8.31. 6.14 8.30↑ If $X_0(\omega) = i_0, \ldots, X_n(\omega) = i_n$ for states $i_0, \ldots, i_n$, put $p_n(\omega) = \pi_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n}$, so that $p_n(\omega)$ is the probability of the observation observed. Show that $-n^{-1} \log p_n(\omega) \to h = -\sum_{ij} \pi_i p_{ij} \log p_{ij}$ with probability 1 if the chain is finite, irreducible, and aperiodic. Extend to this case the notions of source, entropy, and asymptotic equipartition.

8.32. A sequence $\{X_n\}$ is a Markov chain *of second order* if $P[X_{n+1} = j \| X_0 = i_0, \ldots, X_n = i_n] = P[X_{n+1} = j \| X_{n-1} = i_{n-1}, X_n = i_n] = p_{i_{n-1} i_n : j}$. Show that nothing really new is involved because the sequence of pairs (X_n, X_{n+1}) is an ordinary Markov chain (of first order). Compare Problem 8.29. Generalize this idea into chains of order r.

8.33. Consider a chain on $S = \{0, 1, \ldots, r\}$, where 0 and r are absorbing states and $p_{i,i+1} = p_i > 0$, $p_{i,i-1} = q_i = 1 - p_i > 0$ for $0 < i < r$. Identify state i with a point z_i on the line, where $0 = z_0 < \cdots < z_r$ and the distance from z_i to z_{i+1} is q_i/p_i times that from z_{i-1} to z_i. Given a function φ on S, consider the associated function $\hat{\varphi}$ on $[0, z_r]$ defined at the z_i by $\hat{\varphi}(z_i) = \varphi(i)$ and in between by linear interpolation. Show that φ is excessive if and only if $\hat{\varphi}$ is concave. Show that the probability of absorption in r for initial state i is t_{i-1}/t_{r-1}, where $t_i = \sum_{k=0}^{i} q_1 \cdots q_k / p_1 \cdots p_k$. Deduce (7.7). Show that in the new scale the expected distance moved on each step is 0.

8.34. Suppose that a finite chain is irreducible and aperiodic. Show by Theorem 8.9 that an excessive function must be constant.

8.35. *A zero-one law.* Let the state space S contain s points, and suppose that $\epsilon_n = \sup_{ij} |p_{ij}^{(n)} - \pi_j| \to 0$, as holds under the hypotheses of Theorem 8.9. For $a \le b$, let $\mathscr{G}_a^b$ be the σ-field generated by the sets $[X_a = u_a, \ldots, X_b = u_b]$. Let $\mathscr{T}_a = \sigma(\bigcup_{b=a}^{\infty} \mathscr{G}_a^b)$ and $\mathscr{T} = \bigcap_{a=1}^{\infty} \mathscr{T}_a$. Show that $|P(A \cap B) - P(A)P(B)| \le s(\epsilon_n + \epsilon_{b+n})$ for $A \in \mathscr{G}_0^b$ and $B \in \mathscr{G}_{b+n}^{b+m}$; the ϵ_{b+n} can be suppressed if the initial probabilities are the stationary ones. Show that this holds for $A \in \mathscr{G}_0^b$ and $B \in \mathscr{T}_{b+n}$. Show that $C \in \mathscr{T}$ implies that $P(C)$ is either 0 or 1.

8.36[†] Alter the chain in Example 8.13 so that $q_0 = 1 - p_0 = 1$ (the other p_i and q_i still positive). Let $\beta = \lim_n p_1 \cdots p_n$ and assume that $\beta > 0$. Define a payoff function by $f(0) = 1$ and $f(i) = 1 - f_{i0}$ for $i > 0$. If $X_0, \ldots, X_n$ are positive, put $\sigma_n = n$; otherwise let σ_n be the smallest k such that $X_k = 0$. Show that $E_i[f(X_{\sigma_n})] \to 1$ as $n \to \infty$, so that $v(i) \equiv 1$. Thus the support set is $M = \{0\}$, and for an initial state $i > 0$ the probability of ever hitting M is $f_{i0} < 1$.

For an arbitrary finite stopping time τ, choose n so that $P_i[\tau < n = \sigma_n] > 0$. Then $E_i[f(X_\tau)] \le 1 - f_{i+n,0} P_i[\tau < n = \sigma_n] < 1$. Thus no strategy achieves the value $v(i)$ (except of course for $i = 0$).

8.37. ↑ Let the chain be as in the preceding problem, but assume that $\beta = 0$, so that $f_{i0} = 1$ for all i. Suppose that $\lambda_1, \lambda_2, \ldots$ exceed 1 and that $\lambda_1 \cdots \lambda_n \to \lambda < \infty$; put $f(0) = 0$ and $f(i) = \lambda_1 \cdots \lambda_{i-1} / p_1 \cdots p_{i-1}$. For an arbitrary (finite) stopping time τ, the event $[\tau = n]$ must have the form $[(X_0, \ldots, X_n) \in I_n]$ for some set I_n of $(n+1)$-long sequences of states. Show that for each i there is at most one $n \ge 0$ such that $(i, i+1, \ldots, i+n) \in I_n$. If there is no such n, then $E_i[f(X_\tau)] = 0$. If there is one, then

$$E_i[f(X_\tau)] = P_i[(X_0, \ldots, X_n) = (i, \ldots, i+n)] f(i+n),$$

and hence the only possible values of $E_i[f(X_\tau)]$ are

$$0, \quad f(i), \quad p_i f(i+1) = f(i)\lambda_i, \quad p_i p_{i+1} f(i+2) = f(i)\lambda_i \lambda_{i+1}, \ldots.$$

Thus $v(i) = f(i)\lambda / \lambda_1 \cdots \lambda_{i-1}$ for $i \ge 1$; no strategy this value. The support set is $M = \{0\}$, and the hitting time τ_0 for M is finite, but $E_i[f(X_{\tau_0})] = 0$.

8.38. 5.12↑ Consider an irreducible, aperiodic, positive persistent chain. Let τ_j be the smallest n such that $X_n = j$, and let $m_{ij} = E_i[\tau_j]$. Show that there is an r such that $p = P_j[X_1 \ne j, \ldots, X_{r-1} \ne j, X_r = i]$ is positive; from $f_{jj}^{(n+r)} \ge p f_{ij}^{(n)}$ and $m_{jj} < \infty$, conclude that $m_{ij} < \infty$ and $m_{ij} = \sum_{n=0}^{\infty} P_i[\tau_j > n]$. Starting from $p_{ij}^{(t)} = \sum_{s=1}^{t} f_{ij}^{(s)} p_{jj}^{(t-s)}$, show that

$$\sum_{t=1}^{n} \left(p_{ij}^{(t)} - p_{jj}^{(t)}\right) = 1 - \sum_{m=0}^{n} p_{jj}^{(n-m)} P_i[\tau_j > m].$$

Use the M-test to show that

$$\pi_j m_{ij} = 1 + \sum_{n=1}^{\infty} \left(p_{jj}^{(n)} - p_{ij}^{(n)}\right).$$

If $i = j$, this gives $m_{jj} = 1/\pi_j$ again; if $i \ne j$, it shows how in principle m_{ij} can be calculated from the transition matrix and the stationary probabilities.

[†]The final three problems in this section involve expected values for random variables with infinite range.

SECTION 9. LARGE DEVIATIONS AND THE LAW OF THE ITERATED LOGARITHM*

It is interesting in connection with the strong law of large numbers to estimate the rate at which S_n/n converges to the mean m. The proof of the strong law used upper bounds for the probabilities $P[|S_n - m| \ge \alpha]$ for large α. Accurate upper and lower bounds for these probabilities will lead to the law of the iterated logarithm, a theorem giving very precise rates for $S_n/n \to m$.

The first concern will be to estimate the probability of large deviations from the mean, which will require the method of moment generating functions. The estimates will be applied first to a problem in statistics and then to the law of the iterated logarithm.

*This section may be omitted.

Moment Generating Functions

Let X be a simple random variable asssuming the distinct values $x_1, \dots, x_l$ with respective probabilities $p_1, \dots, p_l$. Its *moment generating function* is

$$M(t) = E[e^{tX}] = \sum_{i=1}^{l} p_i e^{tx_i}. \tag{9.1}$$

(See (5.19) for expected values of functions of random variables.) This function, defined for all real t, can be regarded as associated with X itself or as associated with its distribution—that is, with the measure on the line having mass p_i at x_i (see (5.12)).

If $c = \max_i |x_i|$, the partial sums of the series $e^{tX} = \sum_{k=0}^{\infty} t^k X^k / k!$ are bounded by $e^{|t|c}$, and so the corollary to Theorem 5.4 applies:

$$M(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} E[X^k]. \tag{9.2}$$

Thus $M(t)$ has a Taylor expansion, and as follows from the general theory [A29], the coefficient of t^k must be $M^{(k)}(0)/k!$ Thus

$$E[X^k] = M^{(k)}(0). \tag{9.3}$$

Furthermore, term-by-term differentiation in (9.1) gives

$$M^{(k)}(t) = \sum_{i=1}^{l} p_i x_i^k e^{tx_i} = E[X^k e^{tX}];$$

taking $t = 0$ here gives (9.3) again. Thus the moments of X can be calculated by successive differentiation, whence $M(t)$ gets its name. Note that $M(0) = 1$.

Example 9.1. If X assumes the values 1 and 0 with probabilities p and $q = 1 - p$, as in Bernoulli trials, its moment generating function is $M(t) = pe^t + q$. The first two moments are $M'(0) = p$ and $M''(0) = p$, and the variance is $p - p^2 = pq$. ■

If $X_1, \dots, X_n$ are independent, then for each t (see the argument following (5.10)), $e^{tX_1}, \dots, e^{tX_n}$ are also independent. Let M and $M_1, \dots, M_n$ be the respective moment generating functions of $S = X_1 + \cdots + X_n$ and of $X_1, \dots, X_n$; of course, $e^{tS} = \prod_i e^{tX_i}$. Since by (5.25) expected values multiply for independent random variables, there results the fundamental relation

$$M(t) = M_1(t) \cdots M_n(t). \tag{9.4}$$

This is an effective way of calculating the moment generating function of the sum S. The real interest, however, centers on the distribution of S, and so it is important to know that distributions can in principle be recovered from their moment generating functions.

Consider along with (9.1) another finite exponential sum $N(t) = \sum_j q_j e^{ty_j}$, and suppose that $M(t) = N(t)$ for all t. If $x_{i_0} = \max x_i$ and $y_{j_0} = \max y_j$, then $M(t) \sim p_{i_0} e^{tx_{i_0}}$ and $N(t) \sim q_{j_0} e^{ty_{j_0}}$ as $t \to \infty$, and so $x_{i_0} = y_{j_0}$ and $p_{i_0} = q_{i_0}$. The same argument now applies to $\sum_{i \neq i_0} p_i e^{tx_i} = \sum_{j \neq j_0} q_j e^{ty_j}$, and it follows inductively that with appropriate relabeling, $x_i = y_i$ and $p_i = q_i$ for each i. Thus *the function* (9.1) *does uniquely determine the x_i and p_i.*

Example 9.2. If $X_1, \dots, X_n$ are independent, each assuming values 1 and 0 with probabilities p and q, then $S = X_1 + \cdots + X_n$ is the number of successes in n Bernoulli trials. By (9.4) and Example 9.1, S has the moment generating function

$$E[e^{tS}] = (pe^t + q)^n = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} e^{tk}.$$

The right-hand form shows this to be the moment generating function of a distribution with mass $\binom{n}{k} p^k q^{n-k}$ at the integer k, $0 \le k \le n$. The uniqueness just established therefore yields the standard fact that $P[S = k] = \binom{n}{k} p^k q^{n-k}$. ■

The *cumulant generating function of X* (or of its distribution) is

$$C(t) = \log M(t) = \log E[e^{tX}]. \tag{9.5}$$

(Note that $M(t)$ is strictly positive.) Since $C' = M'/M$ and $C'' = (MM'' - (M')^2)/M^2$, and since $M(0) = 1$,

$$C(0) = 0, \qquad C'(0) = E[X], \qquad C''(0) = \operatorname{Var}[X]. \tag{9.6}$$

Let $m_k = E[X^k]$. The leading term in (9.2) is $m_0 = 1$, and so a formal expansion of the logarithm in (9.5) gives

$$C(t) = \sum_{v=1}^{\infty} \frac{(-1)^{v+1}}{v} \left(\sum_{k=1}^{\infty} \frac{m_k}{k!} t^k \right)^v. \tag{9.7}$$

Since $M(t) \to 1$ as $t \to 0$, this expression is valid for t in some neighborhood of 0. By the theory of series, the powers on the right can be expanded and terms with a common factor t^i collected together. This gives an expansion

$$C(t) = \sum_{i=1}^{\infty} \frac{c_i}{i!} t^i, \tag{9.8}$$

valid in some neighborhood of 0.

The c_i are the *cumulants* of X. Equating coefficients in the expansions (9.7) and (9.8) leads to $c_1 = m_1$ and $c_2 = m_2 - m_1^2$, which checks with (9.6). Each c_i can be expressed as a polynomial in $m_1, \dots, m_i$ and conversely, although the calculations soon become tedious. If $E[X] = 0$, however, so that $m_1 = c_1 = 0$, it is not hard to check that

$$c_3 = m_3, \qquad c_4 = m_4 - 3m_2^2. \tag{9.9}$$

Taking logarithms converts the multiplicative relation (9.4) into the additive relation

$$C(t) = C_1(t) + \cdots + C_n(t) \tag{9.10}$$

for the corresponding cumulant generating functions; it is valid in the presence of independence. By this and the definition (9.8), it follows that cumulants add for independent random variables.

Clearly, $M''(t) = E[X^2 e^{tX}] \ge 0$. Since $(M'(t))^2 = E^2[Xe^{tX}] \le E[e^{tX}] \cdot E[X^2 e^{tX}] = M(t)M''(t)$ by Schwarz's inequality (5.36), $C''(t) \ge 0$. Thus *the moment generating function and the cumulant generating function are both convex.*

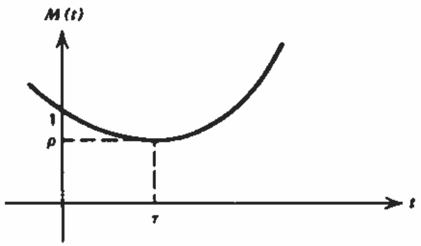
Large Deviations

Let Y be a simple random variable assuming values y_j with probabilities p_j. The problem is to estimate $P[Y \ge \alpha]$ when Y has mean 0 and α is positive. It is notationally convenient to subtract α away from Y and instead estimate $P[Y \ge 0]$ when Y has negative mean.

Assume then that

$$E[Y] < 0, \qquad P[Y > 0] > 0, \tag{9.11}$$

the second assumption to avoid trivialities. Let $M(t) = \sum_j p_j e^{ty_j}$ be the moment generating function of Y. Then $M'(0) < 0$ by the first assumption in



(9.11), and $M(t) \to \infty$ as $t \to \infty$ by the second. Since $M(t)$ is convex, it has its minimum ρ at a positive argument τ:

$$\inf_t M(t) = M(\tau) = \rho, \qquad 0 < \rho < 1, \quad \tau > 0. \tag{9.12}$$

Construct (on an entirely irrelevant probability space) an auxiliary random variable Z such that

$$P[Z = y_j] = \frac{e^{\tau y_j}}{\rho} P[Y = y_j] \tag{9.13}$$

for each y_j in the range of Y. Note that the probabilities on the right do add to 1. The moment generating function of Z is

$$E[e^{tZ}] = \sum_j \frac{e^{\tau y_j}}{\rho} p_j e^{ty_j} = \frac{M(\tau + t)}{\rho}, \tag{9.14}$$

and therefore

$$E[Z] = \frac{M'(\tau)}{\rho} = 0, \qquad s^2 = E[Z^2] = \frac{M''(\tau)}{\rho} > 0. \tag{9.15}$$

For all positive t, $P[Y \ge 0] = P[e^{tY} \ge 1] \le M(t)$ by Markov's inequality (5.31), and hence

$$P[Y \ge 0] \le \rho. \tag{9.16}$$

Inequalities in the other direction are harder to obtain. If $\sum'$ denotes summation over those indices j for which $y_j \ge 0$, then

$$P[Y \ge 0] = \sum{}' p_j = \rho \sum{}' e^{-\tau y_j} P[Z = y_j]. \tag{9.17}$$

Put the final sum here in the form $e^{-\theta}$, and let $p = P[Z \ge 0]$. By (9.16), $\theta \ge 0$. Since $\log x$ is concave, Jensen's inequality (5.33) gives

$$\begin{aligned} -\theta &= \log \sum{}' e^{-\tau y_j} p^{-1} P[Z = y_j] + \log p \\ &\ge \sum{}' (-\tau y_j) p^{-1} P[Z = y_j] + \log p \\ &= -\tau s p^{-1} \sum{}' \frac{y_j}{s} P[Z = y_j] + \log p. \end{aligned}$$

By (9.15) and Lyapounov's inequality (5.37),

$$\sum{}' \frac{y_j}{s} P[Z = y_j] \le \frac{1}{s} E[|Z|] \le \frac{1}{s} E^{1/2}[Z^2] = 1.$$

The last two inequalities give

$$0 \le \theta \le \frac{\tau s}{P[Z \ge 0]} - \log P[Z \ge 0]. \tag{9.18}$$

This proves the following result.

Theorem 9.1. *Suppose that Y satisfies (9.11). Define ρ and τ by (9.12), let Z be a random variable with distribution (9.13), and define s^2 by (9.15). Then $P[Y \ge 0] = \rho e^{-\theta}$, where θ satisfies (9.18).*

To use (9.18) requires a lower bound for $P[Z \ge 0]$.

Theorem 9.2. *If $E[Z] = 0$, $E[Z^2] = s^2$, and $E[Z^4] = \xi^4 > 0$, then $P[Z \ge 0] \ge s^4/4\xi^4$.*†

Proof. Let $Z^+ = ZI_{[Z \ge 0]}$ and $Z^- = -ZI_{[Z<0]}$. Then Z^+ and Z^- are nonnegative, $Z = Z^+ - Z^-$, $Z^2 = (Z^+)^2 + (Z^-)^2$, and

$$s^2 = E\left[(Z^+)^2\right] + E\left[(Z^-)^2\right]. \tag{9.19}$$

Let $p = P[Z \ge 0]$. By Schwarz's inequality (5.36),

$$\begin{aligned} E\left[(Z^+)^2\right] &= E\left[I_{[Z \ge 0]}Z^2\right] \\ &\le E^{1/2}\left[I^2_{[Z \ge 0]}\right]E^{1/2}\left[Z^4\right] = p^{1/2}\xi^2. \end{aligned}$$

By Hölder's inequality (5.35) (for $p = \frac{3}{2}$ and $q = 3$)

$$\begin{aligned} E\left[(Z^-)^2\right] &= E\left[(Z^-)^{2/3}(Z^-)^{4/3}\right] \\ &\le E^{2/3}[Z^-]E^{1/3}\left[(Z^-)^4\right] \le E^{2/3}[Z^-]\xi^{4/3}. \end{aligned}$$

Since $E[Z] = 0$, another application of Hölder's inequality (for $p = 4$ and $q = \frac{4}{3}$) gives

$$\begin{aligned} E[Z^-] &= E[Z^+] = E\left[ZI_{[Z \ge 0]}\right] \\ &\le E^{1/4}\left[Z^4\right]E^{3/4}\left[I^{4/3}_{[Z \ge 0]}\right] = \xi p^{3/4}. \end{aligned}$$

Combining these three inequalities with (9.19) gives $s^2 \le p^{1/2}\xi^2 + (\xi p^{3/4})^{2/3}\xi^{4/3} = 2p^{1/2}\xi^2$. ■

† For a related result, see Problem 25.19.

Chernoff's Theorem†

Theorem 9.3. *Let $X_1, X_2, \ldots$ be independent, identically distributed simple random variables satisfying $E[X_n] < 0$ and $P[X_n > 0] > 0$, let $M(t)$ be their common moment generating function, and put $\rho = \inf_t M(t)$. Then*

$$\lim_{n \to \infty} \frac{1}{n} \log P[X_1 + \cdots + X_n \ge 0] = \log \rho. \tag{9.20}$$

Proof. Put $Y_n = X_1 + \cdots + X_n$. Then $E[Y_n] < 0$ and $P[Y_n > 0] \ge P[X_1 > 0] > 0$, and so the hypotheses of Theorem 9.1 are satisfied. Define ρ_n and τ_n by $\inf_t M_n(t) = M_n(\tau_n) = \rho_n$, where $M_n(t)$ is the moment generating function of Y_n. Since $M_n(t) = M^n(t)$, it follows that $\rho_n = \rho^n$ and $\tau_n = \tau$, where $M(\tau) = \rho$.

Let Z_n be the analogue for Y_n of the Z described by (9.13). Its moment generating function (see (9.14)) is $M_n(\tau + t)/\rho^n = (M(\tau + t)/\rho)^n$. This is also the moment generating function of $V_1 + \cdots + V_n$ for independent random variables $V_1, \ldots, V_n$ each having moment generating function $M(\tau + t)/\rho$. Now each V_i has (see (9.15)) mean 0 and some positive variance σ^2 and fourth moment ξ^4 independent of i. Since Z_n must have the same moments as $V_1 + \cdots + V_n$, it has mean 0, variance $s_n^2 = n\sigma^2$, and fourth moment $\xi_n^4 = n\xi^4 + 3n(n-1)\sigma^4 = O(n^2)$ (see (6.2)). By Theorem 9.2, $P[Z_n \ge 0] \ge s_n^4/4\xi_n^4 \ge \alpha$ for some positive α independent of n. By Theorem 9.1 then, $P[Y_n \ge 0] = \rho^n e^{-\theta_n}$, where $0 \le \theta_n \le \tau_n s_n \alpha^{-1} - \log \alpha = \tau\alpha^{-1}\sigma\sqrt{n} - \log\alpha$. This gives (9.20), and shows, in fact, that the rate of convergence is $O(n^{-1/2})$. ■

This result is important in the theory of statistical hypothesis testing. An informal treatment of the Bernoulli case will illustrate the connection.

Suppose $S_n = X_1 + \cdots + X_n$, where the X_i are independent and assume the values 1 and 0 with probabilities p and q. Now $P[S_n \ge na] = P[\sum_{k=1}^n (X_k - a) \ge 0]$, and Chernoff's theorem applies if $p < a < 1$. In this case $M(t) = E[e^{t(X_1 - a)}] = e^{-ta}(pe^t + q)$. Minimizing this shows that the ρ of Chernoff's theorem satisfies

$$-\log \rho = K(a, p) = a \log \frac{a}{p} + b \log \frac{b}{q},$$

where $b = 1 - a$. By (9.20), $n^{-1} \log P[S_n \ge na] \to -K(a, p)$; express this as

$$P[S_n \ge na] \approx e^{-nK(a,p)}. \tag{9.21}$$

Suppose now that p is unknown and that there are two competing hypotheses concerning its value, the hypothesis H_1 that $p = p_1$ and the hypothesis H_2 that

† This theorem is not needed for the law of the iterated logarithm, Theorem 9.5.

$p = p_2$, where $p_1 < p_2$. Given the observed results $X_1, \ldots, X_n$ of n Bernoulli trials, one decides in favor of H_2 if $S_n \ge na$ and in favor of H_1 if $S_n < na$, where a is some number satisfying $p_1 < a < p_2$. The problem is to find an advantageous value for the threshold a.

By (9.21),

$$P[S_n \ge na | H_1] \approx e^{-nK(a,p_1)}, \tag{9.22}$$

where the notation indicates that the probability is calculated for $p = p_1$—that is, under the assumption of H_1. By symmetry,

$$P[S_n < na | H_2] \approx e^{-nK(a,p_2)}. \tag{9.23}$$

The left sides of (9.22) and (9.23) are the probabilities of erroneously deciding in favor of H_2 when H_1 is, in fact, true and of erroneously deciding in favor of H_1 when H_2 is, in fact, true—the probabilities describing the level and power of the test.

Suppose a is chosen so that $K(a,p_1) = K(a,p_2)$, which makes the two error probabilities approximately equal. This constraint gives for a a linear equation with solution

$$a = a(p_1, p_2) = \frac{\log(q_1/q_2)}{\log(p_2/p_1) + \log(q_1/q_2)}, \tag{9.24}$$

where $q_i = 1 - p_i$. The common error probability is approximately $e^{-nK(a,p_1)}$ for this value of a, and so the larger $K(a,p_1)$ is, the easier it is to distinguish statistically between p_1 and p_2.

Although $K(a(p_1,p_2),p_1)$ is a complicated function, it has a simple approximation for p_1 near p_2. As $x \to 0$, $\log(1+x) = x - \frac{1}{2}x^2 + O(x^3)$. Using this in the definition of K and collecting terms gives

$$K(p+x,p) = \frac{x^2}{2pq} + O(x^3), \qquad x \to 0. \tag{9.25}$$

Fix $p_1 = p$, and let $p_2 = p + t$; (9.24) becomes a function $\psi(t)$ of t, and expanding the logarithms gives

$$\psi(t) = p + \tfrac{1}{2}t + O(t^2), \qquad t \to 0, \tag{9.26}$$

after some reductions. Finally, (9.25) and (9.26) together imply that

$$K(\psi(t),p) = \frac{t^2}{8pq} + O(t^3), \qquad t \to 0. \tag{9.27}$$

In distinguishing $p_1 = p$ from $p_2 = p + t$ for small t, if a is chosen to equalize the two error probabilities, then their common value is about $e^{-nt^2/8pq}$. For t fixed, the nearer p is to $\frac{1}{2}$, the larger this probability is and the more difficult it is to distinguish p from $p + t$. As an example, compare $p = .1$ with $p = .5$. Now $.36nt^2/8(.1)(.9) = nt^2/8(.5)(.5)$. With a sample only 36 percent as large, .1 can therefore be distinguished from $.1 + t$ with about the same precision as .5 can be distinguished from $.5 + t$.

The Law of the Iterated Logarithm

The analysis of the rate at which S_n/n approaches the mean depends on the following variant of the theorem on large deviations.

Theorem 9.4. *Let $S_n = X_1 + \cdots + X_n$, where the X_n are independent and identically distributed simple random variables with mean 0 and variance 1. If a_n are constants satisfying*

$$a_n \to \infty, \qquad \frac{a_n}{\sqrt{n}} \to 0, \tag{9.28}$$

then

$$P\left[S_n \ge a_n\sqrt{n}\right] = e^{-a_n^2(1+\zeta_n)/2} \tag{9.29}$$

for a sequence ζ_n going to 0.

PROOF. Put $Y_n = S_n - a_n\sqrt{n} = \sum_{k=1}^n (X_k - a_n/\sqrt{n})$. Then $E[Y_n] < 0$. Since X_1 has mean 0 and variance 1, $P[X_1 > 0] > 0$, and it follows by (9.28) that $P[X_1 > a_n/\sqrt{n}] > 0$ for n sufficiently large, in which case $P[Y_n > 0] \ge P^n[X_1 - a_n/\sqrt{n} > 0] > 0$. Thus Theorem 9.1 applies to Y_n for all large enough n.

Let $M_n(t)$, ρ_n, τ_n, and Z_n be associated with Y_n as in the theorem. If $m(t)$ and $c(t)$ are the moment and cumulant generating functions of the X_n, then $M_n(t)$ is the nth power of the moment generating function $e^{-ta_n/\sqrt{n}}m(t)$ of $X_1 - a_n/\sqrt{n}$, and so Y_n has cumulant generating function

$$C_n(t) = -ta_n\sqrt{n} + nc(t). \tag{9.30}$$

Since τ_n is the unique minimum of $C_n(t)$, and since $C_n'(t) = -a_n\sqrt{n} + nc'(t)$, τ_n is determined by the equation $c'(\tau_n) = a_n/\sqrt{n}$. Since X_1 has mean 0 and variance 1, it follows by (9.6) that

$$c(0) = c'(0) = 0, \qquad c''(0) = 1. \tag{9.31}$$

Now $c'(t)$ is nondecreasing because $c(t)$ is convex, and since $c'(\tau_n) = a_n/\sqrt{n}$ goes to 0, τ_n must therefore go to 0 as well and must in fact be $O(a_n/\sqrt{n})$. By the second-order mean-value theorem for $c'(t)$, $a_n/\sqrt{n} = c'(\tau_n) = \tau_n + O(\tau_n^2)$, from which follows

$$\tau_n = \frac{a_n}{\sqrt{n}} + O\left(\frac{a_n^2}{n}\right). \tag{9.32}$$

By the third-order mean-value theorem for $c(t)$,

$$\log \rho_n = C_n(\tau_n) = -\tau_n a_n \sqrt{n} + nc(\tau_n)$$
$$= -\tau_n a_n \sqrt{n} + n\left[\tfrac{1}{2}\tau_n^2 + O(\tau_n^3)\right].$$

Applying (9.32) gives

(9.33) $$\log \rho_n = -\tfrac{1}{2}a_n^2 + o(a_n^2).$$

Now (see (9.14)) Z_n has moment generating function $M_n(\tau_n + t)/\rho_n$ and (see (9.30)) cumulant generating function $D_n(t) = C_n(\tau_n + t) - \log \rho_n = -(\tau_n + t)\alpha_n\sqrt{n} + nc(t + \tau_n) - \log \rho_n$. The mean of Z_n is $D_n'(0) = 0$. Its variance s_n^2 is $D_n''(0)$; by (9.31), this is

(9.34) $$s_n^2 = nc''(\tau_n) = n(c''(0) + O(\tau_n)) = n(1 + o(1)).$$

The fourth cumulant of Z_n is $D_n''''(0) = nc''''(\tau_n) = O(n)$. By the formula (9.9) relating moments and cumulants (applicable because $E[Z_n] = 0$), $E[Z_n^4] = 3s_n^4 + D_n''''(0)$. Therefore, $E[Z_n^4]/s_n^4 \to 3$, and it follows by Theorem 9.2 that there exists an α such that $P[Z_n \ge 0] \ge \alpha > 0$ for all sufficiently large n.

By Theorem 9.1, $P[Y_n \ge 0] = \rho_n e^{-\theta_n}$ with $0 \le \theta_n \le \tau_n s_n \alpha^{-1} + \log \alpha$. By (9.28), (9.32), and (9.34), $\theta_n = O(a_n) = o(a_n^2)$, and it follows by (9.33) that $P[Y_n \ge 0] = e^{-a_n^2(1+o(1))/2}$. ∎

The law of the iterated logarithm is this:

Theorem 9.5. *Let $S_n = X_1 + \cdots + X_n$, where the X_n are independent, identically distributed simple random variables with mean 0 and variance 1. Then*

(9.35) $$P\left[\limsup_n \frac{S_n}{\sqrt{2n \log\log n}} = 1\right] = 1.$$

Equivalent to (9.35) is the assertion that for positive ϵ

(9.36) $$P\left[S_n \ge (1+\epsilon)\sqrt{2n\log\log n}\ \text{i.o.}\right] = 0$$

and

(9.37) $$P\left[S_n \ge (1-\epsilon)\sqrt{2n\log\log n}\ \text{i.o.}\right] = 1.$$

The set in (9.35) is, in fact, the intersection over positive rational ϵ of the sets in (9.37) minus the union over positive rational ϵ of the sets in (9.36).

The idea of the proof is this. Write

(9.38) $$\phi(n) = \sqrt{2n\log\log n}.$$

If $A_n^{\pm} = [S_n \ge (1 \pm \epsilon)\phi(n)]$, then by (9.29), $P(A_n^{\pm})$ is near $(\log n)^{-(1\pm\epsilon)^2}$. If n_k increases exponentially, say $n_k \sim \theta^k$ for $\theta > 1$, then $P(A_{n_k}^{\pm})$ is of the order $k^{-(1\pm\epsilon)^2}$. Now $\sum_k k^{-(1\pm\epsilon)^2}$ converges if the sign is + and diverges if the sign is −. It will follow by the first Borel–Cantelli lemma that there is probability 0 that $A_{n_k}^+$ occurs for infinitely many k. In providing (9.36), an extra argument is required to get around the fact that the A_n^+ for $n \ne n_k$ must also be accounted for (this requires choosing θ near 1). If the A_n^- were independent, it would follow by the second Borel–Cantelli lemma that with probability 1, $A_{n_k}^-$ occurs for infinitely many k, which would in turn imply (9.37). An extra argument is required to get around the fact that the $A_{n_k}^-$ are dependent (this requires choosing θ large).

For the proof of (9.36) a preliminary result is needed. Put $M_k = \max\{S_0, S_1, \ldots, S_k\}$, where $S_0 = 0$.

Theorem 9.6. *If the X_k are independent simple random variables with mean 0 and variance 1, then for $\alpha \ge \sqrt{2}$.*

(9.39) $$P\left[\frac{M_n}{\sqrt{n}} \ge \alpha\right] \le 2P\left[\frac{S_n}{\sqrt{n}} \ge \alpha - \sqrt{2}\right].$$

Proof. If $A_j = [M_{j-1} < \alpha\sqrt{n} \le M_j]$, then

$$P\left[\frac{M_n}{\sqrt{n}} \ge \alpha\right] \le P\left[\frac{S_n}{\sqrt{n}} \ge \alpha - \sqrt{2}\right] + \sum_{j=1}^{n-1} P\left(A_j \cap \left[\frac{S_n}{\sqrt{n}} \le \alpha - \sqrt{2}\right]\right).$$

Since $S_n - S_j$ has variance $n - j$, it follows by independence and Chebyshev's inequality that the probability in the sum is at most

$$P\left(A_j \cap \left[\frac{|S_n - S_j|}{\sqrt{n}} > \sqrt{2}\right]\right) = P(A_j)P\left[\frac{|S_n - S_j|}{\sqrt{n}} > \sqrt{2}\right]$$
$$\le P(A_j)\frac{n-j}{2n} \le \tfrac{1}{2}P(A_j).$$

Since $\bigcup_{j=1}^{n-1} A_j \subset [M_n \ge \alpha\sqrt{n}]$,

$$P\left[\frac{M_n}{\sqrt{n}} \ge \alpha\right] \le P\left[\frac{S_n}{\sqrt{n}} \ge \alpha - \sqrt{2}\right] + \tfrac{1}{2}P\left[\frac{M_n}{\sqrt{n}} \ge \alpha\right].$$ ∎

PROOF OF (9.36). Given ϵ, choose θ so that $\theta > 1$ but $\theta^2 < 1 + \epsilon$. Let $n_k = \lfloor \theta^k \rfloor$ and $x_k = \theta(2 \log\log n_k)^{1/2}$. By (9.29) and (9.39),

$$P\left[\frac{M_{n_k}}{\sqrt{n_k}} \ge x_k\right] \le 2\exp\left[-\tfrac{1}{2}\left(x_k - \sqrt{2}\right)^2(1+\xi_k)\right].$$

where $\xi_k \to 0$. The negative of the exponent is asymptotically $\theta^2 \log k$ and hence for large k exceeds $\theta \log k$, so that

$$P\left[\frac{M_{n_k}}{\sqrt{n_k}} \ge x_k\right] \le \frac{2}{k^\theta}.$$

Since $\theta > 1$, it follows by the first Borel–Cantelli lemma that there is probability 0 that (see (9.38))

$$M_{n_k} \ge \theta\phi(n_k) \tag{9.40}$$

for infinitely many k. Suppose that $n_{k-1} < n \le n_k$ and that

$$S_n > (1+\epsilon)\phi(n). \tag{9.41}$$

Now $\phi(n) \ge \phi(n_{k-1}) \sim \theta^{-1/2}\phi(n_k)$; hence, by the choice of θ, $(1+\epsilon)\phi(n) > \theta\phi(n_k)$ if k is large enough. Thus for sufficiently large k, (9.41) implies (9.40) (if $n_{k-1} < n \le n_k$), and there is therefore proability 0 that (9.41) holds for infinitely many n. ■

PROOF OF (9.37). Given ϵ, choose an integer θ so large that $3\theta^{-1/2} < \epsilon$. Take $n_k = \theta^k$. Now $n_k - n_{k-1} \to \infty$, and (9.29) applies with $n = n_k - n_{k-1}$ and $a_n = x_k/\sqrt{n_k - n_{k-1}}$, where $x_k = (1-\theta^{-1})\phi(n_k)$. It follows that

$$P\left[S_{n_k} - S_{n_{k-1}} \ge x_k\right] = P\left[S_{n_k - n_{k-1}} \ge x_k\right] = \exp\left[-\frac{1}{2}\frac{x_k^2}{n_k - n_{k-1}}(1+\xi_k)\right],$$

where $\xi_k \to 0$. The negative of the exponent is asymptotically $(1-\theta^{-1})\log k$ and so for large k is less than $\log k$, in which case $P[S_{n_k} - S_{n_{k-1}} \ge x_k] \ge k^{-1}$. The events here being independent, it follows by the second Borel–Cantelli lemma that with probability 1, $S_{n_k} - S_{n_{k-1}} \ge x_k$ for infinitely many k. On the other hand, by (9.36) applied to $\{-X_n\}$, there is probability 1 that $-S_{n_{k-1}} \le 2\phi(n_{k-1}) \le 2\theta^{-1/2}\phi(n_k)$ for all but finitely many k. These two inequalities give $S_{n_k} \ge x_k - 2\theta^{-1/2}\phi(n_k) > (1-\epsilon)\phi(n_k)$, the last inequality because of the choice of θ. ■

That completes the proof of Theorem 9.5.

PROBLEMS

9.1. Prove (6.2) by using (9.9) and the fact that cumulants add in the presence of independence.

9.2. In the Bernoulli case, (9.21) gives

$$P[S_n \ge np + x_n] = \exp\left[-nK\left(p + \frac{x_n}{n}, p\right)(1+o(1))\right],$$

where $p < a < 1$ and $x_n = n(a-p)$. Theorem 9.4 gives

$$P[S_n \ge np + x_n] = \exp\left[-\frac{x_n^2}{2npq}(1+o(1))\right],$$

where $x_n = a_n\sqrt{npq}$. Resolve the apparent discrepancy. Use (9.25) to compare the two expressions in case x_n/n is small. See Problem 27.17.

9.3. Relabel the binomial parameter p as $\theta = f(p)$, where f is increasing and continuously differentiable. Show by (9.27) that the distinguishability of θ from $\theta + \Delta\theta$, as measured by K, is $(\Delta\theta)^2/8p(1-p)(f'(p))^2 + O(\Delta\theta)^3$. The leading coefficient is independent of θ if $f(p) = \arcsin\sqrt{p}$.

9.4. From (9.35) and the same result for $\{-X_n\}$, together with the uniform boundedness of the X_n, deduce that with probability 1 the set of limit points of the sequence $\{S_n(2n \log\log n)^{-1/2}\}$ is the closed interval from -1 to $+1$.

9.5. ↑ Suppose X_n takes the values ± 1 with probability $\frac{1}{2}$ each, and show that $P[S_n = 0 \text{ i.o.}] = 1$. (This gives still another proof of the persistence of symmetric random walk on the line (Example 8.6).) Show more generally that, if the X_n are bounded by M, then $P[|S_n| \le M \text{ i.o.}] = 1$.

9.6. Weakened versions of (9.36) are quite easy to prove. By a fourth-moment argument (see (6.2)), show that $P[S_n > n^{3/4}(\log n)^{(1+\epsilon)/4} \text{i.o.}] = 0$. Use (9.29) to give a simple proof that $P[S_n > (3n \log n)^{1/2} \text{i.o.}] = 0$.

9.7. Show that (9.35) is true if S_n is replaced by $|S_n|$ or $\max_{k \le n} S_k$ or $\max_{k \le n}|S_k|$.