For a function $f$ of positive integers, let

$$E_n[f] = \frac{1}{n}\sum_{m=1}^{n} f(m) \tag{5.46}$$

be its expected value under the probability measure $P_n$. Show that

$$E_n[\alpha_p] = \sum_{k=1}^{\infty} \frac{1}{n}\left\lfloor \frac{n}{p^k} \right\rfloor \to \frac{1}{p-1}; \tag{5.47}$$

this says roughly that $(p-1)^{-1}$ is the average power of $p$ in the factorization of large integers.

**5.20.** ↑ **(a)** From Stirling's formula, deduce

$$E_n[\log] = \log n + O(1). \tag{5.48}$$

From this, the inequality $E_n[\alpha_p] \le 2/p$, and the relation $\log m = \sum_p \alpha_p(m)\log p$, conclude that $\sum_p p^{-1}\log p$ diverges and that there are infinitely many primes. **(b)** Let $\log^* m = \sum_p \delta_p(m)\log p$. Show that

$$E_n[\log^*] = \sum_p \frac{1}{n}\left\lfloor \frac{n}{p} \right\rfloor \log p = \log n + O(1). \tag{5.49}$$

**(c)** Show that $\lfloor 2n/p \rfloor - 2\lfloor n/p \rfloor$ is always nonnegative and equals 1 in the range $n < p \le 2n$. Deduce $E_{2n}[\log^*] - E_n[\log^*] = O(1)$ and conclude that

$$\sum_{p \le x} \log p = O(x). \tag{5.50}$$

Use this to estimate the error removing the integral-part brackets introduces into (5.49), and show that

$$\sum_{p \le x} p^{-1}\log p = \log x + O(1). \tag{5.51}$$

**(d)** Restrict the range of summation in (5.51) to $\theta x < p \le x$ for an appropriate $\theta$, and conclude that

$$\sum_{p \le x} \log p \asymp x, \tag{5.52}$$

in the sense that the ratio of the two sides is bounded away from 0 and $\infty$.
**(e)** Use (5.52) and truncation arguments to prove for the number $\pi(x)$ of primes not exceeding $x$ that

$$\pi(x) \asymp \frac{x}{\log x}. \tag{5.53}$$



(By the prime number theorem the ratio of the two sides in fact goes to 1.) Conclude that the $r$th prime $p_r$ satisfies $p_r \asymp r\log r$ and that

$$\sum_p \frac{1}{p} = \infty. \tag{5.54}$$

## SECTION 6. THE LAW OF LARGE NUMBERS

### The Strong Law

Let $X_1, X_2, \ldots$ be a sequence of simple random variables on some probability space $(\Omega, \mathscr{F}, P)$. They are *identically distributed* if their distributions (in the sense of (5.12)) are all the same. Define $S_n = X_1 + \cdots + X_n$. The *strong law of large numbers*:

**Theorem 6.1.** *If the $X_n$ are independent and identically distributed and $E[X_n] = m$, then*

$$P\left[\lim_n n^{-1}S_n = m\right] = 1. \tag{6.1}$$

PROOF. The conclusion is that $n^{-1}S_n - m = n^{-1}\sum_{i=1}^n (X_i - m) \to 0$ with probability 1. Replacing $X_i$ by $X_i - m$ shows that there is no loss of generality in assuming that $m = 0$. The set in question does lie in $\mathscr{F}$ (see (5.5)), and by Theorem 5.2(i), it is enough to show that $P[|n^{-1}S_n| \ge \epsilon \text{ i.o.}] = 0$ for each $\epsilon$.

Let $E[X_i^2] = \sigma^2$ and $E[X_i^4] = \xi^4$. The proof is like that for Theorem 1.2. First (see (1.26)), $E[S_n^4] = \sum E[X_\alpha X_\beta X_\gamma X_\delta]$, the four indices ranging independently from 1 to $n$. Since $E[X_i] = 0$, it follows by the product rule (5.25) for independent random variables that the summand vanishes if there is one index different from the three others. This leaves terms of the form $E[X_i^4] = \xi^4$, of which there are $n$, and terms of the form $E[X_i^2X_j^2] = E[X_i^2]E[X_j^2] = \sigma^4$ for $i \ne j$, of which there are $3n(n-1)$. Hence

$$E[S_n^4] = n\xi^4 + 3n(n-1)\sigma^4 \le Kn^2, \tag{6.2}$$

where $K$ does not depend on $n$.

By Markov's inequality (5.31) for $k = 4$, $P[|S_n| \ge n\epsilon] \le Kn^{-2}\epsilon^{-4}$, and so by the first Borel–Cantelli lemma, $P[|n^{-1}S_n| \ge \epsilon \text{ i.o.}] = 0$, as required. ■

*Example 6.1.* The classical example is the strong law of large numbers for Bernoulli trials. Here $P[X_n = 1] = p$, $P[X_n = 0] = 1 - p$, $m = p$; $S_n$ represents the number of successes in $n$ trials, and $n^{-1}S_n \to p$ with probability 1. The idea of probability as frequency depends on the long-range stability of the success ratio $S_n/n$. ■

*Example 6.2.* Theorem 1.2 is the case of Example 6.1 in which $(\Omega, \mathscr{F}, P)$ is the unit interval and the $X_n(\omega)$ are the digits $d_n(\omega)$ of the dyadic expansion of $\omega$. Here $p = \frac{1}{2}$. The set (1.21) of normal numbers in the unit interval has by (6.1) Lebesgue measure 1; its complement has measure 0 (and so in the terminology of Section 1 is negligible). ■

### The Weak Law

Since convergence with probability 1 implies convergence in probability (Theorem 5.2(ii)), it follows under the hypotheses of Theorem 6.1 that $n^{-1}S_n \to_P m$. But this is of course an immediate consequence of Chebyshev's inequality (5.32) and the rule (5.28) for adding variances:

$$P\left[\left|n^{-1}S_n - m\right| \geq \epsilon\right] \leq \frac{\operatorname{Var}[S_n]}{n^2\epsilon^2} = \frac{n \operatorname{Var}[X_1]}{n^2\epsilon^2} \to 0.$$

This is the *weak law of large numbers*.

Chebyshev's inequality leads to a weak law in other interesting cases as well:

*Example 6.3.* Let $\Omega_n$ consist of the $n!$ permutations of $1, 2, \ldots, n$, all equally probable, and let $X_{nk}(\omega)$ be 1 or 0 according as the $k$th element in the cyclic representation of $\omega \in \Omega_n$ completes a cycle or not. This is Example 5.6, although there the dependence on $n$ was suppressed in the notation. The $X_{n1}, \ldots, X_{nn}$ are independent, and $S_n = X_{n1} + \cdots + X_{nn}$ is the number of cycles. The mean $m_{nk}$ of $X_{nk}$ is the probability that it equals 1, namely $(n-k+1)^{-1}$, and its variance is $\sigma_{nk}^2 = m_{nk}(1-m_{nk})$.

If $L_n = \sum_{k=1}^{n} k^{-1}$, then $S_n$ has mean $\sum_{k=1}^{n} m_{nk} = L_n$ and variance $\sum_{k=1}^{n} m_{nk}(1-m_{nk}) < L_n$. By Chebyshev's inequality,

$$P\left[\left|\frac{S_n - L_n}{L_n}\right| \geq \epsilon\right] < \frac{L_n}{\epsilon^2 L_n^2} = \frac{1}{\epsilon^2 L_n} \to 0.$$

Of the $n!$ permutations on $n$ letters, a proportion exceeding $1 - \epsilon^{-2}L_n^{-1}$ thus have their cycle number in the range $(1 \pm \epsilon)L_n$. Since $L_n = \log n + O(1)$, most permutations on $n$ letters have about $\log n$ cycles. For a refinement, see Example 27.3.

Since $\Omega_n$ changes with $n$, it is the nature of the case that there cannot be a strong law corresponding to this result. ■

### Bernstein's Theorem

Some theorems that can be stated without reference to probability nonetheless have simple probabilistic proofs, as the last example shows. Bernstein's approach to the Weierstrass approximation theorem is another example.



Let $f$ be a function on $[0,1]$. The *Bernstein polynomial* of degree $n$ associated with $f$ is

$$B_n(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k} \tag{6.3}$$

**Theorem 6.2.** *If $f$ is continuous, $B_n(x)$ converges to $f(x)$ uniformly on $[0,1]$.*

According to the Weierstrass approximation theorem, $f$ can be uniformly approximated by polynomials; Bernstein's result goes further and specifies an approximating sequence.

PROOF. Let $M = \sup_x |f(x)|$, and let $\delta(\epsilon) = \sup[|f(x) - f(y)|: |x-y| \leq \epsilon]$ be the modulus of continuity of $f$. It will be shown that

$$\sup_x |f(x) - B_n(x)| \leq \delta(\epsilon) + \frac{2M}{n\epsilon^2}. \tag{6.4}$$

By the uniform continuity of $f$, $\lim_{\epsilon \to 0} \delta(\epsilon) = 0$, and so this inequality (for $\epsilon = n^{-1/3}$, say) will give the theorem.

Fix $n \geq 1$ and $x \in [0,1]$ for the moment. Let $X_1, \ldots, X_n$ be independent random variables (on some probability space) such that $P[X_i = 1] = x$ and $P[X_i = 0] = 1 - x$; put $S = X_1 + \cdots + X_n$. Since $P[S = k] = \binom{n}{k} x^k (1-x)^{n-k}$, the formula (5.19) for calculating expected values of functions of random variables gives $E[f(S/n)] = B_n(x)$. By the law of large numbers, there should be high probability that $S/n$ is near $x$ and hence ($f$ being continuous) that $f(S/n)$ is near $f(x)$; $E[f(S/n)]$ should therefore be near $f(x)$. This is the probabilistic idea behind the proof and, indeed, behind the definition (6.3) itself.

Bound $|f(n^{-1}S) - f(x)|$ by $\delta(\epsilon)$ on the set $[|n^{-1}S - x| < \epsilon]$ and by $2M$ on the complementary set, and use (5.22) as in the proof of Theorem 5.4. Since $E[S] = nx$, Chebyshev's inequality gives

$$\begin{aligned} |B_n(x) - f(x)| &\leq E\left[\left|f(n^{-1}S) - f(x)\right|\right] \\ &\leq \delta(\epsilon) P[|n^{-1}S - x| < \epsilon] + 2MP[|n^{-1}S - x| \geq \epsilon] \\ &\leq \delta(\epsilon) + 2M \operatorname{Var}[S]/n^2\epsilon^2; \end{aligned}$$

since $\operatorname{Var}[S] = nx(1-x) \leq n$, (6.4) follows. ■

### A Refinement of the Second Borel–Cantelli Lemma

For a sequence $A_1, A_2, \ldots$ of events, consider the number $N_n = I_{A_1} + \cdots + I_{A_n}$ of occurrences among $A_1, \ldots, A_n$. Since $[A_n \text{ i.o.}] = [\omega: \sup_n N_n(\omega) = \infty]$, $P[A_n \text{ i.o.}]$ can be studied by means of the random variables $N_n$.

Suppose that the $A_n$ are independent. Put $p_k = P(A_k)$ and $m_n = p_1 + \cdots + p_n$. From $E[I_{A_k}] = p_k$ and $\mathrm{Var}[I_{A_k}] = p_k(1-p_k) \le p_k$ follow $E[N_n] = m_n$ and $\mathrm{Var}[N_n] = \sum_{k=1}^{n} \mathrm{Var}[I_{A_k}] \le m_n$. If $m_n > x$, then

$$P[N_n \le x] \le P[|N_n - m_n| \ge m_n - x] \le \frac{\mathrm{Var}[N_n]}{(m_n - x)^2} \le \frac{m_n}{(m_n - x)^2}. \tag{6.5}$$

If $\sum p_n = \infty$, so that $m_n \to \infty$, it follows that $\lim_n P[N_n \le x] = 0$ for each $x$. Since

$$P\Big[\sup_k N_k \le x\Big] \le P[N_n \le x], \tag{6.6}$$

$P[\sup_k N_k \le x] = 0$ and hence (take the union over $x = 1, 2, \ldots$) $P[\sup_k N_k < \infty] = 0$. Thus $P[A_n \text{ i.o.}] = P[\sup_n N_n = \infty] = 1$ if the $A_n$ are independent and $\sum p_n = \infty$, which proves the second Borel–Cantelli lemma once again.

Independence was used in this argument only to estimate $\mathrm{Var}[N_n]$. Even without independence, $E[N_n] = m_n$ and the first two inequalities in (6.5) hold.

**Theorem 6.3.** *If $\sum P(A_n)$ diverges and*

$$\liminf_n \frac{\sum_{j,k \le n} P(A_j \cap A_k)}{\Big(\sum_{k \le n} P(A_k)\Big)^2} \le 1, \tag{6.7}$$

*then $P[A_n \text{ i.o.}] = 1$.*

As the proof will show, the ratio in (6.7) is at least 1; if (6.7) holds, the inequality must therefore be an equality.

PROOF. Let $\theta_n$ denote the ratio in (6.7). In the notation above,

$$\mathrm{Var}[N_n] = E[N_n^2] - m_n^2 = \sum_{j,k \le n} E[I_{A_j} I_{A_k}] - m_n^2 = \sum_{j,k \le n} P(A_j \cap A_k) - m_n^2 = (\theta_n - 1) m_n^2$$

(and $\theta_n - 1 \ge 0$). Hence (see (6.5)) $P[N_n \le x] \le (\theta_n - 1) m_n^2 / (m_n - x)^2$ for $x < m_n$. Since $m_n^2/(m_n - x)^2 \to 1$, (6.7) implies that $\liminf_n P[N_n \le x] = 0$. It still follows by (6.6) that $P[\sup_k N_k \le x] = 0$, and the rest of the argument is as before. ∎



***Example 6.4.*** If, as in the second Borel–Cantelli lemma, the $A_n$ are independent (or even if they are merely independent in pairs), the ratio in (6.7) is $1 + \sum_{k \le n}(p_k - p_k^2)/m_n^2$, so that $\sum P(A_n) = \infty$ implies (6.7). ∎

***Example 6.5.*** Return once again to the run lengths $l_n(\omega)$ of Section 4. It was shown in Example 4.21 that $\{r_n\}$ is an outer boundary ($P[l_n \ge r_n \text{ i.o.}] = 0$) if $\sum 2^{-r_n} < \infty$. It was also shown that $\{r_n\}$ is an inner boundary ($P[l_n \ge r_n \text{ i.o.}] = 1$) if $r_n$ is nondecreasing and $\sum 2^{-r_n} r_n^{-1} = \infty$, but Theorem 6.3 can be used to prove this under the sole assumption that $\sum 2^{-r_n} = \infty$.

As usual, the $r_n$ can be taken to be positive integers. Let $A_n = [l_n \ge r_n] = [d_n = \cdots = d_{n+r_n-1} = 0]$. If $j + r_j \le k$, then $A_j$ and $A_k$ are independent. If $j < k < j + r_j$, then $P(A_j | A_k) \le P[d_j = \cdots = d_{k-1} = 0 | A_k] = P[d_j = \cdots = d_{k-1} = 0] = 1/2^{k-j}$, and so $P(A_j \cap A_k) \le P(A_k)/2^{k-j}$. Therefore,

$$\sum_{j,k \le n} P(A_j \cap A_k) \le \sum_{k \le n} P(A_k) + 2 \sum_{\substack{j < k \le n \\ j + r_j \le k}} P(A_j) P(A_k) + 2 \sum_{\substack{j < k \le n \\ k < j + r_j}} 2^{-(k-j)} P(A_k)$$

$$\le \sum_{k \le n} P(A_k) + \Big(\sum_{k \le n} P(A_k)\Big)^2 + 2 \sum_{k \le n} P(A_k).$$

If $\sum P(A_n) = \sum 2^{-r_n}$ diverges, then (6.7) follows.

Thus $\{r_n\}$ *is an outer or an inner boundary according as $\sum 2^{-r_n}$ converges or diverges*, which completely settles the issue. In particular, $r_n = \log_2 n + \theta \log_2 \log_2 n$ gives an outer boundary for $\theta > 1$ and an inner boundary for $\theta \le 1$. ∎

***Example 6.6.*** It is now possible to complete the analysis in Examples 4.12 and 4.16 of the relative error $e_n(\omega)$ in the approximation of $\omega$ by $\sum_{k=1}^{n} d_k(\omega) 2^{-k}$. If $l_n(\omega) \ge -\log_2 x_n$ $(0 < x_n < 1)$, then $e_n(\omega) \le x_n$ by (4.22). By the preceding example for the case $r_n = -\log_2 x_n$, $\sum x_n = \infty$ implies that $P[\omega: e_n(\omega) \le x_n \text{ i.o.}] = 1$. By this and Example 4.12, $[\omega: e_n(\omega) \le x_n \text{ i.o.}]$ *has Lebesgue measure 0 or 1 according as $\sum x_n$ converges or diverges.* ∎

## PROBLEMS

**6.1.** Show that $Z_n \to Z$ with probability 1 if and only if for every positive $\epsilon$ there exists an $n$ such that $P[|Z_k - Z| < \epsilon, n \le k \le m] > 1 - \epsilon$ for all $m$ exceeding $n$. This describes convergence with probability 1 in "finite" terms.

**6.2.** Show in Example 6.3 that $P[|S_n - L_n| \ge L_n^{1/2+\epsilon}] \to 0$.

**6.3.** As in Examples 5.6 and 6.3, let $\omega$ be a random permutation of $1, 2, \ldots, n$. Each $k$, $1 \le k \le n$, occupies some position in the bottom row of the permutation $\omega$;

let $X_{nk}(\omega)$ be the number of smaller elements (between 1 and $k-1$) lying to the right of $k$ in the bottom row. The sum $S_n = X_{n1} + \cdots + X_{nn}$ is the total number of *inversions*—the number of pairs appearing in the bottom row in reverse order of size. For the permutation in Example 5.6 the values of $X_{71}, \ldots, X_{77}$ are 0, 0, 0, 2, 4, 2, 4, and $S_7 = 12$. Show that $X_{n1}, \ldots, X_{nn}$ are independent and $P[X_{nk} = i] = k^{-1}$ for $0 \le i < k$. Calculate $E[S_n]$ and $\operatorname{Var}[S_n]$. Show that $S_n$ is likely to be near $n^2/4$.

**6.4.** For a function $f$ on $[0,1]$ write $\|f\| = \sup_x |f(x)|$. Show that, if $f$ has a continuous derivative $f'$, then $\|f - B_n\| \le \epsilon \|f'\| + 2\|f\|/n\epsilon^2$. Conclude that $\|f - B_n\| = O(n^{-1/3})$.

**6.5.** Prove *Poisson's theorem*: If $A_1, A_2, \ldots$ are independent events, $\bar{p}_n = n^{-1}\sum_{i=1}^n P(A_i)$, and $N = \sum_{i=1}^n I_{A_i}$, then $n^{-1}N_n - \bar{p}_n \to_P 0$.

In the following problems $S_n = X_1 + \cdots + X_n$.

**6.6.** Prove *Cantelli's theorem*: If $X_1, X_2, \ldots$ are independent, $E[X_n] = 0$, and $E[X_n^4]$ is bounded, then $n^{-1}S_n \to 0$ with probability 1. The $X_n$ need not be identically distributed.

**6.7.** (a) Let $x_1, x_2, \ldots$ be a sequence of real numbers, and put $s_n = x_1 + \cdots + x_n$. Suppose that $n^{-2}s_{n^2} \to 0$ and that the $x_n$ are bounded, and show that $n^{-1}s_n \to 0$. (b) Suppose that $n^{-2}S_{n^2} \to 0$ with probability 1 and that the $X_n$ are uniformly bounded ($\sup_{n,\omega} |X_n(\omega)| < \infty$). Show that $n^{-1}S_n \to 0$ with probability 1. Here the $X_n$ need not be identically distributed or even independent.

**6.8.** † Suppose that $X_1, X_2, \ldots$ are independent and uniformly bounded and $E[X_n] = 0$. Using only the preceding result, the first Borel–Cantelli lemma, and Chebyshev's inequality, prove that $n^{-1}S_n \to 0$ with probability 1.

**6.9.** † Use the ideas of Problem 6.8 to give a new proof of Borel's normal number theorem, Theorem 1.2. The point is to return to first principles and use only negligibility and the other ideas of Section 1, not the apparatus of Sections 2 through 6; in particular, $P(A)$ is to be taken as defined only if $A$ is a finite, disjoint union of intervals.

**6.10.** 5.11 6.7† Suppose that (in the notation of (5.41)) $\beta_n - \alpha_n^2 = O(1/n)$. Show that $n^{-1}N_n - \alpha_n \to 0$ with probability 1. What condition on $\beta_n - \alpha_n^2$ will imply a weak law? Note that independence is not assumed here.

**6.11.** Suppose that $X_1, X_2, \ldots$ are *m-dependent* in the sense that random variables more than $m$ apart in the sequence are independent. More precisely, let $\mathscr{F}_j^k = \sigma(X_j, \ldots, X_k)$, and assume that $\mathscr{F}_{j_1}^{k_1}, \ldots, \mathscr{F}_{j_l}^{k_l}$ are independent if $k_{i-1} + m < j_i$ for $i = 2, \ldots, l$. (Independent random variables are 0-dependent.) Suppose that the $X_n$ have this property and are uniformly bounded and that $E[X_n] = 0$. Show that $n^{-1}S_n \to 0$. *Hint:* Consider the subsequences $X_i, X_{i+m+1}, X_{i+2(m+1)}, \ldots$ for $1 \le i \le m+1$.

**6.12.** † Suppose that the $X_n$ are independent and assume the values $x_1, \ldots, x_l$ with probabilities $p(x_1), \ldots, p(x_l)$. For $u_1, \ldots, u_k$ a $k$-tuple of the $x_i$'s, let



$N_n(u_1, \ldots, u_k)$ be the frequency of the $k$-tuple in the first $n+k-1$ trials, that is, the number of $t$ such that $1 \le t \le n$ and $X_t = u_1, \ldots, X_{t+k-1} = u_k$. Show that with probability 1, all asymptotic relative frequencies are what they should be—that is, with probability 1, $n^{-1}N_n(u_1, \ldots, u_k) \to p(u_1) \cdots p(u_k)$ for every $k$ and every $k$-tuple $u_1, \ldots, u_k$.

**6.13.** † A number $\omega$ in the unit interval is *completely normal* if, for every base $b$ and every $k$ and every $k$-tuple of base-$b$ digits, the $k$-tuple appears in the base-$b$ expansion of $\omega$ with asymptotic relative frequency $b^{-k}$. Show that the set of completely normal numbers has Lebesgue measure 1.

**6.14.** *Shannon's theorem.* Suppose that $X_1, X_2, \ldots$ are independent, identically distributed random variables taking on the values $1, \ldots, r$ with positive probabilities $p_1, \ldots, p_r$. If $p_n(i_1, \ldots, i_n) = p_{i_1} \ldots p_{i_n}$ and $p_n(\omega) = p_n(X_1(\omega), \ldots, X_n(\omega))$, then $p_n(\omega)$ is the probability that a new sequence of $n$ trials would produce the particular sequence $X_1(\omega), \ldots, X_n(\omega)$ of outcomes that happens actually to have been observed. Show that

$$-\frac{1}{n}\log p_n(\omega) \to h = -\sum_{i=1}^r p_i \log p_i$$

with probability 1.

In information theory $1, \ldots, r$ are interpreted as the *letters* of an *alphabet*, $X_1, X_2, \ldots$ are the successive letters produced by an information *source*, and $h$ is the *entropy* of the source. Prove the *asymptotic equipartition property*: For large $n$ there is probability exceeding $1 - \epsilon$ that the probability $p_n(\omega)$ of the observed $n$-long sequence, or *message*, is in the range $e^{-n(h \pm \epsilon)}$.

**6.15.** In the terminology of Example 6.5, show that $\log_2 n + \log_2 \log_2 n + \theta \log_2 \log_2 \log_2 n$ is an outer or inner boundary as $\theta > 1$ or $\theta \le 1$. Generalize. (Compare Problem 4.12.)

**6.16.** 5.20† Let $g(m) = \sum_p \delta_p(m)$ be the number of distinct prime divisors of $m$. For $a_n = E_n[g]$ (see (5.46)) show that $a_n \to \infty$. Show that

$$E_n\left[\left(\delta_p - \frac{1}{n}\left\lfloor\frac{n}{p}\right\rfloor\right)\left(\delta_q - \frac{1}{n}\left\lfloor\frac{n}{q}\right\rfloor\right)\right] \le \frac{1}{np} + \frac{1}{nq} \tag{6.8}$$

for $p \ne q$ and hence that the variance of $g$ under $P_n$ satisfies

$$\operatorname{Var}_n[g] \le 3\sum_{p \le n}\frac{1}{p}. \tag{6.9}$$

Prove the *Hardy–Ramanujan theorem*:

$$\lim_n P_n\left[m: \left|\frac{g(m)}{a_n} - 1\right| \ge \epsilon\right] = 0. \tag{6.10}$$

Since $a_n \sim \log\log n$ (see Problem 18.17), most integers under $n$ have something like $\log\log n$ distinct prime divisors. Since $\log\log 10^7$ is a little less than 3, the typical integer under $10^7$ has about three prime factors—remarkably few.