Empirical Likelihood

1 Mean Case

Suppose $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ are iid with mean $\boldsymbol{\mu}$ and covariance-variance Σ. For simplicity, say we are interested in estimating $\boldsymbol{\mu}$.

To perform *nonparametric* likelihood inference on $\boldsymbol{\mu}$, we can consider a constrained multinomial likelihood, known as the **Empirical Likelihood function of $\boldsymbol{\mu}$:**

$$L_n(\boldsymbol{\mu}|\boldsymbol{Y}) = \sup\left\{\prod_{i=1}^n p_i : p_i \mapsto \boldsymbol{Y}_i, \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n \boldsymbol{Y}_i p_i = \boldsymbol{\mu}\right\}.$$

The largest possible value of $L_n(\boldsymbol{\mu}|\boldsymbol{Y})$ is

2 Statistical Inference

We can form an EL ratio for $\boldsymbol{\mu}$

$$R_n(\boldsymbol{\mu}) = \frac{L_n(\boldsymbol{\mu}|\boldsymbol{Y})}{L_n(\hat{\boldsymbol{\mu}}|\boldsymbol{Y})}$$

Theorem (Wilk's Theorem): If $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n \in \mathbb{R}^q$ are iid with mean $\boldsymbol{\mu}_0$ and covariance-variance Σ where $\text{rank}(\Sigma) = q$, then

$$-2 \log R_n(\boldsymbol{\mu}_0) \stackrel{d}{\to} \chi^2_q \text{ as } n \to \infty.$$

3 EL with Estimating Equations

For EL inference on $\boldsymbol{\theta} \in \mathbb{R}^b$, we make an EL function

Then we can get a point estimate, EL ratio, and corresponding CIs, as well as “profile” EL:

Theorem: Suppose $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \cdots \in \mathbb{R}^q$ are iid with $\mathrm{E}\boldsymbol{\psi}(\boldsymbol{Y}_1, \boldsymbol{\theta}_0) = \mathbf{0}_r$ and $\mathrm{Var}[\boldsymbol{\psi}(\boldsymbol{Y}_1, \boldsymbol{\theta}_0)] = \mathrm{E}\boldsymbol{\psi}(\boldsymbol{Y}_1, \boldsymbol{\theta}_0)\boldsymbol{\psi}(\boldsymbol{Y}_1, \boldsymbol{\theta}_0)^\top$ is positive definite, where $\boldsymbol{\theta}_0$ denotes the true parameter value.

Suppose also that $\partial\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$ and $\partial^2\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$ are continuous in a neighborhood of $\boldsymbol{\theta}_0$ and that, in this neighborhood, $||\boldsymbol{\psi}(\boldsymbol{Y}_1, \boldsymbol{\theta})||^3$, $||\partial\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}||$ and $||\partial^2\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top||$ are bounded by an integrable function $\Psi(\boldsymbol{Y}_1)$.

Finally, suppose the $r \times b$ matrix $D_{\boldsymbol{\psi}} \equiv \mathrm{E}\partial\boldsymbol{\psi}(\boldsymbol{y}, \boldsymbol{\theta})/\partial\boldsymbol{\theta}$ has full column rank b.

Then, as $n \to \infty$,

i. $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}_b, V)$, where $V = (D_{\boldsymbol{\psi}}^\top \mathrm{Var}[\boldsymbol{\psi}(\boldsymbol{Y}_1, \boldsymbol{\theta}_0)] D_{\boldsymbol{\psi}})^{-1}$.

ii. If $r > b$, the asymptotic variance V cannot increase if an estimating function is added.

iii. To test $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$, we may use $-2 \log R_n(\boldsymbol{\theta}_0)$ and when H_0 is true,

$$-2 \log R_n(\boldsymbol{\theta}_0) \xrightarrow{d} \chi^2_b$$

iv. If $r > b$, to test $H_0 : \mathrm{E}\psi(\boldsymbol{Y}_1, \boldsymbol{\theta}) = \mathbf{0}_r$ holds for some $\boldsymbol{\theta}$, we may use

$$-2 \log \frac{L_n(\hat{\boldsymbol{\theta}})}{\prod_{i=1}^{n} (1/n)} =$$

and when H_0 is true this quantity converges in distribution to χ^2_{r-b}.

v. To test the profile assumption $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_1^0 \in \mathbb{R}^q$, we can use the profile EL ratio $-2 \log R_n(\boldsymbol{\theta}_1^0)$ and , when H_0 is true, $-2 \log R_n(\boldsymbol{\theta}_1^0) \xrightarrow{d} \chi^2_q$.

4 Computation

Technically, for a given value of $\boldsymbol{\theta}$, define $L_n(\boldsymbol{\theta}|\boldsymbol{Y}) = 0$ if

$$\mathcal{A}_n(\boldsymbol{\theta}) = \left\{ \prod_{i=1}^{n} p_i : p_i \mapsto \boldsymbol{Y}_i, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} \boldsymbol{p}_i \boldsymbol{\psi}(\mathbf{Y}_i, \boldsymbol{\theta}) = \mathbf{0}_r \right\}$$

is empty.

If $\mathbf{0}_r$ is in the interior convex hull of $\{\psi(\mathbf{Y}_i, \boldsymbol{\theta})\}_{i=1}^{n}$, then $\mathcal{A}_n(\boldsymbol{\theta})$ will not be empty.

The supremum in the definition of $L_n(\boldsymbol{\theta}|\boldsymbol{Y})$ looks nasty, but the form simplifies if $L_n(\boldsymbol{\theta}|\boldsymbol{Y}) > 0$ for a given $\boldsymbol{\theta} \in \mathbb{R}^b$. To see this, fix $\boldsymbol{\theta}$ and let

$$\mathcal{B}_n(\boldsymbol{\theta}) = \left\{ (p_1, \ldots, p_n) : p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} \boldsymbol{p}_i \boldsymbol{\psi}(\mathbf{Y}_i, \boldsymbol{\theta}) = \mathbf{0}_r \right\} \subset [0,1]^n$$

To maximize $\prod_{i=1}^{n} p_i$ on $\mathcal{B}_n(\boldsymbol{\theta})$ and find $(p_1^*, \ldots, p_n^*)$, use Lagrange multipliers $a \in \mathbb{R}$ and $\boldsymbol{\lambda} \in \mathbb{R}^r$ and maximize

$$f(p_1, \ldots, p_n, a, \boldsymbol{\lambda}) = \log \prod_{i=1}^{n} p_i + a \left(1 - \sum_{i=1}^{n} p_i \right) - n\boldsymbol{\lambda}^\top \left(\sum_{i=1}^{n} \boldsymbol{p}_i \boldsymbol{\psi}(\mathbf{Y}_i, \boldsymbol{\theta}) \right)$$

over $p_i \in [0,1], a \in \mathbb{R}$, and $\boldsymbol{\lambda} \in \mathbb{R}^r$.

Take derivatives & set to zero: