

Class 1

Probability Spaces and Random Variables

A discrete **probability space** (Ω, P) is a finite or countably infinite set of outcomes Ω together with probabilities $P(\omega) \geq 0$ that add up to one. An **event** is a subset $A \subset \Omega$, and its probability is

$$P(A) = \sum_{\omega \in A} P(\omega).$$

Unions, intersections and complements of events translate “or”, “and” and “not”. If A and B are disjoint, $P(A \cup B) = P(A) + P(B)$, and in particular $P(A^c) = 1 - P(A)$.

A **random variable** X assigns a real number $X(\omega)$ to each outcome. Its **expectation** and **variance** are

$$E(X) = \sum_{\omega \in \Omega} X(\omega)P(\omega), \quad V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2,$$

and the **covariance** and **correlation** of X and Y are

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y), \quad \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \in [-1, 1].$$

Linear Combinations

For $Z = \alpha X + \beta Y$, expectations are linear while variances pick up the covariance:

$$E(Z) = \alpha E(X) + \beta E(Y), \quad V(Z) = \alpha^2 V(X) + \beta^2 V(Y) + 2\alpha\beta \text{Cov}(X, Y). \quad (1)$$

If $X_1, \dots, X_n$ are uncorrelated with common variance σ^2 , the sum has variance $n\sigma^2$ and the average has variance σ^2/n . A sum of n shocks therefore drifts a distance of order $\sqrt{n}$, the scaling that will turn a random walk into Brownian motion.

Joint and Marginal Distributions

The **probability mass function** (pmf) of X is $p_X(x) = P(X = x)$, so that $E(X) = \sum_x x p_X(x)$. The **joint pmf** $p_{X,Y}(x, y) = P(X = x, Y = y)$ contains everything about X and Y . Summing over one variable recovers the other's pmf,

$$p_X(x) = \sum_y p_{X,Y}(x, y), \quad p_Y(y) = \sum_x p_{X,Y}(x, y),$$

and expectations of functions of both variables are $E(f(X, Y)) = \sum_x \sum_y f(x, y) p_{X,Y}(x, y)$.

Conditioning and Independence

The **conditional probability** of A given B , for $P(B) > 0$, is

$$P(A | B) = \frac{P(A \cap B)}{P(B)}.$$

Learning B discards the outcomes outside B and rescales the rest. If $B_1, B_2, \dots$ partition Ω , the **law of total probability** reads

$$P(A) = \sum_k P(A | B_k) P(B_k). \quad (2)$$

Events A and B are **independent** if $P(A \cap B) = P(A)P(B)$, and random variables are independent if their joint pmf factors, $p_{X,Y}(x, y) = p_X(x) p_Y(y)$. Then $E(f(X)g(Y)) = E(f(X))E(g(Y))$ for any functions f and g , so independent variables are uncorrelated. The converse fails: X uniform on $\{-1, 0, 1\}$ and $Y = X^2$ are uncorrelated, yet Y is a function of X .

Information and Conditional Expectation

A random variable splits Ω into the cells on which it takes each value. Observing it reveals which cell contains the true outcome and nothing more. The events that are unions of cells form the **information set** it generates. A random variable is **known** given an information set $\mathcal{F}$ when it is constant on every cell.

If $\mathcal{F}$ is generated by the partition $C_1, C_2, \dots$, the **conditional expectation** of Y given $\mathcal{F}$ averages Y over the cell that occurred,

$$E(Y | \mathcal{F}) = \frac{E(Y \mathbf{1}_{C_k})}{P(C_k)} \quad \text{on } C_k. \quad (3)$$

It is a random variable known given $\mathcal{F}$: our best forecast of Y with the information in $\mathcal{F}$. Four properties do almost all the work in this course.

1. **Linearity.** $E(aY + bW | \mathcal{F}) = a E(Y | \mathcal{F}) + b E(W | \mathcal{F})$.
2. **Taking out what is known.** If W is known given $\mathcal{F}$, then $E(WY | \mathcal{F}) = W E(Y | \mathcal{F})$.
3. **Independence.** If Y is independent of $\mathcal{F}$, then $E(Y | \mathcal{F}) = E(Y)$.
4. **Tower property.** If $\mathcal{G} \subset \mathcal{F}$, then $E(E(Y | \mathcal{F}) | \mathcal{G}) = E(Y | \mathcal{G})$, and in particular $E(E(Y | \mathcal{F})) = E(Y)$.

Filtrations

A **filtration** $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots$ describes information that accumulates and is never forgotten. With independent coin flips $X_n = \pm 1$ and $\mathcal{F}_n$ generated by the first n flips, each flip splits every cell in two, as in Figure 1. The random walk $S_n = X_1 + \dots + X_n$ is known at time n and the next flip is independent of $\mathcal{F}_n$, so

$$E(S_{n+1} | \mathcal{F}_n) = S_n + E(X_{n+1}) = S_n. \quad (4)$$

A process whose best forecast is its current value is a **martingale**.

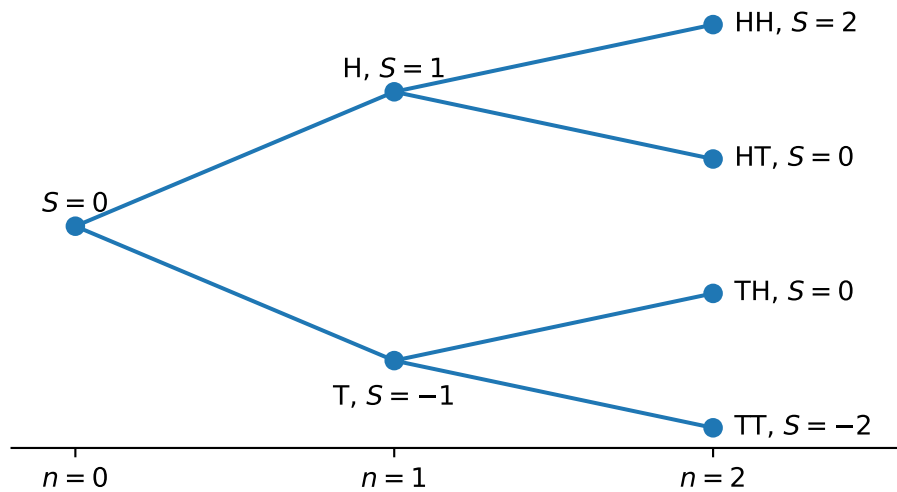


Figure 1: Two flips of a fair coin. Each node is a cell of the partition generating $\mathcal{F}_n$ and shows the value of the random walk S_n .

Practice Problems

1. A portfolio holds two corporate bonds that default independently over the next year with probabilities 0.10 and 0.20. Let N be the number of defaults.
 - a. Find the probability mass function of N .
 - b. Compute $E(N)$ and $V(N)$.
 - c. Given that at least one bond defaults, what is the probability that both default?
2. The economy expands next year with probability 0.7 and enters a recession with probability 0.3. A firm defaults with probability 2% in an expansion and 10% in a recession.
 - a. Use the law of total probability to compute the probability that the firm defaults.
 - b. Given that the firm defaults, what is the probability that the economy was in a recession?
3. Two random variables satisfy $V(X) = 4$, $V(Y) = 9$ and $V(X + Y) = 7$.
 - a. Compute $\text{Cov}(X, Y)$ and $\rho_{X,Y}$.
 - b. Compute $V(X - Y)$ and $V(2X - Y)$.
4. The table below describes two random variables on a sample space with four outcomes.

Outcome	Probability	X	Y
ω_1	0.20	4	1
ω_2	0.30	2	1
ω_3	0.10	6	0
ω_4	0.40	0	0

- a. Write down the partition generated by Y and list the events in $\mathcal{F}_Y$.
- b. Is Y^2 known given $\mathcal{F}_Y$? Is XY ? Explain.
- c. Compute $E(X | \mathcal{F}_Y)$ on each cell and verify that $E(E(X | \mathcal{F}_Y)) = E(X)$.
- d. Use the properties of conditional expectation to compute $E(XY | \mathcal{F}_Y)$.
- e. Are X and Y independent? Explain.