STA 131A: Introduction to Probability Theory

Lecture 10: Conditioning

Dogyoon Song

Spring 2026, UC Davis

Announcements

Homework 3 is due tomorrow (Tue, Apr 21, 11:59 PM)

- Please submit on time and follow the submission instructions
- You can collaborate on Homework, but all submitted work must be your own (also, don't forget to list all collaborators)

Midterm 1 is in class on Fri, Apr 24

- You may bring *one* ***hand-written*** *letter-sized (8.5 $\times$ 11 inches), double-sided sheet of paper* with formulas and brief notes
- **Calculator**: Simple (non-graphing) calculators only
- **No textbooks** or other materials beyond the single cheat sheet
- **SDC accommodations:** Confirm scheduling with AES online
- Practice midterms are available on the course webpage

Agenda

Last time:

- Conditional PMFs
- Conditional expectation
- Total expectation theorem (a.k.a. law of total expectation)

Today: Independence

- Independence of a random variable from an event
- Independence of random variables
- Factorization of expectation
- Variance of sums of independent random variables

Recap: Conditional PMFs and total expectation

If X and Y are discrete random variables, then for any y with $p_Y(y) > 0$,

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}.$$

- Equivalently,

$$p_{X,Y}(x, y) = p_Y(y)\, p_{X|Y}(x \mid y).$$

The law of total expectation states that

$$\mathbb{E}[X] = \sum_y \mathbb{E}[X \mid Y = y]\, p_Y(y).$$

- Conditional PMFs describe the distribution of X after learning $Y = y$.
- The unconditional mean is a weighted average of the conditional means.

Independence of a random variable from an event

Definition

Let A be an event with $P(A) > 0$. We say that X is **independent of** A if

$$p_{X|A}(x) = p_X(x), \qquad \forall x.$$

Equivalently, for every x, the event $\{X = x\}$ is independent of A.

Example (Two fair die rolls)

Let X be the first roll of two fair dice, and let

$$A = \{\text{the sum of the two rolls is even}\}.$$

For each $x \in \{1, \ldots, 6\}$, exactly three values of the second roll make the sum even, so

$$P(X = x \mid A) = \frac{3/36}{18/36} = \frac{1}{6} = p_X(x).$$

Independence of random variables

Definition

Discrete random variables X and Y are **independent** if

$$p_{X,Y}(x, y) = p_X(x)p_Y(y), \qquad \forall (x, y).$$

If $p_Y(y) > 0$, this is equivalent to

$$p_{X|Y}(x \mid y) = p_X(x), \qquad \forall x.$$

- Knowing the value of Y does not change the distribution of X.
- The condition must hold for *every* pair (x, y), not just some of them.
- If X and Y are independent, then $g(X)$ and $h(Y)$ are independent for any functions g and h.

Example (A non-example)

Let X be the first roll of two fair dice, and let $Y =$ the sum of the two rolls.
Then $P(X = 6, Y = 12) = P\big((6, 6)\big) = \frac{1}{36}$, whereas $P(X = 6)\, P(Y = 12) = \frac{1}{6} \cdot \frac{1}{36} = \frac{1}{216}$.
Since these are not equal, X and Y are not independent.

Example: Same experiment, yet independent

Example (Two fair die rolls)

Let

$$X = \mathbf{1}_{\{\text{first roll is even}\}}, \qquad Y = \mathbf{1}_{\{\text{sum of the two rolls is even}\}}.$$

Then $X, Y \in \{0, 1\}$, and the joint PMF[a] is

$p_{X,Y}(x,y)$	$y=0$	$y=1$
$x=0$	$1/4$	$1/4$
$x=1$	$1/4$	$1/4$

Since $p_X(0) = p_X(1) = 1/2$ and $p_Y(0) = p_Y(1) = 1/2$,

$$p_{X,Y}(x,y) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = p_X(x)p_Y(y) \qquad \text{for all } x, y.$$

So X and Y are independent.

[a] Even though Y depends on both rolls, the parity of the sum is still equally likely to be even or odd once the parity of the first roll is fixed.

Pop-up quiz

Two independent fair die rolls are performed. Let X

$$X = \text{first roll}, \qquad A = \{\text{the second roll is even}\}.$$

Question: Which statement is correct?

A) X is not independent of A, because A depends on both rolls.

B) X is independent of A, and $p_{X|A}(x) = 1/6$ for all $x \in \{1, \dots, 6\}$.

C) X is independent of A, but only for $x \in \{1, \dots, 5\}$.

D) X is independent of A only because the dice are fair.

Answer: B.

Conditioned on the event $\{\text{sum} = 7\}$, the six outcomes $(1, 6), (2, 5), \dots, (6, 1)$ are equally likely, so the first roll is still uniform on $\{1, \dots, 6\}$.

Follow-up: Would the same conclusion hold if $A = \{\text{sum is } 8\}$?

Independence implies expectation factorization

Theorem

If X and Y are independent, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

More generally, for any functions g and h,

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)].$$

Proof:

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

Example: Factorization in action

Example (A binomial second moment)

Let $X_1, \ldots, X_n$ be independent Bernoulli(p) random variables, and let

$$S = X_1 + \cdots + X_n.$$

Then

$$S^2 = \sum_{i=1}^{n} X_i^2 + 2 \sum_{1 \leq i < j \leq n} X_i X_j.$$

Taking expectations gives

$$\mathbb{E}[S^2] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i X_j].$$

Since $X_i^2 = X_i$ and the X_i are independent,

$$\mathbb{E}[X_i^2] = p, \qquad \mathbb{E}[X_i X_j] = \mathbb{E}[X_i]\mathbb{E}[X_j] = p^2 \quad (i \neq j).$$

Therefore,

$$\mathbb{E}[S^2] = np + n(n-1)p^2.$$

A common misconception

The identity

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

does *not* by itself imply independence.

Example

Let X be uniform on $\{-1, 0, 1\}$, and let $Y = X^2$. Then

$$\mathbb{E}[X] = 0, \qquad \mathbb{E}[Y] = \frac{2}{3}, \qquad \mathbb{E}[XY] = \mathbb{E}[X^3] = 0.$$

So

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y].$$

But Y is completely determined by X, so X and Y are not independent.

Factorization of $\mathbb{E}[XY]$ is a consequence of independence, but not a characterization of independence.

Independence implies variance adds

Theorem

If X and Y are independent, then

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Proof:

$$\begin{aligned}
\mathrm{Var}(X+Y) &= \mathbb{E}[(X+Y)^2] - \left(\mathbb{E}[X] + \mathbb{E}[Y]\right)^2 \\
&= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X]\mathbb{E}[Y] - \mathbb{E}[Y]^2 \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\left(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\right).
\end{aligned}$$

Noting that if X and Y are independent, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ completes the proof.

More generally, for independent $X_1, \ldots, X_n$,

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Example: Variance of a binomial random variable

Example

Let

$$X = X_1 + \cdots + X_n,$$

where $X_1, \ldots, X_n$ are independent Bernoulli(p) random variables.
Since

$$\mathrm{Var}(X_i) = p(1-p),$$

the variance-addition rule gives

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1-p).$$

So for a binomial random variable $X \sim \mathrm{Binomial}(n, p)$,

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

Example: Mean and variance of the sample mean

Example (Sample mean of i.i.d. observations)

Let $X_1, \dots, X_n$ be independent and identically distributed, with

$$\mathbb{E}[X_i] = \mu, \qquad \mathrm{Var}(X_i) = \sigma^2.$$

Define the sample mean

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Then

$$\mathbb{E}[\overline{X}_n] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \mu, \qquad \text{and} \qquad \mathrm{Var}(\overline{X}_n) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}(X_i) = \frac{\sigma^2}{n}.$$

- Averaging preserves the mean.
- Averaging reduces the variance by a factor of n.
- This is one mathematical reason averaging stabilizes noisy observations.

Same marginals, different dependence

Example (Same marginals, different dependence)

Let $X, Y \in \{0, 1\}$, and suppose both are marginally Bernoulli(1/2).

Model 1: independent

$p_{X,Y}(x,y)$	$y=0$	$y=1$
$x=0$	1/4	1/4
$x=1$	1/4	1/4

Model 2: perfectly dependent

$p_{X,Y}(x,y)$	$y=0$	$y=1$
$x=0$	1/2	0
$x=1$	0	1/2

In both models,

$$\mathbb{E}[X] = \mathbb{E}[Y] = \frac{1}{2}, \qquad \mathbb{E}[X+Y] = 1.$$

But

$$\mathrm{Var}(X+Y) = \frac{1}{2} = \mathrm{Var}(X) + \mathrm{Var}(Y) \quad \text{in Model 1,} \qquad \mathrm{Var}(X+Y) = 1 \quad \text{in Model 2.}$$

- Marginal PMFs alone do not determine the dependence structure.
- Dependence can significantly change quantities such as $P(X=Y)$, $\mathbb{E}[XY]$, and $\mathrm{Var}(X+Y)$.

Pop-up quiz

Suppose X and Y are both marginally Bernoulli(1/2).

Model 1: X and Y are independent.
Model 2: $Y = X$.

Question: Which quantity is the same in both models?

A) $\mathbb{E}[XY]$
B) $\mathrm{Var}(X+Y)$
C) $P(X=Y)$
D) $\mathbb{E}[X+Y]$

Answer: D.
In both models, $\mathbb{E}[X] = \mathbb{E}[Y] = 1/2$, so $\mathbb{E}[X+Y] = 1$; the other quantities depend on the dependence structure.

Follow-up: What extra information beyond the marginal PMFs is needed to determine $\mathrm{Var}(X+Y)$?

Wrap-up

Independence

- A random variable X is independent of an event A if conditioning on A does not change the PMF of X.
- Random variables X and Y are independent if

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \qquad \text{for all } (x,y).$$

Main consequences

- If X and Y are independent, then $\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\,\mathbb{E}[h(Y)]$, however, the converses need not hold.
- In particular,

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y], \qquad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

- For i.i.d. observations, $\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$, so averaging reduces variability.

Suggested reading: [BT08, Ch. 2.7]

References

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