Stat 134 lec 15

Warmup 11:00–11:10

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

$X = I_1 + I_2 + \dots + I_6$ where $I_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ face appears twice out of 14} \\ 0 & \text{else} \end{cases}$

Recall multinomial distribution

roll a die 10 times

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

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

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

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

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~~ $\underbrace{14 * 13}_{6*5} * \binom{14}{2,2,10}(1/6)^2(1/6)^2(4/6)^{10}$

b $\binom{14}{2}(1/6)^2(5/6)^{12}$ ✓

c more than one of the above

d none of the above

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

Announcements:
- remote Midterm 1: Friday October 8
- review sheets and practice test on website soon
- in class review next Monday / Wednesday

Last time Sec 3.6

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

$X = I_1 + \dots + I_n$

$p_1 = E(I_1)$

$p_{12} = E(I_{12})$ ← $I_1 I_2$ $\neq E(I_1 I_2) = E(I_1) \cdot E(I_2)$ by indep
$E(I_1) = p_1$, $E(I_2) = p_1$

$E(X) = np_1$

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

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

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

Today

1. Sec 3.6 Hypergeometric dist.
2. Sec 3.4 geometric distribution
3. Sec 3.4 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

$$E(X) = np_1$$

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

a) Find E(X)

$X = I_1 + \dots + I_n$ where $I_2 = \begin{cases} 1 & \text{if } 2^{nd} \text{ is ace} \\ 0 & \text{else} \end{cases}$ $p_1 = G/N$

$E(X) = n\left(G/N\right)$ or $\left(5 \cdot 4/52\right)$ in this example.

b) Find Var(X)

$I_{12} = \begin{cases} 1 & \text{if } 1^{st} \text{ and } 2^{nd} \text{ card is an ace} \\ 0 & \text{else} \end{cases}$ $p_{12} = G/N \cdot \frac{G-1}{N-1}$

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

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 $\quad *$

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

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

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

$$E(X) = n\,{}^{G}/_{N}$$

$$Var(X) = n\,{}^{G}/_{N}\left(1 - {}^{G}/_{N}\right)\left(\frac{N-n}{N-1}\right)$$

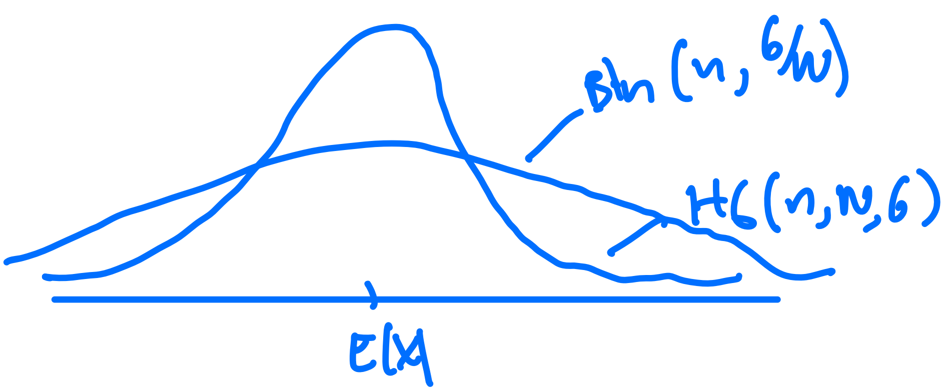
1. The population of a small town is 1000 with 50% democrats. We wish to know what is a more accurate assesment of the % of democrats in the town, to randomly sample with or without replacement? The sample size is 10.

   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 Sec 3.4 Geometric distribution (Geom(p)) on {1, 2, 3, ...}

ex X = # p coin tosses until the first head

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

Find $P(X>k)$ ← we have k failures

$$\underset{q^k p}{P(X=k+1)} + \underset{q^{k+1}p}{P(X=k+2)} + \cdots$$

$$= q^k p\,(1+q+q^2+\cdots) = \frac{q^k p}{p} = \boxed{q^k}$$

$$1+q+q^2+\cdots = \frac{1}{1-q} = 1/p$$

Recall:

$E(X) = P(X\geq 1) + P(X\geq 2) + P(X\geq 3) + \cdots$ Tail Sum Formula

$$= P(X>0) + P(X>1) + \cdots = \sum_{k=0}^{\infty} P(X>k)$$

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

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

see appendix to notes

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

Warning:

Some books define Geom(p) on $\{0, 1, 2, \ldots\}$ 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 $\left(NegBin(r, p)\right)$

generalization of Geom(p)

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

ex $r = 3$

$$\underbrace{qqqp}_{W_1}\underbrace{qqp}_{W_2}\underbrace{p}_{W_3}$$

\# trials until 3^{rd} success

let $T_r \sim NegBin(r, p)$

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

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

$\underbrace{\qquad\qquad\qquad}$ 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:

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

$$\frac{d}{dq}$$

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

$$\frac{d}{dq}$$

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

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

$$E(X(X-1)) = \sum_{k=1}^{\infty} k(k-1) \underbrace{P(X=k)}_{= q^{k-1}p}$$

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

$$= qp \sum_{k=1}^{\infty} k(k-1) q^{k-2} = qp \underbrace{\sum_{k=0}^{\infty} k(k-1) q^{k-2}}_{= \frac{2}{p^3} \text{ (see above)}}$$

$$= \frac{2q}{p^2}$$

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