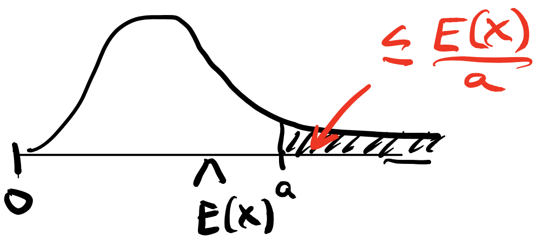
Stat 134, lec 12

Last time

Sec 3.2 Markov's inequality

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

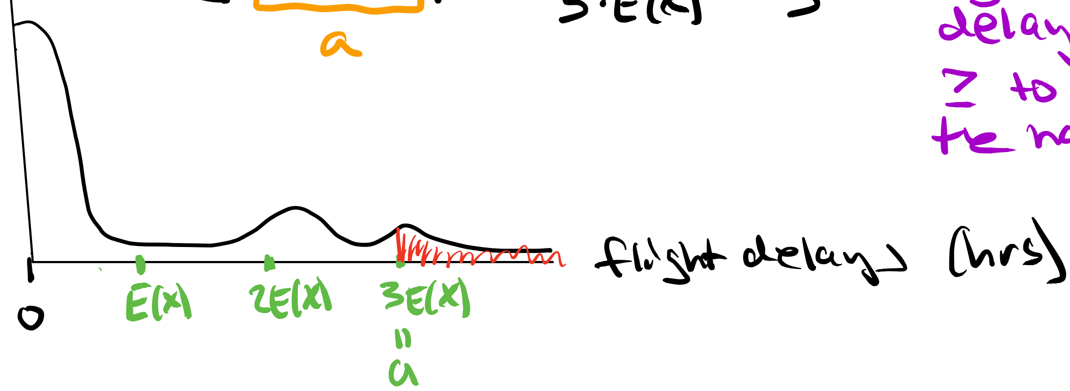
Picture



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.



Today

Sec 3.2 $E(g(X,Y))$

Sec 3.3 $SD(X)$, $Var(X)$

Chebyshev inequality.

Geometric RV — # trials until first success.

Let X = number of p coin tosses until your your first heads.

$X=1$ H p

$X=2$ TH qp

$X=3$ TTH q^2p

$\vdots$

$$\boxed{P(X=k) = q^{k-1}p \quad \text{Geometric formula}}$$

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

X $Y = g(X) = X^2$

$P(X=1) = \frac{1}{4}$ $1 \longrightarrow 1$ $P(Y=1) = 1$

$P(X=-1) = \frac{3}{4}$ $-1 \longrightarrow 1$

$$E(Y) = 1 \cdot \underset{\overset{\parallel}{P(X=1)+P(X=-1)}}{P(Y=1)} = 1 \cdot 1 = 1$$

$$= 1^2\left(\tfrac{1}{4}\right) + (-1)^2\left(\tfrac{3}{4}\right)$$

ex Suppose a RV X has the geometric (p) distribution on $\{1,2,3,\dots\}$ and let $p > 2/3$. Find $E(3^X)$.

If you use an infinite sum you must simplify it.

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

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

$$E(3^X) = \sum_{k=1}^{\infty} 3^k P(X=k) = 3p + 3^2qp + 3^3q^2p + \dots$$

$$= 3p\left(1 + 3q + (3q)^2 + \dots\right)$$

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

for $3q < 1$

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

have $p > 2/3$

$\Rightarrow q < 1/3$

$\Rightarrow 3q < 1$ ✓

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

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

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

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

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

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

Thm If X, Y are independent

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

Pf/ $$E(X,Y) = \sum_x \sum_y xy P(X=x, Y=y)$$

independence

$$= \sum_x \sum_y xy P(X=x) P(Y=y)$$

$\sum_y c P(Y=y) = c \sum_y P(Y=y)$

$$= \sum_x x P(X=x) \sum_y y P(Y=y)$$

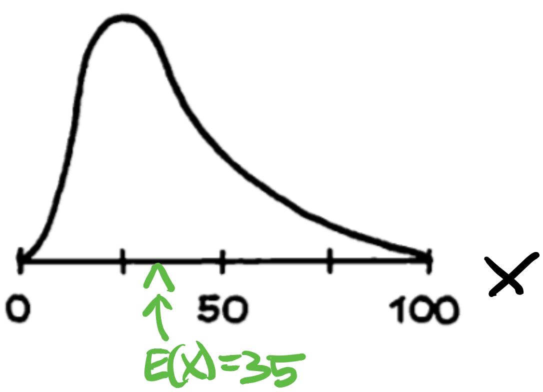
$$= E(X)E(Y)$$

Sec 3.3 Standard deviation (SD)

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

What is the average spread around the mean for the distribution below?

Choises: 5, 15 or 50?



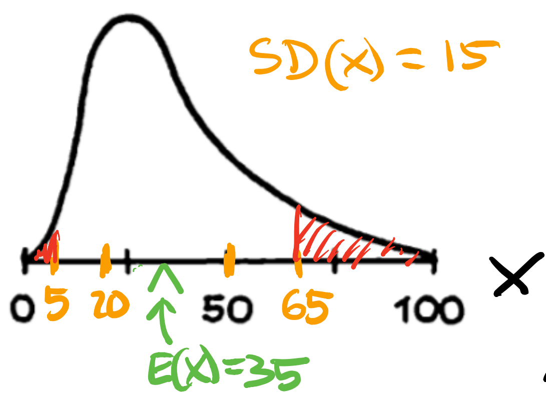
$$SD(X) = \sqrt{E\left(\left(X - E(X)\right)^2\right)}$$

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

Chebyshev's Inequality

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

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



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

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

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

Markov

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

need $X \ge 0$

need $E(X)$

right tail

Chebyshev

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

any X

need $E(X)$, $SD(X)$

one or both tails.

Stat 134

Wednesday September 19 2018

1. A list of incomes has an average of \$70,000 and an SD of \$30,000. Let p be the proportion of incomes over \$100,000. To get an upper bound for p, you should:

   a Assume a normal distribution — not normal

   b Use Markov's inequality — $P(X \geq 100k) \leq \frac{70}{100}$

   c Use Chebyshev's inequality — $P(X \geq 100k) \leq 1$

   d none of the above

since

$P(|x-70| \geq 1 \cdot 30) \leq \frac{1}{1}$

$k=1$

Proof of Chebyshev

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

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

Pf/ By Markov's inequality

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

let

$$Y = (X - E(X))^2 \qquad Y \geq 0.$$

$$a = k^2 \operatorname{var}(X).$$

$$\begin{matrix} P(Y \geq a) & = & \dfrac{\overset{\operatorname{var}(X)}{\overbrace{E\left((X-E(X))^2\right)}}}{k^2 \operatorname{var}(X)} = \dfrac{1}{k^2} \\ \| & & \end{matrix}$$

$$P\left((X-E(X))^2 \geq k^2 \operatorname{var}(X)\right)$$

$$\|$$

$$P\left(\sqrt{(X-E(X))^2} \geq \sqrt{k^2 \operatorname{var}(X)}\right)$$

$$\|$$

$$P(|X-E(X)| \geq k\,SD(X))$$

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

$\square$