Theoretical Statistics. Lecture 23.

Peter Bartlett

1. Recall: QMD and local asymptotic normality. [vdv7]
2. Convergence of experiments, maximum likelihood.
3. Relative efficiency of tests. [vdv14]
Local asymptotic normality

We’ve seen that, for a QMD model $P_\theta$, the log likelihood ratio,

$$\log \frac{dP^n_{\theta_0+h/\sqrt{n}}}{dP^n_{\theta_0}}(X_i),$$

is asymptotically normal. This is useful for:

1. Comparing null $\theta_0$ and shrinking alternative $\theta_0 + h/\sqrt{n}$ with a likelihood ratio test.

2. Understanding the local behavior of a statistic $T_n$.
   If we assume that $\theta$ is fixed, and we understand $T_n$’s asymptotics under $P_\theta$, we can use the asymptotics of the log likelihood ratio to understand the asymptotics of $T_n$ in a local neighborhood of $\theta$. The appropriate local scale is typically $1/\sqrt{n}$. 
Recall: 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. $P_\theta \dot{\ell}_\theta = 0$.
2. $I_\theta = P_\theta \dot{\ell}_\theta \ell_\theta^T$ exists.
3. For every $h_n$ satisfying $\sqrt{nh_n} \to h$,

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

$$
\Rightarrow \theta \rightsquigarrow N \left(- \frac{1}{2} h^T I_\theta h, h^T I_\theta h \right).
$$
Recall: Quadratic mean differentiability

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 $\ell_\theta : \mathcal{X} \rightarrow \mathbb{R}^k$ such that, for $h \rightarrow 0$,

$$\int \left( \sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h^T \ell_\theta \sqrt{p_\theta} \right)^2 d\mu = o(\|h\|^2).$$
Recall: 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 \( \sqrt{n}h_n \to h \),

\[
\left\langle \sqrt{n} (T_n - \mu_\theta), \log \frac{dP_{\theta+h_n}^n}{dP_\theta^n} \right\rangle \xrightarrow{\theta} N \left( \begin{pmatrix} 0 \\ -\frac{1}{2} h^T I_\theta h \end{pmatrix}, \begin{pmatrix} \Sigma & \tau \\ \tau^T & h^T I_\theta h \end{pmatrix} \right),
\]

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

So \( \sqrt{n}(T_n - \mu_\theta)^{\theta+h_n} \xrightarrow{\text{law}} N \left( P_\theta \psi_\theta h^T \dot{\ell}_\theta, \Sigma \right) \).
Asymptotically linear statistics

That is, we know that under $\theta$,

$$\sqrt{n} \left( T_n - \mu_\theta \right) \overset{\theta}{\sim} N(0, \Sigma).$$

And we can use the asymptotics of the log likelihood ratio to determine the asymptotics of this statistic under the shrinking alternative $\theta + h/\sqrt{n}$:

$$\sqrt{n}(T_n - \mu_\theta)^{\theta + h/\sqrt{n}} \overset{\sim}{\sim} N \left( P_\theta \psi_\theta h^T \ell_\theta, \Sigma \right).$$
Asymptotically linear statistics: Example

Location families:
Suppose that

\[ p_\theta(x) = f(x - \theta), \]

where \( f \) is positive, continuously differentiable, and satisfies

\[ \mu = \int x f(x) \, dx = 0, \]
\[ \sigma^2 = \int x^2 f(x) \, dx < \infty, \]
\[ I_\theta = \int \left( \frac{f'(x)}{f(x)} \right)^2 f(x) \, dx < \infty. \]

This family is QMD.
Asymptotically linear statistics: Example

1. Consider the \textit{t-statistic} for the null hypothesis \( \theta = 0 \),

\[
T_n = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{S_n}
\]

\[
\sqrt{n}T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i}{\sigma} + o_{P_0}(1).
\]

Thus, \( T_n \) is an asymptotically linear statistic, with

\[
\psi_\theta(x) = \frac{x}{\sigma},
\]

\[
\ell_\theta(x) = -\frac{f'(x - \theta)}{f(x - \theta)}.
\]
Asymptotically linear statistics: Example

Hence, for $h_n$ satisfying $\sqrt{nh_n} \to h$,

$$\sqrt{n}T_n \overset{h_n}{\sim} N \left( P_0 \psi_0 h \ell_0, P_0 \psi_0^2 \right),$$

$$P_0 \psi_0 h \ell_0 = -P_0 \frac{X f'(X)}{\sigma f(X)} h = -\frac{h}{\sigma} \int x f'(x) \, dx = \frac{h}{\sigma} \int f(x) \, dx = \frac{h}{\sigma}.$$

$$P_0 \psi_0^2 = \frac{1}{\sigma^2} P_0 X^2 = 1.$$

$$\sqrt{n}T_n \overset{h_n}{\sim} N \left( \frac{h}{\sigma}, 1 \right).$$
2. Suppose that $P_0(X > 0) = 1/2$ and consider the sign statistic for the null hypothesis $\theta = 0$,

$$s_n = \frac{1}{n} \sum_{i=1}^{n} \left( 1[X_i > 0] - \frac{1}{2} \right).$$

Thus, $s_n$ is an asymptotically linear statistic, with

$$\psi_0(x) = 1[x > 0] - P_0(X > 0),$$
$$\ell_0(x) = -\frac{f'(x - \theta)}{f(x - \theta)}.$$

Hence, for $h_n$ satisfying $\sqrt{n}h_n \to h$,

$$\sqrt{n}s_n \stackrel{h_n}{\sim} N \left( P_0\psi_0h\ell_0, P_0\psi_0^2 \right).$$
Asymptotically linear statistics: Example

\[ P_0 \psi_0 \dot{\ell}_0 = -P_0 \left( 1[X > 0] - \frac{1}{2} \right) \frac{f'(X)}{f(X)} h \]

\[ = -h \int \left( 1[x > 0] - \frac{1}{2} \right) f'(x) \, dx \]

\[ = \frac{h}{2} \left( \int_{-\infty}^{0} f'(x) \, dx - \int_{0}^{\infty} f'(x) \, dx \right) = h f(0). \]

\[ P_0 \psi_0^2 = \frac{1}{4}. \]

\[ \sqrt{n} s_n \xrightarrow{\text{h.p.}} N \left( h f(0), \frac{1}{4} \right). \]
**Theorem:** If \( \left( P_{\theta} : \theta \in \Theta \subseteq \mathbb{R}^k \right) \) is QMD at \( \theta \) with nonsingular Fisher information \( I_{\theta} \), \( T_n \) are statistics in the local experiments \( \left( P_{\theta + h/\sqrt{n}} : h \in \mathbb{R}^k \right) \), and for every \( h \) there is a law \( L_h \) s.t. \( T_n \overset{h}{\rightsquigarrow} L_h \). Then there is a randomized statistic \( T \) in the experiment \( \left( N \left( h, I_{\theta}^{-1} \right) : h \in \mathbb{R}^k \right) \) such that for each \( h \), \( T_n \overset{h}{\rightsquigarrow} T \).

The proof uses the Le Cam lemmas (change of measure via the asymptotically normal log-likelihood ratio)
Convergence of local statistical experiments

For the local statistical experiment,

\[
P^n_{\theta + h/\sqrt{n}} : h \in \mathbb{R}^k
\]

think of \(\theta\) as a particular parameter value, and \(\theta + h/\sqrt{n}\) as a nearby value. We are interested in the asymptotic behavior of statistics when the parameter is near the value \(\theta\).

Motivation:

- If \(T_n\) defines a test, then the power \(P_h(T_n > c)\) depends on the law of \(T_n\), so we can study its asymptotics via statistics in a normal experiment.

- If \(T_n\) is an estimator, then we can study the asymptotics of the expected squared error \(E_h(T_n - h)^2\) via statistics in a normal experiment.
Consider the maximum likelihood estimator $T_n = \hat{h}_n$ for the local experiment 

$$\left( P_{\theta + h/\sqrt{n}} : h \in \mathbb{R}^k \right).$$

(Notice that $\hat{h}_n = \sqrt{n}(\hat{\theta}_n - \theta)$.) Typically, the matching asymptotic statistic in the limit experiment is the maximum likelihood estimator $T = X \sim N\left(h, I_{\theta}^{-1}\right)$. So we expect the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ to be $N\left(0, I_{\theta}^{-1}\right)$ under $\theta$.

Note that the previous theorem does not imply that this particular statistic in the limit experiment (the maximum likelihood estimator) is the weak limit of the $T_n$. This needs some additional conditions.
Theorem: Suppose

1. $(P_\theta : \theta \in \Theta)$ is QMD at $\theta$ with nonsingular Fisher information $I_\theta$,
2. for every $x$, $\theta \mapsto \log p_\theta(x)$ is Lipschitz, and
3. the maximum likelihood estimator $\hat{\theta}_n$ is consistent.

Then

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\theta} N(0, I_\theta^{-1}).$$
Relative efficiency of tests

**Example:** Suppose $X_1, \ldots, X_n \sim P_\theta$, where

1. $P_\theta$ has density $f(x - \theta)$ on $\mathbb{R}$,
2. $f$ is symmetric about zero (so the mean=median of $P_\theta$ is $\theta$),
3. $f$ has a unique median ($f(0) \neq 0$),
4. $f$ has a finite variance.

We wish to test $H_0 : \theta = 0$ versus $H_1 : \theta > 0$. 
Relative efficiency of tests

Example: Candidate tests:

1. Sign test: \( S_n = \frac{1}{n} \sum_{i=1}^{n} 1[X_i > 0] \).

2. t-test: \( T_n = \frac{1}{n} \sum_{i=1}^{n} \frac{X_i}{S_n} \).

Which is better?
Relative efficiency of tests: sign test

\[ S_n = \frac{1}{n} \sum_{i=1}^{n} 1[X_i > 0]. \]

\[ \frac{\sqrt{n}}{\sigma(\theta)} (S_n - \mu(\theta)) \rightsquigarrow N(0, 1), \]

where \[ \mu(\theta) = 1 - F(-\theta), \]

\[ \sigma^2(\theta) = (1 - F(-\theta))F(-\theta). \]

Thus, \[ 2\sqrt{n} \left( S_n - \frac{1}{2} \right) \rightsquigarrow N(0, 1). \]

Reject \( H_0 \) if \[ 2\sqrt{n}(S_n - 1/2) > z_\alpha. \]
Relative efficiency of tests: sign test

**Definition:** The power function of a test that rejects the null hypothesis when the statistic $T_n$ falls in the critical region $K_n$ is

$$\pi_n(\theta) = P_\theta(T_n \in K_n).$$

For the sign test,

$$\pi_n(\theta) = P_\theta \left( \sqrt{n} (S_n - \mu(0)) > \sigma(0) z_{\alpha_n} \right)$$

$$= P_\theta \left( \frac{\sqrt{n}}{\sigma(\theta)} (S_n - \mu(\theta)) > \frac{\sigma(0) z_{\alpha_n} + \sqrt{n} (\mu(0) - \mu(\theta))}{\sigma(\theta)} \right)$$

$$= 1 - \Phi \left( \frac{\sigma(0) z_{\alpha_n} + \sqrt{n} (\mu(0) - \mu(\theta))}{\sigma(\theta)} \right) + o(1).$$
Relative efficiency of tests: sign test

For $\theta = 0$, we have $\pi_n(0) = 1 - \Phi(z_{\alpha_n}) = \alpha_n$.

For $\theta > 0$, $\mu(0) - \mu(\theta) = F(-\theta) - F(0) < 0$.

Provided $\alpha_n \to 0$ sufficiently slowly,

$$
\pi_n(\theta) = 1 - \Phi \left( \frac{\sigma(0) z_{\alpha_n} + \sqrt{n} (\mu(0) - \mu(\theta))}{\sigma(\theta)} \right) + o(1)
$$

$$
\to \begin{cases} 
0 & \text{if } \theta = 0, \\
1 & \text{if } \theta > 0.
\end{cases}
$$

So the limiting power function is perfect.

This is typical: any reasonable test can distinguish a fixed alternative, given unlimited data.
Relative efficiency of tests

So how do we compare tests? We need to make the problem of discriminating between the null and the alternative more difficult as $n$ increases. It is natural to consider a shrinking alternative, that converges to the null.

Recall our example:
We wish to test $H_0 : \theta = 0$ versus $H_1 : \theta_n > 0$, with $\theta_n \to 0$. 
Relative efficiency of tests

For the sign test,

\[ \pi_n(\theta_n) = 1 - \Phi \left( \frac{\sigma(0)z_\alpha + \sqrt{n}(\mu(0) - \mu(\theta_n))}{\sigma(\theta_n)} \right) + o(1). \]

The level of the test converges:

\[ \pi_n(0) = 1 - \Phi (z_\alpha) + o(1) \to \alpha. \]

What about the power?

It depends on the asymptotics of \( \sqrt{n}(\mu(0) - \mu(\theta_n)) \). Since \( F \) is differentiable at 0,

\[ \sqrt{n}(\mu(0) - \mu(\theta_n)) = \sqrt{n}(F(-\theta_n) - F(0)) = -\sqrt{n}\theta_nf(0) + o(\sqrt{n}\theta_n). \]
Relative efficiency of tests

If $\theta_n \to \theta$ faster than $1/\sqrt{n}$, $\sqrt{n}(\mu(0) - \mu(\theta_n)) \to 0$, so $\pi_n(\theta_n) \to \alpha$. The test fails: these alternatives are too hard.

For $\theta_n \to \theta$ slower than $1/\sqrt{n}$, $\sqrt{n}(\mu(0) - \mu(\theta_n)) \to -\infty$, so $\pi_n(\theta_n) \to 1$. These slowly shrinking alternatives are too easy.

Consider an intermediate rate:

$\sqrt{n}\theta_n \to h$. 