Stat210A: Theoretical Statistics **Lecture Date: October 30, 2014**

Rest of Stein's Paradox and Hypothesis Testing

Lecturer: Michael I. Jordan *Scribe: Yuansi Chen*

1 Rest of Stein's Paradox

We went through the section 3 (Interpretation and Extension) of Stigler (1990)'s paper on Stein's paradox . The regression perspective not only makes the logic of the estimation clear, it also leads to a short, rigorous proof of the phenomenon. Proof of the superiority of the James-Stein estimator is detailed in this section of the paper.

2 Hypothesis Testing

We talked about some history about the development of hypothesis testing and Jerzy Neyman's life. More information about the history could be found on Wikipedia and also the 10th floor of Evans Hall.

2.1 Definitions of terms

In hypothesis testing, data are used to infer which of two competing hypotheses is correct. It is formulated as a decision problem. The following definitions of terms are used throughout the chapter.

Definition 1 (Null hypothesis (H_0)). A null hypothesis (H_0) and alternative hypothesis (H_1) are written

$$H_0 : \theta \in \Omega_0, H_1 : \theta \in \Omega_1; \text{where } \Omega_0 \cap \Omega_1 = \emptyset, \Omega_0 \cup \Omega_1 = \Omega.$$

Definition 2 (Critical Region (S)). The critical region is the set of values of the test statistics for which we accept H_1.

$$\text{“Accept } H_1\text{” if } x \in S$$
$$\text{“Accept } H_0\text{” if } x \notin S$$

Definition 3 (Power Function). The performance of the test if described by its power function $\beta(.)$, which gives the chance of rejecting H_0 as a function of $\theta \in \Omega$:

$$\beta(\theta) = P_\theta(X \in S).$$

In the mathematical formulation for hypothesis testing just presented, the hypothesis H_0 and H_1 have a symmetric role. But in applications H_0 generally represents the status quo, or what someone would believe about θ without compelling evident to the contrary. In view of this, attention is often focused on tests that have a small chance of error when H_0 is correct. Here comes the Neyman-Pearson setting of hypothesis testing.

Definition 4 (Neyman-Pearson setting of hypothesis testing)**.** Set a significance level α defined as:

$$\alpha = \sup_{\theta \in \Omega_0} \beta(\theta)$$

and then maximize power function over Ω_1

For technical reasons it is convenient to allow external randomization to help the researcher decide between H_0 and H_1. Randomized tests are characterized by a test function φ with range a subset of $[0, 1]$. Given $X = x$, $\varphi(x)$ is the probability of rejecting H_0. The power function β still gives the chance of rejecting H_0 and by smoothing:

$$\beta(\theta) = P_\theta(\text{reject } H_0) = E_\theta[P_\theta(\text{reject } H_0|X)] = E_\theta\varphi(X).$$

Note that a nonrandomized test with critical region S can be viewed as a randomized test with $\varphi = 1_S$.

The set of all test functions is convex, for if φ_1 and φ_2 are test functions and $\gamma \in (0, 1)$, then $\gamma\varphi_1 + (1-\gamma)\varphi_2$ is also a critical function. For randomized tests the level α is defined as

$$\alpha = \sup_{\theta \in \Omega_0} \beta(\theta) = \sup_{\theta \in \Omega_0} E_\theta\varphi(X).$$

3 Simple Versus Simple Testing

A hypothesis is called simple if it completely specifies the distribution of the data, so H_i: $\theta \in \Omega_i$ is simple when Ω_i contains a single parameter value θ_i. When both hypotheses, H_0 and H_1 are simple, the Neyman-Pearson lemma (Proposition 12.2 of Keener (2010)) provides a complete characterization of all reasonable tests.

Suppose H_0 and H_1 are both simple, and let p_0 and p_1 denote densities for X under H_0 and H_1, respectively. Since there are only two distributions for the data X, the power function for a test φ has two values,

$$E_0\varphi(X) = \int \varphi(x)p_0(x)\mu(dx)$$

and

$$E_1\varphi(X) = \int \varphi(x)p_1(x)\mu(dx)$$

Then Neyman-Pearson problem can be formulated as follows:

$$\begin{aligned} \text{maximize} \quad & E_1\varphi \\ \text{s.t.} \quad & E_0\varphi \le \alpha. \end{aligned}$$

The following proposition shows that the solutions of unconstrained optimization problems with Lagrange multiplier (k) also solve the Neyman-Pearson optimization problem in this case.

Lemma 5 (Propostion 12.1. of Keener (2010))**.** *Suppose $k \ge 0$, φ^* maximize*

$$E_1\varphi = kE_0\varphi$$

among all test functions, and $E_0\varphi^* = \alpha$. Then φ^* maximizes $E_1\varphi$ under the constraint.

Proof. Suppose φ has level at most α, $E_0\varphi \leq \alpha$. Then

$$\begin{aligned} E_1\varphi &\leq E_1\varphi - kE_0\varphi + k\alpha \\ &\leq E_1\varphi^* - kE_0\varphi^* + k\alpha \\ &= E_1\varphi^* \end{aligned}$$

Maximizing $E_1\varphi - kE_0\varphi$ is fairly easy because

$$\begin{aligned} E_1\varphi - kE_0\varphi &= \int [p_1(x) - kp_0(x)]\varphi(x)\mu(dx) \\ &= \int_{p_1(x) > kp_0(x)} |p_1(x) - kp_0(x)|\varphi(x)\mu(dx) \\ &\quad - \int_{p_1(x) \leq kp_0(x)} |p_1(x) - kp_0(x)|\varphi(x)\mu(dx) \end{aligned}$$

Then the optimal test φ^* must have

$$\begin{aligned} \varphi^*(x) &= 1, \text{ when } p_1(x) > kp_0(x) \\ &\quad 0, \text{ when } p_1(x) < kp_0(x) \end{aligned}$$

□

When division by zero is not an issue, these tests are based on the likelihood ratio $L(x) = p_1(x)/p_0(x)$. When $L(x) > k$, $\varphi = 1$ and $\varphi = 0$ if $L(x) < k$. When $L(x) = k$, $\varphi(x)$ can take any value in $[0, 1]$. Any test of this form is called a likelihood ratio test.

Lemma 6 (Neyman-Pearson Lemma, Propostion 12.2. of Keener (2010))**.** *Given any level $\alpha \in [0, 1]$, there exists a likelihood ratio test φ_α with level α, and any likelihood ratio test with level α maximize $E_1\varphi$ among all tests with level at most α.*

The fact that likelihood ratio tests maximize $E_1\varphi$ among tests with the same or smaller level follows from the previous lemma. The next result shows that if a test is optimal, it must be a likelihood ratio test.

Lemma 7 (Propostion 12.3. of Keener (2010))**.** *Given $\alpha \in [0, 1]$, let k be the critical value for a likelihood ratio test φ_α and let $B = \{x : p_1(x) \neq kp_0(x)\}$. If φ^* maximizes $E_1\varphi$ among all tests with level at most α, then φ^* and φ_α most agree on B.*

References

Keener, R. (2010). *Theoretical Statistics: Topics for a Core Course*. Springer, New York, NY.

Stigler, S. M. (1990). The 1988 neyman memorial lecture: a galtonian perspective on shrinkage estimators. *Statistical Science*, pages 147–155.