Stat 88 lec 22

Warmup 2:00–2:10

$X \geq 0$

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

a) State Markov's inequality.

b) Is it possible that half of US flights have delay times at least 3 times the national average?

$P(X \geq 3E(X)) \leq \frac{E(X)}{3E(X)} = 1/3$

so NO not possible since $1/3$ is less than $1/2$.

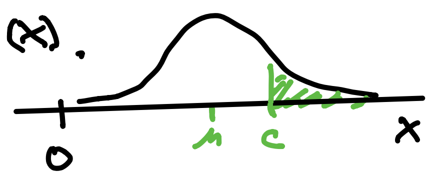
Last time

Upper bounds for tail probability

Sec 6.3 Markov's inequality

$X \geq 0$, $c > 0$, know $\mu = E(X)$.

$P(X \geq c) \leq \dfrac{\mu}{c}$ $\quad c = \mu + k\sigma$



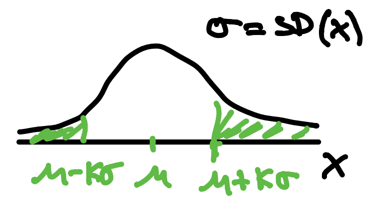
Sec 6.4 Chebyshev's Inequality

any X, $k > 0$, know $\mu = E(X)$, $\sigma = SD(X)$.

→ two tails:

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

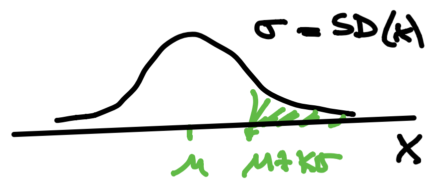
$X \notin (\mu - k\sigma, \mu + k\sigma)$



→ one tail

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

$X \geq \mu + k\sigma$



Alternative formula for $Var(X)$

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

$$Var(X) = E(X^2) - (E(X))^2 \Rightarrow E(X^2) = Var(X) + (E(X))^2$$

Today

1. Sec 6.3, 6.4 Markov, Chebyshev inequalities
2. Sec 7.1 Sums of independent random variables

ex

(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.

M: $P(X \geq c) \leq \frac{E(X)}{c}$

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

$X = \#$ admits (lucky kids!)

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

C: $P(X \geq 22{,}500) \leq \frac{1}{k^2} = \frac{1}{(\frac{3}{2})^2} = \boxed{\frac{4}{9}}$ ← better lower bound.

$22{,}500 = \mu + k\sigma = 15{,}000 + k\sigma$

so $7{,}500 = k\sigma$

$k = \frac{7500}{5000} = \boxed{\frac{3}{2}}$

ex

Suppose a list of numbers $x = \{x_1, \ldots, x_n\}$ has mean μ and standard deviation σ. Let k be the smallest number of standard deviations away from μ we must go to ensure the range $(\mu - k\sigma, \mu + k\sigma)$ contains at least 50% of the data in x. What is k?

$P(X \in (\mu-k\sigma, \mu+k\sigma)) = 1 - P(X \notin (\mu-k\sigma, \mu+k\sigma))$

$\geq 1 - 1/k^2 = 1/2$ (since $P(X \notin (\mu-k\sigma, \mu+k\sigma)) \leq 1/k^2$)

$\Rightarrow k^2 = 1/2 \Rightarrow \boxed{k = \sqrt{2}}$

Tinyurl.com/march13-pt1

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~~ 4. To get an upper bound for p, you should:

a Assume a binomial distribution

b Use Markov's inequality $P(x \geq 5) \leq \frac{1}{5}$ ← better bound

c Use Chebyshev's inequality $P(x \geq 5) \leq \frac{1}{2^2} = \frac{1}{4}$

d none of the above

$\mu + k\sigma$

$5 = 1 + k \cdot 2$

$k = 2$

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?

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

a) $E(X^2) = Var(X) + (E(X))^2$

$= 100 + 10{,}000 = \boxed{10{,}100}$

b) by Chebyshev

note 100 in middle

$$P(70^2 < X^2 < 130^2) = P(70 < X < 130) \geq 1 - \frac{1}{3^2} = \boxed{\frac{8}{9}}$$

$\mu + k\sigma$

$100 + k \cdot 10$

$k = 3$

② Sec 7.1 Sums of independent random variables

We know that

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

but does

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \ ?$$

Let X = # hours a student is awake a day
Y = # hours a student is asleep a day

$X+Y = 24 \Rightarrow \mathrm{Var}(X+Y) = \mathrm{Var}(24) = 0$
$\neq \mathrm{Var}(X) + \mathrm{Var}(Y)$.

So when X and Y are dependent
it is possible that $\mathrm{Var}(X+Y) \neq \mathrm{Var}(X) + \mathrm{Var}(Y)$.

Thm $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ if
X and Y are independent

ex let $X_1, \dots, X_n$ be a i.i.d. random sample with mean μ, SD σ

Let $S_n = \sum_{i=1}^{n} X_i$.

$$Var(S_n) = Var(X_1) + \dots + Var(X_n) = n\sigma^2$$

$\Rightarrow$ $SD(S_n) = \sqrt{n}\,\sigma$

SD of Binomial

ex Let $X \sim$ Bernoulli(p)

X	1	0
$P(X)$	p	$1-p$

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

$E(X^2) = 1^2 \cdot p + 0^2(1-p)$
$= p$
$E(X) = p$
$(E(X))^2 = p^2$

let $X \sim$ Binomial (n, p)

write X as a sum of n indicators and find $SD(X)$.

$X = I_1 + I_2 + \dots + I_n$ where

$I_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ trial success} \\ 0 & \text{else} \end{cases}$ p

$Var(X) = n\,Var(I_1)$
$= npq$

$\Rightarrow$ $\boxed{SD(X) = \sqrt{npq}}$

SD of Poisson

Recall that Binomial (n, p) can be approximated by Poisson(np) for large n and small p.

$$\text{Binomial}(n,p) \xrightarrow[]{\substack{n\to\infty \\ p\to 0 \\ np\to\mu}} \text{Poisson}(\mu = np)$$

so we can find SD of Poisson(μ) from limit of SD of Binomial(n,p) as $n \to \infty$, $p \to 0$, $np \to \mu$

$$\sqrt{npq} \longrightarrow \sqrt{\mu \cdot 1} = \sqrt{\mu}$$

$$\Rightarrow \boxed{SD(\text{Poisson}(\mu)) = \sqrt{\mu}}$$