

ECE 534: Random Processes

Lecture Notes: Large Deviations, Concentration,
Random Processes, and Random Walks

1 Large Deviations and Concentration

Throughout this section, $X_1, X_2, \dots$ are i.i.d. random variables with the same distribution as a generic random variable X , mean $\mu = \mathbb{E}[X]$, and partial sums

$$S_n = X_1 + X_2 + \dots + X_n.$$

The law of large numbers tells us that $S_n/n \rightarrow \mu$. The results in this section quantify *how unlikely* it is for S_n/n to be far from μ .

1.1 The Chernoff bound

Definition 1.1 (Log moment generating function and rate function). The *moment generating function* of X is $M(s) = \mathbb{E}[e^{sX}]$, $s \in \mathbb{R}$, and the *log moment generating function* is $\Lambda(s) = \ln M(s)$ (possibly $+\infty$ for some s). The *rate function* of X is

$$\ell(t) = \sup_{s \in \mathbb{R}} (st - \Lambda(s)), \quad t \in \mathbb{R}.$$

Note that $\ell(t) \geq 0$ for all t (take $s = 0$), and $\ell(\mu) = 0$. For $t > \mu$, the supremum may be restricted to $s \geq 0$.

Remark 1.2 (Legendre transform). The rate function is the *Legendre transform* (more precisely, the Legendre–Fenchel transform, or convex conjugate) of the log moment generating function: $\ell = \Lambda^*$, where

$$\Lambda^*(t) = \sup_{s \in \mathbb{R}} (st - \Lambda(s)).$$

Since Λ is convex (by Hölder’s inequality), ℓ is convex as well, being a supremum of affine functions of t . Geometrically, $\ell(t)$ is the largest gap between the line $s \mapsto st$ and the curve $s \mapsto \Lambda(s)$; when Λ is differentiable, the supremum is attained at the s^* with $\Lambda'(s^*) = t$.

Theorem 1.3 (Chernoff bound). *For any random variable X and any $t \in \mathbb{R}$,*

$$\mathbb{P}\{X \geq t\} \leq \exp\left(-\sup_{s \geq 0} (st - \Lambda(s))\right).$$

Consequently, if $X_1, \dots, X_n$ are i.i.d. copies of X and $t > \mu$,

$$\mathbb{P}\left\{\frac{S_n}{n} \geq t\right\} \leq \exp\left(-n \sup_{s \geq 0} (st - \Lambda(s))\right) = e^{-n \ell(t)}.$$

Remark 1.4. The bound for a single variable follows from Markov's inequality applied to e^{sX} ; the bound for sums uses $\mathbb{E}[e^{sS_n}] = M(s)^n$, which holds by independence. The important message: the probability of a deviation of size $t - \mu > 0$ decays *exponentially* in n , with exponent at least $\ell(t)$. (Similarly, $\mathbb{P}\{S_n/n \leq t\} \leq e^{-n\ell(t)}$ for $t < \mu$.)

Example 1.5 (Two standard rate functions).

(a) **Gaussian.** If $X \sim N(\mu, \sigma^2)$, then $\Lambda(s) = \mu s + \sigma^2 s^2/2$ and

$$\ell(t) = \frac{(t - \mu)^2}{2\sigma^2}.$$

(b) **Bernoulli.** If $X \sim \text{Bernoulli}(p)$, then $\Lambda(s) = \ln(1 - p + pe^s)$ and, for $0 \leq t \leq 1$,

$$\ell(t) = t \ln \frac{t}{p} + (1 - t) \ln \frac{1 - t}{1 - p} = D(\text{Bern}(t) \parallel \text{Bern}(p)),$$

the Kullback–Leibler divergence; $\ell(t) = +\infty$ for $t \notin [0, 1]$.

1.2 Cramér's theorem: the Chernoff exponent is the right one

The Chernoff bound is an *upper* bound. A natural question is whether the exponent $\ell(t)$ can be improved. Cramér's theorem says it cannot: the Chernoff bound captures exactly the right exponential rate.

Theorem 1.6 (Cramér's theorem). *Let $X_1, X_2, \dots$ be i.i.d. with finite mean μ , and let $t > \mu$. Then for every $\varepsilon > 0$ there is an n_ε such that*

$$\mathbb{P}\left\{\frac{S_n}{n} \geq t\right\} \geq e^{-n(\ell(t) + \varepsilon)} \quad \text{for all } n \geq n_\varepsilon.$$

Combined with the Chernoff bound,

$$\boxed{\lim_{n \rightarrow \infty} \frac{1}{n} \ln \mathbb{P}\left\{\frac{S_n}{n} \geq t\right\} = -\ell(t).}$$

The easy-to-remember version. For $t > \mu$,

$$\mathbb{P}\left\{\frac{S_n}{n} \geq t\right\} = e^{-n(\ell(t) + o(1))}, \quad n \rightarrow \infty.$$

In words:

- The **Chernoff bound** gives the upper bound $e^{-n\ell(t)}$, valid for *every* n .
- **Cramér's theorem** gives a matching lower bound $e^{-n(\ell(t) + \varepsilon)}$ for *large* n .
- Together: the probability of the rare event $\{S_n/n \geq t\}$ decays exponentially in n with exponent *exactly* $\ell(t)$. Only sub-exponential factors (such as $1/\sqrt{n}$) are missed by the Chernoff bound.

Remark 1.7. Cramér's theorem is a statement about exponents only. For example, in the Gaussian case, $\mathbb{P}\{S_n/n \geq t\} = Q(\sqrt{n}(t - \mu)/\sigma)$, which behaves like $\frac{c}{\sqrt{n}} e^{-n(t - \mu)^2/(2\sigma^2)}$. The prefactor $c/\sqrt{n}$ is invisible on the scale $\frac{1}{n} \ln(\cdot)$, and the exponent is $\ell(t) = (t - \mu)^2/(2\sigma^2)$, as predicted.

1.3 McDiarmid's inequality (bounded differences)

The Chernoff bound concerns *sums* of independent variables. McDiarmid's inequality extends exponential concentration to *general functions* of independent variables, provided no single variable has too much influence.

Definition 1.8 (Bounded differences property). A function $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \rightarrow \mathbb{R}$ has the *bounded differences property* with constants $c_1, \dots, c_n \geq 0$ if, for each i ,

$$|f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, x'_i, \dots, x_n)| \leq c_i$$

for all $x_1, \dots, x_n$ and x'_i . That is, changing the i -th coordinate alone changes f by at most c_i .

Theorem 1.9 (McDiarmid's inequality). Let $X_1, \dots, X_n$ be independent random variables, $X_i \in \mathcal{X}_i$, and let f have the bounded differences property with constants $c_1, \dots, c_n$. Let $Z = f(X_1, \dots, X_n)$. Then for every $t > 0$,

$$\mathbb{P}\{Z - \mathbb{E}[Z] \geq t\} \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right), \quad \mathbb{P}\{|Z - \mathbb{E}[Z]| \geq t\} \leq 2 \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

Remark 1.10.

- (i) **Hoeffding as a special case.** If $X_i \in [\alpha_i, \beta_i]$ and $f(x) = \sum_i x_i$, then $c_i = \beta_i - \alpha_i$ and McDiarmid's inequality reduces to Hoeffding's inequality. In particular, if $X_i \in [0, 1]$ and $f = S_n/n$, then $c_i = 1/n$ and $\mathbb{P}\{|S_n/n - \mu| \geq t\} \leq 2e^{-2nt^2}$.
- (ii) **Typical applications.** Empirical risk and generalization error in machine learning, the chromatic number or number of isolated vertices of a random graph, the number of empty bins in balls-into-bins, and the length of the longest common subsequence of random strings.
- (iii) **Proof.** We will prove McDiarmid's inequality later in the semester, after we develop martingales. The proof builds the Doob martingale $Z_k = \mathbb{E}[Z \mid X_1, \dots, X_k]$, observes that its increments are bounded, and applies the Azuma–Hoeffding inequality.

2 A Few Concluding Remarks

We close our discussion of convergence of random variables with two facts that are frequently used.

2.1 Limits of Gaussian sequences are Gaussian

Recall that a random variable is called Gaussian with mean μ and variance $\sigma^2 \geq 0$ if its characteristic function is $\Phi(u) = e^{j\mu u - \sigma^2 u^2/2}$. We allow $\sigma^2 = 0$: a *degenerate* Gaussian $N(\mu, 0)$ is simply the constant μ .

Proposition 2.1. Let $(X_n)_{n \geq 1}$ be a jointly Gaussian sequence of random variables, with $X_n \sim N(\mu_n, \sigma_n^2)$.

- (a) If $X_n \rightarrow X$ in any of the four senses (a.s., m.s., p., or d.), then X is Gaussian, possibly degenerate. Moreover $\mu_n \rightarrow \mu$ and $\sigma_n^2 \rightarrow \sigma^2$, where $X \sim N(\mu, \sigma^2)$.
- (b) If $X_n \rightarrow X$ in probability (in particular a.s. or m.s.), then the limit is jointly Gaussian with the sequence: for every k , the vector $(X_1, \dots, X_k, X)$ is jointly Gaussian.

Remark 2.2. Idea of proof of (a). All four modes of convergence imply convergence in distribution, so it suffices to treat that case. Then $\Phi_n(u) = e^{j\mu_n u - \sigma_n^2 u^2/2} \rightarrow \Phi_X(u)$ for every u . Taking absolute values, $e^{-\sigma_n^2 u^2/2} \rightarrow |\Phi_X(u)|$; since Φ_X is continuous with $\Phi_X(0) = 1$, the limit is nonzero near $u = 0$, which forces $\sigma_n^2 \rightarrow \sigma^2$ for some finite $\sigma^2 \geq 0$. One then shows that μ_n converges to a finite μ , so $\Phi_X(u) = e^{j\mu u - \sigma^2 u^2/2}$. For (b), apply (a) to arbitrary linear combinations $a_1 X_1 + \dots + a_k X_k + b X_n$, which are Gaussian and converge in probability to $a_1 X_1 + \dots + a_k X_k + b X$.

Example 2.3. If $X_n \sim N(0, 1/n)$, then $X_n \rightarrow 0$ in m.s. The limit is the degenerate Gaussian $N(0, 0)$. This shows why degenerate Gaussians must be included in the statement.

2.2 Convergence in distribution to a constant

In general, convergence in distribution is the *weakest* of the four modes and does *not* imply convergence in probability. There is, however, one important exception.

Proposition 2.4. If $X_n \xrightarrow{d} c$, where c is a constant, then $X_n \xrightarrow{p} c$.

Proof. The CDF of the constant c is $F(x) = \mathbf{1}_{\{x \geq c\}}$, which is continuous at every $x \neq c$. Hence for any $\varepsilon > 0$, $F_{X_n}(c - \varepsilon) \rightarrow 0$ and $F_{X_n}(c + \varepsilon/2) \rightarrow 1$. Therefore

$$\mathbb{P}\{|X_n - c| \geq \varepsilon\} \leq F_{X_n}(c - \varepsilon) + (1 - F_{X_n}(c + \varepsilon/2)) \rightarrow 0. \quad \square$$

Example 2.5 (Without a constant limit, the implication fails). Let $X \sim N(0, 1)$ and define $X_n = -X$ for all n . By symmetry of the standard Gaussian, $-X \sim N(0, 1)$, so every X_n has exactly the same distribution as X . Hence, trivially,

$$X_n \xrightarrow{d} X.$$

However, $|X_n - X| = 2|X|$ for every n , so for any $\varepsilon > 0$,

$$\mathbb{P}\{|X_n - X| \geq \varepsilon\} = \mathbb{P}\{|X| \geq \varepsilon/2\} = 2Q(\varepsilon/2) > 0,$$

which does not depend on n and does not tend to 0. Thus X_n does *not* converge to X in probability. The point is that convergence in distribution only involves the marginal distributions of X_n and X separately, and says nothing about how they are coupled on the probability space; when the limit is a constant, there is no coupling to worry about.

3 Random Processes

3.1 Definitions

Definition 3.1 (Random process). Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\mathbb{T}$ be an index set. A *random process* (or *stochastic process*) is a collection

$$X = (X_t : t \in \mathbb{T})$$

of random variables, all defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We write $X_t(\omega)$ or $X(t, \omega)$ for the value of X_t at outcome $\omega \in \Omega$.

- If $\mathbb{T}$ is countable (typically $\mathbb{T} = \mathbb{Z}$ or $\mathbb{T} = \mathbb{Z}_+$), X is a *discrete-time* random process, also called a random sequence.

- If $\mathbb{T}$ is an interval (typically $\mathbb{T} = \mathbb{R}$ or $\mathbb{T} = \mathbb{R}_+ = [0, \infty)$), X is a *continuous-time* random process.

Thus a random process is a function of two variables, $X : \mathbb{T} \times \Omega \rightarrow \mathbb{R}$. There are two complementary ways of looking at it:

- **Fix t :** $\omega \mapsto X_t(\omega)$ is a random variable.
- **Fix ω :** $t \mapsto X_t(\omega)$ is an ordinary (deterministic) function of time.

Definition 3.2 (Sample path). For a fixed $\omega \in \Omega$, the function

$$t \mapsto X_t(\omega), \quad t \in \mathbb{T},$$

is called the *sample path* (or *realization*, or *trajectory*) of X corresponding to ω . The process may be viewed as a random experiment whose outcome is an entire function of time.

3.2 Examples of continuous-time processes and their sample paths

Example 3.3 (Random-phase sinusoid). Let $X_t = A \cos(2\pi f_0 t + \Theta)$, $t \in \mathbb{R}$, where $A > 0$ and f_0 are constants and $\Theta \sim \text{Unif}[0, 2\pi]$. Every sample path is a sinusoid of amplitude A and frequency f_0 ; the randomness lies entirely in the phase.

Example 3.4 (Random line). Let $X_t = A + Bt$, where A and B are random variables. Each sample path is a straight line with random intercept $A(\omega)$ and slope $B(\omega)$.

Example 3.5 (Poisson counting process). N_t counts the number of arrivals (customers, packets, photons) in $[0, t]$, where interarrival times are i.i.d. $\text{Exp}(\lambda)$. Sample paths are nondecreasing, integer-valued, right-continuous staircase functions with unit jumps at the arrival times, and $N_t \sim \text{Poisson}(\lambda t)$.

Example 3.6 (Random telegraph signal). $X_t \in \{-1, +1\}$ switches sign at the arrival times of a Poisson process. Sample paths are piecewise-constant functions toggling between ± 1 .

Example 3.7 (Brownian motion / Wiener process). $W = (W_t : t \geq 0)$ with $W_0 = 0$, independent increments, and $W_t - W_s \sim N(0, \sigma^2(t - s))$ for $t > s$. Sample paths are continuous but (with probability one) nowhere differentiable. Brownian motion is the continuous-time limit of a scaled random walk.

Example 3.8 (Discrete-time examples, for contrast). An i.i.d. sequence (X_k) (e.g. Bernoulli trials or white Gaussian noise samples), and the random walk $S_n = X_1 + \cdots + X_n$, whose sample paths are sequences that move up or down at each step.

3.3 Mean, correlation, covariance, and finite-dimensional distributions

Definition 3.9 (Second-order moment functions). Let $X = (X_t : t \in \mathbb{T})$ satisfy $\mathbb{E}[X_t^2] < \infty$ for all t (a *second-order* process). Then:

$$\text{Mean function: } \mu_X(t) = \mathbb{E}[X_t],$$

$$\text{Correlation function: } R_X(s, t) = \mathbb{E}[X_s X_t],$$

$$\text{Covariance function: } C_X(s, t) = \text{Cov}(X_s, X_t) = \mathbb{E}[(X_s - \mu_X(s))(X_t - \mu_X(t))] = R_X(s, t) - \mu_X(s)\mu_X(t),$$

for $s, t \in \mathbb{T}$. Note $C_X(t, t) = \text{Var}(X_t)$.

Remark 3.10. R_X and C_X are symmetric, $R_X(s, t) = R_X(t, s)$, and positive semidefinite: for any n , any $t_1, \dots, t_n \in \mathbb{T}$, and any $a_1, \dots, a_n \in \mathbb{R}$,

$$\sum_{i=1}^n \sum_{k=1}^n a_i a_k C_X(t_i, t_k) = \text{Var}\left(\sum_{i=1}^n a_i X_{t_i}\right) \geq 0.$$

Definition 3.11 (n -th order CDFs). For $n \geq 1$ and times $t_1, \dots, t_n \in \mathbb{T}$, the n -th order CDF of X is the joint CDF of $(X_{t_1}, \dots, X_{t_n})$:

$$F_{X,n}(x_1, \dots, x_n; t_1, \dots, t_n) = \mathbb{P}\{X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n\}.$$

The collection of all of these, for all n and all choices of $t_1, \dots, t_n$, is called the family of *finite-dimensional distributions* of X .

Remark 3.12. The finite-dimensional distributions determine all probabilities involving finitely many (and, by limits, countably many) samples of the process. The mean and correlation functions are determined by the first- and second-order CDFs. Conversely, Kolmogorov's extension theorem says that any consistent family of finite-dimensional distributions arises from some random process.

3.4 Stationarity

Definition 3.13 (Strict-sense stationarity). A random process $X = (X_t : t \in \mathbb{T})$ is *strict-sense stationary* (SSS) if its finite-dimensional distributions are invariant under time shifts: for all n , all $t_1, \dots, t_n$, and all s with $t_i + s \in \mathbb{T}$,

$$(X_{t_1+s}, \dots, X_{t_n+s}) \stackrel{d}{=} (X_{t_1}, \dots, X_{t_n}),$$

i.e. $F_{X,n}(x_1, \dots, x_n; t_1 + s, \dots, t_n + s) = F_{X,n}(x_1, \dots, x_n; t_1, \dots, t_n)$.

Definition 3.14 (Wide-sense stationarity). A second-order random process X is *wide-sense stationary* (WSS) if

- (i) $\mu_X(t) = \mu_X$ does not depend on t , and
- (ii) $R_X(s, t)$ depends on s and t only through $s - t$.

In that case we write $R_X(\tau) = R_X(t + \tau, t)$ and $C_X(\tau) = R_X(\tau) - \mu_X^2$.

Remark 3.15.

- (i) SSS + finite second moments $\Rightarrow$ WSS, because the mean depends only on first-order distributions and the correlation only on second-order distributions.
- (ii) WSS $\nRightarrow$ SSS in general (see below).
- (iii) For *Gaussian* processes, WSS $\Leftrightarrow$ SSS, since Gaussian finite-dimensional distributions are fully determined by means and covariances.

Example 3.16 (i.i.d. sequences are SSS). If (X_k) is i.i.d., then $(X_{k_1+s}, \dots, X_{k_n+s})$ has i.i.d. components with the common marginal, so its distribution does not depend on s . If $\mathbb{E}[X_k^2] < \infty$, it is also WSS, with $C_X(k) = \sigma^2 \mathbf{1}_{\{k=0\}}$ (discrete-time white noise).

Example 3.17 (Random-phase sinusoid is WSS (and in fact SSS)). For $X_t = A \cos(2\pi f_0 t + \Theta)$ with $\Theta \sim \text{Unif}[0, 2\pi]$: $\mu_X(t) = 0$, and using $\cos \alpha \cos \beta = \frac{1}{2}[\cos(\alpha - \beta) + \cos(\alpha + \beta)]$,

$$R_X(s, t) = \frac{A^2}{2} \cos(2\pi f_0(s - t)) + \frac{A^2}{2} \mathbb{E}[\cos(2\pi f_0(s + t) + 2\Theta)] = \frac{A^2}{2} \cos(2\pi f_0(s - t)).$$

So X is WSS with $R_X(\tau) = \frac{A^2}{2} \cos(2\pi f_0 \tau)$. Moreover, shifting time by s just shifts the phase by $2\pi f_0 s$, and $(\Theta + 2\pi f_0 s) \bmod 2\pi$ is again uniform, so X is SSS as well.

Example 3.18 (Random amplitude, fixed phase: not WSS). Let $X_t = A \cos(2\pi f_0 t)$ with $\mathbb{E}[A] = 0$, $\mathbb{E}[A^2] = 1$. Then $\mu_X(t) = 0$ but $R_X(s, t) = \cos(2\pi f_0 s) \cos(2\pi f_0 t)$, which is not a function of $s - t$ alone. Not WSS, hence not SSS.

Example 3.19 (WSS but not SSS). Let $(X_k)_{k \in \mathbb{Z}}$ be independent, with $X_k \sim N(0, 1)$ for even k and X_k Rademacher (± 1 with probability $\frac{1}{2}$ each) for odd k . Then $\mu_X = 0$ and $R_X(k, l) = \mathbf{1}_{\{k=l\}}$ for all k, l , so X is WSS. But the first-order distribution changes with k , so X is not SSS.

Example 3.20 (Random walk and Brownian motion are not stationary). For the random walk $S_n = X_1 + \dots + X_n$ with i.i.d. zero-mean steps of variance σ^2 , $\text{Var}(S_n) = n\sigma^2$ grows with n . For Brownian motion, $R_W(s, t) = \sigma^2 \min(s, t)$. Neither is WSS. Their *increments*, however, are stationary.

4 The Gambler's Ruin Problem and Other Random Walks

4.1 Setup

A gambler starts with an initial capital of k dollars, where $k \in \{0, 1, \dots, b\}$ for some target $b \geq 1$. At each round the gambler bets one dollar on a fair coin. Let $\xi_1, \xi_2, \dots$ be i.i.d. *Rademacher* random variables,

$$\mathbb{P}\{\xi_i = +1\} = \mathbb{P}\{\xi_i = -1\} = \frac{1}{2},$$

where $\xi_i = +1$ means the gambler wins round i . The fortune after n rounds is the simple symmetric random walk

$$S_n = k + \xi_1 + \xi_2 + \dots + \xi_n, \quad S_0 = k.$$

The game stops the first time the gambler is either ruined or reaches the target:

$$\tau = \min\{n \geq 0 : S_n \in \{0, b\}\}.$$

We want

$$p_k = \mathbb{P}\{S_\tau = b \mid S_0 = k\} = \mathbb{P}\{\text{reach } b \text{ dollars before going bankrupt} \mid S_0 = k\}.$$

Remark 4.1 (The game ends with probability one). In any block of b consecutive rounds, if all b coin flips are wins the game must have ended. Such a block occurs with probability 2^{-b} , and disjoint blocks are independent, so $\mathbb{P}\{\tau > mb\} \leq (1 - 2^{-b})^m \rightarrow 0$ as $m \rightarrow \infty$. Hence $\tau < \infty$ a.s., and p_k is well defined. (In fact $\mathbb{E}[\tau] < \infty$.)

4.2 First-step analysis

The idea of *first-step analysis* is to condition on the outcome of the first round. After one round, the walk is at $k + 1$ or $k - 1$, and because the coin flips are i.i.d., the future evolves exactly like a fresh copy of the game started from the new position (the Markov property).

Step 1: Condition on the first step. For $1 \leq k \leq b-1$, by the law of total probability,

$$\begin{aligned} p_k &= \mathbb{P}\{\xi_1 = +1\} \mathbb{P}\{\text{reach } b \text{ first} \mid S_1 = k+1\} + \mathbb{P}\{\xi_1 = -1\} \mathbb{P}\{\text{reach } b \text{ first} \mid S_1 = k-1\} \\ &= \tfrac{1}{2} p_{k+1} + \tfrac{1}{2} p_{k-1}. \end{aligned}$$

Step 2: Boundary conditions. If the gambler starts at 0 the game is already lost; if at b , already won:

$$p_0 = 0, \quad p_b = 1.$$

Step 3: Solve the difference equation. Rewrite the recursion as

$$p_{k+1} - p_k = p_k - p_{k-1}, \quad 1 \leq k \leq b-1.$$

So the increments $\Delta_k = p_k - p_{k-1}$ are all equal, say to Δ : the sequence (p_k) is *linear* in k , $p_k = p_0 + k\Delta = k\Delta$. The condition $p_b = 1$ gives $\Delta = 1/b$. Therefore

$$p_k = \frac{k}{b}, \quad k = 0, 1, \dots, b.$$

The probability of ruin is $1 - p_k = (b-k)/b$.

Remark 4.2 (Uniqueness). The linear system in Steps 1–2 has $b-1$ equations in the $b-1$ unknowns $p_1, \dots, p_{b-1}$, and the argument in Step 3 shows any solution must be k/b ; hence the solution is unique, and the answer is correct.

4.3 More with first-step analysis

Expected duration of the game. Let $m_k = \mathbb{E}[\tau \mid S_0 = k]$. Conditioning on the first step, one round is used and then the game restarts from $k \pm 1$:

$$m_k = 1 + \tfrac{1}{2} m_{k+1} + \tfrac{1}{2} m_{k-1}, \quad 1 \leq k \leq b-1, \quad m_0 = m_b = 0.$$

The second difference of m equals -2 , so m is a quadratic in k vanishing at 0 and b :

$$m_k = k(b-k).$$

For example, starting with \$50 and aiming for \$100, the game lasts 2500 rounds on average.

Biased coin. If instead $\mathbb{P}\{\xi_i = +1\} = p \neq \frac{1}{2}$ and $q = 1 - p$, the same method gives $p_k = p p_{k+1} + q p_{k-1}$, whose solution, with $r = q/p$, is

$$p_k = \frac{1 - r^k}{1 - r^b}.$$

Letting $p \rightarrow \frac{1}{2}$ (so $r \rightarrow 1$) recovers $p_k = k/b$.

Unbounded target. For the fair game, letting $b \rightarrow \infty$ gives $p_k = k/b \rightarrow 0$: a gambler playing a fair game against an infinitely rich opponent is ruined with probability one, yet $m_k = k(b-k) \rightarrow \infty$, so the expected time to ruin is infinite.

Remark 4.3 (A preview of martingales). (S_n) is a martingale. If one can justify $\mathbb{E}[S_\tau] = \mathbb{E}[S_0]$ (the *optional stopping theorem*, later in the semester), then $k = \mathbb{E}[S_\tau] = b p_k + 0 \cdot (1 - p_k)$, giving $p_k = k/b$ in one line. Similarly, $S_n^2 - n$ is a martingale, and optional stopping yields $\mathbb{E}[\tau] = k(b-k)$.

4.4 Other random walk problems in engineering, mathematics, and ML/AI

Random walks and first-step analysis appear far beyond gambling. A representative (non-exhaustive) list:

- 1. Sequential hypothesis testing (Wald’s SPRT).** To decide between H_0 and H_1 from i.i.d. observations $Y_1, Y_2, \dots$, track the log-likelihood ratio $L_n = \sum_{i=1}^n \ln \frac{f_1(Y_i)}{f_0(Y_i)}$, a random walk. Stop and declare H_1 when $L_n \geq B$, or H_0 when $L_n \leq -A$. Error probabilities are gambler’s-ruin probabilities with two absorbing barriers.
- 2. Quickest change detection (CUSUM).** The statistic $W_n = \max(0, W_{n-1} + Z_n)$, with Z_n a log-likelihood ratio increment, is a random walk reflected at 0; an alarm is raised when W_n crosses a threshold. Used in quality control, network intrusion detection, and fault monitoring.
- 3. Queues and buffers (Lindley recursion).** The waiting time of the n -th customer satisfies $W_{n+1} = \max(0, W_n + S_n - A_{n+1})$, a random walk reflected at 0. Buffer overflow probabilities in communication networks are computed via random-walk maxima, and their tails via Chernoff/Cramér-type exponents.
- 4. Recurrence and transience (Pólya’s theorem).** The simple symmetric random walk on $\mathbb{Z}^d$ returns to its starting point with probability one if $d = 1, 2$ (recurrent), but with probability less than one if $d \geq 3$ (transient).
- 5. Reflection principle and the ballot problem.** For the simple symmetric random walk, $\mathbb{P}\{\max_{m \leq n} S_m \geq a\} = 2\mathbb{P}\{S_n \geq a\} - \mathbb{P}\{S_n = a\}$ for integers $a > 0$ (with $S_0 = 0$); the same reflection argument solves the ballot problem and underlies results on Brownian motion.
- 6. Randomized algorithms.** Papadimitriou’s random-walk algorithm for 2-SAT flips a variable from an unsatisfied clause; the number of variables agreeing with a fixed satisfying assignment performs a walk that moves toward the target with probability at least $\frac{1}{2}$. A gambler’s-ruin argument ($m_k = k(b-k)$ -type bound) shows a satisfying assignment is found in $O(n^2)$ expected steps.
- 7. Random walks on graphs: PageRank and node embeddings.** PageRank is the stationary distribution of a random walk on the web graph with random restarts. Hitting times, commute times, and cover times of graph random walks measure connectivity; methods such as DeepWalk and node2vec learn node embeddings from sampled random-walk sequences.
- 8. Markov chain Monte Carlo.** Random-walk Metropolis–Hastings proposes $x' = x + \varepsilon$ and accepts with a probability chosen so that the walk’s stationary distribution is a desired target, e.g. a Bayesian posterior. Mixing-time analysis of such walks is a central theme in statistics and ML.
- 9. Stochastic optimization and diffusion models.** Stochastic gradient descent with small step sizes behaves like a random walk in a potential (Langevin dynamics), and Langevin sampling deliberately adds Gaussian noise. Diffusion generative models corrupt data by a Gaussian random walk (a discretized diffusion) and learn to reverse it.
- 10. Reinforcement learning.** The “random walk” Markov reward process (a chain of states with absorbing ends, exactly the gambler’s ruin chain) is a standard benchmark for temporal-difference learning: the true value of state k is the probability p_k of exiting at the rewarding end, computed by first-step analysis (the Bellman equation).

Remark 4.4. In items 1, 6, and 10, the key computation is a *first-step analysis* of exactly the form above: write an equation for the quantity of interest by conditioning on the first move, impose boundary conditions, and solve. In RL, this equation is precisely the Bellman equation.