MSF100
Statistical Inference Principles - Lecture 9

Rebecka Jörnsten
Mathematical Statistics
University of Gothenburg/Chalmers University of Technology

February 13, 2012

1 Testing - recap

We construct a test statistic $\delta(\tilde{X})$ such that $\delta(\tilde{X}) = 1$ if X falls in the rejection region R', or, equivalently, the sufficient statistic $T(\tilde{X})$ falls in the rejection region R. We want maximum power $\beta(\theta)$ for θ outside the null parameter space, that is maximize

$$\beta(\theta) = P_\theta(T(\tilde{X}) \in R)$$

for $\theta \in \Theta_0^c$. We also want $\beta(\theta)$ to be as small as possible for $\theta in \Theta_0$. Usually, we fix the power or type I error rate at α such that $\beta(\theta \in \Theta_0) \leq \alpha$.

There are many tests that can have $\beta(\theta) \leq \alpha$, or possibly $= \alpha$. We search among these for the most powerful test, which may differ depending on which $\theta \in \Theta_0^c$ we consider. A one-sided test has no power in one direction whereas a two-sided test has less power away from the null.

We call a test with power function β *uniformly most powerful, UMP* if $\beta(\theta) \geq \beta'(\theta), \;\; for \; all \; \theta \in \Theta_0^c$ for any other level α test with power function β'. This is a pretty tall order!

The Neyman-Pearson lemma states that a UMP exists in the simple-vs-simple hypothesis case with $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$. The UMP test is defined by rejection region

$$R = \{x : \frac{f(x|\theta_1)}{f(x|\theta_0)} > k_\alpha\}, \;\; s.t. \;\; P_{\theta_0}(X \in R) = \alpha$$

Proof: The test is a size α test by the above. Further more, from the definition of R we have that

$$x \in R : \;\; f(x|\theta_1) > k_\alpha f(x|\theta_0)$$

$$x \notin R : \;\; f(x|\theta_1) < k_\alpha f(x|\theta_0)$$

We denote the test where the above holds by $\delta(X)$. Now consider another level α test with test function δ'. We denote the power functions of the two tests by β and β' respectively.

- $(\delta(x) - \delta'(x))(f(x|\theta_1) - k_\alpha f(x|\theta_0)) \geq 0$ for all x since $\delta(x) = 1$ if the second term is greater than 0 and 0 otherwise and since δ' is a test function it is always between 0 and 1.
- If we integrate a non-negative integrand we also get a non-negative quantity so we have that

$$0 \leq \int (\delta(x) - \delta'(x))(f(x|\theta_1) - k_\alpha f(x|\theta_0))dx = \beta(\theta_1) - \beta'(\theta_1) - k_\alpha(\beta(\theta_0) - \beta'(\theta_0))$$

by the definition of power functions.

- Now, $k_\alpha(\beta(\theta_0) - \beta'(\theta_0))$ is always non-negative since β is the power function of a size α test and β' of level α test and $k_\alpha > 0$. Therefore, we have

$$0 \leq \beta(\theta_1) - \beta'(\theta_1)$$

and so the test β is UMP.

Notice how much more difficult it would be to prove the above if the hypotheses were composite! We can still do so with the following addendum. First, the NP rejection region is a function of the sufficient statistics by the factorization theorem so we have

$$t \in R_t : \ g(t|\theta_1) > k_\alpha g(t|\theta_0)$$
$$t \notin R_t : \ g(t|\theta_1) < k_\alpha g(t|\theta_0)$$

Now, if we consider composite hypotheses $H_0 : \theta \in \Theta_0$ and $H_1 : \theta \in \Theta_0^c$ and further assume that

- We have a level α test
- There is a $\theta_0 \in \Theta_0$ s.t $P_{\theta_0}(T \in R_t) = \alpha$ (i.e. we achieve equality at an actual parameter value).
- $g(t|\theta)$ is a factor of the pdf f_θ such that for all $\theta' \in \Theta_0^c$ there is a $k' \geq 0$ such that

$$t \in R_t : \ g(t|\theta') > k' g(t|\theta_0)$$
$$t \notin R_t : \ g(t|\theta') < k' g(t|\theta_0$$

then this is a UMP test.

A UMP exists when the likelihood ratio is monotone (monotone $g(t|\theta_1)/g(t|\theta_2)$ for $\theta_2 > \theta_1$) and the test is one-sided. Sometimes you can simply translate a two-sided test to a one-sided one and work things out that way. This added criterion for the LR holds for exponential family subclasses (e.g. when the $w(\theta)$ function is non-decreasing, one-parameter families like poisson or binomial or normal with known variance).

Why doesn't a UMP always exist? Well, the level α class of tests is so large that it is almost impossible to guarantee maximum power *everywhere*. In the estimation problem we restricted ourselves to unbiased estimators in order to find "best" estimators. In testing, there is other restriction you can consider, like invariance to transformation etc (see e.g. Casella and Berger).

1.1 LRT

The likelihood ratio test is a kind of generalization of the NP lemma. We define

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\Theta} L(\theta)} = \frac{L(\hat{\theta}_0)}{L(\hat{\theta})}$$

and find a rejection region R such that

$$R = \{x : \lambda(x) < c_\alpha\} \ \ s.t. \ \ P(\lambda(x) < c_\alpha | \theta \in \Theta_0) \leq \alpha$$

Usually, the LRT is a function of the sufficient statistic $T(x)$ and the cutoff c_α can be derived directly from the distribution of T. If not, we resort to approximation and asymptotic results.

We approximate the likelihood near the unrestricted MLE as follows:

$$l(\theta) \simeq l(\hat{\theta}) + l'(\hat{\theta})(\theta - \hat{\theta}) + l''(\hat{\theta})(\theta - \hat{\theta})^2/2 = l(\hat{\theta}) + l''(\hat{\theta}(\theta - \hat{\theta})^2/2$$

Now, thinking of θ_0 as a fixed value near $\hat{\theta}$ (which would almost be the case if the null is true, true if the null parameter space consisted of one value) then

$$-2 \log \lambda(x) = 2(l(\hat{\theta}) - l(\theta_0)) \simeq -l''(\hat{\theta})(\theta_0 - \hat{\theta})^2$$

This last expression can be rewritten as

$$\frac{-l''(\hat{\theta})}{I(\theta_0)}(\sqrt{I(\theta_0)n}(\hat{\theta}-\theta_0))^2$$

which consists of two terms. The first ratio goes in probability to 1 and the second, by the properties of MLE to a N(0,1) inside the square. We appeal to Slutsky's theorem and get that the LRT (-2 log of) converges in distribution to a χ^2_1.

Now, the proof is not quite complete if $\hat{\theta}_0$ is not a fixed value but a restricted estimate. However, under the null both $\hat{\theta}$ and $\hat{\theta}_0$ are estimates of the truth θ_0 and converge to the same thing (θ_0) and so it all works out. If θ is p-dimensional, we obtain a LRT that is χ^2_p. If the parameter restriction applies to only subset of the parameters, the degrees o freedom of the χ^2 correspond to the difference in number of free parameters in the numerator and denominator.

Another asymptotic test is the Wald test which uses the asymptotic normality of $\hat{\theta}$ directly. We can thus construct a test as follows: Reject θ_0 if the test statistic

$$t = |\frac{\hat{\theta}-\theta_0}{\sqrt{nI(\theta_0)}}| > z_{1-\alpha/2}$$

or a one-sided variant. Which asymptotic approximation is best? Well, the LRT χ^2 approximation is slightly better. You can think of it this way. In the LRT approximation we work directly with the likelihood approximation whereas in the Wald we work with the maximizers.

2 Confidence intervals

The LRT χ^2 approximation can be used to construct confidence intervals for θ. We have

$$-2\log\frac{L(\theta_0|x)}{L(\hat{\theta}|x)} \sim \chi^2_1$$

and so we can look at for which θ and this given sample x would be fail to reject the null:

$$\{\theta : \ -2\log\frac{L(\theta_0|x)}{L(\hat{\theta}|x)} \leq \chi^2_{1-\alpha}\}$$

or

$$\{\theta : \frac{L(\theta_0|x)}{L(\hat{\theta}|x)} \leq e^{-.5\chi^2_{1-\alpha}}$$

i.e., θ's in a neighborhood of the maximizer $\hat{\theta}$

Similarly, the Wald approximation can be used to construct confidence intervals:

$$[\hat{\theta} \pm z_{1-\alpha/2}\sqrt{(nI(\hat{\theta}))^{-1}}]$$

Let's take a step back and discuss confidence interval in more general terms. We want to construct an interval from data such that

$$P_\theta(L \leq \theta \leq U) \geq 1-\alpha$$

for all $\theta \in \Theta$. The end points of the interval, L and U depends on the sample x. We say that this interval *covers* the true θ not that θ falls in the interval as the only thing random above is the sample and therefore the end points of intervals. How can we construct CIs that have this property? There are essentially 3 ways; (1) inverting a test into a confidence interval; (2) using asymptotic properties of the test statistic to do so (see above); and (3) work it out directly from estimators of θ.

Let's look at the inversion of a test. We have a θ_0 and construct a level α test for $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. The test statistic is $\delta_0(\tilde{x})$ with rejection region $R_0 = \{\tilde{x} : \delta_0(\tilde{x}) = 1\}$ and its complement

$R_0^c = \{\delta_0(\tilde{x}) = 0\}$. We now consider values of θ such that the sample $\tilde{x}$ would fall in the corresponding complement of the rejection region, i.e. values of θ that would not be rejected using this test:

$$C(\tilde{x}) = \{\theta : \tilde{x} \in R_\theta^c\} = \{\theta : \delta_\theta(\tilde{x}) = 0\}$$

By construction, since the test δ has level α, we now have for all θ

$$P_\theta(\theta \in C(\tilde{x})) = 1 - \alpha$$

We can likewise start with a confidence interval and turn it into a test. If $C(\tilde{x})$ is a $1-\alpha$ CI, this corresponds to a test that would reject any hypothesis $H_0 : \theta_0$ for $\theta_0 \notin C(\tilde{x})$ at level α. The point is, any level α test applied at value θ that does not lead to a rejection constitute the $1-\alpha$ CI. For the LRT, the inversion looks like this:

$$\delta(\tilde{x}) = 1\{\frac{L(\theta_0)}{L(\hat{\theta(\tilde{x})})} \leq c\}$$

$$C(\tilde{x}) = \{\theta : \frac{L(\theta)}{L(\hat{\theta(\tilde{x})})} \geq c\}$$

We can, as stated above, also start directly with an estimator for θ and from its sampling distribution derive a CI. Usually, we don't work directly with $\hat{\theta}$ since its distribution often depends on θ (the unknown parameter). Instead we work with so-called *pivotal quantities*. A pivotal quantity is a statistic that does not depend on θ in distribution. We denote a pivotal quantity by $Q(\tilde{x}, \theta)$. Note, we can use θ to construct this pivotal quantity. We don't have to actually compute it, only construct its form to derive the CI.

Example:

- $X_i \sim N(\theta, 1)$
- $\bar{X}$ is an estimator of θ
- $\bar{X} - \theta \sim N(0, 1/n)$
- So $C(\tilde{x}, \theta) = \bar{x} - \theta$ is a pivotal quantity
- Now, we want to construct a CI such that

$$P_\theta(a \leq \bar{X} \leq b) \geq 1 - \alpha$$

- Equivalently, we find an interval

$$P_\theta(c \leq \bar{X} - \theta \leq d) \geq 1 - \alpha$$

- But since $\bar{X} - \theta$ is pivotal, this requires only to simply match c and d to the upper and lower $\alpha/2$ quantiles of the $N(0, 1/n)$ distribution

Another example:

- $X_i \sim U[0, \theta]$
- $L(\theta) = \frac{1}{\theta^n} 1\{0 \leq x_{(1)} \leq x_{(n)} \leq \theta\}$
- $\hat{\theta} = x_{(n)}$ is the MLE.
- The pdf of $X_{(n)} = Y$ is $f_\theta(y) = ny^{n-1}/theta^n$
- Change of variables: $Z = Y/\theta$ gives us Z with pdf $f_Z(z) = nz^{n-1}$ on $[0, 1]$ which does not depend on θ
- $Q = X_{(n)}/\theta$ is a pivotal quantity.

- Now, we have

$$1-\alpha = P(c \le Q \le 1) = P(Q \le 1) - P(Q \le c) = 1 - P(Q \le c) = 1 - \int_0^c nz^{n-1}dz = 1 - c^n$$

- We can match c to $\alpha^{1/n}$ and thus the CI is

$$1-\alpha = P(c \le \frac{X_{(n)}}{\theta} \le 1) = P(1/c \ge \frac{\theta}{X_{(n)}} \ge 1) = P(X_{(n)}/c \ge \theta \ge X_{(n)})$$

- CI is $[X_{(n)}, \frac{X_{(n)}}{\alpha^{1/n}}]$

A cool example:

- Non-parametric estimation of the distribution function
- $X_i \sim F$ for unknown F and we want to set up a CI for F.
- CI: $P_F(L(t) \le F(t) \le U(t),\ for\ all\ t) \ge 1-\alpha$
- Now, the MLE of F is $F_n(x) = \frac{1}{n}\sum_i 1\{X_i \le x\}$ the empirical distribution function.
- We construct $K_n = \sup_x |F_n(x) - F(x)|$, the worst case discrepancy between F_n and F.
- Now, we know that $U_i = F(X_i) \sim U[0,1]$ since F is a cdf. (Follows from $F(x) = P(X \le x)$ and $P_F(F(X) \le t) = P_F(X \le F^{-1}(t)) = F(F^{-1}(t)) = t$).
-

$$K_n = \sup_x |F_n(x) - F(x)| = \sup_x |\frac{1}{n}\sum_i 1\{X_i \le x\} - F(x)| = \sup_x |\frac{1}{n}\sum_i 1\{F(X_i) \le F(x)\} - F(x)| =$$

$$= \sup_x |\frac{1}{n} 1\{U_i \le F(x)\} - F(x)| = \sup_{0 \le t \le 1} |\frac{1}{n}\sum_i 1\{U_i \le t\} - t|$$

  which does not depend on F so K_n is pivotal.

- To use this result to set up a cI we need to construct a look-up table for the sup deviation between the sum of uniform random variables less than t and t for different n. However, we don't have to worry about the actual F so this is doable. By simulation we find c such that

$$P(\sup_{0 \le t \le 1} |\frac{1}{n}\sum_i 1\{U_i \le t\} - t| > c) = \alpha$$

  and construct the CI as

$$C = \{F : \sup_x |F_n(x) - F(x)| < c\}$$