STA 131A: Introduction to Probability Theory

Lecture 7: Expectation, Mean, and Variance

Dogyoon Song

Spring 2026, UC Davis

Announcements

Homework 2 is due tomorrow (Tue, Apr 14, 11:59 PM)

- Please submit on time and follow the submission instructions
  - To Gradescope via Canvas in a single PDF
  - Email submissions are not accepted

If you have questions or need help

- TA office hours: Tue, 3:00 PM – 5:00 PM (MSB 1117)
- Feel free to ask questions during lecture, in office hours or post them on Piazza

Recap: Discrete random variable and PMF

A **random variable** is a function $X : \Omega \to \mathbb{R}$

- A random variable assigns a real number to each outcome $\omega \in \Omega$
- The randomness comes from the outcome ω, not from the function X itself
- X is **discrete** if its range is finite or countably infinite

The **probability mass function (PMF)** of X is a function $p_X : \mathbb{R} \to \mathbb{R}$ such that

$$p_X(x) = P(X = x), \qquad x \in \mathbb{R}$$

- $p_X(x) \geq 0$ for all x
- $\sum_x p_X(x) = 1$
- For any set $S \subseteq \mathbb{R}$,

$$P(X \in S) = \sum_{x \in S} p_X(x)$$

Example: Bernoulli, binomial, geometric, Poisson

Agenda

Last time:

- Discrete random variables
- Probability mass function
- Common discrete models: Bernoulli, binomial, geometric, Poisson

Today:

- Functions of random variables
- Expectation and variance
- Examples and applications

Function of a random variable

Given a random variable X, consider its image under a function $g : \mathbb{R} \to \mathbb{R}$:

$$Y = g(X)$$

- Y is a random variable because $Y = g \circ X$ is a map from Ω to $\mathbb{R}$
- Example: $aX + b$, X^2, $|X|$, $\log X$, ...

If X is discrete with PMF p_X, then Y is also discrete with PMF p_Y such that

$$p_Y(y) = \sum_{x \in \{x | g(x) = y\}} p_X(x)$$

- Different values of X may map to the same value of $Y = g(X)$

Example

Let X be a discrete RV with $p_X(x) = 1/5,\ \forall x \in \{-2, -1, 0, 1, 2\}$ and let $Y = X^2$

- Y takes only the values $0, 1, 4 \to Y$ is discrete
- The PMF of Y satisfies $p_Y(0) = 1/5$, $p_Y(1) = 2/5$, $p_Y(4) = 2/5$

Pop-up quiz

Let X satisfy

$$P(X=-1)=\frac{1}{4}, \qquad P(X=0)=\frac{1}{2}, \qquad P(X=1)=\frac{1}{4}.$$

Question: For $Y = X^2$, which statement is correct?

A) $P(Y=0)=\frac{1}{4}$
B) $P(Y=1)=\frac{1}{2}$
C) Y can take the values $-1, 0, 1$
D) Y is not a random variable

Answer: B.
Because both $X=-1$ and $X=1$ map to $Y=1$, so

$$P(Y=1)=P(X=-1)+P(X=1)=\frac{1}{2}.$$

Follow-up: If g is one-to-one on the values of X, do we still need to sum over several x's?

Expectation: Definition

The PMF fully describes the distribution, but often we want a single numerical summary

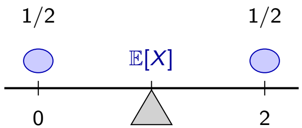
Definition (Expectation)

The **expectation** (also called the **expected value** or the **mean**) of a random variable X with PMF p_X is

$$\mathbb{E}[X] = \sum_x x \, p_X(x)$$

- $\mathbb{E}[X]$ is a probability-weighted average of the values of X
- It can be interpreted as the *center of mass* of the PMF
- $\mathbb{E}[X]$ need not be a value that X can actually take

Center-of-mass picture



Here $P(X = 0) = P(X = 2) = 1/2$, so $\mathbb{E}[X] = 1$, even though $P(X = 1) = 0$.

Expectation: Example

Example (Two biased coin tosses)

Let X be the number of Heads in two independent tosses, each with

$$P(\text{Head}) = \frac{3}{4}.$$

Then

$$P(X = 0) = \frac{1}{16}, \qquad P(X = 1) = \frac{6}{16}, \qquad P(X = 2) = \frac{9}{16}$$

Therefore,

$$\mathbb{E}[X] = 0 \cdot \frac{1}{16} + 1 \cdot \frac{6}{16} + 2 \cdot \frac{9}{16} = \frac{24}{16} = \frac{3}{2}$$

The expected number of Heads is 1.5, even though X can only take the values $0, 1, 2$

Variance: Definition

Definition (Variance)

The **variance** of a random variable X is

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

- Variance is the average squared distance from the mean
- It is always nonnegative because $(X - \mathbb{E}[X])^2 \geq 0$
- The **standard deviation** of X is defined as $\sigma_X = \sqrt{\mathrm{Var}(X)}$

In principle, we could compute variance by first finding the PMF of $Z = (X - \mathbb{E}[X])^2$

Example

Let X be a discrete RV with $p_X(x) = 1/5,\ \forall x \in \{-2, -1, 0, 1, 2\}$

- $Z = (X - \mathbb{E}[X])^2$ is discrete RV with $p_Z(0) = 1/5$, $p_Z(1) = 2/5$, $p_Z(4) = 2/5$
- Thus, $\mathrm{Var}(X) = \mathbb{E}[Z] = 0 \cdot \frac{1}{5} + 1 \cdot \frac{2}{5} + 4 \cdot \frac{2}{5} = 2$

Expectation of function of random variable

Question: Do we really need the PMF of $(X - \mathbb{E}[X])^2$ in order to compute variance?

Lemma

Let X be a random variable with PMF p_X and $g : \mathbb{R} \to \mathbb{R}$ be a function. Then

$$\mathbb{E}\left[g(X)\right] = \sum_x g(x)\, p_X(x)$$

To verify this, let $Y = g(X)$ and observe

$$\begin{aligned}
\mathbb{E}\left[g(X)\right] = \textstyle\sum_y y\, p_Y(y) &= \sum_y y \sum_{x \in \{x | g(x) = y\}} p_X(x) && \because \text{definition of } p_Y \\
&= \sum_y \sum_{x \in \{x \in \mathbb{R} | g(x) = y\}} g(x)\, p_X(x) && \because y = g(x) \\
&= \sum_x g(x)\, p_X(x)
\end{aligned}$$

Therefore, $\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \sum_x (x - \mathbb{E}[X])^2\, p_X(x)$

Expectation of function of random variable: Example

Example

Let X take the values -1 and 1 with probability $1/2$ each, and let $g(x) = x^2$. Then

$$\mathbb{E}[g(X)] = \mathbb{E}[X^2] = 1, \qquad g(\mathbb{E}[X]) = g(0) = 0.$$

Example

Let X take the values in $\{1, 2, 3, 4, 5, 6\}$ uniformly at random, and let $g(x) = 1/x$

Then

$$\mathbb{E}[g(X)] = \mathbb{E}[1/X] = 1 \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{3} \cdot \frac{1}{6} + \frac{1}{4} \cdot \frac{1}{6} + \frac{1}{5} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{147}{360}$$

$$g(\mathbb{E}[X]) = 1/(7/2) = \frac{2}{7}$$

In general, $\mathbb{E}[g(X)] \neq g(\mathbb{E}[X])$

Moments

Definition (Moment)

For a positive integer n, the **n-th moment** of a random variable X is $\mathbb{E}[X^n]$

For a discrete random variable,

$$\mathbb{E}[X^n] = \sum_x x^n \, p_X(x)$$

- The first moment is the mean
- The second *centered* moment is the variance
- Higher moments help describe the shape of a distribution

In this course, our main focus will be the mean and the variance

Properties of mean and variance

Lemma

Let X be a random variable and $a, b \in \mathbb{R}$. Then

$$\mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b, \qquad \mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$$

$$\mathbb{E}[aX + b] = \sum_x (ax + b)\, p_X(x) = a \underbrace{\sum_x x\, p_X(x)}_{=\mathbb{E}[X]} + b \underbrace{\sum_x p_X(x)}_{=1} = a\,\mathbb{E}[X] + b$$

$$\begin{aligned}\mathrm{Var}(aX + b) &= \sum_x (ax + b - \mathbb{E}[aX + b])^2\, p_X(x) = \sum_x (ax + b - a\mathbb{E}[X] - b)^2\, p_X(x) \\ &= a^2 \underbrace{\sum_x (x - \mathbb{E}[X])^2\, p_X(x)}_{=\mathrm{Var}(X)} = a^2\,\mathrm{Var}(X)\end{aligned}$$

Variance revisited

Lemma

For a discrete random variable X,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_x (x - \mathbb{E}[X])^2\, p_X(x) \\
&= \sum_x \left(x^2 - 2x\,\mathbb{E}[X] + (\mathbb{E}[X])^2\right) p_X(x) \\
&= \sum_x x^2\, p_X(x) - 2\mathbb{E}[X] \sum_x x\, p_X(x) + (\mathbb{E}[X])^2 \sum_x p_X(x) \\
&= \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2 \\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2
\end{aligned}$$

Pop-up quiz

Let X take the values 0 and 2 with probability $1/2$ each.

Question: Which statement is correct?

A) $\mathbb{E}[X] = 1$ and $\mathrm{Var}(X) = 0$

B) $\mathbb{E}[X] = 1$ and $\mathrm{Var}(X) = 1$

C) $\mathbb{E}[X] = 2$ and $\mathrm{Var}(X) = 1$

D) $\mathbb{E}[X] = 1$ and $\mathrm{Var}(X) = 2$

Answer: B.
The mean is $0 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = 1$, and the average squared distance from 1 is

$$\frac{(0-1)^2 + (2-1)^2}{2} = 1.$$

Follow-up: Can two random variables have the same mean but different variances?

Example: Bernoulli random variable

Example

If $X \sim \mathrm{Bernoulli}(p)$, then

$$P(X = 1) = p, \qquad P(X = 0) = 1 - p.$$

Therefore,

$$\mathbb{E}[X] = 0 \cdot (1 - p) + 1 \cdot p = p,$$

and since $X^2 = X$,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p).$$

- If $X = \mathbf{1}_A$ is the indicator of an event A, then

$$\mathbb{E}[X] = P(A), \qquad \mathrm{Var}(X) = P(A)(1 - P(A)).$$

Example: Discrete uniform random variable

Example

Let X be the outcome of a fair die roll, so

$$P(X = x) = \frac{1}{6}, \qquad x = 1, 2, 3, 4, 5, 6$$

Then

$$\mathbb{E}[X] = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{7}{2},$$

and

$$\mathbb{E}[X^2] = \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2}{6} = \frac{91}{6}$$

Therefore,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$$

Example: Poisson random variable

Example (Optional; challenging)

If $X \sim \mathrm{Poisson}(\lambda)$, then

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, \ldots$$

Then

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k\, e^{-\lambda}\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} e^{-\lambda}\frac{\lambda^k}{(k-1)!} = \lambda \sum_{j=0}^{\infty} \underbrace{e^{-\lambda}\frac{\lambda^j}{j!}}_{=p_X(j)} = \lambda$$

Also,

$$\mathbb{E}[X(X-1)] = \sum_{k=2}^{\infty} k(k-1)e^{-\lambda}\frac{\lambda^k}{k!} = \lambda^2,$$

so

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda$$

Therefore,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda$$

Decision making with expected values

Expected value is often used to compare decisions with random payoffs

- To maximize long-run average payoff, we typically choose the larger *expected payoff*

Example (Choosing between two lotteries)

Suppose you must choose between

$$A: \text{ receive \$50 for sure}, \qquad B: \text{ receive \$100 with probability 0.6 and \$0 otherwise.}$$

Then

$$\mathbb{E}[\text{payoff under } A] = 50, \qquad \mathbb{E}[\text{payoff under } B] = 60.$$

If the goal is to maximize average payoff, you should choose B.

- Expected value is a useful criterion for repeated decisions
- It is not the only consideration though: two options may have the same mean but different risk (measured in variance)

Wrap-up

Functions of random variables

- If $Y = g(X)$, then Y is again a random variable
- For a discrete X, compute $p_Y(y)$ by adding the probabilities of all x that satisfy $g(x) = y$

Expectation and variance

- $\mathbb{E}[X] = \sum_x x\, p_X(x)$ is a probability-weighted average (center of mass)
- $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$ measures spread around the mean
- For any function g,

$$\mathbb{E}[g(X)] = \sum_x g(x)\, p_X(x)$$

- A useful identity:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Suggested reading: [BT08, Ch. 2.3–2.4]

References

Dimitri Bertsekas and John N Tsitsiklis.
Introduction to probability, volume 1.
Athena Scientific, 2nd edition, 2008.