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

SDS 321: Introduction to Probability and Statistics

Lecture 15: Continuous random variables

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

www.cs.cmu.edu/∼psarkar/teaching

Roadmap

- Continuous random variables
- PDF, CDF, Expectation, Variance
- Common distributions
  - The Uniform
  - The exponential
  - The normal distribution
  - Operations which preserve normality
  - Standardization
- Multiple random variables: joint distributions
  - Joint pdf
  - Joint pdf to a single pdf: Marginalization
  - Conditional pdf

The exponential distribution

- How to model the amount of time until something happens, such as
  - the next email arrives
  - an accident happens
  - a light bulb burns out
  - Notation: $X \sim Exp(\lambda)$

The exponential distribution

- An exponential r.v. has pdf and cdf:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \textit{otherwise} \end{cases}$$

$$\begin{aligned} F_X(x) &= \int_0^x \lambda e^{-\lambda y} dy \\ &= \int_0^{\lambda x} e^{-v} dv = 1 - e^{-\lambda x} \end{aligned}$$

- $$\begin{aligned} E[X] &= \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \int_0^\infty (\lambda x) e^{-\lambda x} d(\lambda x) \\ &= \frac{1}{\lambda} \int_0^\infty u e^{-u} du = \frac{1}{\lambda} \end{aligned}$$

- $\mathrm{var}(X) = E[X^2] - E[X]^2 = \int_0^\infty x^2 \lambda e^{-\lambda x} dx - \frac{1}{\lambda^2} =$
  $\frac{1}{\lambda^2} \int_0^\infty u^2 e^{-u} du - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

The exponential distribution

- Integration by parts anyone?

$$\int f(x)g'(x)dx = f(x)g(x) - \int f'(x)g(x)dx$$

$$\int xe^{-x}dx = x(-e^{-x}) + \int e^{-x}dx = -xe^{-x} - e^{-x}$$

$$\int_0^\infty xe^{-x}dx = -xe^{-x}\Big|_0^\infty - e^{-x}\Big|_0^\infty = 1$$

$$\int x^2e^{-x}dx = x^2(-e^{-x}) + 2\int xe^{-x}dx$$

$$\int_0^\infty x^2e^{-x}dx = -x^2e^{-x}\Big|_0^\infty + 2\int_0^\infty xe^{-x}dx = 2$$

Multiple random variables

We can also have multiple *continuous* random variables associated with the same experiment/sample space.

- For example, our experiment might be selecting a randomly selected person.
- The sample space would be the set of all possible characteristics of this person.
- We can summarize these characteristics into continuous random variables, e.g.
  - The person's height
  - The person's weight
  - The person's age
- Again, multiple random variables stemming from the same sample space!
- These random variables will often depend on each other: Knowing a person is taller than 6'5" tells us something about their expected weight.

Multiple continuous random variables

- Let X and Y be two continuous random variables.
- Each one takes on values on the real line, i.e. $X \in \mathbb{R}$ and $Y \in \mathbb{R}$.
- Together, each possible pair of values describe a point in the real plane, i.e. $(X, Y) \in \mathbb{R}^2$.
- We say X and Y are **jointly continous** if the probability of them jointly taking on values in some subset B of the plane can be described as

$$P((X, Y) \in B) = \iint_{(x,y)\in B} f_{X,Y}(x, y) dx\, dy$$

  using some continuous function $f_{X,Y}$, for all $B \in \mathbb{R}^2$ – i.e. all subsets of the 2-D plane.
- Notation means “integrate over all values of x and y s.t. $(x, y) \in B$

Joint PDF

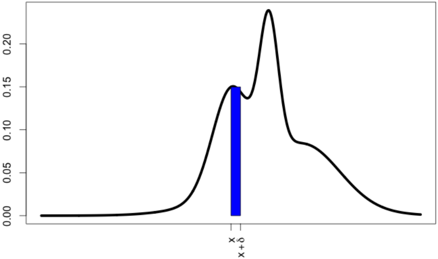
- We call $f_{X,Y}$ the **joint pdf** of X and Y.
- It allows us to calculate the probability of any set of combinations of X and Y
  - e.g. the probability that a person weighs over 200lb and is under 6'
  - e.g. the probability that a person's height in inches is more than twice their weight in pounds.
  - So, this could describe the first scenario above, $P(200 \leq X \leq \infty, -\infty \leq Y \leq 6)$
- What is $\int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} f_{X,Y}(x,y) dx\, dy$?

Joint PDF

- We call $f_{X,Y}$ the **joint pdf** of X and Y.
- It allows us to calculate the probability of any set of combinations of X and Y
  - e.g. the probability that a person weighs over 200lb and is under 6'
  - e.g. the probability that a person's height in inches is more than twice their weight in pounds.
  - So, this could describe the first scenario above, $P(200 \leq X \leq \infty, -\infty \leq Y \leq 6)$
- What is $\int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} f_{X,Y}(x,y) dx\, dy$? 1

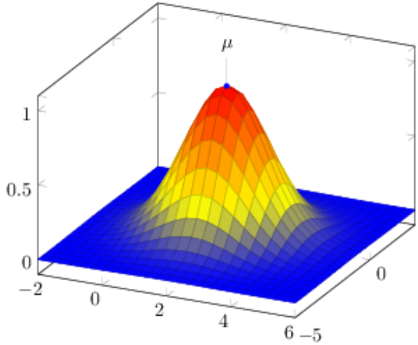
Joint PDF: Intuition

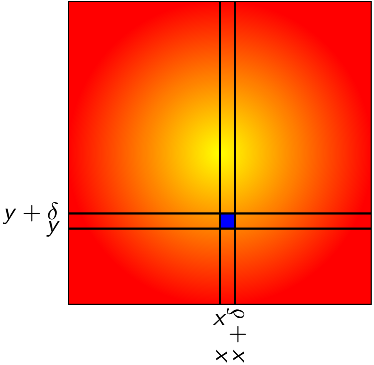
- Remember we could think of $f_X(x)$ as the "probability mass per unit length" near to x?



- Because $f_X(x) = \dfrac{P(x \leq X \leq x + \delta)}{\delta}$

Joint PDF: Intuition





- We can think of the joint PDF $f_{X,Y}(x, y)$ as the "probability mass per unit area" for a small area near X.
- Again, remember, $f_{X,Y}(x, y)$ **is not a probability!**

Multiple random variables to a single random variable

- We can get from the **joint PMF** of X and Y to the **marginal PMF** of X by summing over (marginalizing over) Y:

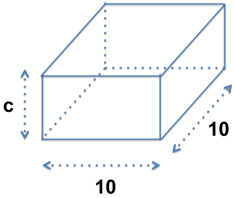
$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

- We can get from the **joint PDF** of X and Y to the **marginal PDF** of X by integrating over (marginalizing over) Y:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

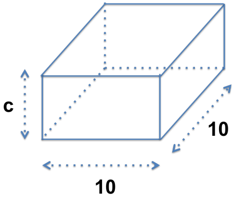
Example: Bivariate uniform random variable

- Anita (X) and Benjamin (Y) both pick a number between 0 and 10, according to a continuous uniform distribution. What is $f_{X,Y}(x, y)$?



Example: Bivariate uniform random variable

- Anita (X) and Benjamin (Y) both pick a number between 0 and 10, according to a continuous uniform distribution. What is $f_{X,Y}(x,y)$?



- Let's see... we know all pairs (x,y) are equally likely, so we know $f_{X,Y} = c$. It must satisfy $\int_{x=0}^{10} \int_{y=0}^{10} f_{X,Y}(x,y) dx\, dy = 1$.
- So, $c \underbrace{\int_{x=0}^{10} \int_{y=0}^{10} dx\, dy}_{100} = 1 \ldots$
- So $c = f_{X,Y}(x,y) = 0.01$ for all $0 \leq x, y \leq 10$.

Example: marginal PDF

- $f_{X,Y}(x,y) = \begin{cases} 0.01 & \text{If } x, y \in [0, 10] \\ 0 & \text{otherwise} \end{cases}$
- What is $f_X(x)$?

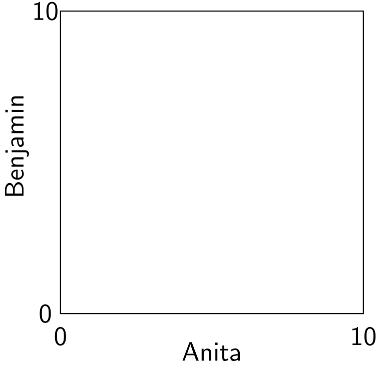
- In general, we will have $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$
- We have **marginalized out** one of our random variables... just like we did when looking at PMFs.
- We call $f_X(x)$ the **marginal PDF** of X

Example: marginal PDF

- $f_{X,Y}(x,y) = \begin{cases} 0.01 & \text{If } x, y \in [0, 10] \\ 0 & \text{otherwise} \end{cases}$
- What is $f_X(x)$?
- $f_X(x) = \begin{cases} \int_{y=0}^{10} 0.01 dy = 0.1 & \text{If } x \in [0, 10] \\ 0 & \text{otherwise} \end{cases}$
- Not surprisingly $X \sim Uniform([0, 10])$ and $Y \sim Uniform([0, 10])$.
- In general, we will have $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
- We have **marginalized out** one of our random variables... just like we did when looking at PMFs.
- We call $f_X(x)$ the **marginal PDF** of X

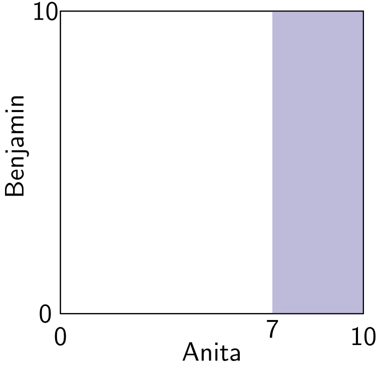
Example: marginalization

- What is the probability that Anita picks a number greater than 7?



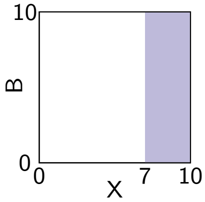
Example: marginalization

- What is the probability that Anita picks a number greater than 7?



- That's going to correspond to the shaded region...
  $P(X > 7) = 0.01(3 \times 10) = 0.3$.
- Or, using calculus: $\int_{x=7}^{10} \int_{y=0}^{10} f_{X,Y}(x, y) dx\, dy$

Marginalization

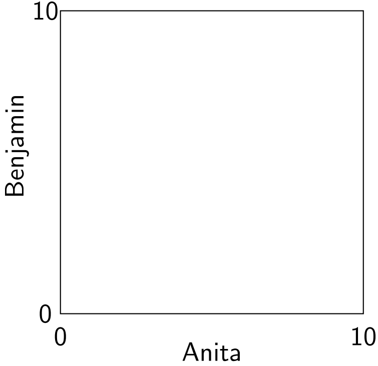


- $P(X > 7) = \int_{x=7}^{10} \int_{y=0}^{10} f_{X,Y}(x, y) dx\, dy$

- But, this doesn't depend on Benjamin at all! It is the same as $P(X > 7) = \int_{x>7} f_X(x) dx$.

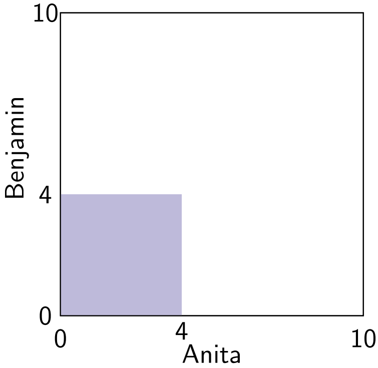
Example: Uniform random variable

- What is the probability that they both pick numbers less than 4?



Example: Uniform random variable

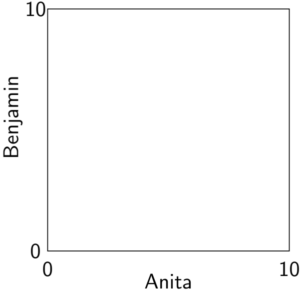
- What is the probability that they both pick numbers less than 4?



- It will be $0.01 \int_0^4 \int_0^4 dx\, dy = 0.01 \times 16 = 0.16$
  – i.e. $0.01 \times$ the shaded area.
  – Or $16/100$!

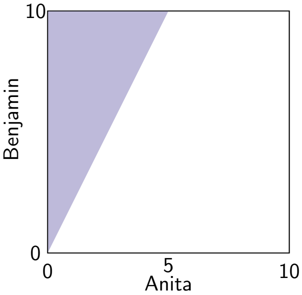
Example: Uniform random variable

- What is the probability that Benjamin picks a number at least twice that of Anita?



Example: Uniform random variable

- What is the probability that Benjamin picks a number at least twice that of Anita?

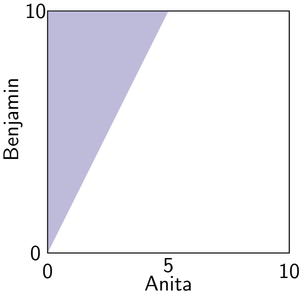


- That's going to correspond to the shaded region...
$P(Y \geq 2X) = 0.01(0.5 \times 5 \times 10) = 0.25.$

- Or, using calculus: $\int_{x=0}^{10} \int_{y=2x}^{10} f_{X,Y}(x,y) dx\, dy = \int_{x=0}^{10} \int_{y=2x}^{10} c \times 1_{0 \leq x \leq 10, 0 \leq y \leq 10} dx\, dy$

Example: Uniform random variable

- What is the probability that Benjamin picks a number at least twice that of Anita?



- That's going to correspond to the shaded region...
  $P(Y \geq 2X) = 0.01(0.5 \times 5 \times 10) = 0.25$.
- Or, using calculus: $\int_{x=0}^{10} \int_{y=2x}^{10} f_{X,Y}(x,y) dx\, dy = \int_{x=0}^{10} \int_{y=2x}^{10} c \times 1_{0 \leq x \leq 10, 0 \leq y \leq 10} dx\, dy$
- $$\int_{x=0}^{10} \int_{y=2x}^{10} c \times 1_{0 \leq 2x \leq 10} dx\, dy = c \int_0^5 dx = c \int_{x=0}^{5} (10 - 2x) dx$$
  $$= c(10 \times 5 - (5^2 - 0)) = 0.01 \times 25 = 0.25$$

Conditional PDFs

- 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

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

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

Conditioning: memoryless property of the exponential

- $X \sim Exp(\lambda)$
- $P(X \geq s + t | X \geq s) = ?$

Conditioning: memoryless property of the exponential

- $X \sim Exp(\lambda)$
- $P(X \geq s + t | X \geq s) = ?$
- 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

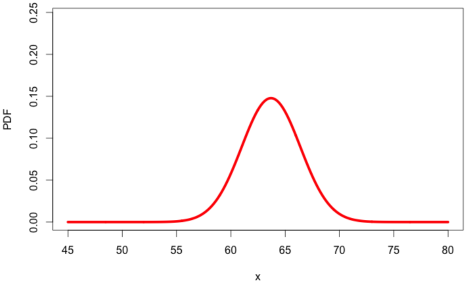
- $X \sim Exp(\lambda)$
- $\begin{cases} f_{X|X>s}(x) = \dfrac{\lambda e^{-\lambda x}}{P(X>s)} = \lambda e^{\lambda(x-s)} & \text{If } x > s \\ = 0 & \text{Otherwise} \end{cases}$
- $P(X > s + t | X > s) = ?$

Conditioning: memoryless property of the exponential

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

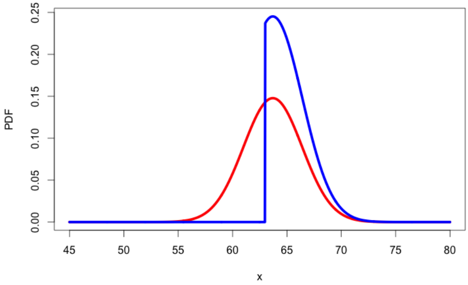
Conditional PDFs: Example

- 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.



Recap

- Last time, we introduced the idea of continuous random variables and PDFs.
- A PDF is a function we can integrate over to get $P(X \in B) = \int_B f_X(x)dx$.
- We extended this to look at **joint PDFs** and **conditional PDFs**.
- We can borrow results from conditional probability and probabilities of intersections!
- But we need to be careful to remember, a PDF is **not** a probability...
- Next time, we will continue looking at continuous probability distributions.