Ec140 - Core Concepts from Statistics

Fernando Hoces la Guardia
06/22/2022

Housekeeping

- Dates for chapter summaries are now on the syllabus.
- Any troubles accessing the book?
- Any first thoughs on PS 1?
- Font size for people in the back ok?

Concepts to Review Today

- Random Variables
- Probability and Probability Distribution/Density Functions

Random Variables 1/3

Imagine you are meeting with a friend later today and let's assume of the three following situations can happen:

- You have the most wonderful time that you could have had today with your friend.
- It was ok, but you were kind of expecting to have a better time today.
- You had a really bad time, your friend did not stop talking about herself and never checked on you.

Roll one Die:

- If you get a 1 or a 2, the first situation happen
- If you get a 3 or a 4, the second situation took place
- 5 or 6 corresponds to the last situation

Each of these situations is what is called an **event** in statistics. Two key characteristics of events are

- They must be mutually elusive
- The collection of all events (called the event space) must contain all the posible outcomes.

This events look a bit hard to keep track of...

Random Variables 2/3

Let's assign a number to each event:

Numbering Events

$Event$	X
You have the most wonderful...	1
It was ok, but...	2
You had a really bad time, ...	3

- This assignment from event to numerical values is what a **random variable** does.
- For this case, where there is a discrete number events, we called this a **discrete random variable**
- You might be used to thinking about variables as collections of numbers in a line (or multiple lines).
- A random variable is similar in the sense that it is a collection of numbers, but it is different in the sense that each number represents an event.

Random Variables 3/3

- Another potential source of confusion is that the idea of connecting something with other numbers is similar of the role of a function, not variables. We need random variables first to move from events into numbers.
- Random variables can also be continued, but for this we cannot talk of an Event-value mapping, instead we combine the event and value in the same object (the number of the random var.).
  - Example: income. In the continuous case, the idea of an event is not so meaningful (what is the likelihood of an income of 33,593.4355?)
  - Hence, when working with continuous random variables we will refer to ranges.
  - Let's go to Seeing Theory and play a bit with a discrete random variable
    - For example: think of that event space as a field where you are throwing a ball

Now we have all we need to define a random variable.

- **Random Variable** is a numerical summary of a random outcome.
- We now know what a random variable is. Notice that we have almost not talk about how likely or unlikely each of these values are, we have not talked about its *proportions in the long run*.

Probability Distribution Functions 1/2

- "Proportions in the long run", is a bit of a mouthful. Let's put a name to it, let's call it **probabilities**.
- Now we need a mapping from an event ("Most wonderful time that you could have had today with your friend", or 1) into a probability.
- This mapping will be described by a function $P(X = x)$ that represents the probability that a random variable X, takes the specific value of x (e.g., 1, or "Most wonderful ...").

Events, Random Variables, and Probabilities

$Event$	X	$P(X = x)$
You have the most wonderful...	1	1/3
It was ok, but...	2	1/3
You had a really bad time, ...	3	1/3

Probability Distribution Functions 2/2

- The **probability distribution function** of a discrete random variable is the list of all possible values of a variable and the probability that each value will occur.
- Notation. Sometimes you will see it express as $P(X = x)$ or as $f(x)$. The former applies only to discrete PDFs, the latter can be used for both continous or discrete random variables.
- A closely related concept describes the acumulated probability of a collection of events. How would you write the probability that people in this room had "At least an ok time with their friend"
- This Cumulative Distribution Function describes the probability that a random variables is less or equal than a particular value. $P(X \leq x)$
- Lets go to Seeing Theory again and explore some PDFs and CDFs of discrete random variables.
    - Bernoulli: success and failure, 0 and 1. Show how p characterizes the entire distributions. Review ranges and values of PDF and CDF.
    - Binomial as example of combining RVs.

Activity 1

- The mythical Stat 110 team shows us how the concepts of Discrete Random Variables and Probability Distributions can be used in practice.
  - Let's watch this video. Then get together in groups of 3 and discuss:
  - What is the event space?
  - What is the random variable?
  - What is the probability distribution that one person is cured?
  - How does the Binomial(10, 0.6) relates to that previous distribution?

Probability Density Functions 1/2

- For the case of continuos random variables, lets think of what is the probability of a specific event (e.g., income 34,680.0003).
- For continued random variables we cannot talk about a (specific) probability distribution, we use instead the concepts of density (as in how much is there in any given range).
- The probability that a continues random variable lies between any given range (of two points) is defined as the area under the **probability density function** of a continuous random variable. [This is a very abstract concept]
- The density function is typically denoted by $f(x)$.
- Note that any specific value of $f(x)$ should not be interpreted as a probability.
- The area under the the probability density function, or density, is the cumulative distribution function between two points:

$$P(x_1 \leq X \leq x_2) = \text{area betwen } f(x_1) \text{ and } f(x_2)$$

Probability Density Functions 2/2

- Lets go to Seeing Theory again and explore some PDFs and CDFs of continous random variables.
    - Uniform: great to represent ignorance.
    - Normal to show how just two parameters can characterize and entire prob distribution.
    - Exponential to show that weird shapes can happen. Check this video for a great demonstration of how this type of distributions appear in real life.
    - Beta to show that one distribution can describe many phenomena. Talks about range.

Activity 2

- Let's watch another Stat 110's video. Then get together in groups of 3 and discuss:
    - What is the type of random variables are represented by Norma and Randy?
    - What is Norma's take on how to address the paradox describe by ancient philosophers?
    - The concept of error tolerance presented in the video correspnds to which concept discuss in previous slides?
    - The last 2 mins on prediction are outside the scope of our course.]

Densities of Continuous Random Variables

Uniform Distribution

The probability density function of a variable uniformly distributed between 0 and 2 is

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{if } x < 0 \text{ or } x > 2 \end{cases}$$

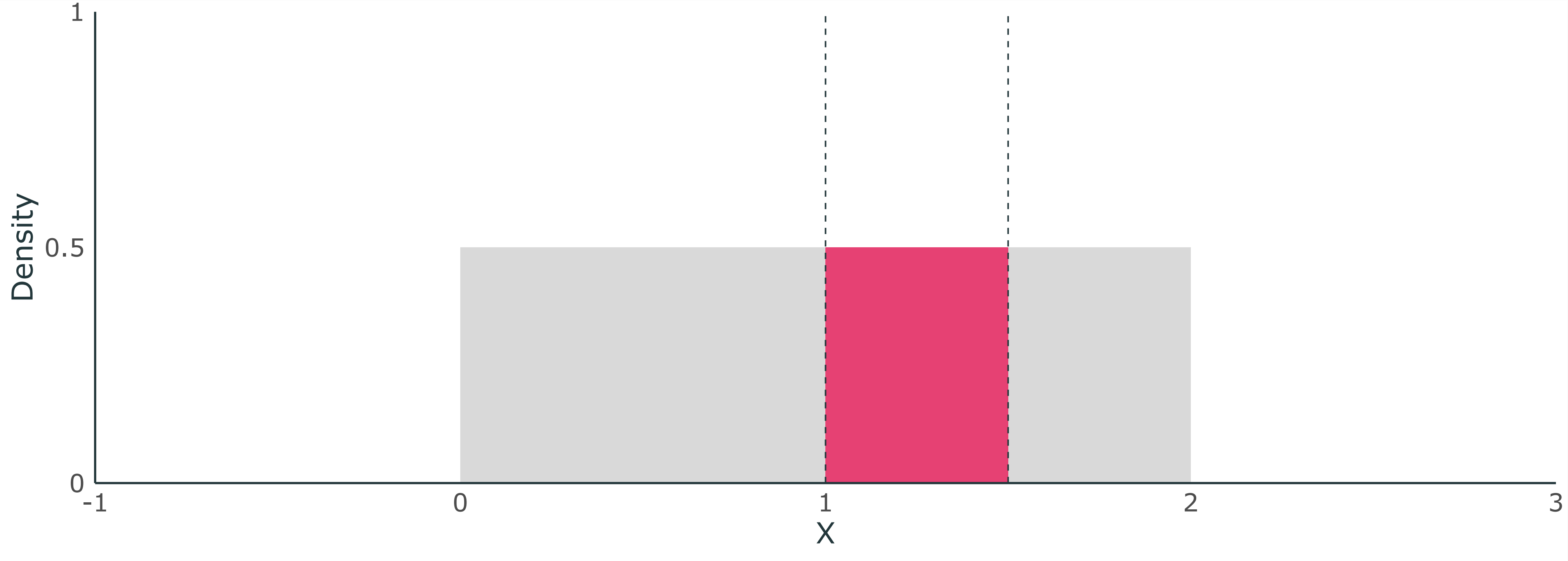
Densities of Continuous Random Variables

Uniform Distribution

By definition, the area under $f(x)$ is equal to 1.

The shaded area illustrates the probability of the event $1 \leq X \leq 1.5$.

- $\mathbb{P}(1 \leq X \leq 1.5) = (1.5 - 1) \times 0.5 = 0.25$.

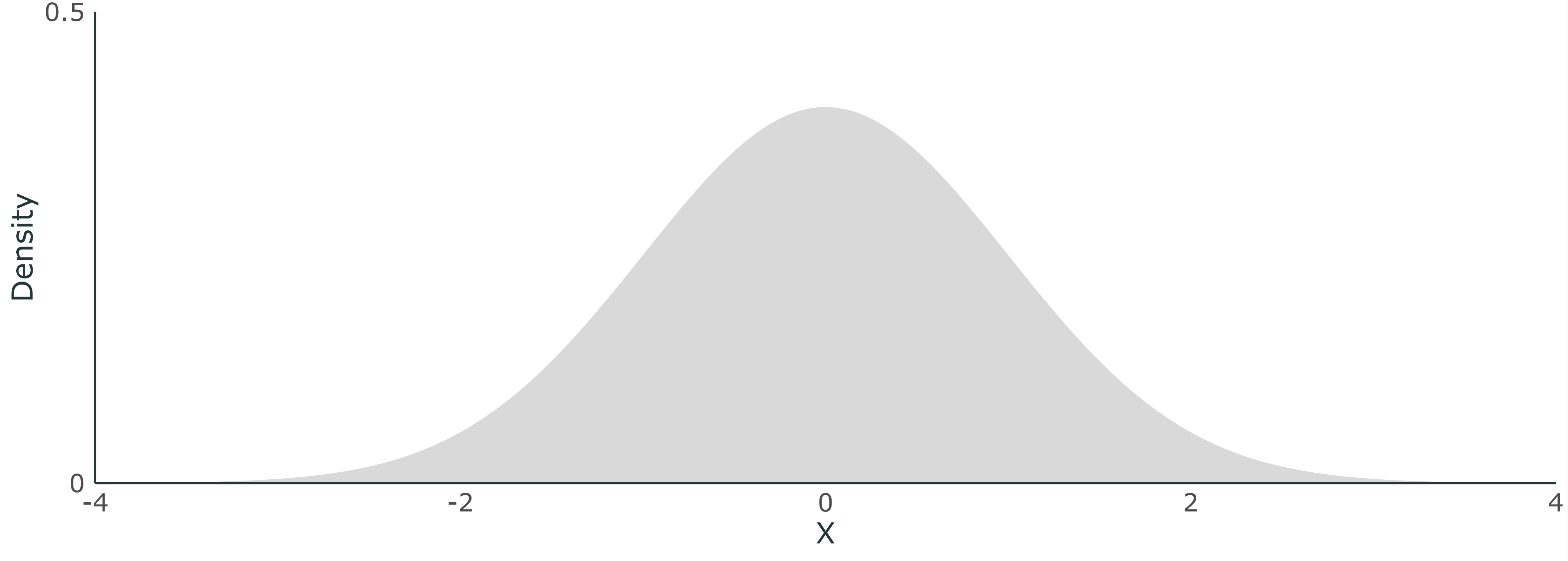


Densities of Continuous Random Variables

Normal Distribution

The "bell curve."

- Symmetric: mean and median occur at the same point (*i.e.*, no skew).
- Low-probability events in tails; high-probability events near center.

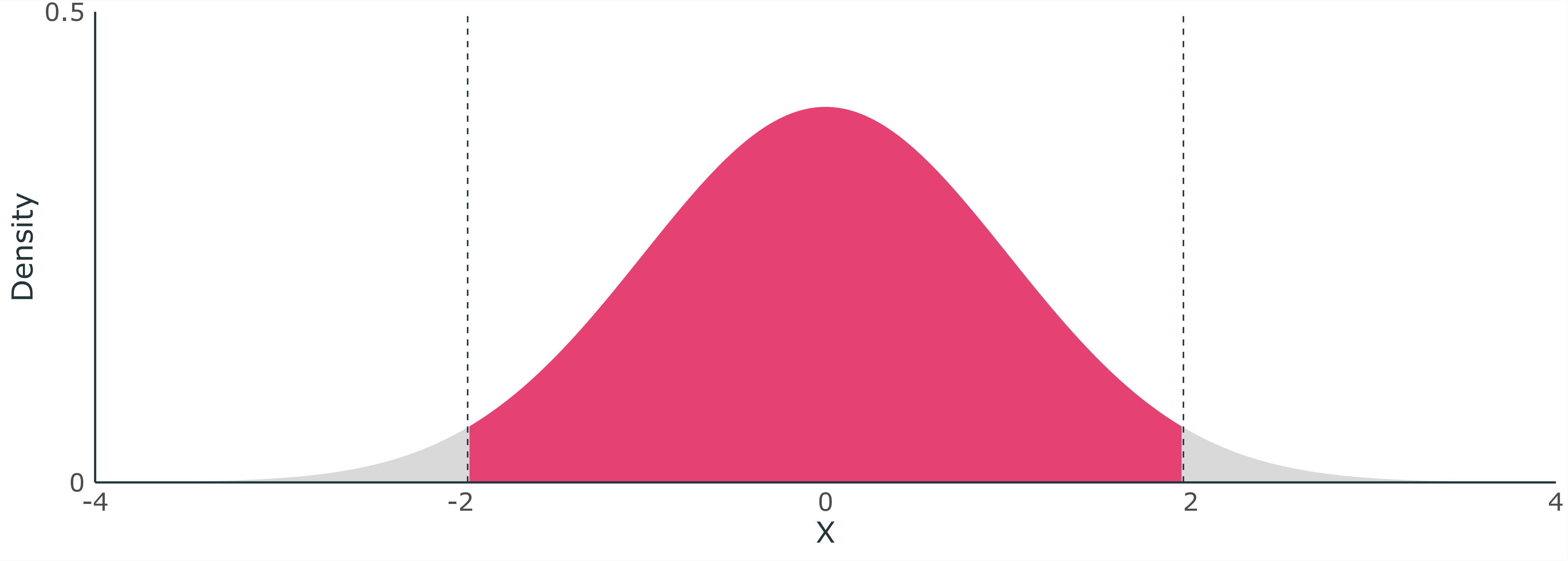


Continuous Random Variables

Normal Distribution

The shaded area illustrates the probability of the event $-2 \leq X \leq 2$.

- "Find area under curve" = use integral calculus (or, in practice, R).
- $\mathbb{P}(-2 \leq X \leq 2) \approx 0.95$.



Acknowledgments

- I started this slides on the basis of Kyle Raze's slides from University of Oregon.
- Seeing Theory.
- Stat 110.
- Stock and Watson 3e. Chapter 2, P61-65.