies that $\Pr\{S_n \geq na\} \to 0$ at least exponentially in n for all $a > \overline{X}$.

Optimizating (9.9) over r is relatively straightforward, since (see Exercise 1.26) $\gamma''(r) > 0$ for $0 < r < r_+$. Figure 9.3 illustrates the optimization and gives a graphical construction for finding $\mu(a)$ in the range of a for which $a = \gamma'(r)$ for some $r \in (0, r_+)$. The optimized bound in this range can then be expressed in the parametric form

$$\Pr\{S_n \geq n\gamma'(r)\} \leq \exp[n(\gamma(r) - r\gamma'(r)] \tag{9.11}$$

Since $\gamma'(r)$ is an increasing function of r from $\overline{X}$ at $r = 0$ to $\sup_{r<r_+} \gamma'(r)$, this specifies the optimized Chernoff bound for all a in the range $\overline{X} < a < \sup_{r<r_+} \gamma'(r)$.

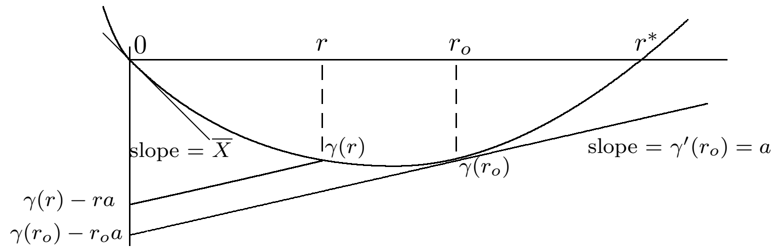


Figure 9.3: Graphical minimization of $\gamma(r) - ra$: Since $\gamma'(0) = \overline{X}$ and $\overline{X} < 0$, $\gamma(r)$ must have a negative slope at $r = 0$. Since $\Pr\{X > 0\} > 0$ by assumption also, X can not be deterministic and $\gamma''(r) > 0$ over the interval where $\gamma(r)$ exists. Except in a very special case to be discussed in Figure 9.4, $\gamma(r)$ grows without bound as $r \to r_+$. To minimize $\gamma(r) - ra$ over r, note that $\gamma(r) - ar$ is the vertical axis intercept of a line of slope a through the point $(r, \gamma(r))$. The minimum occurs when the line of slope a is tangent to the $\gamma(r)$ curve.

For many rv's, $\gamma'(r)$ increases without bound with increasing $r < r_+$, and for such rv's, (9.10) and (9.11) are equivalent. For completeness, however, we should also understand the optimization if $\sup_{r<r_+} \gamma'(r)$ is finite and we are interested in $a > \sup_{r<r_+} \gamma'(r)$. There are two types of such situations, one with $r_+ = \infty$ and the other with $r_+ < \infty$.

Exercise 9.7 shows that if $r_+ = \infty$ and $\sup_{r<\infty} \gamma'(r) < \infty$, then $\mu(a)$ as defined in (9.10) is $-\infty$ for $a > \sup_{r<r_+} \gamma'(r)$. Thus $\Pr\{S_n \geq na\} = 0$ in this case. This case occurs, for example, if X is bounded by $X \leq b$ and a exceeds b. Since $S_n \leq nb$ in this case, it is not surprising that $\Pr\{S_n \geq na\} = 0$.

Exercise 9.8 shows that if $r_+ < \infty$ and $\sup_{r<r_+} \gamma'(r) < \infty$, then $\gamma(r_+) < \infty$. It also shows examples in which this occurs. Figure 9.4 extends the graphical minimization of Figure 9.3 to this case. It is almost obvious from the figure that $\mu(a) = \gamma(r_+) - r_+ a$ in this case if $a > \sup_{r<r_+} \gamma'(r)$.

We can thus summarize the optimization of the Chernoff bound in the following theorem.

Theorem 9.3.1. *The optimized exponent $\mu(a)$ in (9.10) for arbitrary $a > \overline{X}$ in the Chernoff bound is given by*

$$\mu(a) = \begin{cases} \gamma(r) - r\gamma'(r) \quad ; & \text{for } a \in (0, \sup_{r<r_+} \gamma'(r)), \text{ and } \gamma'(r) = a \\ -\infty \quad ; & \text{for } a > \sup_{r<r_+} \gamma'(r), \text{ and } r_+ = \infty \\ \gamma(r_+) - r_+ a \quad ; & \text{for } a > \sup_{r<r_+} \gamma'(r), \text{ and } r_+ < \infty. \end{cases} \tag{9.12}$$

In principle, we could now bound the probability of threshold crossing, $\Pr\{\bigcup_{n=1}^{\infty}\{S_n \geq \alpha\}\}$, by using the union bound over n and then bounding each term by (9.12). This would be

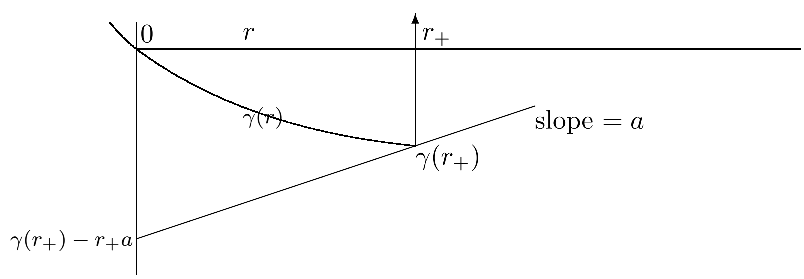


Figure 9.4: Graphical minimization of $\gamma(r) - ra$ for the case where $\gamma(r_+) < 0$. For $r \leq r_+$, $\gamma(r) - ra$ is the vertical axis intercept of a line of slope a through $(r, \gamma(r))$. If $\gamma'(r) = a$ for some $r < r_+$, the minimum occurs at that r. For $a \geq \lim_{r \to r_+} \gamma'(r)$, however (as shown in the figure) the minimum occurs at $r = r_+$.

quite tedious since, for a fixed α, the a in (9.12) such that $\alpha = an$, would vary with n. Instead, we pause to develop the concept of tilted probabilities. We will use these tilted probabilities in three ways, first to get a better understanding of the Chernoff bound, second to prove that the optimized Chernoff bound is exponentially tight in the limit of large n, and third to prove the Wald identity in the next section. The Wald identity in turn will provide an upper bound on $\Pr\{\bigcup_{n=1}^{\infty}\{S_n \geq \alpha\}\}$ that is much simpler than the above use of the union bound.

9.3.2 Tilted probabilities

As above, let $\{X_n; n \geq 1\}$ be IID and let $I(X)$ be the interval over which the MGF of X is finite. We will assume that X is discrete with the PMF $\mathsf{p}_X(x)$, but the case in which X is continuous or arbitrary is essentially handled by replacing PMF's by PDF's or CDF's. For any fixed $r \in I(X)$, define a new PMF (called a tilted PMF) on X by

$$\mathsf{q}_{X,r}(x) = \mathsf{p}_X(x) \exp[rx - \gamma(r)]. \tag{9.13}$$

Note that $\mathsf{q}_{X,r}(x) \geq 0$ for all sample values x and $\sum_x \mathsf{q}_{X,r}(x) = \sum_X \mathsf{p}_X(x) e^{rx} / \mathsf{E}[e^{rx}] = 1$, so this is a valid probability assignment.

Imagine a random walk that sums the IID rv's $X_1, X_2, \dots$, but uses this tilted probability assignment on $X_1, X_2, \dots$. We then have the same mapping from sample points of the underlying sample space to sample values of rv's, but we are dealing with a new probability measure, *i.e.*, we have changed the probability model, and thus changed the probabilities in the random walk. We define this tilted probability measure so that the tilted versions of $X_1, X_2, \dots$ are IID in this new measure.

For notational convenience, $X_1, X_2, \dots$ will denote these rv's both in the original and the tilted probability measure. The expectations and other moments will all change, so for $r \in I(X)$, r is used as a subscript to denote the moments and moment generating functions in the new probability measure. The mean of X in the new, tilted, probability measure is thus denoted as $\mathsf{E}_r[X]$. Using (9.13),

$$\begin{aligned}
\mathsf{E}_r[X] &= \sum_x x\mathsf{q}_{X,r}(x) = \sum_x x\mathsf{p}_X(x)\exp[rx-\gamma(r)] \\
&= \frac{1}{\mathsf{g}(r)}\sum_x \frac{d}{dr}\mathsf{p}_X(x)\exp[rx] \\
&= \frac{\mathsf{g}'(r)}{\mathsf{g}(r)} = \gamma'(r). \qquad (9.14)
\end{aligned}$$

Higher moments of X under the tilted measure can be calculated in the same way, but, more elegantly, the MGF of X under the tilted measure can be seen to be $\mathsf{E}_r[\exp(tX)] = \mathsf{g}(t+r)/\mathsf{g}(r)$. For the semi-invariant MGF then,

$$\gamma_r(t) = \gamma(t+r) - \gamma(r).$$

Using (9.13), using the tilted PMF on each component of an n-tuple $\boldsymbol{X}^n = (X_1, X_2, \dots, X_n)$, we get

$$\mathsf{q}_{\boldsymbol{X}^n,r}(x_1,\dots,x_n) = \mathsf{p}_{\boldsymbol{X}^n}(x_1,\dots,x_n)\exp\left(\sum_{i=1}^n [rx_i - \gamma(r)]\right). \qquad (9.15)$$

We now relate the PMF of the sum, $\sum_{i=1}^n X_i$, in the original probability measure to that in the tilted probability measure. From (9.15), note that for every n-tuple $(x_1,\dots,x_n)$ for which $\sum_{i=1}^n x_i = s_n$ (for any given s_n), we have

$$\mathsf{q}_{\boldsymbol{X}^n,r}(x_1,\dots,x_n) = \mathsf{p}_{\boldsymbol{X}^n}(x_1,\dots,x_n)\exp[rs_n - n\gamma(r)].$$

Summing over all $\boldsymbol{x}^n$ for which $\sum_{i=1}^n x_i = s_n$, we then get a remarkable relation between the PMF for S_n in the original and the tilted probability measure:

$$\mathsf{q}_{\boldsymbol{S}_n,r}(s_n) = \mathsf{p}_{S_n}(s_n)\exp[rs_n - n\gamma(r)]. \qquad (9.16)$$

This equation is a key to large deviation theory applied to sums of IID rv's. The tilted measure of S_n, for positive r, increases the probability of large values of S_n and decreases that of small values. Since S_n is an IID sum under the tilted measure, however, we can use the laws of large numbers and the CLT around the tilted mean to get good estimates and bounds on the behavior of S_n far from the mean for the original measure.

Definition 9.3.1. *Let $\{a_n;\, n \geq 1\}$ be a sequence of numbers and for each n, let b_n be a bound on a_n,* i.e., *$a_n \leq b_n$. This sequence of bounds is* exponentially tight *if the following limits exist with equality as shown.*

$$\lim_{n\to\infty}\frac{\ln a_n}{n} = \lim_{n\to\infty}\frac{\ln b_n}{n}.$$

An equivalent statement is that the sequence of bounds (usually simply called a bound) is exponentially tight if, for any $\epsilon > 0$, $a_n \geq b_n e^{-\epsilon n}$ for all sufficiently large n. In the usual situation, such as that of the Chernoff bound, the bound has the form $a_n \leq e^{-\beta n}$ for all n, and the bound is exponentially tight if β can not be enlarged without violating the bound for sufficiently large n.

It is not quite clear from the definition why exponential tightness is an important concept. For a bound such as $\Pr\{S_n \geq na\} \leq \exp[n\mu(a)]$ for example, the bound might be exponentially tight but be off by orders of magnitude for all n because of a neglected coefficient. For many, or perhaps most, applications for which a small probability of exceeding some level is desired, n is a parameter to be chosen subject to an associated cost. A tight exponent in such a bound then gives a good estimate of the rate at which $\Pr\{S_n \geq na\}$ decreases with increasing n and thus of the corresponding tradeoff between cost and probability. If the bound has the form $a_n \leq c_n e^{-\beta n}$ where c_n is, say, a polynomial in n, then the coefficient c_n, while perhaps important, is less important than the factor $e^{-\beta n}$ for sufficiently large n, since $(\ln[c_n e^{-\beta n}])/n$ approaches $-\beta$ as $n \to \infty$.

The following theorem applies the weak law of large numbers (WLLN) to $\sum_{i=1}^n X_n$ under the tilted measure to essentially show that the Chernoff bound for the original probability measure is exponentially tight.

Theorem 9.3.2. *Let $\{X_n;\, n \geq 1\}$ be an IID sequence of rv's. Let $r_- < 0$ and $r_+ > 0$ be the end points of the interval where $\gamma(r) = \ln(\mathsf{E}\left[e^{rX}\right])$ is finite. Let $S_n = \sum_{i=1}^n X_i$ for each $n \geq 1$. Then for each $r \in (0, r_+)$, and each $\epsilon > 0$, there is an n_o such that for all $n \geq n_o$,*

$$\Pr\big\{S_n \geq n\gamma'(r)\big\} \;\geq\; \exp[n(\gamma(r) - r\gamma'(r) - r\epsilon)]. \tag{9.17}$$

Discussion: Note that $\Pr\{S_n \geq n\gamma'(r)\}$ is the probability of an event in the original probability measure. If we look at that event in the tilted measure for the given r, it becomes the (tilted) probability that S_n is greater than or equal to the tilted mean. From the WLLN, S_n is with high probability close to $n\mathsf{E}_r[X] = \gamma'(r)$ in the tilted measure. The theorem simply uses epsilons carefully to convert this to the original probability measure.

Proof: The WLLN, (1.75), says that $\lim_{n\to\infty} \Pr\big\{|S_n/n - \overline{X}| > \epsilon\big\} = 0$ for each $\epsilon > 0$, or, equivalently, $\lim_{n\to\infty} \Pr\big\{|S_n/n - \overline{X}| \leq \epsilon\big\} = 1$. The meaning of the limit on n is that for any $\epsilon > 0$ and $\delta > 0$, there is an n_o large enough that

$$\Pr\left\{\left|\frac{S_n}{n} - \overline{X}\right| \leq \epsilon\right\} \geq 1 - \delta \qquad \text{for all } \ n \geq n_o. \tag{9.18}$$

The WLLN can be applied to the tilted measure as well as the original measure. In the tilted measure, (9.14) shows that the mean of X is $\gamma'(r)$. Thus, assuming that X is discrete for notional convenience, and assuming $n \geq n_o$ for the n_o required for the given ϵ and δ, we can rewrite (9.18), using the tilted measure, as

$$
\begin{aligned}
1-\delta \quad &\le \sum_{(\gamma'(r)-\epsilon)n \,\le\, s_n \,\le\, (\gamma'(r)+\epsilon)n} \mathsf{q}_{S_n,r}(s_n) && (9.19)\\
&= \sum_{(\gamma'(r)-\epsilon)n \,\le\, s_n \,\le\, (\gamma'(r)+\epsilon)n} \mathsf{p}_{S_n}(s_n)\exp[rs_n - n\gamma(r)] && (9.20)\\
&\le \sum_{(\gamma'(r)-\epsilon)n \,\le\, s_n \,\le\, (\gamma'(r)+\epsilon)n} \mathsf{p}_{S_n}(s_n)\exp[n(r\gamma'(r)+r\epsilon-\gamma(r))] && (9.21)
\end{aligned}
$$

In 9.20, $\mathsf{q}_{S_n,r}(s_n)$ is related to the original probabilities; then (9.21) upper bounds s_n in the exponent by the upper bound in the summation. The resulting sum can then be conveniently upper bounded by extending the sum to arbitrarily large s_n, yielding

$$
\begin{aligned}
1-\delta \quad &\le \sum_{(\gamma'(r)-\epsilon)n \,\le\, s_n} \mathsf{p}_{S_n}(s_n)\exp[n(r\gamma'(r)+r\epsilon-\gamma(r))] && (9.22)\\
&= \exp[n(r\gamma'(r)+r\epsilon-\gamma(r))]\ \Pr\big\{S_n \ge n(\gamma'(r)-\epsilon)\big\}\,. && (9.23)
\end{aligned}
$$

$$
\Pr\big\{S_n \ge n(\gamma'(r)-\epsilon)\big\} \ge (1-\delta)\exp[-n(r\gamma'(r)+r\epsilon-\gamma(r))], \qquad (9.24)
$$

where (9.24) rearranges (9.23). It can be seen from Figure 9.3 that this is essentially equivalent to (9.17). Exercise 9.9 shows how the ϵ on the left and the δ on the right can be absorbed into the ϵ on the right. $\square$

The structure of the above proof can also be used to derive a number of other results. The general approach is to use the tilted probability measure to focus on the tail of the distribution, then to use some known probability result on the tilted distribution, and finally to transfer the result back to the original distribution. The manipulation of epsilons and deltas above becomes automatic after a little practice.

Note that this theorem applies only to the first case in (9.12), *i.e.*, to values of $a \in (0, \sup_{r<r_+}\gamma'(r))$. For the second case in (9.12), where $r_+ = \infty$ and $a > \sup_{r<r_+}\gamma'(r))$, there is no need of tightness, since the upper bound is 0. The theorem is extended by a truncation argument in Exercise 9.10 to show that the third case in (9.12), where $r_+ < \infty$ and $a > \sup_{r<r_+}\gamma'(r)$, is also exponentially tight.

Theorems 9.3.1 and 9.3.2 are important large deviation results since they establish the true exponential decay in n for $\Pr\{S_n \ge na\}$ where S_n is a sum of n IID rv's that have an MGF. Since these results are somewhat abstract, it will be helpful to understand the result in a less abstract but less general context. This was already done for the binomial distibution in Section 1.6.3 and the following section generalizes this to the multinomial distribution, *i.e.*, to rv's that are discrete with a finite number of possible sample values.

9.3.3 Large deviations and compositions

Let $\{a_j;\, 1 \le j \le \mathsf{M}\}$ be the set of sample values for X and let $\boldsymbol{x} = (x_1, \dots, x_n)$ be an n-tuple of these sample values. The *composition* or *type* of $\boldsymbol{x}$ is defined to be the M dimensional vector $\widetilde{\boldsymbol{p}} = (\widetilde{p}_1, \widetilde{p}_2, \dots, \widetilde{p}_{\mathsf{M}})$ where, for each j, $n\widetilde{p}_j$ is the number of occurrences of a_j in $\boldsymbol{x}$. Thus the set of possible sample sequences for $\boldsymbol{x}$ is partitioned into compositions, and each composition $\widetilde{\boldsymbol{p}}$ consists of the set of n-tuples that have the given number $n\widetilde{p}_j$ of occurrences of a_j for each j. The number of n-tuples that have a given composition $\widetilde{\boldsymbol{p}} = (\widetilde{p}_1 \dots, \widetilde{p}_{\mathsf{M}})$ is then the multinomial $n!/[(\widetilde{p}_1 n)!(\widetilde{p}_1 n)! \cdots (\widetilde{p}_{\mathsf{M}} n)!]$. We assume throughout this section that $n\widetilde{p}_j > 0$ for each j; the extension to zero values for one or more $n\widetilde{p}_j$ is essentially accomplished simply by omitting those components, but requires frequent discussion of distracting special cases. The Stirling bounds can be used on the multinomial in the same way as on the binomial in Exercise 1.9 to get

$$\frac{n!}{\prod_j (n\widetilde{p}_j)!} \;<\; \frac{\sqrt{2\pi n}}{\prod_j \sqrt{2\pi n\widetilde{p}_j}} \exp\left[-n\sum_j \widetilde{p}_j \ln \widetilde{p}_j\right] \tag{9.25}$$

$$\frac{n!}{\prod_j (n\widetilde{p}_j)!} \;>\; \frac{\sqrt{2\pi n}}{\prod_j \sqrt{2\pi n\widetilde{p}_j}} \exp\left[-n\sum_j \widetilde{p}_j \ln \widetilde{p}_j\right] \exp\left[\frac{-1}{12n}\sum_j (1/\widetilde{p}_j)\right] \tag{9.26}$$

The ratio of the upper bound to the lower bound clearly approaches 1 with increasing n, uniformly in $\tilde{p}_1, \dots, \widetilde{p}_{\mathsf{M}}$ over any region where each component is bounded away from 0. The bound in (9.25) is thus asymptotically tight. The rate, $\sum_j -\widetilde{p}_j \ln \widetilde{p}_j$, at which the number of n-tuples in a composition $\widetilde{\boldsymbol{p}}$ increases with n is called the entropy of $\widetilde{\boldsymbol{p}}$. Similarly, a discrete rv with PMF $\boldsymbol{p}$ has entropy $\sum_j -p_j \ln p_j$. Entropy is a central concept in information theory; see, for example, Chapter 2 of [10] for a fuller discussion.

Now suppose that $\boldsymbol{X} = (X_1, \dots, X_n)$ is an n-tuple of discrete IID rv's with the set of possible values $\{a_1, \dots, a_{\mathsf{M}}\}$ and the PMF $\boldsymbol{p} = (p_1, \dots, p_{\mathsf{M}})$. We assume $p_j > 0$ for each j throughout. If $\boldsymbol{x}$ is a particular n-tuple of composition $\widetilde{p}_1, \dots, \widetilde{p}_{\mathsf{M}}$, then $\Pr\{\boldsymbol{X} = \boldsymbol{x}\} = p_1^{n\widetilde{p}_1} p_2^{n\widetilde{p}_2} \cdots p_{\mathsf{M}}^{\widetilde{p}_{\mathsf{M}}}$. Multiplying this by the number of n-tuples of composition $\widetilde{\boldsymbol{p}}$, we get

$$\begin{aligned} \Pr\{\boldsymbol{X} \in \widetilde{\boldsymbol{p}}\} &= \frac{n!}{\prod_j (n\widetilde{p}_j)!} \prod_j \exp(n\tilde{p}_j \ln p_j) \\ &< \frac{\sqrt{2\pi n}}{\prod_j \sqrt{2\pi n\widetilde{p}_j}} \exp\left[-n\sum_j \widetilde{p}_j \ln \frac{\widetilde{p}_j}{p_j}\right] \end{aligned} \tag{9.27}$$

This is asymptotically tight since (9.25) is asymptotically tight. The Kullback-Lieber divergence (divergence for short) between two probability vectors $\widetilde{\boldsymbol{p}}$ and $\boldsymbol{p}$ is defined to be

$$D(\widetilde{\boldsymbol{p}}||\boldsymbol{p}) = \sum_j \widetilde{p}_j \ln \frac{\widetilde{p}_j}{p_j} \tag{9.28}$$

Substituting this into (9.27)

$$\Pr\{\boldsymbol{X} \in \widetilde{\boldsymbol{p}}\} \;<\; \frac{\sqrt{2\pi n}}{\prod_j \sqrt{2\pi n\widetilde{p}_j}} \exp\left[-nD(\widetilde{\boldsymbol{p}}||\boldsymbol{p})\right] \tag{9.29}$$

As seen in Exercise 9.12, $D(\widetilde{\boldsymbol{p}}||\boldsymbol{p}) \geq 0$ with strict inequality except for $\boldsymbol{p} = \widetilde{\boldsymbol{p}}$. It is also shown that $D(\widetilde{\boldsymbol{p}}||\boldsymbol{p})$ is convex in $\widetilde{\boldsymbol{p}}$ over the region where $\widetilde{\boldsymbol{p}}$ is a probability vector.

If we temporarily choose $\boldsymbol{p}$ to equal $\widetilde{\boldsymbol{p}}$, then, since $D(\widetilde{\boldsymbol{p}}||\widetilde{\boldsymbol{p}}) = 0$, we see that the coefficient, $\frac{\sqrt{2\pi n}}{\prod_j \sqrt{2\pi n \widetilde{p}_j}}$ in (9.29) is less than or equal to 1. Thus, going back to arbitrary $\boldsymbol{p}$, we have the simpler relation,

$$\Pr\{\boldsymbol{X} \in \widetilde{\boldsymbol{p}}\} \leq \exp\left[-nD(\widetilde{\boldsymbol{p}}||\boldsymbol{p})\right] \tag{9.30}$$

Since the coefficient $\frac{\sqrt{2\pi n}}{\prod_j \sqrt{2\pi n \widetilde{p}_j}}$ is non-exponential in n and (9.29) is asymptotically tight, we see that (9.30) is exponentially tight. Note the difference between being asymptotically tight and exponentially tight; (9.29) (as seen by the upper and lower bounds in (9.25) and (9.26)) is loose only in the factor $\exp\left[\frac{-1}{12n}\sum_j (1/\widetilde{p}_j)\right]$ which approaches 1 as $n \to \infty$. On the other hand, (9.30) is also loose in the factor $\frac{\sqrt{2\pi n}}{\prod_j \sqrt{2\pi n \widetilde{p}_j}}$; this goes to 0 as a power of n, but not exponentially in n. Exponential tightness allows us to ignore terms that might be significant, but less significant asymptotically than the exponential terms.

The expontial tightness of (9.30) is probably the best way to see the intutive significance of divergence. It signifies how far $\widetilde{\boldsymbol{p}}$ is from $\boldsymbol{p}$ by giving the exponent in the probability that an n-tuple drawn according to $\boldsymbol{p}$ 'looks like' an n-tuple of composition $\widetilde{\boldsymbol{p}}$.

Next we look at $\Pr\{S_n \geq na\}$ in terms of compositions. Recall that $S_n = X_1 + \cdots X_n$ where the X_n are IID with sample values $a_1, \ldots, a_{\mathsf{M}}$ and probabilities $p_1, \ldots, p_{\mathsf{M}}$. The event $\{S_n \geq na\}$ consists of all those n-tuples belonging to compositions $\widetilde{\boldsymbol{p}}$ for which $\sum_j \widetilde{p}_j a_j \geq a$. Thus, we have

$$\begin{aligned}
\Pr\{S_n \geq na\} &= \sum_{\widetilde{\boldsymbol{p}}:\, n\widetilde{p}_j \in \mathbb{Z},\, \sum_j \widetilde{p}_j a_j \geq a} \Pr\{\boldsymbol{X} \in \widetilde{\boldsymbol{p}}\} && (9.31)\\
&\leq \sum_{\widetilde{\boldsymbol{p}}:\, n\widetilde{p}_j \in \mathbb{Z},\, \sum_j \widetilde{p}_j a_j \geq a} \exp[-nD(\widetilde{\boldsymbol{p}}||\boldsymbol{p})] && (9.32)\\
&\leq n^{\mathsf{M}+1} \sup_{\widetilde{\boldsymbol{p}}:\, \sum_j \widetilde{p}_j a_j \geq a} \exp[-nD(\widetilde{\boldsymbol{p}}||\boldsymbol{p})] && (9.33)
\end{aligned}$$

In the final expression, the number of compositions in the sum has been upper bounded by the number of all compositions of n-tuples, which in turn is bounded by recognizing that $n\widetilde{p}_j$ is an integer between 0 and M for each j. Each term in the sum has then been upper bounded by the supremum without an integer constraint on $n\widetilde{p}_j$. This result, and its extension to constraints other that $\sum \widetilde{p}_j a_j \leq a$ is called Sanov's theorem.

The supremum in (9.33) could be calculated by minimizing $D(\widetilde{\boldsymbol{p}}||\boldsymbol{p})$ over $\widetilde{\boldsymbol{p}}$ using Lagrange multipliers for the constraints $\sum_j \widetilde{p}_j = 1$ and $\sum_j \widetilde{p}_j a_j \geq a$. The following more elegant approach to this minimization uses tilted probabilities, tilted from $\boldsymbol{p}$ to $\boldsymbol{q}_r$ where r satisfies $\gamma'(r) = a$. Thus $q_{j,r} = p_j \exp[ra_j - \gamma(r)]$ and the mean of S_n, according to the tilted distribution, is na. The motivation for this is that we used these tilted probabilities earlier to evaluate $\Pr\{S_n \geq na\}$ and the tilting emphasized the probabilities of sample n-tuples close to the threshold $s_n \approx na$. Here we do the same thing in terms of compositions rather than individual n-tuples. Writing $\boldsymbol{p}$ in terms of $\boldsymbol{q}_r$, we have

$$\begin{aligned}
D(\widetilde{\boldsymbol{p}}||\boldsymbol{p}) &= \sum_j \widetilde{p}_j \ln\left[\frac{\widetilde{p}_j}{q_{j,r}e^{-ra_j+\gamma(r)}}\right] \\
&= -\gamma(r) + \sum_j \widetilde{p}_j r a_j + \sum_j \widetilde{p}_j \ln\left[\frac{\widetilde{p}_j}{q_{j,r}}\right] \\
&= -\gamma(r) + \sum\nolimits_j \widetilde{p}_j r a_j + D(\widetilde{\boldsymbol{p}}||\boldsymbol{q}_r) \\
&\geq -\gamma(r) + ra \qquad \text{for } \widetilde{\boldsymbol{p}} \text{ such that } \sum\nolimits_j \widetilde{p}_j a_j \geq a.
\end{aligned} \tag{9.34}$$

The final expression is valid since the divergence is nonnegative (see Exercise 9.12). Substituting (9.34) into (9.33) for each term over which the sup is taken, we surprisingly have

$$\Pr\{S_n \geq na\} \leq n^{\mathsf{M}+1} \exp\big(n[\gamma(r) - ra]\big); \qquad \gamma'(r) = a$$

This is the Chernoff bound, except for the additional factor of $n^{\mathsf{M}+1}$, so it seems we have not accomplished much beyond deriving an old result in a less general and somewhat weaker form. What is new is that we can now see that (9.33) is satisfied with equality for $\widetilde{\boldsymbol{p}} = \boldsymbol{q}_r$ and thus the supremum in (9.33) is achieved by $\widetilde{\boldsymbol{p}} = \boldsymbol{q}_r$. In other words, subject to the integer constraint on $\boldsymbol{q}_r n$ the composition most likely to give rise to $S_n \geq na$ is the tilted probability $\boldsymbol{q}_r$ using the r for which $\sum_j q_{j,r} a_j = a$.

To summarize the results of this section, we started by showing that the divergence $D(\widetilde{\boldsymbol{p}}||\boldsymbol{p})$ has a fundamental interpretation as $\lim_n (-1/n) \ln \Pr\{\boldsymbol{X} \in \widetilde{\boldsymbol{p}}\}$ where $\boldsymbol{X}$ is an n-r**v** drawn according to $\boldsymbol{p}$, *i.e.*, D is the exponential rate of decrease in the probability that $\boldsymbol{p}$ 'looks like' $\widetilde{\boldsymbol{p}}$. As indicated by the notation, $D(\widetilde{\boldsymbol{p}}||\boldsymbol{p})$ depends only on $\widetilde{\boldsymbol{p}}$ and $\boldsymbol{p}$ and not on the values $\boldsymbol{a} = (a_1, \dots, a_{\mathsf{M}})$ taken on by X. As indicated by (9.31)-(9.33), $\Pr\{S_n \geq na\}$ is then the probability of the union of compositions bounded by the linear constraint $\boldsymbol{a}^{\mathsf{T}}\widetilde{\boldsymbol{p}} \geq a$. The dominant such compositions are those close to $\boldsymbol{q}_r$. An important use of divergences is to bound or estimate more general unions of compositions.

9.3.4 Back to threshold crossings

Consider again the probability that a random walk crosses a positive threshold α, *i.e.*, $\Pr\{\bigcup_n \{S_n \geq \alpha\}\}$. Assume that $\overline{X} < 0$ and that the end points r_- and r_+ of the interval where $\mathsf{g}(r)$ exists satisfy $r_- < 0 < r_+$. We can use the optimized Chernoff bound to look at each event in this union. For values of n such that $\alpha/n = \gamma'(r)$ for some r in $(0, r_+)$, the first part of (9.12) says that

$$\Pr\{S_n \geq \alpha\} \leq \exp -n[\gamma(r) - r\gamma'(r)] \qquad \text{for } r \text{ such that } \gamma'(r) = \alpha/n.$$

Since $\gamma'(r) = \alpha/n$ in this equation, we can replace n in the exponent with $\alpha/\gamma'(r)$, getting

$$\Pr\{S_n \geq \alpha\} \leq \exp\left\{\alpha\left[\frac{\gamma(r)}{\gamma'(r)} - r\right]\right\} \qquad \text{where } \gamma'(r) = \alpha/n. \tag{9.35}$$

We have seen that $\gamma'(r)$ is continuous and increasing in r for $r \in (0, r_+)$, and we assume temporarily that $\lim_{r\to r_+} \gamma'(r) = \infty$ so that this equation has a solution for $\gamma'(r) = \alpha/n$ for each $\alpha > 0$ and $n \geq 1$.

A nice graphic interpretation of this equation is given in Figure 9.5. Note that the exponent in α, namely $[\gamma(r)/\gamma'(r)] - r$, is the negative of the horizontal axis intercept of the tangent of slope $\gamma'(r) = \alpha/n$ to the curve $\gamma(r)$ in Figure 9.5.

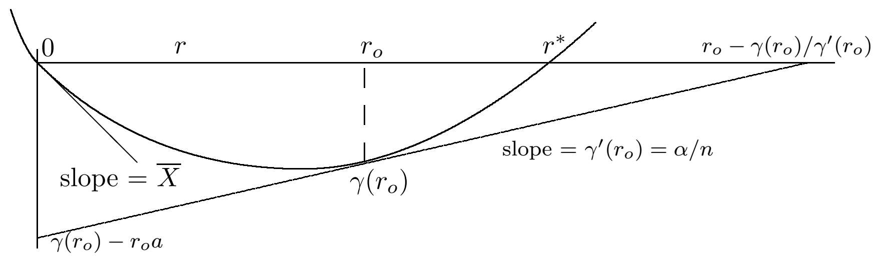


Figure 9.5: The exponent in α for $\Pr\{S_n \geq \alpha\}$, minimized over r. The minimization is the same as that in Figure 9.3, but $\gamma(r_o)/\gamma'(r_o) - r_o$ is the negative of the horizontal axis intercept of the line tangent to $\gamma(r_o)$ at $r_o = r$.

For fixed $\alpha > 0$, (still assuming that $\gamma'(r)$ is unbounded), we see that the slope α/n is small for large n and this horizontal intercept is large. As n is decreased, the slope increases, the point of tangency, r_o, increases, and the horizontal intercept decreases. When r_o increases to the point labeled r^* in the figure, namely the $r > 0$ at which $\gamma(r) = 0$, then the horizontal intercept also reaches r^*. When n decreases even further, r_o becomes larger than r^*, $\gamma(r_0) > 0$, and the horizontal intercept starts to increase again.

Since the horizontal intercept for each n exceeds r^*, we see that (9.35) can be replaced with the looser and simpler upper bound,

$$\Pr\{S_n \geq \alpha\} \leq \exp(-r^*\alpha) \qquad \text{for arbitrary } \alpha > 0, \text{ all } n \geq 1, \tag{9.36}$$

where r^* is the positive root of $\gamma(r)$.

This bound is quite loose for many values of n, but since the bound in (9.35) is exponentially tight, the bound in (9.36) is correspondingly tight for $n \approx \alpha/\gamma'(r^*)$. Our primary interest here is in the probability of threshold crossing, $\Pr\{\bigcup_n\{S_n \geq \alpha\}\}$. The union bound shows that this is upper bounded by a sequence of terms each upper bounded by $e^{-r^*\alpha}$. At the same time, $\Pr\{\bigcup_n\{S_n \geq \alpha\}\}$ is lower bounded by $\Pr\{S_n \geq \alpha\}$ for each n. For $n \approx \alpha/\gamma'(r^*)$, the exponent of this lower bound is essentially $-r^*\alpha$.

We thus conclude, without going through all the epsilons and deltas, that

$$\lim_{\alpha\to\infty} \frac{1}{\alpha} \ln \Pr\left\{\bigcup_n \{S_n \geq \alpha\}\right\} = -r^*. \tag{9.37}$$

We now recall that (9.36) and (9.37) were derived under the assumption that $\sup_{r>0} \gamma'(r) = \infty$. If $\gamma'(r)$ is bounded, we must consider the 2nd and 3rd alternatives in (9.12). The second alternative consists of cases where $r_+ = \infty$ and shows that $\Pr\{S_n \geq \alpha\} = 0$ for $\alpha/n > \sup_r \gamma'(r)$. For these cases, (9.36) is still valid. These cases occur when X is bounded by some b, and the threshold α cannot be reached until n exceeds α/b. Since $\gamma(r)$ is unbounded in this case, the positive root, r^* of $\gamma(r)$ still exists, and there are values of $n \approx \alpha/\gamma'(r^*)$. Thus (9.37) again holds in this case.

The third alternative in (9.12) is viewed graphically in Figure 9.4. In this case, if we redefine r^* as the supremum of values of $r > 0$ such that $\gamma(r) < 0$, we again have (9.36). In this case, the horizontal intercept in Figure 9.4 lies at $r^* - n\gamma(r^*)/\alpha$. We can establish (9.37) in this case by choosing n/α to be arbitrarily small and then increasing n and α together for exponential tightness.

The following theorem summarizes these results.

Theorem 9.3.3. *Let $\{X_n;\, n \geq 1\}$ be IID and let $\{S_n;\, n \geq 1\}$ where $S_n = X_1 + \cdots X_n$. Let 0 be in the interior of the interval where $\gamma(r) = \ln \mathsf{E}\left[e^{rX}\right]$ exists and let $r^* = \sup\{r > 0 : \gamma(r) < 0\}$. Then (9.37) holds and (9.36) is valid for all $n \geq 1$ and $\alpha > 0$.*

In the next section, we develop Wald's identity, which allows us to show (under almost the same conditions) that the probability of threshold crossing, namely $\Pr\{\bigcup_{n=1}^{\infty}\{S_n \geq \alpha\}\}$ is upper bounded by $\exp(-r^*\alpha)$. This is a much simpler result than (9.36), since it bounds the entire union of events over n using the same bound as used in (9.36) for each term. Thus the main purpose of Theorem 9.3.3 is that it establishes the exponential tightness of the bound.

The other purpose of this section has been to provide information on *when* a threshold is crossed as well as *whether* it is crossed. These bounds are also useful in a wide variety of situations involving large deviations other than those regarding threshold crossings.

9.4 Thresholds, stopping rules, and Wald's identity

In this section, we focus on random walks with both a positive and negative threshold and ask questions about which threshold is crossed first and when that first threshold crossing occurs. Let J be the time at which the first of the two thresholds is crossed. To be explicit, Lemma 9.4.1 demonstrates the almost obvious fact that if X is not deterministically 0, then a threshold must be crossed eventually with probability 1. In other words, we show that J is a rv rather than a possibly-defective rv.

Figure 9.6 illustrates two sample paths and how they cross thresholds, say at $\alpha > 0$ and $\beta < 0$. More specifically, the random walk first crosses a threshold at trial n if $\beta < S_i < \alpha$ for $1 \leq i < n$ and either $S_n \geq \alpha$ or $S_n \leq \beta$. For now we make no assumptions about the mean or MGF of each X_i.

Lemma 9.4.1. *Let $\{X_i;\, i \geq 1\}$ be IID rv's, not identically 0. For each $n \geq 1$, let $S_n = X_1 + \cdots + X_n$. Let $\alpha > 0$ and $\beta < 0$ be arbitrary, and let J be the smallest n for which either $S_n \geq \alpha$ or $S_n \leq \beta$. Then J is a random variable (*i.e., $\lim_{m\to\infty} \Pr\{J \geq m\} = 0$*) and $r = 0$ is in the interior of the interval where the MGF of J, $\mathsf{E}\left[e^{rJ}\right]$ exists.*

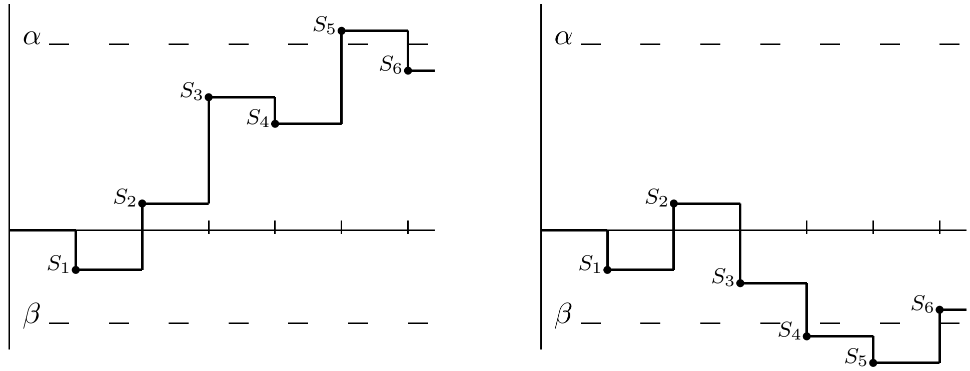


Figure 9.6: Two sample paths of a random walk with two thresholds. In the first, the threshold at α is crossed at $J = 5$. In the second, the threshold at β is crossed at $J = 4$

Proof: Since X is not identically 0, there is some n for which either $\Pr\{S_n \leq -\alpha + \beta\} > 0$ or for which $\Pr\{S_n \geq \alpha - \beta\} > 0$. For any such n, define ϵ by

$$\epsilon = \max[\Pr\{S_n \leq -\alpha + \beta\}\,,\ \Pr\{S_n \geq \alpha - \beta\}].$$

For any integer $k \geq 1$, given that $J > n(k-1)$, and given any value of $S_{n(k-1)}$ in (β, α), a threshold will be crossed by time nk with probability at least ϵ. Thus,

$$\Pr\{J > nk \mid J > n(k-1)\} \leq 1 - \epsilon.$$

Iterating on k,

$$\Pr\{J > nk\} \leq (1-\epsilon)^k.$$

This shows that J is finite with probability 1 and that $\Pr\{J \geq j\}$ goes to 0 at least geometrically in j. It follows that the moment generating function $\mathsf{g}_J(r)$ of J is finite in a region around $r = 0$, and that J has moments of all orders. □

The event $J = n$ (*i.e.*, the event that a threshold is first crossed at trial n), is a function of $X_1, X_2, \ldots, X_n$. Thus, in the notation of Section 5.5, J is a stopping trial for the sequence $\{X_n;\, n \geq 1\}$. In the following subsection, we derive Wald's identity for the two threshold problem at hand. As will be seen, Wald's identity is closely related to Wald's equality. The Wald identity will be derived under more general conditions in Section 9.8.1.

9.4.1 Wald's identity for two thresholds

Theorem 9.4.1 (Wald's identity for 2 thresholds). *Let $\{X_i;\, i \geq 1\}$ be IID and let $\gamma(r) = \ln\{\mathsf{E}\left[e^{rX}\right]\}$. Let $I(X)$ be the interval of r over which $\gamma(r)$ exists. For each $n \geq 1$, let $S_n = X_1 + \cdots + X_n$. Let $\alpha > 0$ and $\beta < 0$ be arbitrary, and let J be the smallest n for which either $S_n \geq \alpha$ or $S_n \leq \beta$. Then for each $r \in I(X)$,*

$$\mathsf{E}\left[\exp(rS_J - J\gamma(r))\right] = 1. \tag{9.38}$$

Discussion: The following proof uses the tilted probability distributions of Section 9.3.2. The theorem and proof are valid without the assumptions that $\mathsf{E}[X] < 0$ and $0 \in I(X)$.

Proof: Assume that each X_i is discrete with the common PMF $\mathsf{p}_X(x)$. For the non-discrete case, the PMF's and their sums can be replaced by CDF's and Stieltjes integrals, thus complicating the technical details but not introducing any new ideas.

For any given $r \in I(X)$, we use the tilted PMF $\mathsf{q}_{X,r}(x)$ given in (9.13) as

$$\mathsf{q}_{X,r}(x) = \mathsf{p}_X(x)\exp[rx - \gamma(r)].$$

Using this tilted measure for each X_i in the n-tuple $\boldsymbol{X}^n = (X_1, X_2, \dots, X_n)$, and taking the X_i to be independent in the tilted probability measure, we have

$$\begin{aligned} \mathsf{q}_{\boldsymbol{X}^n,r}(\boldsymbol{x}^n) &= \prod_{i=1}^n \mathsf{q}_{X_i,r}(x_i) = \prod_{i=1}^n \mathsf{p}_{X_i}(x_i)\exp[rx_i - \gamma(r)] \\ &= \mathsf{p}_{\boldsymbol{X}^n}(\boldsymbol{x}^n)\exp[rs_n - n\gamma(r)] \qquad \text{where } s_n = \sum_{i=1}^n x_i. \end{aligned}$$

Now let $\mathcal{T}_n$ be the set of n-tuples $X_1, \dots, X_n$ such that $\beta < S_i < \alpha$ for $1 \le i < n$ and either $S_n \ge \alpha$ or $S_n \le \beta$. That is, $\mathcal{T}_n$ is the set of $\boldsymbol{x}^n$ for which the stopping trial J has the sample value n. The PMF for the stopping trial J in the tilted measure is then

$$\begin{aligned} \mathsf{q}_{J,r}(n) &= \sum_{\boldsymbol{x}^n \in \mathcal{T}_n} \mathsf{q}_{\boldsymbol{X}^n,r}(\boldsymbol{x}^n) = \sum_{\boldsymbol{x}^n \in \mathcal{T}_n} \mathsf{p}_{\boldsymbol{X}^n}(\boldsymbol{x}^n)\exp[rs_n - n\gamma(r)] \\ &= \mathsf{E}\left[\exp[rS_n - n\gamma(r) \mid J{=}n\right] \Pr\{J = n\}. \end{aligned} \tag{9.39}$$

where the expectation refers to the original measure and the summation over $\mathcal{T}$ corresponds to $J = n$. Lemma 9.4.1 applies to the tilted PMF on this random walk as well as to the original PMF, and thus the sum of $\mathsf{q}_{J,r}(n)$ over n is 1. Summing the expression on the right over n yields $\mathsf{E}[\exp(rS_J - J\gamma(r))]$, completing the proof. □

After understanding the details of this proof, one sees that it is essentially just the statement that J is a non-defective stopping rule in the tilted probability space.

We next give a number of examples of how Wald's identity can be used.

9.4.2 The relationship of Wald's identity to Wald's equality

The trial J at which a threshold is crossed in Wald's identity is a stopping trial in the terminology of Chapter 5. For r in the interior of $I(X)$, the derivative of (9.38) with respect to r is

$$\mathsf{E}\left[(S_J - J\gamma'(r))\exp\{rS_J - J\gamma(r)\}\right] = 0.$$

Assuming that 0 is in the interior of $I(X)$, we can set $r = 0$. Recalling that $\gamma(0) = 0$ and $\gamma'(0) = \overline{X}$, this becomes Wald's equality as established in Theorem 5.5.1,

$$\mathsf{E}[S_J] = \mathsf{E}[J]\,\overline{X}. \tag{9.40}$$

This is somewhat less general than Wald's equality as stated in Theorem 5.5.1, since we assume here that 0 is in the interior of $I(X)$, and we also assume two thresholds (which automatically satisfies the constraint in Wald's equality that $\mathsf{E}[J] < \infty$).

9.4.3 Zero-mean random walks

Wald's equality provides no information about $\mathsf{E}[J]$ when $\overline{X} = 0$, but Wald's identity does provide some useful information since the second derivative of (9.38) with respect to r is

$$\mathsf{E}\left[\left(S_J - J\gamma'(r)\right)^2 - J\gamma''(r)\right] \exp\{rS_J - J\gamma(r)\} = 0.$$

At $r = 0$, this is

$$\mathsf{E}\left[S_J^2 - 2JS_J\overline{X} + J^2\overline{X}^2 + J\,\sigma_X^2\right] = 0. \tag{9.41}$$

This equation is often difficult to use because of the cross term between S_J and J, but its main application comes in the case where $\overline{X} = 0$. In this case, (9.41) simplifies to

$$\mathsf{E}\left[S_J^2\right] = \sigma_X^2\,\mathsf{E}[J]\,. \tag{9.42}$$

Example 9.4.1 (The simple random walk with zero mean). Consider the simple random walk of Section 9.1.1 with $\Pr\{X{=}1\} = \Pr\{X{=}-1\} = 1/2$, and assume that $\alpha > 0$ and $\beta < 0$ are integers. Since S_n takes on only integer values and changes only by ± 1, it takes on the value α or β before exceeding either of these values. Thus S_J is either α or β. Let q_α denote $\Pr\{S_J = \alpha\}$. The expected value of S_J is then $\alpha q_\alpha + \beta(1 - q_\alpha)$. From Wald's equality, $\mathsf{E}[S_J] = 0$, so

$$q_\alpha = \frac{-\beta}{\alpha - \beta}; \qquad 1 - q_\alpha = \frac{\alpha}{\alpha - \beta}. \tag{9.43}$$

Using (9.42) with $\sigma_X^2 = 1$ and evaluating $\mathsf{E}\left[S_J^2\right]$,

$$\mathsf{E}[J] = \mathsf{E}\left[S_J^2\right] = \alpha^2 q_\alpha + \beta^2(1 - q_\alpha). \tag{9.44}$$

Substituting (9.43) into this and simplifying,

$$\mathsf{E}[J] = -\beta\alpha. \tag{9.45}$$

As a sanity check, note that if α and β are each multiplied by some large constant k, then $\mathsf{E}[J]$ increases by k^2. Since $\sigma_{S_n}^2 = n$, we would expect S_n to fluctuate with increasing n, with typical values growing as $\sqrt{n}$, and thus it is reasonable that the expected time to reach a threshold increases with the product of the distances to the thresholds.

We also notice that if β is decreased toward $-\infty$, while holding α constant, then $q_\alpha \to 1$ and $\mathsf{E}[J] \to \infty$. This helps explain Example 5.5.2 where one plays a coin tossing game, stopping when finally ahead. This shows that if the coin tosser has a finite capital β, *i.e.*, stops either on crossing a positive threshold at 1 or a negative threshold at $-\beta$, then the coin tosser wins a small amount with high probability and loses a large amount with small probability.

For more general random walks with $\overline{X} = 0$, there is usually an overshoot when the threshold is crossed. If the magnitudes of α and β are large relative to the range of X, however, it is often reasonable to ignore the overshoots. Repeating the analysis of the simple random walk as an approximation, and including the value of σ_X^2, we get the approximation $\mathsf{E}[J] \approx -\beta\alpha/\sigma_X^2$.

9.4.4 Exponential bounds on the probability of threshold crossing

We next apply Wald's identity to complete the analysis of crossing a threshold at $\alpha > 0$ when $\overline{X} < 0$.

Corollary 9.4.1. *Under the conditions of Theorem 9.4.1, assume that $\overline{X} < 0$ and that $r^* > 0$ exists such that $\gamma(r^*) = 0$. Then*

$$\Pr\{S_J \geq \alpha\} \leq \exp(-r^*\alpha). \tag{9.46}$$

Proof: Wald's identity, with $r = r^*$, reduces to $\mathsf{E}[\exp(r^*S_J)] = 1$. We can express this as

$$\Pr\{S_J \geq \alpha\}\,\mathsf{E}[\exp(r^*S_J) \mid S_J \geq \alpha] + \Pr\{S_J \leq \beta\}\,\mathsf{E}[\exp(r^*S_J) \mid S_J \leq \beta] = 1. \tag{9.47}$$

Since the second term on the left is nonnegative,

$$\Pr\{S_J \geq \alpha\}\,\mathsf{E}[\exp(r^*S_J) \mid S_J \geq \alpha] \leq 1. \tag{9.48}$$

Given that $S_J \geq \alpha$, we see that $\exp(r^*S_J) \geq \exp(r^*\alpha)$. Thus

$$\Pr\{S_J \geq \alpha\}\exp(r^*\alpha) \leq 1, \tag{9.49}$$

which is equivalent to (9.46). □

This bound is valid for all $\beta < 0$ and thus it is clear intuitively that it is also valid in the absence of a lower threshold. When there is no lower threshold, however, the stopping rule becomes defective and the proof of Theorem 9.4.1 no longer holds. Exercise 9.14 verifies that (9.49) is still valid in the absence of a lower threshold.

We see from this that the case of a single threshold is little more than a special case of the two threshold problem, but as seen in the zero-mean simple random walk, having a second threshold is often valuable in further understanding the single threshold case.

Corollary 9.4.1 is also valid in the special case illustrated in Figure 9.4, where $\gamma(r) < 0$ for all $r \in (0, r_+]$ and we have defined r^* to be r_+. This is shown in Exercise 9.15.

The Chernoff bound in (9.36) shows that $\Pr\{S_n \geq \alpha\} \leq \exp(-r^*\alpha)$ for each n; (9.46) is typically much tighter, since it shows that $\exp(-r^*\alpha)$ is an upper bound on the probability of the union of these terms. As discussed earlier, however, the Chernoff bound also shows that the result is exponentially tight and it provides some intuition about the result.

When Corollary 9.4.1 is applied to the G/G/1 queue in Theorem 9.2.1, (9.46) is referred to as the *Kingman Bound*.

Corollary 9.4.2 (Kingman Bound). *Let $\{X_i;\, i \geq 1\}$ and $\{Y_i;\, i \geq 0\}$ be the interarrival intervals and service times of a G/G/1 queue that is empty at time 0 when customer 0 arrives. Let $\{U_i = Y_{i-1} - X_i;\, i \geq 1\}$, and let $\gamma(r) = \ln\{\mathsf{E}\left[e^{rU}\right]\}$ be the semi-invariant moment generating function of each U_i. Assume that $\gamma(r)$ has a root at $r^* > 0$. Then W_n, the queueing delay of the nth arrival, and W, the steady state queueing delay, satisfy*

$$\Pr\{W_n \geq \alpha\} \leq \Pr\{W \geq \alpha\} \leq \exp(-r^*\alpha) \quad ; \qquad \textit{for all } \alpha > 0. \tag{9.50}$$

For a random walk with $\overline{X} > 0$, the exceptional circumstance is $\Pr\{S_J \leq \beta\}$. This can be analyzed by changing the sign of X and β and using the results for a negative expected value. These exponential bounds are not valid for $\overline{X} = 0$, and we will not analyze that case here other than the result in (9.42).

Note that the simple bound on the probability of crossing the upper threshold in (9.46) (and thus also the Kingman bound) is an upper bound (rather than an equality) because, first, the effect of the lower threshold was eliminated (see (9.48)), and, second, the overshoot was bounded by 0 (see (9.49)). The effect of the second threshold can be taken into account by recognizing that $\Pr\{S_J \leq \beta\} = 1 - \Pr\{S_J \geq \alpha\}$. Then (9.47) can be solved, getting

$$\Pr\{S_J \geq \alpha\} = \frac{1 \,-\, \mathsf{E}\left[\exp(r^*S_J) \mid S_J{\leq}\beta\right]}{\mathsf{E}\left[\exp(r^*S_J) \mid S_J{\geq}\alpha\right] \,-\, \mathsf{E}\left[\exp(r^*S_J) \mid S_J{\leq}\beta\right]}. \tag{9.51}$$

Solving for the terms on the right side of (9.51) usually requires analyzing the overshoot upon crossing a barrier, and this is often difficult (See Chapter 12 of [9]), for example). For the case of the simple random walk, overshoots don't occur, since the random walk changes only in unit steps. Thus, for α and β integers, we have $\mathsf{E}\left[\exp(r^*S_J) \mid S_J{\leq}\beta\right] = \exp(r^*\beta)$ and $\mathsf{E}\left[\exp(r^*S_J) \mid S_J{\geq}\alpha\right] = \exp(r^*\alpha)$. Substituting this in (9.51) yields the exact solution for the simple random walk.

$$\Pr\{S_J \geq \alpha\} = \frac{\exp(-r^*\alpha)[1 - \exp(r^*\beta)]}{1 - \exp[-r^*(\alpha - \beta)]}. \tag{9.52}$$

Solving the equation $\gamma(r^*) = 0$ for the simple random walk with probabilities p and q yields $r^* = \ln(q/p)$. This is also valid if X takes on the three values -1, 0, and $+1$ with $p = \Pr\{X = 1\}$, $q = \Pr\{X = -1\}$, and $1 - p - q = \Pr\{X = 0\}$. It can be seen that if α and $-\beta$ are large positive integers, then the simple bound of (9.46) is almost exact for this example.

Equation (9.52) is sometimes used as an approximation for (9.51) for general random walks. Unfortunately, for many applications, the overshoots are more significant than the effect of the opposite threshold. Thus (9.52) is only negligibly better than (9.46) as an approximation, and has the further disadvantage of not being a bound.

9.5 Binary hypotheses with IID observations

The objective of this section is to understand how to make to make a decision between two hypotheses, $X = 0$ or $X = 1$ on the basis of a sequence of observations $Y_1, Y_2, \dots$, with the property that $Y_1, Y_2, \dots$, are IID conditional on $X = 0$ and also IID conditional on $X = 1$. In Subsection 9.5.1 we will analyze the large deviation aspects of binary detection with a large but fixed number of observations. In Subsection 9.5.2 follows this with an analysis of binary detection when the number of observations is variable, and in particular when the observer can choose when to make a decision based on the observations already made. We will see that this choice often reduces to a problem of threshold crossing in a random walk.

9.5.1 Binary hypotheses with a fixed number of observations

Consider the binary hypothesis testing problem of Section 8.2 in which X is a binary hypothesis with *a priori* probabilities p_0 and p_1. The observation $Y_1, Y_2, \dots$ conditional on $X = 0$, is a sequence of IID rv's, each with the probability density $\mathsf{f}_{Y|X}(y \mid 0)$. Conditional on $X = 1$, the observations are IID, each with density $\mathsf{f}_{Y|X}(y \mid 1)$. For any given number n of sample observations, $y_1, \dots, y_n$, the likelihood ratio is

$$\Lambda_n(\boldsymbol{y}) = \prod_{i=1}^{n} \frac{\mathsf{f}_{Y|X}(y_i \mid 1)}{\mathsf{f}_{Y|X}(y_i \mid 0)}.$$

The log-likelihood ratio, $s_n(\boldsymbol{y}) = \ln \Lambda(\boldsymbol{y})$ is then

$$s_n = \sum_{i=1}^{n} z_i; \qquad \text{where } z_i = \ln \frac{\mathsf{f}_{Y|X}(y_i \mid 1)}{\mathsf{f}_{Y|X}(y_i \mid 0)}. \tag{9.53}$$

The MAP test gives the maximum *a posteriori* probability of correct decision based on the n observations, $y_1, \dots, y_n$. Specifically, it is the following threshold test, where the threshold η is given by $\eta = p_0/p_1$:

$$s_n \begin{cases} \geq \ln \eta \quad ; & \text{select } \hat{x}{=}1 \\ < \ln \eta \quad ; & \text{select } \hat{x}{=}0. \end{cases} \tag{9.54}$$

The Chernoff bound can be used to provide an exponentially tight bound on the probability of error, given $X = 0$, resulting from a threshold test. That is, conditional on $X = 0$, S_n is a sum of n IID rv's $Z_1, \dots, Z_n$ whose sample values are given by (9.53) and whose PDF (conditional on $X = 0$) is determined by $\mathsf{f}_{Y|X}(y|0)$. Given $X = 0$, an error is made using the threshold test in (9.54) if $S_n \geq \eta$, *i.e.*,

$$\Pr\{e_\eta \mid X = 0\} = \Pr\{S_n \geq \ln \eta \mid X{=}0\} \tag{9.55}$$

To upper bound this by the Chernoff bound, we use the semi-invariant MGF $\gamma_0(r)$ of Z given $X = 0$.

$$\begin{aligned} \gamma_0(r) &= \ln \int_y \mathsf{f}_{Y|X}(y \mid 0) \exp\left\{ r \left[\ln \frac{\mathsf{f}_{Y|X}(y \mid 1)}{\mathsf{f}_{Y|X}(y \mid 0)} \right] \right\} dy \\ &= \ln \int_y [\mathsf{f}_{Y|X}(y \mid 0)]^{1-r} [\mathsf{f}_{Y|X}(y \mid 1)]^r \, dy. \end{aligned} \tag{9.56}$$

The optimized Chernoff bound (conditional on $X = 0$ is then

$$\Pr\{e_\eta \mid X{=}0\} \le \exp\left\{ n \min_{r\ge 0}[\gamma_0(r) - ra] \right\} \qquad \text{where } a = \frac{1}{n}\ln \eta. \tag{9.57}$$

We can find the Chernoff bound for $\Pr\{e_\eta \mid X{=}1\} = \Pr\{S_n < \eta \mid X = 1\}$ in the same way. The rv $Z = \ln\big(f(y|1)/f(y|0)\big)$ conditional on $X = 1$ has a PDF arising from $f(y|1)$, so the semi-invariant MGF for Z conditional on $X = 1$ is given by

$$\gamma_1(r) = \ln \int_y [\mathsf{f}_{Y|X}(y \mid 0)]^{-r}[\mathsf{f}_{Y|X}(y \mid 1)]^{1+r}\, dy. \tag{9.58}$$

Since we are bounding the lower tail of S_n conditional on $X = 1$, the Chernoff bound is now optimized over $r \le 0$, *i.e.*, the optimized Chernoff bound (conditional on $X = 1$) is

$$\Pr\{e_\eta \mid X{=}1\} \le \exp\left\{ n \min_{r\le 0}[\gamma_1(r) - ra] \right\} \qquad \text{where } a = \frac{1}{n}\ln \eta. \tag{9.59}$$

If we now compare $\gamma_0(r)$ from (9.56) with $\gamma_1(r)$ from (9.58), we find, surprisingly, that $\gamma_1(r) = \gamma_0(r+1)$. If we substitute this into (9.59) and then change $r+1$ throughout to r, this becomes

$$\Pr\{e_\eta \mid X{=}1\} \le \exp\left\{ n \min_{r\le 1}[\gamma_0(r) + (1{-}r)a] \right\} \qquad \text{where } a = \frac{1}{n}\ln \eta. \tag{9.60}$$

If the minimization in (9.59) (for a given threshold η) occurs for $r \in (0,1)$, then it can be seen by comparison of (9.59) and (9.60) that the minimization of (9.60) occurs at the same value of r. Figure 9.7 illustrates these optimizations of (9.57) and (9.60) together for a common threshold η.

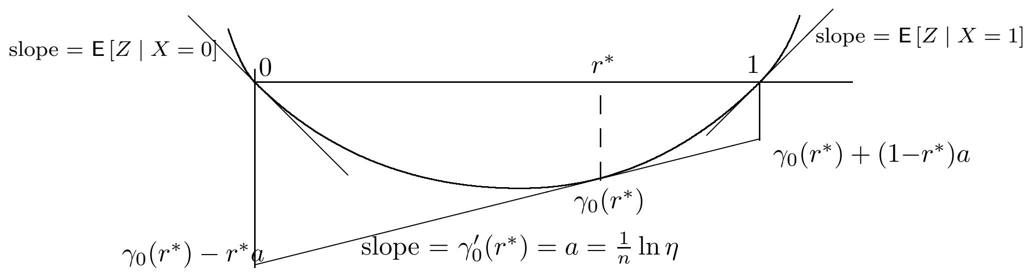


Figure 9.7: Graphical description of the optimization over r for (9.57) and (9.60). Note that since $\gamma_1(r) = \gamma_0(r+1)$ and $\gamma_1(0) = 0$, we must have $\gamma_0(1) = 0$; this can also be seen directly from (9.56). Also, since $\gamma''(r) > 0$ for $0 < r < 1$, we see that $\gamma'(0) < 0$ and $\gamma'(1) > 0$. Thus $\mathsf{E}[Z|X = 0] < 0$ and the optimization for (9.57) is the same as that in Figure 9.3. Similarly, $\mathsf{E}[Z|X{=}1] > 0$ and the optimization for (9.60) uses the same principal. Assuming that $\gamma_0'(r) = a$ for some $r^* \in (0,1)$, both optimizations occur at that same value of r^*.

Recall from Sections (8.3) and (8.4) that binary threshold tests are optimal not only for MAP detection but also min-cost detection and (subject to a tie-breaking randomization) for the Neyman-Pearson rule.

As illustrated in Figure 9.7, the slope $\gamma_0'(r) = \frac{1}{n}\ln\eta$ at which the optimized bound occurs increases with η, varying from $\gamma'(0) = \mathsf{E}[Z \mid X{=}1]$ at $r = 0$ to $\gamma'(1) = \mathsf{E}[Z \mid X{=}0]$ at $r = 1$. The tradeoff between the two error exponents is seen to vary as the two ends of an inverted see-saw. One could in principle achieve a still larger magnitude of exponent for $\Pr\{e_\eta \mid X{=}0\}$ by using $r > 1$, but this would be somewhat silly since $\Pr\{e_\eta \mid X{=}1\}$ would then be very close to 1 and it would usually make more sense to simply decide on $X = 0$ without looking at the observed data at all.

We can view the tradeoff of exponents above as a large deviation form of the Neyman-Pearson rule. The Neyman-Pearson rule is specified for a given number n of observations, and thus does not ask how the rule should be modified for different n. Let $e(n)$ denote the error event as a function of n and let us fix the allowable exponent to the error probability $\Pr\{e(n) \mid X{=}0\}$ to be exponential in n and then choose a test to minimize the exponent for $\Pr\{e(n) \mid X{=}1\}$. Since these bounds are exponentially tight, this gives us an appropriate large deviation result, focusing on behavior for large n. We can summarize these results in the following theorem.

Theorem 9.5.1. *Consider two hypotheses, $X = 0$ and $X = 1$. Conditional on each hypothesis, the observation is an IID n-r*v *$\boldsymbol{Y} = (Y_1, \dots, Y_n)$ with conditional density $\mathsf{f}_{Y|X}(y|\ell)$ for $\ell = 0, 1$. Assume that 0 and 1 are in the interior of the interval where $\gamma_0(r) = \ln \int \mathsf{f}^{1-r}_{Y|X}(y|0)\mathsf{f}^{r}_{Y|X}(y|1)\,dy$ exists. Then for all integer $n \geq 1$ and all $r \in (0,1)$, the following bounds can be met simultaneously by using a threshold test at $\eta = e^{n\gamma_0'(r)}$.*

$$\Pr\{e(n) \mid X{=}0\} \;\leq\; \exp\left\{n[\gamma_0(r) - r\gamma_0'(r)]\right\} \tag{9.61}$$

$$\Pr\{e(n) \mid X{=}1\} \;\leq\; \exp\left\{n[\gamma_0(r) + (1-r)\gamma_0'(r)]\right\} \tag{9.62}$$

Furthermore, these bounds are exponentially tight in the sense that for a given $r \in (0,1)$ and any given $\epsilon > 0$, there is an n_o large enough so that for all $n > n_o$ and all tests, either

$$\Pr\{e(n) \mid X{=}0\} \;\geq\; \exp\left\{n[\gamma_0(r) - r\gamma_0'(r) - \epsilon]\right\} \qquad \text{or} \tag{9.63}$$

$$\Pr\{e(n) \mid X{=}1\} \;\geq\; \exp\left\{n[\gamma_0(r) + (1-r)\gamma_0'(r) - \epsilon]\right\} \tag{9.64}$$

Proof: Note that (9.61) and (9.62) are simply the result of the minimization over r in (9.57) and (9.60). These are Chernoff bounds, and from Theorem (9.3.2) are exponentially tight for fixed r, thus satisfying both (9.63) and (9.64) using threshold tests at $\eta = e^{n\gamma_0'(r)}$. Finally, since the tests here are threshold tests, and the threshold test for each n lies on the error curve, Theorem 8.4.1 implies that one or the other must be valid for arbitrary tests. □

Maximum likelihood (ML) tests provide an important special case of the theorem. We can regard ML as a threshold test with $\eta = 1$ (and *a priori* probabilities can be viewed as either equal or unknown). With $\eta = 1$, $\gamma_0'(r) = 0$ at the optimal r. The following error probabilities are then exponentially tight in the sense of the theorem.

$$\Pr\{e(n) \mid X = 0\} \leq \exp\left[n \min_r \gamma_0(r)\right]; \qquad \Pr\{e(n) \mid X = 1\} \leq \exp\left[n \min_r \gamma_0(r)\right] \tag{9.65}$$

Another special case is the MAP test with given *a priori* probabilities, p_0, p_1. The threshold $\eta = p_0/p_1$ does not vary with increasing n. Theorem 9.5.1 is clearly applicable, but $\gamma_0'(r) = (1/n)\ln(\eta)$ and approaches 0 with increasing n.

In the limit as $n \to \infty$, the exponent $\gamma_0(r) - r\gamma_0'(r)$ approaches $\min_r \gamma_0(r)$ for any fixed η. Thus the simpler ML bound of (9.65) can be used in this case for large n. What is happening here is that the effect of η becomes unimportant in the exponent for very large n. In effect, there is so much information from the observations that the *a priori* probabilities are unimportant.

The opposite extreme from the ML case in Theorem 9.5.1 is that in which an error under one hypothesis, say $X = 1$, is so much more serious than an error under $X = 0$ that we want to maximize the exponential rate at which $\Pr\{e(n) \mid X{=}1\}$ approaches 0 with increasing n subject only to the constraint that $\Pr\{e(n) \mid X{=}0\}$ approaches 0, perhaps subexponentially, as $n \to \infty$.

The solution to this extreme case is known as Stein's lemma and is almost obvious from the geometry of Figure 9.7. By letting the point of tangency in the figure approach 0 with increasing n, the exponent for $\Pr\{e(n) \mid X{=}1\}$ approaches $\mathsf{E}[Z \mid X{=}0]$ and the exponent for $\Pr\{e(n) \mid X = 0\}$ approaches 0. Stein's lemma is usually stated in terms of the divergence $D\big(\mathsf{f}(y|0)\|\mathsf{f}(y|1)\big)$ (see (9.28)), but as seen in the following equation, $\mathsf{E}[Z \mid X{=}0]$ is the negative of this divergence.

$$\mathsf{E}[Z \mid X{=}0] \;=\; \int \mathsf{f}(y|0)\ln\frac{\mathsf{f}(y|1)}{\mathsf{f}(y(0)}\,dy \;=\; -D(\mathsf{f}(y|0)||\mathsf{f}(y|1)) \tag{9.66}$$

Corollary 9.5.1 (Stein's lemma). *For the conditions of Theorem 9.5.1, there is a sequence of tests for $n \geq 1$ for which*

$$\lim_{n\to\infty}(1/n)\ln\Pr\{e(n) \mid X{=}1\} \;=\; -D(\mathsf{f}(y|0)||\mathsf{f}(y|1)) \qquad \text{and} \tag{9.67}$$

$$\lim_{n\to\infty}\Pr\{e \mid X = 0\} \;=\; 0 \tag{9.68}$$

Given (9.68), the result in (9.67) is exponentially tight.

A guided proof, showing how to take the limit as $r \to 0$ and $n \to \infty$ is given in Exercise 9.20.

Divergence has a somewhat broader interpretation here than as the exponent of a composition in Section 9.3.3. In both cases, it is the exponent for the probability that a sequence from one distribution 'looks like' a sequence from another distribution, but in the composition case, it's the exponent for the probability of an exact match, whereas here it's the exponent for the probability of a set of sequences that is large enough to have a probability close to 1 for the opposite hypothesis.

9.5.2 Sequential decisions for binary hypotheses

Common sense tells us that it should be valuable to make additional observations when the current observations do not lead to a clear choice. With such a possibility, there are 3 possible choices at the end of each observation: choose $\hat{x} = 1$, choose $\hat{x} = 0$, or continue with additional observations. We will analyze such a setup by viewing it as a threshold crossing problem. We establish two thresholds, $\alpha > 0$ and $\beta < 0$. We then look at a pair of random walks. The first random walk is the sequence of LLR's given that $X = 0$ and the second is the sequence of LLR's given that $X = 1$.

The observer sees the successive values of the random walk, but does not know whether it is the random walk conditional on $X = 0$ or on $X = 1$. If the random walk first crosses threshold α, however, the decision $\hat{x} = 1$ is made. Conversely, if the threshold β is first crossed, the decision $\hat{x} = 0$ is made.

Let us first analyze the random walk conditional on $X = 0$. The random walk is then the sequence of LLR's $\{S_n; n \geq 1\}$ where $S_n = \sum_{i=1}^{n} Z_i$ and $Z_i = \ln\big(\mathsf{f}(y_i|1)/\mathsf{f}(y_i|0)\big)$ where y_i has the PDF $\mathsf{f}_{Y|X}(y_i|0)$. The stopping rule is to stop when the random walk crosses either a threshold[3] at $\alpha > 0$ or a threshold at $\beta < 0$

Since the decision $\hat{x} = 1$ is made if $S_J \geq \alpha$ and $\hat{x} = 0$ is made if $S_J \leq \beta$. we see that for the random walk conditional on $X = 0$, an error is made if $S_J \geq \alpha$. We denote the error event as $e(J)$. Using the probability distribution for $X = 0$, we apply (9.46), along with $r^* = 1$, to get

$$\Pr\{e(J) \mid X{=}0\} \;=\; \Pr\{S_J \geq \alpha \mid X{=}0\} \;\leq\; \exp(-\alpha). \tag{9.69}$$

Given $X = 1$, $S_n = \sum_{i=1}^{n} Z_i$ is also a random walk, but the probability measure is different. That is $Z_i = \ln\big(\mathsf{f}(y_i|1)/\mathsf{f}(y_i|0)\big)$ where y_i now has the PDF $\mathsf{f}_{Y|X}(y_i|1)$. The same stopping rule must be used, since the decision to stop at n can be based only on $Z_1, \dots, Z_n$ and not on knowledge of X.

For the random walk conditional on $X = 1$, an error is made if the threshold at β is crossed before that at α. Either by recognizing the fundamental symmetry between $X = 0$ and $X = 1$ or by repeating the analysis leading to (9.46) for a lower bound rather than an upper bound, we get

$$\Pr\{e(J) \mid X{=}1\} \;=\; \Pr\{S_J \leq \beta \mid X{=}1\} \;\leq\; \exp(\beta). \tag{9.70}$$

It is rather surprising that these bounds do not depend at all on the likelihoods involved in the decision and that they are so simple. It turns out that the distribution of J conditional on $X = 0$ and $X = 1$ does involve the likelihoods. Also the exact values of $\Pr\{S_J \leq \beta \mid X{=}0\}$ and $\Pr\{S_J \leq \beta \mid X{=}1\}$ depend on the likelihoods.

The error probabilities can be made as small as desired by increasing the magnitudes of α and β, but there is a cost involved in increasing these magnitudes. The cost in increasing α is essentially to increase the number of observations required when $X = 1$. From Wald's equality,

$$\mathsf{E}[J \mid X{=}1] \;=\; \frac{\mathsf{E}[S_J \mid X{=}1]}{\mathsf{E}[Z \mid X{=}1]} \;\approx\; \frac{\alpha + \mathsf{E}[\text{overshoot} \mid S_J \geq \alpha]}{\mathsf{E}[Z \mid X{=}1]}.$$

[3]It is unfortunate that the word 'threshold' has a universally accepted meaning for random walks (*i.e.*, the meaning we are using here), and the word 'threshold test' has a universally accepted meaning for hypothesis testing. Threshold tests for hypothesis testing, as used in in the previous section and in Chapter 8, are non-sequential. The test here, however, is sequential, at each epoch making a ternary choice between continuing or stopping with $H = 0$, or stopping with $H = 1$.

In the approximation, we have ignored the possibility of S_J crossing the threshold at α conditional on $X = 1$ since this is a very small-probability event when α and β have large magnitudes. Thus we see that the expected number of observations (given $X = 1$) is essentially linear in $|\beta|$. Similarly,

$$\mathsf{E}[J \mid X{=}0] \;=\; \frac{\mathsf{E}[S_J \mid X{=}0]}{\mathsf{E}[Z \mid X{=}0]} \;\approx\; \frac{\beta + \mathsf{E}[\text{overshoot} \mid S_J \le \beta]}{\mathsf{E}[Z \mid X{=}0]}.$$

This gives us another interpretation of divergence, since we recall that $\mathsf{E}[Z \mid X{=}1] = D(\mathsf{f}(y|1)\|\mathsf{f}(y|0))$ and $\mathsf{E}[Z \mid X{=}0] = -D(\mathsf{f}(y|0)\|\mathsf{f}(y|1))$. More specifically, the ratio of the expected number of observations when $X = 1$ to the exponent of error probability when $X = 0$ is $D(\mathsf{f}(y|1)\|\mathsf{f}(y|0))$. Similarly, the ratio of the expected number of observations when $X = 0$ to the exponent of error probability when $X = 1$ is $D(\mathsf{f}(y|0)\|\mathsf{f}(y|1))$. In other words, the tradeoff between one error probability and the opposite number of observations is given by the corresponding divergence. Thus the two divergences are the most significant quantities involved in sequential decisions.

We next ask what has been gained quantitatively by using the sequential decision procedure here. Suppose we compare the sequential procedure to a fixed-length test with $n = \alpha/\mathsf{E}[Z \mid X{=}1]$. Referring to Figure 9.7, we see that if we choose the slope $a = \gamma'(1) = \mathsf{E}[Z \mid X{=}1]$, then the (exponentially tight) Chernoff bound on $\Pr\{e \mid X{=}0\}$ is given by $e^{-\alpha}$, but the exponent on $\Pr\{e \mid X{=}1\}$ is 0. In other words, by using a sequential test as described here, we simultaneously get the error exponent for $X = 0$ that a fixed test would provide if we gave up entirely on an error exponent for $X = 1$, and vice-versa.[4]

A final question to be asked is whether any substantial improvement on this sequential decision procedure would result from letting the thresholds at α and β vary with the number of observations. Assuming that we are concerned only with the expected number of observations, the answer is no. We will not carry this argument out here, but it consists of using the Chernoff bound as a function of the number of observations. This shows that there is a typical number of observations at which most errors occur, and changes in the thresholds elsewhere can increase the error probability, but not substantially decrease it.

9.6 Martingales

A martingale is defined as an integer-time stochastic process $\{Z_n;\, n \ge 1\}$ with the properties that $\mathsf{E}[|Z_n|] < \infty$ for all $n \ge 1$ and

$$\mathsf{E}[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1] = Z_{n-1}; \qquad \text{for all } n \ge 2. \tag{9.71}$$

The name martingale comes from gambling terminology where martingales refer to gambling strategies in which the amount to be bet is determined by the past history of winning or losing. If one visualizes Z_n as representing the gambler's fortune at the end of the n^{th}

[4]In the communication context, decision rules are used to detect sequentially transmitted data. The use of a sequential decision rule usually requires feedback from receiver to transmitter, and also requires a variable rate of transmission. Thus the substantial reductions in error probability are accompanied by substantial system complexity.

play, the definition above means, first, that the game is fair (in the sense that the expected increase in fortune from play $n-1$ to n is zero), and, second, that the expected fortune on the n^{th} play depends on the past only through the fortune on play $n-1$.

The important part of the definition of a martingale, and what distinguishes martingales from other kinds of processes, is the form of dependence in (9.71). However, the restriction that $\mathsf{E}[|Z_n|] < \infty$ is also important, particularly since martingales are so abstract and general that one often loses the insight to understand intuitively when this restriction is important. Students are advised to ignore this restriction when first looking at something that might be a martingale, and to check later after acquiring some understanding.

There are two interpretations of (9.71); the first and most straightforward is to view it as shorthand for $\mathsf{E}[Z_n \mid Z_{n-1}{=}z_{n-1},\ Z_{n-2}{=}z_{n-2},\ \dots,\ Z_1{=}z_1] = z_{n-1}$ for all possible sample values $z_1, z_2, \dots, z_{n-1}$. The second is that $\mathsf{E}[Z_n \mid Z_{n-1}{=}z_{n-1}, \dots, Z_1{=}z_1]$ is a function of the sample values $z_1, \dots, z_{n-1}$ and thus $\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1]$ is a random variable which is a function of the random variables $Z_1, \dots, Z_{n-1}$ (and, for a martingale, a function only of Z_{n-1}). Students are encouraged to take the first viewpoint initially and to write out the expanded type of expression in cases of confusion. The second viewpoint, however, is very powerful, and, with experience, is the more useful viewpoint.

It is important to understand the difference between martingales and Markov chains. For the Markov chain $\{X_n;\, n \geq 1\}$, each rv X_n is conditioned on the past only through X_{n-1}, whereas for the martingale $\{Z_n;\, n \geq 1\}$, it is only the expected value of Z_n that is conditioned on the past only through Z_{n-1}. The rv Z_n itself, conditioned on Z_{n-1}, can also be dependent on all the earlier Z_i's. It is very surprising that so many results can be developed using such a weak form of conditioning.

In what follows, we give a number of important examples of martingales, then develop some results about martingales, and then discuss those results in the context of the examples.

9.6.1 Simple examples of martingales

Example 9.6.1 (Random walks). One example of a martingale is a zero-mean random walk, since if $Z_n = X_1 + X_2 + \cdots + X_n$, where the X_i are IID and zero mean, then

$$\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] = \mathsf{E}[X_n + Z_{n-1} \mid Z_{n-1}, \dots, Z_1] \tag{9.72}$$

$$= \mathsf{E}[X_n] + Z_{n-1} = Z_{n-1}. \tag{9.73}$$

Extending this example, suppose that $\{X_i;\, i \geq 1\}$ is an arbitrary sequence of IID random variables with mean $\overline{X}$ and let $\widetilde{X}_i = X_i - \overline{X}$. Then $\{S_n;\, n \geq 1\}$ is a random walk with $S_n = X_1 + \cdots + X_n$ and $\{Z_n;\, n \geq 1\}$ is a martingale with $Z_n = \widetilde{X}_1 + \cdots + \widetilde{X}_n$. The random walk and the martingale are simply related by $Z_n = S_n - n\overline{X}$, and thus general results about martingales can easily be applied to arbitrary random walks.

Example 9.6.2 (Sums of dependent zero-mean rv's). Let $\{X_i;\, i \geq 1\}$ be a sequence of dependent random variables satisfying $\mathsf{E}[X_i \mid X_{i-1}, \dots, X_1] = 0.$ for all $i > 0$. Then

$\{Z_n;\, n \geq 1\}$, where $Z_n = X_1 + \cdots + X_n$, is a zero-mean martingale. This can be seen by induction as follows:

$$\begin{aligned}
\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] &= \mathsf{E}[X_n + Z_{n-1} \mid Z_{n-1}, \dots, Z_1] \\
&= \mathsf{E}[X_n \mid X_{n-1}, \dots, X_1] + \mathsf{E}[Z_{n-1} \mid Z_{n-1}, \dots, Z_1] = Z_{n-1}.
\end{aligned}$$

This is a more general example than it appears, since given any martingale $\{Z_n;\, n \geq 1\}$, we can define $X_n = Z_n - Z_{n-1}$ for $n \geq 2$ with $X_1 = Z_1$. Then $\mathsf{E}[X_n \mid X_{n-1}, \dots, X_1] = 0$ for $n \geq 2$. If the martingale is zero mean (*i.e.*, if $\mathsf{E}[Z_1] = 0$), then $\mathsf{E}[X_1] = 0$ also. This means that results for zero-mean martingales can be translated into results about dependent conditionally zero-mean rv's and vice-versa.

Example 9.6.3 (Product-form martingales). Another example is a product of unit mean IID random variables. Thus if $Z_n = X_1 X_2 \dots X_n$, we have

$$\begin{aligned}
\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] &= \mathsf{E}[X_n Z_{n-1} \mid Z_{n-1}, \dots, Z_1] \\
&= \mathsf{E}[X_n]\,\mathsf{E}[Z_{n-1} \mid Z_{n-1}, \dots, Z_1] \qquad (9.74) \\
&= \mathsf{E}[X_n]\,\mathsf{E}[Z_{n-1} \mid Z_{n-1}] = Z_{n-1}. \qquad (9.75)
\end{aligned}$$

A particularly simple case of this product example is where $X_n = 2$ with probability 1/2 and $X_n = 0$ with probability 1/2. Then for each $n \geq 1$,

$$\Pr\{Z_n = 2^n\} = 2^{-n}; \qquad \Pr\{Z_n = 0\} = 1 - 2^{-n}; \qquad \mathsf{E}[Z_n] = 1. \qquad (9.76)$$

Thus $\lim_{n\to\infty} Z_n = 0$ with probability 1, but $\mathsf{E}[Z_n] = 1$ for all n and $\lim_{n\to\infty} \mathsf{E}[Z_n] = 1$. This is an important example to keep in mind when trying to understand why proofs about martingales are necessary and non-trivial. This type of phenomenon will be clarified somewhat by Lemma 9.8.1 when we discuss stopped martingales in Section 9.8.

An important example of a product-form martingale is as follows: let $\{X_i;\, i \geq 1\}$ be an IID sequence, and let $\{S_n = X_1 + \cdots + X_n;\, n \geq 1\}$ be a random walk. Assume that the semi-invariant MGF $\gamma(r) = \ln\{\mathsf{E}[\exp(rX)]\}$ exists for some given r. For each $n \geq 1$, let Z_n be defined as

$$\begin{aligned}
Z_n &= \exp\{rS_n - n\gamma(r)\} \qquad (9.77) \\
&= \exp\{rX_n - \gamma(r)\}\exp\{rS_{n-1} - (n-1)\gamma(r)\} \\
&= \exp\{rX_n - \gamma(r)\}Z_{n-1}. \qquad (9.78)
\end{aligned}$$

Taking the conditional expectation of this,

$$\begin{aligned}
\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] &= \mathsf{E}[\exp(rX_n - \gamma(r))]\;\mathsf{E}[Z_{n-1} \mid Z_{n-1}, \dots, Z_1] \\
&= Z_{n-1}. \qquad (9.79)
\end{aligned}$$

where we have used the fact that $\mathsf{E}[\exp(rX_n)] = \exp(\gamma(r))$. Thus we see that $\{Z_n;\ n \geq 1\}$ is a martingale of the product-form.

9.6.2 Scaled branching processes

A final simple example of a martingale is a "scaled down" version of a branching process $\{X_n;\, n \geq 0\}$. Recall from Section 6.7 that, for each n, X_n is defined as the aggregate number of elements in generation n. Each element i of generation n, $1 \leq i \leq X_n$ has a number of offspring $Y_{i,n}$ which collectively constitute generation $n+1$, *i.e.*, $X_{n+1} = \sum_{i=1}^{X_n} Y_{i,n}$. The rv's $Y_{i,n}$ are IID over both i and n.

Let $\overline{Y} = \mathsf{E}[Y_{i,n}]$ be the mean number of offspring of each element of the population. Then $\mathsf{E}[X_n \mid X_{n-1}] = \overline{Y} X_{n-1}$, which resembles a martingale except for the factor of $\overline{Y}$. We can convert this branching process into a martingale by scaling it. That is, define $Z_n = X_n/\overline{Y}^n$. It follows that

$$\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] = \mathsf{E}\left[\frac{X_n}{\overline{Y}^n} \mid X_{n-1}, \dots, X_1\right] = \frac{\overline{Y} X_{n-1}}{\overline{Y}^n} = Z_{n-1}. \tag{9.80}$$

Thus $\{Z_n;\, n \geq 1\}$ is a martingale. We will see the surprising result later that this implies that Z_n converges with probability 1 to a limiting rv as $n \to \infty$.

9.6.3 Partial isolation of past and future in martingales

Recall that for a Markov chain, the states at all times greater than a given n are independent of the states at all times less than n, conditional on the state at time n. The following lemma shows that at least a small part of this independence of past and future applies to martingales.

Lemma 9.6.1. *Let $\{Z_n;\, n \geq 1\}$ be a martingale. Then for any $n > i \geq 1$,*

$$\mathsf{E}[Z_n \mid Z_i, Z_{i-1}, \dots, Z_1] = Z_i. \tag{9.81}$$

Proof: For $n = i+1$, $\mathsf{E}[Z_{i+1} \mid Z_i, \dots, Z_1] = Z_i$ by the definition of a martingale. Similarly, for $n = i+2$,

$$\mathsf{E}[Z_{i+2} \mid Z_{i+1}, \dots, Z_1] = Z_{i+1}.$$

The expectation on the left is a rv that is a function of $Z_1, \dots, Z_{i+1}$, and that function, according to the equation, is Z_{i+1}. If we take the expectation of this (*i.e.*, of Z_{i+1}) conditional on $Z_i, \dots, Z_1$, we get

$$\mathsf{E}[Z_{i+2} | Z_i, \dots, Z_1] \;\; = \;\; \mathsf{E}[Z_{i+1} \mid Z_i, \dots, Z_1] \;\; = \;\; Z_i. \tag{9.82}$$

For $n = i+3$, (9.82), with i incremented, shows us that the rv $\mathsf{E}[Z_{i+3} \mid Z_{i+1}, \dots, Z_1]$ (which is in general a function of $Z_1, \dots, Z_{i+1}$) is equal to Z_{i+1}. Taking the conditional expectation of this rv over Z_{i+1} conditional on $Z_i, \dots, Z_1$, we get

$$\mathsf{E}[Z_{i+3} \mid Z_i, \dots, Z_1] = Z_i.$$

This argument can be applied successively to any $n > i$. $\square$

This lemma is particularly important for $i = 1$, where it says that $\mathsf{E}[Z_n \mid Z_1] = Z_1$. The left side of this is a rv which is a function (in fact the identity function) of Z_1. Thus, by taking the expected value of each side, we see that

$$\mathsf{E}[Z_n] = \mathsf{E}[Z_1] \quad \text{for all } n > 1. \tag{9.83}$$

9.7 Submartingales and supermartingales

Submartingales and supermartingales are simple generalizations of martingales that provide many useful results for very little additional work. We will subsequently derive the Kolmogorov submartingale inequality, which is a powerful generalization of the Markov inequality. We use this both to give a simple proof of the strong law of large numbers and also to better understand threshold crossing problems for random walks.

Definition 9.7.1. *A submartingale is an integer-time stochastic process* $\{Z_n;\, n \geq 1\}$ *that satisfies the relations*

$$\mathsf{E}[|Z_n|] < \infty \quad ; \quad \mathsf{E}[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1] \geq Z_{n-1} \quad ; \quad n \geq 1. \tag{9.84}$$

A supermartingale is an integer-time stochastic process $\{Z_n;\, n \geq 1\}$ *that satisfies the relations*

$$\mathsf{E}[|Z_n|] < \infty \quad ; \quad \mathsf{E}[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1] \leq Z_{n-1} \quad ; \quad n \geq 1. \tag{9.85}$$

In terms of our gambling analogy, a submartingale corresponds to a game that is at least fair, *i.e.*, where the expected fortune of the gambler either increases or remains the same. A *supermartingale* is a process with the opposite type of inequality. The subscripts *sub* and *super* are the opposites of what common sense would dictate, but the notation is too standard to change.

Since a martingale satisfies both (9.84) and (9.85) with equality, a martingale is both a submartingale and a supermartingale. Note that if $\{Z_n;\, n \geq 1\}$ is a submartingale, then $\{-Z_n;\, n \geq 1\}$ is a supermartingale, and conversely. Thus, some of the results to follow are stated only for submartingales, with the understanding that they can be applied to supermartingales by changing signs as above.

Lemma 9.6.1, with the equality replaced by inequality, also applies to submartingales and supermartingales. That is, if $\{Z_n;\, n \geq 1\}$ is a submartingale, then

$$\mathsf{E}[Z_n \mid Z_i, Z_{i-1}, \dots, Z_1] \geq Z_i \quad ; \quad 1 \leq i < n, \tag{9.86}$$

and if $\{Z_n;\, n \geq 1\}$ is a supermartingale, then

$$\mathsf{E}[Z_n \mid Z_i, Z_{i-1}, \dots, Z_1] \leq Z_i \quad ; \quad 1 \leq i < n. \tag{9.87}$$

Equations (9.86) and (9.87) are verified in the same way as Lemma 9.6.1 (see Exercise 9.24). Similarly, the appropriate generalization of (9.83) is that if $\{Z_n;\, n \geq 1\}$ is a submartingale, then

$$\mathsf{E}[Z_n] \geq \mathsf{E}[Z_i] \quad ; \quad \text{for all } i,\ 1 \leq i < n. \tag{9.88}$$

and if $\{Z_n;\, n \geq 1\}$ is a supermartingale, then

$$\mathsf{E}[Z_n] \leq \mathsf{E}[Z_i] \quad ; \quad \text{for all } i,\ 1 \leq i < n. \tag{9.89}$$

A random walk $\{S_n;\, n \geq 1\}$ with $S_n = X_1 + \cdots + X_n$ is a submartingale, martingale, or supermartingale respectively for $\overline{X} \geq 0$, $\overline{X} = 0$, or $\overline{X} \leq 0$. Also, if X has a semi-invariant moment generating function $\gamma(r)$ for some given r, and if Z_n is defined as $Z_n = \exp(rS_n)$, then the process $\{Z_n;\, n \geq 1\}$ is a submartingale, martingale, or supermartingale respectively for $\gamma(r) \geq 0$, $\gamma(r) = 0$, or $\gamma(r) \leq 0$. The next example gives an important way in which martingales and submartingales are related.

Example 9.7.1 (Convex functions of martingales). Figure 9.8 illustrates the graph of a convex function h from $\mathbb{R}$ to $\mathbb{R}$. Recall that h is convex if

$$\mu h(x_1) + (1-\mu)h(x_2) \geq h(\mu x_1 + (1-\mu)x_2)$$

for every real x_1, x_2 and every $\mu, 0 < \mu < 1$.

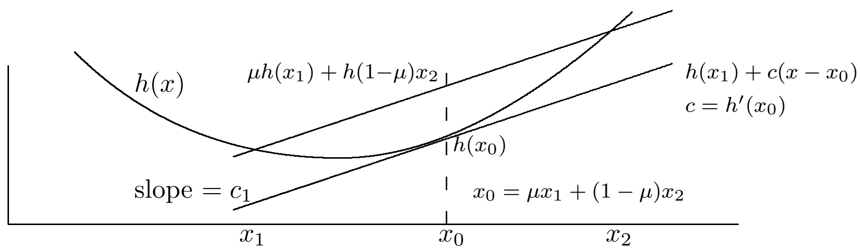


Figure 9.8: Convex function $h(x)$: Every chord of $h(x)$ lies on or above the curve. Equivalently, for each x_0, there is a c such that, for all x, $h(x_0) + c(x - x_0) \leq h(x)$, *i.e.*, all tangents of h lie on or below the curve.. If h is differentiable at x_0, then c is the derivative of h at x_0.

Geometrically, this says that every chord of h lies on or above h. By lowering each chord while holding its slope constant, we see that an equivalent condition is that each tangent to h lies on or below h. If $h(x)$ has a derivative at x_0, then the slope c of the tangent at x_0 is the value of that derivative and $h(x) + c(x - x_0)$ is the equation, as a function of x, for the tangent line at x_0. If $h(x)$ has a discontinuous slope at x_0, then there might be many choices for c; for example, $h(x) = |x|$ is convex, and for $x_0 = 0$, one could choose any c in the range -1 to $+1$ to form a tangent.

A simple condition that implies convexity is a nonnegative second derivative everywhere. This is not a necessary condition, however, and functions (such as $|x|$) are convex even when the first derivative does not exist everywhere.

Lemma 9.7.1 (Jensen's inequality). *If h is a convex function from $\mathbb{R}$ to $\mathbb{R}$ and X is a random variable with $\mathsf{E}[|X|] < \infty$, then*

$$h(\mathsf{E}[X]) \leq \mathsf{E}[h(X)]. \tag{9.90}$$

Proof: Let $x_0 = \mathsf{E}[X]$ and choose c so that $h(x_0) + c(x - x_0) \le h(x)$ for all x. Using the random variable X in place of x and taking expected values of both sides, we get (9.90). □

Note that for any particular event A, this same argument applies to X conditional on A, so that $h(\mathsf{E}[X \mid A]) \le \mathsf{E}[h(X) \mid A]$. Jensen's inequality is very widely used; it is a minor miracle that we have not required it previously.

Theorem 9.7.1. *Assume that h is a convex function from $\mathbb{R}$ to $\mathbb{R}$, that $\{Z_n;\, n \ge 1\}$ is a martingale or submartingale, and that $\mathsf{E}[|h(Z_n)|] < \infty$ for all n. Then $\{h(Z_n);\, n \ge 1\}$ is a submartingale.*

Proof: For any choice of $z_1, \dots, z_{n-1}$, we can use Jensen's inequality with the conditioning event to get

$$\mathsf{E}[h(Z_n)|Z_{n-1}{=}z_{n-1}, \dots, Z_1{=}z_1] \ge h(\mathsf{E}[Z_n \mid Z_{n-1}{=}z_{n-1}, \dots, Z_1{=}z_1]) = h(z_{n-1}). \quad (9.91)$$

For any choice of numbers $h_1, \dots, h_{n-1}$ in the range of the function h, let $z_1, \dots, z_{n-1}$ be arbitrary numbers satisfying $h(z_1){=}h_1, \dots, h(z_{n-1}){=}h_{n-1}$. For each such choice, (9.91) holds, so that

$$\begin{aligned} \mathsf{E}[h(Z_n) \mid h(Z_{n-1}){=}h_{n-1}, \dots, h(Z_1){=}h_1] \quad &\ge \quad h(\mathsf{E}[Z_n \mid h(Z_{n-1}){=}h_{n-1}, \dots, h(Z_1){=}h_1]) \\ &= \quad h(z_{n-1}) = h_{n-1}. \end{aligned} \quad (9.92)$$

completing the proof. □

Some examples of this result, applied to a martingale $\{Z_n;\, n \ge 1\}$, are as follows:

$$\{|Z_n|;\, n \ge 1\} \text{ is a submartingale} \quad (9.93)$$

$$\{Z_n^2;\, n \ge 1\} \text{ is a submartingale if } \mathsf{E}\left[Z_n^2\right] < \infty;\; n \ge 1 \quad (9.94)$$

$$\{\exp(rZ_n);\, n \ge 1\} \text{ is a submartingale if } \mathsf{E}[\exp(rZ_n)] < \infty;\; n \ge 1. \quad (9.95)$$

A function of a real variable $h(x)$ is defined to be concave if $-h(x)$ is convex. It then follows from Theorem 9.7.1 that if h is concave and $\{Z_n;\, n \ge 1\}$ is a martingale, then $\{h(Z_n);\, n \ge 1\}$ is a supermartingale (assuming that $\mathsf{E}[|h(Z_n)|] < \infty$). For example, if $\{Z_n;\, n \ge 1\}$ is a positive martingale and $\mathsf{E}[|\ln(Z_n)|] < \infty$, then $\{\ln(Z_n);\, n \ge 1\}$ is a supermartingale.

9.8 Stopped processes and stopping trials

The definition of stopping trials in Section 5.5 applies to arbitrary integer-time processes $\{Z_n;\, n \ge 1\}$ as well as to IID sequences. Recall that J is a stopping trial for a sequence $\{Z_n;\, n \ge 1\}$ of rv's if $\mathbb{I}_{J=n}$ is a function of $Z_1, \dots, Z_n$ and if J is a rv.

If $\mathbb{I}_{J=n}$ is a function of $Z_1, \dots, Z_n$ and J is a defective rv, then J is called a defective stopping trial. For some of the results to follow, it is unimportant whether J is a random variable or a defective random variable (*i.e.*, whether or not the process stops with probability 1). If it is not specified whether J is a random variable or a defective random variable, we refer to the stopping trial as a *possibly-defective stopping trial*; we consider J to take on the value ∞ if the process does not stop.

Definition 9.8.1. *A stopped process $\{Z_n^*;\ n \geq 1\}$ for a possibly-defective stopping trial J relative to a process $\{Z_n;\ n \geq 1\}$ is the process for which $Z_n^* = Z_n$ for $n \leq J$ and $Z_n^* = Z_J$ for $n > J$.*

As an example, suppose Z_n models the fortune of a gambler at the completion of the nth trial of some game, and suppose the gambler then modifies the game by deciding to stop gambling under some given circumstances (*i.e.*, at the stopping trial). Thus, after stopping, the fortune remains constant, so the stopped process models the gambler's fortune in time, including the effect of the stopping trial.

As another example, consider a random walk with a positive and negative threshold, and consider the process to stop after reaching or crossing a threshold. The stopped process then stays at that point after the threshold crossing as an artifice to simplify analysis. The use of stopped processes is similar to the artifice that we employed in Section 4.5 for first-passage times in Markov chains; recall that we added an artificial trapping state after the desired passage to simplify analysis.

We next show that the possibly-defective stopped process of a martingale is itself a martingale; the intuitive reason is that, before stopping, the stopped process is the same as the martingale, and, after stopping, $Z_n^* = Z_{n-1}^*$. The following theorem establishes this and the corresponding results for submartingales and supermartingales.

Theorem 9.8.1. *Given a stochastic process $\{Z_n;\ n \geq 1\}$ and a possibly-defective stopping trial J for the process, the stopped process $\{Z_n^*;\ n \geq 1\}$ is a submartingale if $\{Z_n;\ n \geq 1\}$ is a submartingale, is a martingale if $\{Z_n;\ n \geq 1\}$ is a martingale, and is a supermartingale if $\{Z_n;\ n \geq 1\}$ is a supermartingale.*

Proof: First we show that, for all three cases, the stopped process satisfies $\mathsf{E}\left[|Z_n^*|\right] < \infty$ for any given $n \geq 1$. Conditional on $J = i$ for some $i < n$, we have $Z_n^* = Z_i$, so

$$\mathsf{E}\left[|Z_n^*| \mid J = i\right] = \mathsf{E}\left[|Z_i| \mid J = i\right] < \infty \quad \text{for each } i < n \text{ such that } \Pr\{J = i\} > 0.$$

The reason for this is that if $\mathsf{E}\left[|Z_i| \mid J = i\right] = \infty$ and $\Pr\{J = i\} > 0$, then $\mathsf{E}\left[|Z_i|\right] = \infty$, contrary to the assumption that $\{Z_n;\ n \geq 1\}$ is a martingale, submartingale, or supermartingale. Similarly, for $J \geq n$, we have $Z_n^* = Z_n$ so

$$\mathsf{E}\left[|Z_n^*| \mid J \geq n\right] = \mathsf{E}\left[|Z_n| \mid J \geq n\right] < \infty \qquad \text{if } \Pr\{J \geq n\} > 0.$$

Averaging,

$$\mathsf{E}\left[|Z_n^*|\right] = \sum_{i=1}^{n-1} \mathsf{E}\left[|Z_n^*| \mid J{=}i\right] \Pr\{J{=}i\} + \mathsf{E}\left[|Z_n^*| \mid J \geq n\right] \Pr\{J \geq n\} < \infty.$$

Next assume that $\{Z_n;\, n \geq 1\}$ is a submartingale. For any given $n > 1$, consider an arbitrary initial sample sequence $(Z_1 = z_1,\, Z_2 = z_2, \dots, Z_{n-1} = z_{n-1})$. Note that z_1 specifies whether or not $J = 1$. Similarly, (z_1, z_2) specifies whether or not $J = 2$, and so forth up to $(z_1, \dots, z_{n-1})$, which specifies whether or not $J = n-1$. Thus $(z_1, \dots, z_{n-1})$ specifies the sample value of J for $J \leq n-1$ and specifies that $J \geq n$ otherwise.

For $(z_1, \dots, z_{n-1})$ such that $\mathbb{I}_{J \geq n} = 0$, we have $z_n^* = z_{n-1}^*$. For all such sample values,

$$\mathsf{E}\left[Z_n^* | Z_{n-1}^*{=}z_{n-1}^*, \ldots, Z_1^*{=}z_1^*\right] = z_{n-1}^*. \tag{9.96}$$

For the remaining case, where $(z_1, \dots, z_{n-1})$ is such that $\mathbb{I}_{J \geq n} = 1$, we have $z_n^* = z_n$. Thus

$$\mathsf{E}\left[Z_n^* | Z_{n-1}^*{=}z_{n-1}^*, \ldots, Z_1^*{=}z_1^*\right] \geq z_{n-1}^*. \tag{9.97}$$

The same argument works for martingales and supermartingales by replacing the inequality in (9.97) by equality for the martingale case and the opposite inequality for the supermartingale case. □

Theorem 9.8.2. Given a stochastic process $\{Z_n;\, n \geq 1\}$ and a possibly-defective stopping trial J for the process, the stopped process $\{Z_n^*;\, n \geq 1\}$ satisfies the following conditions for all $n \geq 1$ if $\{Z_n;\, n \geq 1\}$ is a submartingale, martingale, or supermartingale respectively:

$$\mathsf{E}[Z_1] \;\leq\; \mathsf{E}[Z_n^*] \;\leq\; \mathsf{E}[Z_n] \qquad \text{(submartingale)} \tag{9.98}$$

$$\mathsf{E}[Z_1] \;=\; \mathsf{E}[Z_n^*] \;=\; \mathsf{E}[Z_n] \qquad \text{(martingale)} \tag{9.99}$$

$$\mathsf{E}[Z_1] \;\geq\; \mathsf{E}[Z_n^*] \;\geq\; \mathsf{E}[Z_n] \qquad \text{(supermartingale).} \tag{9.100}$$

Proof: First assume that $\{Z_n;\, n \geq 1\}$ is a submartingale. Theorem 9.8.1 shows that $\{Z_n^*;\, n \geq 1\}$ is a submartingale, so from (9.88), $\mathsf{E}[Z_1^*] \leq \mathsf{E}[Z_n^*]$ for all $n \geq 1$. Since $Z_1 = Z_1^*$, this establishes the first half of (9.98). For the second half, choose any $m \leq n$ and let $z_1, z_2, \dots, z_m$ be any initial sample segment such that $J = m$. Then $z_n^* = z_m$, so

$$\mathsf{E}[Z_n^* | Z_m{=}z_m, \dots, Z_1{=}z_1] = z_m.$$

On the other hand, since $\{Z_n;\, n \geq 1\}$ is a submartingale,

$$\mathsf{E}[Z_n | Z_m{=}z_m, \dots, Z_1{=}z_1] \geq z_m.$$

Since this is true for all $z_1, \dots, z_m$ for which $J = m$, $\mathsf{E}[Z_n \mid J = m] \geq \mathsf{E}[Z_n^* | J = m]$. This is true for all $m \leq n$, and is clearly valid with equality for $m > n$. Averaging over J then yields the second half of (9.98).

Finally, if $\{Z_n;\, n \geq 1\}$ is a supermartingale, then $\{-Z_n;\, n \geq 1\}$ is a submartingale, verifying (9.100). Since a martingale is both a submartingale and supermartingale, (9.99) follows and the proof is complete. □

Consider a (non-defective) stopping trial J for a martingale $\{Z_n;\, n \geq 1\}$. Since the stopped process is also a martingale, we have

$$\mathsf{E}[Z_n^*] = \mathsf{E}[Z_1^*] = \mathsf{E}[Z_1];\; n \geq 1. \tag{9.101}$$

Since $Z_n^* = Z_J$ for all $n \geq J$ and since J is finite with probability 1, we see that $\lim_{n\to\infty} Z_n^* = Z_J$ with probability 1. Surprisingly, $\mathsf{E}[Z_J]$ is not necessarily equal to $\lim_{n\to\infty} \mathsf{E}[Z_n^*] = \mathsf{E}[Z_1]$. The binary product martingale in (9.76) gives an example of inequality. Taking the stopping trial J to be the smallest n for which $Z_n = 0$, we have $Z_J = 0$ with probability 1, and thus $\mathsf{E}[Z_J] = 0$. But $Z_n^* = Z_n$ for all n, and $\mathsf{E}[Z_n^*] = 1$ for all n. The problem here is that, given that the process has not stopped by time n, Z_n and Z_n^* each have the value 2^n. Fortunately, in most situations, this type of bizarre behavior does not occur and $\mathsf{E}[Z_J] = \mathsf{E}[Z_1]$. To get a better understanding of when $\mathsf{E}[Z_J] = \mathsf{E}[Z_1]$, note that for any n, we have

$$\mathsf{E}[Z_n^*] = \sum_{i=1}^{n} \mathsf{E}[Z_n^* \mid J = i] \Pr\{J = i\} + \mathsf{E}[Z_n^* \mid J > n] \Pr\{J > n\} \tag{9.102}$$

$$= \sum_{i=1}^{n} \mathsf{E}[Z_J \mid J = i] \Pr\{J = i\} + \mathsf{E}[Z_n \mid J > n] \Pr\{J > n\}. \tag{9.103}$$

The left side of this equation is $\mathsf{E}[Z_1]$ for all n. If the final term on the right converges to 0 as $n \to \infty$, then the sum must converge to $\mathsf{E}[Z_1]$. If $\mathsf{E}[|Z_J|] < \infty$, then the sum also converges to $\mathsf{E}[Z_J]$. Without the condition $\mathsf{E}[|Z_J|] < \infty$, the sum might consist of alternating terms which converge, but whose absolute values do not converge, in which case $\mathsf{E}[Z_J]$ does not exist (see Exercise 9.27 for an example). Thus we have established the following lemma.

Lemma 9.8.1. Let J be a stopping trial for a martingale $\{Z_n;\, n \geq 1\}$. Then $\mathsf{E}[Z_J] = \mathsf{E}[Z_1]$ if and only if

$$\lim_{n\to\infty} \mathsf{E}[Z_n \mid J > n] \Pr\{J > n\} = 0 \qquad \text{and} \quad \mathsf{E}[|Z_J|] < \infty. \tag{9.104}$$

9.8.1 The Wald identity

Recall the generating function product martingale of (9.77) in which $\{Z_n = \exp[rS_n - n\gamma(r)];\, n \geq 1\}$ is a martingale defined in terms of the random walk $\{S_n = X_1 + \cdots + X_n;\, n \geq 1\}$. From (9.99), we have $\mathsf{E}[Z_n] = \mathsf{E}[Z_1]$, and since $\mathsf{E}[Z_1] = \mathsf{E}[\exp\{rX_1 - \gamma(r\}] = 1$, we have $\mathsf{E}[Z_n] = 1$ for all n. Also, for any possibly-defective stopping trial J, we have $\mathsf{E}[Z_n^*] = \mathsf{E}[Z_1] = 1$. This is the first part of the following theorem, and the second part follows from Lemma 9.8.1.

Theorem 9.8.3 (Wald's identity). *If J is a possibly defective stopping trial, then the stopped process $\{Z_n;\, n \geq 1\}$ satisfies*

$$\mathsf{E}[Z_n^*] = 1 \qquad \textit{for all } n \geq 1 \tag{9.105}$$

If J is a non-defective stopping trial, and if (9.104) holds, then

$$\mathsf{E}[Z_J] = \mathsf{E}[\exp\{rS_J - J\gamma(r)\}] = 1. \tag{9.106}$$

If there are two thresholds, one at $\alpha > 0$, and the other at $\beta < 0$, and the stopping rule is to stop when either threshold is crossed, then (9.106) is just the Wald identity for 2 thresholds, (9.38). If there is only one threshold, with $\alpha > 0$ and $\mathsf{E}[X] < 0$, then J is defective but (9.105) still holds. If we choose $r = r^*$ (the positive root of $\gamma(r)$), then $Z_n = \exp(rS_n - r\gamma(r))$ becomes $Z_n = \exp(r^* S_n)$. In this case, (9.105) shows that $\Pr\{S_J \geq \alpha\} \leq e^{r^*\alpha}$, with no need for the limiting argument used to derive (9.49).

Theorem 9.8.3 can also be used for other stopping rules. For example, for some given integer n, let J_{n+} be the smallest integer $i \geq n$ for which $S_i \geq \alpha$ or $S_i \leq \beta$. Then, in the limit $\beta \to -\infty$, $\Pr\{S_{J_{n+}} \geq \alpha\} = \Pr\{\cup_{i=n}^{\infty}(S_i \geq \alpha)\}$. Assuming $\overline{X} < 0$, we can find an upper bound to $\Pr\{S_{J_{n+}} \geq \alpha\}$ for any $r > 0$ and $\gamma(r) \leq 0$ (*i.e.*, for $0 < r \leq r^*$) by the following steps

$$1 = \mathsf{E}\left[\exp\{rS_{J_{n+}} - J_{n+}\gamma(r)\}\right] \geq \Pr\{S_{J_{n+}} \geq \alpha\}\exp[r\alpha - n\gamma(r)]$$

$$\Pr\{S_{J_{n+}} \geq \alpha\} \leq \exp[-r\alpha + n\gamma(r)]; \qquad 0 \leq r \leq r^*. \tag{9.107}$$

9.9 The Kolmogorov inequalities

We now use the previous theorems to establish Kolmogorov's submartingale inequality, which is a major strengthening of the Markov inequality. Just as the Markov inequality in Section 1.7 was used to derive the Chebychev inequality, the Chernoff bound, and the weak law of large numbers, the Kolmogorov submartingale inequality will be used to develop a number of related inequalities and then to prove the strong law of large numbers (SLLN) and the martingale convergence theorem. The SLLN here assumes only a second moment rather than the fourth moment assumed in Section 5.2. Perhaps more important than the increased generality of the SLLN here is that the proof suggests a number of more general results not requiring IID rv's.

The Kolmogorov submartingale inequality will follow easily from looking at a particular class of stopping rules for nonnegative submartingales. Each such stopping rule is characterized by a threshold a and a trial number m and is denoted by $J(m, a)$. Stopping occurs either at the first crossing of the threshold at a or at trial m, which comes first.

Since stopping occurs by trial m at the latest, we see that $Z_{J(m,a)} \geq a$ if and only if the threshold is crossed at some trial $n \leq m$. In terms of events, this means that

$$\{Z_{J(m,a)} \geq a\} \;=\; \{\max_{1\leq n\leq m} Z_n \geq a\} \tag{9.108}$$

Applying the Markov inequality to $Z_{J(m,a)}$, we then get

$$\Pr\left\{\max_{1\leq n\leq m} Z_n \geq a\right\} \;=\; \Pr\{Z_{J(m,a)} \geq a\} \;\leq\; \frac{\mathsf{E}\left[Z_{J(m,a)}\right]}{a} \tag{9.109}$$

It is only a small step from this to Kolmogorovs's martingale inequality:

Theorem 9.9.1 (Kolmogorov's submartingale inequality). *Let $\{Z_n;\, n \geq 1\}$ be a non-negative submartingale. Then for any positive integer m and any $a > 0$,*

$$\Pr\left\{\max_{1\leq n\leq m} Z_n \geq a\right\} \;\leq\; \frac{\mathsf{E}[Z_m]}{a}. \tag{9.110}$$

Proof: Using first the inequality (9.109),

$$\Pr\left\{\max_{1\le n\le m} Z_n \ge a\right\} \;\le\; \frac{\mathsf{E}\left[Z_{J(m,a)}\right]}{a} \;=\; \frac{\mathsf{E}\left[Z_m^*\right]}{a}$$

The equality above follows because the process must stop by trial m. Finally, from (9.98), $\mathsf{E}\left[Z_m^*\right] \le \mathsf{E}\left[Z_m\right]$, yielding (9.110). □

The following simple corollary shows that (9.110) has a limiting form for nonnegative martingales.

Corollary 9.9.1 (Nonnegative martingale inequality). *Let $\{Z_n;\, n \ge 1\}$ be a nonnegative martingale. Then*

$$\Pr\left\{\sup_{n\ge 1} Z_n \ge a\right\} \le \frac{\mathsf{E}\left[Z_1\right]}{a}; \qquad \textit{for all } a > 0. \tag{9.111}$$

Proof: For a martingale, $\mathsf{E}\left[Z_m\right] = \mathsf{E}\left[Z_1\right]$. Thus, from (9.110), $\Pr\{\max_{1\le i\le m} Z_i \ge a\} \le \frac{\mathsf{E}[Z_1]}{a}$ for all $m > 1$. Passing to the limit $m \to \infty$ essentially yields (9.111). Exercise 9.28 illustrates why the limiting operation is a little tricky, and then shows that it is valid. □

The next 3 corollaries are straightforward consequences of the Kolmogorov submartingale inequality. Their proofs are given in Exercise 9.31. The first bears the same relationship to the Kolmogorov submartingale inequality as the Chebychev inequality does to the Markov inequality. When applied to martingales, it is sometimes called the Kolmogorov martingale inequality.

Corollary 9.9.2. *Let $\{Z_n;\, n \ge 1\}$ be a submartingale with $\mathsf{E}\left[Z_n^2\right] < \infty$ for all $n \ge 1$. Then*

$$\Pr\left\{\max_{1\le n\le m} |Z_n| \ge b\right\} \le \frac{\mathsf{E}\left[Z_m^2\right]}{b^2}; \;\; \textit{for all integer } m \ge 2, \textit{ all } b > 0. \tag{9.112}$$

Corollary 9.9.3 (Kolmogorov's random walk inequality). *Let $\{S_n;\, n \ge 1\}$ be a random walk with $S_n = X_1 + \cdots + X_n$ where $\{X_i;\, i \ge i\}$ is a set of IID random variables with mean $\overline{X}$ and variance σ^2. Then for any positive integer m and any $\epsilon > 0$,*

$$\Pr\left\{\max_{1\le n\le m} |S_n - n\overline{X}| \ge m\epsilon\right\} \le \frac{\sigma^2}{m\epsilon^2}. \tag{9.113}$$

Recall that the simplest form of the weak law of large numbers was given in (1.74) as $\Pr\left\{|S_m/m - \overline{X}| \ge \epsilon\right\} \le \sigma^2/(m\epsilon^2)$. This is strengthened in (9.113) to upper bound the probability that any of the first m terms deviate from the mean by more than $m\epsilon$. It is this strengthening that will allow us to prove the strong law of large numbers assuming only a finite variance.

The following corollary provides a tight exponential bound to the probability of crossing a threshold before a given number of trials.

Corollary 9.9.4. *Let $\{S_n;\, n \geq 1\}$ be a random walk, $S_n = X_1 + \cdots + X_n$ where each X_i has mean $\overline{X} < 0$ and semi-invariant moment generating function $\gamma(r)$. For any $r > 0$ such that $0 < \gamma(r) < \infty$ (*i.e.*, for $r > r^*$), and for any $a > 0$.*

$$\Pr\left\{\max_{1\leq i\leq n} S_i \geq \alpha\right\} \leq \exp\{-r\alpha + n\gamma(r)\}. \tag{9.114}$$

Proof: For $r > r^*$, $\{\exp(rS_n);\, n \geq 1\}$ is a submartingale. Taking a $= \exp(r\alpha)$ in (9.110), we get (9.114). □

The following theorem about supermartingales is, in a sense, the dual of the Kolmogorov submartingale inequality. Note, however, that it applies to the terms $n \geq m$ in the supermartingale rather than $n \leq m$.

Theorem 9.9.2. *Let $\{Z_n;\, n \geq 1\}$ be a nonnegative supermartingale. Then for any positive integer m and any $a > 0$,*

$$\Pr\left\{\bigcup_{i\geq m}\{Z_i \geq a\}\right\} \leq \frac{\mathsf{E}[Z_m]}{a}. \tag{9.115}$$

Proof: For given $m \geq 1$ and $a > 0$, let J be a possibly-defective stopping trial defined as the smallest $i \geq m$ for which $Z_i \geq a$. Let $\{Z_n^*;\, n \geq 1\}$ be the corresponding stopped process, which is also nonnegative and is a supermartingale from Theorem 9.8.1. For any $k > m$, note that $Z_k^* \geq a$ if and only if $\max_{m\leq i\leq k} Z_i \geq a$. Thus

$$\Pr\left\{\max_{m\leq i\leq k} Z_i \geq a\right\} = \Pr\{Z_k^* \geq a\} \leq \frac{\mathsf{E}[Z_k^*]}{a}.$$

Since $\{Z_n^*;\, n \geq 1\}$ is a supermartingale, (9.89) shows that $\mathsf{E}[Z_k^*] \leq \mathsf{E}[Z_m^*]$. On the other hand, $Z_m^* = Z_m$ since the process can not stop before epoch m. Thus $\Pr\{\max_{m\leq i\leq k} Z_i \geq a\}$ is at most $\mathsf{E}[Z_m]/a$. Since k is arbitrary, we can pass to the limit, getting (9.115) and completing the proof. □

9.9.1 The strong law of large numbers (SLLN)

We now use the Kolmogorov sub-martingale inequality to prove the SLLN assuming only a second moment rather than the fourth moment assumed in Section 5.2.1. Perhaps more important than requiring only a second moment is that the proof suggests a number of SLLN type results not requiring IID rv's. The SLLN is also true assuming only a first absolute moment, but the truncation argument we used for the weak law in Theorem 1.7.4 is inadequate here.

Theorem 9.9.3 (SLLN). *Let $\{X_i;\, i \geq 1\}$ be a sequence of IID random variables with zero mean and standard deviation $\sigma < \infty$. Let $S_n = X_1 + \cdots + X_n$ for each $n \geq 1$. Then*

$$\Pr\left\{\lim_{n\to\infty}\frac{S_n}{n} = 0\right\} = 1 \tag{9.116}$$

Discussion: The theorem generalizes trivially to rv's with a non-zero mean. We simply let the X of the theorem be the fluctuation of the original rv. The statement then becomes

$$\Pr\left\{\lim_{n\to\infty} \frac{S_n}{n} - \overline{X} = 0\right\} = 1 \tag{9.117}$$

Proof of theorem: Let $Z_n = S_n^2$ for each n. Then $\{Z_n;\, n \geq 1\}$ is a submartingale. Let $\epsilon > 0$ be fixed and for each $\ell \geq 1$, consider the stopping rule $J(2^\ell, \epsilon 2^{2\ell})$, here abbreviated as J_ℓ. From the Kolmogorov submartingale inequality,

$$\Pr\left\{\max_{1\leq n\leq 2^\ell} Z_n \geq \epsilon 2^{2\ell}\right\} \leq \frac{\mathsf{E}\left[Z_{J_\ell}\right]}{\epsilon 2^{2\ell}} \leq \frac{\mathsf{E}\left[Z_{2^\ell}\right]}{\epsilon 2^{2\ell}} \tag{9.118}$$

For each ℓ, $Z_{2^\ell}/(\epsilon 2^{2\ell})$ is a rv of mean $2^\ell\sigma^2/(\epsilon 2^{2\ell}) = 2^{-\ell}\sigma^2/\epsilon$. Thus

$$\sum_{\ell\geq 1} \mathsf{E}\left[\frac{Z_{J_\ell}}{\epsilon 2^{2\ell}}\right] \leq \sum_{\ell \geq 1} 2^{-\ell}\sigma^2/\epsilon = \sigma^2/\epsilon < \infty$$

It then follows from Lemma 5.2.1 (the lemma on convergence WP1 used to prove the SLLN in Chap 5) that

$$\Pr\left\{\omega : \lim_{\ell\to\infty} \frac{Z_{J_\ell}(\omega)}{\epsilon 2^{2\ell}} = 0\right\} = 1.$$

This means that there is a set of sample points, say Ω_1, of probability 1 such that, for each $\omega \in \Omega_1$ and each $\epsilon > 0$ (and thus in particular the ϵ chosen above), there is an $m(\omega, \epsilon)$ such that $Z_{J_\ell}(\omega)/2^{2\ell} < \epsilon$ for all $\ell \geq m(\omega, \epsilon)$. From (9.108),

$$\max_{1\leq n\leq 2^\ell} \frac{Z_n(\omega)}{2^{2\ell}} < \epsilon \qquad \text{for } \ell \geq m(\omega,\epsilon),\ \omega \in \Omega_1. \tag{9.119}$$

Now (9.119) remains valid if the maximum is restricted to $2^{\ell-1} < n < 2^\ell$, and in this region, $Z_n/n^2 < Z_n/2^{2(\ell-1)} = 4Z_n/2^{2\ell}$. Thus

$$\max_{2^{\ell-1}<n\leq 2^\ell} \frac{Z_n(\omega)}{n^2} < 4\epsilon \qquad \text{for } \ell \geq m(\omega,\epsilon),\ \omega \in \Omega_1. \tag{9.120}$$

This means that $Z_n(\omega)/n^2 < 4\epsilon$ for all $n > 2^{m(\omega,\epsilon)-1}$. Since this is true for all $\epsilon > 0$, it means that $\{Z_n(\omega)/n^2;\, n \geq 1\}$ converges to 0, and thus $\{Z_n/n^2;\, n \geq 1\}$ converges to 0 WP1. Since $Z_n = S_n^2$, we can take the square root of Z_n/n^2 to see that $\{S_n/n;\, n \geq 1\}$ converges to 0 WP1. □

9.9.2 The martingale convergence theorem

Another famous result that follows from the Kolmogorov submartingale inequality is the martingale convergence theorem. This states that if a martingale $\{Z_n;\, n \geq 1\}$ has the property that there is some finite M such that $\mathsf{E}\left[|Z_n|\right] \leq M$ for all n, then $\lim_{n\to\infty} Z_n$ exists (and is finite) with probability 1. This is a powerful theorem, but the limitation that $\mathsf{E}\left[|Z_n|\right] \leq M$ for all n is far more than a technical restriction; for example it is not satisfied by a zero-mean random walk. We prove the theorem with the additional restriction that there is some finite M such that $\mathsf{E}\left[Z_n^2\right] \leq M$ for all n.

Theorem 9.9.4 (Martingale convergence theorem). *Let $\{Z_n;\, n \geq 1\}$ be a martingale and assume that there is some finite M such that $\mathsf{E}\left[Z_n^2\right] \leq M$ for all n. Then there is a random variable Z such that, for all sample sequences except a set of probability 0, $\lim_{n\to\infty} Z_n = Z$.*

Proof*: From Theorem 9.7.1 and the assumption that $\mathsf{E}\left[Z_n^2\right] \leq M$, $\{Z_n^2;\, n \geq 1\}$ is a submartingale. Thus, from (9.88), $\mathsf{E}\left[Z_n^2\right]$ is non-decreasing in n, and since $\mathsf{E}\left[Z_n^2\right]$ is bounded, $\lim_{n\to\infty} \mathsf{E}\left[Z_n^2\right] = M'$ for some $M' \leq M$. For any integer k, the process $\{Y_n = Z_{k+n} - Z_k;\, n \geq 1\}$ is a zero-mean martingale (see Exercise 9.39). Thus from Corollary 9.9.2,

$$\Pr\left\{\max_{1\leq n\leq m} |Z_{k+n} - Z_k| \geq b\right\} \leq \mathsf{E}\left[(Z_{k+m} - Z_k)^2\right]/b^2. \tag{9.121}$$

Next, observe that $\mathsf{E}\left[Z_{k+m}Z_k \mid Z_k = z_k, Z_{k-1} = z_{k-1}, \dots, Z_1 = z_1\right] = z_k^2$, and therefore, $\mathsf{E}\left[Z_{k+m}Z_k\right] = \mathsf{E}\left[Z_k^2\right]$. Thus $\mathsf{E}\left[(Z_{k+m} - Z_k)^2\right] = \mathsf{E}\left[Z_{k+m}^2\right] - \mathsf{E}\left[Z_k^2\right] \leq M' - \mathsf{E}\left[Z_k^2\right]$. Since this is independent of m, we can pass to the limit, obtaining

$$\Pr\left\{\sup_{n\geq 1} |Z_{k+n} - Z_k| \geq b\right\} \leq \frac{M' - \mathsf{E}\left[Z_k^2\right]}{b^2}. \tag{9.122}$$

Since $\lim_{k\to\infty} \mathsf{E}\left[Z_k^2\right] = M'$, we then have, for all $b > 0$,

$$\lim_{k\to\infty} \Pr\left\{\sup_{n\geq 1} |Z_{k+n} - Z_k| \geq b\right\} = 0. \tag{9.123}$$

This means that with probability 1, a sample sequence of $\{Z_n;\, n \geq 1\}$ is a Cauchy sequence, and thus approaches a limit, concluding the proof. □

Example 9.9.1 (Limits of orthonormal expansions of random processes). Assume that $\{X_n;\, n \geq 1\}$ is a sequence of zero-mean independent rv's with finite variances, $\mathsf{E}\left[X_n^2\right] = \sigma_n^2 < \infty$. Let $\{\phi_n(t);\, n \geq 1, t \in \mathbb{R}\}$ be a sequence of orthonormal functions (see Section 3.6.3). Suppose that a random process is defined as an orthonormal expansion, $X(t) = \lim_{\ell\to\infty} X^\ell(t)$ where $X^\ell(t) = \sum_{n=1}^{\ell} X_n\phi_n(t)$.

We want to use the martingale convergence theorem to show that this limit exists WP1 for each t. For a given t, we see that $\{X^\ell(t);\, \ell \geq 1\}$ is simply a sequence of rv's. Since the underlying rv's X_n are independent and zero mean, the sums $X^\ell(t) = \sum_{n=1}^{\ell} X_n\phi_n(t)$ form a martingale (assuming each $\phi_n(t)$ is finite).

If we now make the further assumption that $\sum_{n=1}^{\infty} \sigma_n^2 |\phi_n(t)|^2$ is finite for each t, then the partial sums from 1 to ℓ are bounded independent of ℓ and the martingale convergence theorem applies. Thus for each t, the limit $\lim_{\ell\to\infty} X^\ell(t)$ exists WP1.

It should be clear that this same approach can be used in many situations to show the limit of a sequence of rv's is itself a rv WP1.

Example 9.9.2 (Branching processes). The martingale convergence theorem can also be interpreted relatively easily for branching processes. For a branching process $\{X_n;\, n \geq$

1} where $\overline{Y}$ is the expected number of offspring of an individual, $\{X_n/\overline{Y}^n;\, n \geq 1\}$ is a martingale that satisfies the above conditions. If $\overline{Y} \leq 1$, the branching process dies out with probability 1, so $X_n/\overline{Y}^n$ approaches 0 with probability 1. For $\overline{Y} > 1$, however, the branching process dies out with some probability less than 1 and approaches ∞ otherwise. Thus, the limiting random variable Z is 0 with the probability that the process ultimately dies out, and is positive otherwise. In the latter case, for large n, the interpretation is that when the population is very large, a law of large numbers effect controls its growth in each successive generation, so that $X_n/X_{n-1} \approx \overline{Y}$ for large X_{n-1}. The distribution of Z conditional on $Z > 0$ essentially depends on the random number of trials until X_n becomes large.

9.10 The investment game

Consider a range of possible investments, numbered 1 to ℓ, available to an investor. Each investment fluctuates in value over time, and the investor is faced with the problem of buying and selling those investments over time so as to maintain a portfolio that is desirable in some sense. We introduce a particularly simple model of such portfolio management in this section, partly for the purpose of illustrating some features of stochastic processes and partly as a theoretical introduction to an important application area.[5]

This is an area in which mathematics has sometimes run amuck and where it is difficult to thread a path between common sense, common practice, and mathematical models. The analysis here is not intended to address these practical concerns but only to introduce a theory that might be useful to those who also take the time and effort to fully address the practical side of the area.

We consider a discrete-time model, where the time unit could be microseconds (for programmed trading) or days or months (for more casual investors). Each available investment k, $1 \leq k \leq \ell$ is characterized by a nonnegative rv $X(k)$ which is the ratio of the price of the kth investment at the end of an epoch (a time unit) to that at the beginning of that epoch. As the kth investment evolves over time, we take $X_n(k)$ to be this ratio at epoch n. We assume that $\{X_n(k);\, n \geq 1\}$ is a sequence of IID rv's with the distribution of $X(k)$.

The assumption that $\{X_n(k);\, n \geq 1\}$ is IID is often unrealistic in practice, and in fact much of the work by investment analysts is devoted to understanding how the statistics of these rv's change over time. Our purpose here, however, is to understand how to allocate resources over time among investments, and for this, the simplicity of the IID assumption is desirable.

If 1$ is put into investment k at the beginning of epoch 1 and left to fluctuate over n epochs

[5]Some people view investing in stocks as simply a form of gambling with no redeeming social value. However, venture capitalists play an important role in the formation of new companies, thus providing jobs to many people. Vulture capitalists, on the other hand, often destroy companies for short term personal gain. For the most part, however, when one investor buys a stock, another investor sells it, and the act of buying bids up the price of that stock and thus indirectly helps the corresponding company and its employees. We leave the social aspects of investing as part of the application area, and concentrate here on the stochastic processes involved.

with no further purchases or sales, then the value of that investment (in dollars) at time n is the rv $W_n(k) = \prod_{m=1}^n X_m(k)$. We want to use the laws of large numbers to analyze this later, so it is convenient to define the rv $L_n(k)$ to be the log of $W_n(k)$. Thus

$$L_n(k) = \sum_{m=1}^{n} \ln X_m(k).$$

This means that $L_n(k)$ is the sum of n IID rv's .

For the set of all ℓ investments, we assume there is a joint distribution on the random vector $\boldsymbol{X} = (X(1), \dots, X(\ell))^{\mathsf{T}}$. There is no need to assume that these investments are independent of each other. We do assume however, that each $\boldsymbol{X}_n$ is statistically independent from and identically distributed to $\boldsymbol{X}_1, \dots, \boldsymbol{X}_{n-1}$. In other words, the random vectors $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n, \dots$, are IID. Note that each random vector $\boldsymbol{X}_n$ is a vector of ℓ rv's , $X_n(1), \dots, X_n(\ell)$; the subscript refers to time and the parentheses to the particular investment. The IID assumption is over time and not over investments.

We want to consider what happens if there is a joint allocation, $\boldsymbol{\lambda} = (\lambda(1), \dots, \lambda(\ell))^{\mathsf{T}}$ between investments such that, at the beginning of each epoch n, a fraction $\lambda(k)$ of the investors wealth is in investment k. This means that as the values of the individual investments shift, the portfolio is rebalanced at the end of each epoch to maintain the allocation $\boldsymbol{\lambda}$ for the beginning of the next epoch. We assume there is no commission on the buying and selling required by this rebalancing, so the investor's wealth changes only due to the changes in investment values during each epoch.

Finally assume that $\sum_k \lambda(k) = 1$ and $\lambda(k) \geq 0$ for each k. This mean that there is no investing on margin (borrowing money to invest) and that the investor is always fully invested. We will assume that the ℓth investment is 'cash,' *i.e.*, an investment where $X(\ell) = 1$ with probability 1, so that a cash investment incorporates what is usually meant by being not fully invested.

Let the investor's wealth at the beginning of epoch 1 be denoted as W_0. Then, generalizng the notation above for single investments, we denote the wealth at the end of epoch 1 for the allocation $\boldsymbol{\lambda}$ as the rv $W_1(\boldsymbol{\lambda}) = W_0 \sum_k \lambda(k) X_1(k)$. The wealth at the end of epoch 2 is found by taking the wealth at the end of epoch 1, rebalancing it to the fractions $\lambda_1, \dots, \lambda_\ell$, and investing that rebalanced amount over epoch 2, *i.e.*,

$$W_2(\boldsymbol{\lambda}) = W_0 \big[\sum_k \lambda(k) X_1(k) \big] \big[\sum_k \lambda(k) X_2(k) \Big].$$

The wealth $W_n(\boldsymbol{\lambda})$ at the end of epoch n is similarly W_0 times a product of n IID rv's. Since W_0 is just an extra parameter, we set it to 1, giving us

$$W_n(\boldsymbol{\lambda}) = \prod_{m=1}^{n} \Big[\sum_k \lambda(k) X_m(k) \Big]. \tag{9.124}$$

The log of the wealth at the end of epoch n is the rv $L_n(\boldsymbol{\lambda})$ given by

$$L_n(\boldsymbol{\lambda}) = \sum_{m=1}^{n} \ln \big[\sum_k \lambda(k) X_m(k) \big]. \tag{9.125}$$

Note that $L_n(\boldsymbol{\lambda})$ is the sum of n IID rv's and we can apply the laws of large numbers (given the existence of a mean). As discussed shortly, this means that the wealth of the investor grows (or shrinks) exponentially at a rate given by $\mathsf{E}[\ln \sum_k \lambda(k)X(k)]$. Before being more precise about this, we consider an example.

Example 9.10.1. Suppose there is only one possible investment other than cash, so we omit the index on investments. Assume that $X = 3$ or $X = 0$, each with probability 1/2. This appears to be a very nice investment, since the expected return in one time unit for 1$ invested is 1.5$. If we choose $\lambda = 1$, we see from (9.124) that W_n is 3^n with probability 2^{-n} and 0 with probability $1-2^{-n}$. Thus $\mathsf{E}[W_n] = (3/2)^n$, but $\lim_{n\to\infty} W_n = 0$ with probability 1. In the same way, $L_n = n\ln(3)$ with probability 2^{-n} and is $L_n = -\infty$ otherwise. We have found a goose that lays golden eggs, but in our greed, we go broke WP1.

The solution to this dilemma occurs to most gamblers who like to continue gambling. They keep a fraction λ of their wealth in the given investment and put $1-\lambda$ in cash. They can then continue to invest a fraction λ of their wealth at each epoch, while maintaining a fraction $1-\lambda$ in cash. Using (9.124) and (9.125) to find the expected wealth and log wealth at the end of epoch 1, we get

$$\mathsf{E}[W_1(\lambda)] \;=\; \frac{1}{2}(1-\lambda+3\lambda)+\frac{1}{2}(1-\lambda) \;=\; 1+\frac{\lambda}{2} \tag{9.126}$$

$$\mathsf{E}[L_1(\lambda)] \;=\; \frac{1}{2}\ln(1-\lambda+3\lambda)+\frac{1}{2}\ln(1-\lambda) \;=\; \frac{1}{2}\ln(1+\lambda-2\lambda^2). \tag{9.127}$$

The maximum of $\mathsf{E}[L_1(\lambda)]$ over λ is easily seen to be $(1/2)\ln(9/8)$ achieved at $\lambda = 1/4$. Thus $L_n(1/4)$ is a sum of n IID rv's each of mean $(1/2)\ln(9/8)$. By the law of large numbers, $(1/n)L_n(1/4)$ is close to $(1/2)\ln(9/8)$ with high probability for large n. Since $W_n(1/4) = e^{L_n(1/4)}$, this means that $W_n(1/4)$ should be close to $(9/8)^{n/2}$ with high probability, but we will have to be a little careful about what 'close to' means here. Note also that $\mathsf{E}[W_n(1/4)] = (9/8)^n$ so the behavior of W_n still seems a little peculiar. However, it should be clear that by keeping 3/4 of the wealth in cash, the investment has changed from a lottery type investment to a safe investment which grows exponentially with high probability.

9.10.1 Portfolios with fixed rebalancing

We now return to the general case with ℓ investments, but still consider only situations where the same $\boldsymbol{\lambda}$ is used at each epoch n. Using (9.125),

$$L_n(\boldsymbol{\lambda}) = \sum_{m=1}^{n} Y_m(\boldsymbol{\lambda}) \qquad \text{where } Y_m(\boldsymbol{\lambda}) = \ln\big(\sum_k \lambda(k)X_m(k)\big).$$

We assume that $\mathsf{E}[X(k)] < \infty$ for each k, $1 \le k \le \ell$ and also assume that $\mathsf{E}[|Y(\boldsymbol{\lambda}|] < \infty$. For Example 9.10.1, note from (9.127) that $\mathsf{E}[Y(\boldsymbol{\lambda}]$ approaches $-\infty$ as $\lambda \to 1$, so the assumption $\mathsf{E}[|Y(\boldsymbol{\lambda}|] < \infty$ is not satisfied for this example when $\lambda = 1$. This same problem occurs with any investment for which $X(k)$ takes on the value 0 with positive probability. It is easy to see, however, that whenever the cash investment is used with positive probability, *i.e.*$\lambda(\ell) > 0$ and when $\mathsf{E}[X(k)] < \infty$ for each k, then $\mathsf{E}[|Y(\boldsymbol{\lambda}|] < \infty$.

With the assumption that $\mathsf{E}[|Y(\boldsymbol{\lambda})|] < \infty$, the SLLN applies, *i.e.*,

$$\lim_{n\to\infty} \frac{1}{n} L_n(\boldsymbol{\lambda}) = \mathsf{E}[Y(\boldsymbol{\lambda})] \qquad \text{WP1.}$$

Since $W_n(\boldsymbol{\lambda}) = \exp\big(L_n(\boldsymbol{\lambda})\big)$, we then have

$$\lim_{n\to\infty} W_n(\boldsymbol{\lambda})^{1/n} = \exp \mathsf{E}[Y(\boldsymbol{\lambda})] \qquad \text{WP1.} \tag{9.128}$$

This indicates that $W_n(\boldsymbol{\lambda})$ is exponentially increasing in n at the rate $\mathsf{E}[|Y(\boldsymbol{\lambda})|]$, but it is difficult to be explicit about this interpretation. Thus we use the WLLN, which is easier to interpret in this case. For any $\epsilon > 0$ and $\delta > 0$, there is an n_o such that for all $n \geq n_o$,

$$\Pr\left\{ \Big| \frac{1}{n} L_n(\boldsymbol{\lambda}) - \mathsf{E}[Y(\boldsymbol{\lambda})] \Big| > \epsilon \right\} \leq \delta \qquad \text{for all } n \geq n_o.$$

The probability of the complementary event is then

$$\Pr\left\{ \mathsf{E}[Y(\boldsymbol{\lambda})] - \epsilon \leq \frac{1}{n} L_n(\boldsymbol{\lambda}) \leq \mathsf{E}[Y(\boldsymbol{\lambda})] + \epsilon \right\} > 1 - \delta$$

The event whose probability is being bounded is unchanged by multiplying each side of the inequalities by n and then exponentiating, yielding

$$\Pr\left\{ \exp\Big(n\mathsf{E}[Y(\boldsymbol{\lambda})] - n\epsilon\Big) \leq W_n(\boldsymbol{\lambda}) \leq \exp\Big(n\mathsf{E}[Y(\boldsymbol{\lambda})] + n\epsilon\Big) \right\} > 1 - \delta. \tag{9.129}$$

Thus, for any given fixed rebalancing portfolio $\boldsymbol{\lambda}$, there is an arbitrarily large probability that the investor's wealth changes exponentially in n within an epsilon tolerance of the exponent $\mathsf{E}[Y(\boldsymbol{\lambda})]$. The following theorem summarizes this result.

Theorem 9.10.1. *Let $W_n(\boldsymbol{\lambda})$ in (9.124) be an investor's wealth using a fixed rebalancing vector, $\boldsymbol{\lambda} = (\lambda(1), \dots, \lambda(\ell))^{\mathsf{T}}$ among ℓ investments. Let $Y_m(\boldsymbol{\lambda}) = \ln\big(\sum_k \lambda(k) X_m(j)\big)$ and assume $\mathsf{E}[|Y(\boldsymbol{\lambda})|] < \infty$. Then (9.129) holds.*

Assuming for the moment that a fixed rebalancing portfolio is a sensible strategy, we want to find the maximum exponential growth rate $\mathsf{E}[Y(\boldsymbol{\lambda})]$ over choices of $\boldsymbol{\lambda}$. We will find in the next subsection that understanding this maximization problem is also the key to understanding arbitrary portfolios rather than constant rebalancing portfolios.

The maximization of $\mathsf{E}[Y(\boldsymbol{\lambda})]$ over $\boldsymbol{\lambda}$ is an example of maximizing a concave function over a convex region. To spell this out, a region $\mathcal{R}$ of $\mathbb{R}^\ell$ is said to be convex if $\eta\boldsymbol{\lambda}_1 + (1-\eta)\boldsymbol{\lambda}_2 \in \mathcal{R}$ for all $\boldsymbol{\lambda}_1 \in \mathcal{R}$, $\boldsymbol{\lambda}_2 \in \mathcal{R}$, and $\eta \in (0,1)$, . Geometrically, this says that all points on the straight line segment between $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$ are also in $\mathcal{R}$. For the case here, the maximization of $\mathsf{E}[Y(\boldsymbol{\lambda}]$ is over the region $\mathcal{R} \in \mathbb{R}^\ell$ for which $\lambda(k) \geq 0$ for $1 \leq k \leq \ell$ and $\sum_k \lambda(k) = 1$. It is easy to verify that $\mathcal{R}$ is a convex region.

A function h from a convex region $\mathcal{R} \in \mathbb{R}^\ell$ to $\mathbb{R}$ is said to be concave over a convex region $\mathcal{R}$ if

$$h(\eta\boldsymbol{\lambda}_1 + (1-\eta)\boldsymbol{\lambda}_2) \geq \eta h(\boldsymbol{\lambda}_1) + (1-\eta)h(\boldsymbol{\lambda}_2). \tag{9.130}$$

for all $\boldsymbol{\lambda}_1 \in \mathcal{R}, \boldsymbol{\lambda}_2 \in \mathcal{R}$, and $\eta \in (0,1)$, Geometrically, this says that if we plot the value of the function along the straight line between any two arguments, the value of the function at each point lies on or above the value of a linear function joining the values at those two points. Concave functions are particularly easy to maximize, since if we start at any point and constantly move in a direction to increase the function, we eventually end at the maximum value, *i.e.*, the function does not have multiple peaks.

A function h is convex over a convex region if the inequality in (9.130) is reversed. Thus h is convex if and only if $-h$ is concave. Thus, everything one knows about convex functions translates immediately to concave functions and vice versa. For example concave functions are easy to maximize and convex functions are easy to minimize.

To see that the function $\mathsf{E}[Y(\boldsymbol{\lambda})]$ is concave over the specified region of $\boldsymbol{\lambda}$, we start by the observation that $\ln t$ has a negative second derivative for $t > 0$ and is thus concave for $t > 0$. Using this, the following equation shows that $Y(\boldsymbol{\lambda})$ is concave (*i.e.*, it is concave for each of its sample values),

$$\begin{aligned} Y(\eta\boldsymbol{\lambda}_1 + (1-\eta)\boldsymbol{\lambda}_2) &= \ln\left(\sum_k \big(\eta\lambda_1(k) + (1-\eta)\lambda_2(k)\big)X(k)\right) \\ &\geq \eta\ln\left(\sum_k \big(\lambda_1(k)\big)X(k)\right) + (1-\eta)\ln\left(\sum_k \big(\lambda_2(k)\big)X(k)\right). \end{aligned}$$

Finally, taking the expectation of each side, we see that $\mathsf{E}[Y(\boldsymbol{\lambda})]$ is concave over the specified region of $\boldsymbol{\lambda}$. The argument above was slightly careless about cases where $X(k)$ can take on the value 0 with positive probability. In this case, however, the region of $\mathbb{R}^\ell$ where $\mathsf{E}[Y(\boldsymbol{\lambda})] > -\infty$ is convex, and $\mathsf{E}[Y(\boldsymbol{\lambda})]$ is concave over that region.

There is a large literature about maximizing concave functions over convex regions, and there are necessary and sufficient conditions, called the Kuhn-Tucker conditions, for the maximum. These conditions are particularly simple for the case here in which the convex region is the probability simplex, *i.e.*, the region in which $\sum_k \lambda(k) = 1$ and each $\lambda(k) \geq 0$.

The necessary and sufficient conditions for a maximizing $\boldsymbol{\lambda}$ in this case are that for some constant c, the partial derivatives satisfy

$$\frac{\partial}{\partial\lambda(j)}\mathsf{E}[Y(\boldsymbol{\lambda})] \begin{cases} = c \quad ; & \text{for } \lambda(j) > 0 \\ \leq c \quad ; & \text{for } \lambda(j) = 0. \end{cases} . \tag{9.131}$$

Rather than proving this standard result, we give a convincing intuitive argument that explains it. If for a given $\boldsymbol{\lambda}$, there are investments i and j for which $\partial\mathsf{E}[Y(\boldsymbol{\lambda})]\,\partial\lambda(i) > \partial\mathsf{E}[Y(\boldsymbol{\lambda})]\,\partial\lambda(j)$, and if $\lambda(j) > 0$, then we can increase $\mathsf{E}[Y(\boldsymbol{\lambda})]$ by an incremental increase in $\lambda(i)$ and an incremental decrease in $\lambda(j)$ (thus maintaining $\sum_k \lambda(k) = 1$). This explains why the maximizing partial derivatives must be the same for all $\lambda(j)$ that are positive. It also explains the inequality where $\lambda(j) = 0$, since $\lambda(j)$ cannot be reduced below zero.

We now proceed to look at these partial derivatives. For any $\boldsymbol{\lambda}$ for which $\mathsf{E}[Y(\boldsymbol{\lambda})] > -\infty$, the partial derivative with respect to each $\lambda(j)$ is given by

$$\begin{aligned}\frac{\partial}{\partial\lambda(j)}\mathsf{E}[Y(\boldsymbol{\lambda})] &= \mathsf{E}\left[\frac{\partial}{\partial\lambda(j)}\ln\sum_k \lambda(k)X(k)\right] \\ &= \mathsf{E}\left[\frac{X(j)}{\sum_k \lambda(k)X(k)}\right].\end{aligned} \tag{9.132}$$

Note that the denominator here must be positive for each sample value of $X(1),\dots,X(\ell)$, since otherwise $\mathsf{E}[Y(\boldsymbol{\lambda})]$ would not exist. This, in conjunction with the optimization equations in (9.130), leads to the following theorem:

Theorem 9.10.2. *Let* $\mathcal{R} = \{\boldsymbol{\lambda} : \sum_k \lambda_k = 1 \text{ and } \lambda(k) \geq 0 \text{ for } 1 \leq k \leq \ell\ \}$. *The function* $\mathsf{E}[Y(\boldsymbol{\lambda})] = \mathsf{E}\left[\ln\left(\sum_k \lambda(k)X(k)\right)\right]$ *is maximized over* $\boldsymbol{\lambda} \in \mathcal{R}$ *if and only if* $\boldsymbol{\lambda}$ *satisfies the conditions:*

$$\mathsf{E}\left[\frac{X(j)}{\sum_k \lambda(k)X(k)}\right] \begin{cases} = 1 & ; \qquad \text{for } \lambda(j) > 0 \\ \leq 1 & ; \qquad \text{for } \lambda(j) = 0. \end{cases} \tag{9.133}$$

Proof: This results from substituting (9.132) into (9.131), with the exception that it that it also asserts that the constant c has the value 1. To evaluate c, multiply both sides of (9.132) by $\lambda(j)$ and sum over j, getting

$$\sum_j \lambda(j)\frac{\partial}{\partial\lambda(j)}\mathsf{E}[Y(\boldsymbol{\lambda})] = \sum_j \lambda(j)\mathsf{E}\left[\frac{X(j)}{\sum_k \lambda(k)X(k)}\right] = 1.$$

Substituting (9.131) into the left side of this, we see that $c = 1$. □

It is possible for the maximizing $\boldsymbol{\lambda}$ to be nonunique, and we denote any maximizing choice as $\boldsymbol{\lambda}^*$. The maximum value of $\mathsf{E}[Y(\boldsymbol{\lambda})]$ is of course unique, and will be denoted $\mathsf{E}[Y(\boldsymbol{\lambda}^*)]$. This quantity is important since it is the maximum exponential growth rate for fixed rebalancing strategies. We will soon see that in an important sense, it is also the maximum exponential growth rate for all rebalancing strategies.

Before looking at non-fixed rebalancing, there is a very simple but important question that must be answered. Under what circumstances is $\mathsf{E}[Y(\boldsymbol{\lambda}^*)] > 0$? Since cash is one of the investments and also has zero growth rate, $\mathsf{E}[Y(\boldsymbol{\lambda}^*)] \geq 0$ in all cases. If cash only is an optimizing investment, then we can take $\lambda^*(\ell) = 1$. Since the cash investment is unchanging, $X(\ell) = 1$, so $\sum_k \lambda^*(k)X(k) = 1$. From (9.133), then, $\mathsf{E}[X(j)] \leq 1$ for all j when $\lambda^*(\ell) = 1$. If $\mathsf{E}[X(j)] > 1$ for at least one j however, cash alone can not be the optimizing investment and an incremental move from cash alone toward such an investment increases $\mathsf{E}[Y(\boldsymbol{\lambda})]$ to a positive value. This proves the following theorem.

Theorem 9.10.3. *With the assumption of a cash investment, the maximizing exponential growth rate* $\mathsf{E}[Y(\boldsymbol{\lambda}^*)]$ *is positive if and only if* $\mathsf{E}[X(j)] > 1$ *for at least one investment* j.

As is so often the case in finding conditions for a maximum, the conditions are more valuable for finding other results than for actually computing the maximum. In the situation here, this result allows us essentially to show that this maximum over fixed rebalancing strategies is also the maximum over arbitrary rebalancing strategies.

In order to motivate why non-fixed rebalancing strategies might be important, it will be helpful to look at a particular non-fixed rebalancing strategy for Example 9.10.1. Suppose the investor splits the initial wealth into two piles. The first is then fully invested on successive epochs in the triple or nothing investment and the second is constantly rebalanced using the optimal rebalancing with $\lambda = 1/4$. We then have

$$\mathsf{E}[W_n] = \frac{1}{2} \cdot \left(\frac{3}{2}\right)^n + \frac{1}{2} \cdot \left(\frac{9}{8}\right)^n; \qquad \mathsf{E}[L_n] = \frac{n}{2}\ln(9/8) - \ln 2.$$

By using this investment strategy, which is not a fixed rebalancing strategy, we do essentially as well in maximizing $\mathsf{E}[W_n]$ as the go-for-broke strategy and essentially as well as the optimum constant rebalancing portfolio in maximizing $\mathsf{E}[L_n]$. You can have your cake and eat it too in this game. This example shows that the notion of optimality is murky in the investment game. The next subsection proves an optimality result, applying to all investment strategies. It achieves essentially the same thing as (9.129). It finds the maximum rate at which a portfolio can increase exponentially, subject to any given probability of failing to achieve that rate. The bottom line is that that the maximum exponential rate that can be achieved with non-vanishing probability is the maximum rate with fixed rebalancing of Theorem 9.10.2.

9.10.2 Portfolios with arbitrary rebalancing

As this investment game has been set up, the only choice of the investor is how to rebalance investments at the end of each epoch. In this section, the investor is allowed to use all the results prior to epoch n in the choice of allocation, $\boldsymbol{\lambda}_n = (\lambda_n(1), \dots, \lambda_n(\ell)^{\mathsf{T}}$, to be used during epoch n. Nothing fancy about stopping times is required here since $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots$, are statistically independent. The analysis to follow is based on comparing some arbitrary rebalancing strategy $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \dots$, to the optimal fixed $\boldsymbol{\lambda}$, now denoted $\boldsymbol{\lambda}^*$, of Theorem 9.10.2. Denote the wealth at the end of epoch n using the arbitrary rebalancing as W_n and that using $\boldsymbol{\lambda}^*$ as W_n^*. We compare W_n and W_n^* for the same sequence of changes $\{\boldsymbol{X}_n;\, n \geq 1\}$ for the individual investments.

Define $Z_n = W_n/W_n^*$ as the ratio of wealth using the arbitrary strategy to wealth using the optimal fixed $\boldsymbol{\lambda}^*$.

Theorem 9.10.4. *The sequence $\{Z_n;\, n \geq 1\}$ of rv's with $Z_n = W_n/W_n^*$ is a supermartingale.*

Proof: We can express W_n as

$$W_n = W_{n-1}\sum_j \lambda_n(j)X_n(j).$$

Expressing W_n^* in the same way,

$$Z_n = \frac{W_n}{W_n^*} = \frac{W_{n-1}}{W_{n-1|}^*} \frac{\sum_k \lambda_n(k) X_n(k)}{\sum_k \lambda^*(k) X_n(k)}$$

$$= Z_{n-1} \sum_j \lambda_n(j) \frac{X_n(j)}{\sum_k \lambda^*(k) X_n(k)}. \tag{9.134}$$

Taking the expectation of each side conditional on $Z_1, \ldots, Z_{n-1}$,

$$\mathsf{E}[Z_n \mid Z_1, \ldots, Z_{n-1}] = Z_{n-1} \sum_j \lambda_n(j) \mathsf{E}\left[\frac{X_n(j)}{\sum_k \lambda^*(k) X_n(k)}\right]$$

$$= Z_{n-1} \sum_j \lambda_n(j) \frac{\partial}{\partial \lambda(j)} \mathsf{E}[Y(\boldsymbol{\lambda}^*)] \tag{9.135}$$

$$\leq Z_{n-1}. \tag{9.136}$$

where we have used (9.132) in (9.135) and (9.131) in (9.136). Note that the possible dependence of $\lambda_n(j)$ on $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{n-1}$ does not affect this final inequality. Finally, $\mathsf{E}[|Z_n|]$ is finite for all n. In fact, since Z_n is nonnegative, the argument above shows that $\mathsf{E}[|Z_n|] \leq 1$ for all n. □

We can now use this submartingale property to show that there is little advantage to be gained by the flexibility of using arbitrary rebalancing strategies. Combining the nonnegative supermartingale inequality of (9.115) with the fact that $\mathsf{E}[Z_n] \leq 1$, we see that for any $a > 0$,

$$\Pr\Big\{\bigcup_{n\geq 1} \{Z_n \geq a\}\Big\} \leq \frac{1}{a}.$$

Since $Z_n = W_n/W_n^*$, this becomes

$$\Pr\Big\{\bigcup_{n\geq 1} \{W_n \geq aW_n^*\}\Big\} \leq \frac{1}{a}.. \tag{9.137}$$

This says that an arbitrary strategy cannot be much better than the optimal constant rebalancing strategy. The factor of a looks important until we remember than the optimal fixed strategy has a wealth that is growing exponentially with n, so a constant factor is negligible. We can then combine (9.137) with (9.129), also combining the factor a in (9.137) with the ϵ and δ in (9.129), to see that for any rebalancing strategy, the wealth satisfies

$$\Pr\Big\{W_n \leq \exp\Big(n\mathsf{E}[Y(\boldsymbol{\lambda}^*)] + n\epsilon\Big)\Big\} \geq 1 - \delta \qquad \text{for all } n \geq n_o. \tag{9.138}$$

This could be converted to a strong law statement, but it seems unnecessary. What we have shown is that the optimal fixed rebalancing strategy has a wealth that with high probability grows exponentially with $\mathsf{E}[Y(\boldsymbol{\lambda}^*)]$, and no other strategy has more than a vanishing probability of growing at a greater exponential rate.

9.11 Markov modulated random walks

Frequently it is useful to generalize random walks to allow some dependence between the variables being summed. The particular form of dependence here is the same as the Markov reward processes of Section 4.5. The treatment in Section 4.5 discussed only expected rewards, whereas the treatment here focuses on the random variables themselves. Let $\{Y_m;\, m \geq 0\}$ be a sequence of (possibly dependent) rv's, and let

$$\{S_n;\, n \geq 1\} \quad \text{where } S_n = \sum_{m=0}^{n-1} Y_m. \tag{9.139}$$

be the process of successive sums of these random variables. Let $\{X_n;\, n \geq 0\}$ be a Markov chain, and assume that each Y_n is a function of X_n and X_{n+1}. Conditional on X_n and X_{n+1}, Y_n is independent of $Y_{n-1}, \dots, Y_1$, and of X_i for all $i \neq n, n{-}1$. Assume that Y_n, conditional on X_n and X_{n+1} has a CDF $\mathsf{F}_{ij}(y) = \Pr\{Y_n \leq y \mid X_n = i, X_{n+1} = j\}$. Thus each rv Y_n depends only on the associated transition in the Markov chain, and this dependence is the same for all n.

The process $\{S_n;\, n \geq 1\}$ is called a *Markov modulated random walk*. If each Y_m is positive, it can be viewed as the sequence of epochs in a semi-Markov process. In the general case, Y_m is associated with the transition in the Markov chain from trial m to $m{+}1$, and S_n is the aggregate reward up to but not including time n. Let $\overline{Y}_{ij}$ denote $\mathsf{E}[Y_n \mid X_n = i, X_{n+1} = j]$ and $\overline{Y}_i$ denote $\mathsf{E}[Y_n \mid X_n = i]$. Let $\{P_{ij}\}$ be the set of transition probabilities for the Markov chain, so $\overline{Y}_i = \sum_j P_{ij}\overline{Y}_{ij}$. We may think of the process $\{Y_n;\, n \geq 0\}$ as evolving along with the Markov chain. The distributions of the variables Y_n are associated with the transitions from X_n to X_{n+1}, but the Y_n are otherwise independent random variables.

In order to define a martingale related to the process $\{S_n;\, n \geq 1\}$, we must subtract the mean reward from $\{S_n\}$ and must also compensate for the effect of the state of the Markov chain. The appropriate compensation factor turns out to be the relative-gain vector defined in Section 4.5.

For simplicity, consider only finite-state irreducible Markov chains with M states. Let $\pi = (\pi_1, \dots, \pi_{\mathsf{M}})$ be the steady-state probability vector for the chain, let $\overline{\boldsymbol{Y}} = (\overline{Y}_1, \dots, \overline{Y}_{\mathsf{M}})^{\mathsf{T}}$ be the vector of expected rewards, let $g = \pi\overline{Y}$ be the steady-state gain per unit time, and let $\boldsymbol{w} = (w_1, \dots, w_{\mathsf{M}})^{\mathsf{T}}$ be the relative-gain vector. From Theorem 4.5.1, $\boldsymbol{w}$ is the unique solution to

$$\boldsymbol{w} + g\boldsymbol{e} = \overline{\boldsymbol{Y}} + [P]\boldsymbol{w} \quad ; \quad w_1 = 0. \tag{9.140}$$

We assume a fixed starting state $X_0 = k$. As we now show, the process $Z_n;\, n \geq 1$ given by

$$Z_n = S_n - ng + w_{X_n} - w_k \quad ; \quad n \geq 1 \tag{9.141}$$

is a martingale. First condition on a given state, $X_{n-1} = i$.

$$\mathsf{E}\left[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1, X_{n-1} = i\right]. \tag{9.142}$$

Since $S_n = S_{n-1} + Y_{n-1}$, we can express Z_n as

$$Z_n = Z_{n-1} + Y_{n-1} - g + w_{X_n} - w_{X_{n-1}}. \tag{9.143}$$

Since $\mathsf{E}[Y_{n-1} \mid X_{n-1} = i] = \overline{Y}_i$ and $\mathsf{E}[w_{X_n} \mid X_{n-1} = i] = \sum_j P_{ij} w_j$, we have

$$\mathsf{E}[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1, X_{n-1} = i] = Z_{n-1} + \overline{Y}_i - g + \sum_j P_{ij} w_j - w_i. \tag{9.144}$$

From (9.140) the final four terms in (9.144) sum to 0, so

$$\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1, X_{n-1} = i] = Z_{n-1}. \tag{9.145}$$

Since this is valid for all choices of X_{n-1}, we have $\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] = Z_{n-1}$. Since the expected values of all the reward variables $\overline{Y}_i$ exist, we see that $\mathsf{E}[|Y_n|] < \infty$, so that $\mathsf{E}[|Z_n|] < \infty$ also. This verifies that $\{Z_n;\ n \geq 1\}$ is a martingale. It can be verified similarly that $\mathsf{E}[Z_1] = 0$, so $\mathsf{E}[Z_n] = 0$ for all $n \geq 1$.

In showing that $\{Z_n;\, n \geq 1\}$ is a martingale, we actually showed something a little stronger. That is, (9.145) is conditioned on X_{n-1} as well as $Z_{n-1}, \dots, Z_1$. In the same way, it follows that for all $n > 1$,

$$\mathsf{E}[Z_n \mid Z_{n-1}, X_{n-1}, Z_{n-2}, X_{n-2}, \dots, Z_1, X_1] = Z_{n-1}. \tag{9.146}$$

In terms of the gambling analogy, this says that $\{Z_n;\, n \geq 1\}$ is fair for each possible past sequence of states. A martingale $\{Z_n;\, n \geq 1\}$ with this property (*i.e.*, satisfying (9.146)) is said to be a *martingale relative to the joint process* $\{Z_n, X_n;\, n \geq 1\}$. We will use this martingale later to discuss threshold crossing problems for Markov modulated random walks. We shall see that the added property of being a martingale relative to $\{Z_n, X_n\}$ gives us added flexibility in defining stopping times.

As an added bonus to this example, note that if $\{X_n;\, n \geq 0\}$ is taken as the embedded chain of a Markov process (or semi-Markov process), and if Y_n is taken as the time interval from transition n to $n+1$, then S_n becomes the epoch of the nth transition in the process.

9.11.1 Generating functions for Markov random walks

Consider the same Markov chain and reward variables as in the previous example, again with $X_0 = k$, and assume that for each pair of states, i, j, the moment generating function

$$\mathsf{g}_{ij}(r) = \mathsf{E}[\exp(rY_n) \mid X_n = i, X_{n+1} = j]. \tag{9.147}$$

exists over some open interval (r_-, r_+) containing 0. Let $[\Gamma(r)]$ be the matrix with terms $P_{ij}\mathsf{g}_{ij}(r)$. Since $[\Gamma(r)]$ is an irreducible nonnegative matrix, Theorem 4.4.1 shows that $[\Gamma(r)]$ has a largest real eigenvalue, $\rho(r) > 0$, and an associated positive right eigenvector, $\boldsymbol{\nu}(r) = (\nu_1(r), \dots, \nu_{\mathsf{M}}(r))^\mathsf{T}$ that is unique within a scale factor. We now show that the process $\{M_n(r);\, n \geq 1\}$ defined by

$$M_n(r) = \frac{\exp(rS_n)\nu_{X_n}(r)}{\rho(r)^n \nu_k(r)}. \tag{9.148}$$

is a product type martingale for each $r \in (r_-, r_+)$. Since $S_n = S_{n-1} + Y_{n-1}$, we can express $M_n(r)$ as

$$M_n(r) = M_{n-1}(r)\,\frac{\exp(rY_{n-1})\,\nu_{X_n}(r)}{\rho(r)\nu_{X_{n-1}}(r)}. \tag{9.149}$$

The expected value of the ratio of $M_n(r)/M_{n-1}(r)$ in (9.149), conditional on $X_{n-1} = i$, is

$$\mathsf{E}\left[\frac{\exp(rY_{n-1})\nu_{X_n}(r)}{\rho(r)\nu_i(r)} \mid X_{n-1}{=}i\right]] = \frac{\sum_j P_{ij}\mathsf{g}_{ij}(r)\nu_j(r)}{\rho(r)\nu_i(r)} = 1. \tag{9.150}$$

where, in the last step, we have used the fact that $\nu(r)$ is an eigenvector of $[\Gamma(r)]$. Thus, $\mathsf{E}[M_n(r) \mid M_{n-1}(r), \dots, M_1(r), X_{n-1} = i] = M_{n-1}(r)$. Since this is true for all choices of i, the condition on $X_{n-1} = i$ can be removed and $\{M_n(r);\, n \geq 1\}$ is a martingale. Also, for $n > 1$,

$$\mathsf{E}[M_n(r) \mid M_{n-1}(r), X_{n-1}, \dots, M_1(r), X_1] = M_{n-1}(r). \tag{9.151}$$

so that $\{M_n(r);\, n \geq 1\}$ is also a martingale relative to the joint process $\{M_n(r), X_n;\, n \geq 1\}$.

It can be verified by the same argument as in (9.150) that $\mathsf{E}[M_1(r)] = 1$. It then follows that $\mathsf{E}[M_n(r)] = 1$ for all $n \geq 1$.

One of the uses of this martingale is to provide exponential upper bounds, similar to (9.9), to the probabilities of threshold crossings for Markov modulated random walks. Define

$$\widetilde{M}_n(r) = \frac{\exp(rS_n)\min_j(\nu_j(r))}{\rho(r)^n \nu_k(r)}. \tag{9.152}$$

Then $\widetilde{M}_n(r) \leq M_n(r)$, so $\mathsf{E}\left[\widetilde{M}_n(r)\right] \leq 1$. For any $\mu > 0$, the Markov inequality can be applied to $\widetilde{M}_n(r)$ to get

$$\Pr\left\{\widetilde{M}_n(r) \geq \mu\right\} \leq \frac{1}{\mu}\mathsf{E}\left[\widetilde{M}_n(r)\right] \leq \frac{1}{\mu}. \tag{9.153}$$

For any given α, and any given r, $0 \leq r < r_+$, we can choose $\mu = \exp(r\alpha)\rho(r)^{-n}\min_j(\nu_j(r))/\nu_k(r)$, and for $r > 0$. Combining (9.152) and (9.153),

$$\Pr\{S_n \geq \alpha\} \leq \rho(r)^n \exp(-r\alpha)\nu_k(r)/\min_j(\nu_j(r)). \tag{9.154}$$

This can be optimized over r to get the tightest bound in the same way as (9.9).

9.11.2 Stopping trials for martingales relative to a process

A martingale $\{Z_n;\, n \geq 1\}$ relative to a joint process $\{Z_n, X_n;\, n \geq 1\}$ was defined as a martingale for which (9.146) is satisfied, *i.e.*, $\mathsf{E}[Z_n \mid Z_{n-1}, X_{n-1}, \dots, Z_1, X_1] = Z_{n-1}$. In the same way, we can define a *submartingale or supermartingale* $\{Z_n;\, n \geq 1\}$ *relative to a joint process* $\{Z_n, X_n;\, n \geq 1\}$ as a submartingale or supermartingale satisfying (9.146) with the = sign replaced by $\geq$ or $\leq$ respectively. The purpose of this added complication is to make it easier to define useful stopping rules.

As generalized in Definition 5.5.2, a generalized stopping trial J for a sequence of pairs of rv's $(Z_1, X_1), (Z_2, X_2) \dots$, is a positive integer-valued rv such that, for each $n \geq 1$, $\mathbb{I}_{\{J=n\}}$ is a function of $Z_1, X_1, Z_2, X_2, \dots, Z_n, X_n$.

Theorems 9.8.1, 9.8.2 and Lemma 9.8.1 all carry over to martingales (submartingales or supermartingales) relative to a joint process. These theorems are stated more precisely in Exercises 9.29 to 9.34. To summarize them here, assume that $\{Z_n;\, n \geq 1\}$ is a martingale (submartingale or supermartingale) relative to a joint process $\{Z_n, X_n;\, n \geq 1\}$ and assume that J is a stopping trial for $\{Z_n;\, n \geq 1\}$ relative to $\{Z_n, X_n;\, n \leq 1\}$. Then the stopped process is a martingale (submartingale or supermartingale) respectively, (9.98 —9.100) are satisfied, and, for a martingale, $\mathsf{E}[Z_J] = \mathsf{E}[Z_1]$ is satisfied if and only if (9.104) is satisfied.

9.11.3 Markov modulated random walks with thresholds

We have now developed two martingales for Markov modulated random walks, both conditioned on a fixed initial state $X_0 = k$. The first, given in (9.141), is $\{Z_n = S_n - ng + w_{X_n} - w_k;\, n \geq 1\}$. Recall that $\mathsf{E}[Z_n] = 0$ for all $n \geq 1$ for this martingale. Given two thresholds, $\alpha > 0$ and $\beta < 0$, define J as the smallest n for which $S_n \geq \alpha$ or $S_n \leq \beta$. The indicator function $\mathbb{I}_{J=n}$ of $\{J = n\}$, is 1 if and only if $\beta < S_i < \alpha$ for $1 \leq i \leq n-1$ and either $S_n \geq \alpha$ or $S_n \leq \beta$. Since $S_i = Z_i + ig - w_{X_i} + w_k$, S_i is a function of Z_i and X_i, so the stopping trial is a function of both Z_i and X_i for $1 \leq i \leq n$. It follows that J is a stopping trial for $\{Z_n;\, n \geq 1\}$ relative to $\{Z_n, X_n;\, n \geq 1\}$. From Lemma 9.8.1, we can assert that $\mathsf{E}[Z_J] = \mathsf{E}[Z_1] = 0$ if (9.104) is satisfied, *i.e.*, if $\lim_{n\to\infty} \mathsf{E}[Z_n \mid J > n] \Pr\{J > n\} = 0$ is satisfied. Using the same argument as in Lemma 9.4.1, we can see that $\Pr\{J > n\}$ goes to 0 at least geometrically in n. Conditional on $J > n$, $\beta < S_n < \alpha$, so S_n is bounded independent of n. Also w_{X_n} is bounded, since the chain is finite state, and ng is linear in n. Thus $\mathsf{E}[Z_n \mid J > n]$ varies at most linearly with n, so (9.104) is satisfied, and

$$0 = \mathsf{E}[Z_J] = \mathsf{E}[S_J] - \mathsf{E}[J]\, g + \mathsf{E}[w_{X_n}] - w_k. \tag{9.155}$$

Recall that Wald's equality for random walks is $\mathsf{E}[S_J] = \mathsf{E}[J]\, \overline{X}$. For Markov modulated random walks, this is modified, as shown in (9.155), by the relative-gain vector terms.

The same arguments can be applied to the generating function martingale of (9.148). Again, let J be the smallest n for which $S_n \geq \alpha$ or $S_n \leq \beta$. As before, S_i is a function of $M_i(r)$ and X_i, so $\mathbb{I}_{J=n}$ is a function of $M_i(r)$ and X_i for $1 \leq i \leq n-1$. It follows that J is a stopping trial for $\{M_n(r);\, n \geq 1\}$ relative to $\{M_n(r), X_n;\, n \geq 1\}$. Next we need the following lemma:

Lemma 9.11.1. *For the martingale $\{M_n(r);\, n \geq 1\}$ relative to $\{M_n(r), X_n;\, n \geq 1\}$ defined in (9.148), where $\{X_n;\, n \geq 0\}$ is a finite-state Markov chain, and for the above stopping trial J,*

$$\lim_{n\to\infty} \mathsf{E}[M_n(r) \mid J > n] \Pr\{J > n\} = 0. \tag{9.156}$$

Proof: From lemma 4, slightly modified for the case here, there is a $\delta > 0$ such that for all states i, j, and all $n > 1$ such that $\Pr\{J = n, X_{n-1} = i, X_n = j\} > 0$,

$$\mathsf{E}\left[\exp(rS_n \mid J = n, X_{n-1} = i, X_n = j\right] \geq \delta. \tag{9.157}$$

Since the stopped process, $\{M_n^*(r);\, n \geq 1\}$, is a martingale, we have for each m,

$$1 = \mathsf{E}\left[M_m^*(r)\right] \geq \sum_{n=1}^{m} \frac{\mathsf{E}\left[\exp(rS_n)\nu_{X_n}(r) \mid J = n\right]}{\rho(r)^n \nu_k(r)}. \tag{9.158}$$

From (9.157), we see that there is some $\delta' > 0$ such that

$$\mathsf{E}\left[\exp(rS_n)\nu_{X_n}(r)\right] / \nu_k(r) \mid J = n] \geq \delta'$$

for all n such that $\Pr\{J = n\} > 0$. Thus (9.158) is bounded by

$$1 \geq \delta' \sum_{n \leq m} \rho(r)^n \Pr\{J = n\}.$$

Since this is valid for all m, it follows by the argument in the proof of theorem 9.4.1 that $\lim_{n\to\infty} \rho(r)^n \Pr\{J > n\} = 0$. This, along with (9.157), establishes (9.156), completing the proof. $\square$

From Lemma 9.8.1, we have the desired result:

$$\mathsf{E}\left[M_J(r)\right] = \mathsf{E}\left[\frac{\exp(rS_J)\nu_{X_J}(r)}{[\rho(r)]^J \nu_k(r)}\right] = 1; \qquad r_- < r < r_+. \tag{9.159}$$

This is the extension of the Wald identity to Markov modulated random walks, and is used in the same way as the Wald identity. As shown in Exercise 9.38, the derivative of (9.159), evaluated at $r = 0$, is the same as (9.155).

9.12 Summary

Each term in a random walk $\{S_n;\, n \geq 1\}$ where $S_n = X_1 + \cdots X_n$ is a sum of IID random variables, and thus the study of random walks is closely related to that of sums of IID rv's. The focus in random walks, however, as in most of the processes we have studied, is more in the relationship between the terms (such as which term first crosses a threshold) than in the individual terms. We started by showing that random walks are a generalization of renewal processes and are central to studying the queueing delay for G/G/1 queues.

A major focus of the chapter was on evaluating, bounding, and approximating the probabilities of very unlikely events, a topic known as large deviation theory. We started by studying the Chernoff bound to $\Pr\{S_n \geq \alpha\}$ for $\alpha > 0$ and $\mathsf{E}[X] < 0$. One of the insights gained here was that if a threshold at $\alpha >> 0$ is crossed, it is likely to be crossed at a trial $n \approx= \alpha/\gamma'(r^*)$, where r^* is the positive root of $\gamma(r)$. The overall probability of crossing a threshold at α was next elegantly upper bounded by $\exp(-r^*\alpha)$. This bound resulted from Wald's identity, which is essentially the statement that appropriately tilted versions of the underlying probabilities are in fact random variables themselves. For questions of typical behavior of a random walk, the mean and variance of a random variable are the major quantities of interest, but when interested in atypically large deviations, r^* is a major parameter of interest.

We next introduced martingales, submartingales and supermartingales. These are sometimes regarded as somewhat exotic topics in mathematics, but in fact they are very useful in a large variety of relatively simple processes. For example, all of the random walk issues of earlier sections can be treated as a special case of martingales, and martingales can be used to model both sums and products of random variables. We also showed how Markov modulated random walks can be treated as martingales.

Stopping trials, as first introduced in chapter 5, were then applied to martingales. We defined a stopped process $\{Z_n^*; n \geq 1\}$ to be the same as the original process $\{Z_n; n \geq 1\}$ up to the stopping point, and then constant thereafter. Theorems 9.8.1 and 9.8.2 showed that the stopped process has the same form (martingale, submartingale, or supermartingale) as the original process, and that the expected values $\mathsf{E}[Z_n^*]$ are between $\mathsf{E}[Z_1]$ and $\mathsf{E}[Z_n]$. We also looked at $\mathsf{E}[Z_J]$ and found that it is equal to $\mathsf{E}[Z_1]$ if and only if (9.104) is satisfied. The Wald identity can be viewed as $\mathsf{E}[Z_J] = \mathsf{E}[Z_1] = 1$ for the Wald martingale, $Z_n = \exp\{rS_n - n\gamma(r)\}$. We then found a similar identity for Markov modulated random walks.

The Kolmogorov inequalities were next developed. They are analogs of the Markov inequality and Chebyshev inequality, except they bound initial segments of submartingales and martingales rather than single rv's. They were used, first, to prove the SLLN with only a second moment and, second, the martingale convergence theorem.

A very simple model for allocating investments was developed both as a further example of the use of martingales and as a topic of intrinsic interest. A good reference for the mathematical development of this topic is [5].

A standard reference on random walks, and particularly on the analysis of overshoots is [Fel66]. Dembo and Zeitouni, [6] develop large deviation theory in a much more general and detailed way than the introduction here. The classic reference on martingales is [7], but [4] and [22] are more accessible.

9.13 Exercises

Exercise 9.1. Consider the simple random walk $\{S_n;\, n \geq 1\}$ of Section 9.1.1 with $S_n = X_1 + \cdots + X_n$ and $\Pr\{X_i = 1\} = p;\ \Pr\{X_i = -1\} = 1 - p;$ assume that $p \leq 1/2$.

a) Show that $\Pr\left\{\bigcup_{n\geq 1} \{S_n \geq k\}\right\} = \left[\Pr\left\{\bigcup_{n\geq 1} \{S_n \geq 1\}\right\}\right]^k$ for any positive integer k. Hint: Given that the random walk ever reaches the value 1, consider a new random walk starting at that time and explore the probability that the new walk ever reaches a value 1 greater than its starting point.

b) Find a quadratic equation for $y = \Pr\left\{\bigcup_{n\geq 1} \{S_n \geq 1\}\right\}$. Hint: explore each of the two possibilities immediately after the first trial.

c) For $p < 1/2$, show that the two roots of this quadratic equation are $p/(1-p)$ and 1. Argue that $\Pr\left\{\bigcup_{n\geq 1} \{S_n \geq 1\}\right\}$ cannot be 1 and thus must be $p/(1-p)$.

d) For $p = 1/2$, show that the quadratic equation in part c) has a double root at 1, and thus $\Pr\left\{\bigcup_{n\geq 1} \{S_n \geq 1\}\right\} = 1$. Note: this is the very peculiar case explained in Section 5.5.1.

e) For $p < 1/2$, show that $p/(1-p) = \exp(-r^*)$ where r^* is the unique positive root of $g(r) = 1$ where $g(r) = \mathsf{E}\left[e^{rX}\right]$.

Exercise 9.2. Consider a G/G/1 queue with IID arrivals $\{X_i;\ i \geq 1\}$, IID FCFS service times $\{Y_i;\ i \geq 0\}$, and an initial arrival to an empty system at time 0. Define $U_i = X_i - Y_{i-1}$ for $i \geq 1$. Consider a sample path where $(u_1, \ldots, u_6) = (1, -2, 2, -1, 3, -2)$.

a) Let $Z_i^6 = U_6 + U_{6-1} + \ldots + U_{6-i+1}$. Find the queueing delay for customer 6 as the maximum of the 'backward' random walk with elements $0, Z_1^6, Z_2^6, \ldots, Z_6^6$; sketch this random walk.

b) Find the queueing delay for customers 1 to 5.

c) Which customers start a busy period (*i.e.*, arrive when the queue and server are both empty)? Verify that if Z_i^6 maximizes the random walk in part a), then a busy period starts with arrival $6 - i$.

d) Now consider a forward random walk $V_n = U_1 + \cdots + U_n$. Sketch this walk for the sample path above and show that the queueing delay for each customer is the difference between two appropriately chosen values of this walk.

Exercise 9.3. A G/G/1 queue has a deterministic service time of 2 and interarrival times that are 3 with probability $p < 1/2$ and 1 with probability $1 - p$.

a) Find the distribution of W_1, the wait in queue of the first arrival after the beginning of a busy period.

b) Find the distribution of W_∞, the steady-state wait in queue.

c) Repeat parts a) and b) assuming the service times and interarrival times are exponentially distributed with rates μ and λ respectively.

Exercise 9.4. Let $U = V_1 + \cdots + V_n$ where $V_1, \dots, V_n$ are IID rv's with the MGF $g_V(s)$. Show that $g_U(s) = [g_V(s)]^n$. Hint: You should be able to do this simply in a couple of lines.

Exercise 9.5. Let $\{a_n;\, n \geq 1\}$ and $\{\alpha_; n \geq 1\}$ be sequences of numbers. For some b, assume that $a_n \leq \alpha_n e^{-bn}$ for all $n \geq 1$. For each of the following choices for α_n and arbitrary b, determine whether the bound is exponentially tight. Note: you may assume $b \geq 0$ if you wish, but it really does not matter.

a) $\alpha_n = \sum_{j=1}^{k} c_j n^{-j}$.

b) $\alpha_n = \exp(-\sqrt{n})$.

c) $\alpha_n = e^{-n^2}$.

What you are intended to learn from this is that an exponentially tight bound is not necessarily a reasonable approximation for large n. It is intended for sequences that are essentially exponentially decreasing in n and where one is satisfied with knowing the exponent without concern for the non-exponential coefficients.

Exercise 9.6. Define $\gamma(r)$ as $\ln[\mathsf{g}(r)]$ where $\mathsf{g}(r) = \mathsf{E}[\exp(rX)]$. Assume that X is discrete with possible outcomes $\{a_i;\, i \geq 1\}$, let p_i denote $\Pr\{X = a_i\}$, and assume that $\mathsf{g}(r)$ exists in some open interval (r_-, r_+) containing $r = 0$. For any given r, $r_- < r < r_+$, define a random variable X_r with the same set of possible outcomes $\{a_i;\, i \geq 1\}$ as X, but with a probability mass function $q_i = \Pr\{X_r = a_i\} = p_i \exp[a_i r - \gamma(r)]$. X_r is not a function of X, and is not even to be viewed as in the same probability space as X; it is of interest simply because of the behavior of its defined probability mass function. It is called a tilted random variable relative to X, and this exercise, along with Exercise 9.11 will justify our interest in it.

a) Verify that $\sum_i q_i = 1$.

b) Verify that $\mathsf{E}[X_r] = \sum_i a_i q_i$ is equal to $\gamma'(r)$.

c) Verify that $\mathsf{VAR}[X_r] = \sum_i a_i^2 q_i - (\mathsf{E}[X_r])^2$ is equal to $\gamma''(r)$.

d) Argue that $\gamma''(r) \geq 0$ for all r such that $\mathsf{g}(r)$ exists, and that $\gamma''(r) > 0$ if $\gamma''(0) > 0$.

e) Give a similar definition of X_r for a random variable X with a density, and modify parts a) to d) accordingly.

Exercise 9.7. a) Suppose Z is uniformly distributed over $[-b, +b]$. Find $g_Z(r), g_Z'(r)$, and $\gamma'(r)$ as a function of r.

b) Show that the interval over which $g_Z(r)$ exists is the real line, *i.e.*, $r_+ = \infty$ and $r_- = -\infty$. Show that $\lim_{r\to\infty} \gamma'(r) = b$.

c) Show that if $a > b$, then $\inf_{r\in I(X)} \gamma(r) - ra = -\infty$, so that the optimized Chernoff bound in (9.10) says that $\Pr\{S_n > na\} = 0$. Explain without any mathematics why $\Pr\{S_n > na\} = 0$ must be valid. Explain (with as little mathematical verbiage as possible) why an infimum rather than a minimum must be used in (9.10).

d) Show that for an arbitrary rv X, that if $r_+ = \infty$ and $\lim_{r\to\infty} \gamma'(r) = b$, then (9.10) says that $\Pr\{S_n > na\} = 0$ for $a > b$.

Exercise 9.8. Note that the MGF of the nonnegative exponential rv with density e^{-x} is $(1-r)^{-1}$ for $r < r_+ = 1$. It can be seen from this that, $\mathsf{g}(r_+)$ does not exist (*i.e.*, it is infinite) and both $\lim_{r\to r_+} \mathsf{g}(r) = \infty$ and $\lim_{r\to r_+} \mathsf{g}'(r) = \infty$, where the limit is over $r < r_+$. In this exercise, you are first to assume an arbitrary rv X for which $r_+ < \infty$ and $\mathsf{g}(r_+) = \infty$ and show that both $\lim_{r\to r_+} \mathsf{g}(r) = \infty$ and $\lim_{r\to r_+} \mathsf{g}'(r) = \infty$ You will then use this to show that if $\sup_{r\in I(X)} \mathsf{g}_X(r) < \infty$, then $\gamma(r_+) < \infty$ and $\mu(a)$ in (9.9) is given by $\mu(a) = \gamma(r_+) - r_+a$ for $a > \sup_{r\in I(X)} \mathsf{g}_X(r)$.

a) For X such that $r_+ < \infty$ and $\mathsf{g}(r_+) = \infty$, explain why

$$\lim_{A\to\infty} \int_0^A e^{xr_+}\, d\mathsf{F}(x) = \infty.$$

b) Show that for any $\epsilon > 0$ and any $A > 0$,

$$\mathsf{g}(r_+ - \epsilon) \geq e^{-\epsilon A} \int_0^A e^{xr_+}\, d\mathsf{F}(x).$$

c) Choose $A = 1/\epsilon$ and show that

$$\lim_{\epsilon\to 0} \mathsf{g}(r_+ - \epsilon) = \infty.$$

d) Show that $\lim_{\epsilon\to 0} \mathsf{g}'(r_+ - \epsilon) = \infty$.

e) Use parts a) to d) to show that if $r_+ < \infty$ and $\sup_{r\in I(X)} \mathsf{g}'_X(r) < \infty$, then $\mathsf{g}(r_+) < \infty$.

f) Show that if $r_+ < \infty$ and $\sup_{r\in I(X)} \mathsf{g}_X(r) < \infty$, then $\mu(a) = \gamma(r_+) - r_+a$ for $a > \sup_{r\in I(X)} \mathsf{g}_X(r)$.

Exercise 9.9 (Details in proof of Theorem 9.3.2). a) Show that the two appearances of ϵ in (9.24) can be repaced with two independent arbitrary positive quantities ϵ_1 and ϵ_2, getting

$$\Pr\big\{S_n \geq n(\gamma'(r) - \epsilon_1)\big\} \geq (1-\delta)\exp[-n(r\gamma'(r) + r\epsilon_2 - \gamma(r))].$$

Show that if this equation is valid for ϵ_1 and ϵ_2, then it is valid for all larger values of ϵ_1 and ϵ_2. Hint: Note that the left side of (9.24) is increasing in ϵ and the right side is decreasing.

b) Show that by increasing the n_o such that this equation is valid for all $n \geq n_o$, the factor of $(1-\delta)$ can be eliminated above.

c) For any $r \in (0, r_+)$, let δ_1 be an arbitrary number in $(0, r_+ - r)$, let $r_1 = r + \delta_1$, and let $\epsilon_1 = \gamma'(r_1) - \delta_1$. Show that there is an m such that for all $n \geq m$,

$$\Pr\{S_n \geq n\gamma'(r)\} \geq \exp\{-n\left[(r+\delta_1)\gamma'(r+\delta_1) + (r+\delta_1)\epsilon_2 - \gamma(r+\delta_1)\right]\}. \qquad (9.160)$$

Using the continuity of γ and its derivatives, show that for any $\epsilon > 0$, there is a $\delta_1 > 0$ so that the right side of (9.160) is greater than or equal to $\exp[-n(\gamma'(r) - r\gamma(r) + \epsilon)]$.

Exercise 9.10. In this exercise, we show that the optimized Chernoff bound is tight for the 3rd case in (9.12) as well as the first case. That is, we show that if $r_+ < \infty$ and $a > \sup_{r<r_+} \gamma'(r)$, then for any $\epsilon > 0$, $\Pr\{S_n \geq na\} \geq \exp\{n[\gamma(r_+) - r_+a - \epsilon]\}$ for all large enough n.

a) Let Y_i be the truncated version of X_i, truncated for some given b to $Y_i = X_i$ for $X_i \leq b$ and $Y_i = b$ otherwise. Let $W_n = Y_1 + \cdots + Y_n$. Show that $\Pr\{S_n \geq na\} \geq \Pr\{W_n \geq na\}$.

b) Let $\mathsf{g}_b(r)$ be the MGF of Y. Show that $\mathsf{g}_b(r) < \infty$ and that $\mathsf{g}_b(r)$ is non-decreasing in b for all $r < \infty$.

c) Show that $\lim_{b\to\infty} \mathsf{g}_b(r) = \infty$ for all $r > r_+$ and that $\lim_{b\to\infty} \mathsf{g}_b(r) = \mathsf{g}(r)$ for all $r \leq r_+$.

d) Let $\gamma_b(r) = \ln \mathsf{g}_b(r)$. Show that $\gamma_b(r) < \infty$ for all $r < \infty$. Also show that $\lim_{b\to\infty} \gamma_b(r) = \infty$ for $r > r_+$ and $\lim_{b\to\infty} \gamma_b(r) = \gamma(r)$ for $r \leq r_+$. Hint: Use b) and c).

e) Let $\gamma_b'(r) = \frac{\partial}{\partial r}\gamma_b(r)$ and let $\delta > 0$ be arbitrary. Show that for all large enough b, $\gamma_b'(r_+ + \delta) > a$. Hint: First show that $\gamma_b'(r_+ + \delta) \geq [\gamma_b(r_+ + \delta) - \gamma(r_+)]/\delta$.

f) Show that the optimized Chernoff bound for $\Pr\{W_n \geq na\}$ is exponentially tight for the values of b in part e). Show that the optimizing r is less than $r_+ + \delta$.

g) Show that for any $\epsilon > 0$ and all sufficiently large b,

$$\gamma_b(r) - ra \geq \gamma(r) - ra - \epsilon \geq \gamma(r_+) - r_+a - \epsilon \qquad \text{for } 0 < r \leq r_+.$$

Hint: Show that the convergence of $\gamma_b(r)$ to $\gamma(r)$ is uniform in r for $0 < r < r_+$.

h) Show that for arbitrary $\epsilon > 0$, there exists $\delta > 0$ and b_o such that

$$\gamma_b(r) - ra \geq \gamma(r_+) - r_+a - \epsilon \qquad \text{for } r_+ < r < r_+ + \delta \text{ and } b \geq b_o.$$

i) Note that if we put together parts g) and h) and d), and use the δ of part h) in d), then we have shown that the optimized exponent in the Chernoff bound for $\Pr\{W_n \geq na\}$ satisfies $\mu_b(a) \geq \gamma(r_+) - r_+a - \epsilon$ for sufficiently large b. Show that this means that $\Pr\{S_n \geq na\} \geq \gamma(r_+ - r_+a - 2\epsilon$ for sufficiently large n.

Exercise 9.11. Assume that X is discrete, with possible values $\{a_i; i \geq 1)$ and probabilities $\Pr\{X = a_i\} = p_i$. Let X_r be the corresponding tilted random variable as defined in Exercise 9.6. Let $S_n = X_1 + \cdots + X_n$ be the sum of n IID rv's with the distribution of X, and let $S_{n,r} = X_{1,r} + \cdots + X_{n,r}$ be the sum of n IID tilted rv's with the distribution of X_r. Assume that $\overline{X} < 0$ and that $r > 0$ is such that $\gamma(r)$ exists.

a) Show that $\Pr\{S_{n,r}{=}s\} = \Pr\{S_n{=}s\}\exp[sr - n\gamma(r)]$. Hint: read the text.

b) Find the mean and variance of $S_{n,r}$ in terms of $\gamma(r)$.

c) Define $a = \gamma'(r)$ and $\sigma_r^2 = \gamma''(r)$. Show that $\Pr\big\{|S_{n,r} - na| \le \sqrt{2n}\sigma_r\big\} > 1/2$. Use this to show that

$$\Pr\Big\{\mid S_n - na \mid\le \sqrt{2n}\,\sigma_r\ \Big\} > (1/2)\exp[-r(an + \sqrt{2n}\ \sigma_r) + n\gamma(r)].$$

d) Use this to show that for any ϵ and for all sufficiently large n,

$$\Pr\big\{S_n \ge n(\gamma'(r) - \epsilon)\big\} > \frac{1}{2}\exp[-rn(\gamma'(r) + \epsilon) + n\gamma(r)].$$

Exercise 9.12. a) Let $\boldsymbol{p} = (p_1, \dots, p_{\mathsf{M}})$ and $\widetilde{\boldsymbol{p}} = (\widetilde{p}_1, \dots, \widetilde{p}_{\mathsf{M}})$ be probability vectors. Show that the divergence $D(\widetilde{\boldsymbol{p}}||\boldsymbol{p}) \ge 0$ is 0 if $\widetilde{\boldsymbol{p}} = \boldsymbol{p}$).

b Let $\boldsymbol{u} = u_1, \dots, u_{\mathsf{M}}$ satisfy $\sum_j u_j = 0$ and show that $\widetilde{\boldsymbol{p}} + \epsilon\boldsymbol{u}$ is a probability vector over a sufficiently small region of ϵ around 0. Show that

$$\frac{\partial}{\partial\epsilon}D(\widetilde{\boldsymbol{p}} + \epsilon\boldsymbol{u}||\boldsymbol{p})|_{\epsilon=0} = 0.$$

c Show that

$$\frac{\partial^2}{\partial\epsilon^2}D(\widetilde{\boldsymbol{p}} + \epsilon\boldsymbol{u}||\boldsymbol{p}) \ge 0.$$

where $\widetilde{\boldsymbol{p}} + \epsilon\boldsymbol{u}$ is a probability vector with non-zero components. Explain why this implies that D is convex in $\widetilde{\boldsymbol{p}}$ and nonnegative over the region of probability vectors.

Exercise 9.13. Consider a random walk $\{S_n;\, n \ge 1\}$ where $S_n = X_1 + \cdots X_n$ and $\{X_i;\, i \ge 1\}$ is a sequence of IID exponential rv's with the PDF $\mathsf{f}(x) = \lambda e^{-\lambda x}$ for $x \ge 0$. In other words, the random walk is the sequence of arrival epochs in a Poisson process.

a) Show that for $\lambda a > 1$, the optimized Chernoff bound for $\Pr\{S_n \ge na\}$ is given by

$$\Pr\{S_n \ge na\} \le (a\lambda)^n e^{-n(a\lambda-1)}.$$

b) Show that the exact value of $\Pr\{S_n \ge na\}$ is given by

$$\Pr\{S_n \ge na\} = \sum_{i=0}^{n-1} \frac{(na\lambda)^i\, e^{-na\lambda}}{i!}.$$

c) By upper and lower bounding the quantity on the right above, show that

$$\frac{(na\lambda)^n\, e^{-na\lambda}}{n!\, a\lambda} \ \le\ \Pr\{S_n \ge na\}\ \le\ \frac{(na\lambda)^n e^{-na\lambda}}{n!(a\lambda - 1)}.$$

Hint: Use the term at $i = n - 1$ for the lower bound and note that the term on the right can be bounded by a geometric series starting at $i = n - 1$.

d) Use the Stirling bounds on $n!$ to show that

$$\frac{(a\lambda)^n\, e^{-n(a\lambda-1)}}{\sqrt{2\pi n}\, a\lambda \exp(1/12n)} \;\le\; \Pr\{S_n \ge na\} \;\le\; \frac{(a\lambda)^n e^{-n(a\lambda-1)}}{\sqrt{2\pi n}\,(a\lambda-1)}.$$

The point of this exercise is to demonstrate that the Chernoff bound is not only exponentially tight for this example but captures the important factors in the behavior of $\Pr\{S_n \ge na\}$.

Exercise 9.14. Consider a random walk with thresholds $\alpha > 0$, $\beta < 0$. We wish to find $\Pr\{S_J \ge \alpha\}$ in the absence of a lower threshold. Use the upper bound in (9.46) for the probability that the random walk crosses α before β.

a) Given that the random walk crosses β first, find an upper bound to the probability that α is crossed before a yet lower threshold at 2β is crossed.

b) Given that 2β is crossed before α, upper bound the probability that α is crossed before a threshold at 3β. Extending this argument to successively lower thresholds, find an upper bound to each successive term, and find an upper bound on the overall probability that α is crossed. By observing that β is arbitrary, show that (9.46) is valid with no lower threshold.

Exercise 9.15. This exercise verifies that Corollary 9.4.1 holds in the situation where $\gamma(r_+) < 0$ (see Figure 9.4). We have defined $r^* = r_+$ in this case.

a) Use the Wald identity at $r = r_+$ to show that

$$\Pr\{S_J \ge \alpha\}\, \mathsf{E}\left[\exp(r_+ S_J - J\gamma(r_+)) \mid S_J \ge \alpha\right] \;\le 1.$$

Hint: Look at the first part of the proof of Corollary 9.4.1.

b) Show that

$$\Pr\{S_J \ge \alpha\} \exp[r_+\alpha - \gamma(r_+)) \le 1].$$

c) Show that

$$\Pr\{S_J \ge \alpha\} \le \exp[-r_+\alpha + \gamma(r_+)] \qquad \text{and that } \Pr\{S_J \ge \alpha\} \le \exp[-r^*\alpha].$$

Note that the first bound above is a little stronger than the second, and note that lower bounding J by 1 is not necessarily a very weak bound, since this is the case where if α is crossed, it tends to be crossed for small J.

Exercise 9.16. a) Use Wald's equality to show that if $\overline{X} = 0$, then $\mathsf{E}[S_J] = 0$ where J is the time of threshold crossing with one threshold at $\alpha > 0$ and another at $\beta < 0$.

b) Obtain an expression for $\Pr\{S_J \ge \alpha\}$. Your expression should involve the expected value of S_J conditional on crossing the individual thresholds (you need not try to calculate these expected values).

c) Evaluate your expression for the case of a simple random walk.

d) Evaluate your expression when X has an exponential density, $\mathsf{f}_X(x) = a_1 e^{-\lambda x}$ for $x \ge 0$ and $\mathsf{f}_X(x) = a_2 e^{\mu x}$ for $x < 0$ and where a_1 and a_2 are chosen so that $\overline{X} = 0$.

Exercise 9.17. A random walk $\{S_n; n \geq 1\}$, with $S_n = \sum_{i=1}^{n} X_i$, has the following probability density for X_i

$$\mathsf{f}_X(x) = \begin{cases} \frac{e^{-x}}{e-e^{-1}} \ ; & -1 \leq x \leq 1 \\ \\ = 0 \ ; & \text{elsewhere.} \end{cases}$$

a) Find the values of r for which $\mathsf{g}(r) = \mathsf{E}[\exp(rX)] = 1$.

b) Let P_α be the probability that the random walk ever crosses a threshold at α for some $\alpha > 0$. Find an upper bound to P_α of the form $P_\alpha \leq e^{-\alpha A}$ where A is a constant that does not depend on α; evaluate A.

c) Find a lower bound to P_α of the form $P_\alpha \geq Be^{-\alpha A}$ where A is the same as in part b) and B is a constant that does not depend on α. Hint: keep it simple — you are not expected to find an elaborate bound. Also recall that $\mathsf{E}\left[e^{r^* S_J}\right] = 1$ where J is a stopping trial for the random walk and $\mathsf{g}(r^*) = 1$.

Exercise 9.18. Let $\{X_n; n \geq 1\}$ be a sequence of IID integer-vaued random variables with the probability mass function $\mathsf{p}_X(k) = Q_k$. Assume that $Q_k > 0$ for $|k| \leq 10$ and $Q_k = 0$ for $|k| > 10$. Let $\{S_n; n \geq 1\}$ be a random walk with $S_n = X_1 + \cdots + X_n$. Let $\alpha > 0$ and $\beta < 0$ be integer-vaued thresholds, let J be the smallest value of n for which either $S_n \geq \alpha$ or $S_n \leq \beta$. Let $\{S_n^*; n \geq 1\}$ be the stopped random walk; *i.e.*, $S_n^* = S_n$ for $n \leq J$ and $S_n^* = S_J$ for $n > J$. Let $\pi_i^* = \Pr\{S_J = i\}$.

a) Consider a Markov chain in which this stopped random walk is run repeatedly until the point of stopping. That is, the Markov chain transition probabilities are given by $P_{ij} = Q_{j-i}$ for $\beta < i < \alpha$ and $P_{i0} = 1$ for $i \leq \beta$ and $i \geq \alpha$. All other transition probabilities are 0 and the set of states is the set of integers $[-9 + \beta, 9 + \alpha]$. Show that this Markov chain is ergodic.

b) Let $\{\pi_i\}$ be the set of steady-state probabilities for this Markov chain. Find the set of probabilities $\{\pi_i^*\}$ for the stopping states of the stopped random walk in terms of $\{\pi_i\}$.

c) Find $\mathsf{E}[S_J]$ and $\mathsf{E}[J]$ in terms of $\{\pi_i\}$.

Exercise 9.19. a) Conditional on $X{=}0$ for the hypothesis testing problem, consider the random variables $Z_i = \ln[\mathsf{f}(Y_i|X{=}1)/\mathsf{f}(Y_i|X{=}0)]$. Show that r^*, the positive solution to $\mathsf{g}(r) = 1$, where $g(r) = \mathsf{E}[\exp(rZ_i)]$, is given by $r^* = 1$.

b) Assuming that Y is a discrete random variable (under each hypothesis), show that the tilted random variable Z_r with $r = 1$ has the PMF $\mathsf{p}_Y(y|X{=}1)$.

Exercise 9.20 (Proof of Stein's lemma). a) Use a power series expansion around $r = 0$ to show that $\gamma_0(r) - r\gamma'(r) = -r^2\gamma_0''(0)/2 + o(r^2)$.

b) Choose $r = n^{-1/3}$ in (9.61) and (9.62). Show that (9.61) can then be rewritten as

$$\begin{aligned} \Pr\{e(n) \mid X{=}0\} &\leq \exp\left\{n[-r^2\gamma_0''(0)/2 + o(r^2)]\right\} \\ &= \exp\left\{-n^{1/3}\gamma_0''(0)/2 + o(n^{1/3})]\right\} \end{aligned}$$

Thus $\Pr\{e(n) \mid X = 0\}$ approaches 0 with increasing n exponentially in $n^{1/3}$.

c) Using the same power series expansion for r in (9.62), show that

$$\begin{aligned} \Pr\{e(n) \mid X{=}1\} &\leq \exp\left\{n[r^2\gamma_0''(0)/2 + \gamma_0'(0) + o(r)]\right\} \\ &= \exp\left\{n\gamma_0'(0) + o(n^{2/3})]\right\} \end{aligned}$$

Thus $\lim_n (1/n) \ln \Pr\{e(n) \mid X{=}1\} = \gamma_0'(0) = \mathsf{E}[Z \mid X = 0]$.

d) Explain why the bound in (9.62) is exponentially tight given that $\Pr\{e(n) \mid X = 0\} \to 0$ as $n \to \infty$.

Exercise 9.21 (The secretary problem or marriage problem). This illustrates a very different type of sequential decision problem from those of Section 9.5. Alice is looking for a spouse and dates a set of n suitors sequentially, one per week. For simplicity, assume that Alice must make her decision to accept a suitor for marriage immediately after dating that suitor; she can not come back later to accept a formerly rejected suitor. Her decision must be based only on whether the current suitor is more suitable than all previous suitors. Mathematically, we can view the dating as continuing for all n weeks, but the choice at week m is a stopping rule. Assume that each suitor's suitability is represented by a real number and that all n numbers are different. Alice does not observe the suitability numbers, but only observes at each time m whether suitor m has the highest suitability so far. The suitors are randomly permuted before Alice dates them.

a) A reasonable algorithm for Alice is to reject the first k suitors (for some k to be optimized later) and then to choose the first suitor $m > k$ that is more suitable than all the previous $m - 1$ suitors. Find the probability q_k that Alice chooses the most suitable of all n when she uses this algorithm for given k. Hint 1: What is the probability (as a function of m) that the most suitable of all n is dated on week m. Hint 2: Given that the most suitable of all suitors is dated at week m, what is the probability that Alice chooses that suitor? For $m > k$, show how this involves the location of the most suitable choice from 1 to $m - 1$, conditional on m's overall optimality. Note: if no suitor is selected by the algorithm, it makes no difference whether suitor n is chosen at the end or no suitor is chosen.

b) Approximating $\sum_{i=1}^{j} 1/i$ by $\ln j$, show that for n and k large,

$$q_k \approx \frac{k}{n} \ln(n/k).$$

Ignoring the constraint that n and k are integers, show that the right hand side above is maximized at $k/n = e$ and that $\max_k q_k \approx 1/e$.

c) (Optional) Show that the algorithm of part a), optimized over k, is optimal over all algorithms (given the constraints of the problem). Hint: Let p_m be the maximum probability of choosing the optimal suitor given that no choice has been made before time m. Show that $p_m = \max\left[\frac{1}{n} + \frac{m-1}{m} p_{m+1},\ \ p_{m+1}\right]$; part of the problem here is understanding exactly what p_m means.

d) Explain why this is a poor model for choosing a spouse (or for making a best choice in a wide variety of similar problems). Caution: It is not enough to explain why this is not closely related to these real problems, You should also explain why this gives very little insight into such real problems.

Exercise 9.22. a) Suppose $\{Z_n;\, n \geq 1\}$ is a martingale. Verify (9.83); *i.e.*, $\mathsf{E}[Z_n] = \mathsf{E}[Z_1]$ for $n > 1$.

b) If $\{Z_n;\, n \geq 1\}$ is a submartingale, verify (9.88), and if a supermartingale, verify (9.89).

Exercise 9.23. Suppose $\{Z_n;\, n \geq 1\}$ is a martingale. Show that

$$\mathsf{E}\left[Z_m \mid Z_{n_i}, Z_{n_{i-1}}, \dots, Z_{n_1}\right] = Z_{n_i} \quad \text{for all } 0 < n_1 < n_2 < \dots < n_i < m.$$

Exercise 9.24. a) Assume that $\{Z_n;\, n \geq 1\}$ is a submartingale. Show that

$$\mathsf{E}[Z_m \mid Z_n, Z_{n-1}, \dots, Z_1] \geq Z_n \text{ for all } n < m.$$

b) Show that

$$\mathsf{E}\left[Z_m \mid Z_{n_i}, Z_{n_{i-1}}, \dots, Z_{n_1}\right] \geq Z_{n_i} \text{ for all } \ 0 < n_1 < n_2 < \dots < n_i < m.$$

c) Assume now that $\{Z_n;\, n \geq 1\}$ is a supermartingale. Show that parts a) and b) still hold with $\geq$ replaced by $\leq$.

Exercise 9.25. Let $\{Z_n = \exp[rS_n - n\gamma(r)];\, n \geq 1\}$ be the generating function martingale of (9.77) where $S_n = X_1 + \cdots + X_n$ and $X_1, \dots X_n$ are IID with mean $\overline{X} < 0$. Let J be the possibly-defective stopping trial for which the process stops after crossing a threshold at $\alpha > 0$ (there is no negative threshold). Show that $\exp[r^*\alpha]$ is an upper bound to the probability of threshold crossing by considering the stopped process $\{Z_n^*;\, n \geq 1\}$.

The purpose of this exercise is to illustrate that the stopped process can yield useful upper bounds even when the stopping trial is defective.

Exercise 9.26. This exercise uses a martingale to find the expected number of successive trials $\mathsf{E}[J]$ until some fixed pattern, $a_1, a_2, \dots, a_k$, of successive binary digits occurs within a sequence of IID binary random variables $X_1, X_2, \dots$ (see Example 4.5.1 and Exercise 5.35 for alternate approaches). We take the stopping time J to be the smallest n for which $(X_{n-k+1}, \dots, X_n) = (a_1, \dots, a_k)$. A mythical casino and sequence of gamblers who follow a prescribed strategy will be used to determine $\mathsf{E}[J]$. The outcomes of the plays (trials), $\{X_n;\, n \geq 1\}$ at the casino is a binary IID sequence for which $\Pr\{X_n = i\} = p_i$ for $i \in \{0, 1\}$

If a gambler places a bet s on 1 at play n, the return is s/p_1 if $X_n = 1$ and 0 otherwise. With a bet s on 0, the return is s/p_0 if $X_n = 0$ and 0 otherwise; *i.e.*, the game is fair.

a) Assume an arbitrary choice of bets on 0 and 1 by the various gamblers on the various trials. Let Y_n be the net gain of the casino on trial n. Show that $\mathsf{E}[Y_n] = 0$. Let $Z_n = Y_1 + Y_2 + \cdots + Y_n$ be the aggregate gain of the casino over n trials. Show that for any given pattern of bets, $\{Z_n;\, n \geq 1\}$ is a martingale.

b) In order to determine $\mathsf{E}[J]$ for a given pattern $a_1, a_2, \dots, a_k$, we program our gamblers to bet as follows:

i) Gambler 1 has an initial capital of 1 which is bet on a_1 at trial 1. If $X_1 = a_1$, the capital grows to $1/p_{a_1}$, all of which is bet on a_2 at trial 2. If $X_2 = a_2$, the capital grows to $1/(p_{a_1}p_{a_2})$, all of which is bet on a_3 at trial 3. Gambler 1 continues in this way until either losing at some trial (in which case he leaves with no money) or winning on k successive trials (in which case he leaves with $1/[p_{a_1} \dots p_{a_k}]$).

ii) Gambler ℓ, for each $\ell > 1$, follows the same strategy, but starts at trial ℓ. Note that if the pattern $a_1, \dots, a_k$ appears for the first time at trials $n{-}k{+}1, n{-}k{+}2, \dots, n$, *i.e.*, if $J = n$, then gambler $n - k + 1$ leaves at time n with capital $1/[p(a_1) \dots p(a_k)]$ and gamblers $j < n - k + 1$ have all lost their capital. We will come back later to investigate the capital at time n for gamblers $n - k + 2$ to n.

First consider the string $a_1{=}0$, $a_2{=}1$ with $k = 2$. Find the sample values of Z_1, Z_2, Z_3 for the sample sequence $X_1 = 1$, $X_2 = 0$, $X_3 = 1$, $\dots$. Note that gamblers 1 and 3 have lost their capital, but gambler 2 now has capital $1/p_0p_1$. Show that the sample value of the stopping time for this case is $J = 3$. Given an arbitrary sample value $n \geq 2$ for J, show that $Z_n = n - 1/p_0p_1$.

c) Find $\mathsf{E}[Z_J]$ from part a). Use this plus part b) to find $\mathsf{E}[J]$. Hint: This uses the special form of the solution in part b, not the Wald equality.

d) Repeat parts b) and c) using the string $(a_1, \dots, a_k) = (1, 1)$ and initially assuming $(X_1X_2X_3) = (011)$. Be careful about gambler 3 for $J = 3$. Show that $\mathsf{E}[J] = 1/p_1^2 + 1/p_1$.

e) Repeat parts b) and c) for $(a_1, \dots, a_k) = (1, 1, 1, 0, 1, 1)$.

f) Consider an arbitrary binary string $a_1, \dots, a_k$ and condition on $J = n$ for some $n \geq k$. Show that the sample capital of gambler ℓ is then equal to

- 0 for $\ell < n - k$.
- $1/[p_{a_1}p_{a_2} \cdots p_{a_k}]$ for $\ell = n - k + 1$.
- $1/[p_{a_1}p_{a_2} \cdots p_{a_i}]$ for $\ell = n - i + 1,\ 1 \leq i < k$ if $(a_1, \dots, a_i) = (a_{k-i+1}, \dots, a_k)$.
- 0 for $\ell = n - i + 1,\ 1 \leq i < k$ if $(a_1, \dots, a_i) \neq (a_{k-i+1}, \dots, a_k)$.

Verify that this general formula agrees with parts c), d), and e).

g) For a given binary string $a_1, \dots, a_k$, and each j, $1 \leq j \leq k$ let $\mathbb{I}_j = 1$ for $(a_1, \dots, a_j) = (a_{k-j+1}, \dots, a_k)$ and let $\mathbb{I}_j = 0$ otherwise. Show that

$$\mathsf{E}[J] = \sum_{i=1}^{k} \frac{\mathbb{I}_i}{\prod_{m=1}^{i} p_{a_m}}.$$

Note that this is the same as the final result in Exercise 4.28. The argument is shorter here, but more motivated and insightful there. Both approaches are useful and lead to simple generalizations.

Exercise 9.27. a) This exercise shows why the condition $\mathsf{E}[|Z_J|] < \infty$ is required in Lemma 9.8.1. Let $Z_1 = -2$ and, for $n \geq 1$, let $Z_{n+1} = Z_n[1 + X_n(3n+1)/(n+1)]$ where $X_1, X_2, \dots$ are IID and take on the values $+1$ and -1 with probability 1/2 each. Show that $\{Z_n;\, n \geq 1\}$ is a martingale.

b) Consider the stopping trial J such that J is the smallest value of $n > 1$ for which Z_n and Z_{n-1} have the same sign. Show that, conditional on $n < J$, $Z_n = (-2)^n/n$ and, conditional on $n = J$, $Z_J = -(-2)^n(2n-1)/(n^2 - n)$.

c) Show that $\mathsf{E}[|Z_J|]$ is infinite, so that $\mathsf{E}[Z_J]$ does not exist according to the definition of expectation, and show that $\lim_{n\to\infty} \mathsf{E}[Z_n | J > n] \Pr\{J > n\} = 0$.

Exercise 9.28. This exercise shows why the sup of a martingale can behave markedly differently from the maximum of an arbitrarily large number of the variables. More precisely, it shows that $\Pr\left\{\sup_{n\geq 1} Z_n \geq a\right\}$ can be unequal to $\Pr\left\{\bigcup_{n\geq 1}\{Z_n \geq a\}\right\}$.

a) Consider a martingale where Z_n can take on only the values 2^{-n-1} and $1 - 2^{-n-1}$, each with probability 1/2. Given that Z_n, conditional on Z_{n-1}, is independent of $Z_1, \dots Z_{n-2}$, find $\Pr\{Z_n | Z_{n-1}\}$ for each n so that the martingale condition is satisfied.

b) Show that $\Pr\left\{\sup_{n\geq 1} Z_n \geq 1\right\} = 1/2$ and show that $\Pr\{\bigcup_n \{Z_n \geq 1\}\} = 0$.

c Show that for every $\epsilon > 0$, $\Pr\left\{\sup_{n\geq 1} Z_n \geq a\right\} \leq \frac{\overline{Z}_1}{a-\epsilon}$. Hint: Use the relationship between $\Pr\left\{\sup_{n\geq 1} Z_n \geq a\right\}$ and $\Pr\{\bigcup_n \{Z_n \geq a\}\}$ while getting around the issue in part b).

d Use part c) to establish (9.111).

Exercise 9.29. Show that Theorem 9.7.1 is also valid for martingales relative to a joint process. That is, show that if h is a convex function of a real variable and if $\{Z_n;\, n \geq 1\}$ is a martingale relative to a joint process $\{Z_n, X_n;\, n \geq 1\}$, then $\{h(Z_n);\, n \geq 1\}$ is a submartingale relative to $\{h(Z_n), X_n;\, n \geq 1\}$.

Exercise 9.30. Show that if $\{Z_n;\, n \geq 1\}$ is a martingale (submartingale or supermartingale) relative to a joint process $\{Z_n, X_n;\, n \geq 1\}$ and if J is a stopping trial for $\{Z_n;\, n \geq 1\}$ relative to $\{Z_n, X_n;\, n \geq 1\}$, then the stopped process is a martingale (submartingale or supermartingale) respectively relative to the joint process.

Exercise 9.31. Prove Corollaries 9.9.2 to 9.9.4, *i.e.*, prove the following three statements:

a) Let $\{Z_n;\, n \geq 1\}$ be a submartingale with $\mathsf{E}\left[Z_n^2\right] < \infty$ for all $n \geq 1$. Then

$$\Pr\left\{\max_{1\leq n\leq m} |Z_n| \geq b\right\} \leq \frac{\mathsf{E}\left[Z_m^2\right]}{b^2}; \text{ for all integer } m \geq 2, \text{ and all } b > 0.$$

Hint: Fist show that $\{Z_n^2;\, n \geq 1\}$ is a submartingale.

b) [Kolmogorov's random walk inequality] Let $\{S_n;\, n \geq 1\}$ be a random walk with $S_n = X_1 + \cdots + X_n$ where $\{X_i;\, i \geq i\}$ is a set of IID random variables with mean $\overline{X}$ and variance σ^2. Then for any positive integer m and any $\epsilon > 0$,

$$\Pr\left\{\max_{1\leq n\leq m} |S_n - n\overline{X}| \geq m\epsilon\right\} \leq \frac{\sigma^2}{m\epsilon^2}.$$

Hint: Firt show that $\{S_n - n\overline{X};\, n \geq 1\}$ is a martingale.

c) Let $\{S_n;\, n \geq 1\}$ be a random walk, $S_n = X_1 + \cdots + X_n$, where each X_i has mean $\overline{X} < 0$ and semi-invariant moment generating function $\gamma(r)$. For any $r > 0$ such that $0 < \gamma(r) < \infty$, and for any $\alpha > 0$. show that

$$\Pr\left\{\max_{1\leq i\leq n} S_i \geq \alpha\right\} \leq \exp\{-r\alpha + n\gamma(r)\}.$$

Hint: First show that $\{e^{rS_n};\, n \geq 1\}$ is a submartingale.

Exercise 9.32. a) Let $\{Z_n;\, n \geq 1\}$ be the scaled branching process of Section 9.6.2 and assume that Y, the number of offspring of each element, has finite mean $\overline{Y}$ and finite variance σ^2. Assme that the population at time 0 is 1. Show that

$$\mathsf{E}\left[Z_n^2\right] = 1 + \sigma^2\left[\overline{Y}^{-2} + \overline{Y}^{-3} + \cdots + \overline{Y}^{-n-1}\right].$$

b) Assume that $\overline{Y} > 1$ and find $\lim_{n\to\infty} \mathsf{E}\left[Z_n^2\right]$. Show from this that the conditions for the martingale convergence theorem, Theorem 9.9.4, are satisfied.

Note: This condition is not satisfied if $\overline{Y} \leq 1$. The general martingale convergence theorem requires only a bounded first absolute moment, so it holds for $\overline{Y} \leq 1$

Exercise 9.33. Show that if $\{Z_n;\, n \geq 1\}$ is a martingale (submartingale or supermartingale) relative to a joint process $\{Z_n, X_n;\, n \geq 1\}$ and if J is a stopping trial for $\{Z_n;\, n \geq 1\}$ relative to $\{Z_n, X_n;\, n \geq 1\}$, then the stopped process satisfies (9.98), (9.99), or (9.100) respectively.

Exercise 9.34. Show that if $\{Z_n;\, n1\}$ is a martingale relative to a joint process $\{Z_n, X_n;\, n \geq 1\}$ and if J is a stopping trial for $\{Z_n;\, n \geq 1\}$ relative to $\{Z_n, X_n;\, n \geq 1\}$, then $\mathsf{E}[Z_J] = \mathsf{E}[Z_1]$ if and only if (9.104) is satisfied.

Exercise 9.35 (The double or quarter game). Consider an investment example similar to that of Example 9.10.1 in which there is only one investment other than cash. The ratio X of that investment value at the end of an epoch to that at the beginning is either 2 or 1/4, each with equal probability. Thus $\Pr\{X = 2\} = 1/2$ and $\Pr\{X = 1/4\} = 1/2$. Successive ratios, $X_1, X_2, \ldots,$ are IID.

a) In parts a) to c), ssume a fixed rebalancing strategy where a fraction λ is kept in the double or quarter investment and $1-\lambda$ is kept in cash. Find the expected wealth $\mathsf{E}[W_n(\lambda)]$ and the expected log wealth $\mathsf{E}[L_n(\lambda)]$ as a function of $\lambda \in [0,1]$ and $n \geq 1$. Assume unit initial wealth.

b) For $\lambda = 1$, find the PMF for $W_4(1)$ and give a brief explanation of why $\mathsf{E}[W_n(1)]$ is growing exponentially with n but $\mathsf{E}[L_n(1)]$ is decreasing linearly toward $-\infty$.

c) Find the value of λ^* that maximizes $\mathsf{E}[L_n(1)]$. Show that it satisfies the optimality conditions in (9.133).

d) Find the PMF of the rv Z_n/Z_{n-1} as given in (9.134) for any given λ_n.

Exercise 9.36 (Kelly's horse racing game). An interesting special case of the investment game is a horse racing strategy due to J. Kelly and described in more detail in [5]. There are $\ell - 1$ horses in a race and a fraction $\lambda(j)$ of the gambler's wealth is bet on horse j. Investment ℓ is cash and the fraction $\lambda(j)$ is kept in cash. As usual, $\sum_j \lambda_j = 1$ and $\lambda(j) \geq 0$ for $1 \leq j \leq \ell$. Let $r(j) > 0$ be the payoff reward, per unit bet on j, if j wins and let $p(j)$ be the probability that j wins. Let $X(j)$, be the corresponding rv, where $X(j) = r(j)$ with probability $p(j)$ (*i.e.*, if j wins) and $X(j) = 0$ with probability $1 - p(j)$. Thus $X(1), \dots, X(\ell-1)$ are highly dependent, since only one is nonzero. Note that $\sum_{j=1}^{\ell-1} p(j) = 1$ For cash, $X(\ell) = 1$ as usual.

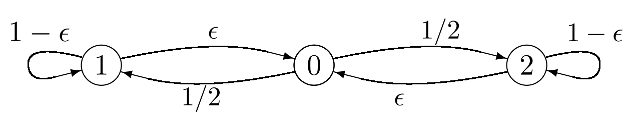
a) For any given gambling allocation $\boldsymbol{\lambda}$ find the expected wealth and the expected log wealth at the end of a race for a unit initial wealth.

b) Assume that a statistically identical sequence of races are run, *i.e.*, $\boldsymbol{X}_1, \boldsymbol{X}_2, \dots,$ are IID where each $\boldsymbol{X}_n = (X_n(1), \dots, X_n(j))^\mathsf{T}$. Assuming constant rebalancing after each race with a given allocation $\boldsymbol{\lambda}$, find the expected log wealth $\mathsf{E}[L_n(\boldsymbol{\lambda})]$ at the end of the nth race.

c) Assume there is an allocation $\boldsymbol{\lambda}$ that maximizes $\mathsf{E}[L_n(\boldsymbol{\lambda}^*)]$ but for which $\lambda(\ell) = 0$. Use the necessary and sufficient conditions for $(\boldsymbol{\lambda}^*)$ to show that $E\lambda^*(j)$ can be expressed solely as a function of $p(j)$ for that j and find that functional relation.

d) Using the relationship in part c), find the further relationship on $r_1, \dots, r_{\ell-1}$ needed for $\lambda^*(\ell)$ to be 0.

Exercise 9.37. Consider the Markov modulated random walk in the figure below. The random variables Y_n in this example take on only a single value for each transition, that value being 1 for all transitions from state 1, 10 for all transitions from state 2, and 0 otherwise. $\epsilon > 0$ is a very small number, say $\epsilon < 10^{-6}$.



a) Show that the steady-state gain per transition is $5.5/(1+\epsilon)$. Show that the relative-gain vector is $\boldsymbol{w} = (0, (\epsilon - 4.5)/[\epsilon(1+\epsilon)], (10\epsilon + 4.5)/[\epsilon(1+\epsilon)])$.

b) Let $S_n = Y_0 + Y_1 + \cdots + Y_{n-1}$ and take the starting state X_0 to be 0. Let J be the smallest value of n for which $S_n \geq 100$. Find $\Pr\{J = 11\}$ and $\Pr\{J = 101\}$. Find an estimate of $\mathsf{E}[J]$ that is exact in the limit $\epsilon \to 0$.

c) Show that $\Pr\{X_J = 1\} = (1 - 45\epsilon + \mathrm{o}(\epsilon))/2$ and that $\Pr\{X_J = 2\} = (1 + 45\epsilon + \mathrm{o}(\epsilon))/2$. Verify, to first order in ϵ that (9.155) is satisfied.

Exercise 9.38. Show that (9.155) results from taking the derivative of (9.159) and evaluating it at $r = 0$.

Exercise 9.39. Let $\{Z_n;\, n \geq 1\}$ be a martingale, and for some integer m, let $Y_n = Z_{n+m} - Z_m$.

a) Show that $\mathsf{E}[Y_n \mid Z_{n+m-1} = z_{n+m-1}, Z_{n+m-2} = z_{n+m-2}, \ldots, Z_m = z_m, \ldots, Z_1 = z_1] = z_{n+m-1} - z_m$.

b) Show that $\mathsf{E}[Y_n \mid Y_{n-1} = y_{n-1}, \ldots, Y_1 = y_1] = y_{n-1}$

c) Show that $\mathsf{E}[|Y_n|] < \infty$. Note that **b)** and **c)** show that $\{Y_n;\, n \geq 1\}$ is a martingale.

Exercise 9.40 (Continuation of continuous-time branching). This exercise views the continuous-time branching process of Exercise 7.15 as a stopped random walk. Recall that the process was specified there as a Markov process such that for each state j, $j \geq 0$, the transition rate to $j+1$ is $j\lambda$ and to $j-1$ is $j\mu$. There are no other transitions, and in particular, there are no transitions out of state 0, so that the Markov process is reducible. Recall that the embedded Markov chain is the same as the embedded chain of an M/M/1 queue except that there is no transition from state 0 to state 1.

a) To model the possible extinction of the population, convert the embedded Markov chain abve to a stopped random walk, $\{S_n;\, n \geq 0\}$. The stopped random walk starts at $S_0 = 0$ and stops on reaching a threshold at -1. Before stopping, it moves up by one with probability $\frac{\lambda}{\lambda+\mu}$ and downward by 1 with probability $\frac{\mu}{\lambda+\mu}$ at each step. Give the (very simple) relationship between the state X_n of the Markov chain and the state S_n of the stopped random walk for each $n \geq 0$.

b) Find the probability that the population eventually dies out as a function of λ and μ. Be sure to consider all three cases $\lambda > \mu$, $\lambda < \mu$, and $\lambda = \mu$.