# 14: Conditional Distributions and Expectation

## PSTAT 120A: Summer 2022

Ethan P. Marzban

July 26, 2022

University of California, Santa Barbara

# Where We've Been

- Axioms of Probability, Probability Spaces, Counting
- Conditional Probabilities, independence, etc.
- Basics of Random Variables (classification, p.m.f., c.m.f., moments)
- Discrete Distributions
- Continuous Distributions
- Transformations of Random Variables
- Double Integrals
- Random Vectors and the basics of multivariate probability
- Independence of random variables, and covariance/correlation
- Sums of Random Variables; Indicators
- Moment Generating Functions
- Tail Bounds
- Limit Theorems

# Leadup

- 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.
- Now, $X$ *sounds* binomial. But it isn't- what's the problem? that's right; the binomial distribution requires a *fixed* number of Bernoulli trials.
- That is to say; information about $N$ is needed in order to ascertain more information about $X$.
- This will lead us into our discussion on conditional distributions.

# Conditioning on an Event; Discrete Case

# Leadup

- Recall the Law of Total Probability: given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a partition $\{B_i\}_{i=1}^n$ of $\Omega$, then for any event $A \in \mathcal{F}$ we have

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$$

- Suppose we replaced the event $A$ with the more specific event $\{X = k\}$, for some random variable $X$ and some fixed value $k$. (Remember that $\{X = k\} := \{\omega \in \Omega : X(\omega) = k\}$ is in fact an event!):

$$\mathbb{P}(X = k) = \sum_{i=1}^{n} \mathbb{P}(X = k \mid B_i)\mathbb{P}(B_i)$$

- This necessitates our first definition:

# Conditional Probability Mass Function

**Definition: Conditional Probability Mass Function**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, an event $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, and a random variable $X$, we define the **conditional probability mass function** (conditional p.m.f.) $p_{X|B}(k)$ to be

$$p_{X|B}(k) = \frac{\mathbb{P}(\{X = k\} \cap B)}{\mathbb{P}(B)}$$

**Theorem: Conditional Probability Mass Function**

The function $p_{X|B}(k)$ as defined in the theorem above is a valid p.m.f.

**Proof.**

- Nonnegativity is trivial.
- To check that $p_{X|B}(k)$ sums to unity, we compute

$$\begin{aligned}\sum_k p_{X|B}(k) &= \sum_k \frac{\mathbb{P}(\{X=k\}\cap B)}{\mathbb{P}(B)} \\ &= \frac{1}{\mathbb{P}(B)} \sum_k \mathbb{P}(\{X=k\}\cap B) \\ &= \frac{1}{\mathbb{P}(B)} \cdot \mathbb{P}(B) = 1 \checkmark\end{aligned}$$

- By the way, what did we use to go from the second-to-last equation to the last one?

□

# Marginal P.M.F.'s from Conditional P.M.F.'s

**Theorem**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $X$, and a partition $\{B_i\}_{i=1}^n$ of $\Omega$ such that $\mathbb{P}(B_i) > 0$, then

$$p_X(k) = \sum_{i=1}^{n} p_{X|B_i}(k)\mathbb{P}(B_i)$$

# Example

Suppose that I am a shopowner, and I know that the number of customers arriving at my shop follows a Poisson distribution. However, I also know that the rate at which customers arrive at my store is much lower on rainy days as opposed to dry days. Specifically, on rainy days customers arrive at a rate $\lambda_r$ per day, and on dry days customers arrive at a rate $\lambda_d$ per day. Further suppose that there is a $p = 10\%$ chance of rain tomorrow..

If $X$ denotes the number of customers that will arrive at my store tomorrow, what is the p.m.f. (probability mass function) of $X$?

- Recall that $p_{X|B}(k)$ is a probability mass function.
- Further recall that we can consider expected values as weighted averages of p.m.f. values. This motivates our next definition:

**Definition**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $X$, an event $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$, we define the **conditional expectation** of $X$, given $B$, to be

$$\mathbb{E}[X \mid B] = \sum_k k \cdot p_{X|B}(k) = \sum_k k \cdot \mathbb{P}(X = k \mid B)$$

- In this way, we can think of $(X \mid B)$ as a random variable in itself, whose expectation is given by the formula above.

# Marginal Expectations from Conditional P.M.F.'s

**Theorem**

Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a random variable $X$, and a partition $\{B_i\}_{i=1}^n$ of $\Omega$ such that $\mathbb{P}(B_i) > 0$, then

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid B_i] \cdot \mathbb{P}(B_i)$$

**Proof.**

- We compute

$$\begin{aligned}
\mathbb{E}[X] &= \sum_k k \cdot p_X(k) \\
&= \sum_k \sum_{i=1}^{n} k \cdot \mathbb{P}(X = k \mid B_i) \cdot \mathbb{P}(B_i) \\
&= \sum_{i=1}^{n} \sum_k k \cdot \mathbb{P}(X = k \mid B_i) \cdot \cdot \mathbb{P}(B_i) = \sum_{i=1}^{n} \left( \sum_k k \cdot \mathbb{P}(X = k \mid B_i) \right) \cdot \mathbb{P}(B_i) \\
&= \sum_{i=1}^{n} \mathbb{E}[X \mid B_i] \cdot \mathbb{P}(B_i)
\end{aligned}$$

# Example (Revisited)

Returning to the shop example from above; what is $\mathbb{E}[X]$, the expected number of customers that will arrive at my store tomorrow?

# Conditioning on a Random Variable; Discrete Case

## Leadup

- We have now discussed the notion of conditioning a random variable on an event, where one of our key definitions was

$$p_{X|B}(k) := \mathbb{P}(X = k \mid B) = \frac{\mathbb{P}(\{X = k\} \cap B)}{\mathbb{P}(B)}$$

- What happens if we replace the event $B$ with... a random variable?
- Before we go that far, let's consider replacing the event $B$ with another *event*, involving a random variable.

**Definition: Conditional P.M.F.**

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two random variables $X$ and $Y$. Then the **conditional probability mass function** of $X$ given $Y = y$ is the following bivariate function:

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

# Conditional Expectation

- We can analogously extend our definition of conditional expectation:

**Definition: Conditional Expectation**

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two random variables $X$ and $Y$. Then the **conditional expectation** of $X$ given $Y = y$ is the quantity

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

where we assume, implicitly, that $y$ is such that $\mathbb{P}(Y = y) > 0$.

- We also get a conditional form of the LOTUS:

**Theorem: Conditional LOTUS**

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and two random variables $X$ and $Y$. Then

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

# Marginal from Conditional

**Theorem**

Given two random variables $X$ and $Y$ we have

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

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

- Of course, the sums above are implicitly ranging over the values of $y$ for which $p_Y(y) > 0$.
- Also, though I have omitted explicit mention of the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ [which I shall continue to do so, for the sake of convenience], we always assume that all random variables are defined on an appropriate probability space.

# Example

The joint probability mass function of the random variables $(X, Y)$ is given by the following table:

| | | $Y=0$ | $Y=1$ |
|---|---|---|---|
| $X$ | 0 | $3/10$ | $2/10$ |
| | 1 | $1/10$ | $4/10$ |

(a) Find $p_{X|Y}(x \mid 0)$

(b) Find $p_{X|Y}(x \mid 1)$

(c) Find $p_{X|Y}(x \mid y)$.

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

# Example: Solutions

**Part (a)**

- $p_{X|Y}(x \mid 0) = \mathbb{P}(X = x \mid Y = 0) = \dfrac{\mathbb{P}(X = x,\ Y = 0)}{\mathbb{P}(Y = 0)} = \dfrac{p_{X,Y}(x, 0)}{4/10}$

  $$p_{X|Y}(0 \mid 0) = \frac{10}{4} \cdot p_{X,Y}(0, 0) = \frac{10}{4} \cdot \frac{3}{10} = \frac{3}{4}$$

  $$p_{X|Y}(1 \mid 0) = \frac{10}{4} \cdot p_{X,Y}(1, 0) = \frac{10}{4} \cdot \frac{1}{10} = \frac{1}{4}$$

- Note that $p_{X|Y}(x \mid 0)$ is in fact a valid probability mass function, as we expected!

**Part (b)**

- $p_{X|Y}(x \mid 1) = \mathbb{P}(X = x \mid Y = 1) = \dfrac{\mathbb{P}(X = x,\ Y = 1)}{\mathbb{P}(Y = 1)} = \dfrac{p_{X,Y}(x, 1)}{6/10}$

  $$p_{X|Y}(0 \mid 1) = \frac{10}{6} \cdot p_{X,Y}(0, 1) = \frac{10}{6} \cdot \frac{2}{10} = \frac{1}{3}$$
  $$p_{X|Y}(1 \mid 1) = \frac{10}{6} \cdot p_{X,Y}(1, 1) = \frac{10}{6} \cdot \frac{4}{10} = \frac{2}{3}$$

- Note that $p_{X|Y}(x \mid 1)$ is in fact a valid probability mass function, as we expected!

# Example: Solutions

**Part (c)**

- Since $p_Y(y) = 0$ for any $y$ that is neither 0 nor 1, we are actually done.
- In other words, for a fixed value of $y$, we have a p.m.f.; said differently, each value of $y$ gives rise to a completely separate conditional mass function.

# Example: Solutions

**Part (c)**

$$\begin{aligned}\mathbb{E}[X \mid Y = 0] &= \sum_{k=0}^{1} k \cdot p_{X|Y}(k, 0) \\ &= (0)p_{X|Y}(0,0) + (1)p_{X|Y}(1,0) = (0)\left(\frac{3}{4}\right) + (1)\left(\frac{1}{4}\right) = \boxed{\frac{1}{4}} \\ \mathbb{E}[X \mid Y = 1] &= \sum_{k=0}^{1} k \cdot p_{X|Y}(k, 1) \\ &= (0)p_{X|Y}(0,1) + (1)p_{X|Y}(1,1) = (0)\left(\frac{1}{3}\right) + (1)\left(\frac{2}{3}\right) = \boxed{\frac{2}{3}}\end{aligned}$$

# Decompress

- I'd like to use the previous example to illustrate some points.
- Firstly: for a fixed value of $y$, $p_{X|Y}(x \mid y)$ is a valid probability mass function.
- $\mathbb{E}[X \mid Y = y]$ **is a function of $y$**. For example, in our example above,

$$\mathbb{E}[X \mid Y = y] = \begin{cases} 1/4 & \text{if } y = 0 \\ 2/3 & \text{if } y = 1 \\ \text{undefined} & \text{otherwise} \end{cases}$$

# The Continuous Realm

- Now that we've talked about conditional distributions/expectations in the discrete setting, let's consider what happens when we dip into the continuous realm.
- Firstly, we cannot directly adapt our definition of $p_{X,Y}(k \mid y)$ into the continuous setting because $\mathbb{P}(Y = y) = 0$! However, we can take the same *idea* and translate it:

**Definition**

Let $(X, Y)$ be a continuous bivariate random vector with joint p.d.f. $f_{X,Y}(x, y)$. Then the **conditional density function** of $X$, given $Y = y$, is defined by

$$f_{X|Y}(x \mid y) := \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad \text{for } y \text{ s.t. } f_Y(y) > 0$$

- Some comments:
- Recall that in the univariate setting, the value $f_X(x)$ of a density has nothing inherently to do with probabilities. Similarly, the value $f_{X|Y}(x \mid y)$ **does not represent a probability in itself!**
- Also, note that $f_{X|Y}(x \mid y)$ is only defined for $y$ s.t. $f_Y(y) > 0$. If we have a point $y'$ such that $f_Y(y') = 0$, then $f_{X|Y}(x \mid y')$ is **undefined** (much like writing $\mathbb{P}(A \mid B)$ for an event $B$ such that $\mathbb{P}(B) = 0$).
- Additionally, for a fixed $y$, $f_{X|Y}(\cdot \mid y)$ is a valid p.d.f.; nonnegativity is trivial, and

$$\begin{aligned}\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)\ \mathrm{d}x &= \int_{-\infty}^{\infty} \frac{f_{X,Y}(x, y)}{f_Y(y)}\ \mathrm{d}x \\ &= \frac{1}{f_Y(y)} \cdot \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ \mathrm{d}x = \frac{1}{f_Y(y)} \cdot f_Y(y) = 1\ \checkmark\end{aligned}$$

# Using Conditional Densities

- Remember how we integrated densities to get probabilities?

**Definition**

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

- Also, we have

**Definition**

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

# Using Conditional Densities

**Theorem**

Given a continuous bivariate random vector $(X, Y)$,

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

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

# 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 $Y$. Use this to compute $\mathbb{E}[Y]$.

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

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

(d) Compute $\mathbb{E}[X \mid Y = 1]$

# Conditional Expectation

- Recall that $\mathbb{E}[X \mid Y = y]$ is in fact a function of $y$. Let us call this function $g(y)$; i.e. $g(y) := \mathbb{E}[X \mid Y = y]$.
- We also know that for any function $h : \mathbb{R} \to \mathbb{R}$ and a random variable $Y$ the quantity $h(Y)$ is a random variable.
- What happens when we take $h$ to be $g$ in the first point; i.e. what happens when we consider the quantity $g(Y)$, where $g(y) := \mathbb{E}[X \mid Y = y]$?

**Definition: Conditional Expectation**

Given a bivariate random vector $(X, Y)$ [discrete or continuous], we define the **conditional expectation** of $X$ given $Y$, denoted $\mathbb{E}[X \mid Y]$, to be $\mathbb{E}[X \mid Y] := g(Y)$ where $g(y) = \mathbb{E}[X \mid Y = y]$.

- Algorithmically, here is now we compute $\mathbb{E}[X \mid Y]$:
  (1) First fix a $y$ [such that $f_Y(y) > 0$ if $(X, Y)$ is continuous or $p_Y(y) > 0$ is $(X, Y)$ is discrete], and compute $g(y) := \mathbb{E}[X \mid Y = y]$ as outlined in the previous sections.
  (2) Then, wherever you have a lowercase $y$ in $g(y)$, replace it with a capital $Y$.

## Example

As an example, let's return to the example where $(X, Y)$ was a discrete bivariate random vector with joint p.m.f.

| | | $Y$ = 0 | $Y$ = 1 |
|---|---|---|---|
| $X$ | 0 | $3/10$ | $2/10$ |
| | 1 | $1/10$ | $4/10$ |

- We saw that $\mathbb{E}[X \mid Y = y] = \begin{cases} 1/4 & \text{if } y = 0 \\ 2/3 & \text{if } y = 1 \\ \text{undefined} & \text{otherwise} \end{cases}$
- Then, $\mathbb{E}[X \mid Y] = \begin{cases} 1/4 & \text{on the event } \{Y = 0\} \\ 2/3 & \text{on the event } \{Y = 1\} \end{cases}$
- Since $\mathbb{P}(Y = 0) = 4/10$ and $\mathbb{P}(Y = 1) = 2/3$, we can write the p.m.f. of $\mathbb{E}[X \mid Y]$ as

$$\mathbb{P}(\mathbb{E}[X \mid Y] = k) = \begin{cases} 4/10 & \text{if } k = 1/4 \\ 6/10 & \text{if } k = 2/3 \\ 0 & \text{otherwise} \end{cases}$$

# Looking Ahead

- Perhaps $\mathbb{E}[X \mid Y]$ seems like an incredibly abstract quantity that is very difficult to interpret.
- However, it will turn out that $\mathbb{E}[X \mid Y]$ has some very incredibly useful properties. Some uses include:
  - Shortening expectation computations
  - Providing a best predictor of $X$, given $Y$
  - Forming the backbone of much of the theory behind Stochastic Processes (PSTAT 160A/B and 174).
- Unfortunately, we do not have time to investigate many of these in PSTAT 120A. So, for now, I hope you will take my word that conditional expectations are very useful, and that you will remember me fondly when you encounter them again :]

# Expectation Computations

- Thankfully, we *do* have the background (and time!) to investigate the first "application" above. I first present a theorem:

**Theorem: Law of Iterated Expectations (a.k.a. Tower Property)**

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

- The inner expectation is taken w.r.t. $(X \mid Y)$ [which, remember, is a random variable!] and the outer expectation is taken w.r.t. $Y$.

# Law of Iterated Expectations

**Proof.**

- I provide the proof in the continuous case, and will ask you to investigate the proof for the discrete case on your own.
- Let $g(y) := \mathbb{E}[X \mid Y = y]$; then

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}[X \mid Y]\right] = \mathbb{E}[v(Y)] = \int_{-\infty}^{\infty} v(y) f_Y(y)\, \mathrm{d}y \\ = \int_{-\infty}^{\infty} \mathbb{E}[X \mid Y = y] \cdot f_Y(y)\, \mathrm{d}y = \mathbb{E}[X]\end{aligned}$$

□

# Law of Total Variance

**Theorem: Law of Total Variance**

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

# Returning to the Beginning

As an example, let's return to the very first motivating example I presented: Suppose I roll a fair six-sided die. Then, whatever number the die lands on, I flip that many $p-$ coins. Let $X$ denote the number of heads.

- Let $N$ denote the number on which the die lands. Then $N \sim \mathrm{DiscUnif}\{1, \cdots, 6\}$.
- Additionally, $(X \mid N = n) \sim \mathrm{Bin}(n, p)$
- Therefore $(X \mid N) \sim \mathrm{Bin}(N, p)$ and $\mathbb{E}[X \mid N] = Np$. (Note that this is a random variable!)
- By the Law of Iterated Expectations:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid N]] = \mathbb{E}[Np] = p \cdot \mathbb{E}[N] = p \cdot \frac{1+6}{2} = \boxed{\frac{7p}{2}}$$

- By the Law of Total Variance:

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{Var}(\mathbb{E}[X \mid N]) + \mathbb{E}[\mathrm{Var}(X \mid N)] \\ &= \mathrm{Var}(Np) + \mathbb{E}[Np(1-p)] = p^2 \mathrm{Var}(N) + p(1-p)\mathbb{E}[N] = \boxed{p^2 \cdot \frac{35}{12} + p(1-p) \cdot \frac{7}{2}} \end{aligned}$$

- As an aside, it turns out that the marginal p.m.f. $p_X(k)$ of $X$ is quite difficult to find!