ECE 534: Random Processes

Gambler's Ruin, Recurrences, Conditional Expectation,
Martingales, Markov Processes, and Brownian Motion

Lecture Notes
University of Illinois Urbana-Champaign

1 The Gambler's Ruin Problem

1.1 Setup

Let $X_1, X_2, \ldots$ be independent and identically distributed (i.i.d.) *Rademacher* random variables:

$$\mathbb{P}(X_i = +1) = \mathbb{P}(X_i = -1) = \tfrac{1}{2}.$$

A gambler starts with k dollars, where $k \in \{0, 1, \ldots, b\}$, and at each round bets one dollar on a fair coin. Her fortune after n rounds is the *simple symmetric random walk*

$$S_0 = k, \qquad S_n = k + X_1 + X_2 + \cdots + X_n, \quad n \geq 1.$$

She stops playing as soon as she is either ruined ($S_n = 0$) or reaches her target ($S_n = b$). Define the *stopping time*

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

We want to compute

$$p_k = \mathbb{P}(S_\tau = b \,|\, S_0 = k) \quad \text{(probability of winning)}, \qquad D_k = \mathbb{E}[\tau \,|\, S_0 = k] \quad \text{(expected duration)}.$$

1.2 First-step analysis

The key idea is to condition on the outcome of the *first* bet. If $1 \leq k \leq b-1$, then after one step the walk is at $k+1$ or $k-1$, each with probability $\frac{1}{2}$, and—because the X_i are i.i.d.—the future looks exactly like a fresh game started from the new position. By the law of total probability,

$$p_k = \tfrac{1}{2}\, p_{k+1} + \tfrac{1}{2}\, p_{k-1}, \qquad 1 \leq k \leq b-1,$$

with *boundary conditions* $p_0 = 0$ and $p_b = 1$.

Similarly, the first step costs one unit of time, so

$$D_k = 1 + \tfrac{1}{2}\, D_{k+1} + \tfrac{1}{2}\, D_{k-1}, \qquad 1 \leq k \leq b-1,$$

with $D_0 = D_b = 0$.

Both are *linear recurrences with constant coefficients*; the first is homogeneous and the second is inhomogeneous. We now develop a general method for solving them.

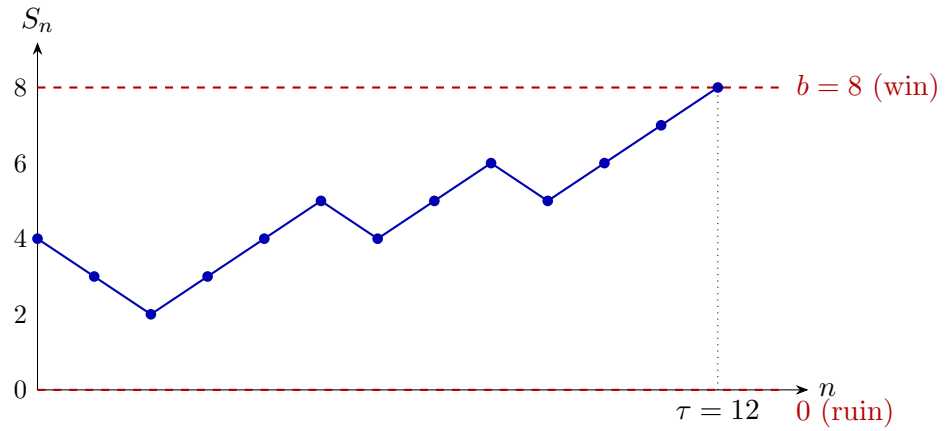


Figure 1: A sample path of the gambler's fortune with $k = 4$, $b = 8$. Here the gambler wins at time $\tau = 12$.

2 Solving Linear Recurrences

2.1 The Fibonacci recurrence

The Fibonacci numbers are defined by

$$F_0 = 0, \quad F_1 = 1, \quad F_{n+1} = F_n + F_{n-1} \quad (n \geq 1),$$

giving $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$ This is a second-order homogeneous recurrence with *initial* (rather than boundary) conditions.

Proposition 2.1 (Binet's formula)**.** *With $\varphi = \frac{1+\sqrt{5}}{2}$ (the golden ratio) and $\psi = \frac{1-\sqrt{5}}{2} = -1/\varphi$,*

$$F_n = \frac{\varphi^n - \psi^n}{\sqrt{5}}, \qquad n \geq 0.$$

Proof. Substituting $F_n = r^n$ gives $r^2 = r + 1$, i.e. $r^2 - r - 1 = 0$, with distinct roots φ, ψ. Hence $F_n = A\varphi^n + B\psi^n$. The initial conditions give

$$F_0 = A + B = 0, \qquad F_1 = A\varphi + B\psi = 1.$$

So $B = -A$ and $A(\varphi - \psi) = A\sqrt{5} = 1$, giving $A = 1/\sqrt{5}$, $B = -1/\sqrt{5}$. □

Remark 2.2. Since $|\psi| < 1$, the term $\psi^n/\sqrt{5}$ vanishes rapidly: F_n is the integer nearest to $\varphi^n/\sqrt{5}$, and $F_{n+1}/F_n \to \varphi$. In general, the root of largest modulus dominates the long-run growth of a linear recurrence. Equivalently, in matrix form

$$\begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F_n \\ F_{n-1} \end{pmatrix},$$

and φ, ψ are exactly the eigenvalues of this matrix.

2.2 The general method

The trick used for Fibonacci—try a geometric sequence r^n—works for every linear recurrence with constant coefficients. A linear recurrence of order d with constant coefficients has the form

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_d a_{n-d} + g(n), \tag{1}$$

where $c_1, \ldots, c_d$ are constants ($c_d \neq 0$) and g is a given forcing term. When $g \equiv 0$ the recurrence is *homogeneous*.

Step 1: Homogeneous solution. Try $a_n = r^n$ in the homogeneous equation. Dividing by r^{n-d} gives the *characteristic equation*

$$r^d - c_1 r^{d-1} - c_2 r^{d-2} - \cdots - c_d = 0.$$

- If the characteristic roots $r_1, \ldots, r_d$ are *distinct*, the general homogeneous solution is
$$a_n^{(h)} = A_1 r_1^n + A_2 r_2^n + \cdots + A_d r_d^n.$$
- If a root r has *multiplicity* m, it contributes the m terms
$$\left(B_0 + B_1 n + B_2 n^2 + \cdots + B_{m-1} n^{m-1}\right) r^n.$$

(This is the exact analogue of solving linear ODEs with constant coefficients, with $e^{\lambda t}$ replaced by r^n.)

Step 2: Particular solution. If $g \not\equiv 0$, find *any* one solution $a_n^{(p)}$ of the full equation. Guess a form similar to g:

- $g(n) =$ constant $\Rightarrow$ try $a_n^{(p)} = C$;
- $g(n) =$ polynomial of degree $q \Rightarrow$ try a polynomial of degree q;
- $g(n) = \beta^n \Rightarrow$ try $C\beta^n$.

Resonance rule: if the guess already solves the homogeneous equation (e.g. β, or 1 for polynomial forcing, is a characteristic root of multiplicity m), multiply the guess by n^m.

Step 3: Combine and fit. The general solution is $a_n = a_n^{(h)} + a_n^{(p)}$. The d free constants are determined by d conditions—either *initial conditions* (e.g. a_0, a_1 given, as for Fibonacci) or *boundary conditions* (e.g. a_0 and a_b given, as in gambler's ruin).

2.3 Winning probability

Proposition 2.3. *For the symmetric gambler's ruin, $p_k = k/b$ for $k = 0, 1, \ldots, b$.*

Proof. Rewrite $p_k = \frac{1}{2}p_{k+1} + \frac{1}{2}p_{k-1}$ as $p_{k+1} - 2p_k + p_{k-1} = 0$. The characteristic equation is

$$r^2 - 2r + 1 = (r-1)^2 = 0,$$

so $r = 1$ is a double root, and the general solution is $p_k = A + Bk$. The boundary conditions give $p_0 = A = 0$ and $p_b = B\,b = 1$, so $p_k = k/b$. □

bote the interpretation: p_k is *linear* in k. A gambler with k dollars playing a fair game against a "house" with $b - k$ dollars wins with probability proportional to her share of the total wealth.

2.4 Expected duration

Proposition 2.4. *For the symmetric gambler's ruin, $D_k = k(b-k)$.*

Proof. The recurrence is $D_{k+1} - 2D_k + D_{k-1} = -2$. The homogeneous solution is again $A + Bk$. The forcing is constant, so we would normally try $D_k^{(p)} = C$; but 1 is a characteristic root of multiplicity 2, so by the resonance rule we try $D_k^{(p)} = Ck^2$:

$$C\big[(k+1)^2 - 2k^2 + (k-1)^2\big] = 2C = -2 \quad \Longrightarrow \quad C = -1.$$

Thus $D_k = A + Bk - k^2$. From $D_0 = 0$ we get $A = 0$; from $D_b = 0$ we get $B\,b - b^2 = 0$, i.e. $B = b$. Hence $D_k = bk - k^2 = k(b-k)$. □

For example, starting from the middle ($k = b/2$) the game lasts on average $b^2/4$ rounds—*quadratic* in the distance to the boundaries. This $\sqrt{\text{time}}$ scaling of a fair random walk reappears in Brownian motion.

2.5 General p

Now let the coin be biased: $\mathbb{P}(X_i = +1) = p$ and $\mathbb{P}(X_i = -1) = q = 1 - p$, with $p \neq \frac{1}{2}$. First-step analysis gives $p_k = p\,p_{k+1} + q\,p_{k-1}$ and $D_k = 1 + p\,D_{k+1} + q\,D_{k-1}$, with the same boundary conditions as before. Solving these recurrences by the method above yields the following.

Proposition 2.5. *For $p \neq \frac{1}{2}$ and $k = 0, 1, \ldots, b$,*

$$p_k = \frac{1 - (q/p)^k}{1 - (q/p)^b}, \qquad D_k = \frac{k}{q-p} - \frac{b}{q-p} \cdot \frac{1 - (q/p)^k}{1 - (q/p)^b}.$$

Letting $p \to \frac{1}{2}$ recovers $p_k = k/b$ and $D_k = k(b-k)$.

The limits $p \to \frac{1}{2}$ are of the indeterminate form $0/0$ (both numerator and denominator vanish when $q/p = 1$). Such limits are computed with the following rule.

Theorem 2.6 (L'Hôpital's rule)**.** *Let $c \in [-\infty, \infty]$ and let I be an open interval containing c (or with c as an endpoint). Suppose f and g are differentiable on $I \setminus \{c\}$, with $g'(x) \neq 0$ for all $x \in I \setminus \{c\}$, and that either*

$$\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = 0 \qquad \text{or} \qquad \lim_{x \to c} |g(x)| = \infty.$$

If $\lim\limits_{x \to c} \dfrac{f'(x)}{g'(x)} = L$ exists, where $L \in [-\infty, \infty]$, then

$$\lim_{x \to c} \frac{f(x)}{g(x)} = L.$$

The same holds for one-sided limits $x \to c^+$ or $x \to c^-$.

Example 2.7 (Recovering the fair-coin winning probability)**.** Write $r = q/p = (1-p)/p$, so that $p \to \frac{1}{2}$ is the same as $r \to 1$. With $f(r) = 1 - r^k$ and $g(r) = 1 - r^b$, both vanish at $r = 1$, and $g'(r) = -b\,r^{b-1} \neq 0$ near $r = 1$. Hence

$$\lim_{p \to 1/2} p_k = \lim_{r \to 1} \frac{1 - r^k}{1 - r^b} = \lim_{r \to 1} \frac{-k\,r^{k-1}}{-b\,r^{b-1}} = \frac{k}{b},$$

which agrees with the fair-coin result.

Example 2.8 (Applying the rule twice)**.** For $\lim_{x\to 0}\dfrac{1-\cos x}{x^2}$, both numerator and denominator tend to 0, and the ratio of derivatives is $\frac{\sin x}{2x}$, again of the form $0/0$. Applying the rule once more gives $\frac{\cos x}{2}\to\frac{1}{2}$. Therefore

$$\lim_{x\to 0}\frac{1-\cos x}{x^2}=\lim_{x\to 0}\frac{\sin x}{2x}=\lim_{x\to 0}\frac{\cos x}{2}=\frac{1}{2}.$$

3 Martingales: A First Definition

In the fair game above, the gambler's expected fortune after the next bet, given everything that has happened so far, equals her current fortune:

$$\mathbb{E}[S_{n+1}\,|\,S_0,S_1,\dots,S_n]=S_n+\mathbb{E}[X_{n+1}]=S_n.$$

This "fair game" property is the defining feature of a martingale.

Definition 3.1 (Martingale, plain vanilla form)**.** A sequence of random variables $(X_n)_{n\ge 0}$ is a *martingale* if for all $n\ge 0$:

(i) $\mathbb{E}|X_n|<\infty$, and

(ii) $\mathbb{E}[X_{n+1}\,|\,X_0,X_1,\dots,X_n]=X_n$.

More generally, (X_n) is a martingale *with respect to* another sequence $(Y_n)_{n\ge 0}$ if each X_n is a function of $(Y_0,\dots,Y_n)$, $\mathbb{E}|X_n|<\infty$, and

$$\mathbb{E}[X_{n+1}\,|\,Y_0,Y_1,\dots,Y_n]=X_n.$$

To make sense of condition (ii)—a conditional expectation given a whole *collection* of random variables, which is itself a random variable—we first review conditional expectation.

4 Review of Conditional Expectation

Throughout, $(\Omega,\mathcal{F},\mathbb{P})$ is a probability space.

4.1 Conditioning on an event

Definition 4.1. Let B be an event with $\mathbb{P}(B)>0$ and X an integrable random variable. Then

$$\mathbb{E}[X\,|\,B]=\frac{\mathbb{E}[X\mathbf{1}_B]}{\mathbb{P}(B)}.$$

For discrete X this is $\sum_x x\,\mathbb{P}(X=x\,|\,B)$; for X with density f_X it is $\frac{1}{\mathbb{P}(B)}\int x f_X(x)\mathbf{1}_{\{x\in B\}}\,dx$ when $B=\{X\in B'\}$.

If $B_1,\dots,B_m$ partition Ω with $\mathbb{P}(B_i)>0$, the *law of total expectation* reads

$$\mathbb{E}[X]=\sum_{i=1}^{m}\mathbb{E}[X\,|\,B_i]\,\mathbb{P}(B_i).$$

(This is exactly what we used in the first-step analysis, with $B_1=\{X_1=+1\}$, $B_2=\{X_1=-1\}$.)

Example 4.2 (Die roll)**.** Let X be a fair die roll and $B=\{X\text{ even}\}$. Then $\mathbb{E}[X\,|\,B]=\frac{2+4+6}{3}=4$, while $\mathbb{E}[X\,|\,B^c]=3$. Indeed $\frac{1}{2}\cdot 4+\frac{1}{2}\cdot 3=3.5=\mathbb{E}[X]$.

Example 4.3 (Memorylessness)**.** Let $X \sim \mathrm{Exp}(\lambda)$ and $a > 0$. Then

$$\mathbb{E}[X \mid X > a] = \frac{\int_a^\infty x\lambda e^{-\lambda x}\,dx}{e^{-\lambda a}} = \frac{(a + 1/\lambda)e^{-\lambda a}}{e^{-\lambda a}} = a + \frac{1}{\lambda}.$$

Having waited a units, the expected *additional* waiting time is still $1/\lambda$.

4.2 Conditioning on a discrete random variable

Let X, Y be discrete. For each y with $p_Y(y) > 0$, the conditional pmf is

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}, \qquad \mathbb{E}[X \mid Y = y] = \sum_x x\, p_{X|Y}(x \mid y) =: g(y).$$

This is a *number* for each fixed y. The *conditional expectation of X given Y* is the *random variable*

$$\mathbb{E}[X \mid Y] := g(Y).$$

That is: first compute $\mathbb{E}[X \mid Y = y]$ as a function of y, then plug in the random Y.

4.3 Conditioning on a continuous random variable

If (X, Y) has joint density $f_{X,Y}$, then for $f_Y(y) > 0$,

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \qquad g(y) = \mathbb{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x\, f_{X|Y}(x \mid y)\,dx,$$

and again $\mathbb{E}[X \mid Y] := g(Y)$. Note $\mathbb{P}(Y = y) = 0$ here, so $\mathbb{E}[X \mid Y = y]$ *cannot* be defined via the event formula; the density-ratio definition is what makes sense (as a limit of conditioning on $\{y \le Y \le y + \delta\}$, $\delta \downarrow 0$).

4.4 Conditioning on several random variables

Conditioning on a collection $Y_0, \ldots, Y_n$ is the same as conditioning on the random vector $\mathbf{Y} = (Y_0, \ldots, Y_n)$: compute $g(y_0, \ldots, y_n) = \mathbb{E}[X \mid Y_0 = y_0, \ldots, Y_n = y_n]$ from the conditional pmf/pdf, and set

$$\mathbb{E}[X \mid Y_0, \ldots, Y_n] = g(Y_0, \ldots, Y_n).$$

This is precisely the object appearing in Definition 3.1.

4.5 Key properties

Let X, Z be integrable and Y a random variable (or vector). Then:

(P1) **Linearity:** $\mathbb{E}[aX + bZ \mid Y] = a\mathbb{E}[X \mid Y] + b\mathbb{E}[Z \mid Y]$.

(P2) **Law of total expectation:** $\mathbb{E}\big[\mathbb{E}[X \mid Y]\big] = \mathbb{E}[X]$.

(P3) **Taking out what is known:** $\mathbb{E}[h(Y)X \mid Y] = h(Y)\,\mathbb{E}[X \mid Y]$; in particular $\mathbb{E}[h(Y) \mid Y] = h(Y)$.

(P4) **Independence:** if X is independent of Y, then $\mathbb{E}[X \mid Y] = \mathbb{E}[X]$.

(P5) **Freezing:** if D is independent of Y, then $\mathbb{E}[\phi(Y, D) \mid Y] = \Phi(Y)$, where $\Phi(y) := \mathbb{E}[\phi(y, D)]$.

(P6) **Monotonicity:** $X \le Z \Rightarrow \mathbb{E}[X \mid Y] \le \mathbb{E}[Z \mid Y]$.

(P7) **Best predictor (MMSE):** if $\mathbb{E}[X^2] < \infty$, then $\mathbb{E}[X \mid Y]$ minimizes $\mathbb{E}[(X - h(Y))^2]$ over all functions h, and the error $X - \mathbb{E}[X \mid Y]$ is orthogonal to every $h(Y)$: $\mathbb{E}\big[(X - \mathbb{E}[X \mid Y])\,h(Y)\big] = 0$.

4.6 Examples

Example 4.4 (Two dice)**.** Let X_1, X_2 be independent fair dice and $S = X_1 + X_2$.

- $\mathbb{E}[S \mid X_1] = X_1 + \mathbb{E}[X_2 \mid X_1] = X_1 + \mathbb{E}[X_2] = X_1 + 3.5$ by (P1), (P3), (P4).
- $\mathbb{E}[X_1 \mid S]$: by symmetry $\mathbb{E}[X_1 \mid S] = \mathbb{E}[X_2 \mid S]$, and their sum is $\mathbb{E}[S \mid S] = S$. Hence $\mathbb{E}[X_1 \mid S] = S/2$.
- Check (P2): $\mathbb{E}[S/2] = 7/2 = \mathbb{E}[X_1]$. ✓

Example 4.5 (Uniform on a triangle)**.** Let (X, Y) be uniform on $T = \{(x, y) : 0 \le y \le x \le 1\}$, so $f_{X,Y} = 2$ on T.

- $f_X(x) = \int_0^x 2\,dy = 2x$ on $[0, 1]$, so $f_{Y|X}(y \mid x) = \frac{2}{2x} = \frac{1}{x}$ for $0 \le y \le x$: given $X = x$, Y is uniform on $[0, x]$. Hence $\mathbb{E}[Y \mid X] = X/2$.
- $f_Y(y) = \int_y^1 2\,dx = 2(1 - y)$, so given $Y = y$, X is uniform on $[y, 1]$ and $\mathbb{E}[X \mid Y] = (1 + Y)/2$.
- Check: $\mathbb{E}[X] = \int_0^1 x \cdot 2x\,dx = 2/3$ and $\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[X]/2 = 1/3$; also $\mathbb{E}[(1+Y)/2] = (1 + 1/3)/2 = 2/3 = \mathbb{E}[X]$. ✓

Example 4.6 (Jointly Gaussian)**.** If (X, Y) is jointly Gaussian with $\mathrm{Var}(Y) > 0$, then

$$\mathbb{E}[X \mid Y] = \mu_X + \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)}\,(Y - \mu_Y),$$

which is *linear* (affine) in Y. For Gaussians, the best predictor coincides with the best linear predictor.

5 σ-Algebras, the Tower Property, and Filtrations

5.1 σ-algebras as information

The σ-algebra *generated* by random variables $Y_0, \ldots, Y_n$, denoted $\sigma(Y_0, \ldots, Y_n)$, is the smallest σ-algebra containing all events of the form $\{Y_i \le y\}$. It represents "the information contained in observing $Y_0, \ldots, Y_n$." Given a σ-algebra $\mathcal{G}$, a random variable Z is *$\mathcal{G}$-measurable* if $\{Z \le z\} \in \mathcal{G}$ for every z ("Z is determined by the information in $\mathcal{G}$"). A basic fact (Doob–Dynkin lemma) is

$$Z \text{ is } \sigma(Y_0, \ldots, Y_n)\text{-measurable} \iff Z = h(Y_0, \ldots, Y_n) \text{ for some (Borel) function } h.$$

Definition 5.1 (Conditional expectation given a σ-algebra)**.** Let X be integrable and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. The conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is the (almost surely unique) random variable Z such that

(i) Z is $\mathcal{G}$-measurable, and

(ii) $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A]$ for every $A \in \mathcal{G}$.

We write $\mathbb{E}[X \mid Y_0, \ldots, Y_n] := \mathbb{E}[X \mid \sigma(Y_0, \ldots, Y_n)]$; this agrees with the pmf/pdf definitions of the previous section.

All properties (P1)–(P7) hold with Y replaced by a σ-algebra $\mathcal{G}$ ("$h(Y)$" becomes "a $\mathcal{G}$-measurable random variable," and "independent of Y" becomes "independent of $\mathcal{G}$").

5.2 The tower property

Theorem 5.2 (Tower property — "the smallest σ-algebra wins")**.** *If $\mathcal{G} \subseteq \mathcal{H}$ are σ-algebras and X is integrable, then*

$$\mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{H}] \,|\, \mathcal{G}\big] = \mathbb{E}[X \,|\, \mathcal{G}] = \mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{H}\big].$$

Proof. Right equality: $\mathbb{E}[X \,|\, \mathcal{G}]$ is $\mathcal{G}$-measurable, hence $\mathcal{H}$-measurable (since $\mathcal{G} \subseteq \mathcal{H}$), so conditioning on $\mathcal{H}$ leaves it unchanged by (P3).

Left equality: Let $Z = \mathbb{E}[X \,|\, \mathcal{H}]$. For any $A \in \mathcal{G}$, we also have $A \in \mathcal{H}$, so by the defining property of Z,

$$\mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{G}]\mathbf{1}_A\big].$$

Since $\mathbb{E}[X \,|\, \mathcal{G}]$ is $\mathcal{G}$-measurable and has the same integrals as Z over all sets $A \in \mathcal{G}$, it equals $\mathbb{E}[Z \,|\, \mathcal{G}]$. □

In words: whichever order you condition in, you end up conditioning on the *coarser* information—the smaller σ-algebra wins.

Example 5.3 (Tower property in action)**.**

1. **Trivial $\mathcal{G}$.** With $\mathcal{G} = \{\emptyset, \Omega\}$, $\mathbb{E}[\,\cdot \,|\, \mathcal{G}] = \mathbb{E}[\,\cdot\,]$, and the tower property reduces to $\mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{H}]\big] = \mathbb{E}[X]$.
2. **Fewer variables.** Since $\sigma(Y) \subseteq \sigma(Y, Z)$,
$$\mathbb{E}\big[\mathbb{E}[X \,|\, Y, Z] \,|\, Y\big] = \mathbb{E}[X \,|\, Y] = \mathbb{E}\big[\mathbb{E}[X \,|\, Y] \,|\, Y, Z\big].$$
3. **Functions of the conditioning variable.** Since $\sigma(Y^2) \subseteq \sigma(Y)$,
$$\mathbb{E}\big[\mathbb{E}[X \,|\, Y] \,|\, Y^2\big] = \mathbb{E}[X \,|\, Y^2].$$
For instance, if $X = Y$ with Y symmetric about 0, then $\mathbb{E}[Y \,|\, Y^2] = 0$ (given $|Y|$, the sign of Y is equally likely to be $\pm$), even though $\mathbb{E}[Y \,|\, Y] = Y$.
4. **Martingales, several steps ahead.** If (X_n) is a martingale with respect to (Y_n), set $\mathcal{G}_n = \sigma(Y_0, \ldots, Y_n)$. Then $\mathcal{G}_n \subseteq \mathcal{G}_{n+1}$, so
$$\mathbb{E}[X_{n+2} \,|\, \mathcal{G}_n] = \mathbb{E}\big[\mathbb{E}[X_{n+2} \,|\, \mathcal{G}_{n+1}] \,|\, \mathcal{G}_n\big] = \mathbb{E}[X_{n+1} \,|\, \mathcal{G}_n] = X_n,$$
and by induction $\mathbb{E}[X_m \,|\, \mathcal{G}_n] = X_n$ for all $m \geq n$. Taking expectations, $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ for all n.

Example 5.4 (Caution: the σ-algebras must be nested)**.** In Example 4.4, $\sigma(X_1)$ and $\sigma(S)$ are *not* nested. Then

$$\mathbb{E}\big[\mathbb{E}[S \,|\, X_1] \,|\, S\big] = \mathbb{E}[X_1 + 3.5 \,|\, S] = \tfrac{S}{2} + 3.5, \qquad \mathbb{E}\big[\mathbb{E}[S \,|\, S] \,|\, X_1\big] = \mathbb{E}[S \,|\, X_1] = X_1 + 3.5.$$

The two orders give different answers; neither is "$\mathbb{E}[S \,|$ common information$]$." The tower property genuinely requires $\mathcal{G} \subseteq \mathcal{H}$.

5.3 Filtrations

Definition 5.5 (Filtration; adapted process)**.** A *filtration* is an increasing sequence of σ-algebras

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}.$$

$\mathcal{F}_n$ models the information available at time n; information is never forgotten. A process (X_n) is *adapted* to $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$-measurable for each n (its value is known at time n).

The *natural filtration* of a process (X_n) is $\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)$, the information obtained by observing the process itself up to time n. In continuous time, a filtration is a family $(\mathcal{F}_t)_{t \geq 0}$ with $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s \leq t$, and $\mathcal{F}_t^X = \sigma(X_u : 0 \leq u \leq t)$.

6 Martingales with Respect to a Filtration

Definition 6.1 (Martingale)**.** Let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration. A process $(X_n)_{n\geq 0}$ is a *martingale with respect to* $(\mathcal{F}_n)$ if

(i) (X_n) is adapted to $(\mathcal{F}_n)$;

(ii) $\mathbb{E}|X_n| < \infty$ for all n;

(iii) $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = X_n$ for all n.

If "=" in (iii) is replaced by "$\geq$" (resp. "$\leq$"), (X_n) is a *submartingale* (resp. *supermartingale*). In continuous time, (iii) becomes $\mathbb{E}[X_t \,|\, \mathcal{F}_s] = X_s$ for all $0 \leq s \leq t$.

Definition 3.1 is the special case $\mathcal{F}_n = \sigma(X_0, \ldots, X_n)$ (or $\sigma(Y_0, \ldots, Y_n)$).

Proposition 6.2. *If (X_n) is a martingale with respect to some filtration $(\mathcal{F}_n)$, then it is also a martingale with respect to its natural filtration $(\mathcal{F}_n^X)$.*

Proof. Adaptedness gives $\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n) \subseteq \mathcal{F}_n$. By the tower property, $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n^X] = \mathbb{E}\big[\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] \,|\, \mathcal{F}_n^X\big] = \mathbb{E}[X_n \,|\, \mathcal{F}_n^X] = X_n$. □

Remark 6.3 (Gambler's ruin revisited (preview of optional stopping))*.* Under mild conditions—satisfied here because $\mathbb{E}\tau < \infty$ and the increments are bounded—the *optional stopping theorem* says that $\mathbb{E}[M_\tau] = \mathbb{E}[M_0]$ for a martingale M. One checks directly (using the independence of X_{n+1} from $X_1, \ldots, X_n$) that S_n and $S_n^2 - n$ are martingales in the fair case, and $(q/p)^{S_n}$ is a martingale in the biased case. Applying optional stopping to each:

- $M_n = S_n$: $k = \mathbb{E}[S_\tau] = bp_k + 0 \cdot (1 - p_k)$, so $p_k = k/b$.
- $M_n = S_n^2 - n$: $k^2 = \mathbb{E}[S_\tau^2] - \mathbb{E}[\tau] = b^2 p_k - D_k = bk - D_k$, so $D_k = k(b-k)$.
- $M_n = (q/p)^{S_n}$ (biased case): $(q/p)^k = (q/p)^b p_k + (1 - p_k)$, which recovers the biased formula.

Martingales thus solve the recurrences of Section 2 in one line.

7 Processes with Independent Increments

Definition 7.1 (Independent increments)**.** A process $(X_t)_{t\in\mathbb{T}}$ (with $\mathbb{T} = \{0, 1, 2, \ldots\}$ or $[0, \infty)$) has *independent increments* if for every n and all times $t_0 < t_1 < \cdots < t_n$ in $\mathbb{T}$, the increments

$$X_{t_1} - X_{t_0},\ X_{t_2} - X_{t_1},\ \ldots,\ X_{t_n} - X_{t_{n-1}}$$

are mutually independent.

In discrete time this is equivalent to $X_n = X_0 + \sum_{i=1}^n D_i$ with $X_0, D_1, D_2, \ldots$ independent. The random walk is the canonical example; the Poisson process and Brownian motion are the canonical continuous-time examples.

Theorem 7.2. *Let (X_t) have independent increments, with $X_0 = x_0$* deterministic*, $\mathbb{E}|X_t| < \infty$, and* zero-mean increments*: $\mathbb{E}[X_t - X_s] = 0$ for all $s \leq t$. Then (X_t) is a martingale with respect to its natural filtration.*

Proof. Discrete time. Let $D_i = X_i - X_{i-1}$. Since X_0 is constant, $X_1, \dots, X_n$ and $D_1, \dots, D_n$ determine each other ($X_m = x_0 + \sum_{i \le m} D_i$ and $D_i = X_i - X_{i-1}$), so

$$\mathcal{F}_n^X = \sigma(X_0, \dots, X_n) = \sigma(D_1, \dots, D_n).$$

By independent increments, D_{n+1} is independent of $(D_1, \dots, D_n)$, hence of $\mathcal{F}_n^X$. Therefore

$$\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n^X] = \mathbb{E}[X_n + D_{n+1} \,|\, \mathcal{F}_n^X] = X_n + \mathbb{E}[D_{n+1}] = X_n,$$

using (P3) for X_n and (P4) for D_{n+1}.

Continuous time. Fix $s < t$. For any times $0 \le u_1 < \cdots < u_m \le s$, the vector $(X_{u_1}, \dots, X_{u_m})$ is a function of the increments $X_{u_1} - X_0$, $X_{u_2} - X_{u_1}, \dots, X_{u_m} - X_{u_{m-1}}$ (since $X_0 = x_0$ is constant), and these are independent of $X_t - X_s$ by assumption. Thus $X_t - X_s$ is independent of every finite-dimensional event $\{X_{u_1} \le a_1, \dots, X_{u_m} \le a_m\}$. These events form a π-system generating $\mathcal{F}_s^X$, so $X_t - X_s$ is independent of $\mathcal{F}_s^X$. As before,

$$\mathbb{E}[X_t \,|\, \mathcal{F}_s^X] = X_s + \mathbb{E}[X_t - X_s] = X_s. \qquad \square$$

Remark 7.3.

- The deterministic starting point ensures that observing $X_0, \dots, X_n$ reveals *only* the past increments. (It suffices that X_0 be independent of all increments.)
- If the increments do not have zero mean, subtracting the mean gives a martingale: $X_t - \mathbb{E}[X_t]$ is a martingale. For example, a Poisson process N_t with rate λ has independent increments with $\mathbb{E}[N_t - N_s] = \lambda(t-s)$, so the *compensated* Poisson process $N_t - \lambda t$ is a martingale.

8 Markov Processes

Definition 8.1 (Markov process)**.** A process $(X_t)_{t \in \mathbb{T}}$ is a *Markov process* if for every n, all times $t_1 < t_2 < \cdots < t_n < t_{n+1}$, and all $x \in \mathbb{R}$,

$$\mathbb{P}\big(X_{t_{n+1}} \le x \,|\, X_{t_1}, \dots, X_{t_n}\big) = \mathbb{P}\big(X_{t_{n+1}} \le x \,|\, X_{t_n}\big).$$

In discrete time this is equivalent to $\mathbb{P}(X_{n+1} \in A \,|\, X_0, \dots, X_n) = \mathbb{P}(X_{n+1} \in A \,|\, X_n)$ for all n and sets A, or to $\mathbb{E}[f(X_{n+1}) \,|\, \mathcal{F}_n^X] = \mathbb{E}[f(X_{n+1}) \,|\, X_n]$ for bounded f.

In words: *given the present, the future is independent of the past.* The most recent observation summarizes everything in the history that is relevant for prediction.

Proposition 8.2. *A process with independent increments and a deterministic starting point is Markov.*

Proof. Let $t_1 < \cdots < t_{n+1}$ and $D = X_{t_{n+1}} - X_{t_n}$. As in the proof of Theorem 7.2, D is independent of $(X_{t_1}, \dots, X_{t_n})$. By the freezing property (P5) with $\phi(y, d) = \mathbf{1}\{y + d \le x\}$,

$$\mathbb{P}(X_{t_{n+1}} \le x \,|\, X_{t_1}, \dots, X_{t_n}) = \mathbb{E}\big[\mathbf{1}\{X_{t_n} + D \le x\} \,|\, X_{t_1}, \dots, X_{t_n}\big] = F_D(x - X_{t_n}),$$

where F_D is the CDF of D. This depends only on X_{t_n}; by the tower property it therefore also equals $\mathbb{P}(X_{t_{n+1}} \le x \,|\, X_{t_n})$. $\square$

Thus random walks (biased or not), the Poisson process, and Brownian motion are all Markov.

Remark 8.3 (Markov and martingale are different properties). Neither property implies the other.

- *Markov but not a martingale:* the biased random walk ($p \neq \frac{1}{2}$) is Markov, but $\mathbb{E}[S_{n+1} \mid \mathcal{F}_n] = S_n + (p - q) \neq S_n$.
- *Martingale but not Markov:* let $V \in \{1, 2\}$ be equally likely, independent of i.i.d. Rademacher ξ_i, and let $X_0 = 0$, $X_n = V(\xi_1 + \cdots + \xi_n)$. Since $V = |X_1|$ is known at time 1, $\mathbb{E}[X_{n+1} \mid \mathcal{F}_n^X] = X_n + V\mathbb{E}[\xi_{n+1}] = X_n$, so X is a martingale. However, $\mathbb{P}(X_3 = 1 \mid X_2 = 0) = \frac{1}{4}$ (step size 1 or 2 equally likely), while $\mathbb{P}(X_3 = 1 \mid X_1 = 1, X_2 = 0) = \frac{1}{2}$ (the past reveals the step size is 1). So X is not Markov.

9 Brownian Motion

9.1 Definition

Brownian motion is the continuous-time analogue of the symmetric random walk: if S_n is the Rademacher walk, then $\sigma S_{\lfloor nt \rfloor}/\sqrt{n}$ converges (in distribution, as a process) to Brownian motion as $n \to \infty$.

Definition 9.1 (Brownian motion / Wiener process)**.** A process $W = (W_t)_{t \geq 0}$ is a *Brownian motion* with parameter $\sigma^2 > 0$ if

(a) $W_0 = 0$;

(b) W has independent increments;

(c) for $0 \leq s < t$, the increment $W_t - W_s \sim \mathcal{N}\big(0, \sigma^2(t - s)\big)$;

(d) with probability one, the sample paths $t \mapsto W_t$ are continuous.

When $\sigma^2 = 1$, W is a *standard* Brownian motion.

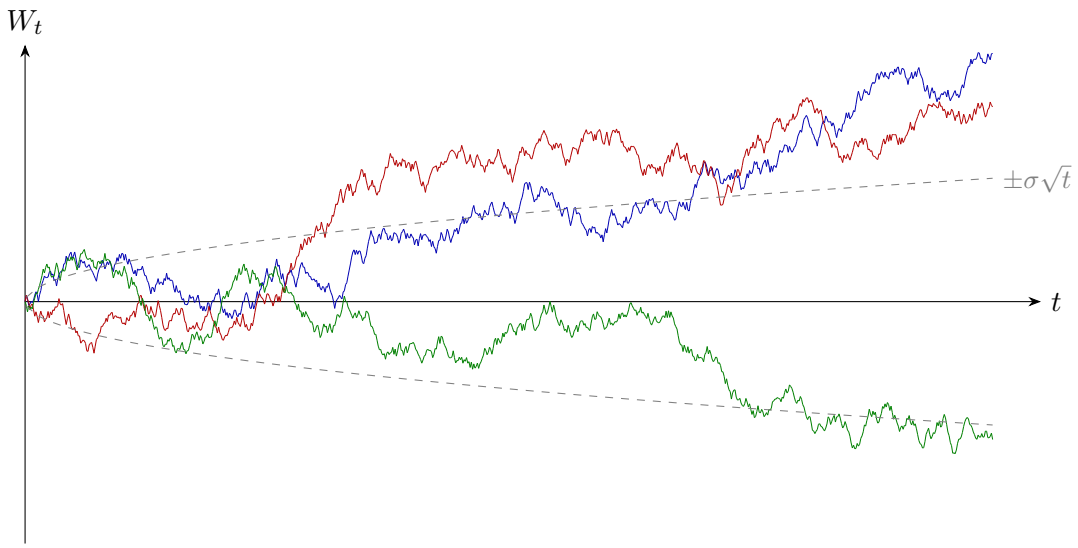


Figure 2: Three simulated sample paths of Brownian motion (fine-mesh random-walk approximations). The dashed curves show $\pm$ one standard deviation, $\sigma\sqrt{t}$.

9.2 Mean and correlation function

Proposition 9.2. *For Brownian motion with parameter σ^2,*

$$\mu_W(t) = \mathbb{E}[W_t] = 0, \qquad R_W(s,t) = \mathbb{E}[W_s W_t] = \sigma^2 \min(s,t).$$

Proof. By (a) and (c), $W_t = W_t - W_0 \sim \mathcal{N}(0, \sigma^2 t)$, so $\mathbb{E}[W_t] = 0$ and $\mathbb{E}[W_t^2] = \sigma^2 t$. For the correlation, take $0 \le s \le t$ and write $W_t = W_s + (W_t - W_s)$. The increment $W_t - W_s$ is independent of $W_s = W_s - W_0$ by (b), so

$$\mathbb{E}[W_s W_t] = \mathbb{E}[W_s^2] + \mathbb{E}[W_s(W_t - W_s)] = \sigma^2 s + \mathbb{E}[W_s]\,\mathbb{E}[W_t - W_s] = \sigma^2 s.$$

By symmetry, $R_W(s,t) = \sigma^2 \min(s,t)$ in general. □

Since $\mu_W \equiv 0$, the autocovariance equals the correlation function: $C_W(s,t) = \sigma^2 \min(s,t)$. In particular W is *not* wide-sense stationary ($\operatorname{Var} W_t = \sigma^2 t$ grows), but its increments are stationary. Moreover, W is a *Gaussian process*: every vector $(W_{t_1}, \dots, W_{t_n})$ is a linear transformation of independent Gaussian increments, hence jointly Gaussian. A zero-mean Gaussian process with continuous paths and $R(s,t) = \sigma^2 \min(s,t)$ is in fact a Brownian motion, so the mean and correlation function characterize W.

Corollary 9.3. *Brownian motion is a martingale and a Markov process with respect to its natural filtration. Also, $W_t^2 - \sigma^2 t$ is a martingale.*

Proof. The first two claims follow from Theorem 7.2 and the previous section, since $W_0 = 0$ and increments have zero mean. For the last, with $s < t$ and $\Delta = W_t - W_s$ independent of $\mathcal{F}_s$,

$$\mathbb{E}[W_t^2 \mid \mathcal{F}_s] = \mathbb{E}[W_s^2 + 2W_s\Delta + \Delta^2 \mid \mathcal{F}_s] = W_s^2 + 0 + \sigma^2(t-s),$$

so $\mathbb{E}[W_t^2 - \sigma^2 t \mid \mathcal{F}_s] = W_s^2 - \sigma^2 s$. (Compare with $S_n^2 - n$ for the random walk.) □

9.3 Brownian motion is nowhere differentiable

The sample paths are continuous, but extremely rough. The heuristic reason is the scaling

$$W_{t+h} - W_t \sim \mathcal{N}(0, \sigma^2 h) \quad \Longrightarrow \quad |W_{t+h} - W_t| \approx \sigma\sqrt{h} \gg h \text{ for small } h,$$

so difference quotients $\frac{W_{t+h}-W_t}{h} \approx \sigma/\sqrt{h}$ blow up as $h \to 0$. We make this precise in three increasingly strong forms.

(1) No mean-square derivative. For fixed t and $h > 0$,

$$\mathbb{E}\left[\left(\frac{W_{t+h} - W_t}{h}\right)^2\right] = \frac{\sigma^2 h}{h^2} = \frac{\sigma^2}{h} \xrightarrow[h\to 0]{} \infty.$$

A family converging in mean square must have bounded second moments, so the difference quotient has no mean-square limit: W is not m.s. differentiable at any t. Equivalently, the m.s. derivative would require $\partial^2 R_W(s,t)/\partial s\,\partial t$ to exist on the diagonal $s = t$, but $R_W(s,t) = \sigma^2 \min(s,t)$ has a kink there.

(2) No derivative at a fixed time, almost surely. Fix t and let $h_n \downarrow 0$. Let $Q_n = (W_{t+h_n} - W_t)/h_n \sim \mathcal{N}(0, \sigma^2/h_n)$, so $Q_n \overset{d}{=} (\sigma/\sqrt{h_n})Z$ with $Z \sim \mathcal{N}(0,1)$. For any $K > 0$,

$$\mathbb{P}(|Q_n| \le K) = \mathbb{P}\big(|Z| \le K\sqrt{h_n}/\sigma\big) \xrightarrow[n\to\infty]{} 0.$$

Since $\{\limsup_n |Q_n| < K\} \subseteq \bigcup_m \bigcap_{n\ge m}\{|Q_n| \le K\}$, and $\mathbb{P}\big(\bigcap_{n\ge m}\{|Q_n| \le K\}\big) \le \inf_{n\ge m} \mathbb{P}(|Q_n| \le K) = 0$ for each m, we get $\mathbb{P}(\limsup_n |Q_n| < K) = 0$. Taking the union over $K \in \mathbb{N}$, $\limsup_n |Q_n| = \infty$ almost surely, so W is a.s. not differentiable at t.

This does *not* yet show that W is nowhere differentiable: there are uncountably many t, and a union of uncountably many null events need not be null. The following theorem handles all t simultaneously.

Theorem 9.4 (Paley–Wiener–Zygmund)**.** *With probability one, the Brownian sample path $t \mapsto W_t$ is not differentiable at any $t \ge 0$.*

Remark 9.5 (White noise as the derivative of the Wiener process)*.* Although W has no derivative in the ordinary sense, engineers routinely write $N_t = \dot{W}_t = dW_t/dt$ and call it *white noise*. A formal computation shows why: differentiating the correlation function in each variable,

$$R_N(s,t) = \frac{\partial^2}{\partial s\,\partial t}\,\sigma^2 \min(s,t) = \frac{\partial}{\partial s}\,\sigma^2\,\mathbf{1}\{t < s\} = \sigma^2\,\delta(t-s).$$

So white noise is zero-mean, its values at distinct times are uncorrelated, and its power spectral density is the constant $S_N(\omega) = \sigma^2$ at all frequencies (hence "white," by analogy with white light). It also has "infinite variance" $R_N(t,t) = \sigma^2\delta(0)$, consistent with the difference-quotient variance $\sigma^2/h \to \infty$ computed above. These statements make sense only if δ and $\dot{W}$ are interpreted as *generalized functions*.

Generalized functions, intuitively. An ordinary function is known by its values $f(t)$ at points. A generalized function (or *distribution*) is instead known only through its "smeared" or averaged values: how it acts on smooth, rapidly vanishing *test functions* φ,

$$\langle T, \varphi\rangle \text{ “}=\text{ ”} \int T(t)\,\varphi(t)\,dt.$$

Think of φ as a measuring instrument with finite resolution: we never observe T at a single instant, only through a small time window. Some examples:

- Every ordinary (locally integrable) function f defines a generalized function by $\langle f, \varphi\rangle = \int f\varphi\,dt$.
- The Dirac delta is *defined* by $\langle \delta, \varphi\rangle = \varphi(0)$. Intuitively it is the limit of ever taller and narrower unit-area pulses, such as $\frac{1}{h}\mathbf{1}_{[0,h]}$; no ordinary function has this property, but the action on test functions is perfectly well defined.
- *Every* generalized function has a derivative, defined by moving the derivative onto the test function (integration by parts, with no boundary terms because φ vanishes at infinity):

$$\langle T', \varphi\rangle := -\langle T, \varphi'\rangle.$$

For example, the derivative of the unit step $u(t) = \mathbf{1}\{t \ge 0\}$ is δ, since $-\int_0^\infty \varphi'(t)\,dt = \varphi(0)$. Applying the last rule to a Brownian path gives white noise rigorously:

$$\langle \dot{W}, \varphi\rangle := -\int_0^\infty W_t\,\varphi'(t)\,dt,$$

which is well defined because W is continuous. It is a zero-mean Gaussian random variable with variance $\sigma^2 \int_0^\infty \varphi(t)^2\,dt$. Thus white noise cannot be sampled at a point, but its averages over any smooth window are finite, well-behaved random variables. This is also exactly how white noise enters linear systems: the output of a filter with impulse response h is $\int h(t-s)\,dW_s$, which is a smeared version of $\dot{W}$.