Stat 134 lec 15

in class midterm next Friday
- review sheet coming today
- in class review next Wed.
- review in section.

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

sec 3.6 Hypergeometric(N, G, n)

sec 3.4 geometric(p)

negative binomial (r, p)

Sec 3.4 Hypergeometric (N, G, n)

ex X = # aces in 5 cards drawn w/o replacement from deck of 52

$N = 52$

$G = 4$

$n = 5$

$X = I_1 + \cdots + I_n$

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

$p_1 = G/N$

$I_1 = \begin{cases} 1 & \text{if } 1^{st} \text{ trial is good} \\ 0 & \text{else} \end{cases}$

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

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

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

algebra

$$= np_1\underbrace{(1-p_1)}_{q_1}\frac{N-n}{N-1}$$

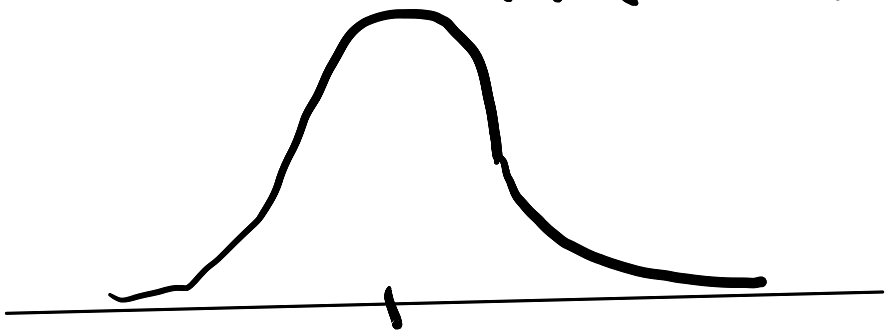
correction factor

Compare with binomial

$$\text{var}(X) = npq$$

Each success (failure) that occurs reduces the probability of subsequent success (failure). So you get values of X closer to the mean.

$\text{var}(X) = n\left(\frac{G}{N}\right)\left(1-\frac{G}{N}\right)\left(\frac{N-n}{N-1}\right)$



$n\frac{G}{N}$

$X = \#$ good in n trials

As $N \to \infty$, hypergeom becomes binomial.

Sec 3.4

$X \sim \text{Geometric}(p)$ $\quad$ $1, 2, 3, \ldots$

ex $X = \#$ p trials before first success

$P(X=k) = q^{k-1}p$

You have a p coin.

$X = \#$ coin tosses until first heads

Find $P(X>k)$?

X has values $x = 1, 2, 3, \ldots$

$$P(X>k) = \underset{\substack{\shortparallel \\ q^k p}}{P(X=k+1)} + \underset{\substack{\shortparallel \\ q^{k+1} p}}{P(X=k+2)} + \ldots$$

$$q^k p\,(\underbrace{1+q+q^2+\dots}_{=\,\frac{1}{1-q}\,=\,\frac{1}{p}}) = \boxed{q^k}$$

A useful identity.

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

$\frac{d}{dq}$

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

$\frac{d}{dq}$

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

Find $E(X)$ and $\mathrm{var}(X)$.

use tail sums formula for $E(X)$ (lec 11)

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=0}^{\infty} \underbrace{P(X > k)}_{=\,q^k} = \frac{1}{1-q} = \boxed{\frac{1}{p}}$$

$$\boxed{\mathrm{var}(X) = \frac{q}{p^2}}$$

Warning:

some books define geom(p) as
Y = # failures until 1^{st} success,

$Y = X - 1$

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

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

ex

qqqqp $X = 5$
$Y = 4$

sum of r indep geom(p)

Negative binomial dist w/ param r, p

generalization of geom(p) — $r = 1$

ex $r = 3$

$\underbrace{qqqp}_{W_1}\underbrace{qqp}_{W_2}\underbrace{p}_{W_3}$ # trials until 3^{rd} success

$T_r \sim \text{Negbin}(r, p)$ — $r, r+1, \ldots$

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

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

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

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

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

$r-1$ p in $k-1$ spots

$$\binom{k-1}{r-1} p^{r-1} q^{k-r}$$

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.

Find $E(X)$, $Var(X)$.

ex $n=3$

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

$$X = X_1 + X_2 + X_3$$

$$\left.\begin{array}{l} X_1 \sim Geom\left(\frac{3}{3}\right) \\ X_2 \sim Geom\left(\frac{2}{3}\right) \\ X_3 \sim Geom\left(\frac{1}{3}\right) \end{array}\right\} \text{indep.}$$

$p_1 = \frac{3}{3}$, $q_1 = 0/3$

$p_2 = \frac{2}{3}$, $q_2 = \frac{1}{3}$

$p_3 = \frac{1}{3}$, $q_3 = \frac{2}{3}$

$$\left.\begin{array}{l} E(X_1) = \frac{1}{p} \\ Var(X_1) = \frac{q}{p^2} \end{array}\right\} =$$

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

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

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

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$\left(\frac{n-1}{n}\right)$

$\vdots$

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

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

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

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

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

Sec 3.6 (Appendix

Thm Let $X \sim \text{Hypergeometric}(N, G, n)$

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

$$\text{Var}(X) = n\left(\frac{G}{N}\right)\left(1-\frac{G}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Proof

let $X = I_1 + \cdots + I_n$, $\quad I_1 = \begin{cases} 1 & \text{if } 1^{st} \text{ elt good} \\ 0 & \text{else} \end{cases}$ $\quad P_1 = \frac{G}{N}$

$\Rightarrow E(X) = nP_1 = \boxed{n\frac{G}{N}}$

$I_{12} = I_1 I_2 = \begin{cases} 1 & \text{if } 1^{st} \text{ and } 2^{nd} \text{ elt. good} \\ 0 & \text{else} \end{cases}$ $\quad P_{12} = \frac{G}{N} \cdot \frac{G-1}{N-1}$

$$E(X^2) = nP_1 + n(n-1)P_{12}$$

$$\text{Var}(X) = \underbrace{nP_1 + n(n-1)P_{12}}_{E(X^2)} - \underbrace{(nP_1)^2}_{E(X)^2}$$

$$= n\frac{G}{N} + n(n-1)\frac{G}{N} \cdot \frac{(G-1)}{N-1} - \left(n\frac{G}{N}\right)^2$$

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

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

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

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

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

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

Sec 3.4 (Appendix)

Thm $X \sim \text{Geometric}(p)$, # trials to 1^{st} success, $x=1,2,\dots$

then $E(X) = \frac{1}{p}$

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

Pf/ $$\underbrace{E(X)}_{\text{tail sum}} = \sum_{k=0}^{\infty} P(X>0) = \sum_{k=0}^{\infty} q^k = \frac{1}{1-q} = \boxed{\frac{1}{p}}$$

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

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

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

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

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