

STA 131A: Introduction to Probability Theory

Lecture 17: More Topics on Conditioning

Dogyoon Song

Spring 2026, UC Davis

Announcement

Midterm 2 on Fri, May 15 (10:00 am–10:50 am in class)

- **Arrive early:** The exam starts at 10:00 am and ends at 10:50 am sharp
- **One hand-written cheat sheet:** Letter-size (8.5"×11"), double-sided, brief formulas/notes
- **Calculator:** A simple (non-graphing) scientific calculator is allowed
- **No other materials** beyond the single cheat sheet (no textbooks, etc.)
- **SDC accommodations:** Confirm scheduling with AES online ASAP

Advice for preparation:

- Primary coverage: Lectures 12–19 (including next Wed)
- Core concepts from earlier may be assumed
- Practice midterm(s) and brief solution keys will be posted on course webpage
- Office hours next week:
 - Instructor: Wed, 2:30–3:30 pm
 - TA: Tue 3:00–5:00 pm

Agenda

Last time: Conditioning and independence in the continuous case

- Conditioning on an event
- Conditioning on another random variable

Today: More on conditioning + derived distributions

- Independence of continuous random variables
- Bayes' rule for continuous random variables
- Conditional expectation and law of total expectation
- Conditional expectation as a random variable

Recap: Conditional PDFs

The **conditional PDF** of X given an event A is the density $f_{X|A}$ satisfying

$$P(X \in B | A) = \int_B f_{X|A}(x) dx.$$

If $A = \{X \in C\}$, then

$$f_{X|A}(x) = \begin{cases} \frac{f_X(x)}{P(X \in C)}, & x \in C, \\ 0, & x \notin C. \end{cases}$$

When X and Y are continuous random variables with joint PDF $f_{X,Y}$, for any y with $f_Y(y) > 0$, the **conditional PDF** of X given that $Y = y$, is defined by

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad f_Y(y) > 0.$$

- Recall the illustration from the previous lecture

Independence of continuous random variables

Definition

Jointly continuous random variables X and Y are **independent** if

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{for all } x, y.$$

Equivalent characterizations:

- CDF:

$$F_{X,Y}(x, y) = F_X(x)F_Y(y) \quad \text{for all } x, y,$$

- Conditional PDF: For every y with $f_Y(y) > 0$,

$$f_{X|Y}(x | y) = f_X(x) \quad \text{for all } x.$$

Interpretation: knowing $Y = y$ does not change the density of X .

Example: Independent normal random variables

Example

Let

$$X \sim N(\mu_X, \sigma_X^2), \quad Y \sim N(\mu_Y, \sigma_Y^2),$$

and suppose X and Y are independent.

Then the joint PDF is

$$f_{X,Y}(x, y) = f_X(x) f_Y(y), \quad \forall (x, y) \in \mathbb{R}^2$$

That is,

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y} \exp \left\{ -\frac{(x - \mu_X)^2}{2\sigma_X^2} - \frac{(y - \mu_Y)^2}{2\sigma_Y^2} \right\}.$$

The density contours are ellipses aligned with the coordinate axes.

Consequences of independence

If X and Y are independent, then for suitable functions g, h ,

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

In particular,

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

If X and Y have finite variances, then

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

Caution: linearity of expectation does not require independence:

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

always holds when the expectations exist.

Bayes' rule for continuous random variables

Think of

$X =$ hidden state, $Y =$ observed measurement.

Suppose X has density f_X , and conditional on $X = x$, Y has density $f_{Y|X}(y | x)$. Then

$$f_{X,Y}(x, y) = f_X(x)f_{Y|X}(y | x).$$

By the definition of conditional density,

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x)f_{Y|X}(y | x)}{f_Y(y)}.$$

Hence, we get

$$f_{X|Y}(x | y) = \frac{f_X(x) f_{Y|X}(y | x)}{\int_{-\infty}^{\infty} f_X(t) f_{Y|X}(y | t) dt}, \quad \because f_Y(y) = \int_{-\infty}^{\infty} f_X(t) f_{Y|X}(y | t) dt$$

posterior \propto prior \times likelihood.

Example: Inferring an exponential rate

Example

Let $\Lambda \sim \text{Uniform}(1, 2)$. Conditional on $\Lambda = \lambda$, suppose

$$Y | \Lambda = \lambda \sim \text{Exponential}(\lambda), \quad \text{so} \quad f_{Y|\Lambda}(y | \lambda) = \lambda e^{-\lambda y}, \quad y \geq 0.$$

Question: Given an observed value $Y = y$, what can we say about Λ ?

By Bayes' rule for continuous variables,

$$f_{\Lambda|Y}(\lambda | y) = \frac{f_{\Lambda}(\lambda) f_{Y|\Lambda}(y | \lambda)}{\int_1^2 f_{\Lambda}(t) f_{Y|\Lambda}(y | t) dt}.$$

Since $f_{\Lambda}(\lambda) = 1$ for $1 \leq \lambda \leq 2$,

$$f_{\Lambda|Y}(\lambda | y) = \frac{\lambda e^{-\lambda y}}{\int_1^2 t e^{-ty} dt}, \quad 1 \leq \lambda \leq 2.$$

Interpretation: observing $Y = y$ updates our density over the unknown rate Λ .

Bayes' rule with a discrete hidden state

Sometimes the hidden state X is discrete, but the measurement Y is continuous.

Let X be discrete with PMF p_X , and suppose $Y | X = x$ has density $f_{Y|X}(y | x)$.

Then

$$P(X = x | Y = y) = \frac{p_X(x)f_{Y|X}(y | x)}{\sum_{x'} p_X(x')f_{Y|X}(y | x')}.$$

Again:

posterior \propto prior \times likelihood.

We can treat the case with continuous X and discrete Y similarly: DIY

Example: Signal/source detection

Example

Suppose the hidden source is

$$X \in \{-1, 1\}, \quad P(X = -1) = P(X = 1) = \frac{1}{2}.$$

Given $X = x$, the measurement is noisy:

$$Y | X = x \sim N(x, 1).$$

Question: After observing $Y = y$, compute $P(X = 1 | Y = y)$.
Using Bayes' rule with a discrete hidden state,

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

Since $\frac{\phi(y + 1)}{\phi(y - 1)} = e^{-2y}$, we get $P(X = 1 | Y = y) = \frac{1}{1 + e^{-2y}}$.

Interpretation: larger y gives stronger evidence that the source was $X = 1$.

Summary: Bayes' rule

Continuous hidden state

$$f_{X|Y}(x | y) = \frac{f_X(x)f_{Y|X}(y | x)}{\int f_X(t)f_{Y|X}(y | t) dt}.$$

Discrete hidden state, continuous observation

$$P(X = x | Y = y) = \frac{p_X(x)f_{Y|X}(y | x)}{\sum_{x'} p_X(x')f_{Y|X}(y | x')}.$$

Continuous hidden state, discrete observation: Same logic

Common structure

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{normalizing constant}}.$$

Pop-up quiz: Bayes' rule with likelihoods

A hidden state H can take three values A, B, C , with prior probabilities

$$P(H = A) = 0.2, \quad P(H = B) = 0.5, \quad P(H = C) = 0.3.$$

A continuous measurement Y is observed. At the observed value $Y = y_0$, the likelihood density values are

$$f_{Y|H}(y_0 | A) = 0.1, \quad f_{Y|H}(y_0 | B) = 0.3, \quad f_{Y|H}(y_0 | C) = 0.6.$$

Question: What is $P(H = C | Y = y_0)$?

A) 0.6

B) $\frac{0.6}{0.1 + 0.3 + 0.6}$

C) $\frac{0.3 \cdot 0.6}{0.2 \cdot 0.1 + 0.5 \cdot 0.3 + 0.3 \cdot 0.6}$

D) $\frac{0.3}{0.2 + 0.5 + 0.3}$

Answer: C. Bayes' rule says $P(H = C | Y = y_0) = \frac{P(H=C)f_{Y|H}(y_0|C)}{\sum_{h \in \{A,B,C\}} P(H=h)f_{Y|H}(y_0|h)}$. Here this equals

$$\frac{0.3 \cdot 0.6}{0.2 \cdot 0.1 + 0.5 \cdot 0.3 + 0.3 \cdot 0.6} = \frac{0.18}{0.35} \approx 0.514.$$

Conditional expectation

Definition

For a conditional density $f_{X|Y}(x | y)$, define the **conditional expectation**

$$\mathbb{E}[X | Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x | y) dx.$$

More generally,

$$\mathbb{E}[g(X) | Y = y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x | y) dx.$$

Interpretation: $\mathbb{E}[X | Y = y]$ is the average value of X under the conditional density after $Y = y$ is fixed.

Total expectation theorem

If A_1, \dots, A_n form a partition, then

$$\mathbb{E}[X] = \sum_{i=1}^n P(A_i) \mathbb{E}[X | A_i].$$

Total expectation theorem

For jointly continuous X, Y ,

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \mathbb{E}[X | Y = y] f_Y(y) dy.$$

This is the continuous analogue of the discrete version

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

Interpretation: The unconditional mean is a weighted average of conditional means.

Example: Total expectation for a piecewise PDF (1/2)

Example

Suppose that

$$f_X(x) = \begin{cases} \frac{1}{3}, & 0 \leq x < 1, \\ \frac{2}{3}, & 1 \leq x \leq 2, \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$A_1 = \{0 \leq X < 1\}, \quad A_2 = \{1 \leq X \leq 2\}.$$

Then

$$P(A_1) = \frac{1}{3}, \quad P(A_2) = \frac{2}{3}.$$

Observe that

$$\mathbb{E}[X | A_1] = \frac{1}{2}, \quad \mathbb{E}[X | A_2] = \frac{3}{2}.$$

Therefore,

$$\mathbb{E}[X] = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{3}{2} = \frac{7}{6}.$$

Example: Total expectation for a piecewise PDF (2/2)

Example

For $X \mid A_1 \sim \text{Uniform}(0, 1)$,

$$\mathbb{E}[X^2 \mid A_1] = \frac{1}{3}.$$

For $X \mid A_2 \sim \text{Uniform}(1, 2)$,

$$\mathbb{E}[X^2 \mid A_2] = \frac{1^2 + 1 \cdot 2 + 2^2}{3} = \frac{7}{3}.$$

By total expectation applied to X^2 ,

$$\mathbb{E}[X^2] = \frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{7}{3} = \frac{5}{3}.$$

Thus,

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{5}{3} - \left(\frac{7}{6}\right)^2 = \frac{11}{36}.$$

Example: Train waiting time, revisited

Example (Expected waiting time)

Let

W = waiting time until the next train.

If $A = \{\text{catch the 7:15 train}\}$, then

$$W = 5 - X, \quad X | A \sim \text{Uniform}(0, 5),$$

so

$$\mathbb{E}[W | A] = 5 - \frac{5}{2} = \frac{5}{2}.$$

If $B = \{\text{catch the 7:30 train}\}$, then

$$W = 20 - X, \quad X | B \sim \text{Uniform}(5, 20),$$

so

$$\mathbb{E}[W | B] = 20 - \frac{25}{2} = \frac{15}{2}.$$

By total expectation,

$$\mathbb{E}[W] = \frac{1}{4} \cdot \frac{5}{2} + \frac{3}{4} \cdot \frac{15}{2} = \frac{25}{4}.$$

Conditional expectation as a random variable

Let (X, Y) be jointly distributed.

The conditional expectation of X given Y , denoted by $\mathbb{E}[X | Y]$ is a random variable that takes value $\mathbb{E}[X | Y = y]$ when $Y = y$

Example (Random biased coins)

Suppose the probability of heads is itself random:

$$Y \sim \text{Uniform}(0, 1).$$

Given $Y = y$, toss a coin n times, each with heads probability y . Let

$$X = \text{number of heads.}$$

Then

$$X | Y = y \sim \text{Binomial}(n, y), \quad \text{so} \quad \mathbb{E}[X | Y = y] = ny.$$

Therefore,

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

Law of iterated expectation

Since $\mathbb{E}[X | Y]$ is a random variable, we can think of its expectation:

$$\mathbb{E}[\mathbb{E}[X | Y]] = \begin{cases} \sum_y \mathbb{E}[X | Y = y] p_Y(y), & Y \text{ discrete,} \\ \int_{-\infty}^{\infty} \mathbb{E}[X | Y = y] f_Y(y) dy, & Y \text{ continuous.} \end{cases}$$

Theorem (Law of iterated expectation)

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

Proof. For continuous Y ,

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \int_{-\infty}^{\infty} \mathbb{E}[X | Y = y] f_Y(y) dy \\ &= \mathbb{E}[X], \end{aligned}$$

by the total expectation theorem.

Interpretation: If we average the conditional mean over the distribution of Y , we recover the unconditional mean.

Example: Random biased coin revisited

Example

In the random coin-bias example,

$$Y \sim \text{Uniform}(0, 1), \quad X | Y = y \sim \text{Binomial}(n, y).$$

We found

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

Therefore, by the law of iterated expectation,

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[nY] = n\mathbb{E}[Y] = \frac{n}{2}.$$

Interpretation: On average, the random bias has mean $1/2$, so the unconditional expected number of heads is $n/2$.

Example: Conditional expectation as a mean trend

Example (A simple regression-style model)

Suppose

$$X \sim \text{Uniform}(0, 1),$$

and conditional on $X = x$,

$$Y = g(x) + \varepsilon,$$

where $\mathbb{E}[\varepsilon] = 0$ and ε is independent of X .

Then

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

Thus,

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

Interpretation: $\mathbb{E}[Y \mid X = x]$ describes the mean trend of Y as a function of x .

Pop-up quiz: Conditional expectation

Suppose

$$Y \sim \text{Uniform}(0, 1), \quad X | Y = y \sim \text{Binomial}(10, y).$$

Question: Which expression is correct?

- A) $\mathbb{E}[X | Y = y] = 10y$, so $\mathbb{E}[X | Y] = 10Y$
- B) $\mathbb{E}[X | Y = y] = y/10$, so $\mathbb{E}[X | Y] = Y/10$
- C) $\mathbb{E}[X | Y] = 10\mathbb{E}[Y]$, so it is not random
- D) $\mathbb{E}[X | Y = y] = 5$ for all y

Answer: A.

Given $Y = y$, X is binomial with parameters 10 and y , so its conditional mean is $10y$. Replacing y by the random variable Y gives $\mathbb{E}[X | Y] = 10Y$.

Follow-up: What is $\mathbb{E}[X]$ by the law of iterated expectation?

Wrap-up

Conditional densities and Bayes

- Conditioning on $Y = y$ is defined by slicing the joint density:

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}.$$

- Bayes' rule has the same structure as before:

posterior \propto prior \times likelihood.

Conditional expectation

- $\mathbb{E}[X | Y = y]$ is an average under the conditional density.
- $\mathbb{E}[X | Y]$ is a random variable: plug in the observed/random value of Y .
- Law of iterated expectation:

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

Suggested reading: [BT08, Ch. 3.5–3.6, Ch. 4.3]

References



Dimitri Bertsekas and John N Tsitsiklis.

Introduction to probability, volume 1.

Athena Scientific, 2nd edition, 2008.