5/16/19 Lecture 14

X discrete pmf $f_X(x) \rightarrow E(X) = \mu = \sum_{\text{all } x \text{ in } S} x P(X = x)$

Support S_X

$\uparrow$ mean $\quad \uparrow$ "mew"

$$= \sum_{\substack{\text{all } x \\ \text{in } S_X}} x f_X(x)$$

X continuous, PDF $f_X(x)$

Support $S_X \rightarrow E(X) = \mu = \int_{S_X} x f_X(x)\, dx$

Expectation of a product when the X_i are independent

$X_1, \ldots, X_n$ n independent rv each with finite $E(X_i) \rightarrow$

$$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$$

Contrast this with a sum: $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$

whether the X_i are independent or not;

$E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)$ only when the X_j are independent

Ex: You have a (Brita) water filter that you use to improve the taste of S.C. water. How much better would the filter do its job if you filtered the water twice instead of once?

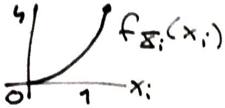
X_1 = proportion of bad stuff removed in the 1st filtering

X_2 = proportion removed in 2nd filtering

Reasonable to assume that X_1, X_2 are independent

Suppose they're IID with common PDF

$$f_{X_i}(x_i) = \begin{cases} 4x^3 & 0 < x < 1 \\ 0 & \text{else} \end{cases}$$



Set Y = proportion of original bad stuff remaining after 2 filtrations $= (1-X_1)(1-X_2)$

Then $E(Y) = E[(1-X_1)(1-X_2)] \overset{\text{independence}}{=} E[(1-X_1)] \cdot E[(1-X_2)]$

X_1, X_2 independent $\rightarrow (1-X_1)(1-X_2)$ independent too

$E(1-X_1) \underset{\substack{\uparrow \\ \text{identical distribution}}}{=} E(1-X_2) \triangleq \mu$ then $E(Y) = \mu^2$

$$\mu = E(1-X_i) = \int_0^1 (1-x_i)\, 4x_i^3\, dx_i = 0.2$$

So 80% of bad stuff expected to be removed in 1st filtering

$E(Y) = \mu^2 = 0.04$ so expect only 4% of bad stuff to remain after 2 filterings

6 a) Suppose X is a discrete rv with possible values 0, 1, 2.... then $E(X) = \sum_{x=0}^{\infty} P(X \geq x)$

b) If X is a continuous rv w/ possible values $(0, \infty)$ then
$$E(X) = \int_0^{\infty} [1 - F_X(x)]\, dx$$

Ex of 6a

I throw a dart at a dart board repeatedly, trying to get a bullseye (success)

X = # of throw on which I 1st succeed

(Ex throws FFS $\to X = 3$ F = failure S = Success)

Suppose that my success probability is constant across the throws and equals p, and throws are independent

Then $E(X)$ should be inversely related to p:
The worse I am, the longer I expect the 1st bullseye to take. $E(X) = ?$

Geometric distribution

At least 1 throw always required so $P(X \geq 1) = 1$ for $n \geq 1$

(at least n tosses required) $\leftrightarrow$ (none of the 1st $(n-1)$ throws succeeded)

so $P(X \geq n) = (1-p)^{n-1}$ and

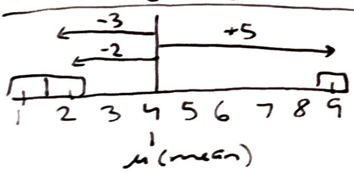
$$E(X) = \sum_{n=1}^{\infty} (1-p)^{n-1} = 1 + (1-p) + (1-p)^2 + \ldots = \frac{1}{1-(1-p)} = \frac{1}{p}$$

geometric series

(inverse relation)

If I'm terrible (e.g. $p = .01$) I expect to succeed on the $\frac{1}{.01} = 100$th throw

Variance and Standard Deviation



$\begin{bmatrix} \$1 \\ \$2 \\ \$9 \end{bmatrix}$ mean $\$4 = \mu$

X discrete rv, Uniform $\{1, 2, 9\}$

$E(X) = 4 = \mu$

Q: How spread out is the dist. of X around its mean μ?

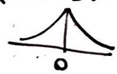
$(X - \mu) \sim$ Uniform $\{-3, -2, +5\}$

↑ deviation from μ

Could try calculating $E(X - \mu)$, but this is 0 for any rv X because of cancellation of $\oplus$ and $\ominus$ deviations

Two different easy fixes:

$E|X - \mu| \triangleq$ mean/average absolute deviation (AAD) (MAD)



doesn't describe reality often

or

$E(X - \mu)^2 \triangleq$ variance of rv X

does describe reality

AAD: not used much

Variance: used constantly

Def: X rv with finite mean $E(X) = \mu$,

Variance of $X = V(X) \triangleq E[(X - \mu)^2]$

If $E(X) = \pm\infty$ or $E(X)$ doesn't exist, $V(X)$ doesn't exist

One problem with variance

The units are wrong: If X is in \$, $V(X)$ is in $\$^2$

Easy fix: standard deviation of $X \triangleq \sqrt{V(X)} \triangleq SD(X)$

Consequences of these definitions

1) $V(X) = E[(X-\mu)^2] = E(X^2 - 2\mu X + \mu^2)$

$$= E(X^2) - 2\mu \underbrace{E(X)}_{\mu} + \mu^2$$

$$= E(X^2) - \mu^2$$

$$= E(X^2) - [E(X)]^2$$

This is a different way to compute variance

So $V(X) = \left(\text{expectation of } X^2\right) - \left(\text{square of expectation of } X\right)$

Toy example

$\begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix}$ mean $\mu = 4$

$X \sim \text{Uniform}\{1, 2, 9\}$

$E(X-\mu)^2 = \frac{1}{3}(1-4)^2 + \frac{1}{3}(2-4)^2 + \frac{1}{3}(9-4)^2 = 12.7$

So $SD(X) = \sqrt{12.7} = 3.6$

This is a reasonable summary of the length of the arrows

2) For any rv X, $V(X) \geq 0$

If X is bounded, $V(X)$ exists and is finite

This is a consequence of Jensen's Inequality:

$g(x) = x^2$ is convex so $E(X^2) \geq [E(X)]^2$

i.e. $V(X) = E(X^2) - [E(X)]^2 \geq 0$

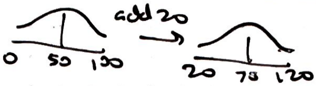
3) $V(X)=0 \overset{iff}{\longleftrightarrow} P(X=c)=1$ for some constant c
(this is a trivial rv)

Notation

In the same way that by convention $E(X)=\mu_X$,
$V(X) \triangleq \sigma_X^2$ and $SD(X) \triangleq \sigma_X$

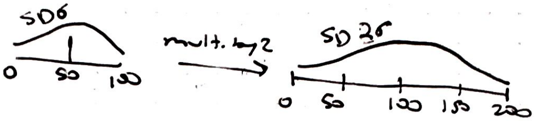
4) X rv, $Y=aX+b \rightarrow V(Y)=a^2V(X)=a^2\sigma_X^2$
and $SD(Y)=|a|\sigma_X$ (for any constants a, b)

$V(Y)=V(aX+b)$



$V(X+c)=V(X)$

$SD(cX)=\ ?$



$E(cX)=c\,E(X)$ $\quad SD(cX)=c\,E(X)$

$V(cX)=c^2V(X)$

$V(Y)=V(aX+b)=a^2V(X)^3$

Special cases $a=1$: $V(X+c)=V(X)$
$SD(X+c)=SD(X)$

$V(aX)=a^2V(X)$
$(b=0)$ $\quad SD(aX)=|a|\,SD(X)$

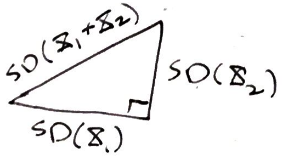
5) If $X_1, \dots, X_n$ are independent rv with finite means,
$$V\left(\sum_{i=1}^{n} X_i\right)=\sum_{i=1}^{n} V(X_i)$$

This is why the concept of variance has endured even though the units of the variance are wrong: for independent rvs, variance is additive whereas SD is not
correct units

X_1, X_2 independent $\rightarrow V(X_1 + X_2) = V(X_1) + V(X_2)$

$$\sqrt{V(X_1 + X_2)} = \sqrt{V(X_1) + V(X_2)}$$

$$SD(X_1 + X_2) = \sqrt{[SD(X_1)]^2 + [SD(X_2)]^2}$$



SD grows like the hypotenuse of a right Δ

Immediately, $\max\left\{ \begin{matrix} SD(X_1), \\ SD(X_2) \end{matrix} \right\} < \underset{\text{(indep)}}{SD(X_1 + X_2)} < SD(X_1) + SD(X_2)$

Consequence of 5

$X_1, \dots, X_n$ independent rv, $a_1, \dots, a_n$, b constants $\rightarrow$

$$V\left[\left(\sum_{i=1}^{n} a_i X_i\right) + b\right] = \sum_{i=1}^{n} a_i^2 V(X_i)$$

Ex: $X \sim$ Binomial (n, p)

We know that $E(X) = np$

What about $V(X)$ and $SD(X)$?

Let $S_i = \begin{cases} 1 & \text{if success on } i\text{th success/failure trial} \\ 0 & \text{else} \end{cases}$

for $(i = 1, \dots, n)$ and suppose as usual that $S_1, \dots, S_n$ are IID Bernoulli (p) — then $X = \sum_{i=1}^{n} S_i$ and we can work out its variance without difficulty

$$V(X) = V\left(\sum_{i=1}^{n} S_i\right) \underset{\uparrow \text{ independence}}{=} \sum_{i=1}^{n} V(S_i)$$

So we need to work out the variance of a Bernoulli rv. We already know that $E(S_i) = p$, so if we use the formula $V(S_i) = E(S_i^2) - [E(S_i)]^2$ we're halfway there.

Bernoulli rvs are funny:

$$S_i = \left\{ \begin{array}{lll} 1 & \text{with probability} & p \\ 0 & \text{"} \quad \text{"} & (1-p) \end{array} \right\}$$

so $$S_i^2 = \left\{ \begin{array}{lll} 1^2 = 1 & \text{with probability} & p \\ 0^2 = 0 & \text{"} \quad \text{"} & (1-p) \end{array} \right\}$$

so $E(S_i^2) = E(S_i) = p$

$V(S_i) = E(S_i^2) - [E(S_i)]^2 = p - p^2 = p(1-p)$

$V(X) = \sum_{i=1}^{n} V(S_i) = \sum_{i=1}^{n} p(1-p) = n\,p(1-p)$

$SD(X) = \sqrt{n\,p(1-p)}$

<u>Ex</u>: T-S disease

Y = (# of babies in family of $n = 5$ both parents carriers so $p = P(\text{T-S baby}) = \frac{1}{4}$)
$\sim$ Binomial (n, p) = Binomial $(5, \frac{1}{4})$

We already worked out that $E(Y) = np = 1.25$

Now $SD(Y) = \sqrt{np(1-p)} = \sqrt{5(\frac{1}{4})(\frac{3}{4})} = 0.97 \simeq 1$

It's useful to summarize this by saying "The # of T-S babies this couple will have will be around 1.25, give or take about 1" $\leftarrow \sigma_Y$

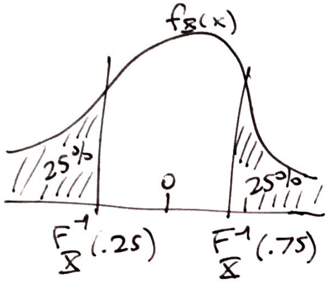
$\uparrow$
μ_Y

How do you measure the spread of a distribution if the variance doesn't exist?

Ex: $X \sim$ Cauchy (standard) $\quad f_X(x) = \dfrac{1}{\pi(1+x^2)}$ for all $-\infty < x < \infty$

We saw that $E(X)$ doesn't exist, so $V(X)$ doesn't exist either.

But we can use the idea of quantiles on any dist. whether its variance exists or not



We defined the interquartile range (IQR) as

$IQR = F_X^{-1}(.75) - F_X^{-1}(.25)$

Standard Cauchy CDF is $F_X(x) = \displaystyle\int_{-\infty}^{x} \frac{1}{\pi(1+t^2)}\,dt$

$$= \frac{1}{2} + \frac{\tan^{-1}(x)}{\pi}$$

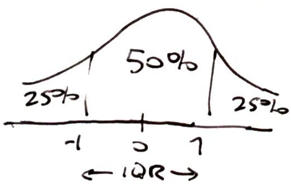
Need to solve $F_X(x) = \frac{1}{2} + \frac{\tan^{-1}(x)}{\pi} = p$ for x

result is $x = F_X^{-1}(p) = \tan\left(\dfrac{p-\frac{1}{2}}{\pi}\right) = -\cot(p\pi)$

So the IQR for the standard Cauchy dist. is

$IQR = F_X^{-1}\left(\frac{3}{4}\right) - F_X^{-1}\left(\frac{1}{4}\right) = \tan\left(\frac{\pi}{4}\right) - \tan\left(-\frac{\pi}{4}\right) = 2$

Standard Cauchy PDF



Moments of a rv

$E(X) = E(X^1)$

$$V(X) = E(X^2) - [E(X^1)]^2 = E(X - \underset{\uparrow}{\overset{E(X^1)}{\mu}})^2$$

With the usual mathematical impulse to generalize

Def: X rv, k integer ≥ 1 $\to$ $E(X^k) \triangleq$ the kth moment of X

of course $E(X^k)$ may not exist, and if it does it may be infinite, but the idea is still useful

You can show that (kth moment of X exists) $\leftrightarrow$ $E(|X|^k) < \infty$

Consequences of the moment definition

1) If $E(|X|^k) < \infty$ for some integer $k \geq 1$, then $E(|X|^j) < \infty$ for all integers $j < k$

If the kth moment of X exists, so do the $(k-1)^{st}$, $(k-2)^{nd}$, ... moments

Def: X rv with expectation $E(X) = \mu$, k integer ≥ 1

$\to E[(X-\mu)^k]$ is called the k^{th} central moment of X or the k^{th} moment of X around its mean.

This generalizes the variance of $X = E((X-\mu)^2)$

2) $E[(X-\mu)^1] = E(X) - \mu = \mu - \mu = 0$

ie every rv has 1st central moment 0

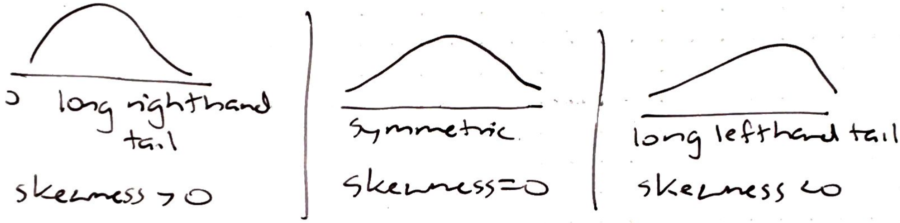
If the dist. of X is symmetric around μ_X, then $E[(X-\mu)^k]=0$ for all odd integers k for which $E[(X-\mu)^k]$ exists

<u>Def:</u> X rv with mean μ_X SD σ_X

If the third moment of X exists and is finite, then

$$\text{skewness}(X) \triangleq E\Bigg(\underbrace{\frac{X-\mu_X}{\sigma_X}}_{\text{converting } X \text{ to standard units}}\Bigg)^3$$

All symmetric distributions with finite 3rd moment have skewness 0



Moment generalizing functions

<u>Def:</u> X rv, t: real #

$\psi_X(t) \triangleq E(e^{tX})$ is called the moment generating function of X (MGF)

psi

<u>Thm:</u> X rv with MGF $\psi_X(t)$, finite for all values of t in an open interval $(-a,b)$ around 0 $(a>0\ b>0)$, then for all integers $n>0$

$-a \quad 0 \quad +b$

$$E(X^n)=\frac{d^n}{dt^n}\psi_X(t)\Big|_{t=0}$$

$\leftarrow$ nth derivative of ψ_X, evaluated at $t=0$

If its premise is satisfied and the calculations are manageable, you get <u>all</u> the moments of $\underline{X}$, just by computing $\psi_{\underline{X}}(t)$ and differentiating it over and over.