Stat 134 lec 14

warmup 1:00-1:10

Let X = number of sixes in 7 tosses of a fair die.

a) write X as a sum of indicators

$X = I_1 + \cdots + I_7$

b) Find $Var(X)$

$p = 1/6$

$I_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ toss is a } 6 \\ 0 & \text{else} \end{cases}$

$$Var(X) = Var(I_1 + \cdots + I_7)$$
$$= \underset{pq}{\underbrace{Var(I_1)}} + \cdots + \underset{pq}{\underbrace{Var(I_7)}} = \boxed{7pq}$$

$X \sim Bin(n,p)$

$Var(X) = npq$

$X = I_1 + \cdots + I_n$

$Var(X) = n\,\underset{pq}{\underbrace{Var(I_1)}}$

If n is large p small and $np \to \mu$

then $X \sim Pois(\mu)$

$Var(X) = npq = \mu$

$np \to \mu$, $q \to 1$

Last time

Sec 3.3 $\quad Var(X) = E\left((X - E(X))^2\right)$

or $\quad Var(X) = E(X^2) - (E(X))^2$

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

$Var(I) = pq$

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

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

$Var(X) = npq$

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

Today

1. Go over concept test responses from last time
2. Sec 3.3 CLT
3. Sec 3.6 (next time sec 3.4) Calculating variance of a sum of dependent indicators
4. Sec 3.6 Hypergeometric distribution

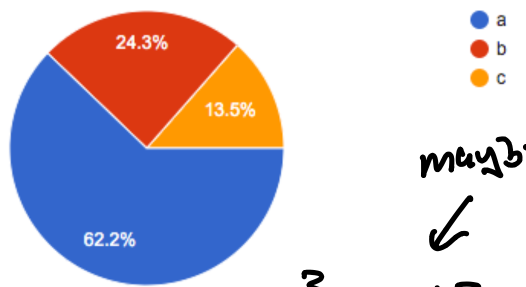
. X is nonnegative random variable with $E(X) = 3$ and $SD(X) = 2$. True, False or Maybe:

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

a True

b False

c Maybe



maybe

$M: P(X \geq \sqrt{40}) \leq \frac{3}{\sqrt{40}} = .47$

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

C: $P(X^2 \geq 40) = P(|X| \geq \sqrt{40})$, X, $= 3 + k \cdot 2$

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

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

Note dont know $SD(X^2)$ so cant use Chebyshev

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.

$SD(X^2)$?

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.

②

Central Limit Thm (CLT)

Let $S_n = X_1 + \dots + X_n$ where $X_1, \dots, X_n$ 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

Let $X_1, X_2, \dots X_{10}$ be i.i.d. Poisson(1).

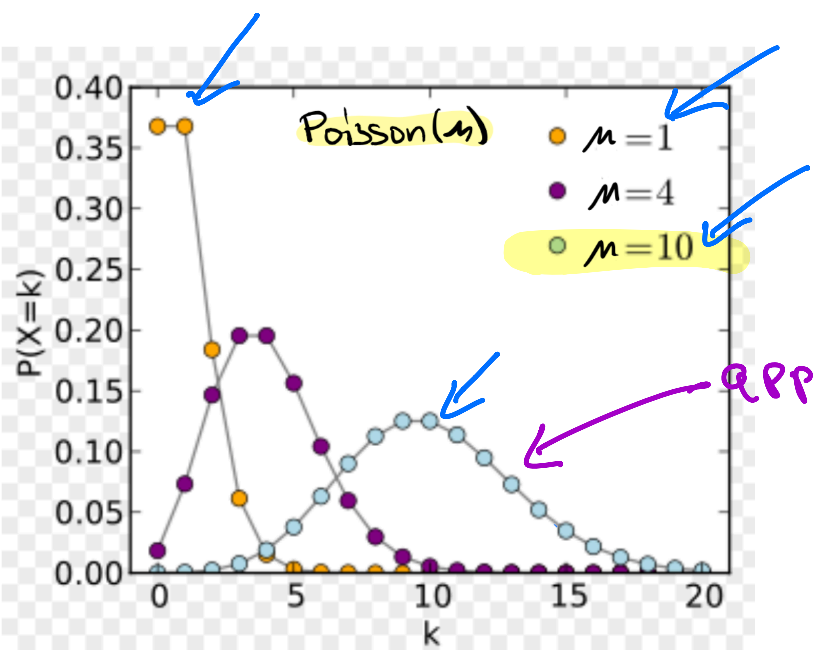
Let $S_{10} = X_1 + \dots + X_{10}$

Facts

if $X \sim Pois(1)$, $E(X) = 1$

$Var(X) = 1$

$E(S_{10}) = E(X_1 + \dots + X_{10}) = 10 E(X_1) = 10$



approx $N(10 \cdot 1, 10 \cdot 1)$

$E(X_1)$

$Var(X_1)$

③ Sec 3.6 Var of 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?

$$\boxed{P_1 = 1-\left(\frac{9}{10}\right)^{12}}$$

X = number of elevator stops.

a) Find $E(X)$

$X = I_1+\dots+I_{10}$

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

$E(X) = 10P_1$

b) Find $Var(X)$.

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

$$E(X^2) = E\left((I_1+\dots+I_{10})^2\right) = \underbrace{\sum_{1\le i,j\le 10}}\, E(I_iI_j)$$

$$\uparrow$$

$$\sum_{1\le i,j\le 10} I_iI_j$$

	I_1	—	…	I_{10}
I_1	I_1			
⋮	I_{12}	I_2	…	
I_{10}				I_{10}

all 100 indicators

I_iI_j ← column ← row

symmetric since $I_1I_2 = I_2I_1$

$$I_1I_1 = I_1 = \begin{cases} 1 & \text{if stop } 1^{st} \text{ floor} \\ 0 & \text{else} \end{cases}$$

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

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

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

$$= 1 - \left[\left(\tfrac{9}{10}\right)^{12} + \left(\tfrac{9}{10}\right)^{12} - \underbrace{\left(\tfrac{8}{10}\right)^{12}}_{\text{Prob dont stop at 1st and 2nd floor,}}\right]$$

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

$$E(x^2) = \underbrace{10E(I_1)}_{\text{diagonal}} + \underbrace{9 \cdot 10\, E(I_{12})}_{\text{non diagonal}} = 10p_1 + 10 \cdot 9\, p_{12}$$

$$(E(x))^2 = (10p_1)^2$$

$$\boxed{\text{var}(x) = 10p_1 + 10 \cdot 9\, p_{12} - (10p_1)^2}$$

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

$X = I_1 + \dots + I_n$

$p_1 = E(I_1)$

$p_{12} = p_1 \cdot p_2 = p_1^2$ — $E(I_1 I_2) = E(I_1)E(I_2)$ since I_1, I_2 indep

$$\mathrm{Var}(X) = \underbrace{np_1 + n(n-1)p_1^2}_{E(X^2)} - \underbrace{(np_1)^2}_{E(X)^2} = np_1 - np_1^2$$

$$= np_1(1-p_1)$$

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/sept28-pt1

Stat 134

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 $Var(X)$

should be 6.5

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

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

c more than one of the above

d none of the above

X = # face that appear twice

$X = I_1 + \cdots + I_6$

P_1

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

P_{12}

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

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