

Lecture 3: Discrete Random Variables and Expectation

EECS 126 (UC Berkeley)

Spring 2026 — January 27, 2026

We transition from events to random variables — functions that assign numerical values to outcomes — and develop the key tools of expectation, linearity, LOTUS, and variance.

1 Random Variables and PMFs

Definition 3.1 (Random variable)

A **(discrete) random variable** is a function $X : \Omega \rightarrow \mathbb{R}$ that assigns a real number to each outcome $\omega \in \Omega$. We write $\mathbb{P}(X = x)$ for the probability of the event $\{\omega \in \Omega : X(\omega) = x\}$.

Definition 3.2 (Probability mass function)

The **probability mass function (PMF)** of a discrete random variable X is the function $p_X : \mathbb{R} \rightarrow [0, 1]$ defined by

$$p_X(x) = \mathbb{P}(X = x). \quad (3.1)$$

The PMF satisfies $\sum_{x \in \mathcal{X}} p_X(x) = 1$, where $\mathcal{X}$ denotes the set of all values taken by X (the *support*).

Example 3.1 (Functions of random variables)

Let X be a random variable with $p_X(x) = 1/5$ for $x \in \{-2, -1, 0, 1, 2\}$. Define $Y = |X|$. Then Y takes values in $\{0, 1, 2\}$ and

$$p_Y(y) = \sum_{\{x : |x|=y\}} p_X(x) = \begin{cases} 1/5 & \text{if } y = 0, \\ 2/5 & \text{if } y = 1, \\ 2/5 & \text{if } y = 2. \end{cases} \quad (3.2)$$

In general, if $Y = g(X)$, then $p_Y(y) = \sum_{\{x : g(x)=y\}} p_X(x)$.

2 Popular Discrete Distributions

Definition 3.3 (Bernoulli distribution)

$X \sim \text{Bernoulli}(p)$: a single coin toss.

$$\mathbb{P}(X = k) = \begin{cases} 1 - p & k = 0, \\ p & k = 1. \end{cases} \quad (3.3)$$

Definition 3.4 (Binomial distribution)

$X \sim \text{Binomial}(n, p)$: number of successes in n independent Bernoulli(p) trials.

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n. \quad (3.4)$$

Note: $\sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1$ by the binomial theorem, confirming this is a valid PMF.

Definition 3.5 (Geometric distribution)

$X \sim \text{Geometric}(p)$: number of trials until the first success.

$$\mathbb{P}(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, 3, \dots \quad (3.5)$$

The geometric distribution is the discrete analog of “waiting time.”

Definition 3.6 (Poisson distribution)

$X \sim \text{Poisson}(\lambda)$: models the number of arrivals in a fixed interval.

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \dots \quad (3.6)$$

The Poisson is the limiting distribution of $\text{Binomial}(n, \lambda/n)$ as $n \rightarrow \infty$, and arises naturally in modeling rare events and arrival processes.

Example 3.2 (Chess match – a PMF computation)

Alice and Bob play a match: the first to win a game wins the match. If there are 10 consecutive draws, the match is declared drawn. Each game independently: $\mathbb{P}(\text{A wins}) = 0.3$, $\mathbb{P}(\text{B wins}) = 0.4$, $\mathbb{P}(\text{draw}) = 0.3$.

Let L be the duration (number of games). Then for $\ell = 1, 2, \dots, 9$:

$$p_L(\ell) = (0.3)^{\ell-1} (1 - 0.3) = (0.3)^{\ell-1} (0.7), \quad (3.7)$$

since the first $\ell - 1$ games must be draws and the ℓ -th decisive. For $\ell = 10$: $p_L(10) = (0.3)^9$ (all 9 games drawn, match ends).

$\mathbb{P}(\text{Alice wins match}) = \sum_{\ell=1}^9 (0.3)^{\ell-1} (0.3) = 0.3 \sum_{k=0}^8 (0.3)^k$. Using $\sum_{k=0}^{\infty} r^k = 1/(1-r)$ for $|r| < 1$, in the limit of no cap on draws: $\mathbb{P}(\text{Alice wins}) = 0.3/(1-0.3) = 3/7$.

3 Expectation

Definition 3.7 (Expected value)

The **expected value** (or **mean**) of a discrete random variable X is

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) = \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}(X = x), \quad (3.8)$$

provided the sum converges absolutely (i.e. $\sum |x| \mathbb{P}(X = x) < \infty$).

Theorem 3.1 (Linearity of expectation)

For *any* random variables X, Y (not necessarily independent) and constant c :

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y], \quad \mathbb{E}[cX] = c\mathbb{E}[X]. \quad (3.9)$$

No independence assumption is needed. This is the single most useful property of expectation.

Definition 3.8 (Indicator random variable)

For an event A , the **indicator** $\mathbf{1}_A$ is

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \notin A. \end{cases} \quad (3.10)$$

Observe that $\mathbf{1}_A \sim \text{Bernoulli}(\mathbb{P}(A))$, so $\mathbb{E}[\mathbf{1}_A] = \mathbb{P}(A)$.

Example 3.3 (Seats on a plane)

A full flight has n passengers with assigned seats, but every passenger ignores their assignment and sits in a uniformly random seat. What is $\mathbb{E}[\text{number who sit in their own seat}]$?

Let $X_i = \mathbf{1}\{\text{passenger } i \text{ sits in their own seat}\}$. Then $X = X_1 + X_2 + \cdots + X_n$. The X_i are *not* independent (if $n = 2$ and passenger 1 is in their seat, passenger 2 must be too). But linearity of expectation does not require independence:

$$\mathbb{E}[X] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n \mathbb{P}(\text{passenger } i \text{ is in their seat}) = n \cdot \frac{1}{n} = 1. \quad (3.11)$$

Regardless of the number of passengers, on average exactly 1 person sits in their own seat.

4 LOTUS (Law of the Unconscious Statistician)

Theorem 3.2 (LOTUS)

If X is a discrete random variable and $g : \mathbb{R} \rightarrow \mathbb{R}$ is any function, then

$$\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) \mathbb{P}(X = x). \quad (3.12)$$

More generally, for a function of several random variables:

$$\mathbb{E}[f(X_1, \dots, X_n)] = \sum_{x_1, \dots, x_n} f(x_1, \dots, x_n) p_{X_1, \dots, X_n}(x_1, \dots, x_n). \quad (3.13)$$

The name reflects that one can compute $\mathbb{E}[g(X)]$ directly from the distribution of X , without first finding the distribution of $Y = g(X)$.

5 Independence of Random Variables

Definition 3.9 (Independent random variables)

Random variables X and Y are **independent** (written $X \perp Y$) if their joint distribution factors into the product of the marginals:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \mathbb{P}(Y = y) \quad \text{for all } x, y. \quad (3.14)$$

Theorem 3.3 (Expectation of a product of independent RVs)

If $X \perp Y$, then

$$\mathbb{E}[XY] = \mathbb{E}[X] \mathbb{E}[Y]. \quad (3.15)$$

Proof. By LOTUS and independence:

$$\begin{aligned} \mathbb{E}[XY] &= \sum_{x,y} xy p_{X,Y}(x,y) = \sum_{x,y} xy p_X(x) p_Y(y) \\ &= \left(\sum_x x p_X(x) \right) \left(\sum_y y p_Y(y) \right) = \mathbb{E}[X] \mathbb{E}[Y]. \end{aligned} \quad (3.16)$$

The key step is factoring the joint PMF into marginals (by independence), which allows separating the double sum into a product of single sums. $\square$

The converse is false. $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ does *not* imply $X \perp Y$. The condition $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ is called *uncorrelation*, which is strictly weaker than independence. As an exercise, construct dependent random variables that are uncorrelated.

6 Tail Sum Formula

Theorem 3.4 (Tail sum formula)

Let X be a random variable taking values in $\mathbb{N} = \{0, 1, 2, \dots\}$. Then

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}(X \geq k). \quad (3.17)$$

Proof. Starting from the definition of expectation:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=1}^{\infty} k \mathbb{P}(X = k) = \sum_{k=1}^{\infty} \sum_{j=1}^k 1 \cdot \mathbb{P}(X = k) \\ &= \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} \mathbb{P}(X = k) = \sum_{j=1}^{\infty} \mathbb{P}(X \geq j). \end{aligned} \quad (3.18)$$

The key step is writing $k = \sum_{j=1}^k 1$ to introduce a double sum, then swapping the order of summation (Tonelli's theorem ensures this is valid for nonnegative terms). Geometrically, this corresponds to “counting vertically then horizontally”: instead of weighting each column k by its height k , we sum across rows $j = 1, 2, \dots$, each contributing $\mathbb{P}(X \geq j)$. $\square$

Intuition for the tail sum. Think of $\mathbb{E}[X]$ as the total area under the “staircase” of the survival function $\mathbb{P}(X \geq k)$. Each unit step at height k contributes $\mathbb{P}(X \geq k)$ to the total. This is the same idea as computing the area under a curve by integrating horizontally rather than vertically.

7 Variance

Definition 3.10 (Variance and standard deviation)

The **variance** of X measures the spread of X around its mean:

$$\text{Var}(X) \triangleq \sigma_X^2 \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2]. \quad (3.19)$$

The **standard deviation** is $\sigma_X = \sqrt{\text{Var}(X)}$.

An equivalent computational formula (often easier to use):

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \quad (3.20)$$

Derivation of the computational formula. Expand the square using linearity:

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + (\mathbb{E}[X])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \end{aligned} \quad (3.21)$$

Note that $\mathbb{E}[X]$ is a *constant*, so $\mathbb{E}[2X \cdot \mathbb{E}[X]] = 2\mathbb{E}[X] \cdot \mathbb{E}[X]$ and $\mathbb{E}[(\mathbb{E}[X])^2] = (\mathbb{E}[X])^2$. $\square$

Theorem 3.5 (Scaling property of variance)

For constants $a, c \in \mathbb{R}$:

$$\text{Var}(aX + c) = a^2 \text{Var}(X). \quad (3.22)$$

Adding a constant shifts the mean but does not change the spread. Scaling by a scales the variance by a^2 (and the standard deviation by $|a|$).

Proof. Let $\mu = \mathbb{E}[X]$. Then $\mathbb{E}[aX + c] = a\mu + c$, so

$$\text{Var}(aX + c) = \mathbb{E}[(aX + c - a\mu - c)^2] = \mathbb{E}[(a(X - \mu))^2] = a^2 \mathbb{E}[(X - \mu)^2] = a^2 \text{Var}(X). \quad (3.23)$$

□

Coming up next lecture: covariance $\text{Cov}(X, Y)$, variance of sums, and the beginning of continuous random variables.

End of Lecture 3.