Topic 01: Conditional Distributions and Expectations

Ethan P. Marzban University of California, Santa Barbara PSTAT 120B

Outline

1. An Introductory Example
2. Conditional Distributions
3. Conditional Expectations
4. The Gamma Distribution

An Introductory Example

Example

Example

Suppose I roll a fair six-sided die. Then, whatever number the die lands on, I flip that many fair coins. Let X denote the number of heads. What is the **PMF** (probability mass function) of X?

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example

Example

Suppose I roll a fair six-sided die. Then, whatever number the die lands on, I flip that many fair coins. Let X denote the number of heads. What is the **PMF** (probability mass function) of X?

- Now, X *sounds* binomial. But, there's a (not-so slight) problem... Can anyone tell me what that problem is?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

Example

Suppose I roll a fair six-sided die. Then, whatever number the die lands on, I flip that many fair coins. Let X denote the number of heads. What is the **PMF** (probability mass function) of X?

- Now, X *sounds* binomial. But, there's a (not-so slight) problem... Can anyone tell me what that problem is? That's right; the binomial distribution requires a *fixed* number of Bernoulli trials.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

Example

Suppose I roll a fair six-sided die. Then, whatever number the die lands on, I flip that many fair coins. Let X denote the number of heads. What is the **PMF** (probability mass function) of X?

- Now, X *sounds* binomial. But, there's a (not-so slight) problem... Can anyone tell me what that problem is? That's right; the binomial distribution requires a *fixed* number of Bernoulli trials.
  - In other words, if the number of coins I tossed remained fixed across repetitions of this experiment, then X would follow a Binomial distribution. But, because the number of coins I toss *itself* potentially changes across repetitions, we can no longer classify X as being binomially distributed.

Notation

- Let's start off (as we should with any problem) by defining some notation.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Notation

- Let's start off (as we should with any problem) by defining some notation.
- Specifically, it seems like I need to keep track of two things: the result of the die roll, and the number of heads in the resulting tosses of the coin.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Notation

- Let's start off (as we should with any problem) by defining some notation.
- Specifically, it seems like I need to keep track of two things: the result of the die roll, and the number of heads in the resulting tosses of the coin.
- As such, let's assign a random variable to each of these quantities:

$$\begin{aligned} N &:= \text{result of the die roll} \\ X &:= \text{number of heads among the coin tosses} \end{aligned}$$

Assumptions

- From the problem statement, it's safe to assume

$$N \sim \text{DiscUnif}\{1, 2, 3, 4, 5, 6\}$$

  that is, that N follows the discrete uniform distribution on the set $\{1, 2, \cdots, 6\}$.

Assumptions

- From the problem statement, it's safe to assume

$$N \sim \text{DiscUnif}\{1, 2, 3, 4, 5, 6\}$$

  that is, that N follows the discrete uniform distribution on the set $\{1, 2, \cdots, 6\}$.
- Now, to reiterate what we said at the beginning of this discussion, it is **NOT** correct to simply say that X is binomially distributed!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Assumptions

- From the problem statement, it's safe to assume

$$N \sim \text{DiscUnif}\{1, 2, 3, 4, 5, 6\}$$

  that is, that N follows the discrete uniform distribution on the set $\{1, 2, \cdots, 6\}$.
- Now, to reiterate what we said at the beginning of this discussion, it is **NOT** correct to simply say that X is binomially distributed!
- But, that doesn't mean we can't get at its PMF directly.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example

- For example, suppose I want to compute $\mathbb{P}(X = 2)$; i.e. say we wanted to compute the probability of observing exactly 2 heads.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- For example, suppose I want to compute $\mathbb{P}(X = 2)$; i.e. say we wanted to compute the probability of observing exactly 2 heads.
- Again, the issue is that we don't know how many coins we tossed!

Example

- For example, suppose I want to compute $\mathbb{P}(X = 2)$; i.e. say we wanted to compute the probability of observing exactly 2 heads.
- Again, the issue is that we don't know how many coins we tossed!
- *If* we knew how many coins we tossed - say, for example, 6 - then we'd be in business! Specifically, the probability of observing 2 heads among six tosses of a fair coin is easily computed using the Binomial PMF: $\binom{6}{2}(1/2)^6$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example

- For example, suppose I want to compute $\mathbb{P}(X = 2)$; i.e. say we wanted to compute the probability of observing exactly 2 heads.
- Again, the issue is that we don't know how many coins we tossed!
- *If* we knew how many coins we tossed - say, for example, 6 - then we'd be in business! Specifically, the probability of observing 2 heads among six tosses of a fair coin is easily computed using the Binomial PMF: $\binom{6}{2}(1/2)^6$.
- In slightly more formal language - specifically, the language of **conditional probabilities**, what we have just shown is that

$$\mathbb{P}(X = 2 \mid N = 6) = \binom{6}{2}\left(\frac{1}{2}\right)^6$$

Example (cont'd)

- Let's get a bit more practice with understanding our notation! What is, say, $\mathbb{P}(X = 2 \mid N = 5)$?

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example (cont'd)

- Let's get a bit more practice with understanding our notation! What is, say, $\mathbb{P}(X = 2 \mid N = 5)$?
- Well, in words, this is asking us to compute the probability of observing 2 heads among 5 tosses of a fair coin.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example (cont'd)

- Let's get a bit more practice with understanding our notation! What is, say, $\mathbb{P}(X = 2 \mid N = 5)$?
- Well, in words, this is asking us to compute the probability of observing 2 heads among 5 tosses of a fair coin.
- We can again use the Binomial formula:

$$\mathbb{P}(X = 2 \mid N = 5) = \binom{5}{2}\left(\frac{1}{2}\right)^5$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Example (cont'd)

- Generalizing a bit, let's see if we can find an expression for $\mathbb{P}(X = 2 \mid N = n)$, where n is an arbitrary integer in the set $\{1, 2, \cdots, 6\}$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example (cont'd)

- Generalizing a bit, let's see if we can find an expression for $\mathbb{P}(X = 2 \mid N = n)$, where n is an arbitrary integer in the set $\{1, 2, \cdots, 6\}$.
- Again, in words this is asking us to compute the probability of observing 2 heads among n tosses of a fair coin.

Example (cont'd)

- Generalizing a bit, let's see if we can find an expression for $\mathbb{P}(X = 2 \mid N = n)$, where n is an arbitrary integer in the set $\{1, 2, \cdots, 6\}$.
- Again, in words this is asking us to compute the probability of observing 2 heads among n tosses of a fair coin.
- Once again, we use the Binomial PMF:

$$\mathbb{P}(X = 2 \mid N = n) = \binom{n}{2} \left(\frac{1}{2} \right)^n$$

Example (cont'd)

- Generalizing one step further:

$$\mathbb{P}(X = x \mid N = n) = \binom{n}{x}\left(\frac{1}{2}\right)^n$$

where $x \in \{1, 2, \cdots, n\}$ and $n \in \{1, 2, \cdots, 6\}$.

Example (cont'd)

- Generalizing one step further:

$$\mathbb{P}(X = x \mid N = n) = \binom{n}{x}\left(\frac{1}{2}\right)^n$$

where $x \in \{1, 2, \cdots, n\}$ and $n \in \{1, 2, \cdots, 6\}$.

  - BTW, can anyone tell me what happens if $x > n$? Think both in terms of intuition, as well as the mathematical formula above!

Example (cont'd)

- Generalizing one step further:

$$\mathbb{P}(X = x \mid N = n) = \binom{n}{x}\left(\frac{1}{2}\right)^n$$

where $x \in \{1, 2, \cdots, n\}$ and $n \in \{1, 2, \cdots, 6\}$.

  - BTW, can anyone tell me what happens if $x > n$? Think both in terms of intuition, as well as the mathematical formula above!

- Also, don't forget:

$$\mathbb{P}(N = n) = \frac{1}{6}, \qquad \text{if } n \in \{1, 2, \cdots, 6\}$$

Clicker Question!

Clicker Question 1

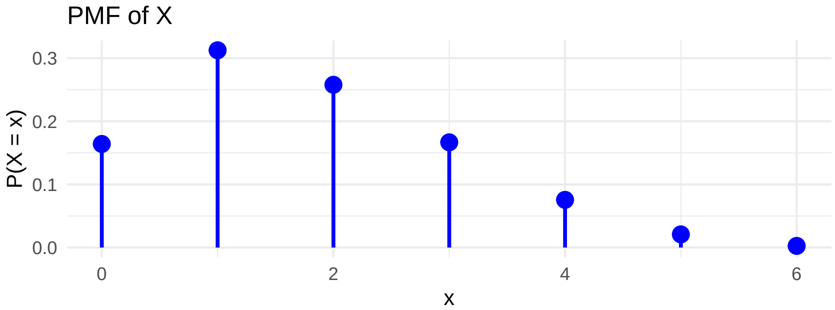
Based on the work we've done so far, which PSTAT 120A topic do you think will help us complete the calculation for $\mathbb{P}(X = x)$?

(A) The Complement Rule
(B) The Inclusion-Exclusion Principle (aka the Addition Rule)
(C) The Law of Total Probability
(D) The Central Limit Theorem
(E) None of the above

Example (cont'd)

- So, once the dust settles, we have

$$\mathbb{P}(X = x) = \frac{1}{6} \sum_{n=1}^{6} \binom{n}{x} \left(\frac{1}{2} \right)^n, \qquad \text{for } x \in \{0, 1, \cdots, 6\}$$



Recap

- So, what did we learn?

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Recap

- So, what did we learn?
- Well, I hope one thing became clear: after conditioning on the result of the die roll, our considerations for the number of heads became much simpler!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Recap

- So, what did we learn?
- Well, I hope one thing became clear: after conditioning on the result of the die roll, our considerations for the number of heads became much simpler!
- In a way, it's tempting to write

$$(X \mid N = n) \sim \text{Bin}(n, 1/2)$$

to indicate the fact that, if we knew the die landed on n, then X becomes binomially distributed.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Recap

- So, what did we learn?
- Well, I hope one thing became clear: after conditioning on the result of the die roll, our considerations for the number of heads became much simpler!
- In a way, it's tempting to write

$$(X \mid N = n) \sim \text{Bin}(n, 1/2)$$

to indicate the fact that, if we knew the die landed on n, then X becomes binomially distributed.

- Indeed, such notation *is* proper - well, it *will* be after we discuss its meaning more carefully!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Conditional Distributions

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Conditional Distributions

- Thinking back to our PSTAT 120A days (as we will often do in this class), recall the notion of a **joint probability density/mass function**.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Conditional Distributions

- Thinking back to our PSTAT 120A days (as we will often do in this class), recall the notion of a **joint probability density/mass function**.
- Essentially, a joint PDF/PMF is a way to jointly specify/quantify the distribution of two random variables that are potentially related in some way.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Conditional Distributions

- Thinking back to our PSTAT 120A days (as we will often do in this class), recall the notion of a **joint probability density/mass function**.
- Essentially, a joint PDF/PMF is a way to jointly specify/quantify the distribution of two random variables that are potentially related in some way.
- Let's consider (temporarily) the discrete and continuous cases separately.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Joint PMF

- Consider two random variables X and Y, both of which are discrete.

Joint PMF

- Consider two random variables X and Y, both of which are discrete.
- Then the joint PMF of X and Y, notated $p_{X,Y}(x, y)$ is defined as

$$p_{X,Y}(x, y) := \mathbb{P}(X = x,\ Y = y)$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Joint PMF

- Consider two random variables X and Y, both of which are discrete.
- Then the joint PMF of X and Y, notated $p_{X,Y}(x, y)$ is defined as

$$p_{X,Y}(x, y) := \mathbb{P}(X = x,\ Y = y)$$

  - For example, letting X and N be defined as they were in our initial die-and-coin example, then $p_{X,N}(x, n)$ is the probability that we observed x heads <u>and</u> the die landed on n.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Joint PMF

- Consider two random variables X and Y, both of which are discrete.
- Then the joint PMF of X and Y, notated $p_{X,Y}(x, y)$ is defined as

$$p_{X,Y}(x, y) := \mathbb{P}(X = x,\ Y = y)$$

  - For example, letting X and N be defined as they were in our initial die-and-coin example, then $p_{X,N}(x, n)$ is the probability that we observed x heads and the die landed on n.

- What happens if we divide both sides of our definition for $p_{X,Y}(x, y)$ by $p_Y(y) := \mathbb{P}(Y = y)$?

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Leadup

- Well, first things first - we need to make sure we're not dividing by zero! So, let's <u>assume</u> that y is such that $\mathbb{P}(Y = y) \neq 0$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Leadup

- Well, first things first - we need to make sure we're not dividing by zero! So, let's <u>assume</u> that y is such that $\mathbb{P}(Y = y) \neq 0$.
- Then, we find that

$$\frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Leadup

- Well, first things first - we need to make sure we're not dividing by zero! So, let's <u>assume</u> that y is such that $\mathbb{P}(Y = y) \neq 0$.
- Then, we find that

$$\frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

- The RHS should look familiar! Specifically, if we let $A := \{X = x\}$ and $B := \{Y = y\}$, then the RHS is simply

$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Leadup

- Indeed, this is just the definition of $\mathbb{P}(A \mid B)$!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Leadup

- Indeed, this is just the definition of $\mathbb{P}(A \mid B)$!
- So,

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

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Leadup

- Indeed, this is just the definition of $\mathbb{P}(A \mid B)$!
- So,

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

- We can use the shorthand $p_{X|Y}(x \mid y)$ to denote the RHS. That is, we define

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

and call this the **conditional PMF** of X given Y.

Leadup

- There are a lot of variables flying around, so maybe it'll be helpful to connect things with our initial die-and-coin example.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Leadup

- There are a lot of variables flying around, so maybe it'll be helpful to connect things with our initial die-and-coin example.
- I previously argued that

$$\mathbb{P}(X = x \mid N = n) = \binom{n}{x} \left(\frac{1}{2} \right)^n$$

based on the setup of the problem.

Leadup

- There are a lot of variables flying around, so maybe it'll be helpful to connect things with our initial die-and-coin example.
- I previously argued that

$$\mathbb{P}(X = x \mid N = n) = \binom{n}{x} \left(\frac{1}{2} \right)^n$$

  based on the setup of the problem.
- Indeed, this is just the conditional PMF of X given $N = n$:

$$p_{X|N}(x \mid n) = \binom{n}{x} \left(\frac{1}{2} \right)^n$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Conditional PMF

- Okay, let's make this a bit more formal.

Conditional PMF

- Okay, let's make this a bit more formal.

Definition (Conditional PMF)

Given a pair of bivariate random variables (X, Y), we define the **conditional PMF** of X given $Y = y$ to be

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

provided that y is such that $p_Y(y) \neq 0$. If $p_Y(y) = 0$, then $p_{X|Y}(x \mid y)$ is undefined.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Some Notes

- Though $p_{X|Y}(x \mid y)$ does involve both x and y, we typically view it as a function of x alone.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Some Notes

- Though $p_{X|Y}(x \mid y)$ does involve both x and y, we typically view it as a function of x alone.
- I'd like to also stress the fact that $p_{X|Y}(x \mid y)$ is undefined whenever $p_Y(y) = 0$. It's not 0, or ∞ - it's just undefined.

Some Notes

- Though $p_{X|Y}(x \mid y)$ does involve both x and y, we typically view it as a function of x alone.
- I'd like to also stress the fact that $p_{X|Y}(x \mid y)$ is undefined whenever $p_Y(y) = 0$. It's not 0, or ∞ - it's just undefined.

Theorem

For any fixed value of y (such that all quantities are defined), $p_{X|Y}(x \mid y)$ is a valid PMF.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Proof.

- Recall that to verify a given function is a valid PMF, we need to establish two things: nonnegativity, and summation to unity.

$\square$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Proof.

- Recall that to verify a given function is a valid PMF, we need to establish two things: nonnegativity, and summation to unity.
- For nonnegativity, it suffices to note that $p_{X|Y}(x \mid y) := \mathbb{P}(X = x \mid Y = y)$ is indeed a *probability*, and is hence always between 0 and 1 (and, therefore, nonnegative).

$\square$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Proof.

- Recall that to verify a given function is a valid PMF, we need to establish two things: nonnegativity, and summation to unity.
- For nonnegativity, it suffices to note that $p_{X|Y}(x \mid y) := \mathbb{P}(X = x \mid Y = y)$ is indeed a *probability*, and is hence always between 0 and 1 (and, therefore, nonnegative).
- For summation to unity:

$$\sum_x p_{X|Y}(x \mid y)$$

$\square$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Proof.

- Recall that to verify a given function is a valid PMF, we need to establish two things: nonnegativity, and summation to unity.
- For nonnegativity, it suffices to note that $p_{X|Y}(x \mid y) := \mathbb{P}(X = x \mid Y = y)$ is indeed a *probability*, and is hence always between 0 and 1 (and, therefore, nonnegative).
- For summation to unity:

$$\sum_x p_{X|Y}(x \mid y) = \sum_x \frac{p_{X,Y}(x, y)}{p_Y(y)} \qquad \text{[Definition of } p_{X|Y}(x \mid y)\text{]}$$

$\square$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Proof.

- Recall that to verify a given function is a valid PMF, we need to establish two things: nonnegativity, and summation to unity.
- For nonnegativity, it suffices to note that $p_{X|Y}(x \mid y) := \mathbb{P}(X = x \mid Y = y)$ is indeed a *probability*, and is hence always between 0 and 1 (and, therefore, nonnegative).
- For summation to unity:

$$\begin{aligned}
\sum_x p_{X|Y}(x \mid y) &= \sum_x \frac{p_{X,Y}(x, y)}{p_Y(y)} && \text{[Definition of } p_{X|Y}(x \mid y)\text{]} \\
&= \frac{1}{p_Y(y)} \sum_x p_{X,Y}(x, y) && \text{[Algebra]}
\end{aligned}$$

$\square$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Proof.

- Recall that to verify a given function is a valid PMF, we need to establish two things: nonnegativity, and summation to unity.
- For nonnegativity, it suffices to note that $p_{X|Y}(x \mid y) := \mathbb{P}(X = x \mid Y = y)$ is indeed a *probability*, and is hence always between 0 and 1 (and, therefore, nonnegative).
- For summation to unity:

$$\begin{aligned} \sum_x p_{X|Y}(x \mid y) &= \sum_x \frac{p_{X,Y}(x,y)}{p_Y(y)} && \text{[Definition of } p_{X|Y}(x \mid y)\text{]} \\ &= \frac{1}{p_Y(y)} \sum_x p_{X,Y}(x,y) && \text{[Algebra]} \\ &= \frac{1}{\cancel{p_Y(y)}} \cdot \cancel{p_Y(y)} \end{aligned}$$

$\square$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Proof.

- Recall that to verify a given function is a valid PMF, we need to establish two things: nonnegativity, and summation to unity.
- For nonnegativity, it suffices to note that $p_{X|Y}(x \mid y) := \mathbb{P}(X = x \mid Y = y)$ is indeed a *probability*, and is hence always between 0 and 1 (and, therefore, nonnegative).
- For summation to unity:

$$
\begin{aligned}
\sum_x p_{X|Y}(x \mid y) &= \sum_x \frac{p_{X,Y}(x, y)}{p_Y(y)} && \text{[Definition of } p_{X|Y}(x \mid y)\text{]} \\
&= \frac{1}{p_Y(y)} \sum_x p_{X,Y}(x, y) && \text{[Algebra]} \\
&= \frac{1}{\cancel{p_Y(y)}} \cdot \cancel{p_Y(y)} = 1 && \text{[Joint PMF to Marginal PMF]}
\end{aligned}
$$

$\square$

Joint PDF

- Alright, let's talk about the continuous case!

Joint PDF

- Alright, let's talk about the continuous case!
- That is, consider a pair of bivariate random variables (X, Y) that are both continuous. Then, information about X and Y is jointly specified through the joint PDF

$$f_{X,Y}(x, y)$$

Joint PDF

- Alright, let's talk about the continuous case!
- That is, consider a pair of bivariate random variables (X, Y) that are both continuous. Then, information about X and Y is jointly specified through the joint PDF
$$f_{X,Y}(x, y)$$
- Now, unlike the discrete case, recall that the values of $f_{X,Y}(x, y)$ do *not* represent probabilities - rather, volumes underneath $f_{X,Y}(x, y)$ represent probabilities.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Joint PDF

- Nevertheless, motivated by our considerations in the discrete case, we can still posit the following definition:

Definition (Conditional PDF)

Given a pair of bivariate random variables (X, Y), we define the **conditional PDF** of X given $Y = y$ to be

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

provided that y is such that $f_Y(y) \neq 0$. If $f_Y(y) = 0$, then $f_{X|Y}(x \mid y)$ is undefined.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Theorem

Theorem

For any fixed value of y (such that all quantities are defined), $f_{X|Y}(x \mid y)$ is a valid PDF.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Theorem

Theorem

For any fixed value of y (such that all quantities are defined), $f_{X|Y}(x \mid y)$ is a valid PDF.

- I encourage you to try the proof of this on your own!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Chalkboard Example

Suppose (X, Y) is a continuous bivariate random vector with joint p.d.f. given by

$$f_{X,Y}(x, y) = \begin{cases} \lambda^3 x e^{-\lambda y} & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

(a) Find $f_Y(y)$, the marginal density of Y.

(b) Find $f_{X|Y}(x \mid y)$, the conditional density of $(X \mid Y = y)$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Working With Conditional Densities

- Once we understand the idea that $f_{X|Y}(x \mid y)$ functions behaves like a PDF (because, in a way, it *is* one), the following definition becomes natural:

UC SANTA BARBARA
Department of Statistics and Applied Probability

Working With Conditional Densities

- Once we understand the idea that $f_{X|Y}(x \mid y)$ functions behaves like a PDF (because, in a way, it *is* one), the following definition becomes natural:

Definition

Given a pair (X, Y) of continuous random variables,

$$\mathbb{P}(X \in A \mid Y = y) = \int_A f_{X|Y}(x \mid y)\, \mathrm{d}y$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Chalkboard Example (cont'd)

Suppose (X, Y) is a continuous bivariate random vector with joint p.d.f. given by

$$f_{X,Y}(x, y) = \begin{cases} \lambda^3 x e^{-\lambda y} & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

(c) Compute $\mathbb{P}(X \geq 1 \mid Y \geq 2)$

(d) Compute $\mathbb{P}(X \geq 1 \mid Y = 2)$

Marginal PMFs/PDFs

- Using the connection between conditional PMFs/PDFs and joint PMFs/PDFs, we can see how one can recover *marginal* PMFs/PDFs from condtitional PMFs/PDFs:

Marginal PMFs/PDFs

- Using the connection between conditional PMFs/PDFs and joint PMFs/PDFs, we can see how one can recover *marginal* PMFs/PDFs from condtitional PMFs/PDFs:

Theorem

(1) If (X, Y) denotes a pair of continuous random variables, then

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y) \, \mathrm{d}y$$

with an analogous formula for $f_Y(y)$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Marginal PMFs/PDFs

- Using the connection between conditional PMFs/PDFs and joint PMFs/PDFs, we can see how one can recover *marginal* PMFs/PDFs from condtitional PMFs/PDFs:

Theorem

(2) If (X, Y) denotes a pair of discrete random variables, then

$$p_X(x) = \sum_y p_{X|Y}(x \mid y) p_Y(y)$$

with an analogous formula for $p_Y(y)$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Proof Outlines

- The proofs for both of these facts are similar: start by writing the integrand/summand as a ratio involving a joint, cancel like terms, and integrate/sum.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Proof Outlines

- The proofs for both of these facts are similar: start by writing the integrand/summand as a ratio involving a joint, cancel like terms, and integrate/sum.
  - I highly encourage you to try these proofs as an exercise in reviewing some PSTAT 120A-related definitions and results!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Proof Outlines

- The proofs for both of these facts are similar: start by writing the integrand/summand as a ratio involving a joint, cancel like terms, and integrate/sum.
  - I highly encourage you to try these proofs as an exercise in reviewing some PSTAT 120A-related definitions and results!

- Now, something interesting happens when we consider the *mixed* case.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Mixed Case

- What do I mean by the "mixed" case?

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Mixed Case

- What do I mean by the “mixed” case?
- Well, for example, consider a discrete random variable X and a continuous random variable Y. Can we define something resembling a conditional PMF/PDF?

UC SANTA BARBARA
Department of Statistics and Applied Probability

Mixed Case

- What do I mean by the "mixed" case?
- Well, for example, consider a discrete random variable X and a continuous random variable Y. Can we define something resembling a conditional PMF/PDF?
- The answer is, perhaps surprisingly, "*yes*"!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Mixed Case

- What do I mean by the “mixed” case?
- Well, for example, consider a discrete random variable X and a continuous random variable Y. Can we define something resembling a conditional PMF/PDF?
- The answer is, perhaps surprisingly, “*yes*”!
- As an example (which you will consider on your homework), suppose Y denotes the number of diseased trees in a forest (and is hence discrete), but that the *rate* of diseased trees (which is continuous) itself varies according to some distribution. Despite the fact that the number and rate of diseased trees are discrete and continuous, respectively, it still makes perfect sense to talk about the *unconditional* distribution of the number of diseased trees.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Results

Theorem

Consider a random vector (X, Y).

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Results

Theorem

Consider a random vector (X, Y).

- If X is discrete and Y is continuous, then
$$p_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y) \ \mathrm{d}y$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Results

Theorem

Consider a random vector (X, Y).

- If X is discrete and Y is continuous, then $p_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y) \, \mathrm{d}y$
- If X is continuous and Y is discrete, then $f_X(x) = \sum_y f_{X|Y}(x \mid y) p_Y(y)$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Results

Theorem

Consider a random vector (X, Y).

- If X is discrete and Y is continuous, then
$$p_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y) f_Y(y) \, \mathrm{d}y$$
- If X is continuous and Y is discrete, then
$$f_X(x) = \sum_{y} f_{X|Y}(x \mid y) p_Y(y)$$

- Moral: for mixed random vectors, integrate/sum according to the type of variable being conditioned on.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Conditional Expectations

Leadup

- Recall that, given a random variable X with density $f_X(x)$, the **Law of the Unconscious Statistician** (LOTUS) states

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \, \mathrm{d}x$$

for well-behaved functions $g : \mathbb{R} \to \mathbb{R}$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Leadup

- Recall that, given a random variable X with density $f_X(x)$, the **Law of the Unconscious Statistician** (LOTUS) states

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \, \mathrm{d}x$$

  for well-behaved functions $g : \mathbb{R} \to \mathbb{R}$.

- Given that, for a pair (X, Y) of continuous random variables, $f_{X|Y}(x \mid y)$ represents a density function [essentially of the "random variable" $(X \mid Y = y)$], it's perhaps natural to define:

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Conditional Expectation; First Pass

Definition (Conditional Expectation; First Pass)

(1) Given a continuous pair (X, Y) of random variables and a well-behaved function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g(X) \mid Y = y] := \int_{-\infty}^{\infty} g(x) f_{X|Y}(x \mid y)\ \mathrm{d}x$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Conditional Expectation; First Pass

Definition (Conditional Expectation; First Pass)

(1) Given a continuous pair (X, Y) of random variables and a well-behaved function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g(X) \mid Y = y] := \int_{-\infty}^{\infty} g(x) f_{X|Y}(x \mid y) \, \mathrm{d}x$$

(2) Given a discrete pair (X, Y) of random variables and a well-behaved function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g(X) \mid Y = y] := \sum_{x} g(x) p_{X|Y}(x \mid y)$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Chalkboard Example

Suppose (X, Y) is a continuous bivariate random vector with joint p.d.f. given by

$$f_{X,Y}(x, y) = \begin{cases} \lambda^3 x e^{-\lambda y} & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute $\mathbb{E}[X \mid Y = y]$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Properties of Conditional Expectations

Theorem (Properties of Conditional Expectations, I)

(I) **(Linearity)**
$$\mathbb{E}[aX + bY + c \mid Z = z] = a\mathbb{E}[X \mid Z = z] + b\mathbb{E}[Y \mid Z = z] + c.$$

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

(III) If $X \perp Y$, then $\mathbb{E}[X \mid Y = y] = \mathbb{E}[X]$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Conditional Expectation

Definition (Conditional Expectation)

Given random variables X and Y, and the function $h(y) := \mathbb{E}[X \mid Y = y]$, we define the **conditional expectation of X given Y**, notated $\mathbb{E}[X \mid Y]$, to be $h(Y)$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Conditional Expectation

Definition (Conditional Expectation)

Given random variables X and Y, and the function $h(y) := \mathbb{E}[X \mid Y = y]$, we define the **conditional expectation of X given Y**, notated $\mathbb{E}[X \mid Y]$, to be $h(Y)$.

- So, in practice, here's how we compute $\mathbb{E}[X \mid Y]$:

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Conditional Expectation

Definition (Conditional Expectation)

Given random variables X and Y, and the function $h(y) := \mathbb{E}[X \mid Y = y]$, we define the **conditional expectation of X given Y**, notated $\mathbb{E}[X \mid Y]$, to be $h(Y)$.

- So, in practice, here's how we compute $\mathbb{E}[X \mid Y]$:
  (1) Compute $h(y) := \mathbb{E}[X \mid Y = y]$ (which will be a function of y)

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Conditional Expectation

Definition (Conditional Expectation)

Given random variables X and Y, and the function $h(y) := \mathbb{E}[X \mid Y = y]$, we define the **conditional expectation of X given Y**, notated $\mathbb{E}[X \mid Y]$, to be $h(Y)$.

- So, in practice, here's how we compute $\mathbb{E}[X \mid Y]$:
  (1) Compute $h(y) := \mathbb{E}[X \mid Y = y]$ (which will be a function of y)
  (2) Substitute Y in place of y in our expression from step (1).

UC SANTA BARBARA
Department of Statistics and Applied Probability

Conditional Expectation

Definition (Conditional Expectation)

Given random variables X and Y, and the function $h(y) := \mathbb{E}[X \mid Y = y]$, we define the **conditional expectation of X given Y**, notated $\mathbb{E}[X \mid Y]$, to be $h(Y)$.

- So, in practice, here's how we compute $\mathbb{E}[X \mid Y]$:
  (1) Compute $h(y) := \mathbb{E}[X \mid Y = y]$ (which will be a function of y)
  (2) Substitute Y in place of y in our expression from step (1).

- Note: $\mathbb{E}[X \mid Y]$ will be a <u>random variable</u>!

UC SANTA BARBARA
Department of Statistics and Applied Probability

Chalkboard Example

Suppose (X, Y) is a continuous bivariate random vector with joint p.d.f. given by

$$f_{X,Y}(x, y) = \begin{cases} \lambda^3 x e^{-\lambda y} & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$

Compute $\mathbb{E}[X \mid Y]$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Properties of Conditional Expectations

Theorem (Properties of Conditional Expectations, II)

(I) **(Linearity)** $\mathbb{E}[aX + bY + c \mid Z] = a\mathbb{E}[X \mid Z] + b\mathbb{E}[Y \mid Z] + c$.

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

(III) If $X \perp Y$, then $\mathbb{E}[X \mid Y] = X$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Properties of Conditional Expectations

Theorem (Properties of Conditional Expectations, II)

(I) **(Linearity)** $\mathbb{E}[aX + bY + c \mid Z] = a\mathbb{E}[X \mid Z] + b\mathbb{E}[Y \mid Z] + c$.

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

(III) If $X \perp Y$, then $\mathbb{E}[X \mid Y] = X$.

- Note how these follow almost directly from the theorem titled (Properties of Conditional Expectations, I), from a few slides ago.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Leadup

- Since $\mathbb{E}[X \mid Y]$ is itself a random variable, it makes sense to take *its* expectation: $\mathbb{E}[\mathbb{E}[X \mid Y]]$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Leadup

- Since $\mathbb{E}[X \mid Y]$ is itself a random variable, it makes sense to take *its* expectation: $\mathbb{E}[\mathbb{E}[X \mid Y]]$.
- It's important we understand what each of these expectations are taken with respect to.

Leadup

- Since $\mathbb{E}[X \mid Y]$ is itself a random variable, it makes sense to take *its* expectation: $\mathbb{E}[\mathbb{E}[X \mid Y]]$.
- It's important we understand what each of these expectations are taken with respect to.
    - The inner expectation is taken with respect to the conditional distribution $(X \mid Y)$

Leadup

- Since $\mathbb{E}[X \mid Y]$ is itself a random variable, it makes sense to take *its* expectation: $\mathbb{E}[\mathbb{E}[X \mid Y]]$.
- It's important we understand what each of these expectations are taken with respect to.
    - The inner expectation is taken with respect to the conditional distribution $(X \mid Y)$
    - The outer expectation is taken with respect to Y.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Leadup

- Since $\mathbb{E}[X \mid Y]$ is itself a random variable, it makes sense to take *its* expectation: $\mathbb{E}[\mathbb{E}[X \mid Y]]$.
- It's important we understand what each of these expectations are taken with respect to.
    - The inner expectation is taken with respect to the conditional distribution $(X \mid Y)$
    - The outer expectation is taken with respect to Y.
    - Hence, it would perhaps be more accurate to write $\mathbb{E}_Y[\mathbb{E}_{X|Y}(X \mid Y)]$, but we often drop the subscripts for convenience.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Continuous Realm

- To be explicit, assume X is a continuous random variable, and define $h(y) := \mathbb{E}[X \mid Y = y]$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Continuous Realm

- To be explicit, assume X is a continuous random variable, and define $h(y) := \mathbb{E}[X \mid Y = y]$.

$$\mathbb{E}[\mathbb{E}[X \mid Y]] =: \mathbb{E}[h(Y)]$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Continuous Realm

- To be explicit, assume X is a continuous random variable, and define $h(y) := \mathbb{E}[X \mid Y = y]$.

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X \mid Y]] &=: \mathbb{E}[h(Y)] \\ &= \int_{\mathbb{R}} h(y) f_Y(y) \, \mathrm{d}y \qquad \text{[LOTUS]} \end{aligned}$$

Continuous Realm

- To be explicit, assume X is a continuous random variable, and define $h(y) := \mathbb{E}[X \mid Y = y]$.

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[X \mid Y]] &=: \mathbb{E}[h(Y)] \\
&= \int_{\mathbb{R}} h(y) f_Y(y)\ \mathrm{d}y && \text{[LOTUS]} \\
&= \int_{\mathbb{R}} \mathbb{E}[X \mid Y = y] f_Y(y)\ \mathrm{d}y && \text{[Def. of } h(y)\text{]}
\end{aligned}
$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Continuous Realm

- To be explicit, assume X is a continuous random variable, and define $h(y) := \mathbb{E}[X \mid Y = y]$.

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[X \mid Y]] &=: \mathbb{E}[h(Y)] \\
&= \int_{\mathbb{R}} h(y) f_Y(y)\ \mathrm{d}y && \text{[LOTUS]} \\
&= \int_{\mathbb{R}} \mathbb{E}[X \mid Y = y] f_Y(y)\ \mathrm{d}y && \text{[Def. of } h(y)\text{]} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X \mid Y}(x \mid y) f_Y(y)\ \mathrm{d}x\ \mathrm{d}y && \text{[Def. of } h(y)\text{]}
\end{aligned}
$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Continuous Realm, cont'd

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X|Y}(x \mid y) f_Y(y) \, \mathrm{d}x \, \mathrm{d}y$$

[From prev. slide]

Continuous Realm, cont'd

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X \mid Y]] &= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X|Y}(x \mid y) f_Y(y) \, \mathrm{d}x \, \mathrm{d}y && \text{[From prev. slide]} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x \cdot \frac{f_{X,Y}(x,y)}{\cancel{f_Y(y)}} \cdot \cancel{f_Y(y)} \, \mathrm{d}x \, \mathrm{d}y && \text{[Def of } f_{X|Y}(x \mid y)\text{]}
\end{aligned}$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Continuous Realm, cont'd

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[X \mid Y]] &= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X \mid Y}(x \mid y) f_Y(y) \, \mathrm{d}x \, \mathrm{d}y && \text{[From prev. slide]} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x \cdot \frac{f_{X,Y}(x,y)}{\cancel{f_Y(y)}} \cdot \cancel{f_Y(y)} \, \mathrm{d}x \, \mathrm{d}y && \text{[Def of } f_{X \mid Y}(x \mid y)\text{]} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X,Y}(x,y) \, \mathrm{d}x \, \mathrm{d}y && \text{[Simplifying]}
\end{aligned}
$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Continuous Realm, cont'd

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[X \mid Y]] &= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X|Y}(x \mid y) f_Y(y) \ \mathrm{d}x \ \mathrm{d}y && \text{[From prev. slide]} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x \cdot \frac{f_{X,Y}(x,y)}{\cancel{f_Y(y)}} \cdot \cancel{f_Y(y)} \ \mathrm{d}x \ \mathrm{d}y && \text{[Def of } f_{X|Y}(x \mid y)\text{]} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X,Y}(x,y) \ \mathrm{d}x \ \mathrm{d}y && \text{[Simplifying]} \\
&= \int_{\mathbb{R}} x \left(\int_{\mathbb{R}} f_{X,Y}(x,y) \ \mathrm{d}y \right) \ \mathrm{d}x && \text{[Rev. Order of int.]}
\end{aligned}$$

Continuous Realm, cont'd

$$
\begin{aligned}
\mathbb{E}[\mathbb{E}[X \mid Y]] &= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X \mid Y}(x \mid y) f_Y(y) \, \mathrm{d}x \, \mathrm{d}y && \text{[From prev. slide]} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x \cdot \frac{f_{X,Y}(x,y)}{\cancel{f_Y(y)}} \cdot \cancel{f_Y(y)} \, \mathrm{d}x \, \mathrm{d}y && \text{[Def of } f_{X \mid Y}(x \mid y)\text{]} \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X,Y}(x,y) \, \mathrm{d}x \, \mathrm{d}y && \text{[Simplifying]} \\
&= \int_{\mathbb{R}} x \left(\int_{\mathbb{R}} f_{X,Y}(x,y) \, \mathrm{d}y \right) \, \mathrm{d}x && \text{[Rev. Order of int.]} \\
&= \int_{\mathbb{R}} x f_X(x) \, \mathrm{d}x = \mathbb{E}[X] && \text{[Simplifying]}
\end{aligned}
$$

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Law of Iterated Expectations

- So, we've shown that $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$.

Law of Iterated Expectations

- So, we've shown that $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$.
- Indeed, this is not a coincidence!

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Law of Iterated Expectations

- So, we've shown that $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$.
- Indeed, this is not a coincidence!

Theorem (Law of Iterated Expectations)

Given random variables X and Y, we have

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

provided these quantities exist.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Law of Iterated Expectations

- So, we've shown that $\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$.
- Indeed, this is not a coincidence!

Theorem (Law of Iterated Expectations)

Given random variables X and Y, we have

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

provided these quantities exist.

- We proved the continuous case above; I'll ask you to prove the discrete case later.

LIE and LOTUS

Theorem

Given random variables X and Y, we have

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

provided these quantities exist.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

LIE and LOTUS

Theorem

Given random variables X and Y, we have

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

provided these quantities exist.

- For example, $\mathbb{E}[X^2] = \mathbb{E}[\mathbb{E}[X^2 \mid Y]]$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Clicker Question!

Clicker Question 2

Let $(X \mid Y = y) \sim \text{Bin}(y, 0.25)$ and $Y \sim \text{Pois}(2)$. What is $\mathbb{E}[X]$?

(A) 0.00
(B) 0.25
(C) 0.50
(D) 2.00
(E) None of the above

UC SANTA BARBARA
Department of Statistics and Applied Probability

Law of Total Variance

- Since $(X \mid Y)$ is a random variable, it makes sense to ask what its variance is. Thankfully, we have a formula for that:

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Law of Total Variance

- Since $(X \mid Y)$ is a random variable, it makes sense to ask what its variance is. Thankfully, we have a formula for that:

Theorem (Law of Total Variance)

Given random variables X and Y, we have

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y])$$

Law of Total Variance

- Since $(X \mid Y)$ is a random variable, it makes sense to ask what its variance is. Thankfully, we have a formula for that:

Theorem (Law of Total Variance)

Given random variables X and Y, we have

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X \mid Y)] + \mathrm{Var}(\mathbb{E}[X \mid Y])$$

- As an exercise, return to the clicker question from a few slides ago and try to compute $\mathrm{Var}(X)$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Example Revisited

Example

Suppose I roll a fair six-sided die. Then, whatever number the die lands on, I flip that many fair coins. Let X denote the number of heads. Compute $\mathbb{E}[X]$ and $\text{Var}(X)$.

UC SANTA BARBARA
Department of Statistics and Applied Probability

One More Formula [NOT COVERED]

Definition (Expectation Conditional on an Event)

Given a random variable X and an event A with $\mathbb{P}(A) \neq 0$,

$$\mathbb{E}[X \mid A] = \frac{\mathbb{E}[X \cdot \mathbb{1}_A]}{\mathbb{P}(A)}$$

Example

The time that Joe spends talking on the phone is exponentially distributed with mean 5 minutes. What is the expected length of his phone call if he talks for more than 2 minutes?

The Gamma Distribution

Gamma Function

- Hopefully, everyone is familiar with the distributions you learned in 120A (e.g. normal, exponential, binomial, etc.)

Gamma Function

- Hopefully, everyone is familiar with the distributions you learned in 120A (e.g. normal, exponential, binomial, etc.)
- There is one distribution that is not always covered in 120A, that ends up being incredibly useful in statistical concepts: the **Gamma Distribution**.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Gamma Function

- Hopefully, everyone is familiar with the distributions you learned in 120A (e.g. normal, exponential, binomial, etc.)
- There is one distribution that is not always covered in 120A, that ends up being incredibly useful in statistical concepts: the **Gamma Distribution**.
- First, I'll introduce the **Gamma function**, which arises in mathematics often:

$$\Gamma(r) := \int_0^\infty x^{r-1} e^{-x} \, dx$$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Gamma Function

- Hopefully, everyone is familiar with the distributions you learned in 120A (e.g. normal, exponential, binomial, etc.)
- There is one distribution that is not always covered in 120A, that ends up being incredibly useful in statistical concepts: the **Gamma Distribution**.
- First, I'll introduce the **Gamma function**, which arises in mathematics often:

$$\Gamma(r) := \int_0^\infty x^{r-1} e^{-x} \, \mathrm{d}x$$

  - $\Gamma(0) := 1$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Gamma Function

- Hopefully, everyone is familiar with the distributions you learned in 120A (e.g. normal, exponential, binomial, etc.)
- There is one distribution that is not always covered in 120A, that ends up being incredibly useful in statistical concepts: the **Gamma Distribution**.
- First, I'll introduce the **Gamma function**, which arises in mathematics often:

$$\Gamma(r) := \int_0^\infty x^{r-1} e^{-x} \, dx$$

  - $\Gamma(0) := 1$
  - recursive property: $\Gamma(r) = (r-1)\Gamma(r-1)$

UC SANTA BARBARA
Department of Statistics and Applied Probability

Gamma Function

- Hopefully, everyone is familiar with the distributions you learned in 120A (e.g. normal, exponential, binomial, etc.)
- There is one distribution that is not always covered in 120A, that ends up being incredibly useful in statistical concepts: the **Gamma Distribution**.
- First, I'll introduce the **Gamma function**, which arises in mathematics often:

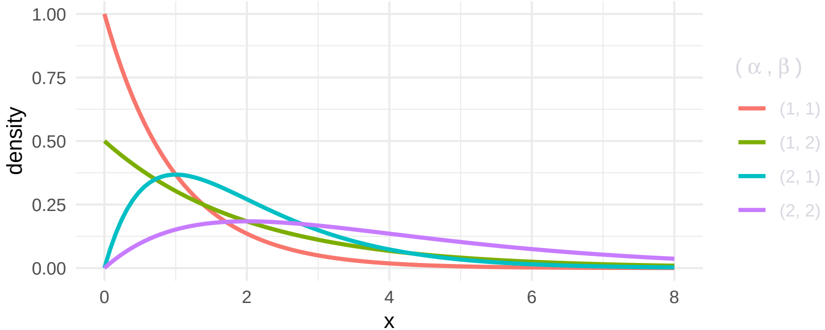
$$\Gamma(r) := \int_0^\infty x^{r-1} e^{-x} \, dx$$

  - $\Gamma(0) := 1$
  - recursive property: $\Gamma(r) = (r-1)\Gamma(r-1)$
  - $\Gamma(n) = (n-1)!$ for $n \in \mathbb{N}$.

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Gamma Distribution

- **Notation:** $X \sim \text{Gamma}(\alpha, \beta)$
- **PDF:** $f_X(x) = \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \cdot \mathbb{1}_{\{x \geq 0\}}$
- **Expectation and Variance:** $\mathbb{E}[X] = \alpha\beta; \quad \text{Var}(X) = \alpha\beta^2$



UC SANTA BARBARA
Department of Statistics
and Applied Probability

Special Cases of the Gamma Distribution

- Note that the Gamma$(1, \beta)$ distribution is equivalent to the Exp(β) distribution.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Special Cases of the Gamma Distribution

- Note that the Gamma$(1, \beta)$ distribution is equivalent to the Exp(β) distribution.
  - **PLEASE NOTE:** in PSTAT 120B, we adopt the convention that the parameter of the exponential distribution is its mean. In other words, saying $X \sim \text{Exp}(\beta)$ means $\mathbb{E}[X] = \beta$ and X has a density given by $f_X(x) = (1/\beta)e^{-x/\beta} \cdot \mathbb{1}_{\{x \geq 0\}}$.

Special Cases of the Gamma Distribution

- Note that the Gamma$(1, \beta)$ distribution is equivalent to the Exp(β) distribution.
  - **PLEASE NOTE:** in PSTAT 120B, we adopt the convention that the parameter of the exponential distribution is its mean. In other words, saying $X \sim \text{Exp}(\beta)$ means $\mathbb{E}[X] = \beta$ and X has a density given by $f_X(x) = (1/\beta)e^{-x/\beta} \cdot \mathbb{1}_{\{x \geq 0\}}$.

- Another special case of the Gamma distribution is the so-called χ^2 **distribution** (pronounced "kai-squared").

UC SANTA BARBARA
Department of Statistics
and Applied Probability

Special Cases of the Gamma Distribution

- Note that the Gamma$(1, \beta)$ distribution is equivalent to the Exp(β) distribution.
  - **PLEASE NOTE:** in PSTAT 120B, we adopt the convention that the parameter of the exponential distribution is its mean. In other words, saying $X \sim \text{Exp}(\beta)$ means $\mathbb{E}[X] = \beta$ and X has a density given by $f_X(x) = (1/\beta)e^{-x/\beta} \cdot \mathbb{1}_{\{x \geq 0\}}$.

- Another special case of the Gamma distribution is the so-called χ^2 **distribution** (pronounced "kai-squared").
  - Specifically, the χ^2_ν distribution is equivalent to the $\mathrm{Gamma}(\nu/2\ ,\ 2)$ distribution.

UC SANTA BARBARA
Department of Statistics and Applied Probability

Special Cases of the Gamma Distribution

- Note that the Gamma$(1, \beta)$ distribution is equivalent to the Exp(β) distribution.
  - **PLEASE NOTE:** in PSTAT 120B, we adopt the convention that the parameter of the exponential distribution is its mean. In other words, saying $X \sim \text{Exp}(\beta)$ means $\mathbb{E}[X] = \beta$ and X has a density given by $f_X(x) = (1/\beta)e^{-x/\beta} \cdot \mathbb{1}_{\{x \geq 0\}}$.

- Another special case of the Gamma distribution is the so-called χ^2 **distribution** (pronounced "kai-squared").
  - Specifically, the χ^2_ν distribution is equivalent to the $\mathrm{Gamma}(\nu/2\ ,\ 2)$ distribution.
  - Question for you: what are the expectation and variance of the χ^2_ν distribution?

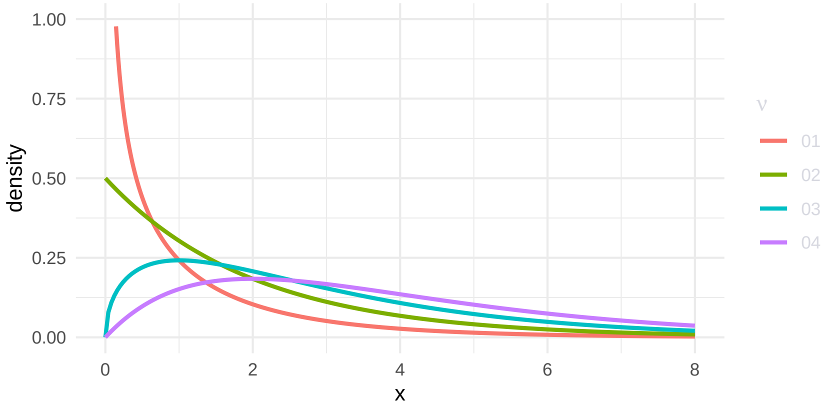
UC SANTA BARBARA
Department of Statistics and Applied Probability

Special Cases of the Gamma Distribution

- Note that the Gamma$(1, \beta)$ distribution is equivalent to the Exp(β) distribution.
    - **PLEASE NOTE:** in PSTAT 120B, we adopt the convention that the parameter of the exponential distribution is its mean. In other words, saying $X \sim \text{Exp}(\beta)$ means $\mathbb{E}[X] = \beta$ and X has a density given by $f_X(x) = (1/\beta)e^{-x/\beta} \cdot \mathbb{1}_{\{x \geq 0\}}$.

- Another special case of the Gamma distribution is the so-called χ^2 **distribution** (pronounced "kai-squared").
    - Specifically, the χ^2_ν distribution is equivalent to the $\mathrm{Gamma}(\nu/2\ ,\ 2)$ distribution.
    - Question for you: what are the expectation and variance of the χ^2_ν distribution?
    - Also, while we're at it, let's derive the density of the χ^2_ν distribution on the board.

χ^2_ν Distribution



UC SANTA BARBARA
Department of Statistics
and Applied Probability

More to Come

- You'll talk a bit more about the Gamma distribution during section this week.

More to Come

- You'll talk a bit more about the Gamma distribution during section this week.
- You'll also show that if $X \sim \text{Gamma}(\alpha, \beta)$, then

$$M_X(t) = \begin{cases} (1 - \beta t)^{-\alpha} & \text{if } t < 1/\beta \\ \infty & \text{otherwise} \end{cases}$$

which will, in turn, allow you to derive the MGF of the χ^2_ν distribution.

UC SANTA BARBARA
Department of Statistics and Applied Probability