Peter Bartlett

1. Relative efficiency of tests [vdv14]: Rescaling rates.
2. Likelihood ratio tests [vdv15].
Recall: Relative efficiency of tests

Theorem: Suppose that (1) $T_n$, $\mu$, and $\sigma$ are such that, for all $h$ and $\theta_n = \theta_0 + h/\sqrt{n}$,

$$\frac{\sqrt{n}(T_n - \mu(\theta_n))}{\sigma(\theta_n)} \xrightarrow{\theta_n} N(0, 1),$$

(2) $\mu$ is differentiable at 0, (3) $\sigma$ is continuous at 0.
Then a test that rejects $H_0 : \theta = \theta_0$ for large values of $T_n$ and is asymptotically of level $\alpha$ satisfies, for all $h$,

$$\pi_n(\theta_n) \rightarrow 1 - \Phi\left(z_\alpha - h\frac{\mu'(\theta_0)}{\sigma(\theta_0)}\right).$$

So the slope $\mu'(\theta_0)/\sigma(\theta_0)$ determines the asymptotic power.
Rescaling rates

So far, we’ve considered alternatives of the form

\[ \theta_n = \theta_0 + \frac{h}{\sqrt{n}}. \]

This corresponds to choosing a sequence \( \theta_n \) such that the difference, \( \theta_n - \theta_0 \), when appropriately rescaled, approaches a constant:

\[ \sqrt{n}(\theta_n - \theta_0) \to h. \]

This rescaling rate is appropriate for regular cases. But other rates are possible.
**Rescaling rates: \( L_1 \)-distance**

**Definition:** The \( L_1 \)-distance [not total variation] between two distributions \( P \) and \( Q \) with densities \( p = dP/d\mu \) and \( q = dQ/d\mu \) is

\[
\|P - Q\| = \int |p - q| \, d\mu.
\]

**Lemma:** For a sequence of models \( P_{n,\theta} \) with null hypothesis \( H_0 : \theta = \theta_0 \) and alternatives \( H_1 : \theta = \theta_n \), the power function of any test satisfies

\[
\pi_n(\theta_n) - \pi_n(\theta_0) \leq \frac{1}{2} \|P_{n,\theta_n} - P_{n,\theta_0}\|.
\]

Furthermore, there is a test for which equality holds.
Consequences:

1. If $\|P_{n,\theta_n} - P_{n,\theta_0}\| \to 2$: Some sequence of tests is perfect, that is, $\pi_n(\theta_n) \to 1$ and $\pi_n(\theta_0) \to 0$.

2. If $\|P_{n,\theta_n} - P_{n,\theta_0}\| \to 0$: Any sequence of tests is worthless, because $\pi_n(\theta_n) - \pi_n(\theta_0) \to 0$.

3. If $\|P_{n,\theta_n} - P_{n,\theta_0}\|$ is bounded away from 0 and 2: There is no perfect sequence of tests, but not all tests are worthless.

This result reveals the appropriate rescaling rate: we need $\theta_n$ to approach $\theta_0$ at a rate than ensures an intermediate value of $\|P_{n,\theta_n} - P_{n,\theta_0}\|$.
**Rescaling rates: $L_1$-distance**

**Proof:** First, for any densities $p$ and $q$,

$$0 = \int (p - q) \, d\mu$$

$$= \int_{p > q} (p - q) \, d\mu + \int_{p < q} (p - q) \, d\mu$$

$$= \int_{p > q} |p - q| \, d\mu - \int_{p < q} |p - q| \, d\mu,$$

so [notice relationship with total variation distance]

$$\int |p - q| \, d\mu = \int_{p > q} |p - q| \, d\mu + \int_{p < q} |p - q| \, d\mu$$

$$= 2 \int_{p > q} |p - q| \, d\mu.$$
Rescaling rates: $L_1$-distance

So we have

$$\pi_n(\theta_n) - \pi_n(\theta_0) = \int 1[T_n \in K_n](p_{n,\theta_n} - p_{n,\theta_0}) \, d\mu_n$$

$$\leq \int 1[p_{n,\theta_n} > p_{n,\theta_0}](p_{n,\theta_n} - p_{n,\theta_0}) \, d\mu_n$$

$$= \int 1[p_{n,\theta_n} > p_{n,\theta_0}]|p_{n,\theta_n} - p_{n,\theta_0}| \, d\mu_n$$

$$= \frac{1}{2} \|P_{n,\theta_n} - P_{n,\theta_0}\|,$$

where the upper bound is achieved by the test

$$1[T_n \in K_n] = 1[p_{n,\theta_n} > p_{n,\theta_0}].$$
Rescaling rates: Hellinger distance

It’s convenient to relate the $L_1$-distance to Hellinger distance (because then product measures are easy to deal with).

Definition: The Hellinger distance between $P$ and $Q$ (which have densities $p$ and $q$) is

$$h(P, Q) = \left( \frac{1}{2} \int \left( p^{1/2} - q^{1/2} \right)^2 \, d\mu \right)^{1/2}.$$  

(The $1/2$ ensures $0 \leq h(P, Q) \leq 1$. It is defined without it in vdV.)
Theorem:

\[ nh^2(P_{\theta_n}, P_{\theta_0}) \to \infty \quad \Rightarrow \quad \| P_{\theta_n}^n - P_{\theta_0}^n \| \to 2, \]
\[ nh^2(P_{\theta_n}, P_{\theta_0}) \to 0 \quad \Rightarrow \quad \| P_{\theta_n}^n - P_{\theta_0}^n \| \to 0, \]
\[ h^2(P_{\theta_n}, P_{\theta_0}) = \Theta \left( \frac{1}{n} \right) \quad \Rightarrow \quad \| P_{\theta_n}^n - P_{\theta_0}^n \| \not\to \{0, 2\}. \]
Rescaling rates: Hellinger distance

Proof:
Useful properties:

\[ 2h^2(P, Q) \leq \|P - Q\| \leq 2\sqrt{2}h(P, Q). \]

Also, \( A(P^n, Q^n) = A^n(P, Q) \),

Where

\[ A(P, Q) = 1 - h^2(p, q) = \int p^{1/2}q^{1/2} \, d\mu \]

is the Hellinger affinity.
Proof (continued):

\[ nh^2(P_{\theta_n}, P_{\theta_0}) \to \infty \]

\[ \Rightarrow \quad A(P_{\theta_n}, P_{\theta_0}) = 1 - \omega \left( \frac{1}{n} \right) \]

\[ \Rightarrow \quad A(P_{\theta_n}^n, P_{\theta_0}^n) \to 0 \]

\[ \Rightarrow \quad h^2(P_{\theta_n}^n, P_{\theta_0}^n) \to 1 \]

\[ \Rightarrow \quad \| P_{\theta_n}^n - P_{\theta_0}^n \| \to 2. \]
Rescaling rates: Hellinger distance

Proof (continued):

\[ nh^2(P_{\theta_n}, P_{\theta_0}) \to 0 \]

\[ \Rightarrow \quad A(P_{\theta_n}, P_{\theta_0}) = 1 - o\left(\frac{1}{n}\right) \]

\[ \Rightarrow \quad A(P_{\theta_n}^n, P_{\theta_0}^n) \to 1 \]

\[ \Rightarrow \quad h^2(P_{\theta_n}^n, P_{\theta_0}^n) \to 0 \]

\[ \Rightarrow \quad \|P_{\theta_n}^n - P_{\theta_0}^n\| \to 0. \]
Rescaling rates: Hellinger distance

Thus, if \( h^2(P_\theta, P_{\theta_0}) = \Theta(|\theta - \theta_0|^{\alpha}) \), then the critical quantity is the limit of

\[
 nh^2(P_{\theta_n}, P_{\theta_0}) = \Theta \left( \left( n^{1/\alpha} |\theta_n - \theta_0| \right)^{\alpha} \right).
\]

If \( P_\theta \) is QMD at \( \theta_0 \), then

\[
 h^2(P_\theta, P_{\theta_0}) = \Theta(|\theta - \theta_0|^2),
\]

that is, \( \alpha = 2 \), so we consider a shrinking alternative with

\[
 \sqrt{n}(\theta_n - \theta_0) \to h.
\]
**Rescaling rates: Hellinger distance**

**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 $h \to 0$,

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

**Theorem:** If $P_\theta$ is QMD at $\theta$ and $I_\theta = P_\theta \dot{\ell}_\theta \dot{\ell}_\theta^T$ exists, then

$$
h^2(P_{\theta+h}, P_\theta) = \frac{1}{8} h^T I_\theta h + o(\|h\|^2).
$$
Rescaling rates: Hellinger distance

Proof:

\[ 2h^2(P_{\theta+h}, P_\theta) = \int (\sqrt{p_{\theta+h}} - \sqrt{p_\theta})^2 \, d\mu \]

\[ = \|\sqrt{p_{\theta+h}} - \sqrt{p_\theta}\|_{L_2(\mu)}^2. \]

But QMD implies

\[ \left\| \sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h^T \ell_\theta \sqrt{p_\theta} \right\|_{L_2(\mu)}^2 = o(\|h\|^2), \]

and

\[ \left\| \frac{1}{2} h^T \ell_\theta \sqrt{p_\theta} \right\|_{L_2(\mu)}^2 = \frac{1}{4} h^T P_\theta \left( \ell_\theta \ell_\theta^T \right) h \]

\[ = \frac{1}{4} h^T I_\theta h = O(\|h\|^2). \]
Rescaling rates: Hellinger distance

So

\[ 2h^2(P_{\theta+h}, P_\theta) = \left\| \sqrt{p_{\theta+h}} - \sqrt{p_\theta} \right\|_{L_2(\mu)}^2 \]

\[ = \left\| \frac{1}{2} h^T \ell_\theta \sqrt{p_\theta} + \left( \sqrt{p_{\theta+h}} - \sqrt{p_\theta} - \frac{1}{2} h^T \ell_\theta \sqrt{p_\theta} \right) \right\|_{L_2(\mu)}^2 \]

\[ = \frac{1}{4} h^T I_\theta h + o (\| h \|^2) + \left( o(\| h \|^2) O(\| h \|^2) \right)^{1/2} \quad \text{(Cauchy-Schwarz)} \]

\[ = \frac{1}{4} h^T I_\theta h + o (\| h \|^2) . \]
Consider $P_{\theta}$ uniform on $[0, \theta]$. Recall that this model is not QMD. A straightforward calculation shows that

$$h^2(P_{\theta}, P_{\theta_0}) = \frac{|\theta - \theta_0|}{\theta \lor \theta_0}.$$

So the appropriate shrinking alternative has $n(\theta_n - \theta_0) \to h$. 

**Rescaling rates: Hellinger distance**
Suppose we observe $X_1, \ldots, X_n$, with density $p_\theta$, $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$.

For $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$, the optimal test statistic is

$$
\log \prod_{i=1}^{n} \frac{p_{\theta_1}(X_i)}{p_{\theta_0}(X_i)}.
$$

If we have composite hypotheses, we could instead use

$$
\tilde{\Lambda}_n = \log \frac{\sup_{\theta \in \Theta_1} \prod_{i=1}^{n} p_\theta(X_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^{n} p_\theta(X_i)}.
$$
Likelihood ratio tests

Notice that, for a minimal sufficient statistic $T$, we can write

$$\tilde{\Lambda}_n = \log \frac{\sup_{\theta \in \Theta_1} \prod_{i=1}^{n} h(X_i) f_\theta(T(X_i))}{\sup_{\theta \in \Theta_0} \prod_{i=1}^{n} h(X_i) f_\theta(T(X_i))}$$

$$= \log \frac{\sup_{\theta \in \Theta_1} \prod_{i=1}^{n} f_\theta(T(X_i))}{\sup_{\theta \in \Theta_0} \prod_{i=1}^{n} f_\theta(T(X_i))},$$

so $\tilde{\Lambda}_n$ depends only on the minimal sufficient statistic.

Since the critical value will be positive, it will not change the test if we replace this statistic by $\tilde{\Lambda}_n \vee 0$. We will also scale it by a factor of 2. (We’ll see that this gives a neater test.)
Define

\[ \Lambda_n = 2(\tilde{\Lambda}_n \lor 0) \]

\[ = 2 \log \frac{(\sup_{\theta \in \Theta_1} \prod_{i=1}^n p_\theta(X_i)) \lor (\sup_{\theta \in \Theta_0} \prod_{i=1}^n p_\theta(X_i))}{\sup_{\theta \in \Theta_0} \prod_{i=1}^n p_\theta(X_i)} \]

\[ = 2 \log \frac{\sup_{\theta \in \Theta_0 \cup \Theta_1} \prod_{i=1}^n p_\theta(X_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^n p_\theta(X_i)} \]

\[ = 2 \sum_{i=1}^n \left( \ell_{\hat{\theta}_n}(X_i) - \ell_{\hat{\theta}_{n,0}}(X_i) \right), \]

where \( \hat{\theta}_n \) is the maximum likelihood estimator for \( \theta \) over \( \Theta = \Theta_0 \cup \Theta_1 \), and \( \hat{\theta}_{n,0} \) is the maximum likelihood estimator over \( \Theta_0 \).
**Likelihood ratio tests**

We’ll focus on cases where $\Theta = \Theta_0 \cup \Theta_1$ is a subset of $\mathbb{R}^k$, and where $\Theta$ and $\Theta_0$ are locally linear spaces. Then under $H_0$, we’ll see that $\Lambda_n$ is asymptotically chi-square distributed with $m$ degrees of freedom, where $m = \dim(\Theta) - \dim(\Theta_0)$. So we can get a test that is asymptotically of level $\alpha$ by comparing $\Lambda_n$ to the upper $\alpha$-quantile of a chi-square distribution.