7.6 Martingales and submartingales

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

$$\mathsf{E}[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1] = Z_{n-1}; \qquad \text{for all } n \geq 2 \tag{7.57}$$

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} 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 $n-1$.

There are two interpretations of (7.57); 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}). The student is encouraged to take the first viewpoint initially and to write out the expanded type of expression in cases of confusion.

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 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.

7.6.1 Simple examples of martingales

Example 7.6.1 (Zero-mean random walk). One example of a martingale is a zero-mean random walk, since if $Z_n = X_1 + X_2 + \dots + X_n$, where the X_i are IID and zero mean, then

$$\begin{aligned} \mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] &= \mathsf{E}[X_n + Z_{n-1} \mid Z_{n-1}, \dots, Z_1] &(7.58)\\ &= \mathsf{E}[X_n] + Z_{n-1} \ = \ Z_{n-1} &(7.59)\end{aligned}$$

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 $S_n = X_1 + \dots + X_n$ 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 random walks.

Example 7.6.2 (Sums of dependent zero-mean variables). Let $\{X_i;\ i \geq 1\}$ be a set of dependent random variables satisfying $\mathsf{E}[X_i \mid X_{i-1}, \dots, X_1] = 0$. Then $\{Z_n;\ n \geq 1\}$, where $Z_n = X_1 + \cdots + X_n$, is a zero mean martingale. To see this, note that

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

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 $n \geq 2$ and define $X_1 = Z_1$. 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.

Example 7.6.3 (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

$$\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] = \mathsf{E}[X_n Z_{n-1} \mid Z_{n-1}, \dots, Z_1] = \mathsf{E}[X_n]\, Z_{n-1} = Z_{n-1} \tag{7.60}$$

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

$$\mathrm{P}\{Z_n = 2^n\} = 2^{-n}; \qquad \mathrm{P}\{Z_n = 0\} = 1 - 2^{-n}; \qquad \mathsf{E}[Z_n] = 1 \tag{7.61}$$

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.

An important example of a product-form martingale is as follow: let $\{X_i;\ i \geq 1\}$ be an IID sequence, and let $\{S_n = X_1 + \ldots + X_n;\ n \geq 1\}$ be a random walk. Assume that the semi-invariant generating function $\gamma(r) = \ln\{\mathsf{E}[\exp(rX)]\}$ exists in some region of r around 0. For each $n \geq 1$, let Z_n be defined as

$$\begin{aligned} Z_n &= \exp\{rS_n - n\gamma(r\} & (7.62) \\ &= \exp\{rX_n - \gamma(r)\} \exp\{rS_{n-1} - (n-1)\gamma)r)\} \\ &= \exp\{rX_n - \gamma(r)\} Z_{n-1} & (7.63) \end{aligned}$$

Taking the conditional expectation of this,

$$\begin{aligned} \mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] &= \mathsf{E}[rX_n - \gamma(r)]\ \mathsf{E}[Z_{n-1} \mid Z_{n-1}, \dots, Z_1] \\ &= Z_{n-1} & (7.64) \end{aligned}$$

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.

7.6.2 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

$$\{S_n;\, n \geq 1\} \quad \text{where } S_n = \sum_{m=0}^{n-1} Y_m \tag{7.65}$$

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 can depend on X_n and X_{n+1}. Conditional on X_n and X_{n+1}, however, Y_n is independent of $Y_{n-1}, \dots, Y_1$, and of X_i for all $i \neq n\}$. Assume that Y_n, conditional on X_n and X_{n+1} has a distribution function $F_{ij}(y) = \mathrm{P}\{Y_n \leq y \mid X_n = i, X_{n+1} = j\}$. 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*. For each m, Y_m is associated with the transition in the Markov chain from time m to $m+1$, and S_n is the aggregate reward up to but not including time n. Let $\overline{X}_{ij}$ denote $\mathsf{E}[Y_n \mid X_n = i, X_{n+1} = j]$ and $\overline{X}_i$ denote $\mathsf{E}[Y_n \mid X_n = i]$. Let $\{P_{ij}\}$ be the set of transition probabilities for the Markov chain, so $\overline{Y}_i = \sum_j P_{ij}\overline{X}_{ij}$. 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 $\mathcal{J}$ states. Let $\pi = (\pi_{\mathbf{1}}, \dots, \pi_{\mathcal{J}})$ be the steady state probability vector for the chain, let $\overline{\boldsymbol{Y}} = (\overline{Y}_1, \dots, \overline{Y}_{\mathcal{J}}$ be the vector of expected rewards, let $g = \pi\overline{Y}$ be the steady state gain per unit time, and let $\boldsymbol{w} = (w_1, \dots, w_{\mathcal{J}})^T$ be the relative gain vector. From (4.41), $\boldsymbol{w}$ is the unique solution to

$$\boldsymbol{w} + g\boldsymbol{e} = \overline{\boldsymbol{Y}} + [P]\boldsymbol{w} \quad ; \quad w_1 = 0 \tag{7.66}$$

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

$$Z_n = S_n - ng + w_{X_n} - w_k \quad ; \quad n \geq 1 \tag{7.67}$$

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

$$\mathsf{E}[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1, X_{n-1} = i] \tag{7.68}$$

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

$$Z_n = Z_{n-1} + Y_{n-1} - g + w_{X_n} - w_{X_{n-1}} \tag{7.69}$$

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

$$\mathsf{E}[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1, X_{n-1} = i] = Z_{n-1} + \overline{Y}_i - g + \sum_j P_{ij} w_j - w_i \tag{7.70}$$

From (7.66) the final four terms in (7.70) sum to 0, so

$$\mathsf{E}\left[Z_n \mid Z_{n-1}, \dots, Z_1, X_{n-1} = i\right] = Z_{n-1}. \tag{7.71}$$

Since this is valid for all choices of X_{n-1}, we have $\mathsf{E}\left[Z_n \mid Z_{n-1}, \dots, Z_1\right] = Z_{n-1}$. Since the expected values of all the reward variables $\overline{Y}_i$ exist, we see that $\mathsf{E}\left[|Y_n|\right] < \infty$, so that $\mathsf{E}\left[|Z_n|\right] < \infty$ also. This verifies that $\{Z_n;\ n \geq 1\}$ is a martingale. It can be verified similarly that $\mathsf{E}\left[Z_1\right] = 0$, so $\mathsf{E}\left[Z_n\right] = 0$ 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, (7.71) is conditioned on X_{n-1} as well as $Z_{n-}, \dots, Z_1$. In the same way, it follows that for all $n > 1$,

$$\mathsf{E}\left[Z_n \mid Z_{n-1}, X_{n-1}, Z_{n-2}, X_{n-2}, \dots, Z_1, X_1\right] = Z_{n-1} \tag{7.72}$$

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 (7.72)) 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 rules.

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.

7.6.3 Generating functions for Markov random walks

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

$$g_{ij}(r) = \mathsf{E}\left[\exp(rY_n) \mid X_n = i, X_{n+1} = j\right] \tag{7.73}$$

exists over some open interval (r_-, r_+) containing 0. Let $[\Gamma(r)]$ be the matrix with terms $P_{ij}g_{ij}(r)$. Since $[\Gamma(r)]$ is an irreducible non-negative matrix, Theorem 4.6 shows that $[\Gamma(r)]$ has a largest real eigenvalue, $\rho(r) > 0$, and an associated positive right eigenvector, $\nu(r) = (\nu_1(r), \dots, \nu_J(r))^T$ that is unique within a scale factor. We now show that the process $\{M_n(r);\, n \geq 1\}$ defined by

$$M_n(r) = \frac{\exp(rS_n)\nu_{X_n}(r)}{\rho(r)^n \nu_k(r)} \tag{7.74}$$

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

$$M_n(r) = M_{n-1}(r)\, \frac{\exp(rY_{n-1})\, \nu_{X_n}(r)}{\rho(r)\nu_{X_{n-1}}(r)} \tag{7.75}$$

The expected value of the ratio in (7.75), conditional on $X_{n-1} = i$, is

$$E\left[\frac{\exp(rY_{n-1})\nu_{X_n}(r)}{\rho(r)\nu_i(r)} \mid X_{n-1}{=}i\right] = \frac{\sum_j P_{ij}g_{ij}(r)\nu_j(r)}{\rho(r)\nu_i(r)} = 1 \tag{7.76}$$

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 $n > 1$,

$$\mathsf{E}[M_n(r) \mid M_{n-1}(r), X_{n-1}, \dots, M_1(r), X_1] = M_{n-1}(r) \tag{7.77}$$

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 (7.76) 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 (7.16), to the probabilities of threshold crossings for Markov modulated random walks. Define

$$\widetilde{M}_n(r) = \frac{\exp(rS_n)\min_j(\nu_j(r))}{\rho(r)^n\nu_k(r)} \tag{7.78}$$

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

$$\mathrm{P}\left\{\widetilde{M}_n(r) \geq \mu\right\} \leq \frac{1}{\mu}\mathsf{E}\left[\widetilde{M}_n(r)\right] \leq \frac{1}{\mu} \tag{7.79}$$

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 (7.78) and (7.79),

$$\mathrm{P}\{S_n \geq \alpha\} \leq \rho(r)^n \exp(-r\alpha)\nu_k(r)/\min_j(\nu_j(r)) \tag{7.80}$$

This can be optimized over r to get the tightest bound in the same way as (7.16).

7.6.4 Scaled branching processes

A final example of a martingale is a "scaled down" version of a branching process $\{X_n;\, n \geq 0\}$. Recall from Section 5.2) 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 $Y_{i,n}$ of offspring 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, however. That is, define $Z_n = X_n/\overline{Y}^n$. It follows that

$$\mathsf{E}[Z_n \mid Z_{n-1}, \dots, Z_1] = E\left[\frac{X_n}{\overline{Y}^n} \mid X_{n-1}, \dots, X_1\right] = \frac{\overline{Y}X_{n-1}}{\overline{Y}^n} = Z_{n-1} \tag{7.81}$$

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 $n \to \infty$.

7.6.5 Isolating the past from the future in a martingale

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

$$E[Z_n \mid Z_i, Z_{i-1}, \dots, Z_1] = Z_i \tag{7.82}$$

Proof: By definition of a martingale, $\mathsf{E}[Z_{i+1} \mid Z_i, \dots, Z_1] = Z_i$. Next consider Z_{i+2}:

$$\begin{aligned}
\mathsf{E}[Z_{i+2}|Z_i, \dots, Z_1] &= \int_{z_{i+1}} \mathsf{E}[Z_{i+2} \mid Z_{i+1}{=}z_{i+1}, Z_i, \dots, Z_1]\, d\mathrm{P}\{Z_{i+1} \leq z_{i+1} \mid Z_i, \dots, Z_1\} \\
&= \int_{z_i+1} z_{i+1} d\mathrm{P}\{Z_{i+1} \leq z_{i+1} \mid Z_i, \dots, Z_1\} \\
&= \mathsf{E}[Z_{i+1} \mid Z_i, \dots, Z_1] = Z_i
\end{aligned}$$

The same argument can be applied successively from Z_{i+3} to Z_n, completing the proof. □

The above proof becomes more transparent if we view $\mathsf{E}[Z_{i+2}|Z_{i+1}, Z_i, \dots, Z_1]$ as a random variable that is a function of $Z_{i+1}, \dots, Z_1$. Since this random variable equals Z_{i+1}, we can take its expectation, conditional on $(Z_i, \dots, Z_1)$, to get $\mathsf{E}[Z_{i+1} \mid Z_i, \dots, Z_1] = Z_i$. The same argument can be used (see Exercise 7.19) to show that

$$\mathsf{E}[Z_n] = \mathsf{E}[Z_1] \qquad \text{for all } n > 1 \tag{7.83}$$

7.6.6 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 7.1. *A submartingale is an integer time stochastic process $\{Z_n;\, n \geq 1\}$ that satisfies the relations*

$$E[|Z_n|] < \infty \quad ; \quad E[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1] \geq Z_{n-1} \quad ; \quad n \geq 1. \tag{7.84}$$

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

$$E[|Z_n|] < \infty \quad ; \quad E[Z_n \mid Z_{n-1}, Z_{n-2}, \dots, Z_1] \leq Z_{n-1} \quad ; \quad n \geq 1 \tag{7.85}$$

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.[3]

Since a martingale satisfies both (7.84) and (7.85) 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 7.5, with the equality replaced by inequality, also applies to submartingales and supermartingales. That is, if $\{Z_n;\, n \geq 1\}$ is a submartingale, then

$$\mathsf{E}\left[Z_n \mid Z_i, Z_{i-1}, \dots, Z_1\right] \geq Z_i \quad ; \quad 1 \geq i < n, \tag{7.86}$$

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

$$\mathsf{E}\left[Z_n \mid Z_i, Z_{i-1}, \dots, Z_1\right] \leq Z_i \quad ; \quad 1 \leq i < n. \tag{7.87}$$

Equations (7.86) and (7.87) are verified in the same way as Lemma 7.5 (see Exercise 7.21). Similarly, the appropriate generalization of (7.83) is that if $\{Z_n;\, n \geq 1\}$ is a submartingale, then

$$\mathsf{E}\left[Z_n\right] \geq \mathsf{E}\left[Z_i\right] \quad ; \quad \text{for all } i,\ 1 \leq i < n. \tag{7.88}$$

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

$$\mathsf{E}\left[Z_n\right] \leq \mathsf{E}\left[Z_i\right] \quad ; \quad \text{for all } i,\ 1 \leq i < n. \tag{7.89}$$

A random walk $\{S_n;\, n \geq 1\}$ with $S_n = X_1 + \dots + X_n$ 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 7.6.4 (Convex functions of martingales). Figure 7.7 illustrates the graph of a convex function h of a real variable x. A function h of a real variable is defined to be *convex* if, for each point x_1, there is a real number c with the property that $h(x_1) + c(x - x_1) \leq h(x)$ for all x.

Geometrically, this says that every tangent to $h(x)$ lies on or below $h(x)$. If $h(x)$ has a derivative at x_1, then c is the value of that derivative and $h(x_1) + c(x - x_1)$ is the tangent line at x_1. If $h(x)$ has a discontinuous slope at x_1, then there might be many choices for c; for example, $h(x) = |x|$ is convex, and for $x_1 = 0$, one could choose any c in the range -1 to $+1$.

[3]The easiest way to remember the difference between a submartingale and a supermartingale is to remember that it is contrary to what common sense would dictate. That is, a submartingale is bigger than a supermartingale. Why this terminology became standard is a mystery.

A simple condition that implies convexity is a nonnegative second derivative everywhere. This is not necessary, however, and functions can be convex even when the first derivative does not exist everywhere. For example, the function $\gamma(r)$ in Figure 7.5 is convex even though it blows up at $r = r_+$.

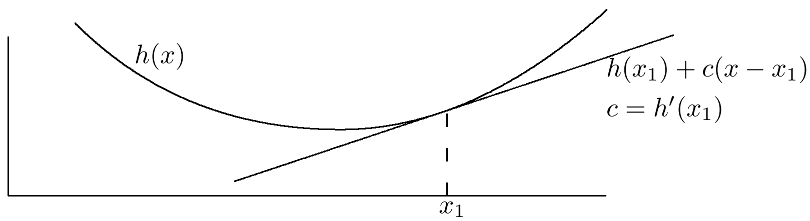

Figure 7.7: Convex functions: For each x_1, there is a value of c such that, for all x, $h(x_1) + c(x - x_1) \le h(x)$. If h is continuous at x_1, then c is the derivative of h at x_1.

Jensen's inequality states that if h is convex and X is a random variable with an expectation, then $h(\mathsf{E}[X]) \le \mathsf{E}[h(X)]$. To prove this, let $x_1 = \mathsf{E}[X]$ and choose c so that $h(x_1) + c(x - x_1) \le h(x)$. Using the random variable X in place of x and taking expected values of both sides, we get Jensen's inequality. 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 7.3. *If h is a convex function of a real variable, $\{Z_n;\, n \ge 1\}$ is a martingale, and $\mathsf{E}[|h(Z_n)|] < \infty$ 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 probabilities to get

$$\mathsf{E}[h(Z_n)|Z_{n-1}{=}z_{n-1}, \dots, Z_1{=}z_1] \ge h(\mathsf{E}[Z_n \mid Z_{n-1}{=}z_{n-1}, \dots, Z_1{=}z_1]) = h(z_{n-1}) \quad (7.90)$$

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, (7.90) holds, so that

$$\begin{aligned} \mathsf{E}[h(Z_n) \mid h(Z_{n-1}){=}h_{n-1}, \dots, h(Z_1){=}h_1] &\ge h(\mathsf{E}[Z_n \mid h(Z_{n-1}){=}h_{n-1}, \dots, h(Z_1){=}h_1]) \\ &= h(z_{n-1}) = h_{n-1} \end{aligned} \quad (7.91)$$

completing the proof. □

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

$$\{|Z_n|;\, n \ge 1\} \text{ is a submartingale} \quad (7.92)$$
$$\{Z_n^2;\, n \ge 1\} \text{ is a submartingale if } \mathsf{E}\left[Z_n^2\right] < \infty \quad (7.93)$$
$$\{\exp(rZ_n);\, n \ge 1\} \text{ is a submartingale for } r \text{ such that } \mathsf{E}[\exp(rZ_n)] < \infty. \quad (7.94)$$

A function of a real variable $h(x)$ is defined to be concave if $-h(x)$ is convex. It then follows from Theorem 7.3 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.

7.6.7 Exercises

Exercise 7.19. **a)** Suppose $\{Z_n;\, n \geq 1\}$ is a martingale. Verify (7.83); *i.e.*, $\mathsf{E}[Z_n] = \mathsf{E}[Z_1]$ for $n > 1$.

b) If $\{Z_n;\, n \geq 1\}$ is a submartingale, verify (7.88), and if a supermartingale, verify (7.89).

Exercise 7.20. Suppose $\{Z_n;\, n \geq 1\}$ is a martingale. Show that

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

Exercise 7.21. **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 7.22. Let $\{Z_n = \exp[rS_n - n\gamma(r)];\, n \geq 1\}$ be the generating function martingale of (7.62) where $S_n = X_1 + \dots + X_n$ and $X_1, \dots X_n$ are IID with mean $\overline{X} < 0$. Let N be the possibly defective stopping rule 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 rule is defective.

Exercise 7.23. We use martingales to find the expected number of trials $\mathsf{E}[N]$ before a fixed pattern, $a_1, a_2, \dots, a_k$, of binary digits occurs in a sequence of IID binary random variables $X_1, X_2, \dots$ (see Exercises 3.25 and 4.23 for alternate approaches). A mythical casino and set of gamblers who follow a prescribed strategy will be used to determine $\mathsf{E}[N]$. The casino has a game where, on the ith trial, gamblers bet money on either 1 or 0. After bets are placed, X_i above is used to select the outcome 0 or 1. Let $p(1) = \mathrm{P}\{X_i = 1\}$ and $p(0) = 1 - p_1 = \mathrm{P}\{X_i = 0\}$. If an amount s is bet on 1, the casino receives s if $X_i = 0$, and pays out $s/p(1) - s$ (plus returning the bet s) if $X_i = 1$. If s is bet on 0, the casino receives s if $X_1 = 1$, and pays out $s/p(0) - s$ (plus the bet s) if $X_i = 0$.

a) Assume an arbitrary pattern of bets by various gamblers on various trials (some gamblers might bet arbitrary amounts on 0 and some on 1 at any given trial). Let Y_i be the net gain of the casino on trial i. Show that $\mathsf{E}[Y_i] = 0$ (*i.e.*, show that the game is fair). Let $Z_n = Y_1 + Y_2 + \dots + Y_n$ be the aggregate gain of the casino over n trials. Show that for the given pattern of bets, $\{Z_n;\, n \geq 1\}$ is a martingale.

b) In order to determine $\mathsf{E}[N]$ 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 he wins, his capital grows to $1/p(a_1)$, which is bet on a_2 at trial 2. If he wins again, he bets his entire capital, $1/[p(a_1)p(a_2)]$, on a_3 at trial 3. He continues, at each trial i, to bet his entire capital on a_i until he loses at some trial (in which case he leaves with no money) or he wins on k successive trials (in which case he leaves with $1/[p(a_1)\dots p(a_k)]$.

ii) Gambler j, $j > 1$, follows exactly the same strategy but starts at trial j. Note that if the pattern $a_1, \dots, a_k$ appears for the first time at $N = n$, then gambler $n - k + 1$ leaves at time n with capital $1/[p(a_1)\dots p(a_k)]$ and gamblers $j < n - k + 1$ all lose their capital.

Suppose the string $(a_1, \dots, a_k)$ is $(0, 1)$. Let $\{Z_n; n \geq 1\}$ be the martingale of the casino's gain for the above gambling strategy. Given that $N = 3$ (*i.e.*, that $X_2 = 0$ and $X_3 = 1$), note that gambler 1 loses his money at either trial 1 or 2, gambler 2 leaves at time 3 with $1/[p(0)p(1)]$ and gambler 3 loses his money at time 3. Show that $Z_N = 3 - 1/[p(0)p(1)]$ given $N = 3$. Find Z_N given $N = n$ for arbitrary $n \geq 2$ (note that the condition $N = n$ uniquely specifies Z_N).

c) Find $\mathsf{E}[Z_N]$ from part **(a)**. Use this plus part **(b)** to find $\mathsf{E}[N]$.

d) Repeat parts **(b)** and **(c)** using the string $(a_1, \dots, a_k) = (1, 1)$. Be careful about gambler 3 for $N = 3$. Show that $\mathsf{E}[N] = 1/[p(1)p(1)] + 1/p(1)$

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