Stat210B: Theoretical Statistics **Lecture Date: April 12, 2007**

Power of the LRT, Bartlett Correction, Efficiency of estimators

Lecturer: Michael I. Jordan *Scribe: Daniel Ting*

1 Asymptotic Power of the LRT (van der Vaart, 1998, Sec 16.4)

Recall the setup from last time. Consider a distribution P_η such that $\{P^n_{\eta+h/\sqrt{n}}\}$ satisfies local asymptotic normality. e.g. the distribution has a density that is qmd. Then the likelihood ratio statistics Λ_n converge in distribution to a statistic in the limit experiment $N(h, I_\eta^{-1})$. Let X be from the limit experiment. So in the limit we may write

$$X \sim N(h, I_\eta^{-1})$$

$$\Lambda_n \xrightarrow{d} \Lambda = ||Z + I_\eta^{1/2} h - I_\eta^{1/2} H_0||_2^2$$

where $Z = I_\eta^{1/2} X$ is a standard multidimensional normal, and the norm is understood to be taken as the infimum over all values in the null hypothesis H_0. So we see that Λ is the squared norm of a standard k dimensional normal projected onto a $k-l$ dimensional subspace where k is the dimension of X and l is the dimension of H_0. Thus Λ_n converges to a χ-squared distribution with non-centrality parameter $\delta = ||I_\eta^{1/2} h - I_\eta^{1/2} H_0||$. Under the null hypothesis, $\delta = 0$ so the asymptotic level α test is to reject when $\Lambda_n > \chi^2_{k-l,\alpha}$. The asymptotic power function under the alternatives is given by

$$\pi_n(\eta + h/\sqrt{n}) = P_{\eta+h/\sqrt{n}}(\Lambda_n > \chi^2_{k-l,\alpha}) \to P_h(\Lambda > \chi^2_{k-l,\alpha}) = \pi(h)$$

where

$$\pi(h) = P(\chi^2_{k-l}(\delta) > \chi^2_{k-l,\alpha})$$

Example 1 (Power and eigenvalues of the Fisher matrix)**.** Consider the simple null hypothesis $H_0 = 0$. Then the non-centrality parameter is simply

$$\delta = \sqrt{h^T I_\eta h}$$

Consider the alternatives in the direction of an eigenvector h_e of I_η i.e. let $h = \mu h_e$. Then

$$\delta = \mu\sqrt{\lambda_e}$$

and

$$\pi(h) = P(\chi^2_{k-l}(\mu\sqrt{\lambda_e}) > \chi^2_{k-l,\alpha}).$$

We see that the power is the greatest on the biggest eigenvalue since the corresponding non-centrality parameter is the largest.

2 Bartlett Correction (van der Vaart, 1998, Sec 16.5)

In the previous section we apply the properties of the asymptotic distribution of the test statistic Λ directly to Λ_n. However, the properties of Λ_n for any particular n are different from Λ, and we may consider "correcting" Λ_n to make it more similar to Λ. In particular, we may consider correcting the mean. If $\Lambda \sim \chi^2_r$ then it has mean r. We may try to "correct" Λ_n by taking

$$\Lambda'_n = \frac{r\Lambda_n}{E_{\theta_0}\Lambda_n}$$

(Note that θ_0 is the η in the previous section.) However, $E_{\theta_n}\Lambda_n$ is generally hard to compute. We may try to replace is with some expansion

$$E_{\theta_0}\Lambda_n = 1 + b(\theta_0)/n + ...$$

Note that this series is typically divergent. However, for a given truncation of the series, it is often accurate for a range of values before diverging. This expansion is related to Edgeworth expansions and saddlepoint approximations.

If we have an estimator $\hat{b}_n$ for $b(\theta_0)$, we obtain the corrected statistic

$$\Lambda'_n = \frac{r\Lambda_n}{1 + \hat{b}_n/n}$$

3 Estimation and Efficiency of Estimation (van der Vaart, 1998, Ch 8)

Outline:

- estimate $\psi(\theta)$ with a sequence of estimators T_n
- derive a Gaussian limit as the "best" within a minimax framework
- First consider an easier problem, asymptotic relative efficiency

3.1 Asymptotic Relative Efficiency (ARE)

Consider an estimator that satisfies

$$\sqrt{n}(T_n - \psi(\theta)) \overset{\theta}{\rightsquigarrow} N(0, \sigma^2(\theta))$$

Let us rescale "time" to get a $N(0,1)$ limit.
Let ν denote time and let n_ν observations be taken at times ν so that

$$\sqrt{\nu}(T_{n_\nu} - \psi(\theta)) \overset{\theta}{\rightsquigarrow} N(0,1)$$

Then we have

$$\sqrt{\frac{\nu}{n_\nu}}\sqrt{n_\nu}(T_{n_\nu} - \psi(\theta)) \overset{\theta}{\rightsquigarrow} N(0,1)$$

so

$$\sqrt{\frac{\nu}{n_\nu}} \to \sigma(\theta)$$

We see that n_ν represents how many samples we need to take in order to achieve a fixed level of accuracy. As with Pitman efficiency of tests, we can compare estimators by taking a ratio of the n_ν's. Define the aymptotic relative efficiency to be

$$ARE = \lim_{\nu \to \infty} \frac{n_{\nu,2}}{n_{\nu,1}} = \frac{\sigma_2(\theta)^2}{\sigma_1(\theta)^2}$$

Example 2 (ARE of median)**.** Consider a location family with density f where f is symmetric about 0 and iid draws from the family.

$$X_i \stackrel{iid}{\sim} f(x-\theta)$$

Then

$$\sqrt{n}(\overline{X}_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

$$\sqrt{n}(\widetilde{X}_n - \theta) \xrightarrow{d} N(0, \frac{1}{4f(0)^2})$$

where $\widetilde{X_n}$ denotes the median. We now consider the ARE under the normal location and Laplace location families.
Under the normal location family with $\sigma^2 = 1$, we have $1/4f(0)^2 = \pi/2$, so the ARE is

$$ARE = \frac{\sigma^2}{1/4f(0)^2} = \pi/2$$

Under the Laplace location family we have $f = \frac{1}{2}e^{|x|}$.

$$\sigma_1^2 = 1/2 \int x^2 e^{-|x|} dx = \int_0^\infty x^2 e^{-x} dx = \Gamma(3) = 2$$

$$\sigma_2^2 = \frac{1}{4f(0)^2} = 1$$

so the $ARE = 1/2$, and the median requires half the number of samples as the mean.

3.2 Hodges' estimator and superefficiency (van der Vaart, 1998, Example 8.1)

Consider

$$X_i \stackrel{iid}{\sim} N(\theta, 1)$$

$$T_n = \overline{X}_n$$

Define Hodges' estimator to be

$$S_n = \begin{cases} T_n & \text{if } |T_n| \geq n^{-1/4} \\ 0 & \text{else} \end{cases}$$

We have $\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0,1)$, but for S_n, we have

1. $r_n S_n \stackrel{\theta}{\rightsquigarrow} 0$ for any sequence $\{r_n\}_n$ if $\theta = 0$

2. $\sqrt{n}(S_n - \theta) \xrightarrow{d} N(0,1)$

In other words, for any $\theta \neq 0$, the asymptotic behavior of S_n is the same as T_n, and for $\theta = 0$, S_n converges arbitrarily fast to the truth.

To show (2), note that

$$P(T_n \in (\theta - M/\sqrt{n}, \theta + M/\sqrt{n})) \to L_\theta(-M, M)$$

where L_θ is the measure for a $N(\theta, 1)$. Note that we may choose M large to make $L_\theta(-M, M)$ arbitrarily close to 1. If $\theta \neq 0$ then the intervals $(\theta - M/\sqrt{n}, \theta + M/\sqrt{n})$ and $(-n^{-1/4}, n^{1/4})$ are eventually disjoint and hence

$$P(T_n = S_n) \to 1$$

To show (1), note that the interval $(\theta - M/\sqrt{n}, \theta + M/\sqrt{n}) \subset (-n^{-1/4}, n^{1/4})$ eventually, so

$$P(S_n = 0) \to 1$$

References

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.