Stat 134 Lec 14

warmup 9:00-9:10

X = # of coin tosses until the first head

$\sim$ Geom(p)

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

Find $P(X>k) = P(X=k+1) + P(X=k+2) + \cdots\cdots$

$$= q^k p + q^{k+1}p + \cdots\cdots$$

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

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

Announcements:

- Midterm 1: Chaps 1–3. Wednesday October 5
- review sheets and practice test on website soon
- in class review Friday/Monday before test.

Last time sec 3.6

Variance of sum of dependent i.d. indicators:

identically distributed

$$X = I_1 + \dots + I_n$$

$$p_1 = E(I_1)$$

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

$I_1 I_2$

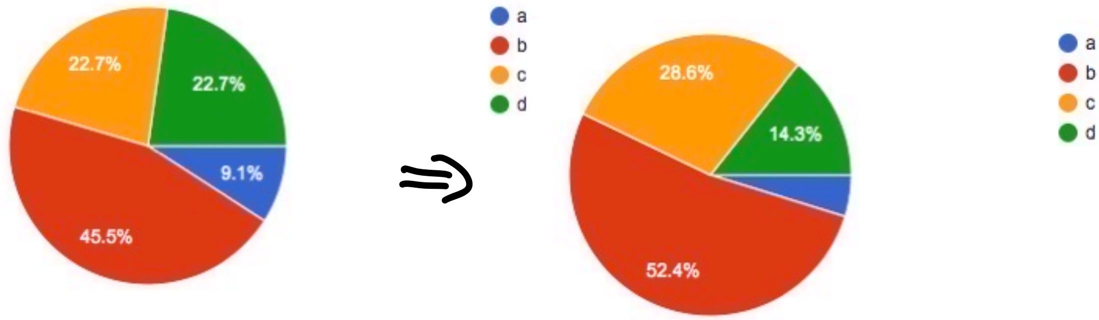
$$E(X) = np_1$$

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

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

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

$$= np_1(1-p_1)$$



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

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

   c more than one of the above

   d none of the above

b	Independent

b	Not 14*13, but 6*5

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

$$E(X) = np_1$$

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

a) Find $E(X) = I_1 + \cdots + I_n$ $p = G/N$

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

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

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

b) Find $Var(X)$

$$I_{12} = \begin{cases} 1 & \text{if first and } 2^{nd} \text{ card both aces} \\ 0 & \text{else} \end{cases}$$

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

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

constant

then $X = I_1 + \dots + I_N = G$

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 $*$

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

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

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

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

correction factor ≤ 1

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

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

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

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

tinyurl.com/sept26-2022

Stat 134

Wednesday October 2 2019

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.

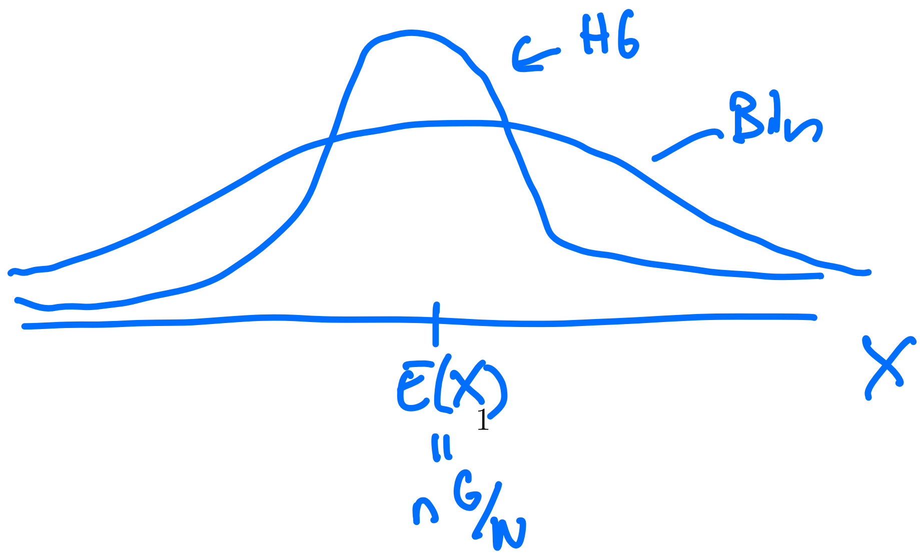
   a with replacement

   b without replacement

   c same accuracy with or without replacement

   d not enough info to answer the question

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



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

ex X = # p coin tosses until the first head

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

You showed in the warm up that

$$P(X>k) = q^k$$

Recall:

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

$= P(X>0) + P(X>2) + \cdots = \sum_{k=0}^{\infty} P(X>k)$

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 $\{0,1,2,\dots\}$ as

Y = # failures until 1^{st} success

ex $P(Y=4) = qqqqp$
$= P(X=5)$

$Y = X - 1$

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

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

ex

Coupon Collector's problem

You have a collection of boxes each containing a coupon. There are n different coupons. Each box is equally likely to contain any coupon independent of the other boxes.

X = # boxes needed to get all n different coupons.

ex $n=3$ $X = X_1 + X_2 + X_3$

$\underbrace{c}_{X_1}\underbrace{c\,c\,c}_{X_2}\underbrace{c\,c\,c\,c}_{X_3}$

$\sim Geom(3/3)$ $\sim Geom(2/3)$ $\sim Geom(1/3)$

a) What is the distribution of X_1, X_2, X_3?
Are they independent? Yes

b) What is $E(X) = E(X_1) + E(X_2) + E(X_3) = \boxed{3\left(1+\frac{1}{2}+\frac{1}{3}\right)}$

$E(X_1) = \frac{1}{3/3}$, $E(X_2) = \frac{1}{\frac{2}{3}}$, $E(X_3) = \frac{1}{\frac{1}{3}}$

$$\text{Var}(X_1) = \frac{q}{p^2}$$

c) What is Var(X)?

$$= \frac{\frac{0}{3}}{\left(\frac{3}{3}\right)^2} + \frac{\frac{1}{3}}{\left(\frac{2}{3}\right)^2} + \frac{\frac{2}{3}}{\left(\frac{1}{3}\right)^2} = 3\left(0+\frac{1}{2^2}+\frac{2}{1^2}\right)$$

Appendix

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

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

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

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

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

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

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

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

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

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

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