Lecture 12: Continuous random variables

Statistics 251

Yi Sun and Zhongjian Wang

Department of Statistics
The University of Chicago

Midterm

- Solution to midterm is ready on Canvas
- Our TA is planning to hold a special session for midterm this afternoon at 5PM during his office our.
- Re-grade is available until end of this week.
- Median 74.0 Maximum 85.0 Mean 68.5

Where are we?

Continuous random variables

Expectation and variance of continuous random variables

Do all sets have probabilities?

Continuous random variables

We say that X is a **continuous** random variable if there is a **probability density function** $f(x) = f_X(x)$ on $\mathbb{R}$ such that for any $A \subseteq \mathbb{R}$ we have

$$\mathbb{P}(X \in A) = \int_A f_X(x)dx = \int_{-\infty}^{\infty} 1_A(x) f_X(x)dx.$$

Notice that

- $1 = \mathbb{P}(X \in \mathbb{R}) = \int_{-\infty}^{\infty} f_X(x)dx$
- $f_X(x)$ takes non-negative values

Can we have $f_X(x) > 1$ for some x?

Continuous random variables

We say that X is a **continuous** random variable if there is a **probability density function** $f(x) = f_X(x)$ on $\mathbb{R}$ such that for any $A \subseteq \mathbb{R}$ we have

$$\mathbb{P}(X \in A) = \int_A f_X(x)dx = \int_{-\infty}^{\infty} 1_A(x) f_X(x)dx.$$

Notice that

- $\mathbb{P}(X = a) = \int_{\{a\}} f_X(x)dx = 0$ so the probability of any single point is 0.
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$.

Continuous random variables

We say that X is a **continuous** random variable if there is a **probability density function** $f(x) = f_X(x)$ on $\mathbb{R}$ such that for any $A \subseteq \mathbb{R}$ we have

$$\mathbb{P}(X \in A) = \int_A f_X(x)dx = \int_{-\infty}^{\infty} 1_A(x) f_X(x)dx.$$

Notice that

- $\mathbb{P}(X = a) = \int_{\{a\}} f_X(x)dx = 0$ so the probability of any single point is 0.
- $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x)dx$.

The **cumulative distribution function** is

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(x)dx.$$

So, given F_X is differentiable at x_0,

$$f_X(x_0) = F_X'(X_0)$$

Example

Suppose that the density function is

$$f_X(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$$

What are the values of:

- $\mathbb{P}(X < 3/2)$
- $\mathbb{P}(X \leq 3/2)$
- $\mathbb{P}(1/2 < X < 3/2)$

What is the CDF of X?

In this case, we say that X is **uniformly distributed** on $[0,2]$.

Example

Suppose that the density function is

$$f_X(x) = \begin{cases} x/2 & x \in [0,2] \\ 0 & x \notin [0,2]. \end{cases}$$

What are the values of:

- $\mathbb{P}(X < 3/2)$
- $\mathbb{P}(X \leq 3/2)$
- $\mathbb{P}(1/2 < X < 3/2)$

What is the CDF of X?

Where are we?

Continuous random variables

Expectation and variance of continuous random variables

Do all sets have probabilities?

Expectation

Recall that if X is discrete, we have that

$$\mathbb{E}[X] = \sum_{x:\mathbb{P}(X=x)>0} x \cdot \mathbb{P}(X = x).$$

To modify it to the continuous setting, we write

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

and correspondingly

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx.$$

Variance

If X is a continuous random variable, we define

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_{-\infty}^{\infty} (x - \mathbb{E}[X])^2 p_X(x) dx.$$

As before, we find that

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[(X - \mathbb{E}[X])^2] \\ &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.\end{aligned}$$

Examples

If the density function is

$$f_X(x) = \begin{cases} 1/2 & x \in [0,2] \\ 0 & x \notin [0,2], \end{cases}$$

what is $\mathrm{Var}(X)$?

Examples

If the density function is

$$f_X(x) = \begin{cases} x/2 & x \in [0,2] \\ 0 & x \notin [0,2], \end{cases}$$

what is $\text{Var}(X)$?

Where are we?

Continuous random variables

Expectation and variance of continuous random variables

Do all sets have probabilities?

Uniform measure on $[0, 1]$

Consider the uniform measure on $[0, 1]$ with $f_X(x) = 1_{[0,1]}(x)$.

Consider the translations $T_r : [0, 1) \to [0, 1)$ so that

$$T_r(x) = x + r \pmod 1.$$

Call $x, y \in [0, 1)$ equivalent if $x - y$ is rational.
Now we try to find $A \subset [0, 1)$ such that each point in $[0, 1)$ is equivalent to exactly 1 point in A.

Uniform measure on $[0, 1]$

Consider the uniform measure on $[0, 1]$ with $f_X(x) = 1_{[0,1]}(x)$.

Consider the translations $T_r : [0, 1) \to [0, 1)$ so that

$$T_r(x) = x + r \pmod 1.$$

Call $x, y \in [0, 1)$ equivalent if $x - y$ is rational.
Now we try to find $A \subset [0, 1)$ such that each point in $[0, 1)$ is equivalent to exactly 1 point in A. Such set is called Vitali set.

The existence of Vitali set can be proved by axiom of choice. In modern algebra, $A = \mathbb{R}/\mathbb{Q}$.

Uniform measure on $[0, 1]$

For this choice, we have that $T_r(A)$ and $T_s(A)$ are disjoint for $r \neq s$ and that

$$[0, 1) = \bigcup_{r \in \mathbb{Q}} T_r(A).$$

Let $A_r = T_r A$, first $\mathbb{P}(A_r) = \mathbb{P}(A)$ and

$$\mathbb{P}([0, 1]) = \sum_{r \in \mathbb{Q}} \mathbb{P}(A_r) = \sum_{r \in \mathbb{Q}} \mathbb{P}(A).$$

What does this mean for $\mathbb{P}(A)$?

Not all sets have probabilities

Possible answers:

1. Axioms of mathematics are wrong. In other words, we cannot chose A so that

   each point in $[0, 1)$ is equivalent to exactly 1 point in A.

   This requires something called the **axiom of choice**.

2. Accept that not all sets can have probabilities. Instead, define some sets to be **measurable** and allow only those to have probabilities. In this case, the relevant sets are countable unions of intervals and points.

Most of mainstream mathematics takes Answer 2.