ENEE 621
SPRING 2016
DETECTION AND ESTIMATION THEORY

THE PARAMETER ESTIMATION PROBLEM

1 The basic setting

Throughout, p, q and k are positive integers.

The setup

With Θ being a Borel subset of $\mathbb{R}^p$, consider a parametrized family $\{F_\theta,\ \theta \in \Theta\}$ of probability distributions on $\mathbb{R}^k$. The problem considered here is that of estimating θ on the basis of some $\mathbb{R}^k$-valued observation whose statistical description depends on θ.

The setting is alway understood as follows: Given $(\Omega, \mathcal{F})$ some measurable space, consider a rv $\boldsymbol{Y} : \Omega \to \mathbb{R}^k$ defined on it. With $\{F_\theta,\ \theta \in \Theta\}$, we associate a collection of probability measures $\{\mathbb{P}_\theta,\ \theta \in \Theta\}$ defined on $\mathcal{F}$ such that

$$\mathbb{P}_\theta\left[\boldsymbol{Y} \in B\right] = \int_B dF_\theta(\boldsymbol{y}), \qquad \begin{array}{c} B \in \mathcal{B}(\mathbb{R}^k), \\ \theta \in \Theta. \end{array}$$

Sufficient statistics

It is customary to refer to any Borel mapping $T : \mathbb{R}^k \to \mathbb{R}^q$ as a *statistic*.

A statistic $T : \mathbb{R}^k \to \mathbb{R}^q$ is said to be *sufficient* for $\{F_\theta,\ \theta \in \Theta\}$, or alternatively, for estimating θ on the basis of $\boldsymbol{Y}$, if there exists a mapping $\gamma : \mathbb{R}^q \times \mathcal{B}(\mathbb{R}^k) \to [0,1]$ which satisfies the following conditions:

(i) For every B in $\mathcal{B}(\mathbb{R}^k)$, the mapping $\mathbb{R}^q \to [0,1] : \boldsymbol{t} \to \gamma(B; \boldsymbol{t})$ is Borel measurable;

(ii) For every $\boldsymbol{t}$ in $\mathbb{R}^q$, the mapping $\mathcal{B}(\mathbb{R}^k) \to [0,1] : B \to \gamma(B; \boldsymbol{t})$ is a probability measure on $\mathcal{B}(\mathbb{R}^k)$; and

(iii) For every θ in Θ, the property

$$\mathbb{P}_\theta\left[\boldsymbol{Y} \in B | T(\boldsymbol{Y}) = \boldsymbol{t}\right] = \gamma(B; \boldsymbol{t}) \quad \mathbb{P}_\theta - \text{a.s.} \qquad \begin{array}{c} B \in \mathcal{B}(\mathbb{R}^k) \\ \boldsymbol{t} \in \mathbb{R}^q \end{array}$$

holds.

In other words, the statistic $T : \mathbb{R}^k \to \mathbb{R}^q$ is sufficient for $\{F_\theta,\ \theta \in \Theta\}$ if the conditional distribution of $\boldsymbol{Y}$ under $\mathbb{P}_\theta$ given $T(\boldsymbol{Y})$ is *independent* of θ.

Completeness

The family $\{F_\theta,\ \theta \in \Theta\}$ is *complete* if whenever we consider a Borel mapping $\psi : \mathbb{R}^k \to \mathbb{R}$ such that

$$\mathbb{E}_\theta\left[|\psi(\boldsymbol{Y})|\right] < \infty, \quad \theta \in \Theta$$

the condition

$$\mathbb{E}_\theta\left[\psi(\boldsymbol{Y})\right] = 0, \quad \theta \in \Theta$$

implies

$$\mathbb{P}_\theta\left[\psi(\boldsymbol{Y}) = 0\right] = 1, \quad \theta \in \Theta.$$

Lemma 1.1 *If the family $\{F_\theta,\ \theta \in \Theta\}$ is complete, then there exists no non-trivial sufficient statistic for estimating θ on the basis of $\boldsymbol{Y}$.*

2 Finite variance estimators

An estimator for θ on the basis of $\boldsymbol{Y}$ is any Borel mapping $g : \mathbb{R}^k \to \mathbb{R}^p$. We define the estimation error at θ (in Θ) associated with the estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ as the rv $\varepsilon_g(\theta; \boldsymbol{Y})$ given by

$$\varepsilon_g(\theta; \boldsymbol{Y}) = g(\boldsymbol{Y}) - \theta.$$

Finite mean estimators

An estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is said to be a *finite mean* estimator if

$$\mathbb{E}_\theta\left[|g_i(\boldsymbol{Y})|\right] < \infty, \qquad \begin{array}{c} i = 1, \ldots, p \\ \theta \in \Theta. \end{array}$$

The *bias* of the finite mean estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ at θ is well defined and given by

$$b_\theta(g) = \mathbb{E}_\theta\left[\varepsilon_g(\theta; \boldsymbol{Y})\right] = \mathbb{E}_\theta\left[g(\boldsymbol{Y})\right] - \theta.$$

The finite mean estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is said to be *unbiased* at θ if $b_\theta(g) = 0$. Furthermore, the finite mean estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is said to be *unbiased* if

$$\mathbb{E}_\theta\left[g(\boldsymbol{Y})\right] = \theta, \quad \theta \in \Theta.$$

Finite variance estimators

An estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is a *finite variance* estimator if

$$\mathbb{E}_\theta \left[|g_i(\boldsymbol{Y})|^2 \right] < \infty, \qquad \begin{matrix} i = 1, \ldots, p \\ \theta \in \Theta. \end{matrix}$$

Obviously, a finite variance estimator is also a finite mean estimator. The *error covariance* of the finite variance estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ at θ is the $p \times p$ matrix $\Sigma_\theta(g)$ given by

$$\Sigma_\theta(g) = \mathbb{E}_\theta \left[\varepsilon_g(\theta; \boldsymbol{Y}) \varepsilon_g(\theta; \boldsymbol{Y})' \right].$$

In general, the matrix $\Sigma_\theta(g)$ is *not* the covariance matrix of the error $g(\boldsymbol{Y})$; in fact we have

$$\Sigma_\theta(g) = \text{Cov}_\theta \left[g(\boldsymbol{Y}) \right] + b_\theta(g) b_\theta(g)', \quad \theta \in \Theta.$$

MVUEs

A finite variance estimator $g^\star : \mathbb{R}^k \to \mathbb{R}^p$ is said to be a *Minimum Variance Unbiased Estimator* (MVUE) if it is *unbiased* and

$$\Sigma_\theta(g^\star) \leq \Sigma_\theta(g), \quad \theta \in \Theta$$

for any other finite variance *unbiased* estimator $g : \mathbb{R}^k \to \mathbb{R}^p$. Alternatively, a finite variance estimator $g^\star : \mathbb{R}^k \to \mathbb{R}^p$ is said to be an MVUE if it is an unbiased estimator and

$$\text{Cov}_\theta \left[g^\star(\boldsymbol{Y}) \right] \leq \text{Cov}_\theta \left[g(\boldsymbol{Y}) \right], \quad \theta \in \Theta$$

for any other finite variance unbiased estimator $g : \mathbb{R}^k \to \mathbb{R}^p$.

Under the completeness of the family $\{F_\theta, \ \theta \in \Theta\}$, unbiased estimators for θ on the basis of $\boldsymbol{Y}$ are essentially unique in the following sense.

Lemma 2.1 *Assume the family $\{F_\theta, \ \theta \in \Theta\}$ to be complete. If the finite mean estimators $g_1, g_2 : \mathbb{R}^k \to \mathbb{R}^p$ are unbiased, then*

$$\mathbb{P}_\theta \left[g_1(\boldsymbol{Y}) = g_2(\boldsymbol{Y}) \right] = 1, \quad \theta \in \Theta.$$

3 The Rao-Blackwell Theorem

A basic step in the search for MVUEs is provided by the Rao-Blackwell Theorem.

Complete sufficient statistic

A statistic $T : \mathbb{R}^k \to \mathbb{R}^q$ is said to be a *complete sufficient* statistic for $\{F_\theta,\ \theta \in \Theta\}$ if it is a sufficient statistic for $\{F_\theta,\ \theta \in \Theta\}$ such that the family $\{H_\theta,\ \theta \in \Theta\}$ of probability distributions on $\mathbb{R}^q$ is complete where

$$H_\theta(\boldsymbol{t}) = \mathbb{P}_\theta\left[T(\boldsymbol{Y}) \leq \boldsymbol{t}\,\right], \qquad \begin{array}{l} \boldsymbol{t} \in \mathbb{R}^q \\ \theta \in \Theta. \end{array}$$

Rao-Blackwell Theorem

The Rao-Blackwell Theorem given next can be viewed as providing a "variance" reduction algorithm.

Theorem 3.1 *Let $T : \mathbb{R}^k \to \mathbb{R}^q$ be a sufficient statistic for $\{F_\theta,\ \theta \in \Theta\}$. With any finite variance estimator $g : \mathbb{R}^k \to \mathbb{R}^p$, define the mapping $\hat{g} : \mathbb{R}^k \to \mathbb{R}^q$ given by*

$$\widehat{g}(\boldsymbol{t}) = \int_{\mathbb{R}^k} g(\boldsymbol{y}) d\gamma(\boldsymbol{y}, \boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{R}^q$$

where the mapping $\gamma : \mathbb{R}^q \times \mathcal{B}(\mathbb{R}^k) \to [0,1]$ is the one appearing in the definition of the sufficiency of the statistic $T : \mathbb{R}^k \to \mathbb{R}^q$.

The mapping $\widehat{g} \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is a finite variance estimator for θ on the basis of $\boldsymbol{Y}$ such that

$$b_\theta(\widehat{g} \circ T) = b_\theta(g)$$

and

$$\Sigma_\theta(\widehat{g} \circ T) \leq \Sigma_\theta(g)$$

for every θ in Θ. Moreover,

$$\Sigma_\theta(\widehat{g} \circ T) = \Sigma_\theta(g)$$

at some θ in Θ iff

$$\mathbb{P}_\theta\left[g(\boldsymbol{Y}) = \widehat{g}(T(\boldsymbol{Y}))\right] = 1.$$

The "algorithm" that takes the estimator g into the estimator $\widehat{g} \circ T$ does not change the bias but reduces variability. These properties are simple consequences of Jensen's inequality (for conditional expectations) and of the law of iterated conditioning applied to the fact that

$$\hat{g}(T(\boldsymbol{Y})) = \mathbb{E}_\theta \left[g(\boldsymbol{Y}) | T(\boldsymbol{Y}) \right], \quad \mathbb{P}_\theta - a.s.$$

for every θ in Θ. The Rao-Blackwell Theorem has the following consequence.

Corollary 3.1 *Let $T : \mathbb{R}^k \to \mathbb{R}^q$ be a sufficient statistic for $\{F_\theta,\ \theta \in \Theta\}$. Assume that there exists a Borel mapping $\tilde{g} : \mathbb{R}^q \to \mathbb{R}^p$ such that $\tilde{g} \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is a finite variance unbiased estimator for θ on the basis of $\boldsymbol{Y}$.*

Then, the estimator $\tilde{g} \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is an MVUE for θ on the basis of $\boldsymbol{Y}$ whenever the Borel mapping $\tilde{g} : \mathbb{R}^q \to \mathbb{R}^p$ is essentially unique in the following sense: If Borel mappings $\tilde{g}_1, \tilde{g}_2 : \mathbb{R}^q \to \mathbb{R}^p$ have the property that for each $i = 1, 2$, the estimator $\tilde{g}_i \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is a finite variance unbiased estimator for θ on the basis of $\boldsymbol{Y}$, then

$$\mathbb{P}_\theta \left[\tilde{g}_1(T(\boldsymbol{Y})) = \tilde{g}_2(T(\boldsymbol{Y})) \right] = 1, \quad \theta \in \Theta.$$

Finding MVUEs

The needed uniqueness condition in Corollary 3.1 can be guaranteed by asking for a stronger form of sufficiency for the statistic $T : \mathbb{R}^k \to \mathbb{R}^q$, namely its complete sufficiency.

Lemma 3.1 *Let $T : \mathbb{R}^k \to \mathbb{R}^q$ be a complete sufficient statistic for $\{F_\theta,\ \theta \in \Theta\}$. If there exists a Borel mapping $\tilde{g} : \mathbb{R}^q \to \mathbb{R}^p$ such that $\tilde{g} \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is a finite variance unbiased estimator for θ on the basis of $\boldsymbol{Y}$, then the following holds:*

(i) The Borel mapping $\tilde{g} : \mathbb{R}^q \to \mathbb{R}^p$ is essentially unique in the following sense: If the Borel mappings $\tilde{g}_1, \tilde{g}_2 : \mathbb{R}^q \to \mathbb{R}^p$ have the property that for each $i = 1, 2$, the estimator $\tilde{g}_i \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is a finite variance unbiased estimator for θ on the basis of $\boldsymbol{Y}$, then

$$\mathbb{P}_\theta \left[\tilde{g}_1(T(\boldsymbol{Y}) = \tilde{g}_2(T(\boldsymbol{Y})) \right] = 1, \quad \theta \in \Theta.$$

(ii) The estimator $\tilde{g} \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is MVUE.

Part (i) is a consequence of the fact that $T : \mathbb{R}^k \to \mathbb{R}^q$ is a complete sufficient statistic for $\{F_\theta,\ \theta \in \Theta\}$, and Part (ii) follows then by Corollary 3.1. Taken together, these results lead to the following strategy for finding MVUEs:

(i) Find a complete sufficient statistic $T : \mathbb{R}^k \to \mathbb{R}^q$ for $\{F_\theta,\ \theta \in \Theta\}$;

(ii) Find a finite variance *unbiased* estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ for θ on the basis of $\boldsymbol{Y}$ – This step is often implemented by guessing $g = \tilde{g} \circ T$ for some Borel mapping $\tilde{g} : \mathbb{R}^q \to \mathbb{R}^p$;

(iii) Absent such a guess, generate from g the Borel mapping $\hat{g} : \mathbb{R}^q \to \mathbb{R}^p$ as per the Rao-Blackwell Theorem. The estimator $\hat{g} \circ T$ is MVUE.

4 Exponential families

Recall that the family $\{F_\theta,\ \theta \in \Theta\}$ is an *exponential* family (with respect to F) if its absolutely continuous with respect to F, and the corresponding density functions $\{f_\theta,\ \theta \in \Theta\}$ are of the form

$$f_\theta(\boldsymbol{y}) = C(\theta)q(\boldsymbol{y})e^{Q(\theta)'K(\boldsymbol{y})} \quad F - \text{a.e.} \tag{1}$$

for every θ in Θ with Borel mappings $C : \Theta \to \mathbb{R}_+$, $Q : \Theta \to \mathbb{R}^q$, $q : \mathbb{R}^k \to \mathbb{R}_+$ and $K : \mathbb{R}^k \to \mathbb{R}^q$. The requirement

$$\int_{\mathbb{R}^k} f_\theta(\boldsymbol{y})dF(\boldsymbol{y}) = 1$$

reads

$$C(\theta)\int_{\mathbb{R}^k} q(\boldsymbol{y})e^{Q(\theta)'K(\boldsymbol{y})}dF(\boldsymbol{y}) = 1.$$

This is equivalent to

$$C(\theta) > 0$$

and

$$0 < \int_{\mathbb{R}^k} q(\boldsymbol{y})e^{Q(\theta)'K(\boldsymbol{y})}dF(\boldsymbol{y}) < \infty.$$

Exponential families and sufficient statistics

An exponential family always admits at least one sufficient statistic.

Theorem 4.1 *Assume $\{F_\theta,\ \theta \in \Theta\}$ to be an exponential family (with respect to F) with representation (1). Then, the mapping $K : \mathbb{R}^k \to \mathbb{R}^q$ is a sufficient statistic for $\{F_\theta,\ \theta \in \Theta\}$.*

The sufficient statistic $K : \mathbb{R}^k \to \mathbb{R}^q$ for $\{F_\theta,\ \theta \in \Theta\}$ admits a simple characterization as a complete sufficient statistic.

Theorem 4.2 *Assume $\{F_\theta,\ \theta \in \Theta\}$ to be an exponential family (with respect to F) with representation (1). Then, the mapping $K : \mathbb{R}^k \to \mathbb{R}^q$ is a complete sufficient statistic for $\{F_\theta,\ \theta \in \Theta\}$ if the set*

$$Q(\Theta) = \{Q(\theta) :\ \theta \in \Theta\}\,.$$

contains a q-dimensional rectangle.

A proof

Consider a Borel mapping $\psi : \mathbb{R}^q \to \mathbb{R}$ such that

$$\mathbb{E}_\theta\left[|\psi(K(\boldsymbol{Y}))|\right] < \infty, \quad \theta \in \Theta.$$

We need to show that if

$$\mathbb{E}_\theta\left[\psi(K(\boldsymbol{Y}))\right] = 0, \quad \theta \in \Theta$$

then

$$\mathbb{P}_\theta[\psi(K(\boldsymbol{Y})) = 0] = 1, \quad \theta \in \Theta.$$

The integrability conditions are equivalent to

$$\int_{\mathbb{R}^k} |\psi(K(\boldsymbol{y}))| q(\boldsymbol{y}) e^{Q(\theta)' K(\boldsymbol{y})} dF(\boldsymbol{y}) < \infty, \quad \theta \in \Theta.$$

With $\boldsymbol{u} = (u_1, \ldots, u_q)'$ in $\mathbb{C}^q$, we note that

$$\int_{\mathbb{R}^k} |\psi(K(\boldsymbol{y})) q(\boldsymbol{y}) e^{\boldsymbol{u}' K(\boldsymbol{y})}| dF(\boldsymbol{y}) < \infty$$

as soon as $\Re(\boldsymbol{u}) = ((\Re(u_1), \ldots, \Re(u_q))'$ lies in $Q(\Theta)$. This is a consequence of the fact that

$$|\psi(K(\boldsymbol{y})) q(\boldsymbol{y}) e^{\boldsymbol{u}' K(\boldsymbol{y})}| = q(\boldsymbol{y}) |\psi(K(\boldsymbol{y}))| \cdot |e^{\boldsymbol{u}' K(\boldsymbol{y})}|$$

where

$$
\begin{aligned}
|e^{\boldsymbol{u}'K(\boldsymbol{y})}| &= |\prod_{i=1}^{q} e^{u_i K_i(\boldsymbol{y})}| \\
&= |\prod_{i=1}^{q} e^{(\Re(u_i)+j\Im(u_i))K_i(\boldsymbol{y})}| \\
&= |\prod_{i=1}^{q} e^{\Re(u_i)K_i(\boldsymbol{y})}| \\
&= \prod_{i=1}^{q} e^{\Re(u_i)K_i(\boldsymbol{y})} \qquad (2)
\end{aligned}
$$

so that

$$\int_{\mathbb{R}^k} |\psi(K(\boldsymbol{y}))q(\boldsymbol{y})e^{\boldsymbol{u}'K(\boldsymbol{y})}|dF(\boldsymbol{y}) = \int_{\mathbb{R}^k} |\psi(K(\boldsymbol{y}))|q(\boldsymbol{y})e^{\Re(\boldsymbol{u})'K(\boldsymbol{y})}dF(\boldsymbol{y}).$$

Let R denotes a q-dimensional rectangle contained in $Q(\Theta)$, i.e.,

$$R = \prod_{i=1}^{q}[a_i, b_i] \subseteq Q(\Theta).$$

The arguments given above then show that on the subset $R^\star$ given by

$$R^\star = \prod_{i=1}^{q}\left([a_i, b_i] + j\mathbb{R}\right),$$

the $\mathbb{C}$-valued integral

$$\widehat{\Psi}(\boldsymbol{u}) \equiv \int_{\mathbb{R}^k} \psi(K(\boldsymbol{y}))q(\boldsymbol{y})e^{\boldsymbol{u}'K(\boldsymbol{y})}dF(\boldsymbol{y})$$

is *well defined* as soon as $\boldsymbol{u} = (u_1, \ldots, u_q)'$ lies in $R^\star$ (hence in R).

Under the enforced assumptions on the mapping $\Psi : \mathbb{R}^q \to \mathbb{R}$, we have

$$\widehat{\Psi}(\boldsymbol{u}) = 0, \quad \boldsymbol{u} \in R.$$

Standard properties of functions of complex variables imply that

$$\widehat{\Psi}(\boldsymbol{u}) = 0, \quad \boldsymbol{u} \in R^\star.$$

In particular, given the form of $R^\star$, we also have

$$\widehat{\Psi}(\boldsymbol{a} + j\boldsymbol{u}) = 0, \quad \boldsymbol{u} \in \mathbb{R}^q$$

where $\boldsymbol{a} = (a_1, \ldots, a_q)$. It now follows the theory of Fourier transforms that

$$\psi(K(\boldsymbol{y}))q(\boldsymbol{y})e^{\boldsymbol{a}'K(\boldsymbol{y})} = 0 \quad F - a\text{a.e.}$$

and the desired conclusion is readily obtained.

5 The Cramèr-Rao bounds

The Cramèr-Rao bound requires certain technical conditions to be satisfied by the family $\{F_\theta,\ \theta \in \Theta\}$.

CR1 The parameter set Θ is an open set in $\mathbb{R}^p$;

CR2a The probability distributions $\{F_\theta,\ \theta \in \Theta\}$ are all absolutely continuous with respect to the same distribution $F : \mathbb{R}^k \to \mathbb{R}_+$. Thus, for each θ in Θ, there exists a Borel mapping $f_\theta : \mathbb{R}^k \to \mathbb{R}_+$ such that

$$F_\theta(\boldsymbol{y}) = \int_\infty^{\boldsymbol{y}} f_\theta(\boldsymbol{\eta})dF(\boldsymbol{\eta}), \quad \boldsymbol{y} \in \mathbb{R}^k;$$

CR2b The density functions $\{f_\theta,\ \theta \in \Theta\}$ all have the same support in the sense that the set $\{\boldsymbol{y} \in \mathbb{R}^k :\ f_\theta(\boldsymbol{y}) > 0\}$ is the same for all θ in Θ. Let S denote this common support;

CR3 For each θ in Θ, the gradient $\nabla_\theta f_\theta(\boldsymbol{y})$ exists and is finite on S;

CR4 For each θ in Θ, the square integrability condition

$$\mathbb{E}_\theta\left[\left|\frac{\partial}{\partial\theta_i}\log f_\theta(\boldsymbol{Y})\right|^2\right] < \infty, \quad i = 1, \ldots, p$$

holds;

CR5 For each θ in Θ, the regularity condition

$$\frac{\partial}{\partial\theta_i}\int_S f_\theta(\boldsymbol{y})dF(\boldsymbol{y}) = \int_S \left(\frac{\partial}{\partial\theta_i} f_\theta(\boldsymbol{y})\right) dF(\boldsymbol{y}), \quad i = 1, \ldots, p$$

holds. This is equivalent to asking

$$\int_S \left(\frac{\partial}{\partial\theta_i} f_\theta(\boldsymbol{y})\right) dF(\boldsymbol{y}) = 0, \quad i = 1, \ldots, p$$

since

$$\int_S f_\theta(\boldsymbol{y})dF(\boldsymbol{y}) = 1.$$

The Fisher information matrix

Under Conditions **(CR1)–(CR4)**, define the *Fisher information matrix* $M(\theta)$ at parameter θ as the $p \times p$ matrix given entrywise by

$$M_{ij}(\theta) = \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta_i}\log f_\theta(\boldsymbol{Y}) \cdot \frac{\partial}{\partial\theta_j}\log f_\theta(\boldsymbol{Y})\right], \quad i, j = 1, \ldots, p,$$

or equivalently,

$$M(\theta) = \mathbb{E}_\theta\left[\left(\nabla_\theta \log f_\theta(\boldsymbol{Y})\right)\left(\nabla_\theta \log f_\theta(\boldsymbol{Y})\right)'\right].$$

Regular estimators

A finite variance estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is a *regular* estimator (with respect to the family $\{F_\theta,\ \theta \in \Theta\}$) if the regularity conditions

$$\frac{\partial}{\partial\theta_i}\left(\int_S g(\boldsymbol{y})f_\theta(\boldsymbol{y})dF(\boldsymbol{y})\right) = \int_S g(\boldsymbol{y})\left(\frac{\partial}{\partial\theta_i} f_\theta(\boldsymbol{y})\right) dF(\boldsymbol{y}), \quad i = 1, \ldots, p$$

hold for all θ in Θ.

The regularity of an estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ amounts to

$$\frac{\partial}{\partial\theta_i}\left(\mathbb{E}_\theta\left[g(\boldsymbol{Y})\right]\right) = \mathbb{E}_\theta\left[g(\boldsymbol{Y})\left(\frac{\partial}{\partial\theta_i}\log f_\theta(\boldsymbol{Y})\right)\right], \quad i = 1, \ldots, p.$$

The bounds

The generalized Cramèr-Rao bound is given first

Theorem 5.1 *Assume Conditions* **(CR1)–(CR5)**. *If the Fisher information matrix $M(\theta)$ is invertible for each θ in Θ, then every regular estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ (with respect to the family $\{F_\theta,\ \theta \in \Theta\}$) obeys the lower bound*

$$\Sigma_\theta(g) \geq b_\theta(g)b_\theta(g)' + (\boldsymbol{I}_p + \nabla_\theta b_\theta(g))\, M(\theta)^{-1} (\boldsymbol{I}_p + \nabla_\theta b_\theta(g))'\,.$$

Equality holds at θ in Θ if and only if there exists a $p \times p$ matrix $K(\theta)$ such that

$$g(\boldsymbol{Y}) - \theta = b_\theta(g) + K(\theta)\nabla_\theta \log f_\theta(\boldsymbol{Y}) \quad F - \text{a.e.}$$

with

$$K(\theta) = (\boldsymbol{I}_p + \nabla_\theta b_\theta(g))\, M(\theta)^{-1}.$$

The classical Cramèr-Rao bound holds for unbiased estimators, and is now a simple corollary of Theorem 5.1.

Theorem 5.2 *Assume Conditions* **(CR1)–(CR5)**. *If the Fisher information matrix $M(\theta)$ is invertible for each θ in Θ, then every unbiased regular estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ (with respect to the family $\{F_\theta,\ \theta \in \Theta\}$) obeys the lower bound*

$$\Sigma_\theta(g) \geq M(\theta)^{-1}.$$

Equality holds at θ in Θ if and only if there exists a $p \times p$ matrix $K(\theta)$ such that

$$g(\boldsymbol{Y}) - \theta = K(\theta)\nabla_\theta \log f_\theta(\boldsymbol{Y}) \quad F - \text{a.e.}$$

with

$$K(\theta) = M(\theta)^{-1}.$$

The Fisher information matrix is often computed through an alternate expression given next. It requires three additional conditions. The first one provides smoothness beyond **(CR3)**.

CR6 For each θ in Θ, the partial derivatives

$$\frac{\partial^2}{\partial\theta_i \partial\theta_i j} f_\theta(\boldsymbol{y}), \qquad i, j = 1, \ldots, p$$

all exist and are finite on S;

CR7 For each θ in Θ,

$$\mathbb{E}_\theta\left[\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j} f_\theta(\boldsymbol{Y})\right|\right] < \infty, \quad i,j = 1,\ldots,p;$$

CR8 For each θ in Θ, the regularity conditions

$$\frac{\partial^2}{\partial\theta_i\partial\theta_i j}\int_S f_\theta(\boldsymbol{y})dF(\boldsymbol{y}) = \int_S \frac{\partial^2}{\partial\theta_i\partial\theta_j} f_\theta(\boldsymbol{y})dF(\boldsymbol{y}), \quad i,j = 1,\ldots,p$$

hold. This is equivalent to asking

$$\int_S \frac{\partial^2}{\partial\theta_i\partial\theta_j} f_\theta(\boldsymbol{y})dF(\boldsymbol{y}) = 0, \quad i,j = 1,\ldots,p$$

Lemma 5.1 *Assume Conditions* **(CR1)–(CR8)** *to hold. Then, the Fisher information matrix takes the form*

$$M_{ij}(\theta) = -\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f_\theta(\boldsymbol{Y})\right], \quad i,j = 1,\ldots,p$$

Facts and arguments

Two key facts flow from the assumptions: Fix θ in Θ. From **(CR3)** and **(CR5)**, we get

$$\mathbb{E}_\theta\left[\nabla_\theta \log f_\theta(\boldsymbol{Y})\right] = \mathbf{0}_p.$$

Recall that

$$\mathbb{E}_\theta\left[g(\boldsymbol{Y})\right] = \theta + b_\theta(g), \quad \theta \in \Theta.$$

Thus, if the estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is regular, differentiating and using **(CR3)**, we conclude that

$$\boldsymbol{I}_p + \nabla_\theta b_\theta(g) = \mathbb{E}_\theta\left[g(\boldsymbol{Y})\left(\nabla_\theta \log f_\theta(\boldsymbol{Y})\right)'\right].$$

Therefore,

$$\begin{aligned} &\boldsymbol{I}_p + \nabla_\theta b_\theta(g) \\ &= \mathbb{E}_\theta\left[\left(g(\boldsymbol{Y}) - \mathbb{E}_\theta\left[g(\boldsymbol{Y})\right]\right)\cdot\left(\nabla_\theta \log f_\theta(\boldsymbol{Y})\right)'\right] \\ &= \mathbb{E}_\theta\left[\left(g(\boldsymbol{Y}) - \theta - b_\theta(g)\right)\cdot\left(\nabla_\theta \log f_\theta(\boldsymbol{Y})\right)'\right]. \end{aligned}$$

When $p = 1$, this last relation forms the basis for a proof via the Cauchy-Schwarz inequality. An alternate proof, valid for arbitrary p, can be obtained as follows: Introduce the $\mathbb{R}^p$-valued rv $U(\theta, \boldsymbol{Y})$ given by

$$U(\theta, \boldsymbol{Y}) = g(\boldsymbol{Y}) - \theta - b_\theta(g) - (\boldsymbol{I}_p + \nabla_\theta b_\theta(g))\, M(\theta)^{-1} \nabla_\theta \log f_\theta(\boldsymbol{Y}), \quad \theta \in \Theta.$$

Note that the rv $U(\theta, \boldsymbol{Y})$ has zero mean since

$$\begin{aligned} &\mathbb{E}_\theta\left[U(\theta, \boldsymbol{Y})\right] \\ &= \mathbb{E}_\theta\left[g(\boldsymbol{Y})\right] - \theta - b_\theta(g) - (\boldsymbol{I}_p + \nabla_\theta b_\theta(g))\, M(\theta)^{-1} \mathbb{E}_\theta\left[\nabla_\theta \log f_\theta(\boldsymbol{Y})\right] \\ &= \mathbf{0}_p. \end{aligned}$$

The Cramèr-Rao bound is equivalent to the statement that the covariance matrix $\mathrm{Cov}_\theta[U(\theta, \boldsymbol{Y})]$ is positive semi-definite! Indeed, it is straightforward to check that

$$\begin{aligned} \mathrm{Cov}_\theta[U(\theta, \boldsymbol{Y})] &= \mathrm{Cov}_\theta\left[U(\theta, \boldsymbol{Y})\right] \\ &= \Sigma_\theta(g) - b_\theta(g) b_\theta(g)' \\ &\quad - (\boldsymbol{I}_p + \nabla_\theta b_\theta(g))\, M(\theta)^{-1} \left(\boldsymbol{I}_p + \nabla_\theta b_\theta(g)\right)'. \end{aligned} \tag{3}$$

In particular, $\mathrm{Cov}_\theta[U(\theta, \boldsymbol{Y})] = \boldsymbol{O}_p$ iff

$$\mathbb{P}_\theta\left[U(\theta, \boldsymbol{Y})\right] = \mathbf{0}_p] = 1.$$

Efficient estimators

A finite variance *unbiased* estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is an *efficient* estimator if it achieves the Cràmer-Rao bound, namely

$$\Sigma_\theta(g) = M(\theta)^{-1}, \quad \theta \in \Theta.$$

Efficiency is meaningless for unbiased estimators!

Lemma 5.2 *Assume Conditions* **(CR1)–(CR5)** *to hold. A regular estimator that is also efficient satisfies the relations*

$$g(y) - \theta = M(\theta)^{-1} \nabla_\theta \log f_\theta(\boldsymbol{y}) \quad F - \text{a.e. on } S$$

for each θ on Θ. Conversely, any estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ which satisfies

$$g(y) - \theta = M(\theta)^{-1} \nabla_\theta \log f_\theta(\boldsymbol{y}) \quad F - \text{a.e. on } S$$

on Θ is an efficient regular estimator.

As an immediate corollary we have the following.

Corollary 5.1 *Assume Conditions* **(CR1)–(CR5)** *to hold. If an efficient regular estimator* $g : \mathbb{R}^k \to \mathbb{R}^p$ *exists, it is essentially unique on* S *in the sense that if* $g_1, g_2 : \mathbb{R}^k \to \mathbb{R}^p$ *are two efficient regular estimators, then* $g_1(\boldsymbol{y}) = g_2(\boldsymbol{y})$ F*-a.e. on* S.

6 Cramèr-Rao bounds for exponential families

Assume the family $\{F_\theta,\ \theta \in \Theta\}$ to be an exponential family (with respect to F) with density functions of the form

$$f_\theta(\boldsymbol{y}) = C(\theta) q(\boldsymbol{y}) e^{Q(\theta)' K(\boldsymbol{y})} \quad F - \text{a.e.}$$

for every θ in Θ with Borel mappings $C : \Theta \to \mathbb{R}_+$, $Q : \Theta \to \mathbb{R}^q$, $q : \mathbb{R}^k \to \mathbb{R}_+$ and $K : \mathbb{R}^k \to \mathbb{R}^q$. Conditions **(CR1)–(CR8)** can now be expressed more simply as follows:

Condition **(CR2a)** is obviously satisfied . Note that $f_\theta(\boldsymbol{y}) > 0$ if and only if $q(\boldsymbol{y}) > 0$, whence

$$\left\{\boldsymbol{y} \in \mathbb{R}^k :\ f_\theta(\boldsymbol{y}) > 0\right\} = \left\{\boldsymbol{y} \in \mathbb{R}^k :\ q(\boldsymbol{y}) > 0\right\}$$

for each θ in Θ, and **(CR2b)** holds.

Next, observe that here

$$\text{(4)} \qquad \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) = \frac{\partial}{\partial \theta_i} \log C(\theta) + \frac{\partial}{\partial \theta_i} Q(\theta)' K(\boldsymbol{y}), \qquad \begin{array}{c} i = 1, \ldots, p \\ \boldsymbol{y} \in S \end{array}$$

upon assuming the existence of the various derivatives. Therefore, **(CR3)** is equivalent to the differentiability of the mappings $C : \Theta \to \mathbb{R}_+$ and $Q : \Theta \to \mathbb{R}^q$. It follows that **(CR4)** is equivalent to

$$\mathbb{E}_\theta \left[|K_\ell(\boldsymbol{Y})|^2 \right] < \infty, \quad \ell = 1, \ldots, p.$$

Furthermore, the regularity condition **(CR5)** is easily seen to be equivalent to

$$\text{(5)} \qquad \left(\frac{\partial}{\partial \theta_i} Q(\theta) \right)' \mathbb{E}_\theta \left[K(\boldsymbol{Y}) \right] = -\frac{\partial}{\partial \theta_i} \log C(\theta), \quad \begin{array}{c} \theta_i \in \Theta \\ i = 1, \ldots, p. \end{array}$$

Combining (4) and (5) we get

$$\text{(6)} \quad \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) = \left(\frac{\partial}{\partial \theta_i} Q(\theta) \right)' (K(\boldsymbol{y}) - \mathbb{E}_\theta [K(\boldsymbol{Y})]), \qquad \begin{array}{c} i = 1, \ldots, p \\ \boldsymbol{y} \in S \end{array}$$

It is now straightforward to see that

$$\text{(7)} \quad M_{ij}(\theta) = \left(\frac{\partial}{\partial \theta_i} Q(\theta) \right)' \mathrm{Cov}_\theta [K(\boldsymbol{Y})] \left(\frac{\partial}{\partial \theta_i j} Q(\theta) \right), \qquad \begin{array}{c} i, j = 1, \ldots, p \\ \theta \in \Theta \end{array}$$

The regularity of the estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ can be expressed as the equalities

$$\begin{aligned} & \frac{\partial}{\partial \theta_i} (\theta + b_\theta(g)) \\ = \; & \mathbb{E}_\theta \left[(g(\boldsymbol{Y}) - \mathbb{E}_\theta [g(\boldsymbol{Y})]) \left(\frac{\partial}{\partial \theta_i} Q(\theta) \right)' K(\boldsymbol{Y}) \right], \quad i = 1, \ldots, p \end{aligned}$$

on Θ.

The case $p = 1$ with $q = 1$

Under these conditions $K : \mathbb{R} \to \mathbb{R}$ is a bone fide estimator of θ on the basis of Y. In view of the last equality above (with $g = K$), this estimator is regular if

$$\begin{aligned} \frac{d}{d\theta} \mathbb{E}_\theta [K(Y)] &= \frac{d}{d\theta} Q(\theta) \cdot \mathbb{E}_\theta [(K(\boldsymbol{Y}) - \mathbb{E}_\theta [K(\boldsymbol{Y})]) K(\boldsymbol{Y})] \\ \text{(8)} \qquad &= \frac{d}{d\theta} Q(\theta) \cdot \mathrm{Var}_\theta [K(Y)], \quad \theta \in \Theta. \end{aligned}$$

The estimator $K : \mathbb{R} \to \mathbb{R}$ then satisfies the Cramèr-Rao bound with equality, as this is equivalent to

$$\begin{aligned} \text{(9)} \quad & K(y) - \theta \\ = \; & b_\theta(K) + \left(1 + \frac{d}{d\theta} b_\theta(K) \right) M(\theta)^{-1} \frac{\partial}{\partial \theta} \log f_\theta(y) \quad F - a.e. \text{ on } S \end{aligned}$$

for every θ in Θ. The validity of (9) follows by direct substitution as we note that

$$1 + \frac{d}{d\theta} b_\theta(K) = \frac{d}{d\theta} \mathbb{E}_\theta [K(Y)] = \frac{d}{d\theta} Q(\theta) \cdot \mathrm{Var}_\theta [K(Y)],$$

$$M(\theta) = \left(\frac{d}{d\theta} Q(\theta) \right)^2 \cdot \mathrm{Var}_\theta [K(Y)].$$

and

$$\frac{\partial}{\partial\theta}\log f_\theta(y) = \frac{d}{d\theta}Q(\theta)\left(K(y) - \mathbb{E}_\theta\left[K(Y)\right]\right).$$

If the estimator $K : \mathbb{R} \to \mathbb{R}$ is also unbiased, that is

$$\mathbb{E}_\theta\left[K(Y)\right] = \theta, \quad \theta \in \Theta$$

the condition for it being regular reads

$$\frac{d}{d\theta}Q(\theta) \cdot \mathrm{Var}_\theta\left[K(Y)\right] = 1, \quad \theta \in \Theta$$

whence

$$M(\theta) = \frac{d}{d\theta}Q(\theta), \quad \theta \in \Theta$$

and the estimator $K : \mathbb{R} \to \mathbb{R}$ indeed achieves the Cramèr-Rao bound since

$$\mathrm{Var}_\theta\left[K(Y)\right] = M(\theta)^{-1}, \quad \theta \in \Theta.$$

Therefore, the (assumed) regular unbiased estimator $K : \mathbb{R} \to \mathbb{R}$ is MVUE amongst all *regular* unbiased estimators (upon applying the Cramèr-Rao bound). If in addition, it is also a complete sufficient statistic for the family $\{F_\theta,\ \theta \in \Theta\}$, then it is also MVUE (among all unbiased finite variantce estimators) by virtue of Lemma 3.1 (with $\tilde{g}(t) = t$).

7 The i.i.d. case

In many situations the data to be used for estimating the parameter θ is obtained by collecting i.i.d. samples from the underlying distribution. Formally, let $\{F_\theta,\ \theta \in \Theta\}$ denote the usual collection of probability distributions on $\mathbb{R}^k$. With positive integer n, let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ be *i.i.d.* $\mathbb{R}^k$-valued rvs, each distributed according to F_θ under $\mathbb{P}_\theta$. Thus, for each θ in Θ we have

$$\mathbb{P}_\theta[\boldsymbol{Y}_1 \in B_1, \ldots, \boldsymbol{Y}_n \in B_n] = \prod_{i=1}^{n} \mathbb{P}_\theta[\boldsymbol{Y}_i \in B_i], \qquad \begin{array}{l} B_i \in \mathcal{B}(\mathbb{R}^k), \\ i = 1, \ldots, n. \end{array}$$

Let $F_\theta^{(n)}$ denote the corresponding probability distributions on $\mathbb{R}^{nk}$, namely

$$\begin{aligned} F_\theta^{(n)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) &= \mathbb{P}_\theta[\boldsymbol{Y}_1 \le \boldsymbol{y}_1, \ldots, \boldsymbol{Y}_n \le \boldsymbol{y}_n] \\ &= = \prod_{i=1}^n \mathbb{P}_\theta[\boldsymbol{Y}_i \le \boldsymbol{y}_i] \\ &= = \prod_{i=1}^n F_\theta(\boldsymbol{y}_i), \qquad \begin{array}{c} \boldsymbol{y}_i \in \mathbb{R}^k \\ i = 1, \ldots, n \end{array} \end{aligned} \tag{10}$$

When $n \ge 2$ the family $\{F_\theta^{(n)},\ \theta \in \Theta\}$ is *never* complete even if the family $\{F_\theta,\ \theta \in \Theta\}$ is complete.

The following hereditary properties are easily shown.

1. If the family $\{F_\theta,\ \theta \in \Theta\}$ is absolutely continuous with respect to the distribution F on $\mathbb{R}^k$ with density functions $\{f_\theta,\ \theta \in \Theta\}$, then the family $\{F_\theta^{(n)},\ \theta \in \Theta\}$ is also absolutely continuous but with respect to the distribution $F^{(n)}$ on $\mathbb{R}^{nk}$ given by

$$F^{(n)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \prod_{i=1}^n F(\boldsymbol{y}_i), \qquad \begin{array}{c} \boldsymbol{y}_i \in \mathbb{R}^k \\ i = 1, \ldots, n \end{array}$$

For each θ in Θ, he corresponding density function $f_\theta^{(n)} : \mathbb{R}^{nk} \to \mathbb{R}_+$ is given by

$$f^{(n)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \prod_{i=1}^n f(\boldsymbol{y}_i), \qquad \begin{array}{c} \boldsymbol{y}_i \in \mathbb{R}^k \\ i = 1, \ldots, n. \end{array}$$

2. Assume the family $\{F_\theta,\ \theta \in \Theta\}$ to be an exponential family (with respect to F) with density functions of the form

$$f_\theta(\boldsymbol{y}) = C(\theta) q(\boldsymbol{y}) e^{Q(\theta)' K(\boldsymbol{y})} \quad F - \text{a.e.}$$

for every θ in Θ with Borel mappings $C : \Theta \to \mathbb{R}_+$, $Q : \Theta \to \mathbb{R}^q$, $q : \mathbb{R}^k \to \mathbb{R}_+$ and $K : \mathbb{R}^k \to \mathbb{R}^q$. Then, the family $\{F_\theta^{(n)},\ \theta \in \Theta\}$ is also an exponential family (with respect to $F^{(n)}$) with density functions of the form

$$f_\theta^{(n)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = C(\theta)^n q^{(n)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) e^{Q(\theta)' K^{(n)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)} \quad F^{(n)} - \text{a.e.}$$

for each θ in Θ, where

$$q^{(n)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \prod_{i=1}^{n} q(\boldsymbol{y}_i), \qquad \begin{array}{c} \boldsymbol{y}_i \in \mathbb{R}^k \\ i = 1, \ldots, n \end{array}$$

and

$$K^{(n)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) = \sum_{i=1}^{n} K(\boldsymbol{y}_i), \qquad \begin{array}{c} \boldsymbol{y}_i \in \mathbb{R}^k \\ i = 1, \ldots, n. \end{array}$$

3. Assuming **(CR1)**, if the family $\{F_\theta, \ \theta \in \Theta\}$ satisfies Conditions **(CR2)–(CR5)** (with respect to F), then the family $\{F_\theta^{(n)}, \ \theta \in \Theta\}$ also satisfies Conditions **(CR2)–(CR5)** (with respect to $F^{(n)}$), and the Fisher information matrices are related through the relation

$$M^{(n)}(\theta) = nM(\theta), \quad \theta \in \Theta.$$

8 Asymptotic theory – Types of estimators

We are often interested in situations where the parameter θ is estimated on the basis of multiple $\mathbb{R}^k$-valued samples, say $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ for n large. The most common situation is that when the incoming observations form a sequence $\{\boldsymbol{Y}_n, \ n = 1, 2, \ldots\}$ of i.i.d. $\mathbb{R}^k$-valued rvs (as described earlier). However, in some applications the variates $\{\boldsymbol{Y}_n, \ n = 1, 2, \ldots\}$ may be correlated, e.g., the rvs $\{\boldsymbol{Y}_n, \ n = 1, 2, \ldots\}$ form a Markov chain.

In general, for each $n = 1, 2, \ldots$, let $g_n : \mathbb{R}^{nk} \to \mathbb{R}^k$ be an estimator for θ on the basis of the $\mathbb{R}^k$-valued observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$. We shall write

$$\boldsymbol{Y}^{(n)} = \begin{pmatrix} \boldsymbol{Y}_1 \\ \vdots \\ \boldsymbol{Y}_n \end{pmatrix}, \quad n = 1, 2, \ldots$$

The estimators $\{g_n, \ n = 1, 2, \ldots\}$ are *(weakly) consistent* at θ (in Θ) if the rvs $\{g_n(\boldsymbol{Y}^{(n)}), \ n = 1, 2, \ldots\}$ converge in probability to θ under $\mathbb{P}_\theta$, i.e., for every $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_\theta \left[\|g_n(\boldsymbol{Y}^{(n)}) - \theta\| > \varepsilon \right]. = 0$$

The estimators $\{g_n, \ n = 1, 2, \ldots\}$ are *(strongly) consistent* at θ (in Θ) if the rvs $\{g_n(\boldsymbol{Y}^{(n)}), \ n = 1, 2, \ldots\}$ converge a.s. to θ under $\mathbb{P}_\theta$, i.e.,

$$\lim_{n\to\infty} g_n(\boldsymbol{Y}^{(n)}) = \theta \quad \mathbb{P}_\theta - \text{a.s.}$$

As expected, strong consistency implies (weak) consistency.

The estimators $\{g_n,\ n = 1, 2, \ldots\}$ are *asymptotically normal* at θ (in Θ) if there exists a $p \times p$ positive semi-definite matrix $\mathbf{\Sigma}(\theta)$ with the property that

$$\sqrt{n}\left(g_n(\boldsymbol{Y}^{(n)}) - \theta\right) \Longrightarrow_n \mathrm{N}(\mathbf{0}_p, \mathbf{\Sigma}(\theta)).$$

The estimators $\{g_n,\ n = 1, 2, \ldots\}$ are *asymptotically unbiased* at θ (in Θ) if for each $n = 1, 2, \ldots$, the estimator is a finite mean estimator and

$$\lim_{n\to\infty} \mathbb{E}_\theta\left[g_n(\boldsymbol{Y}^{(n)})\right] = \theta.$$

This is equivalent to

$$\lim_{n\to\infty} b_\theta(g_n) = \mathbf{0}_p.$$

Assume that for each $n = 1, 2, \ldots$, the family of distributions $\{F_\theta^{(n)},\ \theta \in \Theta\}$ satisfies the appropriate conditions **(CR2)–(CR5)**. The estimators $\{g_n,\ n = 1, 2, \ldots\}$ are *asymptotically efficient* at θ (in Θ) if

$$\lim_{n\to\infty}\left(\Sigma_\theta(g_n) - M^{(n)}(\theta)^{-1}\right) = \boldsymbol{O}_{p\times p}, \quad \theta \in \Theta$$

provided the Fisher information matrices $\{M^{(n)}(\theta),\ n = 1, 2, \ldots\}$ are invertible for each θ in Θ.

9 Maximum likelihood estimation methods

Assume **(CR2a)** to hold. A Borel mapping $g_{\mathrm{ML}} : \mathbb{R}^k \to \Theta$ is called a *maximum likelihood estimator* of θ on the basis of $\boldsymbol{Y}$ if

$$f_{g_{\mathrm{ML}}(\boldsymbol{y})}(\boldsymbol{y}) = \max\left(f_\theta(\boldsymbol{y}),\ \theta \in \Theta\right), \quad \boldsymbol{y} \in \mathbb{R}^k.$$

This definition implicitly assumes that at the observation point $\boldsymbol{y}$, the supremum

$$\sup\left(f_\theta(\boldsymbol{y}),\ \theta \in \Theta\right)$$

is indeed achieved at some point in Θ. Note that (i) maximum likelihood estimators may not exist or (ii) may not be unique. Often these problems are handled by altering the selection of the density functions $\{f_\theta,\ \theta \in \Theta\}$.

ML equation

Note that the maximum likelihood estimator of θ on the basis of $\boldsymbol{Y}$ can equivalently be defined by

$$\log f_{g_{\mathrm{ML}}(\boldsymbol{y})}(\boldsymbol{y}) = \max\left(\log f_\theta(\boldsymbol{y}),\ \theta \in \Theta\right), \quad \boldsymbol{y} \in \mathbb{R}^k$$

under the convention $\log 0 = -\infty$. This equation is known as the *maximum likelihood equation*.

This observation leads to the following characterization: Assume $\mathrm{Int}(\Theta)$ to be non-empty and that condition **(CR2)** holds. Also assume that condition **(CR3)** holds for all θ in $\mathrm{Int}(\Theta)$ (rather than for all θ in Θ). Then

$$\nabla_\theta \log f_\theta(\boldsymbol{y})\Big|_{\theta = g_{\mathrm{ML}}(\boldsymbol{y})} = \mathbf{0}_p \quad \boldsymbol{y} \in S$$

provided

$$g_{\mathrm{ML}}(\boldsymbol{y}) \in \mathrm{Int}(\Theta).$$

When a sufficient statistics exists, the ML estimates can always expressed in terms of it. This is a consequence of the Factorization Theorem.

Theorem 9.1 *Assume that Condition* **(CR2a)** *holds and that for each $\boldsymbol{y}$ in S, the ML estimate $g_{\mathrm{ML}}(\boldsymbol{y})$ exists. If the statistic $T : \mathbb{R}^k \to \mathbb{R}^q$ is sufficient for the family $\{F_\theta,\ \theta \in \Theta\}$, then there exists a Borel mapping $G_{\mathrm{ML}} : \mathbb{R}^q \to \Theta$ such that*

$$g_{\mathrm{ML}}(\boldsymbol{y}) = G_{\mathrm{ML}}(T(\boldsymbol{y})) \quad F - \text{a.e. on } S.$$

ML estimators and efficiency

There are relationships between efficiency and ML estimators.

Theorem 9.2 *Assume Conditions* **(CR1)–(CR5)** *to hold, and that for each θ in Θ, the Fisher information matrix $M(\theta)$ is invertible. Assume further that for each $\boldsymbol{y}$ in S, the ML estimate $g_{\mathrm{ML}}(\boldsymbol{y})$ exists. Then every regular estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ which achieves the generalized Cramér-Rao bound must necessarily satisfy the equality*

$$g(\boldsymbol{y}) = g_{\mathrm{ML}}(\boldsymbol{y}) + b_{g_{\mathrm{ML}}(\boldsymbol{y})}(g) \quad F - \text{a.e. on } S.$$

Corollary 9.1 *Under the assumptions of Theorem 9.2, if the regular estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is efficient, then it must necessarily be an ML estimator.*

Asymptotic theory for ML estimators

The result given next assumes the availability of i.i.d. samples

Theorem 9.3 *Assume Conditions* **(CR1)–(CR8)** *to hold. For each* $n = 1, 2, \ldots$, *assume that for each* $\boldsymbol{y}^{(n)}$ *in* S^n, *the ML estimate* $g_{n,\mathrm{ML}}(\boldsymbol{y}^{(n)})$ *exists. Then the following statements hold.*

(i) The ML estimators $\{g_{n,\mathrm{ML}},\ n = 1, 2, \ldots\}$ *are strongly consistent, i.e., for each* θ *in* Θ,

$$\lim_{n\to\infty} g_{n,\mathrm{ML}}(\boldsymbol{Y}^{(n)}) = \theta \quad \mathbb{P}_\theta - \text{ a.s.}$$

(ii) The ML estimators $\{g_{n,\mathrm{ML}},\ n = 1, 2, \ldots\}$ *are asymptotically normal, i.e., for each* θ *in* Θ,

$$\sqrt{n}\left(g_{n,\mathrm{ML}}(\boldsymbol{Y}^{(n)}) - \theta\right) \Longrightarrow_n \mathrm{N}(\mathbf{0}_p, M(\theta)^{-1})$$

under $\mathbb{P}_\theta$ *provided the Fisher information matrix* $M(\theta)$ *is invertible.*