Stat 134 lec 13

warmup 11:00–11:10

$X \sim Pois(\frac{1}{3})$

Find $E(X!)$

$X \sim Pois(\mu)$

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

$k=0,1,2,\ldots$

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

$E(X!) = \sum_{k=0}^{\infty} \cancel{k!}\, P(X=k) = \frac{e^{-1/3}(1/3)^k}{\cancel{k!}} = e^{-1/3} \underbrace{\sum_{k=0}^{\infty}(1/3)^k}$

$= 1 + 1/3 + (1/3)^2 + \cdots = \frac{1}{1-1/3} = \frac{3}{2}$

$= \boxed{\frac{3}{2} \cdot e^{-1/3}}$

Announcement: Remote Q2 Monday sec 2.1, 2.2, 2.4, 2.5, 3.1, 3.2
Download on gradescope. Logistics TBA soon.

last time

sec 3.3 SD(X) is the average deviation from the mean

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

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

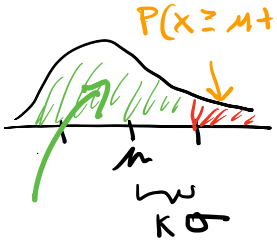
often write $E(X-\mu)^2$

Tail bounds

Markov's inequality

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

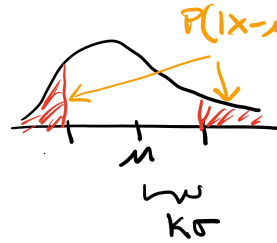
$P(X \geq \mu + k\sigma) \leq \frac{\mu}{\mu + k\sigma}$



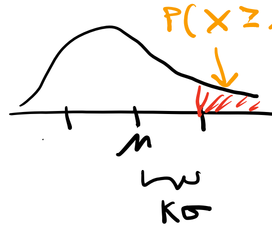
$P(X < \mu + k\sigma) \geq 1 - \frac{\mu}{\mu + k\sigma}$

Chebyshev's inequality

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



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



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

(3 pts) Suppose that each year, Berkeley admits 15,000 students on average, with an SD of 5,000 students. Assuming that the application pool is roughly the same across years, find the smallest upper bound you can on the probability that Berkeley will admit at least 22,500 students in 2019. (Hint: you should be comparing two possible bounds.)

Let X = # students Cal admits in 2019

M

$$P(X \geq 22{,}500) \leq \frac{15000}{22500} = 2/3$$

C

$$P(X \geq 22{,}500) \leq \frac{1}{(1.5)^2} = 4/9$$

$$" = 15{,}000 + 1.5\,(5000)$$

Today Sec 3.3

1. another formula for variance
2. Properties of variance

1

Sec 3.3 Another formula for Var(X).

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

so $E(\,E(X)\,X\,) = E(X)\,E(X)$

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

FOIL

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

$E(cX) = cE(X)$

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

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

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

$(E(X))^2$

ex let $X = \begin{cases} 1 & \text{with prob } p \\ 0 & \text{with prob } q \end{cases}$

$$E(X^2) = \sum_{\text{all } x} x^2 P(X = x) = 1^2 \cdot p + 0 \cdot q = \boxed{p}$$

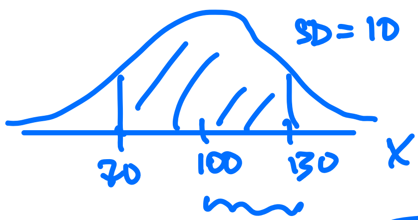
$$\mathrm{Var}(X)\,? = E(X^2) - (E(X))^2 = p - p^2 = p(1-p) = pq$$

ex: Let X be a nonnegative RV such that $E(X) = 100 = Var(X)$

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

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

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



$k\sigma = 30 \Rightarrow \boxed{k = 3}$

$$P(70^2 < X^2 < 130^2)$$
$$= P(70 < X < 130) \geq 1 - \frac{1}{k^2} = \frac{8}{9}$$

since X is nonneg.

Stat 134

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

C: $P(X^2 \ge 40) = P(X \ge \sqrt{40}) \le \dfrac{1}{\left(\frac{\sqrt{40}-3}{2}\right)^2} = 0.36$

$\sqrt{40} = \mu + k\sigma$, $\mu = 3$, $\sigma = 2$ $\Rightarrow k = \dfrac{\sqrt{40}-3}{2}$

M: $P(X \ge \sqrt{40}) \le \dfrac{3}{\sqrt{40}} = 0.47$

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

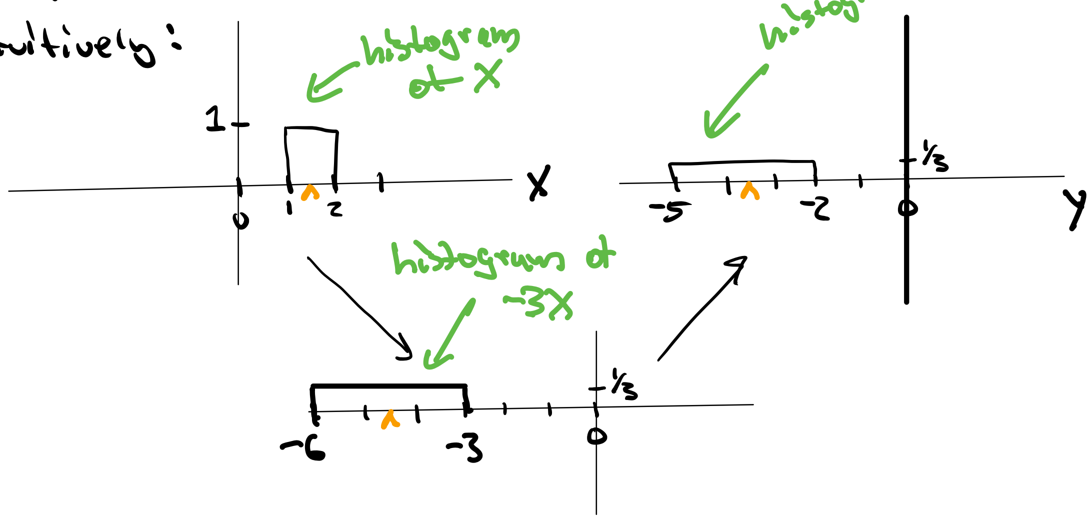
$P(X^2 \ge 40) \le \dfrac{E(X^2)}{40} = \dfrac{13}{40} < \dfrac{13}{39} = \boxed{1/3}$

② 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 for the proof.

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

ex X = # hours a student is awake a day

Y = # hours a student is asleep a day.

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

so Variance formula needs X, Y to be independent.

ex

Let $X_1, X_2, \ldots$ be independent and identically distributed, and for each $n \geq 1$ let $S_n = X_1 + X_2 + \cdots + X_n$. Suppose $E(S_{100}) = x$ and $SD(S_{100}) = y$. Let $W = S_{900} - 40$. Fill in the blanks with formulas in terms of x and y.

$E(W) =$ ________________ $SD(W) =$ ________________

2 is an index

$$S_{900} = \underbrace{X_1 + \cdots + X_{100}}_{S^1_{100}} + \underbrace{X_{101} + \cdots + X_{200}}_{S^2_{100}} + \cdots + \underbrace{X_{801} + \cdots + X_{900}}_{S^9_{100}}$$

Note $S^1_{100}, S^2_{100}, \ldots S^9_{100}$ are independent and identically distributed.

$$E(S_{900}) = \underbrace{E(S^1_{100})}_{=x} + \cdots + \underbrace{E(S^9_{100})}_{=x} = 9x$$

$$E(W) = E(S_{900} - 40) = \boxed{9x - 40}$$

$$\operatorname{Var}(W) = \operatorname{Var}(S_{900} - 40) = \operatorname{Var}(S_{900})$$

$$= 9\underbrace{\operatorname{Var}(S^1_{100})}_{=y^2}$$

$$\Rightarrow SD(W) = \boxed{3y}$$