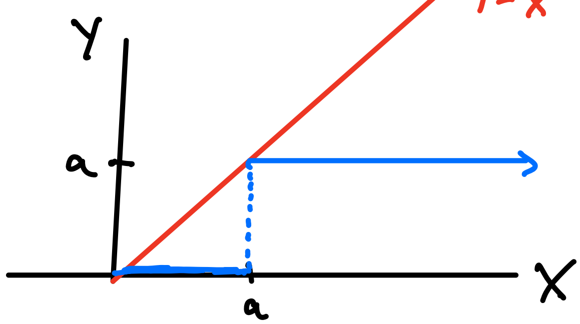
Stat 134 lec 12

Warmup: 11:00–11:10

This question asks you to graph

Let $Y = aI(X \geq a) = \begin{cases} a & \text{if } X \geq a \\ 0 & \text{else} \end{cases}$

$I(X \geq a)$ is an indicator RV and $X > 0$ and $a > 0$



Y=X (diagonal line drawn for convenience)

$$aI(X \geq a) \leq X$$

$$\Rightarrow E(aI(X \geq a)) \leq E(X)$$

$$\| $$

$$a\underbrace{E(I(X \geq a))}_{\substack{\| \\ P(X \geq a)}} \leq E(X)$$

Assumes $X \geq 0$ and $a > 0$

$$\Rightarrow \boxed{P(X \geq a) \leq \frac{E(X)}{a}}$$

Markov's inequality

last time

$E(X) = P(X\geq 1) + P(X \geq 2) + P(X \geq 3) + \dots$ Tail Sum Formula

This is useful when $X =$ min or max,

Discrete Distributions

1. Ber(p)
2. Bin(n,p)
3. HG(n,N,G)
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

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

Note trials are independent

Today

1. Sec 3.2 Markov inequality
2. Sec 3.2 $E(g(X,Y))$
3. Sec 3.3 $SD(X)$, $Var(X)$, Chebyshev's inequality

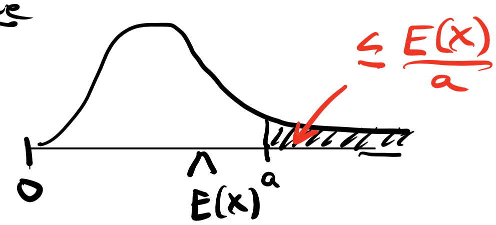
(1) Sec 3.2 Markov Inequality

Proved in warmup

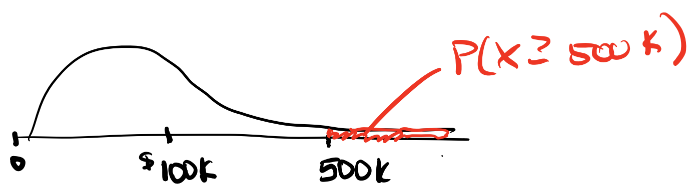
Markov's inequality:

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

Picture

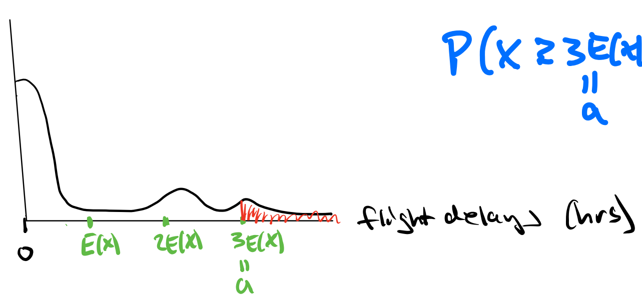


ex Let X be the yearly income of Bay area residents. $E(X) = \$100K$. Find an upper bound for $P(X \geq 500K)$



$$P(X \geq 500k) \leq \frac{100}{500} = \frac{1}{5}$$

ex Give an upper bound for the fraction of all US flights that have delay times greater than 3 or more times the national average.



$$P(X \geq \underset{a}{\underset{\parallel}{3E(X)}}) \leq \frac{\cancel{E(X)}}{3\cancel{E(X)}} = \frac{1}{3}$$

$X \sim \text{Pois}(\mu)$

$P(X=k) = \dfrac{e^{-\mu}\mu^{k}}{k!}$

ex Let $X_1, X_2, \dots, X_{100}$ be independent and identically distributed (iid) $\text{Pois}(.01)$.

Let $S = X_1 + X_2 + \dots + X_{100}$

a) What distribution is S? $\text{Pois}(\underbrace{100(.01)}_{=1})$

b) Find an upperbound for $P(S \geq 3)$ using Markov's inequality.

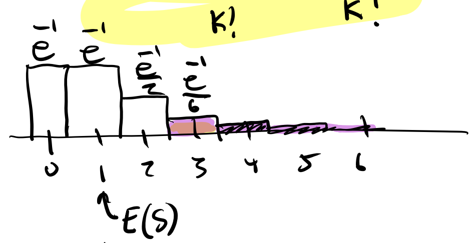
$S \sim \text{Pois}(1)$ $E(S) = 1$

$P(S \geq 3) \leq \boxed{1/3}$

Note Exact: $P(S \geq 3) = 1 - P(0) - P(1) - P(2)$

$\qquad P(0) = e^{-1}, \quad P(1) = e^{-1}, \quad P(2) = \dfrac{e^{-1}}{2}$

$P(S=k) = \dfrac{e^{-1}1^{k}}{k!} = \dfrac{e^{-1}}{k!}$



$= 1 - e^{-1}(1 + 1 + \frac{1}{2})$

$= .08$

② Sec 3.2 Expectation of a function of a RV.

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

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

ex Suppose $X \sim \text{Geom}(p)$ on $\{1, 2, \cdots\}$ with $p > 2/3$.
Find $E(3^X)$.

$X \sim \text{Geom}(p)$
trials to 1st failure
$P(X=k) = q^{k-1} p$

Picture

$y = 3^x$

$$1 \longrightarrow 3^1 \quad p$$
$$2 \longrightarrow 3^2 \quad qp$$
$$3 \longrightarrow 3^3 \quad q^2 p$$
$$\vdots$$

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

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

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

if $3q < 1$

yes since $p > 2/3$

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

Several variables

(X,Y) joint distribution

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

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

see appendix to notes

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

see appendix to notes

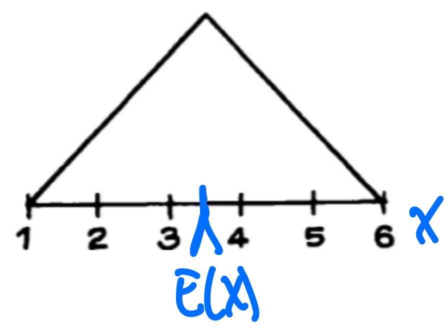
Thm if X and Y are independent

$$E(XY) = E(X)E(Y)$$

③ Sec 3.3 Standard deviation (SD)

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

What is the SD of the following figure?



a 0.5

b 1

c 2

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

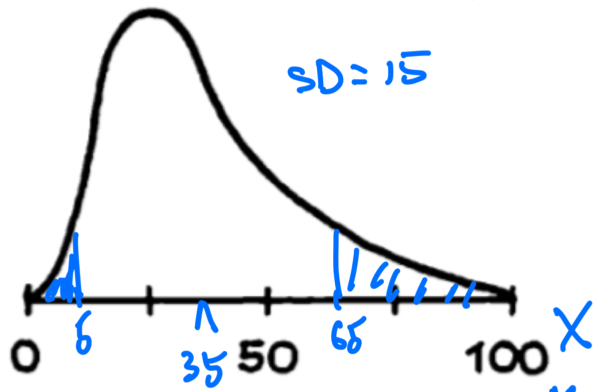
$$Var(x) = (SD(x))^2 = E\left((x-E(x))^2\right)$$

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 $E(X)=35$, $SD(X)=15$.



$k \cdot 15 \Rightarrow \boxed{k=2}$

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

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

What can you say about $P(X \geq 65)$? $\leq 1/4$

$$P(X \geq 65) = P(X \geq \underbrace{\mu + k\sigma}_{= 65}) \leq \frac{1}{2^2}$$

SD

Stat 134

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

   a Assume a normal distribution

   b Use Markov's inequality

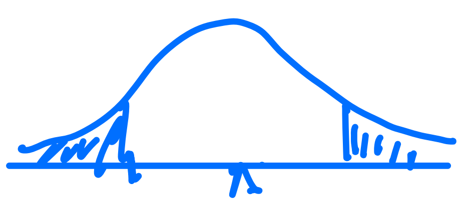
   c Use Chebyshev's inequality

   d none of the above

$P(X \ge 5) \le \frac{1}{5}$

$P(X \ge 5) \le \frac{1}{4}$

$5 = \mu + k \cdot \sigma$, $\mu = 1$, $\sigma = 2$



Proof of Chebyshev

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

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

By Markov's inequality

$$P(Y \geq a) \leq \frac{E(Y)}{a} \quad \text{for } Y \geq 0,\ a > 0$$

$$P\left(|X - E(X)| \geq K\,SD(X)\right)$$

$$\| $$

$$P\Big(\underbrace{(X - E(X))^2}_{=\,Y} \geq \underbrace{(K\,SD(X))^2}_{=\,a}\Big) \overset{\text{Markov}}{\leq} \frac{E(Y)}{a} = \frac{E\left((X - E(X))^2\right)}{K^2 \cdot \cancel{SD(X)^2}} = \boxed{\frac{1}{K^2}}$$

$Var(X) = \cancel{SD(X)^2}$

$\square$

Appendix

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

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

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

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

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

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

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

$= P(X=x)P(Y=y)$ by independence

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

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