Theoretical Statistics. Lecture 22.
Peter Bartlett

1. Recall: Asymptotic testing.
2. Quadratic mean differentiability.
3. Local asymptotic normality. [vdv7]
Consider the asymptotics of a test. We have

- A parametric model $P_{\theta}$ 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

$$
\lambda = \frac{nh_n}{\sigma^2}(\bar{X} - \theta_0) - \frac{nh_n^2}{2\sigma^2}
$$

$$
\theta_0 \sim N \left( -\frac{nh_n^2}{2\sigma^2}, \frac{nh_n^2}{\sigma^2} \right).
$$

For $\sqrt{nh_n} \to h \neq 0$, the normal parameters approach $(-h^2/(2\sigma^2), h^2/\sigma^2)$. 
Recall: Asymptotic testing

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

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

$$= h_n \sum_{i=1}^{n} (T(X_i) - P_{\theta_0}T(X_i)) - \frac{n}{2} A''(\theta_0) h_n^2 + o(nh_n^2)$$

$$\theta_0 \xrightarrow{} N \left(-\frac{h^2 \text{var}_{\theta_0}(T(X_1))}{2}, h^2 \text{var}(T(X_1)) \right),$$

for $h_n = h/\sqrt{n}$. 
Local asymptotic normality: Taylor series

Suppose that we have a density $p_\theta$ wrt some measure, and the log likelihood, $\ell_\theta(x) = \log p_\theta(x)$ is twice differentiable wrt $\theta$, 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

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

$$= \sum_{i=1}^{n} (\log p_{\theta+h_n}(X_i) - \log p_{\theta}(X_i))$$

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

\[
\int \dot{p}_\theta(x) \, d\mu(x) = 0, \quad \int \ddot{p}_\theta(x) \, d\mu(x) = 0.
\]

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

\[
P_\theta \ddot{\ell}_\theta = \int \ddot{\ell}_\theta p_\theta \, d\mu = \int \left( \frac{\ddot{p}_\theta}{p_\theta} - \frac{\dot{p}_\theta \dot{p}_\theta^T}{p_\theta^2} \right) p_\theta \, d\mu = - \int \ell_\theta \dot{\ell}_\theta^T p_\theta \, d\mu = -I_\theta.
\]
Local asymptotic normality: Taylor series

Thus,

$$\frac{1}{n^{1/2}} \sum_{i=1}^{n} \ell_\theta(X_i) \xrightarrow{P_\theta} N(0, I_\theta),$$

$$\frac{1}{n} \sum_{i=1}^{n} \ell_\theta(X_i) \xrightarrow{P_\theta} -I_\theta.$$

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

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

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

This behavior is known as local asymptotic normality.
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 $\theta$ 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 $\mu$—squared) error should be small.
**QMD and local asymptotic normality**

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

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

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

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

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

11
QMD sufficient conditions

**Theorem:** If

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

Then $\sqrt{p_\theta}$ is QMD at $\theta$, with $\ell_\theta = \dot{p}_\theta / p_\theta$. 
• 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 \( \theta \), \( I_\theta \) does not depend on \( \theta \).)
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, \( \theta \), has measure zero).

- Uniform distribution \( p_\theta \) on \([0, \theta]\) is not QMD. Indeed, QMD requires

\[
o(h^2) = \int \left( \sqrt{p_\theta + h} - \sqrt{p_\theta} - \frac{1}{2} h^T \ell_\theta \sqrt{p_\theta} \right)^2 d\mu \\
\geq \int_0^{\theta+h} \left( \sqrt{p_\theta + h} - \sqrt{p_\theta} - \frac{1}{2} h^T \ell_\theta \sqrt{p_\theta} \right)^2 d\mu \\
= \frac{h}{\theta + h}, \quad \text{which is a contradiction.}
\]
Recall: Contiguity

**Theorem:** For
\[
\log \frac{dQ_n}{dP_n} \overset{P_n}{\sim} N(\mu, \sigma^2),
\]
\(Q_n \prec P_n\) iff \(\mu = -\sigma^2/2\). (Also, \(P_n \prec Q_n\) for any \(\mu, \sigma^2\).)

But for QMD families, if \(h_n\) satisfies \(\sqrt{n}h_n \rightarrow h\),
\[
\log \prod_{i=1}^{n} \frac{p_{\theta+h_n}}{p_{\theta}}(X_i) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h^T \ell_\theta(X_i) - \frac{1}{2} h^T I_\theta h + o_{P_\theta}(1)
\]
\(\overset{\theta}{\sim} N \left(-\frac{1}{2} h^T I_\theta h, h^T I_\theta h \right)\).

So \(P_{\theta+h_n} \prec\succ P_{\theta}^n\).
Recall: Contiguity and change of measure

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

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

Then $X_n \overset{Q_n}{\rightsquigarrow} N(\mu + \tau, \Sigma)$. 
Suppose the model \( \{ P_\theta : \theta \in \Theta \} \) is QMD, and a statistic \( T_n \) satisfies

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

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

\[
\log \left( \frac{dP_{\theta+h_n}^n}{dP_\theta^n} \right)(X_1, \ldots, X_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h^T \dot{\ell}_\theta(X_i) - \frac{1}{2} h^T I_\theta h + o_{P_\theta}(1).
\]
Asymptotically linear statistics

Thus, the central limit theorem implies

$$\left( \sqrt{n} (T_n - \mu_\theta), \log \frac{dP_{\theta+h_n}^n}{dP_{\theta}^n} \right) \xrightarrow{\theta} N \left( \left( \begin{array}{c} 0 \\ -\frac{1}{2} h^T I_\theta h \end{array} \right), \left( \begin{array}{cc} \Sigma & \tau \\ \tau^T & h^T I_\theta h \end{array} \right) \right),$$

where $\tau = P_\theta \psi_\theta h^T \ell_\theta$.

Then $\sqrt{n}(T_n - \mu_\theta) \xrightarrow{\theta+h_n} N \left( P_\theta \psi_\theta h^T \ell_\theta, \Sigma \right)$. 