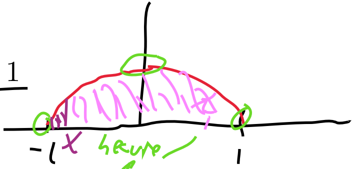
CSCI 3022 Intro to Data Science

nb09, 10.

Example: A random variable X has pdf:

$$\boxed{f(x) = \frac{3}{4}(1 - x^2);} \quad -1 \le X \le 1$$



light   heavy   light

2 in long

Review: What is $F(x)$? input x-value returns "area up to x"

$$= \int_{-1}^{x} \text{density} = \int_{-1}^{x} \frac{3}{4}(1-a^2)\,da = \int_{-1}^{x} \frac{3}{4} - \frac{3}{4}a^2\,da$$

$$= \frac{3}{4}\cdot a^1 - \frac{3}{4}\cdot\frac{a^3}{3}\Big|_{-1}^{x}$$

What is $E[X]$? : center of mass

input random variable

output "center" of randomness

$E[X] = 0$ by symmetry. or check: $\int_{-1}^{1} x \cdot \text{density} = 0 = \int_{-1}^{1} x \cdot \frac{3}{4}(1-x^2)\,dx$

$= 0$ b/c odd

CSCI 3022 Intro to Data Science

new r.v.!

$Y = X^3$

Example: A random variable X has pdf:

$$f(x) = \frac{3}{4}(1 - x^2); \quad -1 \le X \le 1$$

X has outcomes

x	density $(3/4)(1-x^2)$
outcome x^3	"probability" $(3/4)(1-x^2)$

Review: What is $F(x)$?

$$F(x) = \int_{-1}^{x} f(t)\, dt = \frac{3t}{4} - \frac{3t^3}{12}\Big|_{-1}^{x} = \frac{3x}{4} - \frac{3x^3}{12} - \left(\frac{-3}{4} - \frac{3(-1)^3}{12}\right)$$

What is $E[X]$?

$$E(X) = \int_{-1}^{1} x\frac{3}{4}(1 - x^2)\, dx = \frac{3x}{4} - \frac{3x^3}{12}\Big|_{-1}^{1} = 0$$

$E[X^3] = \int$ or $\sum$ of $x^3 \cdot$ density

What is $E(X^3)$?

E[if we take random X's & cube them].

CSCI 3022 Intro to Data Science

Example: A random variable X has pdf:

$$f(x) = \frac{3}{4}(1 - x^2); \quad -1 \le X \le 1$$

Review: What is $F(x)$?

What is $E[X]$?

$$E(X) = \int_{-1}^{1} x\frac{3}{4}(1 - x^2)\, dx = \frac{3x}{4} - \frac{3x^3}{12}|_{-1}^{1} = 0$$

What is $E(X^3)$?

$$E(X^3) = \int_{-1}^{1} x^3\frac{3}{4}(1 - x^2)\, dx = \frac{3x^4}{16} - \frac{3x^6}{24}|_{-1}^{1} = 0$$

Mean/Expected Value

Definition: *Expected Value:*

For a random variable X with pdf $f(x)$, the *expected* value or *mean* value of X is denoted as E(X) and is calculated as:

if X is discrete. If X is continuous, this becomes:

One way to interpret expected value of a discrete distribution (especially on a finite support) is the *sample mean* if we managed to observe observations that *exactly* mirror the probability mass function.

The mean is the "center of mass" of a density object.

Mean/Expected Value

Definition: *Expected Value:*
For a random variable X with pdf $f(x)$, the *expected* value or *mean* value of X is denoted as E(X) and is calculated as:

$$E[X] = \sum_{x \in \Omega} x \cdot P(X = x)$$

if X is discrete. If X is continuous, this becomes:

$$E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$$

One way to interpret expected value of a discrete distribution (especially on a finite support) is the *sample mean* if we managed to observe observations that *exactly* mirror the probability mass function.
The mean is the "center of mass" of a density object.

Mean/Expected Value

Example:

Consider a university having 15,000 students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01,	.03	.13	.25	.39	.17	.02

What is $E[X]$?

Mean/Expected Value

Example:

Consider a university having 15,000 students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01,	.03	.13	.25	.39	.17	.02

What is $E[X]$?

$$E[X] = \sum_{x \in \Omega} x \cdot P(X = x) = 1 \cdot .01 + 2 \cdot .03 + 3 \cdot .13 + 4 \cdot .25 + 5 \cdot .39 + 6 \cdot .17 + 7 \cdot .02$$

$$E[X] = 4.57$$

Expected Value of a Function

$\nearrow f(x)$

If a discrete r.v. X has a a density $P(X = x)$, then the expected value of any function $\underline{g(X)}$ is computed as:

1. Continuous: $\int$ outcome $\cdot$ probability

   $= \int_{\text{outcomes of } x} g(x) \cdot \text{density} = \int_x g(x) \cdot f(x) \cdot dx$

2. Discrete:

   $\sum_{\substack{\text{all } x \\ \text{outcomes}}} \underbrace{g(x)}_{\text{"transformed outcome"}} \cdot \overbrace{P(X=x)}^{\text{pmf/pdf/}f(x)}$

Note that $E[g(X)]$ is computed in the same way that $E(X)$ itself is, except that $g(x)$ is substituted in place of x.

Expected Value of a Function

If a discrete r.v. X has a a density $P(X = x)$, then the expected value of any function $g(X)$ is computed as:

1. Continuous:

$$E[X] = \int_{\infty}^{\infty} x f(x) \; dx$$

2. Discrete:

$$E[X] = \sum_{x} x f(x) \; dx$$

Note that $E[g(X)]$ is computed in the same way that $E(X)$ itself is, except that $g(x)$ is substituted in place of x.

Expected Value of a Linear Function

linear: $y = mx + b$.

If $g(X)$ is a linear function of X (i.e., $g(X) = \boxed{aX + b}$) then $E[g(X)]$ can be easily computed from $E(X)$.

$\int (ax+b)\,dx$

$= a\int x\,dx + \int b\,dx$

$\sum_{x=1}^{n} (ax+b)$

Theorem:
Let $a, b \in \mathbb{R}$ and X be a random variable with density f. Then:

$$E[aX+b] = a\,E[X] + b.$$

average of "X times a + b"

a · "average of X"

plus b.

Proof:

Note: This works for continuous and discrete random variables.

Expected Value of a Linear Function

If $g(X)$ is a linear function of X (i.e., $g(X) = aX + b$) then $E[g(X)]$ can be easily computed from $E(X)$.

$\int f(x)dx = 1$
over all
x.
is a pdf

Theorem:
Let $a, b \in \mathbb{R}$ and X be a random variable with density f. Then:

$$E[g(X)] = g(E[X])$$

$$E[aX + b] = aE[X] + b$$

Proof:
$E[aX + b] = \int (ax + b)f(x)\, dx \overset{\text{split integral}}{=} a \int xf(x)\, dx + b \int f(x)\, dx = aE[X] + b$, since integration is also linear!
Note: This works for continuous and discrete random variables.

Course load : random = X

Linear Expectation

tuition = $500 \cdot \underbrace{\boxed{\text{courses}}}_{X} + 100 = aX + b$

Example:
Consider a university having 15,000 students and let X equal the number of courses for which a randomly selected student is registered.
The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01,	.03	.13	.25	.39	.17	.02

Earlier, we calculated that $E(X)$ was 4.57. If students pay \$500 per course plus a \$100 per-semester registration fee, what is the average amount of money the university can expect a student to pay each a semester?

$$E[500 \cdot X + 100] = 500 \underbrace{E[X]}_{\text{avg. courses}} + 100$$

Linear Expectation

Example:

Consider a university having 15,000 students and let X equal the number of courses for which a randomly selected student is registered.

The pdf of X is given to you as follows:

x	1	2	3	4	5	6	7
$f(x) = P(X = x)$	.01,	.03	.13	.25	.39	.17	.02

Earlier, we calculated that $\boxed{E(X)}$ was 4.57. If students pay $500 per course plus a $100 per-semester registration fee, what is the average amount of money the university can expect a student to pay each a semester?

$Money = 500 \cdot Courses + 100 = 500X + 100 = g(X)$. Then,

NOT X^2
X^3
X^J . . $\sin(X)$.

$$E[g(X)] = g(E[X]) = 500 \cdot 4.57 + 100 = 2385.$$

Ënote this only works for linear operations

Expectation and Spread

The idea of **Expected value** can be extended to describe all kind of notions of "what should happen if we have a (arbitrarily large) sample.

> Suppose we wish to know the variance or standard deviation of the population. For a *sample*, recall that

x from mean(x)

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n-1}$$

Expectation and Spread

The idea of **Expected value** can be extended to describe all kind of notions of "what should happen if we have a (arbitrarily large) sample.

Suppose we wish to know the variance or standard deviation of the population. For a *sample*, recall that

$$s = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n-1}$$

Another way: sample variance is $\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}}_{\text{averaged out}} \underbrace{\left(X_i - \bar{X}\right)^2}_{\text{squared deviations}}$

Expectation and Spread

The idea of **Expected value** can be extended to describe all kind of notions of "what should happen if we have a (arbitrarily large) sample.

Suppose we wish to know the variance or standard deviation of the population. For a *sample*, recall that

$$s = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n-1}$$

Another way: sample variance is $\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}}_{\text{averaged out}} \underbrace{\left(X_i - \bar{X}\right)^2}_{\text{squared deviations}}$

We might ask: what is the *expected* value of how spread out x-value are?

Population variance is this idea expressed as an expectation:

Expectation and Spread

The idea of **Expected value** can be extended to describe all kind of notions of "what should happen if we have a (arbitrarily large) sample.

Suppose we wish to know the variance or standard deviation of the population. For a *sample*, recall that

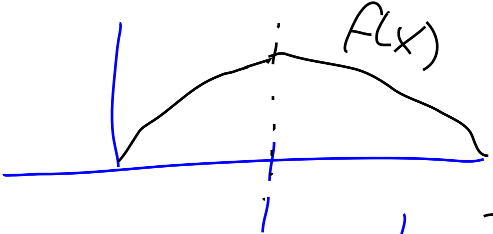
$$s = \frac{\sum_{i=1}^{n} (x_i - \bar{x})}{n-1}$$

Another way: sample variance is $\underbrace{\frac{1}{n-1}\sum_{i=1}^{n}}_{\text{averaged out}} \quad \underbrace{\left(X_i - \bar{X}\right)^2}_{\text{squared deviations}}$

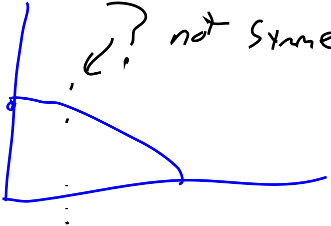
Population variance is this idea expressed as an expectation:

$$Var[X] = E[\underbrace{(X - E[X])^2}_{\text{squared deviations}}] = E[(X - \mu_X)^2]$$

Daily Recap



E[X] = 1



not symmetry

Today we learned

1. Expectation

Moving forward:

- nb day Friday!

Next time in lecture:

- Expected dispersion/spread: calculating variances from pdfs!