11/23/20

Michael Velez

Lec 20 Proof of the one parameter (one θ) Likelihood Ratio Test (LRT) being asymptotically distributed as χ^2_1. This proof follows C&B p.g. 489. Let the DGP $\overset{iid}{\sim} f(x;\theta)$ and we're testing against $H_0: \theta = \theta_0$

$$\hat{\Lambda} = 2\ln(\hat{LR}) = 2(l(\hat{\theta}^{MLE}; x_1, \dots, x_n) - l(\theta_0; x_1, \dots, x_n))$$

Just like in lec 11 when we proved the asymptotic normality of the MLE, we considered a Taylor series. This one is a bit different than the one in lec 11. We want to approx. $l(\theta_0;\dots)$ centered at $\hat{\theta}^{MLE}$

$$l(\theta_0;x_1,\dots,x_n) = l(\hat{\theta}^{MLE};x_1,\dots,x_n) + \overbrace{(\theta_0-\hat{\theta}^{MLE})}^{0}\, \underbrace{l'(\hat{\theta}^{MLE};x_1,\dots,x_n)}_{0} + \tfrac{1}{2}(\theta_0-\hat{\theta}^{MLE})^2 \cdot l''(\hat{\theta}^{MLE};x_1,\dots,x_n) + \dots$$

$$\Rightarrow l(\hat{\theta}^{MLE};x_1,\dots,x_n) - l(\theta_0;x_1,\dots,x_n) \approx \tfrac{1}{2}(\theta_0-\hat{\theta}^{MLE})^2\, l''(\hat{\theta}^{MLE};x_1,\dots,x_n)$$

we ignore all terms past the second order just like the MLE normality proof by assuming technical conditions (see L8B)

$$\Rightarrow \hat{A} \approx -(\theta_0-\hat{\theta}^{MLE})^2\, l''(\hat{\theta}^{MLE};x_1,\dots,x_n)\cdot \frac{l''(\theta;x_1,\dots,x_n)}{l''(\theta;x_1,\dots,x_n)}$$

$$= -(\theta_0-\hat{\theta}^{MLE})^2\, l''(\theta;x_1,\dots,x_n)\, \overbrace{\frac{l''(\hat{\theta}^{MLE};x_1,\dots,x_n)}{l''(\theta;x_1,\dots,x_n)}}^{\hat{A}_1}$$

$$= \frac{(\theta_0-\hat{\theta}^{MLE})^2}{\frac{1}{-l''(\theta;x_1,\dots,x_n)}}\hat{A}_1 = \frac{(\theta_0-\hat{\theta}^{MLE})^2}{\frac{1}{\frac{I(\theta)}{I(\theta)}\, n\frac{1}{n}\sum_{i=1} -l''(\theta;x_i)}}\hat{A}_1$$

$$= \frac{(\theta_0 - \hat{\theta}_{MLE})^2}{\frac{1}{nI(\theta)}} \overbrace{\frac{\frac{1}{n}\sum -l''(\theta; x_i)}{I(\theta)}}^{\hat{A}_2} \hat{A}_1$$

Recall: $I(\theta) = E[-l''(\theta; x)]$

$$= \Bigg(\underbrace{\frac{\theta_0 - \hat{\theta}^{MLE}}{\sqrt{\frac{I(\theta)^{-1}}{n}}}}_{\hat{B}} \Bigg)^2 \hat{A}_1 \hat{A}_2 \xrightarrow{d} \chi_1^2$$

- $\hat{B} \xrightarrow{d} N(0,1)$ by the asymptotic normality of MLE thm.

- $\hat{B}^2 \longrightarrow B^2 \sim \chi_1^2$ due to a $\overbrace{\text{continuous mapping thm}}^{CMT}$:
  $\hat{B} \xrightarrow{d} B \Rightarrow h(\hat{B}) \xrightarrow{d} h(B)$

- $\frac{1}{n}\sum -l''(\theta; x_i) \xrightarrow{p} I(\theta)$ by LLN $\Rightarrow \hat{A}_2 \xrightarrow{p} 1$ by CMT

- $\hat{\theta}^{MLE} \xrightarrow{p} \theta \Rightarrow l''(\hat{\theta}^{MLE}; x_1, \ldots, x_n) \xrightarrow{p} l''(\theta, x_1, \ldots, x_n)$
  $\Rightarrow \hat{A}_1 \xrightarrow{p} 1$ by CMT

- Thm: $\hat{C}_1 \xrightarrow{p} c_1$, $\hat{C}_2 \xrightarrow{p} c_2 \Rightarrow \hat{C}_1 \hat{C}_2 \xrightarrow{p} c_1 c_2 \Rightarrow A$

The LRT is much more flexible and general than just one parameter example.
For ex), assume $\overset{iid}{\sim} f(x; \theta_1, \dots \theta_K)$ i.e. K parameters. And you want to test against:

$$H_0: \theta_1 = \theta_{1_0} \text{ and } \theta_2 = \theta_{2_0} \text{ and} \dots \theta_K = \theta_{K_0}$$

v.s

H_a: at least one of these inequalities is false

$$\hat{\Lambda} = 2\ln\left(\frac{l(\hat{\theta}_1^{MLE}, \dots, \hat{\theta}_K^{MLE}; x_1, \dots, x_n)}{l(\theta_{1_0}, \dots, \theta_{K_0}; x_1, \dots, x_n)}\right) \xrightarrow{d} \chi^2_K$$

Let's see an example of this. Remember the die roll setting. As Robin said there are only 5 parameters since $\theta_1 + \theta_2 + \dots + \theta_6 = 1$ and thus if you know 5 of these θs, you automatically know the 6^{th} one. If we wish to prove the die is unfair, then the H_0 hypothesis is:

$$H_0: \theta_1 = \theta_2 = \dots = \theta_6 = \frac{1}{6}$$

H_a: at least one inequality is wrong

$$\hat{\Lambda} = 2\ln(\hat{LR}) \xrightarrow{d} \chi^2_5 \quad , \quad F_{\chi^2}(11.07) = 95\%$$

$$\hat{LB} = \prod_{i=1}^{n} \frac{l(\hat{\theta}_1^{MLE}, \dots, \hat{\theta}_5^{MLE}; x_i)}{l(\frac{1}{6}, \dots, \frac{1}{6}; x_i)}$$

$n_1 = \#\,1\text{'s}, \dots, n_5 = \#\,5\text{'s}$

$\hat{\theta}_1^{MLE} = \frac{n_1}{n}, \dots, \hat{\theta}_5^{MLE} = \frac{n_5}{n}$

$\Rightarrow \hat{\theta}_6^{MLE} = 1 - (\hat{\theta}_1^{MLE} + \dots + \hat{\theta}_5^{MLE})$

$$= \underbrace{\frac{\left(\frac{n_1}{n}\right)^{n_1} + \left(\frac{n_2}{n}\right)^{n_2} \cdots \left(\frac{n_5}{n}\right)^{n_5} \left(1 - \left(\frac{n_1}{n} + \dots \frac{n_5}{n}\right)\right)^{n-(n_1+\dots+n_5)}}{\left(\frac{1}{6}\right)^{n_1} \left(\frac{1}{6}\right)^{n_2} \cdots \left(\frac{1}{6}\right)^{n_5} \left(\frac{1}{6}\right)^{n-(n_1+\dots+n_5)}}}$$

$$\left(\frac{1}{6}\right)^n = 6^{-n}$$

$$\hat{\Lambda} = 2\left(n_1 \ln\left(\frac{n_1}{n}\right) + \dots + n_5 \ln\left(\frac{n_5}{n}\right) + (1 - (\ ))\ln\left(\frac{n-(\ )}{n}\right)\right)$$

$$= 2\Big(4\ln\left(\tfrac{4}{15}\right) + 1\ln\left(\tfrac{1}{15}\right) + 3\ln\left(\tfrac{3}{15}\right) + 2\ln\left(\tfrac{2}{15}\right) + 1\ln\left(\tfrac{1}{15}\right) + 4\ln\left(\tfrac{4}{15}\right) + 15\ln(6)$$

$= 4.056 \approx 3.8 < 11.07 \Rightarrow$ Retain H_0

from previous lecture

They're not the same test (LRT statistic is not the same as the Pearson GOF statistic) both with their own advantages / dis.

The most general LRT is for iid $f(x; \theta_1, ..., \theta_k)$ DGP and you wish to test an arbitrary subset of the k parameters of size $k_0 \leq k$

e.g. for $k = 20$, $H_0: \theta_2 = \theta_{2_0}$ and $\theta_7 = \theta_{7_0}$ and $\theta_{17} = \theta_{17_0} \Rightarrow k_0 = 3$

$$\hat{LR} = \frac{l(\hat{\theta}_1^{MLE}, ..., \hat{\theta}^{MLE})}{l(\hat{\theta}_1^{MLE}, \theta_{2_0}, \hat{\theta}_3^{MLE} ... \hat{\theta}_6^{MLE}, \theta_{7_0}, \hat{\theta}_8^{MLE} ... \hat{\theta}_{16}^{MLE}, \theta_{17_0}, \hat{\theta}_{18}^{MLE}, \hat{\theta}_{20}^{MLE}}$$

$$\xrightarrow{d} \chi^2_{k_0 = 3}$$

The numerator has k "degrees of freedom" i.e parameters to fit and the denominator has $k - k_0$ "degrees of freedom". So the difference in dimension between top and bottom is $k - (k - k_0) = k_0$

This situation is a very classic and famous situation. The top is called the "full model" and the bottom is called the "reduced model" and the reduced model is "nested in" the full model because the reduced model has a parameter space which is a subspace of the full model's parameter space.

One thing to be careful of: the MLE's in the reduced model sometimes will be functions of the pinned values (the theory in H_0). The MLE's in the reduced model are conditional on these values.

Let's see this in action. Test
H_0: DGP is $N(0, \theta_2)$ i.e. $H_0: \theta_1 = 0$
and the full model is $\stackrel{iid}{\sim} N(\theta_1, \theta_2)$.
This yields a beautiful test statistic:

$$\hat{LR} = \frac{L(\hat{\theta}_1^{MLE}, \hat{\theta}_2^{MLE}; x_1, \dots, x_n)}{L(0, \hat{\theta}_2^{MLE} \mid \theta_1 = 0; x_1, \dots, x_n)} = \prod_{i=1}^{n} \frac{\frac{1}{\sqrt{2\pi(\bar{a}-\bar{x}^2)}} e^{-\frac{1}{2(\bar{a}-\bar{x}^2)}(x_i-\bar{x})^2}}{\frac{1}{\sqrt{2\pi\bar{a}}} e^{-\frac{1}{2\bar{a}} x_i^2}}$$

$$= \left(\frac{\bar{a}}{\bar{a}-\bar{x}^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2}\left(\frac{1}{\bar{a}-\bar{x}^2} \underbrace{\sum(x_i-\bar{x})^2}_{n(\bar{a}-\bar{x}^2)} - \frac{1}{\bar{a}} \underbrace{\sum x_i^2}_{n\bar{a}}\right)}$$

In lecture 7,

$$= \left(\frac{\bar{a}}{\bar{a}-\bar{x}^2}\right)^{\frac{n}{2}} \Rightarrow \boxed{\hat{\Lambda} = n \ln\left(\frac{\bar{a}}{\bar{a}-\bar{x}^2}\right)} \sim \chi^2_1 = 3.84$$

$\hat{\theta}_1^{MLE} = \bar{x}$

$$\hat{\theta}_2^{MLE} = \hat{\sigma}^2 = \frac{1}{n} \underbrace{\sum (x_i - \bar{x})^2}_{n(\bar{a}-\bar{x}^2)} \stackrel{\text{lec 7}}{=} \frac{1}{n} \underbrace{\sum x_i^2}_{\bar{a}} - \bar{x}^2$$

$$\hat{\theta}_2^{MLE} \mid \theta_1 = \frac{1}{n} \sum (x_i - \theta_1)^2 = \frac{1}{n} \overbrace{\sum x_i^2}^{n\bar{a}} = \bar{a}$$