Stat 134 lec 13

warmup 10:00-10:10

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

ex

$p = 1/6$

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

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

a) write X as a sum of indicators

$X = I_1 + \dots + I_7$

b) Find $Var(X)$

$Var(X) = Var(I_1 + \dots + I_7) = Var(I_1) + \dots + Var(I_7)$

$= 7 \cdot 1/6 \cdot 5/6$

$pq = p(1-p) = 1/6 \cdot 5/6$

c) let $X \sim Bin(n,p)$

$Var(X) = npq$

d) let $X \sim Bin(n,p)$ with n large and p small and $np \to \mu$.

Then X is approx. $Pois(\mu)$.

$Var(X) \approx \underbrace{np}_{\to \mu} \underbrace{q}_{\to 1} \to \mu$

From warmup

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

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

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

$var(I) = pq$

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

so variance formula needs X, Y to be independent.

ex $X \sim Bin(n, p)$
$Var(X) = npq$
$SD(X) = \sqrt{npq}$

Today

(0) Properties of variance

(1) Sec 3.3 Central Limit theorem (CLT)

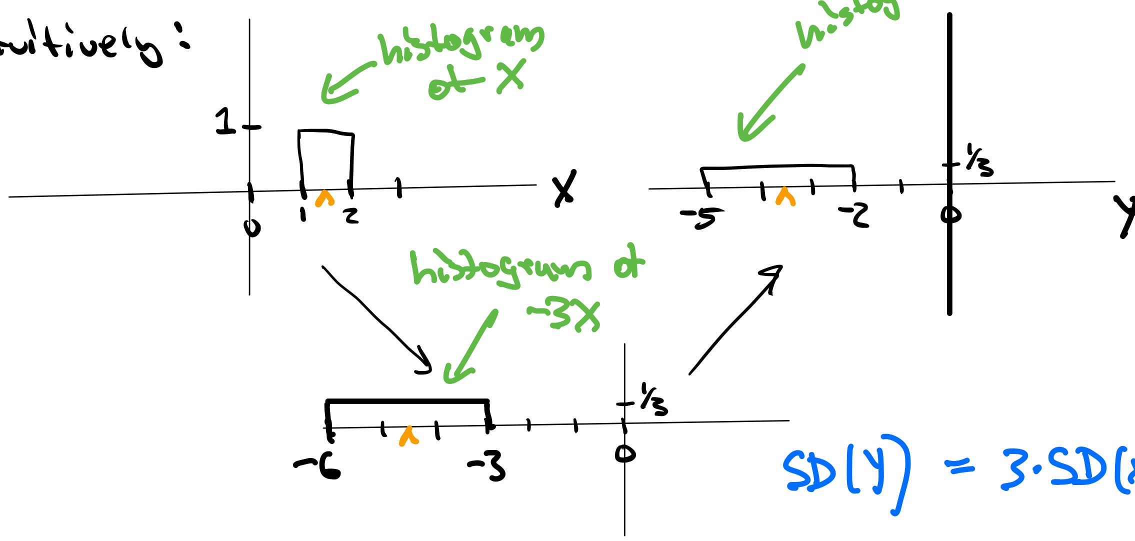
(2) Sec 3.6 (next time sec 3.4) Calculating the variance of a sum of dependent indicators.

⓪ Properties of Variance

Let $Y = -3X + 1$

How does $SD(Y)$ compare to $SD(X)$?

intuitively:



$SD(Y) = 3 \cdot SD(X)$

$$\boxed{\begin{aligned} SD(aX+b) &= |a| SD(X) \\ Var(aX+b) &= a^2 Var(X) \end{aligned}}$$

1) Sec 3.3

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,

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

approximately

often ≥ 10

ex

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

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

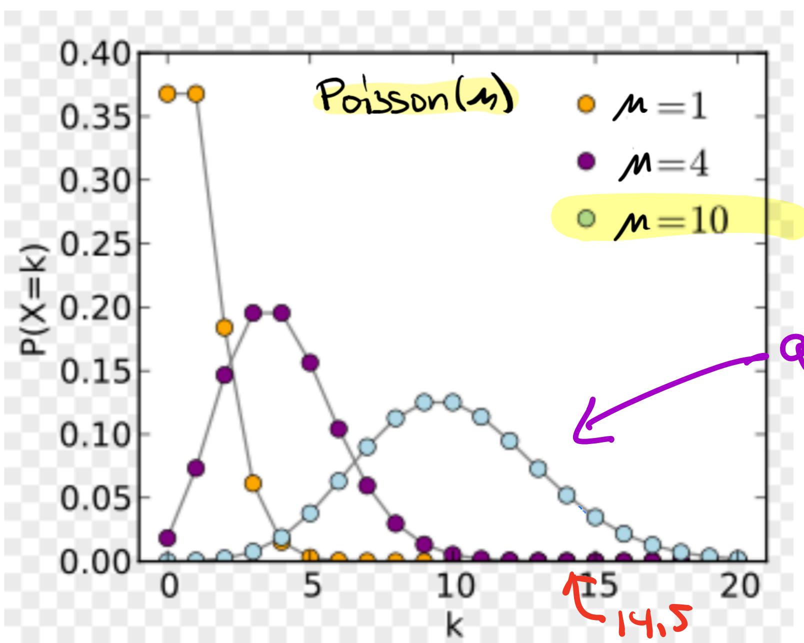
Facts

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

$Var(X) = 1$

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

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



$E(X_1)$

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

$Var(X_1)$

Approximate $P(S_{10} \geq 15)$ with continuity correction:

$$1 - \Phi\left(\frac{14.5 - 10}{\sqrt{10}}\right)$$

② 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?

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

X = number of elevator stops.

$I_2 = \begin{cases} 1 & \text{if at least 1 person gets off 2nd floor} \\ 0 & \text{else} \end{cases}$

a) Find $E(X)$ $\quad X = I_1 + \dots + I_{10}$

$E(X) = 10 P_1$

b) Find $Var(X)$.

$X = I_1 + \dots + I_{10}$ sum of dependent indicators

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

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

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

$\|$

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

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

$P_{12} = 1 - P(\text{no one gets off at 1st or 2nd floor})$

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

$\| $
I_{12}

$\vdots$

$P(A \cap B) = P\left(1 - (A \cap B)^c\right)$

$\|$
$A^c \cup B^c$

	I_1	$\cdots$	I_{10}
I_1	I_1	I_{12}	
$\vdots$		I_2	
$\vdots$		$\ddots$	
I_{10}			I_{10}

i

$\leftarrow$ 10 indicators

$I_{22} = I_2$

symmetric since $I_1 I_2 = I_2 I_1$

$E(X^2) = 10 E(I_1) + 9 \cdot 10 E(I_{12}) = 10 P_1 + 9 \cdot 10 P_{12}$

$\quad E(I_1) = P_1, \quad E(I_{12}) = P_{12}$

$Var(X) = E(X^2) - (E(X))^2 = \boxed{10 P_1 + 9 \cdot 10 P_{12} - (10 P_1)^2}$

Summary

Identically Distributed

Variance of sum of dependent i.d. indicators

$X = I_1 + \dots + I_n$

$p_1 = E(I_1)$

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

$E(X) = np_1$

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

Variance of sum of i.d. independent indicators

$X = I_1 + \dots + I_n$

$p_1 = E(I_1)$

$p_{12} = p_1 \cdot p_2 = p_1^2$

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

$$= np_1(1-p_1)$$

6. A drawer contains s black socks and s white socks, where s is a positive integer. I pull two socks out at random without replacement and call that my first pair. Then I pull two socks out at random without replacement from the remaining socks in the drawer, and call that my second pair. I proceed in this way till I have s pairs and the drawer is empty.

Let D be the number of pairs in which the two socks are of different colors.

a) Find $E(D)$.

b) Find $Var(D)$.

$D = I_1 + \cdots + I_s$

$$P_1 = 2 \cdot \frac{s}{2s} \frac{s}{2s-1} = \frac{\overset{w}{\binom{s}{1}} \overset{B}{\binom{s}{1}}}{\binom{2s}{2}}$$

$$I_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ pair diff colors} \\ 0 & \text{else} \end{cases}$$

$E(D) = sP_1$

$$P_{12} = \frac{\binom{s}{1}\binom{s}{1}}{\binom{2s}{2}} \frac{\binom{s-1}{1}\binom{s-1}{1}}{\binom{2s-2}{2}}$$

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

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

$$Var(D) = sP_1 + s(s-1)P_{12} - (sP_1)^2$$

ex (practice properties of variance)

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.

$E(W) =$ $9x - 40$ $\qquad SD(W) =$ $3y$

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.

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

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

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

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

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