Stat 134 lec 14

warmup 10:00–10:10

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

   wrong should be 6.5

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

   c more than one of the above

   d none of the above

$p_1 = \binom{14}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{12}$

$X = I_1 + \cdots + I_6$

$I_2 = \begin{cases} 1 & \text{if 2 appears twice} \\ 0 & \text{else} \end{cases}$

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

Note

you can find P_{12} using the multiplication rule

$P_{12} = P(\text{1 appears twice})\, P(\text{2 appears twice} \mid \text{1 appears twice})$

$= \binom{14}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{12} \cdot \binom{12}{2}\left(\frac{1}{5}\right)^2\left(\frac{4}{5}\right)^{10}$

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

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

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

Announcements:

- Midterm 1: Wednesday March 2
- review sheets and practice test on website soon
- in class review Friday/Monday before test.

<u>Last time</u> <u>Sec 3.6</u>

Variance of sum of dependent i.d. indicators: identically distributed

$X = I_1 + \cdots + I_n$

$p_1 = E(I_1)$

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

$E(X) = np_1$

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

Variance of sum of i.i.d. indicators:

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

$= np_1(1-p_1)$

<u>Today</u>

1. <u>Sec 3.6</u> Hypergeometric dist.
2. <u>Sec 3.4</u> geometric distribution
3. <u>Sec 3.4</u> Negative Binomial distribution

① Sec 3.6 Hypergeometric Distribution

ex

A deck of cards has G aces.

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

above $N = 52$
$G = 4$
$n = 5$

$$E(X) = np_1$$

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

a) Find $E(X) = I_1 + \dots + I_n$

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

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

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

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

b) Find $Var(X)$

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

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

let $X \sim HG(n, N, G)$

identically distributed

$X = I_1 + \dots + I_n$ sum of dependent i.d. indicators

From above

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

where

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

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

(*)

A more useful formula for $Var(X)$:

Suppose $n = N$ then

then $X = I_1 + \dots + I_N = G$ ← constant.

So $Var(X) = 0$

so $Np_1 + N(N-1)p_{12} - (Np_1)^2 = 0$

$$\Rightarrow p_{12} = \frac{Np_1(Np_1 - 1)}{N(N-1)}$$

Note that $Np_1 = N \cdot \frac{G}{N} = G$

This is another way to write $\frac{G}{N} \cdot \frac{G-1}{N-1}$

Plug this into (*)

$$Var(X) = np_1 + n(n-1)\frac{\cancel{N}p_1(Np_1-1)}{\cancel{N}(N-1)} - (np_1)^2$$

$$= np_1\left[1 + \frac{(n-1)(Np_1-1)}{N-1} - np_1\right]$$

$$= \frac{np_1}{N-1}\left[\underbrace{(N-1) + (n-1)(Np_1-1) - np_1(N-1)}_{\substack{\| \\ N-n-Np_1+np_1 \\ \| \\ (N-n)(1-p)}}\right]$$

$$\boxed{Var(X) = np_1(1-p_1)\frac{N-n}{N-1}}$$

correction factor ≤ 1

Compare with $\boxed{Var(X) = np_1(1-p_1)}$ for $X \sim Bin(n, p_1)$

So $X \sim HG(n, N, G)$

$E(X) = n\,G/N$

$$Var(X) = n\underbrace{G/N}_{p}\Big(\underbrace{1 - G/N}_{1-p}\Big)\left(\frac{N-n}{N-1}\right)$$

1. The population of a small town is 1000 with 50% democrats. We wish to know what is a more accurate assesment of the % of democrats in the town, to randomly sample with or without replacement? The sample size is 10.

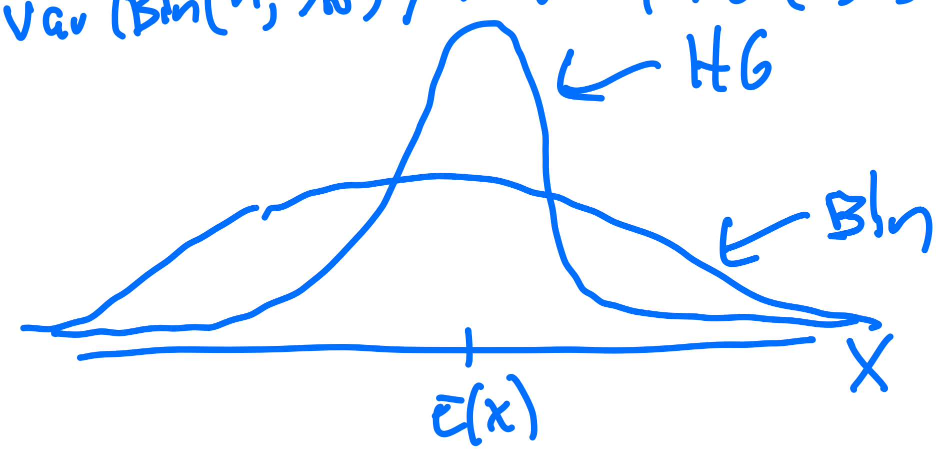
   a with replacement

   b without replacement

   c same accuracy with or without replacement

   d not enough info to answer the question

$\mathrm{Var}(\mathrm{Bin}(n, G/N)) > \mathrm{Var}(\mathrm{HG}(n, N, G))$



② $\underline{\text{Sec 3.4}}$ Geometric distribution (Geom(p))
on $\{1, 2, 3, \ldots\}$

$\underline{\text{ex}}$ $X = \#$ p coin tosses until the first head

$$P(X=k) = \underbrace{qq\cdots q}_{k-1}p = q^{k-1}p$$

Find $P(X > k)$

$$P(X=k+1) + P(X=k+2) + \ldots$$
$$\quad \overset{\shortparallel}{q^k p} \qquad\qquad \overset{\shortparallel}{q^{k+1}p}$$

$$= q^k p \left(1 + q + q^2 + \ldots\right) = \frac{q^k \cancel{p}}{\cancel{p}} = q^k$$

$$1 + q + q^2 + \ldots = \frac{1}{1-q} \quad \text{Geometric Sum}$$

Recall:

$$E(X) = P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \cdots \quad \text{Tail Sum Formula}$$
$$= P(X > 0) + P(X > 2) + \ldots = \sum_{k=0}^{\infty} P(X > k)$$

Find $E(X)$ using the tail sum formula

$$E(X) = \sum_{k=0}^{\infty} q^k = \frac{1}{1-q} = \boxed{\frac{1}{p}}$$

see appendix to notes

$\underline{\text{Fact}}$ $Var(X) = \frac{q}{p^2}$

Warning:

Some books define Geom(p) on $\{0,1,2,\dots\}$ as

Y = # failures until 1^{st} success

ex $P(Y=4) = qqqqp$

$\quad = P(X=5)$

$Y = X-1$

$E(Y) = E(X) - 1 = \frac{1}{p} - 1 = \frac{1}{p} - \frac{p}{p} = \boxed{\frac{q}{p}}$

$Var(Y) = Var(X) = \boxed{\frac{q}{p^2}}$

(4) Negative Binomial Distribution $\left(\text{NegBin}(r,p)\right)$

generalization of Geom(p)

Sum of r indep Geom(p) on $\{r, r+1, r+2, \dots\}$

ex $r=3$

$\underbrace{qqqp}_{w_1}\underbrace{qqp}_{w_2}\underbrace{p}_{w_3}$ # trials until 3^{rd} success

let $T_r \sim \text{NegBin}(r,p)$

T_r = # indep p-trials until r^{th} success

$r-1$ p in $k-1$ slots

$\underbrace{\qquad\qquad\qquad\qquad}\;p$

$$P(T_r = k) = \underbrace{\binom{k-1}{r-1} p^{r-1} q^{\overset{k-r}{k-1-(r-1)}}}\, p = \binom{k-1}{r-1} p^{r} q^{k-r}$$

$$T_r = W_1 + \cdots + W_r \quad \text{where } W_1, \ldots, W_r \overset{iid}{\sim} \text{Geom}(p)$$

$$E(T_r) = r\,E(W_1) = \boxed{\frac{r}{p}}$$

$$\text{Var}(T_r) = r\,\text{Var}(W_1) = \boxed{\frac{rq}{p^2}}$$

Appendix

Fact $Var(X) = \frac{q}{p^2}$

To find $Var(X)$ we need an identity:

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q} \quad \text{geometric sum}$$

$$\xrightarrow{\frac{d}{dq}} \quad \sum_{k=0}^{\infty} k q^{k-1} = \frac{1}{(1-q)^2}$$

$$\xrightarrow{\frac{d}{dq}} \quad \boxed{\sum_{k=0}^{\infty} k(k-1) q^{k-2} = \frac{2}{(1-q)^3} = \frac{2}{p^3}}$$

$$\begin{aligned} Var(X) &= E(X^2) - E(X)^2 \\ &= E(X^2) - E(X) + E(X) - E(X)^2 \\ &= E(\underbrace{X^2 - X}_{X(X-1)}) + \underset{\frac{1}{p}}{\underset{\shortparallel}{E(X)}} - \underset{\frac{1}{p^2}}{\underset{\shortparallel}{E(X)^2}} \end{aligned}$$

$E(g(X)) = \sum_{x \in X} g(x) P(X=x)$

$$\begin{aligned} E(X(X-1)) &= \sum_{k=1}^{\infty} k(k-1) \underset{q^{k-1}p}{\underset{\shortparallel}{P(X=k)}} \\ &= qp \sum_{k=1}^{\infty} k(k-1) q^{k-2} = qp \underbrace{\sum_{k=0} k(k-1) q^{k-2}}_{\shortparallel \; \frac{2}{p^3} \text{ (see above)}} \\ &= \frac{2q}{p^2} \end{aligned}$$

so $Var(X) = \underbrace{\frac{2q}{p^2}}_{E(X(X-1))} + \underbrace{\frac{1}{p}}_{E(X)} - \underbrace{\frac{1}{p^2}}_{E(X)^2} = \boxed{\frac{q}{p^2}}$