Lecture 10

Thursday, February 6, 2020 1:59 PM

- Section 12.9 (conditional tests)
- Chapter 13
  - Binomial
  - Multinomial
- Goodness of fit tests

Recall

- In the presence of nuissance parameters, we have two paths for hypothesis testing.
  - → GLR based asymptotic test
  - → Exact conditional test (if we have sufficient statistic for nuissance parameter)

Ex

$$\begin{cases} X \sim Bin(n_1, p_1) \\ Y \sim Bin(n_2, p_2) \\ X \perp Y \end{cases}$$

- Want to construct a size α test for:

$$\begin{cases} H_0: p_1 = p_2 = p \\ H_A: p_1 < p_2 \end{cases}$$

- In general, since $X \perp Y$, $f_{XY}(x,y) = \binom{n_1}{x}\binom{n_2}{y} p_1^x (1-p_1)^{n_1-x} \; p_2^y (1-p_2)^{n_2-y}$
- Under H_0:

$$f_{XY}(x,y) = \binom{n_1}{x}\binom{n_2}{y}\left(\frac{p}{1-p}\right)^{x+y}(1-p)^{n_1+n_2}$$

$$= \underbrace{\binom{n_1}{x}\binom{n_2}{y}}_{h(x,y)} \exp\left\{ \underbrace{(x+y)}_{S=s(x,y)} \underbrace{\ln\left(\frac{p}{1-p}\right)}_{\eta(p)} \right\} \underbrace{(1-p)^{n_1+n_2}}_{c(p)}$$

- we want something of form $f_{X,Y|S}$ ...
- we know that given $S=s$, $Y = s - X$.
  - ↳ Thus given s, we can find the value of Y & X

ie, $P\{(X=x) \cap (Y=y) \cap (S=s)\}$

$$= P\{(X=x) \cap (Y = s-x)\}$$

- $$f_{X|S} = \frac{f_{X,S}(x,s)}{f_S(s)}$$

- $$f_{X,S}(x,s) = f_{X,Y}(x, s-x)$$

$$= \binom{n_1}{x}\binom{n_2}{s-x}\left(\frac{p}{1-p}\right)^{x+(s-x)} \cdot (1-p)^{n_1+n_2}$$

$$= \binom{n_1}{x}\binom{n_2}{y}\left(\frac{p}{1-p}\right)^{s}(1-p)^{n_1+n_2} \quad \text{for} \begin{cases} s = 0, \ldots, n_1+n_2 \\ x = 0, \ldots, s \end{cases}$$

$$= \binom{n_1}{x}\binom{n_2}{y}\left(\frac{p}{1-p}\right)^s (1-p)^{n_1+n_2} \quad \text{for } \begin{cases} s = 0, \dots, n_1+n_2 \\ x = 0, \dots, s \end{cases}$$

- then the marginal pdf for S gives:

$$f_S(s) = \sum_{x=0}^{s} \binom{n_1}{x}\binom{n_2}{s-x}\left(\frac{p}{1-p}\right)^s (1-p)^{n_1+n_2}$$

↳ then since $\sum_{x=0}^{s} \binom{n_1}{x}\binom{n_2}{s-x} = \binom{n_1+n_2}{s}$,

we can simplify $f_S(s)$ like so...

$$f_S(s) = \binom{n_1+n_2}{s}\left(\frac{p}{1-p}\right)^s (1-p)^{n_1+n_2}$$

- Then, $$f_{X|S} = \frac{\binom{n_1}{x}\binom{n_2}{s-x}\cancel{\left(\frac{p}{1-p}\right)^s}\cancel{(1-p)^{n_1+n_2}}}{\binom{n_1+n_2}{s}\cancel{\left(\frac{p}{1-p}\right)^s}\cancel{(1-p)^{n_1+n_2}}}$$

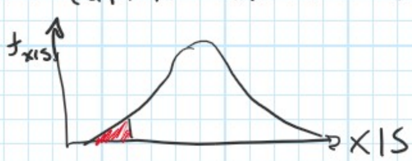
$$= \frac{\binom{n_1}{x}\binom{n_2}{s-x}}{\binom{n_1+n_2}{s}} \quad \text{for } \begin{matrix} x = 0, 1, \dots, s \\ s = 0, 1, \dots, n_1+n_2 \end{matrix}$$

↳ Hyper-Geometric Distribution
(not dependent on nuissance (p))

- We can find an exact (albeit conservative test) based on $f_{X|S}$

$H_0: p_1 = p_2 = p$

$H_A: p_1 < p_2$



- Thus, the form of the rejection region is:

$$F_{X|S}(x|s) < \alpha$$

Theorem: 12.9.1

(Recall) Def 10.2.1 of Jointly Sufficient Statistics

- Let $\underline{X} \sim f_{\underline{x}}(\underline{x}; \theta)$ and let $S = (S_1, \dots, S_k)$ be a k dimensional statistic.
- Then, $S_1, \dots, S_k$ are jointly sufficient for θ if for any other vector $\underline{T}$, the conditional pdf given $S = s$, $f_{\underline{T}|\underline{S}}(\underline{t}|\underline{s})$, does not depend on θ.

Thm:

- Suppose we are concerned with θ in the presence of nuissance parameters $\kappa = (\kappa_1, \dots, \kappa_m)$.

If $\underline{X} \sim f_{\underline{x}}(x; \theta, \underline{\kappa}) = c(\theta, \kappa)\, h(\underline{x}) \exp\left\{\sum_{i=1}^{m} \kappa_i s_i(\underline{x}) + \theta t(\underline{x})\right\}$, of nuissance parameters $\underline{K} = (K_1, \ldots, K_m)$:

If $\underline{X} \sim f_{\underline{x}}(\underline{x}; \theta, \underline{K}) = c(\theta, \underline{K})\, h(\underline{x}) \exp\left\{ \sum_{i=1}^{m} K_i\, s_i(\underline{x}) + \theta t(\underline{x}) \right\}$,

then $\underline{S} = (S_1, \ldots, S_m)$ is jointly sufficient for $\underline{K}$ for each θ and the conditional pdf of $f_{T|\underline{S}}(t|s; \theta)$ does not depend on $\underline{K}$.

① A size α test for $H_0: \theta \le \theta_0$, $H_A: \theta > \theta_0$ rejects H_0 if $t(\underline{x}) \ge k(s)$ where;

$$P[T \ge k(\underline{s}) \mid \underline{s}] = \alpha$$

② The inequalities are flipped for the opposite hypothesis.

Chapter 13

"How well does our data support our assumption of our model??"

Contingency Table

Ex. Type of pet

Pet \ Owner	Female	Not-Female	Total
Cat	16	28	44
Dog	24	32	56
Total	40	60	100

Question: does the proportion of females who own dogs equal the proportion that own cats.

- Let p_1 = the proportion who own dogs

H_0: $p_1 = 1 - p_1$ OR ($p_{1,0} = 0.5$)

H_A: $p_1 \ne p_{1,0}$

- If x is the count of females who own dogs...

$X \sim Bin(n = 40, p_{1,0} = 0.5)$ under H_0

$E(x) = np_{10}$, $Var(x) = np_{10}(1 - p_{10})$

... we can come up with this test via the asymptotic exact test.

... the 2nd approach

$\hat{p} = \frac{x}{n} \sim N\left(p_{10}, var(\hat{p}) = \frac{p_{10}(1-p_{10})}{n}\right)$ under H_0.

$$Z = \frac{\frac{x}{n} - p_{10}}{\sqrt{\frac{p_{10}(1-p_{10})}{n}}} \sim N(0,1)$$