Recap :→ LRT for 1-parameter Case

9th Feb' 2024.

*) Show that the cross product $\sum_{i=1}^{10} 2[(x_i - \bar{x})(\bar{x} - \mu_0)]$ is equal to zero.

$$\sum_{i=1}^{10} 2[(x_i - \bar{x})(\bar{x} - \mu_0)]$$

$$= 2(\bar{x} - \mu_0) \sum_{i=1}^{10} (x_i - \bar{x})$$

$$= 2(\bar{x} - \mu_0) \left[\sum_{i=1}^{10} x_i - n\bar{x} \right]$$

$$= 2(\bar{x} - \mu_0) [n\bar{x} - n\bar{x}] \quad \cdots\cdots \left[\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow \sum_{i=1}^{n} x_i = n\bar{x} \right]$$

$$= 2(\bar{x} - \mu_0) * 0$$

$$= 0$$

So, $\sum_{i=1}^{10} 2[(x_i - \bar{x})(\bar{x} - \mu_0)] = 0$

Multi-parameter Case

LRS for testing, $H_0 : \underset{\sim}{\theta} = \underset{\sim}{\theta}_0$; where $\underset{\sim}{\theta}_0$ is a vector of numerical values

$$D \equiv -2r(\underset{\sim}{\theta}_0) = 2[l(\hat{\underset{\sim}{\theta}}) - l(\underset{\sim}{\theta}_0)]$$

where, $\hat{\underset{\sim}{\theta}}$ is the joint MLE of $\underset{\sim}{\theta}$, and $D \geq 0$.

E.g. in Example 4.2.1, if δ was unknown and we wanted to estimate it as well (i.e., get joint MLE$(\hat{\mu},\hat{\delta})$), an example of a simple hypothesis could be

$$H_0 : \mu = 226, \text{ and } \delta = 1.$$

Assuming $H_0 : \underset{\sim}{\theta} = \underset{\sim}{\theta}_0$ is true, $D \approx \chi^2_{(k)}$ where k is the # of functionally independent parameters.

e.g. if $\underset{\sim}{\theta} = (\theta_1, \theta_2, \cdots, \theta_b)$, then $1 \leq k \leq b$.

p-value $\simeq \mathbb{P}(\chi^2_{(k)} \geq d_{obs})$, assuming H_0 is true.

For Example:- In Ex 4.2.1, if δ was not assumed known, one possible simple hypothesis is,

$$H_0 : \mu = 226,\ \delta = 1 \Rightarrow D \approx \chi^2_{(2)}$$

For Example :- In case of Multinomial Distribution,

Suppose, $(x_1, x_2, x_3, x_4) \sim \text{Multinomial}(n, \underbrace{p_1, p_2, p_3, p_4}_{})$

↳ 4 parameters

If 3 are known, then 4^{th} is defined by $\sum_{i=1}^{4} p_i = 1$.

$$\therefore k = 4 - 1 = 3$$

So, $D \approx \chi^2_{(3)}$ $\boxed{\text{Not, } \chi^2_{(4)}}$

Let's look at an Example; Example 4.2.3 from the Complete Lecture Notes.

Here, $(x_1, x_2, \cdots, x_7) \sim \text{Multinomial}(n, p_1, p_2, \cdots, p_7)$

$H_0 : p_1 = p_2 = \cdots = p_7 = 1/7 \quad ; \; p_0 = (1/7, 1/7, \cdots, 1/7)$

$D = 2[\ell(\hat{\underset{\sim}{p}}) - \ell(\underset{\sim}{p_0})]$

We need to maximize $\ell(\underset{\sim}{p})$ subject to $\sum_{i=1}^{7} p_i = 1$.

$\ell(\underset{\sim}{p}) = \sum_{i=1}^{7} x_i \ln(p_i) \quad ; \quad p_i \in (0, 1]$

$$\frac{\partial \ell(\underset{\sim}{p})}{\partial p_i} = x_i / p_i \quad \text{for } i = 1, 2, \cdots, 7.$$

Substitute, $p_7 = 1 - \sum_{i=1}^{6} p_i \longrightarrow \left[p_n = 1 - \sum_{i=1}^{n-1} p_i\right]$.

$$\ell(\underset{\sim}{p}) = \sum_{i=1}^{6} x_i \ln(p_i) + x_7 \ln\left(1 - \sum_{i=1}^{6} p_i\right)$$

Set of Equations $\left\{ \frac{\partial \ell(\underset{\sim}{p})}{\partial p_i} = \frac{x_i}{p_i} - \frac{x_7}{1 - \sum_{i=1}^{6} p_i} \right.$, for $i = 1, 2, \cdots, 6$

$$= \frac{x_i}{p_i} - \frac{x_7}{p_7}$$

At joint MLE $\hat{\underset{\sim}{p}}$, all partials $= 0$

$$\therefore \frac{x_i}{\hat{p}_i} = \frac{x_7}{\hat{p}_7} \quad , \text{ for } i = 1, 2, \cdots, 6$$

$$\Rightarrow \hat{p}_i = \frac{x_i}{x_7}\hat{p}_7 \text{, for } i=1,2,\cdots,6 \quad \cdots (\star)$$

$\hookrightarrow$ Dependent MLEs, so we have to find one and the replace to get another one.

We know, $\hat{p}_7 = 1 - \sum_{i=1}^{6} \hat{p}_i = 1 - \frac{\hat{p}_7}{x_7}\sum_{i=1}^{6} x_i$

$$\Rightarrow \hat{p}_7 = 1 - \frac{\hat{p}_7}{x_7}(n - x_7) \quad \cdots\cdots \left[n = \sum_{i=1}^{7} x_i = \sum_{i=1}^{6} x_i + X_7 \Rightarrow \sum_{i=1}^{6} x_i = n - x_7 \right]$$

$\Rightarrow$ $\vdots$ [Algebra]

$$\Rightarrow \hat{p}_7 = x_7/n \quad \cdots \text{ Substitute this in } (\star)$$

we get, $\hat{p}_i = \frac{x_i}{\cancel{x_7}} * \frac{\cancel{x_7}}{n} = x_i/n$

So, $\boxed{\hat{p}_i = x_i/n}$; for $i = 1,2,\cdots,6$

$$D = 2\left[l(\hat{\underset{\sim}{p}}) - l(\underset{\sim}{p_0}) \right]$$

$$d_{obs} = 2\left[\sum_{i=1}^{7} x_i \ln\left(\frac{x_i}{n}\right) - \sum_{i=1}^{7} x_i \ln(1/7) \right]$$

plug in $n = 63$, and all x_i's from table,

$d_{obs} = 23.27$,

$\therefore$ p-value $\Rightarrow \mathbb{P}\left[\chi^2_{(6)} \geq 23.27\right] \simeq 0.0007$

Conclusion:-

Under the null hypothesis, the estimated expected frequencies are $E(x_i) = np_0 = 63(1/7) = 9$. The p-value of our test was 0.0007, showing very strong evidence against H_0. The data do not support the hypothesis that fatal heart attacks are equally likely to occur on any day of the week.

Next week:- Sec 4.3 [LRT for Composite Hypothesis]