Theoretical Statistics. Lecture 26.

Peter Bartlett

1. Likelihood ratio tests [vdv15].
   (a) Taylor series.
   (b) $\Lambda_n \stackrel{\theta \in \Theta_0}{\rightsquigarrow} \chi^2_{k-l}$.
   (c) Asymptotic power function.

Recall: Likelihood ratio tests

Suppose we observe $X_1, \ldots, X_n$, with density p_θ,
$H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$.
NB: composite hypotheses.

Define

$$\Lambda_n = 2 \log \frac{\sup_{\theta \in \Theta_0 \cup \Theta_1} \prod_{i=1}^n p_\theta(X_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^n p_\theta(X_i)}$$
$$= 2 \sum_{i=1}^n \left(\ell_{\hat{\theta}_n}(X_i) - \ell_{\hat{\theta}_{n,0}}(X_i) \right),$$

where $\hat{\theta}_n$ is the maximum likelihood estimator for θ over $\Theta = \Theta_0 \cup \Theta_1$,
and $\hat{\theta}_{n,0}$ is the maximum likelihood estimator over Θ_0.

Likelihood ratio tests

Notice that, for a sufficient statistic T, $p_\theta(x)$ depends on x only through $T(x)$:

$$p_\theta(x) = h(x) f_\theta(T(x)),$$

so

$$\begin{aligned}\Lambda_n &= 2 \log \frac{\sup_{\theta\in\Theta_1} \prod_{i=1}^n h(X_i) f_\theta(T(X_i))}{\sup_{\theta\in\Theta_0} \prod_{i=1}^n h(X_i) f_\theta(T(X_i))} \\ &= 2 \log \frac{\sup_{\theta\in\Theta_1} \prod_{i=1}^n f_\theta(T(X_i))}{\sup_{\theta\in\Theta_0} \prod_{i=1}^n f_\theta(T(X_i))},\end{aligned}$$

hence Λ_n depends only on a **minimal** sufficient statistic.

Likelihood ratio tests

We'll focus on cases where $\Theta = \Theta_0 \cup \Theta_1$ is a subset of $\mathbb{R}^k$, and where Θ and Θ_0 are locally linear spaces. Then under H_0, we'll see that Λ_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 α by comparing Λ_n to the upper α-quantile of a chi-square distribution.

Likelihood ratio tests: Taylor series

Under P_θ, where $\theta \in \Theta_0$ is in the interior of Θ,

$$\begin{aligned}
\Lambda_n &= 2\sum_{i=1}^n \left(\ell_{\hat\theta_n}(X_i) - \ell_{\hat\theta_{n,0}}(X_i)\right) \\
&= -2\sum_{i=1}^n \left(\ell_{\hat\theta_{n,0}}(X_i) - \ell_{\hat\theta_n}(X_i)\right) \\
&= -2\left(\hat\theta_{n,0} - \hat\theta_n\right)^T \underbrace{\sum_{i=1}^n \dot\ell_{\hat\theta_n}(X_i)} - \left(\hat\theta_{n,0} - \hat\theta_n\right)^T \sum_{i=1}^n \ddot\ell_{\tilde\theta_n}(X_i)\left(\hat\theta_{n,0} - \hat\theta_n\right),
\end{aligned}$$

where $\tilde\theta_n$ is between $\hat\theta_n$ and $\hat\theta_{n,0}$, and we have assumed that, for all x, $\theta \mapsto \ell_\theta(x)$ is twice continuously differentiable.

Likelihood ratio tests: Taylor series

$$\Lambda_n = -\sqrt{n}\left(\hat{\theta}_{n,0} - \hat{\theta}_n\right)^T \frac{1}{n}\sum_{i=1}^{n} \ddot{\ell}_{\tilde{\theta}_n}(X_i)\sqrt{n}\left(\hat{\theta}_{n,0} - \hat{\theta}_n\right) + o_{P_\theta}(1),$$

because $\hat{\theta}_n$ maximizes $P_n \ell_\theta$, and asymptotically this is in the interior of Θ, so $P_n \dot{\ell}_{\hat{\theta}_n} = 0$.

$$\Lambda_n = \sqrt{n}\left(\hat{\theta}_{n,0} - \hat{\theta}_n\right)^T I_\theta \sqrt{n}\left(\hat{\theta}_{n,0} - \hat{\theta}_n\right) + o_{P_\theta}(1),$$

where we have assumed that the sequence $\sqrt{n}\left(\hat{\theta}_{n,0} - \hat{\theta}_n\right)$ is uniformly tight, and that

$$\frac{1}{n}\sum_{i=1}^{n} \ddot{\ell}_{\tilde{\theta}_n}(X_i) = -I_\theta + o_{P_\theta}(1).$$

Likelihood ratio tests: Taylor series

(Here, $\theta \in \Theta_0$, that is, under the null, so we have $\tilde{\theta}_n \xrightarrow{P} \theta$.)

Thus,

$$\Lambda_n = \sqrt{n}\left(\hat{\theta}_{n,0} - \hat{\theta}_n\right)^T I_\theta \sqrt{n}\left(\hat{\theta}_{n,0} - \hat{\theta}_n\right) + o_{P_\theta}(1)$$

is a quadratic form defining a squared distance between $\hat{\theta}_{n,0}$ and $\hat{\theta}_n$.

Likelihood ratio tests: Simple null

Suppose $\Theta_0 = \{\theta_0\}$ and $\theta = \theta_0$.

$$\Lambda_n = \sqrt{n}\left(\hat{\theta}_n - \theta_0\right)^T I_\theta \sqrt{n}\left(\hat{\theta}_n - \theta_0\right) + o_{P_\theta}(1).$$

Under general conditions (we saw them for maximum likelihood estimators),

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \rightsquigarrow X,$$

where $X \sim N\left(0, I_\theta^{-1}\right)$, so

$$\Lambda_n \rightsquigarrow X^T I_\theta X = Z^T I_\theta^{-1/2} I_\theta I_\theta^{-1/2} Z = Z^T Z,$$

where $Z = I_\theta^{1/2} X \sim N(0, I_k)$. Thus, $\Lambda_n \rightsquigarrow \chi_k^2$.

Recall: Maximum likelihood

Theorem: Suppose

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

Then

$$\sqrt{n}(\hat\theta_n - \theta) \overset{\theta}{\rightsquigarrow} N(0, I_\theta^{-1}).$$

Likelihood ratio tests: Composite null

What if Θ_0 is a linear subspace (of dimension more than 0)?

We might expect $\sqrt{n}(\hat{\theta}_{n,0} - \theta, \hat{\theta}_n - \theta)$ to converge jointly to a normal vector (X_0, X), in which case

$$\Lambda_n \rightsquigarrow (X - X_0)^T I_\theta (X - X_0).$$

We'll see that this has a χ^2_{k-l} distribution, where $k = \dim(\Theta)$ and $l = \dim(\Theta_0)$.

Likelihood ratio tests: Composite null

Write Λ_n in terms of local likelihood ratios, for the true θ in Θ_0:

$$\begin{aligned}\Lambda_n &= 2\log\frac{\sup_{\theta\in\Theta}\prod_{i=1}^n p_\theta(X_i)}{\sup_{\theta\in\Theta_0}\prod_{i=1}^n p_\theta(X_i)}\\ &= 2\sup_{h\in H_n}\log\frac{\prod_{i=1}^n p_{\theta+h/\sqrt{n}}(X_i)}{\prod_{i=1}^n p_\theta(X_i)}\\ &\qquad - 2\sup_{h\in H_{n,0}}\log\frac{\prod_{i=1}^n p_{\theta+h/\sqrt{n}}(X_i)}{\prod_{i=1}^n p_\theta(X_i)},\end{aligned}$$

where

$$H_n = \sqrt{n}\,(\Theta-\theta)\,,$$

$$H_{n,0} = \sqrt{n}\,(\Theta_0-\theta)$$

are the **local parameter spaces**.

Likelihood ratio tests: Composite null

Theorem: Suppose (1) $(P_\theta : \theta \in \Theta)$ is QMD at $\theta \in \Theta_0$ with I_θ nonsingular, (2) for a function $\dot\ell$ with $P_\theta \dot\ell^2 < \infty$, for every θ_1, θ_2 in a neighborhood of θ,

$$|\log p_{\theta_1}(x) - \log p_{\theta_2}(x)| \le \dot\ell(x) \left\|\theta_1 - \theta_2\right\|,$$

(3) the estimators $\hat\theta_{n,0}$ and $\hat\theta_n$ are consistent under P_θ, and (4) the sets $H_{n,0}$ and H_n converge to sets H_0 and H. Then

$$\Lambda_n \overset{\theta+h/\sqrt{n}}{\rightsquigarrow} \left\| I_\theta^{1/2} X - I_\theta^{1/2} H_0 \right\|^2 - \left\| I_\theta^{1/2} X - I_\theta^{1/2} H \right\|^2.$$

where $X \sim N(h, I_\theta^{-1})$.

Likelihood ratio tests: Composite null

Here, we say that a sequence H_n of sets converges to a set H if

$$H = \left\{ \lim_{i\to\infty} h_{n_i} : h_{n_i} \text{ convergent}, h_n \in H_n \right\}.$$

Also, we write

$$\left\| I_\theta^{1/2} X - I_\theta^{1/2} H_0 \right\|^2 = \inf_{h \in H_0} \left\| I_\theta^{1/2} X - I_\theta^{1/2} h \right\|^2.$$

Likelihood ratio tests: Composite null

Idea of Proof:

Λ_n is the difference of two rescaled maximum likelihood ratio processes,

$$2 \sup_{h \in H_n} \log \frac{\prod_{i=1}^n p_{\theta+h/\sqrt{n}}(X_i)}{\prod_{i=1}^n p_\theta(X_i)} - 2 \sup_{h \in H_{n,0}} \log \frac{\prod_{i=1}^n p_{\theta+h/\sqrt{n}}(X_i)}{\prod_{i=1}^n p_\theta(X_i)}.$$

Just as we saw for maximum likelihood, this statistic for the local experiment converges to the corresponding asymptotic statistic in the normal experiment,

$$2 \sup_{h \in H} \log \frac{dN(h, I_\theta^{-1})}{dN(0, I_\theta^{-1})}(X) - 2 \sup_{h \in H_0} \log \frac{dN(h, I_\theta^{-1})}{dN(0, I_\theta^{-1})}(X).$$

where $X \sim N(0, I_\theta^{-1})$. (And under $\theta + g/\sqrt{n}$, Λ_n converges in distribution to the same thing, with $X \sim N(g, I_\theta^{-1})$.)

Likelihood ratio tests: Composite null

But this is

$$2\sup_{h\in H}\log\frac{dN(h,I_\theta^{-1})}{dN(0,I_\theta^{-1})}(X)-2\sup_{h\in H_0}\log\frac{dN(h,I_\theta^{-1})}{dN(0,I_\theta^{-1})}(X)$$

$$=\sup_{h\in H}-(X-h)^TI_\theta(X-h)-\sup_{h\in H_0}-(X-h)^TI_\theta(X-h)$$

$$=\inf_{h\in H_0}(X-h)^TI_\theta(X-h)-\inf_{h\in H}(X-h)^TI_\theta(X-h)$$

$$=\inf_{h\in H_0}\left\|I_\theta^{1/2}(X-h)\right\|^2-\inf_{h\in H}\left\|I_\theta^{1/2}(X-h)\right\|^2$$

$$=\left\|I_\theta^{1/2}X-I_\theta^{1/2}H_0\right\|^2-\left\|I_\theta^{1/2}X-I_\theta^{1/2}H\right\|^2.$$

Likelihood ratio tests: Composite null

Theorem: If $\theta \in \Theta_0$ is an interior point of Θ, then H_n converges to $H = \mathbb{R}^k$, and

$$\left\| I_\theta^{1/2} X - I_\theta^{1/2} H \right\|^2 = 0.$$

If, in addition, H_0 is a linear subspace of dimension l, then

$$\Lambda_n \overset{\theta}{\rightsquigarrow} \left\| I_\theta^{1/2} X - I_\theta^{1/2} H_0 \right\|^2 \sim \chi^2_{k-l},$$

where $X \sim N(0, I_\theta^{-1})$.

Likelihood ratio tests: Composite null

Proof:

Write $Z = I_\theta^{1/2} X \sim N(0, I_k)$. Write $Z = (Z_1, \ldots, Z_k)$ in a basis where the first l basis vectors lie in H_0. (And notice that, in this basis, it is still a standard normal.) Then the squared distance from Z to H_0 is

$$\|Z - H_0\|^2 = \sum_{i=l+1}^{k} Z_i^2 \sim \chi^2_{k-l}.$$

Likelihood ratio tests: Examples

Example:

Suppose $\theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+$ and

$$p_\theta = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right),$$

where f is a fixed density on $\mathbb{R}$.

Consider $H_0 : \mu = 0$ versus $H_1 : \mu \neq 0$. Fix $\theta = (0, \sigma)$.

$$\Theta_0 = \{0\} \times (0, \infty),$$
$$H_{n,0} = \sqrt{n}(\Theta_0 - \theta) = \{0\} \times (-\sqrt{n}\sigma, \infty) \to \{0\} \times \mathbb{R} = H_0.$$

So $\dim(H_0) = 1$, $\dim(H) = 2$. For suitably regular f, the likelihood ratio statistic is asymptotically χ^2_1. [PICTURE]

Likelihood ratio tests: Examples

Consider $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$. Fix $\theta = (\mu, \sigma)$ with $\mu < 0$.

$$\Theta_0 = (-\infty, 0] \times (0, \infty),$$
$$H_{n,0} = \sqrt{n}(\Theta_0 - \theta) \to \mathbb{R} \times \mathbb{R} = H_0.$$

So $\dim(H_0) = 2 = \dim(H)$. For suitably regular f, the likelihood ratio statistic is asymptotically 0.

Likelihood ratio tests: Examples

Consider $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$. Fix $\theta = (0, \sigma)$.

$$\Theta_0 = (-\infty, 0] \times (0, \infty),$$
$$H_{n,0} = \sqrt{n}(\Theta_0 - \theta) \to (-\infty, 0] \times \mathbb{R} = H_0.$$

The weak limit is

$$\left\| Z - I_\theta^{1/2} H_0 \right\|^2 .$$

Notice that $I_\theta^{1/2} H_0$ is a half-space. [PICTURE] So this asymptotic distribution is the distribution of $(Z_1 \vee 0)^2$, where $Z_1 \sim N(0, 1)$. Because $\Pr((Z_1 \vee 0)^2 > c) = (1/2) \Pr(Z_1^2 > c)$, we can set the critical value as the upper 2α-quantile of a χ_1^2 variable.

Likelihood ratio tests: Asymptotic power function

If $\theta \in \Theta_0$ is an interior point of Θ, we have seen that

$$\Lambda_n \overset{\theta+h/\sqrt{n}}{\rightsquigarrow} \left\| I_\theta^{1/2} X - I_\theta^{1/2} H_0 \right\|^2,$$

where $X \sim N(h, I_\theta^{-1})$.

If H_0 is a linear subspace of dimension l, then under the null ($h = 0$), this is χ^2_{k-l}.

Likelihood ratio tests: Asymptotic power function

Setting the critical value $\chi^2_{k-l,\alpha}$, we have

$$\begin{aligned}
\pi_n\left(\theta + \frac{h}{\sqrt{n}}\right) &= P_{\theta+h/\sqrt{n}}\left(\Lambda_n > \chi^2_{k-l,\alpha}\right) \\
&\to P_{N(h,I_\theta^{-1})}\left(\left\| I_\theta^{1/2} X - I_\theta^{1/2} H_0 \right\|^2 > \chi^2_{k-l,\alpha}\right) \\
&= P_{N(0,I)}\left(\left\| X - I_\theta^{1/2}(-h + H_0) \right\|^2 > \chi^2_{k-l,\alpha}\right) \\
&= P\left(\chi^2_{k-l}\left(\left\| I_\theta^{1/2}(h - H_0) \right\|\right) > \chi^2_{k-l,\alpha}\right),
\end{aligned}$$

where $\chi^2_{k-l}(\delta)$ is a random variable with a noncentral chi-squared distribution with noncentrality parameter δ...

Likelihood ratio tests: Asymptotic power function

That is, $\chi^2_{k-l}(\delta)$ has the distribution of the squared distance between a standard normal in $\mathbb{R}^k$ and an affine subspace of dimension l that is distance δ from the origin.

$$P\left(\chi^2_{k-l}\left(\left\| I_\theta^{1/2}(h - H_0)\right\|\right) > \chi^2_{k-l,\alpha}\right)$$

is an increasing function of $\left\| I_\theta^{1/2}(h - H_0)\right\|$, and hence of $\|h\|$.

Likelihood ratio tests: Asymptotic power function

First, think of $H_0 = \{0\}$:

$$\left\| I_\theta^{1/2}(h - H_0) \right\| = \sqrt{h^T I_\theta h}.$$

Decomposing I_θ into outer products of its eigenvectors, we have

$$h^T I_\theta h = \sum_{i=1}^{k} \lambda_i (e_i^T h)^2.$$

So we get highest power in the directions that align with eigenvectors e_i that have largest eigenvalues λ_i. If the log likelihood is twice differentiable, these are the directions with a large second derivative: the variance of the score function is large in these directions.

And if H_0 is a subspace, replace h here with the difference between h and its projection on H_0, which is in the space orthogonal to H_0.