Math 3070, Applied Statistics

Section 1

September 21, 2019

Lecture Outline, 9/21

Section 4.1

- Expected Value and Standard Deviation Revisited
- Probability Density Functions
- Uniform Random Variables
- Examples

Expected Value and Standard Deviation Revisited

Suppose that you have a data set x_i where the number 1 shows up 200 times, the number 2 shows up 300 times and the number 3 shows up 300 times. This data set has a 800 observations.

Let's compute the sample mean.

$$\frac{1}{800}\sum_{i=1}^{800} x_i = 1\frac{200}{800} + 2\frac{300}{800} + 3\frac{300}{800} = 2.125$$

Now compute the expected value of the random variable X with the folowing PMF

$$f(x) = P(X = x) = \begin{cases} 2/8, & x = 1 \\ 3/8, & x = 2 \\ 3/8, & x = 3 \\ 0, & \text{otherwise} \end{cases}$$

$$E[X] = 1\frac{2}{8} + 2\frac{3}{8} + 3\frac{3}{8} == 2.125$$

Expected Value and Standard Deviation Revisited

In the frequentist view of statistics, the $E[X]$ measures the center or average of a random variable. When sample size increases, the sample mean becomes closer to this theoretical mean which is related to other parameters depending on distribution.

Expected Value and Standard Deviation Revisited

Suppose that you have a data set x_i where the number 1 shows up 200 times, the number 2 shows up 300 times and the number 3 shows up 300 times. This data set has a 800 observations.

Let's compute the sample variance.

$$\frac{1}{800}\sum_{i=1}^{800} x_i = (1-2.125)^2\frac{200}{800}+(2-2.125)^2\frac{300}{800}+(3-2.125)^2\frac{300}{800} = 2.125$$

Each squared term is the distance from the mean or the deviation away from the mean.

$$V(X) = E[(X-E[X])^2] =$$
$$(1-2.125)^2\frac{2}{8}+(2-2.125)^2\frac{3}{8}+(3-2.125)^2\frac{3}{8} = 2.125$$

Both variances measure spread from the mean. In the same way as with the means, the sample variance becomes closer to the theoretical variance.

Continuous Random Variables

So far, we have only discussed discrete random variables, which have only a sequence of possible values (usually whole numbers):

- The number of defective widgets in a batch.
- The number of widgets inspected before finding one defective.
- The number of customers who visit a store in an hour.

However, many quantities in real life vary continuously:

- The length of a metal rod.
- The strength of a specimen of concrete.
- The weight of a bottled drink.
- The amount of time until the next customer arrives.

We will need different techniques to deal with continuous random variables.

Continuous Random Variable

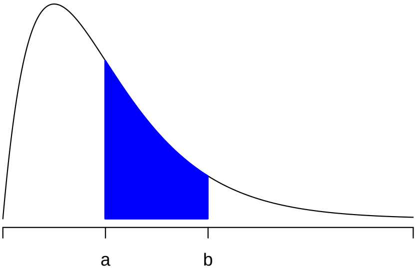
We say that a random variable X is **continuous** if $P(X = x) = 0$ for every x. If there is a function $f(x)$ such that for all $a \leq b$,

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx$$

then we call $f(x)$ a **probability density function** (pdf) of X.

To be a valid pdf, we must have

1. $f(x) \geq 0$ for all x.
2. $\int_{-\infty}^{\infty} f(x) = 1$.



Standard Uniform Random Variable

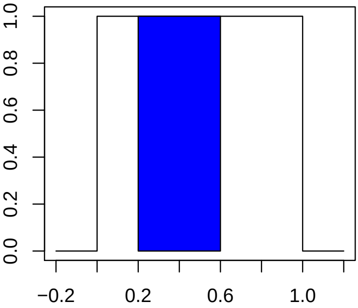
Define a pdf by

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

The continuous random variable X with this pdf is called a **standard uniform** random variable; it takes values uniformly on the interval [0,1]. $X \sim unif(0,1)$

For example, the probability that X is between .2 and .6 is

$$\begin{aligned} & P(.2 \leq X \leq .6) \\ &= \int_{.2}^{.6} 1 \; dx \\ &= x\Big|_{.2}^{.6} \\ &= .6 - .2 \\ &= .4 \end{aligned}$$



Uniform Random Variable

We say that X is a **uniform** random variable on the interval $[a, b]$ if X has pdf

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

$X \sim Unif(a, b)$

Example: Suppose that the time we have to wait at a bus stop is a uniform random variable X between 0 and 15 minutes. What is the probability that we will have to wait more than 10 minutes?

$$\begin{aligned} P(X \geq 10) &= \int_{10}^{\infty} f(x)\, dx \\ &= \int_{10}^{15} \frac{1}{15-0}\, dx \\ &= \frac{1}{15} x \Big|_{10}^{15} = \frac{15-10}{15} = 1/3 \end{aligned}$$

Summary

- X is continuous if $P(X = x) = 0$ for all x.
- The probability density function PDF f of X satisfies

$$P(a \leq X \leq b) = \int_a^b f(x)dx.$$

  This means that the proability of events involving on continuous random variables can be computed used integrals.
- Need $f(x) \geq 0$ for all x and $\int_{-\infty}^{\infty} f(x) = 1$.
- The PDF identifies the continuous random variable.
- $X \sim unif(a, b)$ or a random variable X is uniformly distributed on the interval $[a, b]$ means X has

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

  as it's PDF. Note, a, b from this bullet point is not the same as the previous a, b

Example, Non-PDFs

Which of the following could be PDFs?

$$f(x) = \begin{cases} \frac{\cos(x)}{2} & \text{if } 0 \le x \le \pi \\ 0 & \text{otherwise} \end{cases} \qquad g(x) = \begin{cases} \frac{\sin(x)}{2} & \text{if } 0 \le x \le \pi \\ 0 & \text{otherwise} \end{cases}$$

$$f(3\pi/4) = -1/\sqrt{2} < 0$$

$f(x) \ge 0$ is false. $f(x)$ cannot be a PDF.

$g(x) \ge 0$ is true. $g = 0$ outisde $[0, \pi]$.

$$\int_{-\infty}^{\infty} g(x) = \int_0^{\pi} \frac{\sin(x)}{2} dx = -\frac{\cos(x)}{2}\bigg|_{x=0}^{\pi} = 1$$

$g(x)$ could be a PDF.

If fact, it is one since the PDF identifies the random variable.

Example, Unbounded X and Complement Example

Show that

$$f(x) = \begin{cases} e^{-x} & \text{if } 0 \leq x \\ 0 & \text{otherwise} \end{cases}$$

is a PDF. Compute the probability of $X > 1$.

$f(x) \geq 0$ for all x.

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} e^{-x}dx = -e^{-x}\Big|_{x=0}^{\infty} = 0 - (-1) = 1$$

$f(x)$ is a PDF since it satisfies the conditions.

Example, Unbounded X and Complement Example

Show that

$$f(x) = \begin{cases} e^{-x} & \text{if } 0 \leq x \\ 0 & \text{otherwise} \end{cases}$$

is a PDF. Compute the probability of $X > 1$.

$$P(X > 1) = \int_{x=1}^{\infty} e^{-x} dx = -e^{-x} \Big|_{x=1}^{\infty} = 0 - (-e^{-1}) = e^{-1}$$

$$\begin{aligned} P(X > 1) &= 1 - P(X \leq 1) = 1 - \int_{-\infty}^{1} f(x) dx \\ &= 1 - \int_{0}^{1} e^{-x} dx = 1 - \left(- e^{-x} \Big|_{x=0}^{1} \right) \\ &= 1 - \left(- e^{-1} + 1 \right) = e^{-1} \end{aligned}$$

Example, Unbounded X and Complement Example

Show that

$$f(x) = \begin{cases} e^{-x} & \text{if } 0 \leq x \\ 0 & \text{otherwise} \end{cases}$$

is a PDF. Compute the probability of $X > 1$.

The take away is that both work and the laws of probability work the same way.

Here's a detail that can be used for continuous random variables.

$$P(X \leq 1) = P(X = 1) + P(X < 1) = 0 + P(X < 1)$$

Here we are breaking up disjoint sets and using the fact that the probability of observing any single value is zero.

Example, Another Complement Example

For the random variable X which has the follow PMF

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

Compute the probability that $0 > X$ or $1 < X$.

$$\begin{aligned} \{0 > X \text{ or } 1 < X\}' &= (\{0 > X\} \cup \{1 < X\})' \\ &= \{0 > X\}' \cap \{1 < X\}' \\ &= \{0 \leq X\} \cap \{1 \geq X\} \\ &= \{0 \leq X \leq 1\} \end{aligned}$$

Example, Another Complement Example

For the random variable X which has the follow PMF

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

Compute the probability that $0 > X$ or $1 < X$.

$$\begin{aligned} P(0 > X \text{ or } 1 < X) &= 1 - P(0 \leq X \leq 1) \\ &= 1 - \int_0^1 f(x)dx = 1 - \int_0^1 \frac{1}{2}e^{-x}dx \\ &= 1 - \frac{1}{2}\left(-e^{-x}\Big|_{x=0}^{1} \right) \\ &= 1 - \frac{1}{2}\left(-e^{-1} + 1 \right) = \frac{1 + e^{-1}}{2} \end{aligned}$$

The ratio of served and leftover food is uniformly distributed on the interval $[0,1]$. Compute the probablity that less than a quarter of the of food remains given that half of the food is leftover.

$X \sim unif(0,1)$

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

Condition the event $\{X < 0.25\}$ on the event $\{X < 0.5\}$.

$$P(X < 0.25 | X < 0.5) = \frac{P(X < 0.25, X < 0.5)}{P(X < 0.5)}$$

$$\text{(use containment)} = \frac{P(X < 0.25)}{P(X < 0.5)} = \frac{\int_{-\infty}^{0.25} f(x)dx}{\int_{-\infty}^{0.5} f(x)dx}$$

$$= \frac{\int_0^{0.25} 1dx}{\int_0^{0.5} 1dx} = \frac{0.25}{0.5} = 0.5$$

Summary

- Sets are typically written in terms of inequalities.
- Complement and Conditioning work the same way.
- Unions and Intersections are useful to know, but, in this class, it is easier to reduce the set to simpler sets. For example:

$$\{X < 0\} \cup \{X < 1\} = \{X < 1\}$$

$$\{X < 0\} \cap \{X < 1\} = \{X < 0\}$$