Last time

Stat 134 lec 14

tail bounds Markov $X \ge 0,\ a > 0$ $P(X \ge a) \le \frac{E(X)}{a}$

Chebyshev $P(|X-\mu| \ge k\sigma) \le \frac{1}{k^2}$

sec 3.3 $Var(X+Y) = Var(X) + Var(Y)$ if X, Y are independent

Central Limit Thm (CLT)

Let $S_n = X_1 + \cdots + X_n$ where $X_1, \ldots, X_n$ are iid RVs, $E(X) = \mu$, $Var(X) = \sigma^2$.

Then,

$S_n \approx N(n\mu, n\sigma^2)$ for "large" n.

approximately

often ≥ 10

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 $P(T < 60)$:

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

$Var(T) = 4 \cdot 7^2 \Rightarrow SD = 14$

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

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

$\underset{24}{\mu} + k\underset{14}{\sigma} \Rightarrow k = 2.57$

better lower bound.

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

Today

1. Go over concept test responses from last time
2. Sec 3.6 (next time sec 3.4) Calculating variance of a sum of dependent indicators
3. Sec 3.6 Hypergeometric distribution

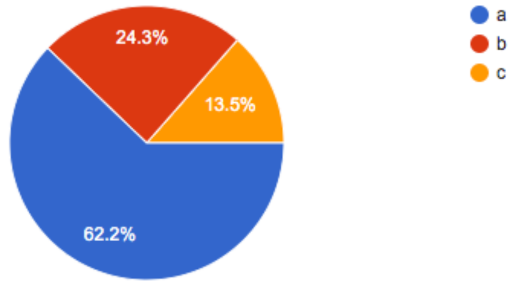
. X is nonnegative random variable with $E(X) = 3$ and $SD(X) = 2$. True, False or Maybe:

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

a True

b False

c Maybe



a	False for Markova, true for Chebysheva

b	using chebyshev on P(X >= 40) we get k = 1.66. thus, it is less than 1/1.66^2, which is greater than 1/3

$$P(X^2 \geq 40) = P(|X| \geq \sqrt{40})$$

$$\sqrt{40} = 3 + k \cdot 2$$

$$k = \frac{\sqrt{40} - 3}{2} = 1.66$$

$$\leq \frac{1}{(1.66)^2} = .36 \Rightarrow \text{maybe}$$

a	If we calculate with Markov's inequality using E(X^2), we find that the above probability is less than or equal to 13/40, and since that is less than 1/3, the above statement is true.

a	Var(X) = (SD(X))^2 = 2^2 = 4 E(X^2) = Var(X) + E(X)^2 = 4 + 3^2 = 13 P(X^2 >= 40) <= 13/40 < 1/3 Thus, the statement is true.

Sec 3.6 Var of sum of dependent indicators.

But first:
The variance of a sum of independent indicators

ex $I = \begin{cases} 1 & \text{Prob } p \\ 0 & \text{prob } 1-p \end{cases}$

$Var(I) = E(I^2) - E(I)^2$

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

$E(I) = p$

$Var(I) = p - p^2 = p(1-p)$

ex $X \sim Bin(n,p)$

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

$E(X) = nE(I_1) = \boxed{np}$

$Var(X) = n\,var(I_1) = \boxed{np(1-p)}$

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

Next we will look at an example where X is a sum of dependent indicators ex

14. A building has 10 floors above the basement. If 12 people get into an elevator at the basement, and each chooses a floor at random to get out, independently of the others, at how many floors do you expect the elevator to make a stop to let out one or more of these 12 people?

X = number of elevator stops. $p_1 = 1 - \left(\frac{9}{10}\right)^{12}$

a) Find $E(X)$

$X = I_1 + \dots + I_{10}$ $\qquad I_2 = \begin{cases} 1 & \text{if stop at } 2^{nd} \text{ floor} \\ 0 & \text{else} \end{cases}$

$E(X) = 10p_1$

b) Find $Var(X)$.

$Var(X) = E(X^2) - (E(X))^2$

$E(X^2) = E((I_1 + \dots + I_{10})^2) = \sum_{1 \le i,j \le 10} E(I_i I_j)$

$$\begin{array}{c|cccc} & I_1 & \cdots & \cdots & I_{10} \\ \hline I_1 & I_1 & I_{12} & & \\ I_2 & I_{21} & I_2 & & \\ \vdots & & & \ddots & \\ I_{10} & & & & I_{10} \end{array}$$

9·10 nondiagonals

$I_1 = \begin{cases} 1 & \text{if stop } 1^{st} \text{ floor} \\ 0 & \text{else} \end{cases}$

$I_2 = \begin{cases} 1 & \text{if stop } 2^{nd} \text{ floor} \\ 0 & \text{else} \end{cases}$

p_{12}

$I_{12} = I_1 I_2 = \begin{cases} 1 & \text{if stop } 1^{st} \text{ and } 2^{nd} \text{ floor} \\ 0 & \text{else} \end{cases}$

Note $I_{11} = I_1$

$P_{12} = 1 -$ Prob(dont stop at 1^{st} floor or dont stop 2nd floor)

$$= 1 - \left[\left(\frac{9}{10}\right)^{12} + \left(\frac{9}{10}\right)^{12} - \left(\frac{8}{10}\right)^{12} \right]$$

Prob dont stop at 1st and 2nd floor.

$$= \boxed{1 - 2\left(\frac{9}{10}\right)^{12} + \left(\frac{8}{10}\right)^{12}}$$

$$E(X^2) = \underbrace{10E(\overset{"P_1}{I_1})}_{\text{diagonals}} + \underbrace{9 \cdot 10E(\overset{"P_{12}}{I_{12}})}_{\text{non diagonals}}$$

$$(E(X))^2 = (10P_1)^2$$

$$\text{var}(X) = \underbrace{10P_1 + 9 \cdot 10P_{12}}_{E(X^2)} - \underbrace{(10P_1)^2}_{(E(X))^2}$$

Summary

Variance of sum of dependent indicators

$X = I_1 + \dots + I_n$

$P_1 = E(I_1)$

$P_{12} = E(I_{12}) = E(I_1 I_2)$

$E(X) = nP_1$

$$\mathrm{var}(X) = \underbrace{nP_1 + n(n-1)P_{12}}_{E(X^2)} - \underbrace{(nP_1)^2}_{E(X)^2}$$

Variance of sum of independent indicators

$X = I_1 + \dots + I_n$

$P_1 = E(I_1)$

$P_{12} = P_1 \cdot P_2 = P_1^2$

$$\mathrm{var}(X) = \underbrace{nP_1 + n(n-1)P_1^2}_{E(X^2)} - \underbrace{(nP_1)^2}_{E(X)^2} = nP_1 - nP_1^2 = nP_1(1-P_1)$$

Recall multinomial distribution

roll a die 10 times

$$P(2 \text{ ones}, 3 \text{ twos}) = \binom{10}{2,3,5}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^3\left(\frac{4}{6}\right)^5$$

$\frac{10!}{2!\,3!\,5!}$

tinyurl.com/sept30-pt1

Stat 134

Monday February 24 2019

1. A fair die is rolled 14 times. Let X be the number of faces that appear exactly twice. Which of the following expressions appear in the calculation of $Var(X)$

   a $14 * 13 * \underbrace{\binom{14}{2,2,10}(1/6)^2(1/6)^2(4/6)^{10}}_{= P_{12}}$

   b $\underbrace{\binom{14}{2}(1/6)^2(5/6)^{12}}_{= P_1}$

   c more than one of the above

   d none of the above

X = # face that appear twice

$X = I_1 + I_2 + \dots + I_6$

$I_2 = \begin{cases} 1 & 2^{nd} \text{ face twice} \\ 0 & \text{else} \end{cases}$

$P_1 = \binom{14}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{12}$

$I_{12} = \begin{cases} 1 & 1^{st} \text{ and } 2^{nd} \text{ face twice} \\ 0 & \text{else} \end{cases}$

$P_{12} = \binom{14}{2,2,10}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{4}{6}\right)^{10}$

$E(X^2) = 6P_1 + 6 \cdot 5 P_{12}$

$Var(X) = E(X^2) - (E(X))^2 = 6P_1 + 30P_{12} - (6P_1)^2$

Hypergeometric (n, N, G)

ex

A deck of cards has G aces.

$X = \#$ aces in n cards drawn without replacement from a deck of N cards

$N = 52$

$G = 4$

$n = 5$

$p_1 = \frac{G}{N}$

a) Find $E(X)$

$X = I_1 + I_2 + \dots + I_5$

$I_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ card ace} \\ 0 & \text{else} \end{cases}$

$E(X) = 5\left(\frac{4}{52}\right)$

$= \boxed{n\left(\frac{G}{N}\right)}$

$p_{12} = \frac{G}{N} \cdot \frac{G-1}{N-1}$

b) Find $Var(X)$

$I_{12} = \begin{cases} 1 & \text{if } 1^{st} \text{ and } 2^{nd} \text{ ace} \\ 0 & \text{else} \end{cases}$

$Var(X) = np_1 + n(n-1)p_{12} - (np_1)^2$