2/16/18

Stat 134 lec 14.

last time $E(XY) = E(X)E(Y)$ if X, Y indep.

Next, $Var(X+Y) = Var(X) + Var(Y)$ if X, Y indep

Recall, $Var(X) = E((X-E(X))^2) = E((X-\mu_X)^2)$

$\mu_X = E(X)$.

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

$$= E\left(\left((X+Y) - (\mu_X+\mu_Y)\right)^2\right)$$

$$= E\left(\left((X-\mu_X) + (Y-\mu_Y)\right)^2\right)$$

FOIL

$$= E\left[(X-\mu_X)^2 + 2(X-\mu_X)(Y-\mu_Y) + (Y-\mu_Y)^2\right]$$

$$= \underbrace{E\left((X-\mu_X)^2\right)}_{= Var(X)} + \underbrace{2E\left((X-\mu_X)(Y-\mu_Y)\right)}_{= 2E(X-\mu_X)E(Y-\mu_Y) = 0} + \underbrace{E\left((Y-\mu_Y)^2\right)}_{= Var(Y)}$$

indep. of X, Y

Note $E(X-\mu_X) = E(X) - \mu_X = \mu_X - \mu_X = 0$

So $\boxed{Var(X+Y) = Var(X) + Var(Y)}$ when X, Y indep.

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

$X = I_1 + \cdots + I_n$ sum of indep Bernoulli (p) trials,

$E(X) = \underset{\overset{\parallel}{p}}{E(I_1)} + \cdots + \underset{\overset{\parallel}{p}}{E(I_n)} = np$

$Var(X) = \underset{\overset{\parallel}{p(1-p)}}{Var(I_1)} + \cdots + Var(I_n) = \boxed{np(1-p)}$

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

Expectation and SD of Sample sum and average

$X_1, X_2, \ldots, X_n$ iid $E(X_1) = \mu$, $SD(X_1) = \sigma$

let $S_n = X_1 + \cdots + X_n$ sample sum,

$\boxed{E(S_n) = n\mu}$ (identically distributed)

$SD(S_n) = \sqrt{n\sigma^2} = \boxed{\sqrt{n}\,\sigma}$ $\quad Var(S_n) = n\sigma^2$

let $A_n = \frac{S_n}{n}$ sample avg.

$E(A_n) = E\left(\frac{S_n}{n}\right) = \frac{E(S_n)}{n} = \boxed{\mu}$

$SD(A_n) = SD\left(\frac{S_n}{n}\right) = \frac{1}{n} SD(S_n) = \frac{1}{n}\sqrt{n}\,\sigma = \boxed{\frac{\sigma}{\sqrt{n}}}$

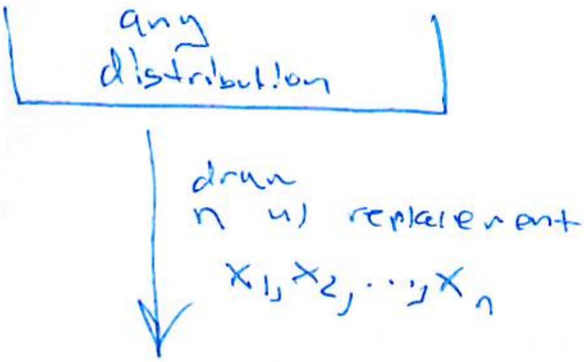
Central Limit Theorem (CLT)

$X_1, \dots, X_n$ iid mean $= \mu$, $SD = \sigma$

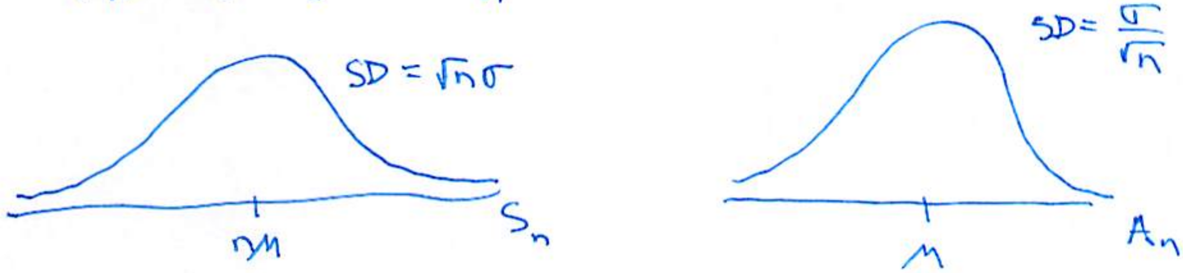
$S_n = X_1 + \dots + X_n$, $A_n = \frac{S_n}{n}$

Then for "large" n, the distribution of S_n and A_n are normal,

Picture



$S_n = X_1 + X_2 + \dots + X_n$



For CLT app:
Google Brown CLT Seeing Theory

ex (3.3.17)

Let X be a RV w/ $P(X=-100) = P(X=0) = 1/4$, $P(X=100) = 1/2$

Let S be sum of 25 indep RV each with the same dist as X.

Find $P(S=0)$

$$\begin{array}{ccc} 1 & 1 & 2 \\ \boxed{-100} & \boxed{0} & \boxed{100} \end{array}$$

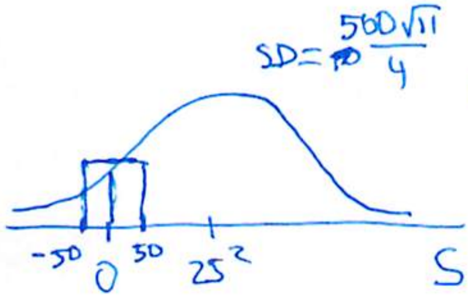
draw 25

$X_1, X_2, \ldots, X_{25}$

$$S = X_1 + \cdots + X_{25}$$

values
-2500 to 2500
increment of
100

bars
width
100.



$$E(X) = -100 \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 100 \cdot \frac{1}{2}$$
$$= \boxed{25}$$

$$SD(X) = \sqrt{E(X^2) - E(X)^2}$$

$$E(X^2) = (-100)^2 \cdot \frac{1}{4} + 0^2 \cdot \frac{1}{4} + (100)^2 \cdot \frac{1}{2}$$

$$E(X)^2 = 25^2$$

$$SD(X) = 100\frac{\sqrt{11}}{4}$$

$$E(S) = 25^2$$

$$SD(S) = 100\frac{\sqrt{11}}{4} \cdot 5$$

$$\Phi\left(\frac{50 - 25^2}{\frac{500\sqrt{11}}{4}}\right) - \Phi\left(\frac{-50 - 25^2}{\frac{500\sqrt{11}}{4}}\right) = .031$$

Finished with section 3.3.

Next Sec 3.6. See Reading Guide what to read in this section.

Understand how to find variance of a sum of non independent indicators as in next example.

Section 3.6

5

Stat 134

Friday February 16 2018

1. 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?

   a $12(1-(9/10)^{10})$

   b $10(1-(9/10)^{12})$

   c none of the above

$E(I_1) = 1 - \left(\frac{1}{10}\right)^0\left(\frac{9}{10}\right)^{12}$

X = # elevator stops

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

I_1 = indicator for at least 1 person gets off at floor 1

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

Find Var(x).

$Var(x) \neq Var(I_1) + \cdots Var(I_{10})$ since I_j not indep.

$$E(x^2) = E\left(\left(\sum_{j=1}^{10} I_j\right)^2\right)$$

$$\left(\sum_{j=1}^{10} I_j\right)^2$$

$$= \sum_{j=1}^{10} I_j + \sum_{1 \le i \ne j \le 10} I_i I_j$$

$$\begin{array}{c|cccc}
 & I_1 & I_2 & \cdots & I_{10} \\
\hline
I_1 & I_1^2 = I_1 & & & I_1 I_{10} \\
\vdots & & I_2 & & \\
 & & & \ddots & \\
I_{10} & & & & I_{10}
\end{array}$$

$I_1 I_2$ = indicator stop at 1st and 2nd floor

$\quad \parallel$

I_{12} = at least one person gets off at 1st and 2nd floor

$E(I_{12}) = 1 - \underbrace{\text{Prob(no one gets off 1st floor } \underline{\text{or}} \text{ no one gets off 2nd floor)}}$

$\qquad \parallel$

Prob (no one off 1st floor) + $\leftarrow \left(\frac{9}{10}\right)^{12}$

Prob (" " " 2nd floor) $\leftarrow \left(\frac{9}{10}\right)^{12}$ −

Prob (" " 1st and 2nd floor)

$$\left(\frac{9}{10}\right)^{12} \cdot \left(\frac{8}{9}\right)^{12}$$

$$\parallel$$

$$\left(\frac{8}{10}\right)^{12}$$

$$\underline{E(X^2)} = E\left(\sum_{j=1}^{10} I_j\right) + E\left(\sum_{1 \le i \ne j \le 10} I_i I_j\right)$$

$$= n E(I_1) + 10 \cdot 9 \, E(I_1 I_2)$$

$$Var(X) = E(X^2) - E(X)^2 = \boxed{.34}$$

<u>Next</u> understand derivation of variance of hypergeometric distribution. I worked out the details for you on the next page.

Variance of hypergeometric (N, G, n)

Prob $= G/N$

$X = I_1 + \cdots + I_n$ where $I_1 = \begin{cases} 1 & \text{If } 1^{st} \text{ element is Good.} \\ 0 & \text{els.} \end{cases}$

Know $E(X) = n\left(\frac{G}{N}\right)$.

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

$= I_j$

$$= nE(I_1) + n(n-1)E(I_1 I_2)$$

$I_1 I_2 = I_{12} = \begin{cases} 1 & \text{If } 1^{st} \text{ and } 2^{nd} \text{ element is Good} \\ 0 & \text{else} \end{cases}$ ← prob $\frac{G}{N}\frac{G-1}{N-1}$

$$\Rightarrow E(X^2) = n\frac{G}{N} + n(n-1)\frac{G}{N}\frac{(G-1)}{N-1}$$

$$\Rightarrow \text{Var}(X) = E(X^2) - E(X)^2 = n\frac{G}{N} + n(n-1)\frac{G}{N}\frac{(G-1)}{N-1} - \left(n\frac{G}{N}\right)^2$$

$$= \frac{nG}{N}\left[1 + \frac{(n-1)(G-1)}{(N-1)} - \frac{nG}{N}\right]$$

$$= \frac{nG}{N}\left[\frac{N(N-1) + (n-1)(G-1)N - nG(N-1)}{N(N-1)}\right]$$

$$= \frac{nG}{N}\left[\frac{N^2 - \cancel{N} + \cancel{nGN} - GN - nN + \cancel{N} - \cancel{nGN} + nG}{N(N-1)}\right]$$

$$= \frac{nG}{N}\left[\frac{N^2 - GN - nN + nG}{N(N-1)}\right]$$

$$= \frac{nG}{N}\left[\frac{N(N-G) - n(N-G)}{N(N-1)}\right]$$

$$= \frac{nG}{N}\left(\frac{(N-G)(N-n)}{N \cdot N-1}\right)$$

$$= \boxed{n\left(\frac{G}{N}\right)\left(\frac{N-G}{N}\right)\left(\frac{N-n}{N-1}\right)}$$

$$\Rightarrow \boxed{SD(X) = \sqrt{n\left(\frac{G}{N}\right)\left(1-\frac{G}{N}\right)} \cdot \sqrt{\frac{N-n}{N-1}}}$$