STAT 88: Lecture 22

Contents

Warm up:

(a) State Markov's inequality.

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

(a) X nonnegative RV,

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

(b) $X =$ delay time of a flight ← Random number.

$E(X) =$ national average.

$P(X \geq \underbrace{3 \cdot E(X)}_{= c}) \leq \frac{E(X)}{3 \cdot E(X)} = \frac{1}{3}$

So it's not possible since $\frac{1}{3}$ less than $\frac{1}{2}$.

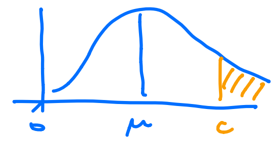
Last time

"Expectation & Var put constraint on tail probability"

Upper bounds for tail probability:

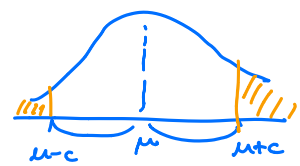
Markov's inequality For a non-negative random variable X and a positive constant $c > 0$,

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



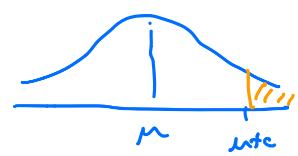
Chebyshev's inequality For a random variable X with mean μ and SD σ and a positive constant $c > 0$,

$$P(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2} = \frac{\mathrm{Var}(X)}{c^2}.$$



One tail bound:

$$P(X - \mu \geq c) \leq \frac{\sigma^2}{c^2} = \frac{\mathrm{Var}(X)}{c^2}.$$



<u>Alternative formula for $\mathrm{Var}(X)$:</u>

- $\mathrm{Var}(X) = E((X - \mu)^2)$.
- $\mathrm{Var}(X) = E(X^2) - E(X)^2$. This implies $E(X^2) = E(X)^2 + \underbrace{\mathrm{Var}(X)}_{\geq 0}$.

$\geq (E(X))^2$

(Jensen's Inequality)

$Var(X) = 0$

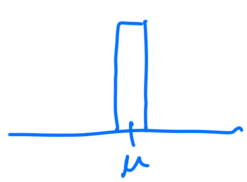
$\Rightarrow E((X-\mu)^2) = 0$

$$\sum_{\text{all } x} \underbrace{(x-\mu)^2 P(X=x)}_{=0}$$

$\Rightarrow (x-\mu)^2 = 0$ for all possible x

$\Rightarrow x = \mu$ for all possible x.

X takes value on μ w./ probability 1.



Example: 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.

X = # admits in 2019, $E(X) = 15000$, $SD(X) = 5000$

$P(X \geq 22500)$?

① $P(X \geq \underbrace{22500}_{=c}) \leq \frac{E(X)}{c} = \frac{15000}{22500} = \boxed{\frac{2}{5}}$ (Markov's)

② $P(X \geq 22500) = P(X - \underbrace{15000}_{=\mu} \geq \underbrace{7500}_{=c})$

$\leq \frac{Var(X)}{c^2}$

$= \frac{5000^2}{7500^2}$

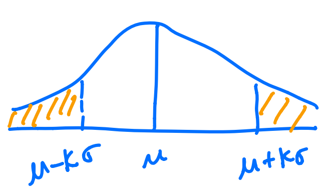
$= \frac{4}{9}$ (Chebyshev's)

Example: 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?

X = number from x

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

$= 1 - \underbrace{P(X \notin (\mu - k\sigma, \mu + k\sigma))}_{= P(|X-\mu| \geq k \cdot \sigma) \leq \frac{1}{k^2} \text{ (Chebyshev's)}} \geq \boxed{1 - \frac{1}{k^2}} = \frac{1}{2} \rightsquigarrow k = \sqrt{2}$



Example: A list of non negative numbers has an average of 1 and an SD of 2. Let p be the proportion of numbers over 4. To get an upper bound for p, you should:

(a) Assume a binomial distribution.

(b) Use Markov's inequality.

(c) Use Chebyshev's inequality.

(d) None of the above.

X = number from the list.

$E(X)=1$, $SD(X)=2$.

$p = P(X \geq 5)$ Need upper bound.

① Markov's. $p = P(X \geq 5) \leq \frac{E(X)}{5} = \boxed{\frac{1}{5}}$

② Chebyshev's. $p = P(X \geq 5)$

$= P(X - \underset{\mu}{1} \geq \underset{c}{4})$

$\leq \frac{Var(X)}{c^2} = \frac{4}{16} = \frac{1}{4}$

Example: Let X be a non negative random variable such that $E(X) = 100 = \text{Var}(X)$.

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

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

(a) $\text{Var}(X) = E(X^2) - (EX)^2$

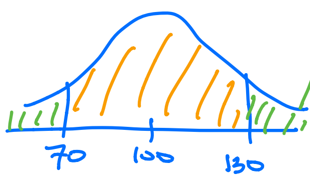
$$\rightarrow E(X^2) = \text{Var}(X) + (EX)^2$$
$$= 100 + (100)^2$$
$$= 10100$$

(b) $P(70 < X < 130) = 1 - P(X \notin (70, 130))$

$$= 1 - \underbrace{P(|X - \overset{\mu}{\overset{\shortparallel}{100}}| \geq \overset{c}{\overset{\shortparallel}{30}})}_{\leq \frac{\text{Var}(X)}{c^2} = \frac{100}{900} = \frac{1}{9}} \geq 1 - \frac{1}{9} = \frac{8}{9}$$

① $P(|X - \mu| \geq c) \leq \frac{\text{Var}(X)}{c^2}$

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



7.1. Sums of Independent Random Variables

We know that $E(X+Y) = E(X)+E(Y)$ but does $\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y)$?

Not always true.

Let X be the number of hours a student is awake a day and let Y be the number of hours a student is asleep a day. Then $X+Y = 24$, so trivially

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

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

Theorem If X and Y are independent,

$$\text{Var}(X+Y) = \text{Var}(X)+\text{Var}(Y).$$

Ex: Let $X_1, X_2, \ldots, X_n$ be a i.i.d. random sample with mean μ and SD σ. Let $S_n = \sum_{i=1}^{n} X_i$. Then

$$\text{Var}(S_n) = \underbrace{\text{Var}(X_1)}_{=\sigma^2} + \underbrace{\text{Var}(X_2)}_{=\sigma^2} + \cdots + \underbrace{\text{Var}(X_n)}_{=\sigma^2} = n\sigma^2.$$

So, $\text{SD}(S_n) = \sigma\sqrt{n}$. : grows $\sim \sqrt{n}$

$E(S_n) = \mu \cdot n$: grows $\sim n$

$n = 10{,}000$ $\quad SD(S_n) \sim 100$

$E(S_n) \sim 10{,}000$

SD of Binomial Let $X \sim \text{Bernoulli}(p)$. Then

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

$$\text{Var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p).$$

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

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

Now let $X \sim \text{Binomial}(n,p)$. Write X as a sum of n Bernoulli random variables and find $\text{SD}(X)$.

X = # successes out of n trials, $\quad I_j = \begin{cases} 1 & \text{if } j\text{th trial is success} \\ 0 & \text{o.w.} \end{cases}$

$= I_1 + \cdots + I_n$ $\quad I_j \sim \text{Bernoulli}(p)$

$\text{Var}(X) = \text{Var}(I_1) + \cdots + \text{Var}(I_n)$

$= n \cdot p \cdot (1-p)$

$SD(X) = \sqrt{np(1-p)}$

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

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

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

$$\sqrt{\underbrace{np}_{\mu}\underbrace{(1-p)}_{1}} \xrightarrow[np\to\mu]{n\to\infty,\ p\to 0} \sqrt{\mu}$$

$\Rightarrow$ If $X \sim \text{pois}(\mu)$,

$$SD(X) = \sqrt{\mu}$$

SD of Geometric Fact: If $X \sim \text{Geom}(p)$,

$$\text{SD}(X) = \frac{\sqrt{1-p}}{p}.$$

<u>Ex:</u> (Waiting till the 10th success) Suppose you roll a die until the 10th success. Let R be the number of rolls required. Find SD(R).

$$\underbrace{\_\ \_\ \_\ \overset{S}{\_}}_{X_1}\quad \underbrace{\_\ \_\ \overset{S}{\_}}_{X_2}\quad \underbrace{\_\ \_\ \_\ \overset{S}{\_}}_{X_3}\quad \cdots \quad \underbrace{\_\ \overset{S}{\_}}_{X_{10}}$$

$\leftarrow$ 10th success

$\rightarrow$ Independent

X_2 = # trials after first success until second success.

$X_i \sim \text{Geom}(p)$

$R = X_1 + X_2 + \cdots + X_{10}$.

$\text{Var}(R) = \text{Var}(X_1) + \cdots + \text{Var}(X_{10})$

$\qquad = 10 \cdot \frac{1-p}{p^2}$

$SD(R) = \sqrt{\quad ''\quad}$