Stat 134 lec 15

- in class Midterm Friday March 8
- review sheet coming today
- in class review next Wednesday.

Last time sec 3.6

Variance of sum of dependent indicators

$X = I_1 + \cdots + I_n$

$p_1 = E(I_1)$

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

$E(X) = np_1$

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

Today

1. student responses from concept test
2. finish sec 3.6 Hypergeometric dist.
3. sec 3.4 geometric distribution
   negative binomial distribution

Monday February 24 2019

1

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

c	a is the nondiagonals and b is the diagonals.

$P_1 = \binom{14}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^{12}$

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

	I_1	I_2	I_3	I_4	I_5	I_6
I_1	I_1	I_{12}				
I_2	I_{12}	I_2				
I_3			I_3			
I_4				I_4		
I_5					I_5	
I_6						I_6

b	Not 14x13, should be 6x5 (36 - 6 diagonal) for E(X^2)

② <u>Sec 3.6</u> Hypergeometric Distribution

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

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

last time we saw:

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

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

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

Plug this into ✱

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

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

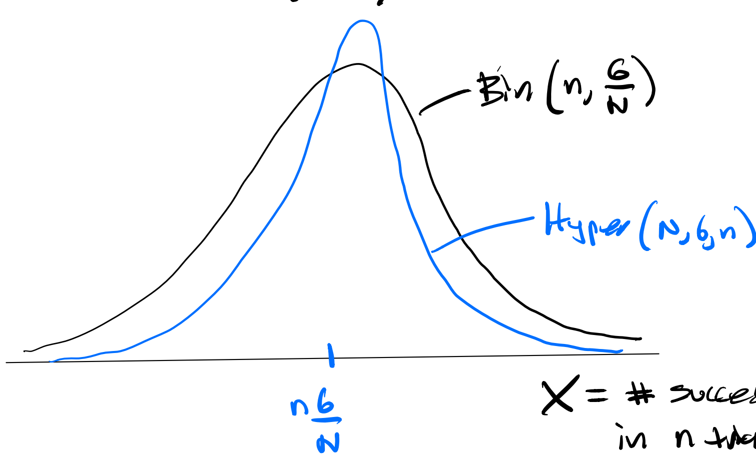
$$= \frac{np_1}{N-1}\Big[\underbrace{(N-1) + (n-1)(Np_1-1) - np_1(N-1)}_{= N-n-Np_1+np_1 \;=\; (N-n)(1-p)}\Big]$$

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

correction factor ≤ 1

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

$$Var(hyper(N,6,n)) \le Var\left(Bin\left(n, \frac{6}{N}\right)\right)$$



X = # successes (good) in n trials

Each success that occurs reduces the probability of a subsequent success, so you get values of X closer to the mean than the binomial.

③ Sec 3.4 Geometric distribution (Geom(p))

ex $X =$ # p coin tosses until the first head

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

Find $P(X>k)$

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

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

$$= q^{k}p\left(1+q+q^{2}+\cdots\right) = q^{k}\cancel{p}\cdot\frac{1}{\underset{\underset{\cancel{p}}{\parallel}}{1-q}} = \boxed{q^{k}}$$

Find

$$E(X) = \sum_{k=1}^{\infty} P(X\geq k) = P(X\geq 1)+P(X\geq 2)+\cdots$$

$$= P(X>0)+P(X>1)+\cdots$$

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

$$= \frac{1}{1-q} = \boxed{\frac{1}{p}}$$

To find Var(X) we need an identity:

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

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

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

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

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

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

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

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

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

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

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

Warning:

Some books define $\text{Geom}(p)$ as

$Y = \#$ failures until 1^{st} success

ex $P(Y=4) = qqqqp$

$\quad \shortparallel$

$P(X=5)$

$Y = X - 1$

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

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

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

generalization of Geom(p)

sum of r indep Geom(p).

<u>ex</u> $r = 3$

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

trials until 3rd success

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

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

$$\underbrace{\qquad\qquad\qquad\qquad}_{}\ p$$

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

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

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

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

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

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.

eg $n=3$ $\quad X = X_1 + X_2 + X_3$

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

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

$$\left.\begin{aligned} X_1 &\sim \text{Geom}\left(\tfrac{3}{3}\right) \\ X_2 &\sim \text{Geom}\left(\tfrac{2}{3}\right) \\ X_3 &\sim \text{Geom}\left(\tfrac{1}{3}\right) \end{aligned}\right\} \text{indep}$$

b) What is $E(\overbrace{X}^{X_1+X_2+X_3})$

$E(X_1) = \dfrac{1}{p_1}$

$$E(X) = \frac{1}{\frac{3}{3}} + \frac{1}{\frac{2}{3}} + \frac{1}{\frac{1}{3}} = 3\left(1 + \frac{1}{2} + \frac{1}{3}\right)$$

c) What is $\text{Var}(X)$?

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

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

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

soln for n coupons:

X_1 = # boxes to 1^{st} coupon $\sim Geom\left(\frac{n}{n}\right)$

$X_1 + X_2$ = # boxes to 2^{nd} coupon so $X_2 \sim Geom\left(\frac{n-1}{n}\right)$

$\vdots$

$X_1 + \cdots + X_n$ = # boxes to n^{th} coupon so $X_n \sim Geom\left(\frac{1}{n}\right)$

$X = X_1 + \cdots + X_n$ sum of indep Geom with diff. p.

$$E(X) = \underset{\frac{n}{n}}{E(X_1)} + \underset{\frac{n}{n-1}}{E(X_2)} + \underset{\frac{n}{n-2}}{E(X_3)} + \cdots + \underset{\frac{n}{1}}{E(X_n)}$$

$$\boxed{E(X) = n\left(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)}$$

$$\boxed{Var(X) = n\left(\frac{0}{n^2} + \frac{1}{(n-1)^2} + \cdots \frac{n-1}{1^2}\right)}$$