THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics
Lecture 17: Continuous random variables: conditional PDF

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

Roadmap

- Two random variables: joint distributions
  - Joint pdf
  - Joint pdf to a single pdf: Marginalization
  - Conditional pdf
    - Conditioning on an event
    - Conditioning on a continuous r.v
    - Total probability rule for continuous r.v's
    - Bayes theorem for continuous r.v's
    - Conditional expectation and total expectation theorem
  - Independence
- More than two random variables.

Conditional PDFs–conditioning on an event

- For *discrete* random variables, we looked at marginal PMFs $p_X(X)$, conditional PMFs $p_{X|Y}(x|y)$, and joint PMFs $p_{X,Y}(x,y)$.
- These corresponded to the probability of an event, $P(A)$, the conditional probability of an event given some other event, $P(A|B)$, and probability of the intersection of two events, $P(A \cap B)$.
- We've looked at marginal PDFs, $f_X(x)$ and joint PDFs, $f_{X,Y}(x,y)$.
- These don't directly give us probabilities of events, but we can use them to calculate such probabilities by integration.
- We can also look at conditional PDFs! These allow us to calculate the probability of events given extra information.

Conditional PDFs

- Recall, the PDF of a continuous random variable X is the non-negative function $f_X(x)$ that satisfies

$$P(X \in B) = \int_B f_X(x)dx$$

  for any subset B of the real line.
- Let A be some event with $P(A) > 0$
- The **conditional PDF** of X, given A, is the non-negative function $f_{X|A}$ that satisfies

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

  for any subset B of the real line.
- If B is the entire line, then we have

$$\int_{-\infty}^{\infty} f_{X|A}(x)dx = 1$$

- So, $f_{X|A}(x)$ is a valid PDF.

Conditional PDFs

- The event we are conditioning on can also correspond to a range of values of our continuous random variable.
- **Definition-**

$$f_{X|\{X\in A\}}(x) = \begin{cases} \frac{f_X(x)}{P(X\in A)} & \text{if } X \in A \\ 0 & \text{otherwise.} \end{cases}$$

- In this case, we can write the conditional probability as

$$\begin{aligned} \int_B f_{X|A}(x)dx &= \int_B \frac{f_X(x)1(x\in A)}{P(X\in A)}dx \\ &= \frac{\int_{A\cap B} f_X(x)dx}{P(X\in A)} = \frac{P(\{X\in A\}\cap\{X\in B\})}{P(X\in A)} \\ &= P(X\in B|X\in A) \end{aligned}$$

- This is a valid PDF–non-negative and integrates to one. Check?

Conditioning: memoryless property of the exponential

- $X \sim Exp(\lambda)$
- $f_X(x) = \lambda e^{-\lambda x}$ when $x \geq 0$, and zero otherwise.
- $P(X > s + t | X > s) =?$

Conditioning: memoryless property of the exponential

- $X \sim Exp(\lambda)$
- $f_X(x) = \lambda e^{-\lambda x}$ when $x \geq 0$, and zero otherwise.
- $P(X > s + t | X > s) =?$
- Remember the exponential? $F_X(x) = 1 - e^{-\lambda x}$.
- $$\begin{aligned} P(X > s + t | X > s) &= \frac{P(X > s + t, X > s)}{P(X > s)} \\ &= \frac{P(X > s + t)}{P(X > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t} = P(X > t) \end{aligned}$$

Conditioning: memoryless property of the exponential

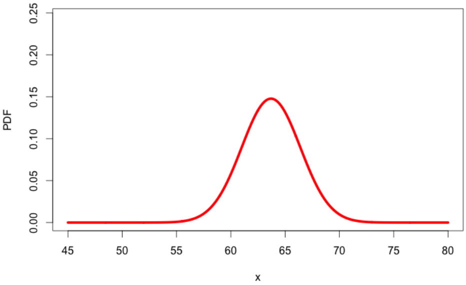
- $X \sim Exp(\lambda)$
- $f_{X|X>s}(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}}{P(X>s)} = \lambda e^{\lambda(x-s)} & \text{If } x > s \\ 0 & \text{Otherwise} \end{cases}$
- $P(X > s + t | X > s) = ?$

Conditioning: memoryless property of the exponential

- $X \sim Exp(\lambda)$
- $f_{X|X>s}(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}}{P(X>s)} = \lambda e^{\lambda(x-s)} & \text{If } x > s \\ 0 & \text{Otherwise} \end{cases}$
- $P(X > s + t | X > s) = ?$
- Remember the exponential? $F_X(x) = 1 - e^{-\lambda x}$.
- $$\begin{aligned} P(X > s+t|X>s) &= \int_{s+t}^{\infty} f_{X|X>s}(x)dx = \lambda \int_{s+t}^{\infty} e^{-\lambda(x-s)} dx \\ &= \lambda \int_{t}^{\infty} e^{-\lambda u} du = e^{-\lambda t} \end{aligned}$$

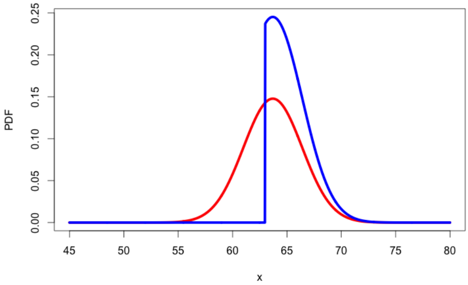
Conditional PDFs: Example

- The height X of a randomly picked american woman can be modeled by $X \sim N(63.7, 2.7^2)$
- Whats the conditional PDF given that the randomly picked woman is at least 63 inches tall?
- The PDF of heights (X) is shown in red.



Conditional PDFs: Example

- The height X of a randomly picked american woman can be modeled by $X \sim N(63.7, 2.7^2)$
- Whats the conditional PDF given that the randomly picked woman is at least 63 inches tall?
- The PDF of heights (X) is shown in red.
- The conditional PDF given $X > 63$, shown in blue, is the same shape for $X > 63$... but scaled up to integrate to one.



Conditioning on a different random variable

- So far, we conditioned X on an arbitrary event A, or on a range of values of X.

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

- We can also condition on the outcome of a second random variable Y.
- We know we could condition on a range of outcomes of Y, by replacing the arbitrary event A with the event $\{Y \in A\}$

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

- What about conditioning on a specific value of $Y = y$?
- Even though any outcome $Y = y$ has $P(Y = y) = 0$, we know that *some* value has to happen.
  - Pick some number, say 0.6777, now generate 100 $N(0, 1)$ random variables. I will bet a 100$ that you won't see that number.
  - But when you simulate from the standard normal, you will get a 100 different values, right?

Conditioning on a different random variable

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

  provided $f_Y(y) > 0$.

- What does this mean?

- $$\begin{aligned} f_{X|Y}(x|y)dx &= \frac{f(x,y)dxdy}{f(y)dy} \\ &= \frac{P(x \le X \le x+dx, y \le Y \le y+dy)}{P(y \le Y \le y+dy)} \\ &= P(x \le X \le x+dx | y \le Y \le y+dy) \end{aligned}$$

Multiplication rule: Calculating the joint PDF

- We can use the same relationship, $f_{X|Y}(x|y) = \dfrac{f_{X,Y}(x,y)}{f_Y(y)}$, to calculate the joint PDF from the conditional and the marginal PDF.
- i.e., $f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$.
- This is a PDF version of our multiplication rule.
- We can extend it to more than 2 random variables:

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

Lets remember all the rules

We've now got a lot of ways to go between our various PDFs!

- If we know $f_{X,Y}(x, y)$, we can get $f_X(x)$
  - How?
- If we know $f_{X,Y}(x, y)$ and $f_Y(y)$, if $f_Y(y) > 0$ we can get $f_{X|Y}(x|y)$
- If we know $f_X(x)$ and $f_{Y|X}(y|x)$, we can get $f_{X,Y}(x, y)$

Lets remember all the rules

We've now got a lot of ways to go between our various PDFs!

- If we know $f_{X,Y}(x,y)$, we can get $f_X(x)$
  - How? marginalization! $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
- If we know $f_{X,Y}(x,y)$ and $f_Y(y)$, if $f_Y(y) > 0$ we can get $f_{X|Y}(x|y)$
- If we know $f_X(x)$ and $f_{Y|X}(y|x)$, we can get $f_{X,Y}(x,y)$

Lets remember all the rules

We've now got a lot of ways to go between our various PDFs!

- If we know $f_{X,Y}(x, y)$, we can get $f_X(x)$
  - How? marginalization! $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$
- If we know $f_{X,Y}(x, y)$ and $f_Y(y)$, if $f_Y(y) > 0$ we can get $f_{X|Y}(x|y)$
  - How?
- If we know $f_X(x)$ and $f_{Y|X}(y|x)$, we can get $f_{X,Y}(x, y)$

Lets remember all the rules

We've now got a lot of ways to go between our various PDFs!

- If we know $f_{X,Y}(x,y)$, we can get $f_X(x)$
  - How? marginalization! $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
- If we know $f_{X,Y}(x,y)$ and $f_Y(y)$, if $f_Y(y) > 0$ we can get $f_{X|Y}(x|y)$
  - How? $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
- If we know $f_X(x)$ and $f_{Y|X}(y|x)$, we can get $f_{X,Y}(x,y)$

Lets remember all the rules

We've now got a lot of ways to go between our various PDFs!

- If we know $f_{X,Y}(x, y)$, we can get $f_X(x)$
  - How? marginalization! $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$
- If we know $f_{X,Y}(x, y)$ and $f_Y(y)$, if $f_Y(y) > 0$ we can get $f_{X|Y}(x|y)$
  - How? $f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$
- If we know $f_X(x)$ and $f_{Y|X}(y|x)$, we can get $f_{X,Y}(x, y)$
  - How?

Lets remember all the rules

We've now got a lot of ways to go between our various PDFs!

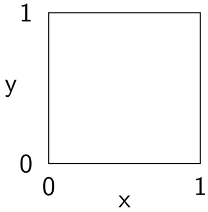
- If we know $f_{X,Y}(x, y)$, we can get $f_X(x)$
  - How? marginalization! $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$
- If we know $f_{X,Y}(x, y)$ and $f_Y(y)$, if $f_Y(y) > 0$ we can get $f_{X|Y}(x|y)$
  - How? $f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$
- If we know $f_X(x)$ and $f_{Y|X}(y|x)$, we can get $f_{X,Y}(x, y)$
  - How? multiplication rule! $f_{X,Y}(x, y) = f_X(x)f_{Y|X}(y|x)$

Example: Calculating the conditional PDF

- Let $f_{X,Y}(x,y) = \begin{cases} c & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$
- What is the conditional PDF of X given Y, $f_{X|Y}(x|y)$?

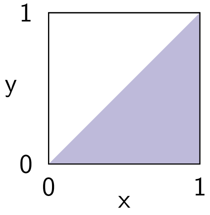
Example: Calculating the conditional PDF

- Let $f_{X,Y}(x,y) = \begin{cases} c & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$
- What is the conditional PDF of X given Y, $f_{X|Y}(x|y)$?
- First things first... what is c? Well, what does our joint PDF look like?



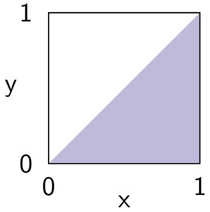
Example: Calculating the conditional PDF

- Let $f_{X,Y}(x,y) = \begin{cases} c & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$
- What is the conditional PDF of X given Y, $f_{X|Y}(x|y)$?
- First things first... what is c? Well, what does our joint PDF look like?



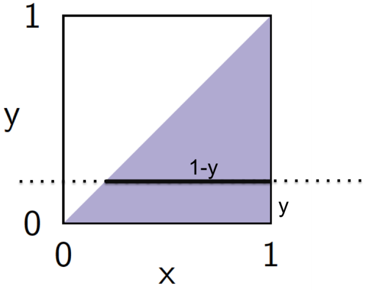
Example: Calculating the conditional PDF

- Let $f_{X,Y}(x,y) = \begin{cases} c & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$
- What is the conditional PDF of X given Y, $f_{X|Y}(x|y)$?
- First things first... what is c? Well, what does our joint PDF look like?



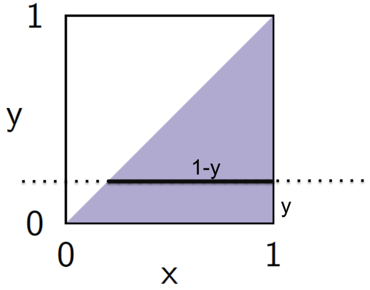
- The total area where $0 \leq x \leq 1$ and $0 \leq y \leq x$ is 0.5, so $c = 2$.
- What is the marginal PDF of Y, $f_Y(y)$?

Example: Calculating the conditional PDF



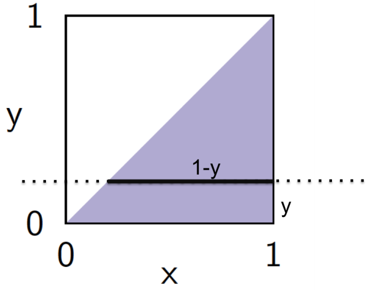
- To get the marginal PDF of Y, we take the joint PDF and marginalize out X.
- $f_Y(y) = \int_0^1 f_{X,Y}(x,y)dx = 2\int_0^1 \mathbf{1}_{0\leq x\leq 1, 0\leq y\leq x} dx$

Example: Calculating the conditional PDF



- To get the marginal PDF of Y, we take the joint PDF and marginalize out X.
- $f_Y(y) = \int_0^1 f_{X,Y}(x,y)dx = 2\int_0^1 \mathbf{1}_{0\leq x\leq 1, 0\leq y\leq x} dx$
  $= 2\int_{x=y}^1 dx = 2(1-y)$

Example: Calculating the conditional PDF



- To get the marginal PDF of Y, we take the joint PDF and marginalize out X.
- $f_Y(y) = \int_0^1 f_{X,Y}(x,y)dx = 2\int_0^1 \mathbf{1}_{0\le x\le 1, 0\le y\le x} dx$

$$= 2\int_{x=y}^1 dx = 2(1-y)$$

- So, the conditional PDF of X given $Y = y$ is

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \frac{1}{1-y} & \text{if } y \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Total probability theorem for continuous random variables

- We know that conditional probabilities must obey the total probability theorem.
- If $B_1, \ldots, B_n$ form a partition of Ω, such that $P(B_i) > 0$ for each i, then for any event A,

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

- In terms of discrete r.v's we have:

$$P(X = x) = \sum_{i} P(X = x|B_i) P(B_i)$$

- How about continuous r.v.'s? Replace $P(X = x|B_i)$ by conditional pdf.

$$f_X(x) = \sum_{i} f_{X|B_i}(x) P(B_i)$$

Bayes' law with continuous outcomes but discrete hidden causes

- Sometimes our hidden cause is inherently discrete.
  - e.g. I may be interested in whether I have flu or not – a binary choice.
  - My observation might be my temperature – a continuous random variable.
- We want $P(A|Y = y) =$ e.g. $P(\text{flu}|Y = 100)$
- Pretend Y is a discrete r.v.

$$P(A|Y = y) = \frac{P(Y = y|A)P(A)}{P(Y = y|A)P(A) + P(Y = y|A^c)P(A^c)}$$

All that changes for a continuous r.v. is:

$$P(A|Y = y) = \frac{f_{Y|A}(y)P(A)}{f_{Y|A}(y)P(A) + f_{Y|A^c}(y)P(A^c)}$$

Bayes' law with continuous outcomes but discrete hidden causes

- The probability that anyone has flu (event A) is 20%.
- Body temperature is Y.
- Without flu, Y is a normal random variable with $\mu = 98.6$ degrees and $\sigma = .5$.
- With flu, Y is a normal random variable with $\mu = 102$ and $\sigma = 2$.
- My temperature is 100. If A is the event "has flu" and Y is temp.

Bayes' law with continuous outcomes but discrete hidden causes

- The probability that anyone has flu (event A) is 20%.
- Body temperature is Y.
- Without flu, Y is a normal random variable with $\mu = 98.6$ degrees and $\sigma = .5$.
- With flu, Y is a normal random variable with $\mu = 102$ and $\sigma = 2$.
- My temperature is 100. If A is the event "has flu" and Y is temp.
- $$\boxed{\begin{aligned} f_{Y|A}(y) &= \frac{1}{\sqrt{2\pi \times 4}} \exp - \frac{(y-102)^2}{2 \times 4} \\ f_{Y|A^c}(y) &= \frac{1}{\sqrt{2\pi \times .25}} \exp - \frac{(y-98.6)^2}{2 \times .25} \end{aligned}}$$

Bayes' law with continuous outcomes but discrete hidden causes

- The probability that anyone has flu (event A) is 20%.
- Body temperature is Y.
- Without flu, Y is a normal random variable with $\mu = 98.6$ degrees and $\sigma = .5$.
- With flu, Y is a normal random variable with $\mu = 102$ and $\sigma = 2$.
- My temperature is 100. If A is the event "has flu" and Y is temp.
- $$f_{Y|A}(y) = \frac{1}{\sqrt{2\pi \times 4}} \exp - \frac{(y-102)^2}{2 \times 4}$$ $$f_{Y|A^c}(y) = \frac{1}{\sqrt{2\pi \times .25}} \exp - \frac{(y-98.6)^2}{2 \times .25}$$
- $$P(A|Y=y) = \frac{P(A) f_{Y|A}(y)}{f_Y(y)} = \frac{f_{Y|A}(y)P(A)}{f_{Y|A}(y)P(A) + f_{Y|A^c}(y)P(A^c)}$$ $$P(A|Y=100) = \frac{0.2 \frac{1}{2\sqrt{2\pi}} e^{-(100-102)^2/8}}{0.2 \frac{1}{2\sqrt{2\pi}} e^{-(100-102)^2/8} + 0.8 \frac{1}{0.5\sqrt{2\pi}} e^{-(100-98.6)^2/0.5}} = 0.65$$

Continuous Bayes' rule

- Discrete X, Y.
- $P(X = x|Y = y) = \dfrac{P(Y = y|X = x)P(X = x)}{\sum_x P(Y = y|X = x)P(X = x)}$
- What is $f_{X|Y}(x|y)$?

Continuous Bayes' rule

- Discrete X, Y.
- $P(X = x|Y = y) = \frac{P(Y = y|X = x)P(X = x)}{\sum_x P(Y = y|X = x)P(X = x)}$
- What is $f_{X|Y}(x|y)$?
- $f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x)dx}$

Conditional Expectation

- When we were looking at discrete random variables, we looked at **conditional expectations**.
- The conditional expectation, $E[X|A]$, of a random variable X given an event A is the value of X we expect to get out, on average, when A is true.
- We could calculate it by summing over all values x that X can take on, and scaling them by the conditional PMF $p_{X|A}(x) = P(X = x|A)$.

$$E[X|A] = \sum_x x p_{X|A}(x)$$

Conditional Expectation

- We can also look at the conditional expectation of a continuous random variable.
- If $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$, what do you think the conditional expectation of X given some event A looks like?

Conditional Expectation

- We can also look at the conditional expectation of a continuous random variable.
- If $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$, what do you think the conditional expectation of X given some event A looks like?
- $E[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x) dx$

Conditional Expectation

- We can also look at the conditional expectation of a continuous random variable.
- If $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$, what do you think the conditional expectation of X given some event A looks like?
- $E[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x) dx$
- How about the conditional expectation of some function $g(X)$ given some event A?

Conditional Expectation

- We can also look at the conditional expectation of a continuous random variable.
- If $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$, what do you think the conditional expectation of X given some event A looks like?
- $E[X|A] = \int_{-\infty}^{\infty} x f_{X|A}(x) dx$
- How about the conditional expectation of some function $g(X)$ given some event A?
- $E[g(X)|A] = \int_{-\infty}^{\infty} g(x) f_{X|A}(x) dx$

Total expectation theorem

- More generally, if $A_1, A_2, \ldots, A_n$ are a partition of Ω, we have a continuous version of the **total expectation theorem**:

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

- Or, if we are conditioning on specific values $Y = y$,

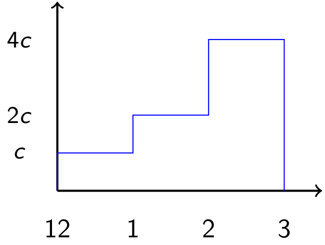
$$E[X] = \int_{-\infty}^{\infty} E[X|Y = y] f_Y(y) dy$$

Conditional expectation

- I am expecting an email, that will definitely arrive between midday and 3pm.
- Within a given hour (midday-1, 1-2, 2-3), each time is equally likely.
- It is twice as likely to arrive between 1 and 2 as it is to arrive between midday and 1.
- It is twice as likely to arrive between 2 and 3 as it is to arrive between 1 and 2.
- What does the PDF look like?

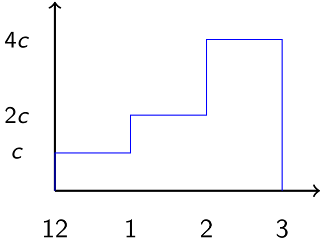
Conditional expectation

- I am expecting an email, that will definitely arrive between midday and 3pm.
- Within a given hour (midday-1, 1-2, 2-3), each time is equally likely.
- It is twice as likely to arrive between 1 and 2 as it is to arrive between midday and 1.
- It is twice as likely to arrive between 2 and 3 as it is to arrive between 1 and 2.
- What does the PDF look like?



Conditional expectation

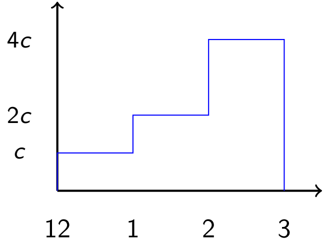
- I am expecting an email, that will definitely arrive between midday and 3pm.
- Within a given hour (midday-1, 1-2, 2-3), each time is equally likely.
- It is twice as likely to arrive between 1 and 2 as it is to arrive between midday and 1.
- It is twice as likely to arrive between 2 and 3 as it is to arrive between 1 and 2.
- What does the PDF look like?



- What is c?

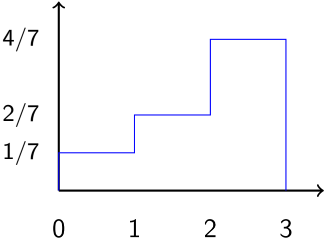
Conditional expectation

- I am expecting an email, that will definitely arrive between midday and 3pm.
- Within a given hour (midday-1, 1-2, 2-3), each time is equally likely.
- It is twice as likely to arrive between 1 and 2 as it is to arrive between midday and 1.
- It is twice as likely to arrive between 2 and 3 as it is to arrive between 1 and 2.
- What does the PDF look like?



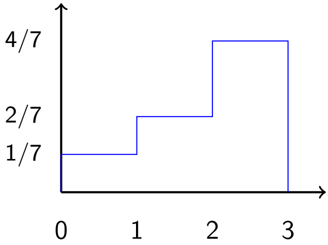
- What is c? $1/7$

Conditional expectation



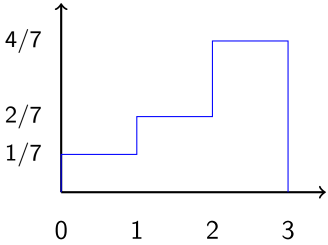
- I wait until 2pm. It still hasn't arrived. What is the expected value of the arrival time?
- What is the expected time without any conditioning?

Conditional expectation



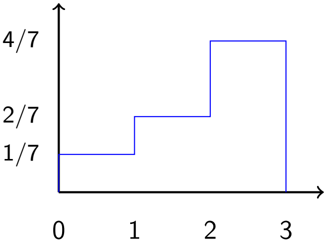
- I wait until 2pm. It still hasn't arrived. What is the expected value of the arrival time?
- What is the expected time without any conditioning?

Conditional expectation



- I wait until 2pm. It still hasn't arrived. What is the expected value of the arrival time?
- What is the expected time without any conditioning?
- First, what is the conditional probability, $f_{X|X>2}(x)$?

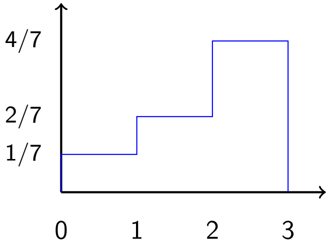
Conditional expectation



- I wait until 2pm. It still hasn't arrived. What is the expected value of the arrival time?
- What is the expected time without any conditioning?
- First, what is the conditional probability, $f_{X|X>2}(x)$?

$$f_{X|X>2}(x) = \begin{cases} 1 & \text{if } 2 < x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Conditional expectation

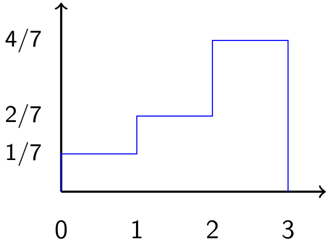


- I wait until 2pm. It still hasn't arrived. What is the expected value of the arrival time?
- What is the expected time without any conditioning?
- First, what is the conditional probability, $f_{X|X>2}(x)$?

$$f_{X|X>2}(x) = \begin{cases} 1 & \text{if } 2 < x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

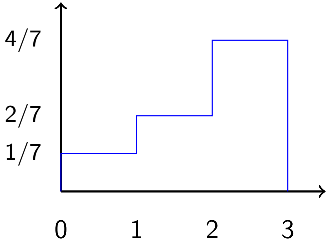
- So, $E[X|X > 2] = \int_{-\infty}^{\infty} x f_{X|X>2}(x) dx = \int_2^3 x \, dx = 2.5.$

Conditional expectation



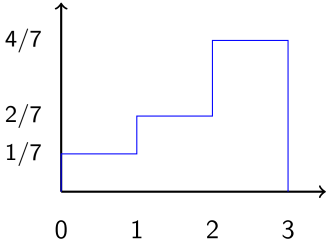
- What is the (unconditional) probability that $X > 2$?

Conditional expectation



- What is the (unconditional) probability that $X > 2$?
- $P(X > 2) = \int_2^3 f_X(x)dx = 4/7$

Conditional expectation



- What is the (unconditional) probability that $X > 2$?
- $P(X > 2) = \int_2^3 f_X(x)dx = 4/7$
- Similarly, $P(X < 1) = \int_0^1 f_X(x)dx = 1/7$ and $P(1 \leq X \leq 2) = 2/7$.

Total expectation theorem

- What is the total expectation of X?

Total expectation theorem

- What is the total expectation of X?
- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$

Total expectation theorem

- What is the total expectation of X?
- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- By the total probability theorem,

$$\begin{aligned} f_X(x) =& P(X \leq 1) f_{X|0 \leq X \leq 1}(x) \\ &+ P(1 \leq X \leq 2) f_{X|1 \leq X \leq 2}(x) + P(X > 2) f_{X|X>2}(x) \end{aligned}$$

Total expectation theorem

- What is the total expectation of X?
- $E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$
- By the total probability theorem,

$$\begin{aligned} f_X(x) =& P(X \leq 1) f_{X|0 \leq X \leq 1}(x) \\ &+ P(1 \leq X \leq 2) f_{X|1 \leq X \leq 2}(x) + P(X > 2) f_{X|X>2}(x) \end{aligned}$$

- So, we can write the total expectation as

$$\begin{aligned} E[X] = & \int_0^1 x P(X \leq 1) f_{X|X \leq 1}(x) + \int_1^2 x P(1 \leq X \leq 2) f_{X|1 \leq X \leq 2}(x) \\ &+ \int_2^3 x P(X > 2) f_{X|X>2}(x) \\ =& E[X|X \leq 1] P(X \leq 1) + E[X|1 \leq X \leq 2] P(1 \leq X \leq 2) \\ &+ E[X|X > 2] P(X > 2) \\ =& 0.5 \cdot 1/7 + 1.5 \cdot 2/7 + 2.5 \cdot 4/7 = 27/14 \end{aligned}$$

Total expectation theorem: Example

- John's tank holds 15 gallons of gas, and he always refills his tank when he gets down to 5 gallons.
- John's car gets 30MPG on average, with a standard deviation of 2MPG.
- I plan on borrowing John's car tomorrow. I don't know how much gas he will have. How far should I expect to be able to drive it?

- I want $E[M]$.

Total expectation theorem: Example

- John's tank holds 15 gallons of gas, and he always refills his tank when he gets down to 5 gallons.
- John's car gets 30MPG on average, with a standard deviation of 2MPG.
- I plan on borrowing John's car tomorrow. I don't know how much gas he will have. How far should I expect to be able to drive it?

- Let's set up some reasonable modeling assumptions.
- Let G be the random volume of gas. Assume

$$f_G(g) = \begin{cases} 0.1 & \text{if } 5 < g \leq 15 \\ 0 & \text{otherwise.} \end{cases}$$

- I want $E[M]$.
- Let M be the random number of miles. Assume $M \sim N(30g, 4)$.

Total expectation theorem: Example

- John's tank holds 15 gallons of gas, and he always refills his tank when he gets down to 5 gallons.
- John's car gets 30MPG on average, with a standard deviation of 2MPG.
- I plan on borrowing John's car tomorrow. I don't know how much gas he will have. How far should I expect to be able to drive it?

- Let's set up some reasonable modeling assumptions.
- Let G be the random volume of gas. Assume

$$f_G(g) = \begin{cases} 0.1 & \text{if } 5 < g \leq 15 \\ 0 & \text{otherwise.} \end{cases}$$

- I want $E[M]$.
- Let M be the random number of miles. Assume $M \sim N(30g, 4)$.
- If we have exactly g gallons, what is $E[M|G = g]$?

Total expectation theorem: Example

- John's tank holds 15 gallons of gas, and he always refills his tank when he gets down to 5 gallons.
- John's car gets 30MPG on average, with a standard deviation of 2MPG.
- I plan on borrowing John's car tomorrow. I don't know how much gas he will have. How far should I expect to be able to drive it?

- Let's set up some reasonable modeling assumptions.
- Let G be the random volume of gas. Assume

$$f_G(g) = \begin{cases} 0.1 & \text{if } 5 < g \leq 15 \\ 0 & \text{otherwise.} \end{cases}$$

- I want $E[M]$.
- Let M be the random number of miles. Assume $M \sim N(30g, 4)$.
- If we have exactly g gallons, what is $E[M|G = g]$? 30g
- So, we can use the total expectation theorem to get:

$$E[M] = \int_{-\infty}^{\infty} E[M|G = g] f_G(g) dg = \int_{5}^{15} 30g \times 0.1\, dg = [1.5g^2]_5^{15} = 300$$

Independent random variables

- For discrete random variables, we said two random variables X and Y are independent if

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

- Just like in the discrete case, we say two continous random variables are independent if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \qquad \forall x, y$$

- If $f_Y(y) > 0$, this is the same as saying $f_X(x) = f_{X|Y}(x|y)$ – i.e. knowing that $Y = y$ doesn't tell us anything about X.
- Just like with discrete random variables, we if X and Y are independent we have $E[XY] = E[X]E[Y]$ and $\text{var}(X+Y) = \text{var}(X) + \text{var}(Y)$.
  - For two functions $f(X)$ and $g(Y)$ we have $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$.

More than two random variables

- For multiple random variables we have:
  $P((X, Y, Z) \in B) = \int_{(x,y,z)\in B} f_{X,Y,Z}(x, y, z) dxdydz$
- Marginalization: $f_{X,Y}(x, y) =$

More than two random variables

- For multiple random variables we have:
  $P((X, Y, Z) \in B) = \int_{(x,y,z)\in B} f_{X,Y,Z}(x, y, z)dxdydz$
- Marginalization: $f_{X,Y}(x, y) = \int f_{X,Y,Z}(x, y, z)dz$

More than two random variables

- For multiple random variables we have:
  $P((X, Y, Z) \in B) = \int_{(x,y,z)\in B} f_{X,Y,Z}(x, y, z) dxdydz$
- Marginalization: $f_{X,Y}(x, y) = \int f_{X,Y,Z}(x, y, z) dz$
- Marginalization: $f_X(x) = \int \int f_{X,Y,Z}(x, y, z) dy \ dz$

More than two random variables

- For multiple random variables we have:
  $P((X, Y, Z) \in B) = \int_{(x,y,z)\in B} f_{X,Y,Z}(x, y, z) dxdydz$
- Marginalization: $f_{X,Y}(x, y) = \int f_{X,Y,Z}(x, y, z)dz$
- Marginalization: $f_X(x) = \int \int f_{X,Y,Z}(x, y, z)dy\ dz$
- Conditional PDF: $f_{X,Y|Z}(x, y|z) = \frac{f_{X,Y,Z}(x, y, z)}{f_Z(z)}$, For $f_Z(z) > 0$

More than two random variables

- For multiple random variables we have:
$P((X, Y, Z) \in B) = \int_{(x,y,z) \in B} f_{X,Y,Z}(x, y, z) dxdydz$
- Marginalization: $f_{X,Y}(x, y) = \int f_{X,Y,Z}(x, y, z) dz$
- Marginalization: $f_X(x) = \int \int f_{X,Y,Z}(x, y, z) dy \ dz$
- Conditional PDF: $f_{X,Y|Z}(x, y|z) = \frac{f_{X,Y,Z}(x, y, z)}{f_Z(z)}$, For $f_Z(z) > 0$
- Conditional PDF: $f_{X|Y,Z}(x|y, z) = \frac{f_{X,Y,Z}(x, y, z)}{f_{Y,Z}(y, z)}$, For $f_{Y,Z}(y, z) > 0$

More than two random variables

- For multiple random variables we have:
  $P((X,Y,Z) \in B) = \int_{(x,y,z)\in B} f_{X,Y,Z}(x,y,z)dxdydz$
- Marginalization: $f_{X,Y}(x,y) = \int f_{X,Y,Z}(x,y,z)dz$
- Marginalization: $f_X(x) = \int \int f_{X,Y,Z}(x,y,z)dy\ dz$
- Conditional PDF: $f_{X,Y|Z}(x,y|z) = \frac{f_{X,Y,Z}(x,y,z)}{f_Z(z)}$, For $f_Z(z) > 0$
- Conditional PDF: $f_{X|Y,Z}(x|y,z) = \frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)}$, For $f_{Y,Z}(y,z) > 0$
- Multiplication rule:
  $f_{X,Y,Z}(x,y,z) = f_{X|Y,Z}(x|y,z)f_{Y|Z}(y|z)f_Z(z)$, For $f_{Y,Z}(y,z) > 0$

More than two random variables

- For multiple random variables we have:
  $P((X,Y,Z) \in B) = \int_{(x,y,z)\in B} f_{X,Y,Z}(x,y,z)dxdydz$
- Marginalization: $f_{X,Y}(x,y) = \int f_{X,Y,Z}(x,y,z)dz$
- Marginalization: $f_X(x) = \int\int f_{X,Y,Z}(x,y,z)dy\ dz$
- Conditional PDF: $f_{X,Y|Z}(x,y|z) = \frac{f_{X,Y,Z}(x,y,z)}{f_Z(z)}$, For $f_Z(z) > 0$
- Conditional PDF: $f_{X|Y,Z}(x|y,z) = \frac{f_{X,Y,Z}(x,y,z)}{f_{Y,Z}(y,z)}$, For $f_{Y,Z}(y,z) > 0$
- Multiplication rule:
  $f_{X,Y,Z}(x,y,z) = f_{X|Y,Z}(x|y,z)f_{Y|Z}(y|z)f_Z(z)$, For $f_{Y,Z}(y,z) > 0$
- Independence: $f_{X,Y,Z}(x,y,z) = f_X(x)f_Y(y)f_Z(z)$ For all x,y,z

More than two random variables

- For two random variables X, Y arising out of the same experiment, we define their CDF as:

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) =$$

- How do I get $f_{X,Y}(x,y)$ back? $f_{X,Y}(x,y) = \dfrac{d^2 F_{X,Y}(x,y)}{dx\ dy}$
- Let X and Y be jointly uniform on the unit square. $F_{X,Y}(x,y) = xy$ for $0 \leq x, y \leq 1$
- What is $f_{X,Y}(x,y)$?. Differentiate! $\dfrac{d}{dx}\left(\dfrac{d}{dy}(xy)\right)$
- This equals 1 for all $0 \leq x, y \leq 1$!

More than two random variables

- For two random variables X, Y arising out of the same experiment, we define their CDF as:

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v)du\ dv$$

- How do I get $f_{X,Y}(x,y)$ back? $f_{X,Y}(x,y) = \dfrac{d^2 F_{X,Y}(x,y)}{dx\ dy}$
- Let X and Y be jointly uniform on the unit square. $F_{X,Y}(x,y) = xy$ for $0 \leq x, y \leq 1$
- What is $f_{X,Y}(x,y)$?. Differentiate! $\dfrac{d}{dx}\left(\dfrac{d}{dy}(xy)\right)$
- This equals 1 for all $0 \leq x, y \leq 1$!

Practice problem

- Let $Y = g(X) = X^2$. X is a random variable with a known PDF $f_X(x)$. Whats the PDF of Y?
- Solution: See example 3.23 of Bertsekas and Tsitsiklis.