Stat 134 lec 13

Quiz 3 Monday sec 3.1–3.3

last time sec 3.3

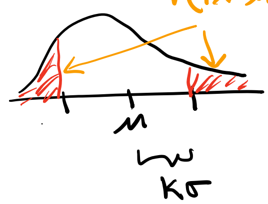
$$SD = \sigma = \sqrt{E(X-\mu)^2}$$

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

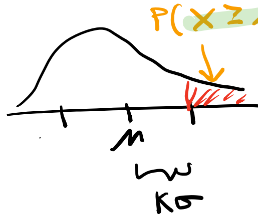
Tail bounds

Chebyshev's inequality

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



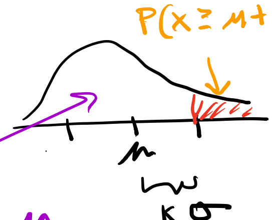
$$P(X \geq \mu + k\sigma) \leq \frac{1}{k^2}$$



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).
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)

<u>Proof of Chebyshev</u>

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

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

Pf/ By Markov's inequality

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

let $Y = (X - E(X))^2$ ← non negative

$a = (K\,SD(X))^2$ ← positive

So we have

$$P\left((X-E(X))^2 \geq (K\,SD(X))^2\right) \leq \frac{\overset{\mathrm{var}(X)}{\overset{\shortparallel}{E\left((X-E(X))^2\right)}}}{\underset{K^2\mathrm{var}(X)}{\underset{\shortparallel}{(K\,SD(X))^2}}}$$

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

$\square$

(3)

Sec 3.3 Another formula for $Var(X)$.

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

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

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

FOIL

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

$E(cX)=cE(X)$

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

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

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

e_x Let X be a non negative RV such that
$E(X) = 100 = Var(X)$

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

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

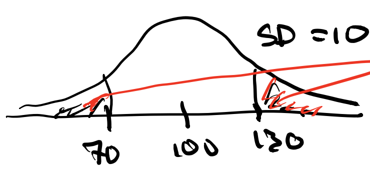
b) Can ~~you~~ find $P(70^2 < X^2 < 130^2)$ exactly? If not what can we say? NO

$P(70^2 < X^2 < 130^2) \geq .89$ by Chebyshev

You could use Chebyshev:
We don't know $SD(X^2)$ so we will need to apply Chebyshev to X

Note that:
$$70^2 < X^2 < 130^2 \iff 70 < |X| < 130$$
$|X| = X$ since $X \geq 0$

So,
$$P(70^2 < X^2 < 130^2) = P(70 < X < 130)$$



$K = 3$
$$P(|X - 100| \geq 30) \leq \frac{1}{9}$$

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

You could use Markov:
$$P(X^2 \geq 130^2) \leq \frac{E(X^2)}{130^2} = \frac{10100}{130^2} = .60$$
$$\Rightarrow P(70^2 < X^2 < 130^2) \leq P(X^2 < 130^2) \geq 1 - .60 = .40$$

Clearly Chebyshev is a better lower bound.

Stat 134

Chapter 3 Friday February 22 2019

1. X is 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 $E(X^2) = var(X) + (E(X))^2 = 4 + 9 = 13$

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

Note that we can't use Chebyshev since

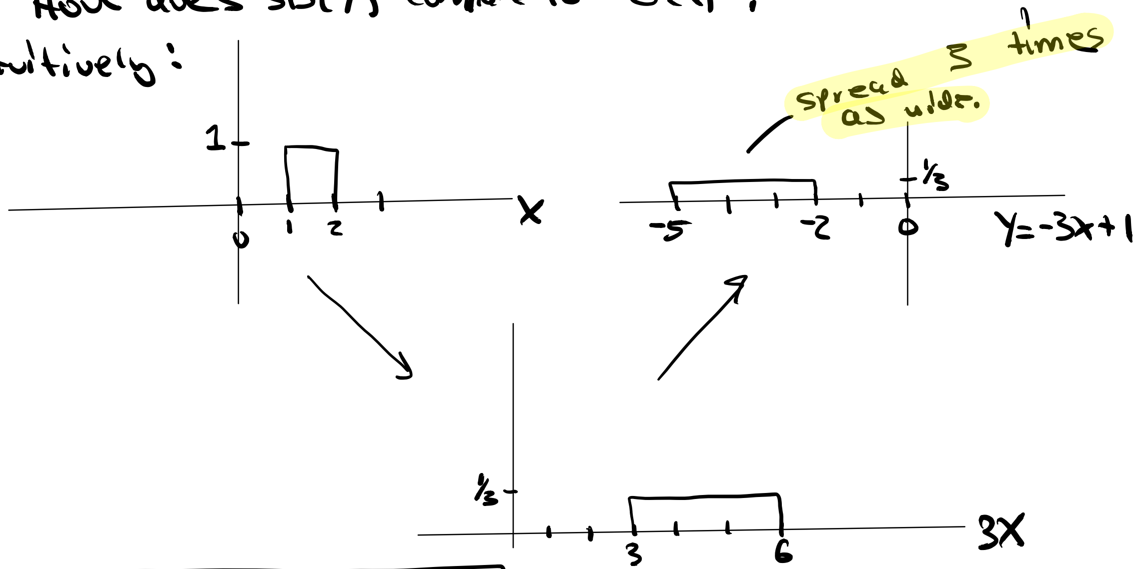
$P(X^2 \geq 40) \neq P(X \geq \sqrt{40})$ if X not nonnegative.

④ Properties of Variance

Let $Y = -3X + 1$

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

intuitively:



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

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.

ex Suppose X_1, X_2, X_3, X_4 are lengths of independent calls, with expectation 6 minutes and SD 7 min.
Let $T = X_1 + X_2 + X_3 + X_4$ be the total length of the calls.

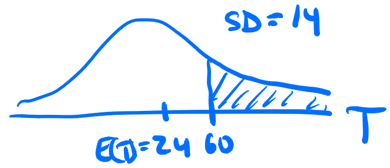
Give a lower bound for $P(T < 60)$

Soln We know how to find an upper bound for $P(T \geq 60)$ using both Markov and Chebyshev inequality.

$E(T) = 4 \cdot 6 = 24$

$\text{var}(T) = 4 \cdot 7^2$

$SD(T) = 14$



Markov $P(T \geq 60) \leq \frac{24}{60} = .40 \Rightarrow P(T < 60) \geq 1 - .4 = .6$

Chebyshev $P(T \geq 60) \leq \frac{1}{(2.57)^2} = .15 \Rightarrow P(T < 60) \geq 1 - .15 = .85$

better lower bound

$\mu + k\sigma$

$60 = 24 + k \cdot 14 \Rightarrow k = \frac{60 - 24}{14} = 2.57$

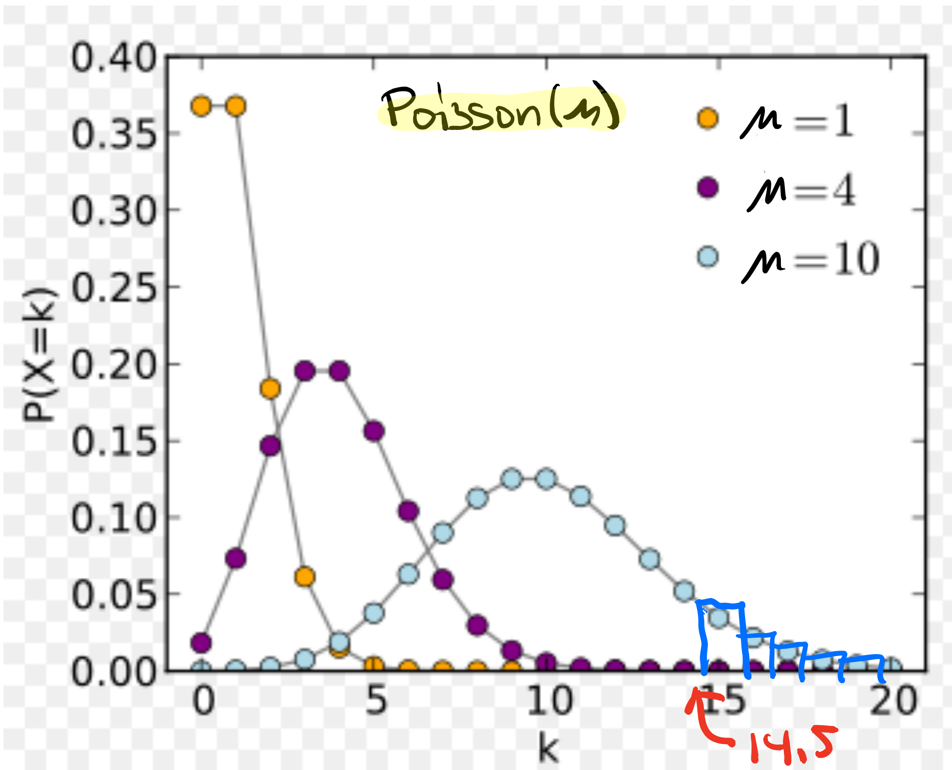
$\Rightarrow \boxed{P(T < 60) \geq .85}$

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

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



CLT says $S_{10} \sim N(10, 10)$
μ σ^2

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