Theoretical Statistics. Lecture 22.

Peter Bartlett

1. Recall: Asymptotic testing.
2. Quadratic mean differentiability.
3. Local asymptotic normality. [vdv7]

Recall: Asymptotic testing

Consider the asymptotics of a test. We have

- A parametric model P_θ for $\theta \in \Theta$.
- A null hypothesis $\theta = \theta_0$.
- An alternative hypothesis $\theta = \theta_0 + h_n$.

Test: compute the log likelihood ratio,

$$\lambda = \log \prod_{i=1}^{n} \frac{dP_{\theta_0+h_n}}{dP_{\theta_0}}(X_i),$$

and reject the null hypothesis if it is sufficiently large.

Recall: Asymptotic testing

For example, suppose $P_\theta = N(\theta, \sigma^2)$. Then we saw that

$$\lambda = \frac{nh_n}{\sigma^2}(\bar{X} - \theta_0) - \frac{nh_n^2}{2\sigma^2}$$
$$\overset{\theta_0}{\sim} N\left(-\frac{nh_n^2}{2\sigma^2}, \frac{nh_n^2}{\sigma^2}\right).$$

For $\sqrt{n}h_n \to h \neq 0$, the normal parameters approach $(-h^2/(2\sigma^2), h^2/\sigma^2)$.

Recall: Asymptotic testing

Another example. The exponential family with sufficient statistic T: $p_\theta(x) = \exp\left(T(x)\theta - A(\theta)\right)$. We have

$$\begin{aligned}\lambda &= \log \prod_{i=1}^n \frac{dP_{\theta_0+h_n}}{dP_{\theta_0}}(X_i)\\ &= h_n \sum_{i=1}^n \left(T(X_i) - P_{\theta_0} T(X_i)\right) - \frac{n}{2} A''(\theta_0) h_n^2 + o(nh_n^2)\\ &\overset{\theta_0}{\rightsquigarrow} N\left(-\frac{h^2 \operatorname{var}_{\theta_0}(T(X_1))}{2}, h^2 \operatorname*{var}_{\theta_0}(T(X_1))\right),\end{aligned}$$

for $h_n = h/\sqrt{n}$.

Local asymptotic normality: Taylor series

Suppose that we have a density p_θ wrt some measure, and the log likelihood, $\ell_\theta(x) = \log p_\theta(x)$ is twice differentiable wrt θ, and can be approximated by its second order Taylor series,

$$\ell_{\theta+h}(x) = \ell_\theta(x) + h^T \dot{\ell}_\theta(x) + \frac{1}{2} h^T \ddot{\ell}_\theta(x) h + o(\|h\|^2).$$

Then

$$\begin{aligned}
\lambda &= \log \prod_{i=1}^{n} \frac{dP_{\theta+h_n}}{dP_\theta}(X_i) \\
&= \sum_{i=1}^{n} \left(\log p_{\theta+h_n}(X_i) - \log p_\theta(X_i)\right) \\
&= h_n^T \sum_{i=1}^{n} \dot{\ell}_\theta(X_i) + \frac{1}{2} h_n^T \sum_{i=1}^{n} \ddot{\ell}_\theta(X_i) h_n + o(n\|h_n\|^2).
\end{aligned}$$

Score functions

Consider the log likelihood function $\ell_\theta(x) = \log p_\theta(x)$. Its derivative $\dot{\ell}_\theta$ is called the score function. For $X \sim P_\theta$ (and for ℓ_θ satisfying regularity conditions), we have

1. The score function has mean zero: $P_\theta \dot{\ell}_\theta = 0$,
2. The mean curvature of the log likelihood is the negative Fisher information: $P_\theta \ddot{\ell}_\theta = -I_\theta$, where $I_\theta = P_\theta \dot{\ell}_\theta \dot{\ell}_\theta^T$.

Score functions: Proof

Notice that $\int p_\theta(x)\,d\mu(x) = 1$ implies

$$\int \dot p_\theta(x)\,d\mu(x) = 0, \qquad \int \ddot p_\theta(x)\,d\mu(x) = 0.$$

But

$$P_\theta \dot\ell_\theta = \int \dot\ell_\theta\,dp_\theta = \int \frac{\dot p_\theta}{p_\theta} p_\theta\,d\mu = \int \dot p_\theta\,d\mu = 0$$

and

$$P_\theta \ddot\ell_\theta = \int \ddot\ell_\theta p_\theta\,d\mu = \int \left(\frac{\ddot p_\theta}{p_\theta} - \frac{\dot p_\theta \dot p_\theta^T}{p_\theta^2}\right) p_\theta\,d\mu = -\int \dot\ell_\theta \dot\ell_\theta^T p_\theta\,d\mu = -I_\theta.$$

Local asymptotic normality: Taylor series

Thus,

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} \dot{\ell}_\theta(X_i) \overset{P_\theta}{\rightsquigarrow} N(0, I_\theta),$$

$$\frac{1}{n} \sum_{i=1}^{n} \ddot{\ell}_\theta(X_i) \overset{P_\theta}{\to} -I_\theta.$$

So if $\sqrt{n} h_n \to h$,

$$\lambda = h_n^T \sum_{i=1}^{n} \dot{\ell}_\theta(X_i) + \frac{1}{2} h_n^T \sum_{i=1}^{n} \ddot{\ell}_\theta(X_i) h_n + o(n\|h_n\|^2)$$

$$\overset{P_\theta}{\rightsquigarrow} N\left(-\frac{1}{2} h^T I_\theta h, h^T I_\theta h\right).$$

This behavior is known as **local asymptotic normality**.

Quadratic mean differentiability

What conditions make this argument rigorous? A weaker condition than twice differentiability suffices: $\theta \mapsto \sqrt{p}_\theta$ differentiable for most x.

Definition: The root density $\theta \mapsto \sqrt{p_\theta}$ (for $\theta \in \mathbb{R}^k$) is **differentiable in quadratic mean** at θ if there exists a vector-valued measurable function $\dot{\ell}_\theta : \mathcal{X} \to \mathbb{R}^k$ such that, for $h \to 0$,

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

Quadratic mean differentiability

Why the strange notation? If $\theta \mapsto p_\theta$ is differentiable, then

$$\nabla_\theta \sqrt{p_\theta} = \frac{1}{2}\frac{\nabla_\theta p_\theta}{\sqrt{p_\theta}} = \frac{1}{2}\sqrt{p_\theta}\frac{\nabla_\theta p_\theta}{p_\theta} = \frac{1}{2}\sqrt{p_\theta}\nabla_\theta \ell_\theta = \frac{1}{2}\sqrt{p_\theta}\dot{\ell}_\theta.$$

Notice that we do not need differentiability at every x. Rather, the $L_2(\mu)$ (average—under μ—squared) error should be small.

QMD and local asymptotic normality

Theorem: If Θ is an open subset of $\mathbb{R}^k$, and P_θ is QMD at $\theta \in \Theta$, then

1. $P_\theta \dot{\ell}_\theta = 0$.
2. $I_\theta = P_\theta \dot{\ell}_\theta \ell_\theta^T$ exists.
3. For every h_n satisfying $\sqrt{n} h_n \to h$,

$$\log \prod_{i=1}^n \frac{p_{\theta+h_n}}{p_\theta}(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^n h^T \dot{\ell}_\theta(X_i) - \frac{1}{2} h^T I_\theta h + o_{P_\theta}(1)$$

$$\overset{\theta}{\rightsquigarrow} N\left(-\frac{1}{2} h^T I_\theta h, h^T I_\theta h\right).$$

QMD of $\sqrt{p_\theta}$ is elegant: $\int (\sqrt{p})^2 \, d\mu = 1$; we can use inner prods in $L_2(\mu)$.

QMD sufficient conditions

Theorem: If

1. Θ is an open subset of $\mathbb{R}^k$.
2. $\theta \mapsto \sqrt{p_\theta(x)}$ is continuously differentiable at μ-almost all x.
3. $I_\theta = \int \dot{p}_\theta \dot{p}_\theta^T / p_\theta \, d\mu$ is continuous in θ.

Then $\sqrt{p_\theta}$ is QMD at θ, with $\dot{\ell}_\theta = \dot{p}_\theta / p_\theta$.

QMD Examples

- Exponential families are QMD. (See earlier example).
- Location families.

$$p_\theta(x) = f(x - \theta),$$

where f is positive, continuously differentiable, with

$$I_\theta = \int \left(\frac{f'(x)}{f(x)} \right)^2 f(x)\, dx < \infty,$$

are QMD. (Note that, because we can shift x by θ, I_θ does not depend on θ.)

QMD Examples

- Laplace location model is QMD:

$$p_\theta(x) = \frac{1}{2}\exp\left(-|x-\theta|\right).$$

  Notice that $\sqrt{p_\theta}$ is not differentiable. But it is QMD (because the single point of non-differentiability, θ, has measure zero).

- Uniform distribution p_θ on $[0,\theta]$ is not QMD. Indeed, QMD requires

$$\begin{aligned} o(\|h\|^2) &= \int \left(\sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2}h^T\dot{\ell}_\theta\sqrt{p_\theta}\right)^2 d\mu \\ &\geq \int_\theta^{\theta+h} \left(\sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2}h^T\dot{\ell}_\theta\sqrt{p_\theta}\right)^2 d\mu \\ &= \frac{h}{\theta+h}, \qquad \text{which is a contradiction.} \end{aligned}$$

Recall: Contiguity

Theorem: For
$$\log \frac{dQ_n}{dP_n} \overset{P_n}{\rightsquigarrow} N(\mu, \sigma^2),$$
$Q_n \triangleleft P_n$ iff $\mu = -\sigma^2/2$. (Also, $P_n \triangleleft Q_n$ for any μ, σ^2.)

But for QMD families, if h_n satisfies $\sqrt{n} h_n \to h$,

$$\log \prod_{i=1}^{n} \frac{p_{\theta+h_n}}{p_\theta}(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h^T \dot{\ell}_\theta(X_i) - \frac{1}{2} h^T I_\theta h + o_{P_\theta}(1)$$
$$\overset{\theta}{\rightsquigarrow} N\left(-\frac{1}{2} h^T I_\theta h, h^T I_\theta h\right).$$

So $P^n_{\theta+h_n} \triangleleft \triangleright P^n_\theta$.

Recall: Contiguity and change of measure

Lemma: [Le Cam's Third Lemma] Suppose, for $X_n \in \mathbb{R}^k$,

$$\left(X_n, \log \frac{dQ_n}{dP_n}\right) \overset{P_n}{\rightsquigarrow} N\left(\begin{pmatrix} \mu \\ -\frac{\sigma^2}{2} \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^T & \sigma^2 \end{pmatrix}\right).$$

Then $X_n \overset{Q_n}{\rightsquigarrow} N(\mu + \tau, \Sigma)$.

Asymptotically linear statistics

Suppose the model $\{P_\theta : \theta \in \Theta\}$ is QMD, and a statistic T_n satisfies

$$\sqrt{n}\,(T_n - \mu_\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\theta(X_i) + o_{P_\theta}(1),$$

where $P_\theta \psi_\theta = 0$ and $P_\theta \psi_\theta \psi_\theta^T = \Sigma$. Then for h_n satisfying $\sqrt{n} h_n \to h$, the sequence of log likelihood ratios satisfies

$$\log \frac{dP^n_{\theta+h_n}}{dP^n_\theta}(X_1, \ldots, X_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h^T \dot{\ell}_\theta(X_i) - \frac{1}{2} h^T I_\theta h + o_{P_\theta}(1).$$

Asymptotically linear statistics

Thus, the central limit theorem implies

$$\left(\sqrt{n}\,(T_n - \mu_\theta)\,, \log \frac{dP^n_{\theta+h_n}}{dP^n_\theta}\right) \overset{\theta}{\rightsquigarrow} N\left(\begin{pmatrix} 0 \\ -\frac{1}{2}h^T I_\theta h \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^T & h^T I_\theta h \end{pmatrix}\right),$$

where $\tau = P_\theta \psi_\theta h^T \dot{\ell}_\theta$.

Then $\sqrt{n}(T_n - \mu_\theta) \overset{\theta+h_n}{\rightsquigarrow} N\left(P_\theta \psi_\theta h^T \dot{\ell}_\theta, \Sigma\right)$.