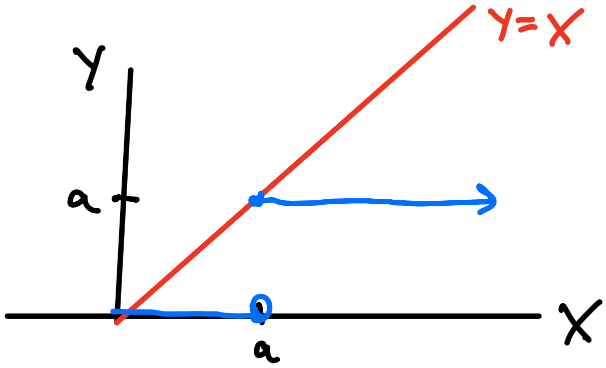
Stat 134 lec 11

Warmup: 10:00–10:10

This question asks you to graph

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

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



$$aI(X \ge a) \le X$$

$$\Rightarrow E(aI(X \ge a)) \le E(X)$$

$$= aE(I(X \ge a))$$

$$= P(X \ge a)$$

$$\Rightarrow \boxed{P(X \ge a) \le \frac{E(X)}{a}}$$

Markov's Inequality

assumes $X \ge 0$, $a > 0$

Announcement: Q2 in section next Wednesday 2/16. Coverage: Sections 2.1, 2.2, 2.4, 2.5, 3.1, 3.2

last time

$E(X) = P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \cdots$ 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,...,n}
6. Geom(p) on {1,2,...}

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	$q^2 p$

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

Note trials are independent

Today

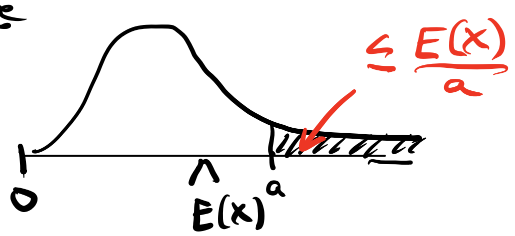
1. Sec 3.2 Markov inequality
2. Sec 3.2 $E(g(X,Y))$
3. Sec 3.3 $SD(X)$, $Var(X)$, Chebyshev's Inequality

① Sec 3.2 **Markov Inequality**

Markov's inequality: Proved in warmup

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

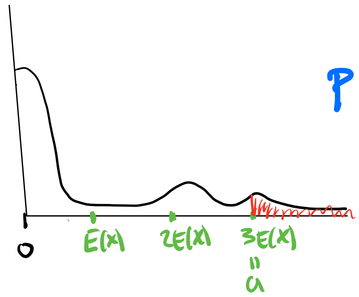
Picture



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

$P(X \geq 500K) \leq \frac{E(X)}{500K} = \boxed{1/5}$

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.



$P(X \geq 3E(X)) \leq \frac{E(X)}{3E(X)} = \boxed{1/3}$

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

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

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

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

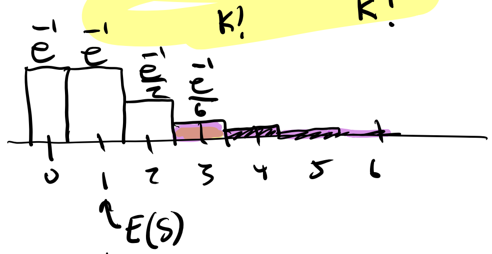
a) What distribution is S? $S \sim \text{Pois}\Big(100(.01)\Big) = \text{Pois}(1)$

b) Find an upperbound for $P(S \geq 3)$ using Markov's inequality. $\mu = 1$

$$P(S \geq 3) \leq \frac{E(S)}{3} = \boxed{\frac{1}{3}}$$

Note Exact: $P(S \geq 3) = 1 - \underset{e^{-1}}{\underset{\shortparallel}{P(0)}} - \underset{e^{-1}}{\underset{\shortparallel}{P(1)}} - \underset{\frac{e^{-1}}{2}}{\underset{\shortparallel}{P(2)}}$

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



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

$= .08$

② Sec 3.2 Expectation of a function of a RV.

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

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

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 1^{st} failure

$P(X=k) = q^{k-1} p$

Picture

$y = 3^x$

$1 \longrightarrow 3^1$

$2 \longrightarrow 3^2$

$3 \longrightarrow 3^3$

$\vdots$

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

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

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

$\frac{1}{1-3q}$ 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

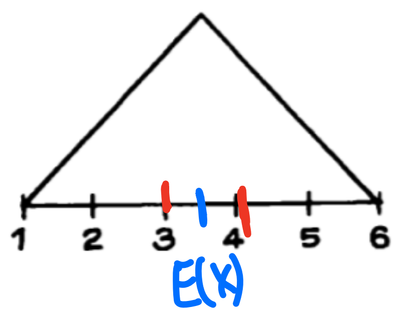
Thm if X and Y are independent

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

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

$$SD = E\left(|X - E(X)|\right)$$

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

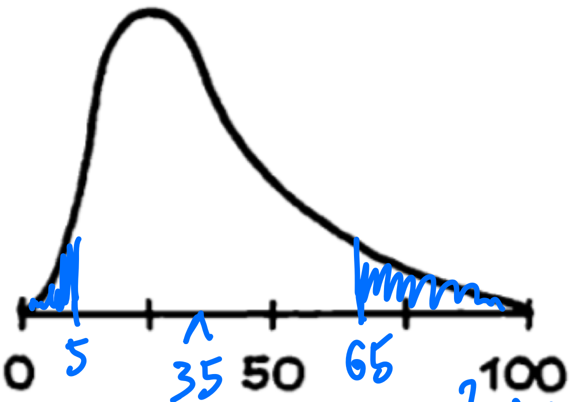
$$Var(X) = (SD(X))^2 = E\left((X - E(X))^2\right)$$

Chebyshev's Inequality

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

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

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



$$\underbrace{k}_{2}\cdot\underbrace{SD}_{15} = 30 \qquad \boxed{k=2}$$

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

$$\le \frac{1}{2^2} = \boxed{\frac{1}{4}}$$

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

$$P(X \ge 65) = P(X \ge \underbrace{\mu}_{35} + \underbrace{k}_{2}\underbrace{\sigma}_{15}) \le \left(\frac{1}{2}\right)^2$$

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 $\geq$ 5. To get an upper bound for p, you should:

   X = non neg number
   $E(X) = 1$

   ~~**a**~~ Assume a normal distribution

   b Use Markov's inequality $a=5$ M: $P(X \geq 5) \leq 1/5$

   c Use Chebyshev's inequality C: $P(X \geq 5) \leq \frac{1}{4}$

   $1+2\cdot 2 \Rightarrow \boxed{k=2}$

   d none of the above

since $1/5 < 1/4$

Proof of Chebyshev

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

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

By Markov

$$P(Y \geq a) \leq \frac{E(Y)}{a} \quad \text{for } \begin{array}{l} Y \geq 0 \\ a > 0 \end{array}$$

$Y = (X-E(X))^2$ ← nonnegative

$a = (K\,SD(X))^2$ ← pos

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

$= \cancel{Var(X)}$

$$\parallel$$

$$P\left(\sqrt{(X-E(X))^2} \geq \sqrt{(K\,SD(X))^2}\right.$$

$\sqrt{(X-E(X))^2} = |X-E(X)|$, $\sqrt{(K\,SD(X))^2} = K\,SD(X)$

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

$\square$

Appendix

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

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

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

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

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

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

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

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