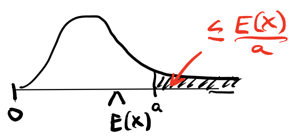
Stat 134, lec 12

Last time

Sec 3.2 Markov's inequality

If $X \geq 0$, then $P(X \geq a) \leq \frac{E(X)}{a}$ for every $a>0$.

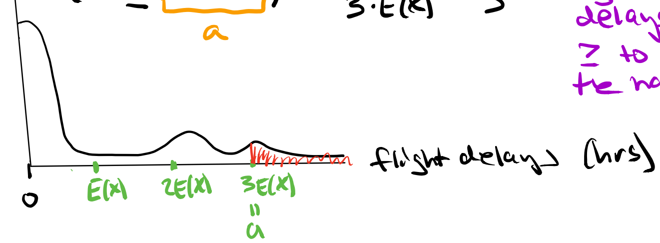
Picture



ex Is it possible that half of all US flights have delay times greater than 3 or more times the national average?

$$P(X \geq \underbrace{3\cdot E(X)}_{a}) \leq \frac{E(X)}{3\cdot E(X)} = \frac{1}{3}$$

NO at most $\frac{1}{3}$ of flight delays can be $\geq$ to 3 times the national avg.



Discrete Distributions

1. Ber(p)
2. Bin(n,p)
3. hyper(N,G,n)
4. Pois(μ)
5. Unif $\{1,\dots,n\}$
6. Geom(p) on $\{1,2,\dots\}$

Geometric RV

trials until first success

ex X = number of p coin tosses until your first heads

$X=1$	H	p
$X=2$	TH	qp
$X=3$	TTH	q^2p
$\vdots$		

$P(X=k) = q^{k-1}p$ Geom(p) formula on $\{1,2,\dots\}$

Note trials are independent

Today

1. review student responses to concept test
2. Sec 3.2 $E(g(X,Y))$
3. Sec 3.3 $SD(X)$, $Var(X)$ Chebyshev inequality.

(1) concept test responses.

2. Consider a well shuffled deck of cards. The expected number of cards between the first and second ace (not counting either ace)?

a 52/5

b 48/5

c 48/4

d none of the above

b	Think of slots: _ A _ A _ A _ A _ --> 5 slots. Every non-ace card had a 1/5 chance of being before the first ace ==> 48* 1/5 = 48/5 = B

c	spread out aces over 48 non-aces, means on average there should be one after every ~~12~~ theoretically

$\frac{48}{5} = 9.6$

c	because expected value of geometric is ~~1/p~~ $\frac{1-p}{p}$

$P =$ prob you draw an ace $= \frac{4}{52}$

$X =$ # nonaces before first ace (if deal cards with replacement)

$X \sim \text{Geom}(p)$ on $0, 1, 2, \ldots$

Fact (sec 3.4) $E(X) = \frac{1-p}{p} = \frac{1-\frac{4}{52}}{\frac{4}{52}} = \frac{\frac{48}{52}}{\frac{4}{52}} = \frac{48}{4}$

② <u>Sec 3.2</u> Expectation of a function of a RV.

$$E(X) = \sum_{x \in X} x P(X=x)$$

$$E(g(X)) = \sum_{x \in X} g(x) P(X=x)$$

<u>ex</u> Suppose $X \sim \text{Geom}(p)$ on $\{1, 2, \cdots\}$ with $p > 2/3$

Find $E(3^X)$.

<u>Picture</u>

$$\begin{array}{lll} p & 1 & \xrightarrow{y=3^x} 3^1 \\ qp & 2 & \longrightarrow 3^2 \\ q^2p & 3 & \longrightarrow 3^3 \\ & \vdots & \end{array}$$

$$E(3^X) = \sum_{k=1}^{\infty} 3^k P(X=k) = \sum_{k=1}^{\infty} 3^k q^{k-1} p$$

$$= 3p + 3^2 qp + 3^3 q^2 p + \cdots$$

$$= 3p\Big(\underbrace{1 + 3q + (3q)^2 + \cdots}_{= \frac{1}{1-3q} \text{ if } 3q < 1}\Big)$$

yes since $p > 2/3$

$$\Rightarrow \boxed{E(3^X) = 3p\left(\frac{1}{1-3q}\right)}$$

ex $X \sim \text{Pois}\left(\frac{1}{3}\right)$

Find $E(X!)$

$$\left(\begin{array}{l} X \sim \text{Pois}(\mu) \\ P(X=k) = \dfrac{e^{-\mu}\mu^k}{k!} \\ k=0,1,2,\dots \end{array} \right)$$

Soln

$$E(X!) = \sum_{k=0}^{\infty} \cancel{k!}\, P(X=k) = \sum_{k=0}^{\infty} \cancel{k!} \frac{e^{-\frac{1}{3}} \left(\frac{1}{3}\right)^k}{\cancel{k!}} = e^{-\frac{1}{3}} \sum_{k=0}^{\infty} \left(\frac{1}{3}\right)^k$$

$$= e^{-1/3}\Big[\underbrace{1 + \tfrac{1}{3} + \left(\tfrac{1}{3}\right)^2 + \cdots}_{= \frac{1}{1-1/3} = \frac{3}{2}}\Big] = \boxed{\frac{3}{2} e^{-1/3}}$$

Several variables

(X, Y) joint distribution

$$E(g(X)) = \sum_{\text{all } x} g(x) P(X = x)$$

$$E(g(X,Y)) = \sum_{\text{all } x,y} g(x,y) P(X = x, Y = y)$$

Ex (X, Y) joint distribution

	$\frac{1}{2}$	$\frac{1}{2}$	$P(x)$
Y			$P(y)$
2	0	$\frac{1}{4}$	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$
0	$\frac{1}{4}$	0	$\frac{1}{4}$
X	0	1	

Find $E(X)$:

$$E(X) = \sum_{\text{all } x,y} x P(X = x, Y = y) = 0 \cdot \tfrac{1}{4} + 0 \cdot \tfrac{1}{4} + 0 \cdot 0 + 1 \cdot 0 + 1 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{4} = \tfrac{1}{2}$$

$g(X,Y) = X$

Thm $E(X+Y) = E(X) + E(Y)$

Pf/ $E(X) = \sum_{\text{all } x,y} x P(X=x, Y=y)$

$E(Y) = \sum_{\text{all } x,y} y P(X=x, Y=y)$

$E(X+Y) = \sum_{\text{all } x,y} (x+y) P(X=x, Y=y)$

$= \underbrace{\sum_{\text{all } x,y} x P(X=x, Y=y)}_{= E(X)} + \underbrace{\sum_{\text{all } x,y} y P(X=x, Y=y)}_{= E(Y)}.$ $\square$

ex Thm if X and Y are independent
$E(XY) = E(X)E(Y)$

a) Write the formula for $E(XY)$

$E(XY) = \sum_{\text{all } x,y} xy P(X=x, Y=y)$

b) $E(X)E(Y) = \sum_{\text{all } x} x P(X=x) \sum_{\text{all } y} y P(Y=y)$

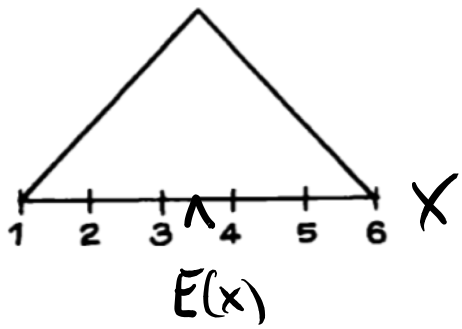
finish the proof.

$= \sum_{\text{all } x} \sum_{\text{all } y} xy P(X=x) P(Y=y) = \underbrace{\sum_{\text{all } x} x P(X=x)}_{E(X)} \underbrace{\sum_{\text{all } y} y P(Y=y)}_{E(Y)}$

③ <u>Sec 3.3</u> <u>Standard deviation</u> (SD)

SD is the average spread of your data around the mean.

What is the SD of the following figure?



a 0.5 — too small

b 1 — about right (I am not expecting you to calculate this)

c 2 — too large

$$SD(x) = \sqrt{E\left((x - E(x))^2\right)}$$

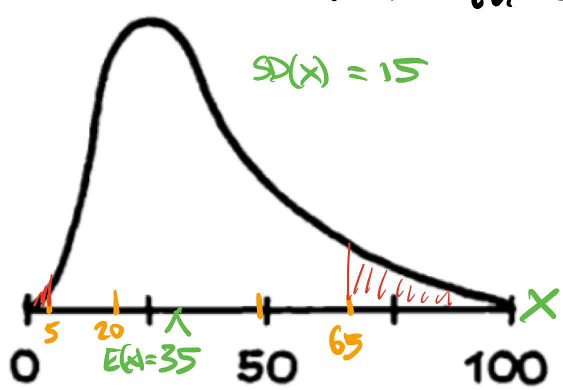
$$Var(x) = (SD(x))^2 = E\left((x - E(x))^2\right)$$

Chebyshev's Inequality

For any random variable X, and any $k>0$,

$$P\left(|X - E(X)| \geq k \cdot SD(X) \right) \leq \frac{1}{k^2}$$

ex Let X have distribution with $E(X)=35$, $SD(X)=15$.



Find $P(|X-35| \geq 30)$?

$$\leq \frac{1}{2^2} = \frac{1}{4}$$

What can you say about $P(X \geq 65) \leq \frac{1}{4}$

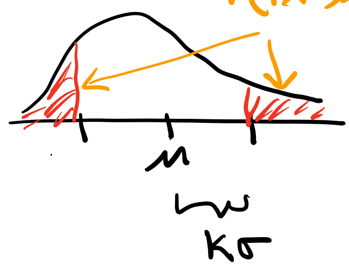
$$\mu = E(X)$$

$$\sigma = SD(X).$$

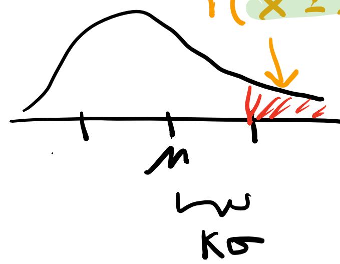
Summary of Tail bounds

Chebyshev's inequality

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

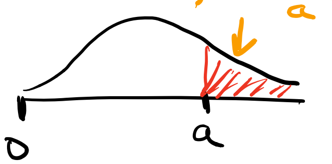


$$P(X \geq \mu + k\sigma) \leq \frac{1}{k^2}$$

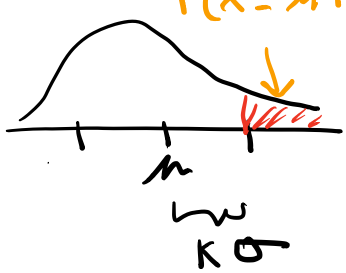


Markov's inequality

$$P(X \geq a) \leq \frac{\mu}{a}$$



$$P(X \geq \mu + k\sigma) \leq \frac{\mu}{\mu + k\sigma}$$



Stat 134

Wednesday February 20 2019

1. A list of non negative numbers has an average of 1 and an SD of 2. Let p be the proportion of numbers over 5. To get an upper bound for p, you should:

   a Assume a normal distribution

   b Use Markov's inequality $P(x \geq 5) \leq \frac{1}{5}$ smaller

   c Use Chebyshev's inequality $P(x \geq 5) \leq \frac{1}{2^2} = \frac{1}{4}$

   d none of the above

$\left.\begin{array}{c} \mu + k\sigma \\ \shortparallel \\ 5 = 1 + 2 \cdot 2 \end{array}\right\} \Rightarrow k = 2$