Lecture 7

9.8. Generalized Likelihood Ratio Tests

In 9.1-9.7 we have been concerned with defining optimality of tests and defending optimality properties of specifically constructed tests. We have also seen that **uniformly most powerful tests** in the set of **all** α-tests exist only in **exceptionally simple situations**. Therefore, in more complicated situations, we introduced the *unbiasedness* requirement to discard some unwanted tests in our search and to finally be able to find one that we would like to call "best". Other approaches (not discussed but just listed here) to limit the number of tests that can compete for the "best" title include:

- looking at the power function *only locally around* θ_0 (which leads to *locally most powerful* α*-tests*);
- asking for certain invariance properties to be satisfied by the test statistic

We stress once again on the fact that the reason to consider so many different criteria comes from the lack of universally best α test in situations *more general* than just the one of simple hypothesis versus simple alternative.

Now we have reached a stage, similar to the one we had earlier in the context of estimation: we introduced further restrictions (unbiasedness), considered relatively complicated procedures (Lehmann-Scheffe approach) to derive optimal estimators. But recall that in **estimation** context, at the end of our discussion, these optimal estimators turned out to be asymptotically very close to an *easy-to-derive "universal"* estimator - the **MLE**. The question is now in **Hypothesis testing** context: Is there again a similar "universal" and easy to apply test with good (meaning close to optimal asymptotically) properties?

The answer is **Yes** and it is given be the **Generalized Likelihood Ratio Test** (GLRT). Let us discuss this test briefly.

9.8.1 General formulation

When no points in the parameter space specified by H_0 are preferred to others, the likelihood function can be maximized under the null and alternative hypotheses:

Assume, for example, that $\Theta \subseteq R^{r+s} = R^k; \Theta_0 = \{\theta \in \Theta \mid \theta_1 = \theta_{10}, .., \theta_r = \theta_{r0}\}, H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta \backslash \Theta_0$. Let us define the statistic

If Θ_0 is true : ratio $\to$ 1 (very close together)

$$\lambda(\mathbf{X}) = \frac{\sup_{\theta \in \Theta_0} L(\mathbf{X}, \theta)}{\sup_{\theta \in \Theta} L(\mathbf{X}, \theta)}$$

restricted MLE

MLE

which is obviously in the interval [0,1]. Intuitively, it makes sense to define the rejection region as $S = \{\mathbf{x} \mid \lambda(\mathbf{x}) \leq C\}$ for a certain constant C. Indeed, small value of $\lambda(\mathbf{x})$ would imply that other values of theta, different from the ones specified under the null hypothesis, are more likely to have generated our data set! Therefore, we should reject H_0.

However, the optimum properties of likelihood ratios for simple hypotheses, as discussed in the NP lemma, no longer apply in the present context, except asymptotically. In addition, the exact distribution of $\lambda(\mathbf{X})$ is needed in order to be able to determine the constant C that would force the test to be a level- α test. This exact distribution is only rarely possible to find analytically and hence the need arises to use *asymptotic approximation* for it to make it usable at least for large samples. Hereby, the fact that the *deviance statistic* has a well known asymptotic distribution turns out to be very helpful in our endeavours to get the constant C for a given level α.

Bearing in mind the derivations in 8.6, one can formulate the following, slightly more general, theorem:

9.8.2. Theorem: Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a random sample from $f(x, \theta), \theta \in R^{r+s}$. Suppose the regularity conditions for consistency and asymptotic normality of MLE under H_0 and H_1 hold. Then under $H_0 : -2\ln\lambda(\mathbf{X}) \to^d \chi^2_r$

Note 1: r is interpreted as the difference between the number of free parameters when there is no restriction and the number of free parameters under the zero hypothesis H_0. Indeed when H_0 does not hold, all $k = r + s$ components of the parameter θ can vary "freely" in Θ. In the case where H_0 is assumed to hold, the first r components are **fixed** and only the remaining s components are allowed to vary, that is, "are free". The difference between the number of free components without the restrictions of the null hypothesis and the number of free components under the restrictions of the null is therefore equal to $k - s = r$ (and determines the degrees of freedom for the limiting χ^2 distribution).

Note 2: The case of $r = 1, s = 0$, i.e. $H_0 : \theta = \theta_0 \in R^1$ versus $H_1 : \theta \neq \theta_0$ was, in fact, derived already in 8.6. In this case taking sup over the Θ_0 values is not necessary since $\Theta_0 = \{\theta_0\}$(has a single element only).

9.8.3. **Examples** (see lectures)**:**

i) For testing $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$ for a sample of n i.i.d. $N(\mu, \sigma^2)$, (σ^2 known) one has $-2\ln\lambda(\mathbf{x}) = \frac{n(\bar{\mathbf{x}}-\mu_0)^2}{\sigma^2} = D(\mu_0) \sim \chi^2_1$ (and the result is *exact)*.

ii) Normal sample with both μ and σ^2 unknown. Testing $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. In this case, the rejection constant can be determined precisely (not just asymptotically (!)) by examining **directly** the statistic $\lambda(\mathbf{X})$ rather than $-2ln\lambda(\mathbf{X})$. The GLR test turns out to be *equivalent* to the classical t-test with a rejection region

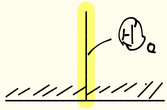
$$S_1 = \{\mathbf{X} || \frac{(\bar{\mathbf{X}} - \mu_0)\sqrt{n}}{s} | \geq k\}$$

Here $s = (s^2)^{1/2} = [\frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{\mathbf{x}})^2]^{1/2}$ and to make the size equal to α, we must choose $k = t_{\alpha/2,(n-1)}$(the upper $(\alpha/2)$.100% point of the t distribution with $(n-1)$ degrees of freedom.) Thus, a popular test (known to be also the UMPU α-test for the above problem) could be constructed using the GLR Test methodology.

If instead of the exact calculations, we utilized the approximate test-construction given by Theorem 9.8.2, we have to use the statistic $-2ln\lambda(\mathbf{X})$ instead of $\lambda(\mathbf{X})$ and in that case we end up with a test with a rejection region $S_2 = \{\mathbf{X} : n\log\frac{\hat{\sigma}^2}{\tilde{s}^2} > \chi^2_{1,\alpha}\}$ (here $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2, \tilde{s}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$.) The two rejection regions S_1 and S_2 are not quite the same but when the sample size is large, they are getting quite close.

Indeed, it is a good exercise for you to show that:

$H_0 : \mu = \mu_0$ (σ unknown) | $H_1 : \mu \neq \mu_0$ σ unknown



$$\tilde{S} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad \sup_{\theta \in \Theta} L(X, \theta) = \frac{1}{(\sqrt{2\pi})^n ((\tilde{S})^2)^{n/2}} \exp\left(-\frac{n}{2}\right)$$

$$\hat{\mu} = \bar{X} \qquad \theta = \begin{pmatrix}\mu \\ \sigma\end{pmatrix}$$

$$\left(-\frac{1}{2\tilde{S}^2} \boxed{\sum_{i=1}^{n}(X_i - \bar{X})^2 = n\tilde{S}^2}\right.$$

$$= -\frac{n}{2}$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$S = \sqrt{S^2}$$

$$\Downarrow$$

$H_0 : \mu = \mu_0$, σ unknown

$$T = \frac{\sqrt{n}\,|\bar{X} - \mu_0|}{S} > t$$

$$\ln L = -\frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_0)^2$$

$$\frac{\partial}{\partial\sigma^2}\ln L = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(X_i - \mu_0)^2 = 0$$

$$\rightarrow \hat{\sigma}^2_{H_0} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu_0)^2 \qquad \left(-\frac{1}{2\hat{\sigma}^2_{H_0}}\sum_{i=1}^{n}(X_i - \mu_0)^2 = -\frac{n}{2}\right)$$

$$\sup_{\substack{\theta \in \Theta_0 \\ \mu \to \mu_0}} L(X, \theta) = \frac{1}{(\sqrt{2\pi})^n}\frac{1}{(\hat{\sigma}_{H_0})^{n/2}}\exp\left(-\frac{n}{2}\right)$$

$$\Lambda = \left(\frac{\tilde{S}^2}{\hat{\sigma}^2_{H_0}}\right)^{n/2} \text{ reject} \rightarrow \text{small}$$

$$\Leftrightarrow \frac{\sum_{i=1}^{n}(X_i - \mu_0)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \overset{?}{>} k$$

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu_0)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} \overset{?}{>} k$$

$$\Leftrightarrow \frac{n(\bar{X} - \mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \overset{?}{>} K$$

$\hookrightarrow$ same as

$$T = \frac{\sqrt{n}\,|\bar{X} - \mu_0|}{\sqrt{\frac{1}{n-1}\sum(X_i - \bar{X})^2}} \overset{?}{>} K$$

$\hookrightarrow$ $t_{\frac{\alpha}{2}, n-1}$

Asymptotic:

$$-2\ln\Lambda = -2\log\left[\frac{(\tilde{S}^2)^{n/2}}{(\hat{\sigma}^2_{H_0})^{n/2}}\right]$$

$$= n\log\left(\frac{\hat{\sigma}^2_{H_0}}{\tilde{S}^2}\right) = W$$

Test reject H_0 if $W > \chi^2_{1,\alpha}$

$$T = \frac{\text{stanr.}}{\sqrt{\chi^2/df}} \Rightarrow T^2 = \frac{\overbrace{\text{std}^2/1}^{\chi^2_1}}{\chi^2/df} \sim F_{1,df}$$

$$-2\ln\Lambda \quad , \quad n\ln\left(1+\frac{T^2}{n-1}\right) \approx n\frac{T^2}{n-1} \approx T^2$$

$$\ln(1+x) \approx x$$

- S_1 means that $T^2 > \chi^2_{1,\alpha}$ where $T = \frac{(\bar{\mathbf{X}}-\mu_0)\sqrt{n}}{s}$ is the usual T statistic.
- S_2 means that $nln(1+\frac{T^2}{n-1}) > \chi^2_{1,\alpha}$. However, for large n, we have $nln(1+\frac{T^2}{n-1}) \approx \frac{nT^2}{n-1} \approx T^2$

which convinces us that the two rejection regions coincide asymptotically.

Further examples, where exact calculations of the GLRT are not possible but the methodology of Theorem 9.8.2 does work, can also be given (time permitting).

9.9. Alternatives to the GLRT. The GLRT is widely used but there are circumstances where other test procedures may be preferred. We shall define two of these test procedures for the case $s = 0$, that is, $k = r$.

9.9.1. Score test. It uses $S = V(\mathbf{X},\theta_0)' I^{-1}_{\mathbf{X}}(\theta_0) V(\mathbf{X},\theta_0)$ instead of $-2\log\lambda(\mathbf{X})$.

9.9.2. Wald test. It uses $(\hat{\theta}-\theta_0)' I_{\mathbf{X}}(\hat{\theta})(\hat{\theta}-\theta_0)$ instead of $-2\log\lambda(\mathbf{X})$ where θ_0 is the hypothetical vector and $\hat{\theta}$ is the MLE.

For both the Score test, and the Wald test, the *asymptotic* distribution of the test statistic under H_0 is the *same* as the distribution of the GLRT statistic (that is, chi-square with $r = k$ degrees of freedom). Score tests have a numerical advantage in comparison to GLRT and the Wald test, that they do *not* require the MLE to be calculated! Specifically in the econometrics literature, the Score test is known as **Lagrange Multiplier Test**. The name comes from its alternative derivation in which the Likelihood function is maximized subject to the restrictions of H_0 and the maximization method uses Lagrange multipliers.

Much research has been devoted to selecting one of the three tests as a preferred test in a particular situation for relatively small sample sizes. We shall only say that this is a difficult task and will not be discussed in our course.

It is left for you as an exercise (cf. Example 4.10 in the textbook but be aware of the misprint there) that the following is true:

Let X be the number of successes in a binomial experiment with a probability of success p. We wish to test $H_0 : p = p_0$ versus $H_1 : p \neq p_0$. Denote $\hat{p} = \frac{X}{n}$. Then the Score statistic is

$$S = \frac{(X - np_0)^2}{np_0(1-p_0)},$$

the Wald statistic is

$$W = \frac{(X - np_0)^2}{n\hat{p}(1-\hat{p})}$$

and

$$-2\ln\lambda = 2\{X \ln\frac{\hat{p}(1-p_0)}{p_0(1-\hat{p})} + n\ln\frac{1-\hat{p}}{1-p_0}\}.$$