Stat210A: Theoretical Statistics **Lecture Date: November 6, 2014**

Hypothesis Testing and Confidence Region Duality; Generalized LR Tests

Lecturer: Michael I. Jordan *Scribe: Arturo Fernandez*

1 Type I/II Error

Consider a test of the form: If the data observed X is in some set C, then accept $H_1 : \theta \in \Omega_1$. Otherwise, accept $H_0 : \theta \in \Omega_0$. Given this rule, C is often called the *rejection region*. Thus, our test would take the form $\varphi(X) = \mathbb{1}\{X \in C\}$. Generally in hypothesis testing, we recognize that we can make two types of mistakes.

- Type I Error: This is the probability that we reject the null hypothesis H_0 (accept H_1) when H_0 is indeed true. We write
$$\text{Type I Error} = \mathbb{P}_\theta(X \in C), \qquad \theta \in \Omega_0$$
Sometimes this is also denoted as $\mathbb{P}(X \in C \,|\, \text{null})$ but this alludes to a Bayesian paradigm that is not consistent with the frequentist framework that has so far been presented.

- Type II Error: This is the probability that we accept the null hypothesis H_0 when it is false (when H_1 is true). That is
$$\text{Type II Error} = \mathbb{P}_\theta(X \notin C), \qquad \theta \in \Omega_1$$
Moreover, type II Error can also be summarized by the power, which measure the probability that we will reject H_0 (accept H_1) when H_1 is true. We often denote this value by β and we see this earlier in Chapter 12 defined more technically as
$$\beta_\varphi(\theta) = \beta_{\mathbb{1}_C}(\theta) = \mathbb{E}_\theta[\mathbb{1}_C(X)] = \mathbb{P}_\theta(X \in C), \qquad \theta \in \Omega_1$$
and with that the connection is clear since
$$\text{Type II Error} = \mathbb{P}_\theta(X \notin C) = 1 - \mathbb{P}_\theta(X \in C) = 1 - \beta_{\mathbb{1}_C}(\theta), \qquad \theta \in \Omega_1$$
In summary, we see that minimizing Type II Error is equivalent to maximizing power.

We know make the observation that Hypothesis Testing is just a special case of Decision Theory.

1.1 Decision Theory

To make our previous statement clear, let's begin by defining our loss function as
$$\mathcal{L}(\,\theta, \delta(x)\,) = |\mathbb{1}\{\theta \in \Omega_1\} - \mathbb{1}\{X \in C\}|$$
In essence, this is the indicator that we make a mistake. But what's the risk? It's easy to see that
$$\mathcal{R}(\theta) = \mathbb{E}_\theta[\,\mathcal{L}(\,\theta, \delta(x)\,)\,] = \begin{cases} \beta_{\mathbb{1}_C}(\theta), & \theta \in \Omega_0 \\ 1 - \beta_{\mathbb{1}_C}(\theta), & \theta \in \Omega_1 \end{cases}$$
This observation motivates the following thought experiment. When we do interval estimation we are talking about the uncertainty that comes with point estimation. When we do Hypothesis Testing, under set criteria, we use the data to make a decision. Is there a connection? It turns out that from one you can get the other.

2 Duality between Testing and Interval Estimation

Let $S(X)$ be a $1-\alpha$ confidence region (CR) for $\xi = \xi(\theta)$. That is

$$\mathbb{P}_\theta(\xi \in S(X)\,) \geq 1-\alpha, \qquad \forall \theta \in \Omega$$

For every ξ_0, let $A(\xi_0)$ be the **Acceptance Region** (AR) for a level α test of $X_0 : \xi(\theta) = \xi_0$ versus $H_1 : \xi(\theta) \neq \xi_0$. So it satisfies

$$\mathbb{P}_\theta\left\{X \in A(\xi(\theta))\,\right\} \geq 1-\alpha, \qquad \forall \theta \in \Omega$$

We will begin with the transition from acceptance regions to confidence regions.

2.1 Acceptance Region to Confidence Region

Given the acceptance region $A(\xi)$, let $S(X) = \{\xi : X \in A(\xi)\,\}$. We can motivate this selection by noting that this set of parameters ξ are those which are consistent with the data. If were making a CR for ξ based on the observed data X, we would take these ξ's.

More formally, note that

$$\xi(\theta) \in S(X) \iff X \in A(\xi(\theta))$$

from which it follows that

$$\mathbb{P}_\theta\Big(\,\xi(\theta) \in S(X)\,\Big) = \mathbb{P}_\theta\Big(\,X \in A(\xi(\theta))\,\Big) \geq 1-\alpha$$

2.2 Confidence Region to Acceptance Region

Let $S(X)$, a $1-\alpha$ confidence region, be given. Define

$$\varphi = \begin{cases} 1, & \xi_0 \notin S(X) \\ 0, & o.w. \end{cases}$$

It follows that if $\xi(\theta) = \xi_0$, then

$$\mathbb{E}_\theta[\varphi] = \mathbb{P}_\theta\Big(\xi_0 \notin S(x)\Big) = 1 - \underbrace{\mathbb{P}_\theta\Big(\xi_0 \in S(x)\Big)}_{\geq 1-\alpha} \leq \alpha$$

With this duality in mind, we might wonder if Uniformly Most Powerful (UMP) are optimal for CR estimation in any sense? That is, all we have assumed are level α tests, but among these tests are there "good" ones for CR estimation.

2.3 UMP Tests and CR Estimation

Let $S(X)$ be derived from a UMP test φ. Let $S^*(X)$ be a competitor corresponding to φ^*, where

$$\varphi = \begin{cases} 1 & \theta_0 \notin S^*(X) \\ 0 & o.w. \end{cases}$$

It follows that φ^* is test of $H_0 : \theta = \theta_0$ versus $H_1 : \theta > \theta_0$ with level α.

By the UMP property,

$$\mathbb{E}_\theta[\varphi] \geq \mathbb{E}_\theta[\varphi^*], \qquad \forall \theta \in \Omega_1 = \{\theta : \theta > \theta_0\}$$

by construction the inequality above is equivalent to

$$\mathbb{P}_\theta\Big(\theta_0 \notin S(X)\Big) \geq \mathbb{P}_\theta\Big(\theta_0 \notin S^*(X)\Big)$$

implying that

$$\mathbb{P}_\theta\Big(\theta_0 \in S(X)\Big) \leq \mathbb{P}_\theta\Big(\theta_0 \in S^*(X)\Big) \tag{1}$$

both for $\theta \in \Omega_1$. Thus we see that if $\theta \in \Omega_1 = \{\theta : \theta > \theta_0\}$, then $S(X)$ has a smaller chance probability of containing the incorrect parameter $\theta_0 \notin \Omega_1$.

And although the optimality noted above is nice, in practice we might actually care about the "length" of a CR as well. To formalize this, let λ be the Lebesgue measure and let $\Omega = (\underset{\sim}{\omega}, \bar{\omega})$ then we have that the *left hand side of length is*

$$\begin{aligned}\mathbb{E}_\theta[\,\lambda\{\,S(X) \cap (\underset{\sim}{\omega}, \theta)\,\}\,] &= \mathbb{E}_\theta\left[\int_{\underset{\sim}{\omega}}^{\theta} \mathbb{1}\{\theta_0 \in S(x)\}\lambda(dP\theta_0)\right] \\ &\overset{Fubini}{=} \int_{\underset{\sim}{\omega}}^{\theta} \mathbb{E}_\theta\left[\mathbb{1}\{\theta_0 \in S(x)\}\right] d\theta_0 = \int_{\underset{\sim}{\omega}}^{\theta} \mathbb{P}_\theta\Big(\theta_0 \in S(x)\Big)\, d\theta_0\end{aligned}$$

By a similar argument we have that

$$\mathbb{E}_\theta[\,\lambda\{\,S^*(X) \cap (\underset{\sim}{\omega}, \theta)\,\}\,] = \int_{\underset{\sim}{\omega}}^{\theta} \mathbb{P}_\theta\Big(\theta_0 \in S^*(x)\Big)\, d\theta_0$$

Thus by (1)

$$\mathbb{E}_\theta[\,\lambda\{\,S(X) \cap (\underset{\sim}{\omega}, \theta)\,\}\,] \leq \mathbb{E}_\theta[\,\lambda\{\,S^*(X) \cap (\underset{\sim}{\omega}, \theta)\,\}\,]$$

Which tells that that hypothesis testing on the right gives you confidence interval control on the left.

We then went over what was coming up: Generalized Likelihood ratio test, Ch14 General Linear Model but not lecturing on it, Bootstrap, End Book material, and then some high-dimensional statistics lectures.

3 Generalized Likelihood Ratio Test

We would like to test Hypotheses in the Composite vs Composite setup. $X_i \overset{iid}{\sim} f_\theta$ with likelihood function $\mathcal{L}(\theta) = \prod_{i=1}^n f_\theta(X_i)$. Moreover let $\Omega = \Omega_0 \cup \Omega_1$. Then the generalized likelihood ratio test statistic for testing $H_0 : \theta \in \Omega_0$ versus $H_1 : \theta \in \Omega_1$ is

$$\lambda = \frac{\sup_{\Omega_1} \mathcal{L}(\theta)}{\sup_{\Omega_0} \mathcal{L}(\theta)}$$

and we reject if λ is greater than some k. Typically Ω_0 is a smaller smooth manifold of lower dimension than Ω, in which was we write

$$\lambda = \frac{\sup_{\Omega} \mathcal{L}(\theta)}{\sup_{\Omega_0} \mathcal{L}(\theta)} \overset{if achieved}{=} \frac{\mathcal{L}(\hat{\theta}_n)}{\mathcal{L}(\tilde{\theta}_n)}$$

We now give an example.

Example 1. Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$ then we can write the log-likelihood as

$$\ell(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{2\sigma^2}$$

and the MLE's as

$$\hat{\mu}_{MLE} = \bar{X} \qquad \text{and} \qquad \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i\bar{X})^2$$

thus the likelihood of the MLE's becomes

$$\text{lik}(\hat{\mu}, \hat{\sigma}^2) = \frac{1}{(2\pi\hat{\sigma}^2)^{n/2}} \exp\left\{-\frac{1}{2}\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\hat{\sigma}^2}\right\} = \left(2\pi n^{-1}(X_i - \bar{X})^2\right)^{-n/2} e^{-n/2}$$

Suppose we would like to test $H_0 : \mu = 0$ vs. $H_1 : \mu \neq 0$. First note that the MLEs in Ω_0 are

$$\tilde{\mu} = 0 \qquad \text{and} \qquad \tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

with associated likelihood

$$\text{lik}(\tilde{\mu}, \tilde{\sigma}^2) = \left(2\pi\tilde{\sigma}^2\right)^{-n/2} e^{-n/2}$$

Thus the generalized likelihood ratio test statistic is just

$$\lambda = \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)^{n/2} = \left(\frac{\sum_{i=1}^{n} X_i^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)^{n/2}$$

Intuitively this makes sense but note that we can also rewrite above as

$$\left(\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + n\bar{X}^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}\right)^{n/2} = \left(1 + \frac{T^2}{n-1}\right)^{n/2}$$

where $T = \bar{X}/(S/\sqrt{n})$. So rejecting for $\lambda > k$ is equivalent to rejecting for $|T| > k_T$.