Stat 134 lec 12

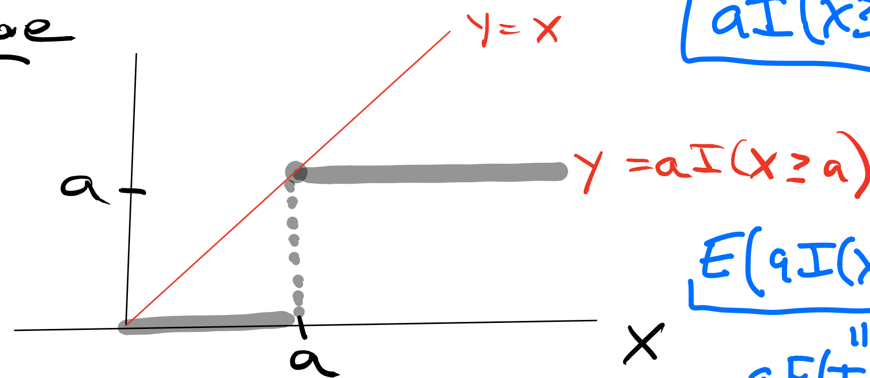
Warmup: 1:00–1:10

For a nonneg RV $X \geq 0$ and $a > 0$,

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

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

Picture



$$\boxed{aI(X \geq a) \leq X}$$

$$\underline{E(aI(X \geq a)) \leq E(X)}$$

$$\| $$

$$aE(I(X \geq a)) = aP(X \geq a)$$

Which of the following is true

a) $E(aI(X \geq a)) \leq E(X)$ ✓

b) $P(X \geq a) \leq \frac{E(X)}{a}$ ✓

c) more than one of the above

d) none of the above

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

Markov's inequality,

Quiz 3 Monday Sec 3.1–3.2 and Chebyshev's inequality

last time

Discrete Distributions

1. Ber(p)
2. Bin(n,p)
3. HG(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
$\vdots$		

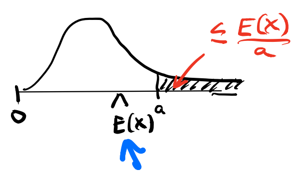
$P(X=k) = q^{k-1}p$ Geom(p) formula on $\{1,2,\dots\}$

Note trials are independent

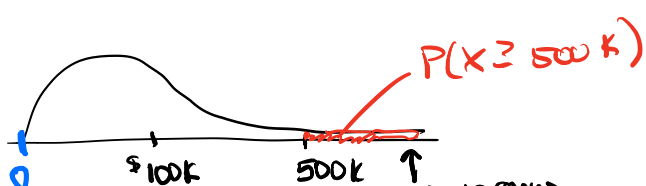
Sec 3.2 Markov's inequality

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

Picture



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

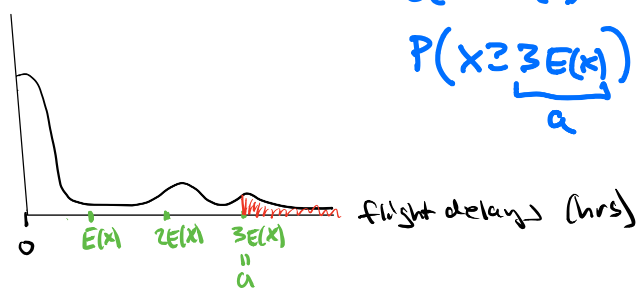


By Markov's inequality $P(X \geq \$500k) \leq \frac{100k}{500k} = 1/5$

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

$a = 3E(X)$

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



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

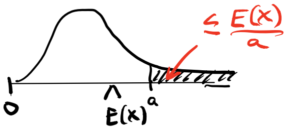
Today

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

1) Markov Inequality example

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

Picture ?



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

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

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

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

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

Soln $S = X_1 + X_2 + \cdots + X_{100} \sim \text{Pois}(100(.01)) = \text{Pois}(1)$.

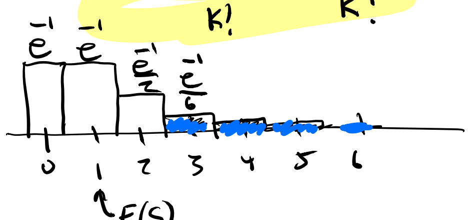
$S \sim \text{Pois}(1)$

$E(S) = 1$

By Markov's inequality, $P(S \geq 3) \leq \frac{1}{3}$

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

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



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

$= .08$

② Sec 3.2 Expectation of a function of a RV.

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

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

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

Picture

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

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

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

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

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 $E(\underbrace{X+Y}_{= g(x,y)}) = E(X) + E(Y)$

see appendix to notes

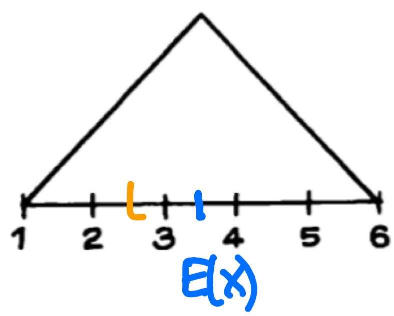
Thm if X and Y are independent

$$E(\underbrace{XY}_{= g(x,y)}) = E(X)E(Y)$$

(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

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

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

Chebyshev's Inequality

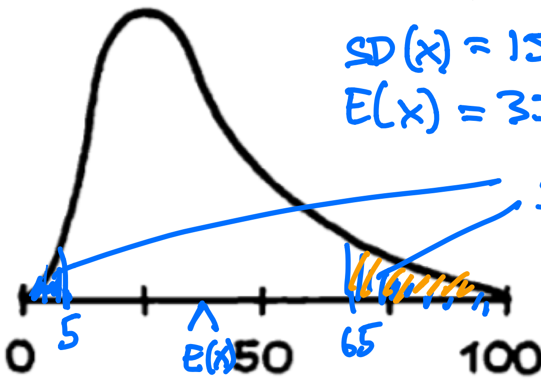
For any random variable X, and any $k>0$,

$$P\left(|X - \underset{\substack{\| \\ 35}}{E(X)}| \geq \underset{\substack{\| \\ 2}}{k}\cdot \underset{\substack{\| \\ 15}}{SD(X)}\right) \leq \frac{1}{k^2}$$

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

$\left.\begin{array}{l} SD(X) = 15 \\ E(X) = 35 \end{array}\right\}$ Given

$\leq 1/4$



$\underset{\substack{\| \\ 35}}{E(X)}$ $\qquad$ $\underset{\substack{\| \\ 30}}{k\cdot SD(X)}$

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

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

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

$$P(\underset{\substack{\| \\ \mu + k\sigma}}{X \geq 65}) \leq \frac{1}{2^2}$$

$\sigma = SD(X)$

$\mu = E(X)$

$$65 = 35 + \underset{\substack{\| \\ 2}}{k}\cdot 15$$

tinyurl.com/sept23-pt1

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

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

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

   c Use Chebyshev's inequality

   d none of the above

$P(X \geq 5) \leq 1/4$

$\mu + k\sigma$

$5 = 1 + 2 \cdot 2 \Rightarrow k = 2$

Appendix

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

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

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

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

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

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

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

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