

Large Sample Confidence Ellipsoids Based on Maximum Likelihood
Stat 543 Spring 2005

Basic Fact: If $\mathbf{Y} \sim N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then $(\mathbf{Y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \sim \chi_k^2$

So, if c is the upper α point of the χ_k^2 distribution,

$$P[(\mathbf{Y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) < c] = 1 - \alpha \quad ,$$

and thus, if $\mathbf{Y} = \mathbf{y}$, the set

$$\{\boldsymbol{\mu} \in \mathfrak{R}^k \mid (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu}) < c\}$$

is a $(1 - \alpha) \times 100\%$ confidence set for $\boldsymbol{\mu}$.

Application in ML Estimation Problems:

Suppose that $l \leq k$ and that

$$\boldsymbol{\theta}_{k \times 1} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta} \end{pmatrix}, \quad \mathbf{I}_1(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{I}_{111}(\boldsymbol{\theta}) & \mathbf{I}_{112}(\boldsymbol{\theta}) \\ \mathbf{I}_{121}(\boldsymbol{\theta}) & \mathbf{I}_{122}(\boldsymbol{\theta}) \end{pmatrix} \text{ and}$$

$$\mathbf{H}_n(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{H}_{n11}(\boldsymbol{\theta}) & \mathbf{H}_{n12}(\boldsymbol{\theta}) \\ \mathbf{H}_{n21}(\boldsymbol{\theta}) & \mathbf{H}_{n22}(\boldsymbol{\theta}) \end{pmatrix} = \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} l_n(\boldsymbol{\theta}) \right).$$

Then with a nicely behaved "MLE" $\widehat{\boldsymbol{\theta}}_n$

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N_k(\mathbf{0}, (\mathbf{I}_1(\boldsymbol{\theta}))^{-1})$$

so that

$$(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})' (n\mathbf{I}_1(\boldsymbol{\theta})) (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \chi_k^2$$

and under appropriate conditions, the same convergence holds for $n\mathbf{I}_1(\widehat{\boldsymbol{\theta}}_n)$ or $-\mathbf{H}_n(\widehat{\boldsymbol{\theta}}_n)$ replacing $n\mathbf{I}_1(\boldsymbol{\theta})$. So, large sample confidence sets for $\boldsymbol{\theta} \in \mathfrak{R}^k$ are

$$\{\boldsymbol{\theta} \in \mathfrak{R}^k \mid (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})' (n\mathbf{I}_1(\boldsymbol{\theta})) (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) < c\}$$

with either $n\mathbf{I}_1(\widehat{\boldsymbol{\theta}}_n)$ or $-\mathbf{H}_n(\widehat{\boldsymbol{\theta}}_n)$ replacing $n\mathbf{I}_1(\boldsymbol{\theta})$. The first possibility is the use of the "expected Fisher information," while the second is the use of the "observed Fisher information" and these confidence ellipsoids are sometimes called the "Wald" confidence sets.

One may also make confidence ellipsoids for a part of the parameter vector, say $\boldsymbol{\theta}_1$. This is based on the fact that if $\widehat{\boldsymbol{\theta}}_n$ is approximately k -dimensional normal, then $\widehat{\boldsymbol{\theta}}_{n1}$ is approximately l -dimensional normal. In fact,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_1) \xrightarrow{\mathcal{L}} N_l(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$$

for $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ the upper left $l \times l$ block of $(\mathbf{I}_1(\boldsymbol{\theta}))^{-1}$. So, for c' the upper α point of the χ_l^2 distribution, large sample confidence sets for $\boldsymbol{\theta}_1 \in \mathfrak{R}^l$ are

$$\{\boldsymbol{\theta}_1 \in \mathfrak{R}^l \mid n(\widehat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_1)' (\boldsymbol{\Sigma}(\boldsymbol{\theta}))^{-1} (\widehat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_1) < c'\}$$

where one must make an appropriate substitution for $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. One possibility is $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}_n)$ which is the use of the "expected Fisher information." Now using results about partitioned matrices

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \left(\mathbf{I}_{111}(\boldsymbol{\theta}) - \mathbf{I}_{112}(\boldsymbol{\theta}) (\mathbf{I}_{122}(\boldsymbol{\theta}))^{-1} \mathbf{I}_{121}(\boldsymbol{\theta}) \right)^{-1}$$

so these confidence sets are

$$\{\boldsymbol{\theta}_1 \in \mathfrak{R}^l \mid n(\widehat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_1)' \left(\mathbf{I}_{111}(\widehat{\boldsymbol{\theta}}_n) - \mathbf{I}_{112}(\widehat{\boldsymbol{\theta}}_n) (\mathbf{I}_{122}(\widehat{\boldsymbol{\theta}}_n))^{-1} \mathbf{I}_{121}(\widehat{\boldsymbol{\theta}}_n) \right)^{-1} (\widehat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_1) < c'\}$$

(Notice that this is more complicated than just taking the formula for the whole parameter vector and using the upper left block of the Fisher information matrix.) A second possibility is to note that $-\frac{1}{n}\mathbf{H}_n(\hat{\boldsymbol{\theta}}_n)$ can serve as an estimator of $\mathbf{I}_1(\boldsymbol{\theta})$. The "observed Fisher information" confidence sets are then

$$\{\boldsymbol{\theta}_1 \in \mathfrak{R}^d \mid -(\hat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_1)' \left(\mathbf{H}_{n11}(\hat{\boldsymbol{\theta}}_n) - \mathbf{H}_{n12}(\hat{\boldsymbol{\theta}}_n) \left(\mathbf{H}_{n22}(\hat{\boldsymbol{\theta}}_n) \right)^{-1} \mathbf{H}_{n21}(\hat{\boldsymbol{\theta}}_n) \right) (\hat{\boldsymbol{\theta}}_{n1} - \boldsymbol{\theta}_1) < c'\}$$