Stat 134 lec 13

Quiz 3 Monday sec 3.1–3.2 and Chebyshev Inequality

last time

sec 3.3

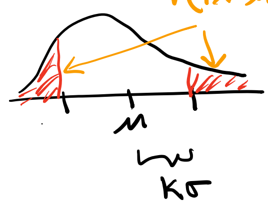
$SD = \sigma = \sqrt{E(X-\mu)^2}$ $\mu = E(X)$

$Var = \sigma^2 = E(X-\mu)^2$

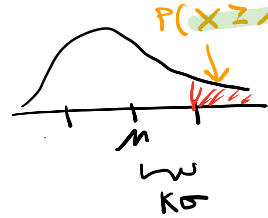
Tail bounds

Chebyshev's inequality

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



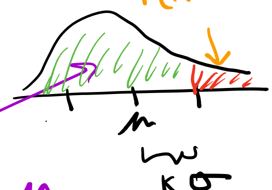
$P(X \geq \mu + k\sigma) \leq \frac{1}{k^2}$



Markov's inequality

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

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



$P(X < \mu + k\sigma) \geq 1 - \frac{\mu}{\mu + k\sigma}$

← lower bound.

Today finish sec 3.3
1. review student comments concept test
2. Proof of Chebyshev inequality
3. another formula for Variance
4. Properties of variance
5. Central limit theorem (CLT)

(1) Concept test

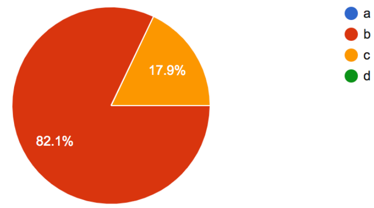
1. A list of non negative numbers has an average of 1 and an SD of 2. Let p be the proportion of numbers over 5. To get an upper bound for p, you should:

a Assume a normal distribution

b Use Markov's inequality

c Use Chebyshev's inequality

d none of the above



c	Can't use markov's because the data falls below 0

b	bc we are only interested in the right tail region?

b	The markov gives a smaller upper bound in comparison to Chebyshev (1/5 is a tighter upper bound than 1/4).

Markov: $E(X)/5 = 1/5$

Chebyshev: $1/2^2 = 1/4$

b	Chebyshev would have k=2 so 2 SD's above/below this would give an upper bound of 1/4. But you cannot be 2 SD's below so Cheby's doesn't work (I think) Markov's is meant purely for upper tail, so E(x)/a = 1/5 So you would use Markov's

②

Proof of Chebyshev

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

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

Pf/ By Markov

$$P(Y \ge a) \le \frac{E(Y)}{a} \qquad \text{for } Y \ge 0,\ a > 0$$

let $Y = (X-E(X))^2$ ← nonnegative

$a = (K\,SD(X))^2$ ← positive

so

$$P\left((X-E(X))^2 \ge (K\,SD(X))^2\right) \le \frac{\overbrace{E\left((X-E(X))^2\right)}^{\mathrm{var}(X)}}{K^2\underbrace{SD(X)^2}_{\mathrm{var}(X)}} = \frac{1}{K^2}$$

$$\|$$

$$P\Big(\underbrace{\sqrt{(X-E(X))^2}}_{|X-E(X)|} \ge \underbrace{\sqrt{(K\,SD(X))^2}}_{K\,SD(X)}\Big)$$

□

③

Sec 3.3 Another formula for Var(X).

Recall $E(cX) = cE(X)$

so $E(E(X)X) = E(X)E(X)$

$$\operatorname{Var}(X) = E\left((X - E(X))^2\right)$$

$$\overset{\text{FOIL}}{=} E\left(X^2 - 2E(X)X + (E(X))^2\right)$$

E(cX) = cE(X)

$$= E(X^2) \underbrace{- 2E(X)E(X) + E(X)^2}_{= -E(X)^2}$$

$$\Rightarrow \boxed{\operatorname{Var}(X) = E(X^2) - E(X)^2}$$

$$\text{(or)} \quad \boxed{E(X^2) = \operatorname{Var}(X) + E(X)^2}$$

ex: Let X be a nonnegative RV such that
$E(X) = 100 = \text{var}(X)$

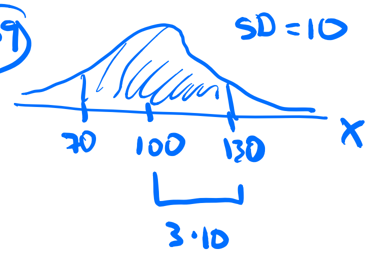
a) Can you find $E(X^2)$ exactly? If not what can you say.

$E(X^2) = \text{var}(X) + E(X)^2$
$= 100 + 10{,}000 = \boxed{10{,}100}$

b) Can you find $P(70^2 < X^2 < 130^2)$ exactly? If not what can we say?

Chebyshev

$P(70 < X < 130) \geq 1 - \frac{1}{9} = .89$



$\shortparallel$

$P(70^2 < X^2 < 130^2)$

Markov doesn't work here:

Markov:

$P(70^2 < X^2 < 130^2) \leq P(X^2 < 130^2) \geq 1 - \frac{E(X^2)}{130^2} = 1 - .6 = .4$

$E(X^2) = 10{,}100$

Markov

this inequality goes the wrong way!

tinyurl.com/sept27-pt1

Stat 134

Chapter 3 Friday February 22 2019

1. X is nonnegative random variable with $E(X) = 3$ and $SD(X) = 2$. True, False or Maybe:

$$P(X^2 \geq 40) \leq \frac{1}{3}$$

a True

b False

c Maybe

Markov:

$P(X \geq \sqrt{40}) \leq \frac{E(X)}{\sqrt{40}} = \frac{3}{\sqrt{40}} = .47$ $\boxed{\text{Maybe}}$

Markov:

$E(X^2) = Var(X) + E(X)^2 = 4 + 9 = 13$

$P(X^2 \geq 40) \leq \frac{13}{40} < \frac{13}{39} = 1/3$ $\boxed{\text{True}}$

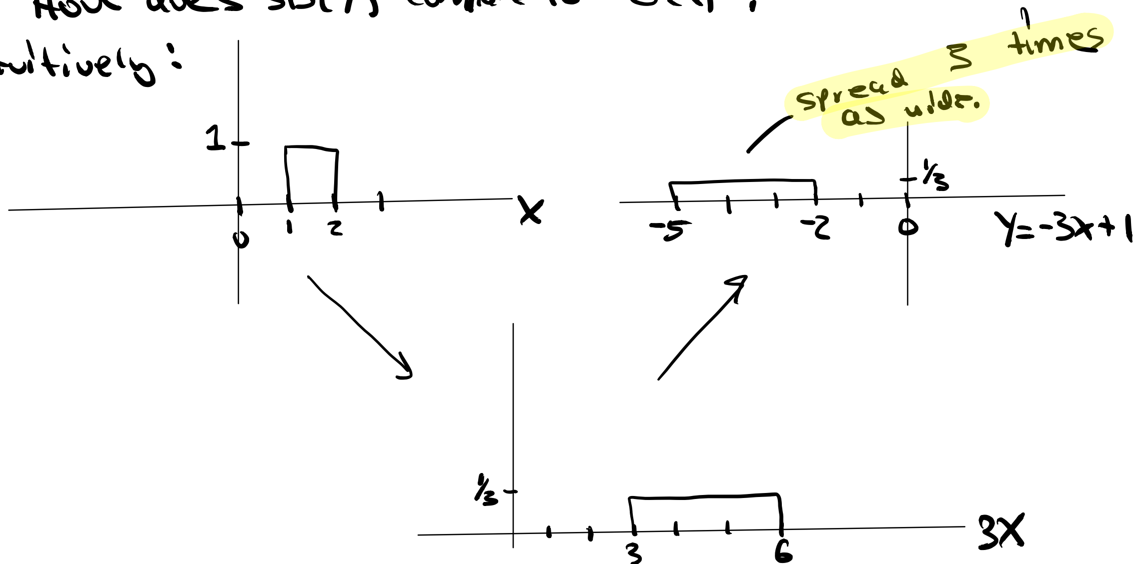
So answer is true.

④ Properties of Variance

Let $Y = -3X + 1$

How does $SD(Y)$ compare to $SD(X)$?

intuitively:



$$\boxed{\begin{aligned} SD(aX+b) &= |a|\,SD(X) \\ Var(aX+b) &= a^2\,Var(X) \end{aligned}}$$

see p 193 Pitman

Thm $Var(X+Y) = Var(X) + Var(Y)$ if X, Y are independent.

ex $\left.\begin{array}{l} X \sim Bin(n,p) \\ Y \sim Bin(m,p) \end{array}\right\}$ indep.

$\Rightarrow X+Y \sim Bin(n+m,p)$

$\Rightarrow Var(X+Y) = (n+m)pq = npq + mpq = Var(X) + Var(Y)$.

ex X = # hours a student is awake a day

Y = # hours a student is asleep a day.

$X+Y = 24 \Rightarrow Var(X+Y) = Var(24) = 0 \neq Var(X) + Var(Y)$

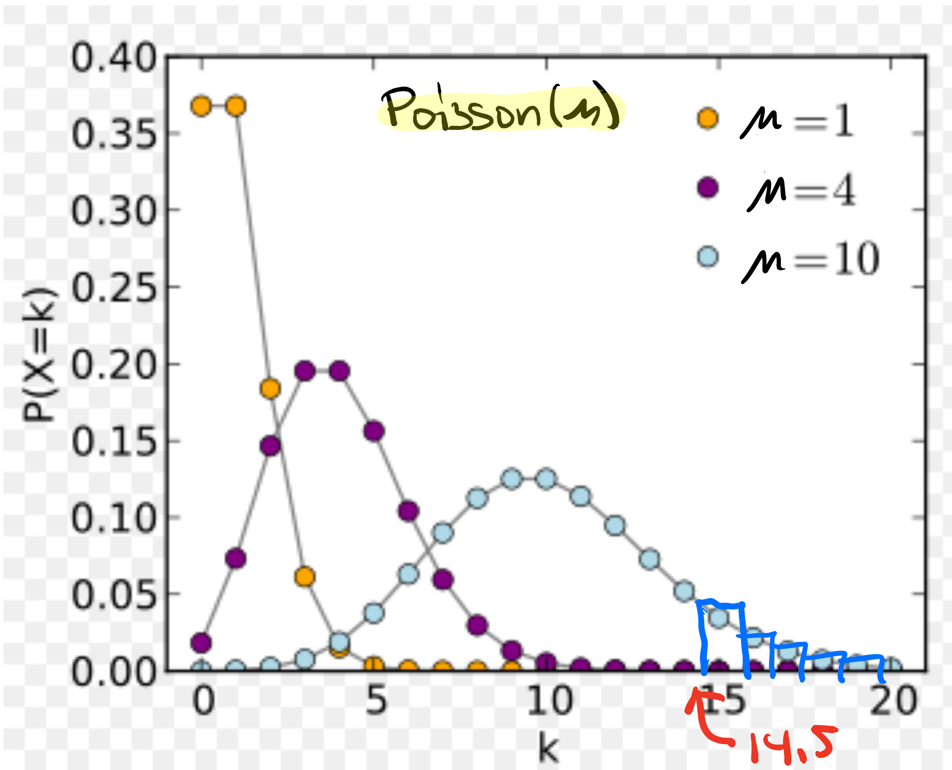
so variance formula needs X, Y to be independent.

(5) Central Limit theorem (CLT)

ex Let $X_1, X_2, \ldots X_{10}$ be i.i.d. Poisson(1).
Let $S_{10} = X_1 + \cdots + X_{10}$ Find $P(S_{10} \geq 15)$

Facts
if $X \sim \text{Pois}(\mu)$, $E(X) = \mu$
$\text{Var}(X) = \mu$

$$E(S_{10}) = E(X_1 + \cdots + X_{10}) = 10\,E(X_1) = 10$$
$$\text{Var}(S_{10}) = \text{Var}(X_1 + \cdots + X_{10}) = 10\,\text{Var}(X_1) = 10$$
$$SD(S_{10}) = \sqrt{\text{Var}(S_{10})} = \sqrt{10}$$



CLT says $S_{10} \sim N(\underset{\mu}{10}, \underset{\sigma^2}{10})$

$$P(S_{10} \geq 15) = 1 - \Phi\left(\frac{14.5 - 10}{\sqrt{10}}\right)$$
$$= 1 - \Phi(1.42) = .077$$