Stat210B: Theoretical Statistics **Lecture Date: 3/8/2007**

Lecture 15: Asymptotic Testing

Lecturer: Michael I. Jordan *Scribe: Dapo Omidiran*

Reading: Chapter 7 of van der Vaart (1998).

1 Asymptotic Testing

Setup: We are given:

- A parametric model: $P_\theta, \theta \in \Theta$
- A null hypothesis: $\theta = \theta_0$
- An alternative hypothesis: $\theta = \theta_1$

Our test then consists of computing the log likelihood ratio:

$$\lambda = \log \left(\frac{\prod_i p_{\theta_1}(X_i)}{\prod_i p_{\theta_0}(X_i)} \right),$$

and accepting the alternative hypothesis if λ is sufficiently large.

Example 1. (Normal location model)

Let $P_\theta = N(\theta, \sigma^2)$, with σ^2 known. After some algebra, we see that

$$\lambda = \frac{n}{\sigma^2}[(\theta_1 - \theta_0)\bar{X}_n - \frac{1}{2}(\theta_1^2 - \theta_0^2)].$$

We can study the distribution under each hypothesis.

Under θ_0, we can use the WLLN to conclude:

$$\lambda \xrightarrow{P} -\frac{n}{\sigma^2}\frac{1}{2}(\theta_1 - \theta_0)^2 \to -\infty$$

Notice that this a good thing. Asymptotically, we will never reject the null hypothesis; our test is "consistent". However, this is also somewhat vacuous, as almost any reasonable test will give the same result.

We should instead look at the rates at which our test converges. One approach is to use large deviations (pioneered by Hoeffding in the '60s?) However, we won't go that route. Instead, we will "shrink" θ towards θ_0 as n increases (e.g., $\theta_1 = \theta_0 + \frac{h}{\sqrt{n}}$.)

In some sense, this $\sqrt{n}$ behavior is the right shrinkage factor for "regular" data, such as iid data.

This approach was first developed for testing, but is applicable to estimation as well.

So, let's study shrinking alternatives:

Example 2. (Normal location model revisited)

$\lambda = h\sqrt{n}\frac{\bar{X_n}-\theta_0}{\sigma^2} - \frac{h^2}{2\sigma^2} = h\bar{Z_n} - \frac{h^2}{2\sigma^2}$ (where $Z_n = \sqrt{n}\frac{\bar{X_n}-\theta_0}{\sigma^2} \overset{H_0}{\sim} N(0, \frac{1}{\sigma^2})$)

Note that this is a quadratic in h. Hence:

$$\lambda \sim N(-\frac{h^2}{2\sigma^2}, \frac{h^2}{\sigma^2})$$

The mean is -$\frac{1}{2}$ the variance!

Is this behavior specific to the Normal distribution? Let's check the exponential family:

Example 3. (Exponential Family)

$$\begin{aligned}
p_\theta(x) &= \exp\left[\theta T(x) - A(\theta)\right] \\
\lambda &= hn^{-\frac{1}{2}}\sum_i T(X_i) - n[A(\theta_0 + hn^{-\frac{1}{2}}) - A(\theta_0)] \\
&= hn^{-\frac{1}{2}}\sum_i T(X_i) - n[A'(\theta_0)hn^{-\frac{1}{2}} + \frac{1}{2}A''(\theta_0)h^2n^{-1} + o(n^{-1})] \\
&= hZ_n - \frac{1}{2}h^2A''(\theta_0) + o(1)
\end{aligned}$$

Where $Z_n = n^{\frac{1}{2}}\sum_i T(X_i) - E_{\theta_0}[T(X_i)]$ (As $A'(\theta_0) = E_{\theta_0}[T(X_i)]$).

Asymptotically, the mean is again -1/2 the variance.

How much further can we go?

The key property is quadratic mean differentiability (QMD), essentially a notion of smoothness relevant for asymptotic statistics.

In particular, we want a smoothness condition. However, we are constrained by the following:

- We want to avoid assuming that derivatives exist for all x (i.e., for each x, a derivative exists at each value of θ)
- We also want to avoid explicit conditions on higher derivatives.

Solution: We will work with square roots of densities. Classical (Frechet) differentiability of $\sqrt{p_\theta}$ (Again, note that x is held fixed, and θ is the variable):

$$\sqrt{p_{\theta_0+h}} - \sqrt{p_{\theta_0}} - h^T\eta_{\theta_0}(x) = o(||h||).$$

To weaken this somewhat stringent condition, we only ask that it hold in the quadratic mean:

Definition 4. QMD

P_θ is QMD at θ_0 if

$$\int \left(\sqrt{p_{\theta_0+h}} - \sqrt{p_{\theta_0}} - \frac{1}{2}h^T\dot{\ell}_{\theta_0}\sqrt{p_{\theta_0}}\right)^2 d\mu = o(||h^2||)$$

for some function $\dot{\ell}_{\theta_0}$.

Keep in mind that these (p_{θ_0}, p_θ, etc) are all functions of x. $\dot{\ell}_{\theta_0}$ is *not* the derivative of some ℓ_{θ_0}, but is instead some function.

Why do we define things in this weird way? If classical derivatives *do* exist:

$$\frac{\partial}{\partial\theta}\sqrt{p_\theta} = \frac{1}{2}\sqrt{p_\theta}\frac{\partial}{\partial\theta}\log p_\theta.$$

So we associate $\dot{\ell}_\theta$ is the score function, in this case.

Theorem 5 (Theorem 7.2, van der Vaart (1998) p94)**.** *If Θ is an open subset of $\mathcal{R}^K$ and $P_\theta,\ \theta \in \Theta$ is QMD. Then:*

- $P_\theta \dot{\ell}_\theta = 0$ *(Like score functions),*
- *and* $I_\theta = P_\theta \dot{\ell}_\theta \dot{\ell}_\theta^T$ *exists (Fisher information),*
- *and* $\lambda = \dfrac{\prod_i p_{\theta+\frac{h_n}{\sqrt{n}}}(X_i)}{\prod_i p_\theta(X_i)} = \frac{1}{\sqrt{n}}\sum_i h^T \dot{\ell}_\theta(X_i) - \frac{1}{2}h^T I_\theta h + o_{p_\theta}(1).$

Where $h_n \to h \neq 0$. Note that this implies that:

$$\lambda \xrightarrow{d} N(-\frac{1}{2}h^T I_\theta h, h^T I_\theta h)\ .$$

Proof. (Partial Proof) Let

$$\begin{aligned} p_n &= p_\theta + \frac{h_n}{\sqrt{n}}\ , \\ p &= p_\theta\ , \\ g &= h^T \dot{\ell}_\theta\ . \end{aligned}$$

By the definition of QMD, it follows that:

$$\begin{aligned} &\int (\sqrt{p_n} - \sqrt{p} - \frac{1}{2}g\sqrt{p})^2 d\mu = o(n^{-1})\ , \\ \implies\quad & n^{1/2}(\sqrt{p_n} - \sqrt{p}) \xrightarrow{QM} \frac{1}{2}g\sqrt{p}\ , \\ \implies\quad & \sqrt{p_n} \xrightarrow{QM} \sqrt{p}\ . \end{aligned}$$

We recall that $\int f_n g_n \to \int fg$ if $f_n \to$ f and $g_n \to g$.

By continuity of the inner product:

$$Pg = \int gp d\mu = \int \frac{1}{2}g\sqrt{p}2\sqrt{p}d\mu = \lim_{n\to\infty}\int \sqrt{n}(\sqrt{p_n} - \sqrt{p})(\sqrt{p_n} + \sqrt{p})d\mu = 0\ .$$

Define:

$$W_{n,i} = 2\left(\frac{\sqrt{p_n(X_i)}}{\sqrt{p(X_i)}} - 1\right)$$

We use the fact that $\log(1+x) = x - \frac{1}{2}x^2 + x^2 R(x)$ (where R(x) $\to 0$ as x$\to 0$) to conclude that:

$$\log \prod_i p_n(X_i)/p(X_i) = 2\sum_i \log(1+\frac{1}{2}W_{n,i}) = \sum_i W_{n,i} - \frac{1}{4}\sum_i W_{n,i}^2 + \frac{1}{2}\sum_i W_i^2 R(W_{n,i}) \ .$$

As:

$$E_p\left(\sum_i W_{n,i}\right) = 2n\left(\int \sqrt{p_n}\sqrt{p}d\mu - 1\right) = -n\int (\sqrt{p_n}-\sqrt{p})^2 d\mu \to -\int \frac{1}{4}g^2 p d\mu \ ,$$

where $Pg^2 = \int \frac{1}{4}g^2 p d\mu = h^T(\int \dot{\ell}_\theta \dot{\ell}_\theta^T dP)h = h^T I_\theta h$.

Look at the remainder of the proof in van der Vaart (1998). □

References

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