Lecture 20 11/23/2020

Proof of the one parameter (one theta) likelihood Ratio Test (LRT) being asymplotically distributed as a χ^2 with 1 df. This proof follows C&B p489. Let the DGP be iid $F(X;\theta)$ and we're testing against $H_0: \theta = \theta_0$.

$$\hat{\lambda} = 2\ln(\hat{LR}) = 2(\ell(\hat{\theta}^{MLE}; X_1, \dots, X_n) - \ell(\theta_0; X_1, \dots, X_n))$$

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

$$\ell(\theta_0; X_1, \dots, X_n) = \ell(\hat{\theta}^{MLE}; X_1, \dots, X_n) + (\theta_0 - \hat{\theta}^{MLE})\ell'(\hat{\theta}^{MLE}; X_1, \dots, X_n) + \tfrac{1}{2}(\theta_0 - \hat{\theta}^{MLE})^2 \cdot \ell''(\hat{\theta}^{MLE}; X_1, \dots, X_n) + \dots$$

$$\Longrightarrow \ell(\hat{\theta}^{MLE}; X_1, \dots, X_n) - \ell(\theta_0; X_1, \dots, X_n) \approx -\tfrac{1}{2}(\theta_0 - \hat{\theta}^{MLE})^2 \ell''(\hat{\theta}^{MLE}; X_1, \dots, X_n)$$

We ignore all terms past the second order just like in the MLE normality proof by assuming technical conditions (see C&B)

$$\Longrightarrow \hat{\lambda} \approx -(\theta_0 - \hat{\theta}^{MLE})^2 \ell''(\hat{\theta}^{MLE}; X_1, \dots, X_n) \cdot \frac{\ell''(\theta; X_1, \dots, X_n)}{\ell''(\theta; X_1, \dots, X_n)}$$

$$= -(\theta_0 - \hat{\theta}^{MLE})^2 \ell''(\theta; X_1, \dots, X_n) \underbrace{\frac{\ell''(\hat{\theta}^{MLE}; X_1, \dots, X_n)}{\ell''(\theta; X_1, \dots, X_n)}}_{\hat{A}_1}$$

Recall: $\mathcal{I}(\theta) = E[-\ell''(\theta; X)]$

$$= \frac{(\theta_0 - \hat{\theta}^{MLE})^2}{\frac{1}{-\ell''(\theta; X_1, \dots, X_n)}} \hat{A}_1 = \frac{(\theta_0 - \hat{\theta}^{MLE})^2}{\frac{\mathcal{I}(\theta)}{\mathcal{I}(\theta)} \frac{1}{n \frac{1}{n}\sum_{i=1}^{n} -\ell''(\theta; X_i)}} \hat{A}_1$$

$$= \frac{(\theta_0 - \hat{\theta}^{MLE})^2}{\frac{1}{n\mathcal{I}(\theta)}} \cdot \overbrace{\frac{\frac{1}{n}\sum -\ell''(\theta; X_i)}{\mathcal{I}(\theta)}}^{\hat{A}_2} \hat{A}_1$$

$$= \Bigg(\overbrace{\frac{\theta_0 - \hat{\theta}^{MLE}}{\sqrt{\frac{\mathcal{I}(\theta)}{n}}}}^{\hat{B}}\Bigg)^2 \hat{A}_1 \hat{A}_2 \xrightarrow{d} \chi^2_1$$

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

  $\hat{B}^2 \xrightarrow{d} B^2 \sim \chi^2_1$ due to a $\overbrace{\text{cont. mapping thm}}^{\text{CMT}}$: $\hat{B} \xrightarrow{d} B$
  $\Rightarrow h(\hat{B}) \xrightarrow{d} h(B)$

- $\frac{1}{n}\sum -\ell''(\theta; X_i) \xrightarrow{P} \mathcal{I}(\theta)$ by LLN $\Rightarrow \hat{A}_2 \xrightarrow{P} 1$ by CMT

- $\hat{\theta}^{MLE} \xrightarrow{P} \theta \Rightarrow \ell''(\hat{\theta}^{MLE}; X_1, \ldots, X_n) \xrightarrow{P} \ell''(\theta; X_1, \ldots, X_n)$
  $\Rightarrow \hat{A}_1 \xrightarrow{P} 1$ by CMT

- Thm: $\hat{C}_1 \xrightarrow{P} C_1$, $\hat{C}_2 \xrightarrow{P} C_2$, $\Rightarrow \hat{C}_1\hat{C}_2 \xrightarrow{P} C_1C_2$
  $\Rightarrow \hat{A}_1\hat{A}_2 \xrightarrow{P} 1$

The LRT is much more general and flexible than just this one parameter example. For example, assume iid $f(X; \theta_1, \ldots, \theta_k)$. i.e. k parameters. And you want to test against:

$$H_0: \theta_1 = \theta_{1_0} \text{ and } \theta_2 = \theta_{2_0} \text{ and} \ldots \text{and } \theta_k = \theta_{k_0} \text{ vs}$$

H_a: at least one of these inequalities is false.

$$\hat{\Lambda} = 2\ln\left(\frac{\mathcal{L}(\hat{\theta}_1^{MLE}, \ldots, \hat{\theta}_k^{MLE}; X_1, \ldots, X_n)}{\mathcal{L}(\theta_{1_0}, \ldots, \theta_{k_0}; X_1, \ldots, X_n)}\right) \xrightarrow{d} \chi^2_k.$$

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

$H_0: \theta_1 = \theta_2 = \ldots = \theta_6 = 1/6$ and H_a: at least one inequality is incorrect.

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

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

$$\hat{LR} = \prod_{i=1}^{n} \frac{f(\hat{\theta}_1^{MLE}, \ldots, \hat{\theta}_5^{MLE}; x_i)}{f(1/6, \ldots, 1/6; x_i)}$$

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

$$= \frac{\left(\frac{n_1}{n}\right)^{n_1}\left(\frac{n_2}{n}\right)^{n_2} \cdot \ldots \cdot \left(\frac{n_5}{n}\right)^{n_5} \left(1 - \left(\frac{n_1}{n} + \ldots + \frac{n_5}{n}\right)\right)^{n-(n_1+\ldots+n_5)}}{\underbrace{\left(1/6\right)^{n_1}\left(1/6\right)^{n_2} \cdot \ldots \cdot \left(1/6\right)^{n_5}\left(1/6\right)^{n-(n_1+\ldots+n_5)}}_{(1/6)^n = 6^{-n}}}$$

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

$= \text{numerator} \cdot 6^n$.

$$\hat{\lambda} = 2\left(n_1 \ln(n_1/n) + \ldots + n_5 \ln(n_5/n) + (\quad)\ln(\quad) + n\ln(6)\right)$$

$$= 2\left(4\ln(4/15) + 1\ln(1/15) + 3\ln(3/15) + 2\ln(2/15) + 1\ln(1/15) + 4\ln(4/15) + 15\ln(6)\right)$$

$$= 4.056 \approx 3.8 < 11.07 \Rightarrow \text{Retain } H_0.$$

The LRT statistic is not the same as the pearson GOF statistic because they're different testing procedures, both asymptotic and both with their own dis/advantages.

The most general LRT is for an iid $f(x; \theta_1, \ldots, \theta_K)$ 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 $\theta_{17} = \theta_{17_0} \Rightarrow K_0 = 3$

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

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

The numerator has K "degrees of freedom" i.e. parameters to fit and the denominator has $K - K_0$ "degrees of freedom" i.e. parameters to fit. 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 theories in H_0). The MLE's in the reduced model are conditional on those values.

Let's see this in action. Test H_0: DGP is normal with mean zero i.e. $H_0: \theta_1 = 0$ which means the reduced model is iid $N(0, \theta_2)$ and the full model is iid $N(\theta_1, \theta_2)$. This yields a beautiful test statistic:

$$\hat{LR} = \frac{\mathcal{L}(\hat{\theta}_1^{MLE}, \hat{\theta}_2^{MLE}; x_1, \ldots, x_n)}{\mathcal{L}(0, \hat{\theta}_2^{MLE} \mid \theta_1 = 0; x_1, \ldots, x_n)}$$

$$= \prod_{i=1}^{n} \frac{\frac{1}{\sqrt{2\pi(\bar{a}-\bar{x}^2)}} e^{-\frac{1}{2(\bar{a}-\bar{x}^2)}(x_i-\bar{x})^2}}{\frac{1}{\sqrt{2\pi\bar{a}}} e^{-\frac{1}{2\bar{a}} x_i^2}}$$

In lec 7.

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

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

$\hat{\theta}_2^{MLE} = \hat{\sigma}^2 = \frac{1}{n} \underbrace{\sum(x_i-\bar{x})^2}_{n(\bar{a}-\bar{x}^2)} \overset{\text{lec 7}}{=} \frac{1}{n} \overbrace{\sum x_i^2}^{\bar{a}} - \bar{x}^2 .$ $= \left(\frac{\bar{a}}{\bar{a}-\bar{x}^2}\right)^{n/2} \Rightarrow \hat{\Lambda} = n \ln \left(\frac{\bar{a}}{\bar{a}-\bar{x}^2}\right)$

$\hat{\theta}_2^{MLE} | \theta_1 = \frac{1}{n} \sum (x_i - \theta_1)^2 \overset{\text{if } \theta_1 = 0}{=} \frac{1}{n} \underbrace{\sum x_i^2}_{n\bar{a}} = \bar{a}$

which is compared to the quantile of χ_1^2, which is 3.84.