ENEE 621
SPRING 2017
DETECTION AND ESTIMATION THEORY

THE PARAMETER ESTIMATION PROBLEM

Throughout, p, q and k are positive integers.

1 The basic setting

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 $\boldsymbol{Y}$ whose statistical description depends on θ.

The setting is alway understood as follows: Given some measurable space $(\Omega, \mathcal{F})$, 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

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

The following concrete construction is standard: Take $\Omega = \mathbb{R}^k$ and $\mathcal{F} = \mathcal{B}\left(\mathbb{R}^k\right)$, and define the rv $\boldsymbol{Y} : \Omega \to \mathbb{R}^k$ to be the identity mapping given by

$$\boldsymbol{Y}(\omega) = \omega, \quad \omega \in \mathbb{R}^k.$$

For each θ in Θ, the probability measure $\mathbb{P}_\theta$ on $\mathcal{F}$ is the unique probability measure induced by the probability distribution function F_θ through the requirement

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

We will often assume the following absolute continuity assumption on the family of probability distributions $\{F_\theta,\ \theta \in \Theta\}$:

Condition (**A**): For each θ in Θ, the probability distribution F_θ on $\mathbb{R}^k$ is absolutely continuous with respect to some distribution F on $\mathbb{R}^k$.

This is equivalent to requiring that for each θ in Θ, there exists a Borel mapping $f_\theta : \mathbb{R}^k \to \mathbb{R}_+$ such that

$$F_\theta(\boldsymbol{y}) = \int_{-\infty}^{\boldsymbol{y}} f_\theta(\boldsymbol{\eta}) dF(\boldsymbol{\eta}), \quad \boldsymbol{y} \in \mathbb{R}^k. \tag{1}$$

For mathematical reasons we require that the mapping

$$\Theta \times \mathbb{R}^k \to \mathbb{R}_+ : (\theta, \boldsymbol{y}) \to f_\theta(\boldsymbol{y})$$

be Borel measurable. This condition is satisfied in all applications of interest.

Many results and statistical concepts take a very pleasing from in the context of so-called exponential families.

Assume the family $\{F_\theta,\ \theta \in \Theta\}$ to satisfy Condition **(A)** with respect to the probability distribution F on $\mathbb{R}^k$. The family $\{F_\theta,\ \theta \in \Theta\}$ is said to be an *exponential* family (with respect to F) if the corresponding density functions $\{f_\theta,\ \theta \in \Theta\}$ are of the form

$$f_\theta(\boldsymbol{y}) = C(\theta) q(\boldsymbol{y}) e^{Q(\theta)' K(\boldsymbol{y})} \quad F - \text{a.e.} \tag{2}$$

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 $K : \mathbb{R}^k \to \mathbb{R}^q$.

For each θ in Θ, 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

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

2 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 *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 $T(\boldsymbol{Y})$ does *not* depend on θ. Several observations to keep in mind:

(i) The statistic $T : \mathbb{R}^k \to \mathbb{R}^k$ given by

$$T(\boldsymbol{y}) = \boldsymbol{y}, \quad \boldsymbol{y} \in \mathbb{R}^k$$

is always a sufficient statistic; we shall refer to it as the *trivial* sufficient statistic.

(ii) Consider a statistic $T : \mathbb{R}^k \to \mathbb{R}^q$. If it is sufficient for $\{F_\theta,\ \theta \in \Theta\}$, then the statistic $\widetilde{T} : \mathbb{R}^k \to \mathbb{R}^r$ given by

$$\widetilde{T}(\boldsymbol{y}) = g(T(\boldsymbol{y})), \quad \boldsymbol{y} \in \mathbb{R}^k$$

for some Borel mapping $g : \mathbb{R}^q \to \mathbb{R}^r$ is *not* necessarily sufficient – Just take

$$g(\boldsymbol{t}) = \boldsymbol{0}_r, \quad \boldsymbol{t} \in \mathbb{R}^q.$$

(iii) On the other hand, if for some Borel mapping $g : \mathbb{R}^q \to \mathbb{R}^r$, the statistic $\widetilde{T} : \mathbb{R}^k \to \mathbb{R}^r$ given above is sufficient for $\{F_\theta,\ \theta \in \Theta\}$, then the statistic $T : \mathbb{R}^k \to \mathbb{R}^q$ is necessarily sufficient for $\{F_\theta,\ \theta \in \Theta\}$.

Two statistics $T_1 : \mathbb{R}^k \to \mathbb{R}^{q_1}$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^{q_2}$ (with q_1 and q_2 possibly different) are said to be (essentially) equivalent under the family $\{F_\theta,\ \theta \in \Theta\}$ if there exists Borel mappings $g_{12} : \mathbb{R}^{q_1} \to \mathbb{R}^{q_2}$ and $g_{21} : \mathbb{R}^{q_2} \to \mathbb{R}^{q_1}$ such that

$$\mathbb{P}_\theta \left[T_2(\boldsymbol{Y}) = g_{21}(T_1(\boldsymbol{Y})\right] = 1, \quad \theta \in \Theta \tag{3}$$

and

$$\mathbb{P}_\theta \left[T_1(\boldsymbol{Y}) = g_{12}(T_2(\boldsymbol{Y})\right] = 1, \quad \theta \in \Theta \tag{4}$$

In many cases, $q_1 = q_2 = q$ and the mappings $g_{12}, g_{21} : \mathbb{R}^q \to \mathbb{R}^q$ can be taken to be bijections which are inverses of each other, say $g_{12} = g$ with bijective Borel mapping $g : \mathbb{R}^q \to \mathbb{R}^q$ and $g_{21} = g^{-1}$.

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

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

The next result gives a simple implication for complete families in terms of possible sufficient statistics.

Lemma 2.1 *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 $\boldsymbol{Y}$ in the sense that for each $i = 1, \ldots, k$, we have*

$$\mathbb{P}_\theta \left[Y_i = \mathbb{E}_\theta \left[Y_i | T(\boldsymbol{Y}) \right] \right] = 1, \quad \theta \in \Theta \tag{5}$$

Proof. We shall assume first that the finite mean condition

$$\mathbb{E}_\theta \left[|Y_i| \right] < \infty, \qquad \begin{array}{c} i = 1, \ldots, k \\ \theta \in \Theta. \end{array}$$

holds. It follows that for each $\boldsymbol{t}$ in $\mathbb{R}^q$, the conditional expectations

$$\mathbb{E}_\theta\left[Y_i | T(\boldsymbol{Y}) = \boldsymbol{t}\right], \qquad \begin{array}{c} i = 1, \ldots, k \\ \theta \in \Theta. \end{array}$$

are all well defined and finite.

Consider now a sufficient statistic $T : \mathbb{R}^k \to \mathbb{R}^q$ which is sufficient for $\{F_\theta,\ \theta \in \Theta\}$. Thus, for each $\boldsymbol{t}$ in $\mathbb{R}^q$, we conclude that

$$\mathbb{E}_\theta\left[Y_i | T(\boldsymbol{Y}) = \boldsymbol{t}\right] = \int_{\mathbb{R}^k} \eta_i d\gamma(\boldsymbol{\eta}, \boldsymbol{t}), \qquad \begin{array}{c} i = 1, \ldots, k \\ \theta \in \Theta. \end{array}$$

where the notation is the one appearing in the definition of sufficiency for the statistic $T : \mathbb{R}^k \to \mathbb{R}^q$.

We define the Borel mapping $h : \mathbb{R}^k \times \mathbb{R}^q \to \mathbb{R}^k$ componentwise by

$$h_i(\boldsymbol{y}; \boldsymbol{t}) \equiv y_i - \int_{\mathbb{R}^k} \eta_i d\gamma(\boldsymbol{\eta}, \boldsymbol{t}), \qquad \begin{array}{c} i = 1, \ldots, k \\ \boldsymbol{y} \in \mathbb{R}^k \\ \boldsymbol{t} \in \mathbb{R}^q \end{array}$$

Note that

$$h_i(\boldsymbol{Y}; T(\boldsymbol{Y})) = Y_i - \mathbb{E}_\theta\left[Y_i | T(\boldsymbol{Y})\right], \quad \theta \in \Theta$$

and by iterated conditioning, we can conclude that

$$\mathbb{E}_\theta\left[h_i(\boldsymbol{Y}; T(\boldsymbol{Y}))\right] = \mathbb{E}_\theta\left[Y_i\right] - \mathbb{E}_\theta\left[\mathbb{E}_\theta\left[Y_i | T(\boldsymbol{Y})\right]\right] = 0, \qquad \begin{array}{c} i = 1, \ldots, k \\ \theta \in \Theta \end{array}$$

It is also plain that

$$\mathbb{E}_\theta\left[|h_i(\boldsymbol{Y}; T(\boldsymbol{Y}))|\right] \leq 2\mathbb{E}_\theta\left[|Y_i|\right] < \infty, \qquad \begin{array}{c} i = 1, \ldots, k \\ \theta \in \Theta. \end{array}$$

For each $i = 1, \ldots .k$, consider the Borel mapping $\psi_i : \mathbb{R}^k \to \mathbb{R}$ given by

$$\psi_i(\boldsymbol{y}) \equiv h_i(\boldsymbol{y}, T(\boldsymbol{y})), \quad \boldsymbol{y} \in \mathbb{R}^k$$

The discussion so far implies

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

with

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

The completeness of the family $\{F_\theta,\ \theta \in \Theta\}$ now gives

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

and this establishes (5).

If the finite mean condition fails to hold, i.e., $\mathbb{E}_\theta\left[|Y_i|\right] = \infty$ for some $i = 1, \ldots, k$ and some θ in Θ, then we proceed as follows: Let $g : \mathbb{R} \to \mathbb{R}$ denote a strictly increasing bijection that maps $\mathbb{R}$ into $(-1, 1)$, i.e., $g(x) = \frac{2}{\pi} \arctan x$ for each x in $\mathbb{R}$. Then, define the $\mathbb{R}^k$-valued rv $\boldsymbol{Z}$ componentwise by

$$Z_i = g(Y_i), \quad i = 1, \ldots, k.$$

It is plain that $\mathbb{E}_\theta\left[|Z_i|\right] < \infty$ since $|Z_i| \leq 1$.

Write

$$G_\theta(\boldsymbol{z}) \equiv \mathbb{P}_\theta\left[\boldsymbol{Z} \leq \boldsymbol{z}\right], \quad \begin{array}{l} \boldsymbol{z} \in \mathbb{R}^k \\ \theta \in \Theta. \end{array} \tag{6}$$

It is also easy to see that the family $\{G_\theta,\ \theta \in \Theta\}$ is also a complete family. Next, if $T : \mathbb{R}^k \to \mathbb{R}^q$ is a sufficient statistic for $\{F_\theta,\ \theta \in \Theta\}$, then $\widetilde{T} : \mathbb{R}^k \to \mathbb{R}^q$ given by

$$\widetilde{T}(\boldsymbol{z}) \equiv T(g^{-1}(z_1), \ldots, g^{-1}(z_k)), \quad \boldsymbol{z} \in \mathbb{R}^k$$

is also a sufficient statistic for $\{G_\theta,\ \theta \in \Theta\}$: Indeed for each θ in Θ, we have

$$\begin{aligned} \mathbb{P}_\theta\left[\boldsymbol{Z} \in B \middle| \widetilde{T}(\boldsymbol{Z}) = \widetilde{\boldsymbol{t}}\right] &= \mathbb{P}_\theta\left[(g(Y_1), \ldots, g(Y_k)) \in B \middle| T(\boldsymbol{Y}) = \widetilde{\boldsymbol{t}}\right] \\ &= \int_{B_g} d\gamma(\boldsymbol{\eta}; \widetilde{\boldsymbol{t}}), \quad B \in \mathcal{B}(\mathbb{R}^q) \end{aligned} \tag{7}$$

where B_g is the Borel subset of $\mathbb{R}^k$ given by

$$B_g \equiv \left\{\boldsymbol{y} \in \mathbb{R}^k :\ (g(y_1), \ldots, g(y_k)) \in B\right\}.$$

By the first part of the proof it follows that

$$\mathbb{P}_\theta\left[Z_i = \mathbb{E}_\theta\left[Z_i | \widetilde{T}(\boldsymbol{Z})\right]\right] = 1, \quad \theta \in \Theta \tag{8}$$

Equivalently,

$$\mathbb{P}_\theta\left[Y_i = g^{-1}\left(\mathbb{E}_\theta\left[g(Y_i) | T(\boldsymbol{Y})\right]\right)\right] = 1, \quad \theta \in \Theta \tag{9}$$

■

3 Finite mean 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

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

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

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

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 3.1 *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*

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

Proof. With finite mean estimators $g_1, g_2 : \mathbb{R}^k \to \mathbb{R}^p$, introduce the Borel mapping $\psi : \mathbb{R}^k \to \mathbb{R}^p$ given by

$$\psi(\boldsymbol{y}) \equiv g_1(\boldsymbol{y}) - g_2(\boldsymbol{y}), \quad \boldsymbol{y} \in \mathbb{R}^k.$$

Fix $i = 1, \ldots, p$ and θ in Θ. The finite mean assumption implies

$$\mathbb{E}_\theta \left[|\psi_i(\boldsymbol{Y})| \right] \le \mathbb{E}_\theta \left[|g_{1,i}(\boldsymbol{Y})| \right] + \mathbb{E}_\theta \left[|g_{2,i}(\boldsymbol{Y})| \right] < \infty$$

while the fact that these estimators are unbiased yields

$$\mathbb{E}_\theta\left[\psi_i(\boldsymbol{Y})\right] = \mathbb{E}_\theta\left[g_{1,i}(\boldsymbol{Y})\right] - \mathbb{E}_\theta\left[g_{2,i}(\boldsymbol{Y})\right] = \theta_i - \theta_i = 0,$$

By the completeness of the the family $\{F_\theta,\ \theta \in \Theta\}$ we conclude

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

and the desired result is obtained. ■

4 Finite variance estimators

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

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

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

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

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

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

for any other finite variance unbiased estimator $g : \mathbb{R}^k \to \mathbb{R}^p$. The basic question of finding MVUEs is addressed next.

5 The Rao-Blackwell Theorem

A basic step in the search for MVUEs is provided by the Rao-Blackwell Theorem. But first an additional concept is needed.

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 $\{F_\theta,\ \theta \in \Theta\}$ with the property that the family $\{H_\theta,\ \theta \in \Theta\}$ of probability distributions on $\mathbb{R}^q$ is complete where

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

This notion has an easy and important consequence for finding MVUEs.

Lemma 5.1 *Let $T : \mathbb{R}^k \to \mathbb{R}^q$ be a complete sufficient statistic for $\{F_\theta,\ \theta \in \Theta\}$. 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 mean unbiased estimator for θ on the basis of $\boldsymbol{Y}$, then*

$$\mathbb{P}_\theta\left[\tilde{g}_1(T(\boldsymbol{Y}) = \tilde{g}_2(T(\boldsymbol{Y}))\right] = 1, \quad \theta \in \Theta. \tag{10}$$

Proof. Under the foregoing assumptions, we note that

$$\mathbb{E}_\theta\left[\tilde{g}_i(T(\boldsymbol{Y})\right] = \theta, \quad \begin{array}{c} i = 1, 2 \\ \theta \in \Theta \end{array}$$

whence $\mathbb{E}_\theta\left[\tilde{g}_1(T(\boldsymbol{Y}) - \tilde{g}_1(T(\boldsymbol{Y})\right] = 0$ for each θ in Θ. The conclusion (10) now follows by the complete sufficiency of the statistic $T : \mathbb{R}^k \to \mathbb{R}^q$. ■

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

Theorem 5.1 *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 $\widehat{g} : \mathbb{R}^q \to \mathbb{R}^p$ given by*

$$\widehat{g}(\boldsymbol{t}) = \int_{\mathbb{R}^k} g(\boldsymbol{y}) d\gamma(\boldsymbol{y}, \boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{R}^q \tag{11}$$

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) applied to the rv

$$\boldsymbol{v}' \left(g(\boldsymbol{Y}) - \theta \right) \left(g(\boldsymbol{Y}) - \theta \right)' \boldsymbol{v}, \quad \theta \in \Theta$$

(combined with the law of iterated conditioning) once it is observed that

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

for every θ in Θ.

6 Finding MVUEs

The basic idea relies on Lemma 5.1 and on the Rao-Blackwell Theorem: Start with a statistics $T : \mathbb{R}^k \to \mathbb{R}^q$ which is a sufficient statistic for $\{F_\theta, \ \theta \in \Theta\}$. By the Rao-Blackwell Theorem above, if there exists an unbiased estimator $g : \mathbb{R}^k \to \mathbb{R}^p$, then with the definition (11), the estimator $\widehat{g} \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is also an unbiased estimator for θ on the basis of $\boldsymbol{Y}$ with

$$\Sigma_\theta(\widehat{g} \circ T) \leq \Sigma_\theta(g), \quad \theta \in \Theta.$$

On the other hand, if $g_{\text{Other}} : \mathbb{R}^k \to \mathbb{R}^p$ is another arbitrary unbiased estimator, then define the mapping $\widehat{g}_{\text{Other}} : \mathbb{R}^k \to \mathbb{R}^q$ given by

$$\widehat{g}_{\text{Other}}(\boldsymbol{t}) = \int_{\mathbb{R}^k} g_{\text{Other}}(\boldsymbol{y}) d\gamma(\boldsymbol{y}, \boldsymbol{t}), \quad \boldsymbol{t} \in \mathbb{R}^q. \tag{12}$$

By the Rao-Blackwell Theorem the estimator $\widehat{g}_{\text{Other}} \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is also an unbiased estimator for θ on the basis of $\boldsymbol{Y}$ with

$$\Sigma_\theta(\widehat{g}_{\text{Other}} \circ T) \leq \Sigma_\theta(g_{\text{Other}}), \quad \theta \in \Theta.$$

Now, *if* the statistic $T : \mathbb{R}^k \to \mathbb{R}^q$ is a complete sufficient statistic for $\{F_\theta, \ \theta \in \Theta\}$, then Lemma 5.1 implies

$$\mathbb{P}_\theta \left[\widehat{g} \circ T(\boldsymbol{Y}) = \widehat{g}_{\text{Other}} \circ T(\boldsymbol{Y}) \right] = 1, \quad \theta \in \Theta,$$

whence

$$\Sigma_\theta(\widehat{g} \circ T) = \Sigma_\theta(\widehat{g}_{\text{Other}} \circ T), \quad \theta \in \Theta.$$

Therefore,

$$\begin{aligned} \Sigma_\theta(\widehat{g} \circ T) &= \Sigma_\theta(\widehat{g}_{\text{Other}} \circ T) \\ &\leq \Sigma_\theta(g_{\text{Other}}), \quad \theta \in \Theta \end{aligned} \tag{13}$$

and the estimator $\widehat{g} \circ T : \mathbb{R}^k \to \mathbb{R}^p$ is indeed MVUE since $g_{\text{Other}} : \mathbb{R}^k \to \mathbb{R}^p$ is an *arbitrary* unbiased estimator. These observations lead to the following strategy to finding MVUEs:

(i) Find a *complete* sufficient statistic $T : \mathbb{R}^k \to \mathbb{R}^q$ for $\{F_\theta, \ \theta \in \Theta\}$. In the context of exponential families, Theorem 8.2 can be invoked;

(ii) Find a finite variance *unbiased* estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ for θ on the basis of $\boldsymbol{Y}$. As there are no general procedure for doing so, this step often involves some guessing based on the structure of the problem. However, there are many situations where it it is possible to find rather easily a Borel mapping $\tilde{g} : \mathbb{R}^q \to \mathbb{R}^p$ such that $\tilde{g} \circ T$ is an unbiased finite variance estimator for θ on the basis of $\boldsymbol{Y}$;

(iii) From the estimator g obtained in (ii), generate the Borel mapping $\widehat{g} : \mathbb{R}^q \to \mathbb{R}^p$ as per the Rao-Blackwell Theorem.

As argued earlier, the estimator $\widehat{g} \circ T$ is MVUE by the uniqueness result of Lemma 5.1; this also implies that the essential uniqueness of the MVUE.

---

7 An example

Consider the situation where the observation $\boldsymbol{Y}$ is given componentwise by

$$Y_i = \mu a_i + N_i, \quad i = 1, \ldots, k$$

where the rvs $N_1, \ldots, N_k$ are i.i.d. zero mean rvs with variance $\sigma^2 > 0$, μ is a scaling factor and the amplitudes $a_1, \ldots, a_k$ are *known* non-zero constants – We shall assume $\sum_{i=1}^k a_i^2 > 0$ so that $a_i \neq 0$ for at least one index $i = 1, \ldots, k$. Under these assumptions,

$$\boldsymbol{Y} \sim \mathrm{N}\left(\mu \boldsymbol{a}, \sigma^2 \boldsymbol{I}_k\right)$$

with $\boldsymbol{a} \equiv (a_1, \ldots, a_k)'$. Therefore,

$$\begin{aligned} f_{\mu,\sigma^2}(\boldsymbol{y}) &= \prod_{i=1}^k \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \mu a_i)^2}{2\sigma^2}} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^k e^{-\frac{1}{2\sigma^2}\sum_{i=1}^k (y_i - \mu a_i)^2} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^k e^{-\frac{1}{2\sigma^2}\sum_{i=1}^k \left(y_i^2 - 2\mu a_i y_i + \mu^2 a_i^2\right)}, \quad \boldsymbol{y} \in \mathbb{R}^k \end{aligned} \tag{14}$$

Here $\mu \neq 0$ in order to avoid trivial cases of limited interest. With $\theta = \mu$ and $\Theta \subseteq \mathbb{R}$ (with σ^2 known), it is easy to check that

(i) The family $\{F_\theta,\ \theta \in \Theta\}$ is an exponential family with

$$Q(\theta) = \frac{\theta}{\sigma^2}, \quad \theta \in \mathbb{R}$$

and

$$K(\boldsymbol{y}) = \sum_{i=1}^n a_i y_i, \quad \boldsymbol{y} \in \mathbb{R}^k.$$

(ii) The statistic $K : \mathbb{R}^k \to \mathbb{R}$ is a complete sufficient statistic as soon as Θ contains a non-trivial interval; see Theorem 8.2.

(iii) Fix $i = 1, \ldots, k$. For each c in $\mathbb{R}$, the Borel mapping $g_{i,c} : \mathbb{R}^k \to \mathbb{R}$ given by

$$g_{i,c}(\boldsymbol{y}) = c y_i, \quad \boldsymbol{y} \in \mathbb{R}^k$$

yields

$$\mathbb{E}_\theta\left[g_{i,c}(\boldsymbol{Y})\right] = c\mathbb{E}_\theta\left[Y_i\right] = c\mu a_i.$$

Thus, the estimator $g_{i,c} : \mathbb{R}^k \to \mathbb{R}$ will be a finite variance unbiased estimator of μ on the basis $\boldsymbol{Y}$ if i and c are selected so that $a_i \neq 0$ and $ca_i = 1$.

(iv) Fix θ in $\mathbb{R}$. Having in mind to apply the Rao-Blackwell Theorem, we note that

$$\begin{aligned}
& \mathbb{E}_\theta\left[g_{i,c}(\boldsymbol{Y})\middle| K(\boldsymbol{Y})\right] \\
= \; & c\mathbb{E}_\theta\left[Y_i\middle| K(\boldsymbol{Y})\right] \\
= \; & c\mu a_i + c\mathrm{Cov}\left[Y_i, K(\boldsymbol{Y})\right]\mathrm{Cov}\left[K(\boldsymbol{Y})\right]^{-1}\left(\sum_{j=1}^k a_j\left(Y_j - \mu a_j\right)\right) \qquad (15)
\end{aligned}$$

with

$$\mathrm{Cov}\left[Y_i, K(\boldsymbol{Y})\right] = \sum_{j=1}^k a_j \mathrm{Cov}\left[Y_i, Y_j\right] = a_i \mathrm{Var}\left[Y_i\right] = a_i\sigma^2$$

and

$$\mathrm{Cov}\left[K(\boldsymbol{Y})\right] = \mathrm{Var}\left[\sum_{j=1}^k a_j Y_j\right] = \sum_{j=1}^k a_j^2\sigma^2.$$

Therefore,

$$\begin{aligned}
c\mathbb{E}_\theta\left[Y_i\middle| K(\boldsymbol{Y})\right] &= c\mu a_i + \frac{ca_i}{\sum_{j=1}^k a_j^2}\left(\sum_{j=1}^k a_j\left(Y_j - \mu a_j\right)\right) \\
&= \frac{ca_i}{\sum_{j=1}^k a_j^2}\cdot\sum_{j=1}^k a_j Y_j \\
&= \frac{1}{\sum_{j=1}^k a_j^2}\cdot\sum_{j=1}^k a_j Y_j \qquad (16)
\end{aligned}$$

(v) The calculations above show that here the estimator $\widehat{g} : \mathbb{R} \to \mathbb{R}$ is given by

$$\widehat{g}(t) = \frac{t}{\sum_{j=1}^k a_j^2}, \quad t \in \mathbb{R}. \qquad (17)$$

Regardless of θ in $\mathbb{R}$, it is plain that

$$\mathbb{E}_\theta\left[g_{i,c}(\boldsymbol{Y})\Big|K(\boldsymbol{Y})\right] = \widehat{g}(K(\boldsymbol{Y})) \quad \mathbb{P}_\theta\text{-}a.s.$$

with $\mathbb{E}_\theta\left[\widehat{g}(K(\boldsymbol{Y}))\right] = \mu$. It follows that the finite variance estimator $\widehat{g} \circ K$ is an unbiased estimator for μ on the basis of $\boldsymbol{Y}$, and is MVUE.

With $\theta = (\mu, \sigma^2)$ and $\Theta \subseteq \mathbb{R} \times (0, \infty)$, it is easy to check the following:

(i) As before the family $\{F_\theta,\ \theta \in \Theta\}$ is an exponential family but this time we have

$$Q(\theta) = \begin{bmatrix} \frac{\mu}{\sigma^2} \\ \\ -\frac{1}{2\sigma^2} \end{bmatrix}, \quad \theta = (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$$

and

$$K(\boldsymbol{y}) = \begin{bmatrix} K_\mu(\boldsymbol{y}) \\ \\ K_{\sigma^2}(\boldsymbol{y}) \end{bmatrix}, \quad \boldsymbol{y} \in \mathbb{R}^k$$

with

$$K_\mu(\boldsymbol{y}) = \sum_{i=1}^n a_i y_i$$

and

$$K_{\sigma^2}(\boldsymbol{y}) = \sum_{i=1}^n y_i^2.$$

(ii) By Theorem 8.2, the two-dimensional statistic $K : \mathbb{R}^k \to \mathbb{R}^2$ is a complete sufficient statistic as soon as the set

$$Q(\Theta) = \left\{ \frac{1}{2\sigma^2} \begin{bmatrix} 2\mu \\ \\ -1 \end{bmatrix}, \quad (\mu, \sigma^2) \in \Theta \right\}$$

contains a non-trivial rectangle. This will happen if both μ and σ^2 lie in intervals.

(iii) Obviously the estimator $\widehat{g} \circ K_\mu : \mathbb{R}^k \to \mathbb{R}$ where $\widehat{g} : \mathbb{R} \to \mathbb{R}$ is given by (17) is an unbiased finite variance estimator of μ on the basis of $\boldsymbol{Y}$. Applying the Rao-Blackwell Theorem to it (with complete sufficient two-dimensional statistic $K : \mathbb{R}^k \to \mathbb{R}^2$ we readily conclude that the estimator $\widehat{g} \circ K_\mu$ is still an MVUE in this new setting

(iv) We need to find an unbiased finite variance estimator of σ^2 on the basis of $\boldsymbol{Y}$. To do this, we note the following. Fix θ in Θ. We have

$$\begin{aligned}
\mathbb{E}_\theta\left[K_{\sigma^2}(\boldsymbol{Y})\right] &= \mathbb{E}_\theta\left[\sum_{i=1}^k Y_i^2\right] \\
&= \sum_{i=1}^k \left(\sigma^2 + \mu^2 a_i^2\right) \\
&= k\sigma^2 + \mu^2\left(\sum_{i=1}^k a_i^2\right) \qquad (18)
\end{aligned}$$

while

$$\begin{aligned}
\mathbb{E}_\theta\left[K_\mu(\boldsymbol{Y})^2\right] &= \mathbb{E}_\theta\left[\left(\sum_{i=1}^k a_i Y_i\right)^2\right] \\
&= \mathrm{Var}_\theta\left[\sum_{i=1}^k a_i Y_i\right] + \left(\mathbb{E}_\theta\left[K_\mu(\boldsymbol{Y})\right]\right)^2 \\
&= \sum_{i=1}^k \mathrm{Var}_\theta\left[a_i Y_i\right] + \left(\sum_{i=1}^k \mu a_i^2\right)^2 \\
&= \sigma^2\left(\sum_{i=1}^k a_i^2\right) + \mu^2\left(\sum_{i=1}^k a_i^2\right)^2 \\
&= \left(\sum_{i=1}^k a_i^2\right)\left(\sigma^2 + \mu^2\left(\sum_{i=1}^k a_i^2\right)\right) \qquad (19)
\end{aligned}$$

It follows that

$$\mathbb{E}_\theta\left[\frac{K_\mu(\boldsymbol{Y})^2}{\sum_{i=1}^k a_i^2}\right] = \sigma^2 + \mu^2\left(\sum_{i=1}^k a_i^2\right)$$

so that

$$\mathbb{E}_\theta\left[K_{\sigma^2}(\boldsymbol{Y})\right] - \mathbb{E}_\theta\left[\frac{K_\mu(\boldsymbol{Y})^2}{\sum_{i=1}^k a_i^2}\right] = (k-1)\sigma^2$$

whence

$$\mathbb{E}_\theta\left[\frac{1}{k-1}\cdot\left(K_{\sigma^2}(\boldsymbol{Y}) - \frac{K_\mu(\boldsymbol{Y})^2}{\sum_{i=1}^k a_i^2}\right)\right] = \sigma^2.$$

The estimator $g_{\mathrm{MVU}} : \mathbb{R}^k \to \mathbb{R}^2$ given by

$$g_{\mathrm{MVU}}(\boldsymbol{y}) \equiv \begin{bmatrix} \frac{K_\mu(\boldsymbol{y})}{\sum_{i=1}^k a_i^2} \\ \frac{1}{k-1} \cdot \left(K_{\sigma^2}(\boldsymbol{y}) - \frac{K_\mu(\boldsymbol{y})^2}{\sum_{i=1}^k a_i^2} \right) \end{bmatrix}, \quad \boldsymbol{y} \in \mathbb{R}^k$$

is MVUE.

8 Exponential families and sufficient statistics

An exponential family always admits at least one sufficient statistic.

Theorem 8.1 *Assume $\{F_\theta,\ \theta \in \Theta\}$ to be an exponential family (with respect to F) with representation (2). 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 8.2 *Assume $\{F_\theta,\ \theta \in \Theta\}$ to be an exponential family (with respect to F) with representation (2). 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.

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

$$\begin{aligned} \left| e^{\boldsymbol{u}' K(\boldsymbol{y})} \right| &= \left| \prod_{i=1}^q e^{u_i K_i(\boldsymbol{y})} \right| \\ &= \left| \prod_{i=1}^q e^{(\Re(u_i) + j\Im(u_i)) K_i(\boldsymbol{y})} \right| \\ &= \prod_{i=1}^q \left| e^{(\Re(u_i) + j\Im(u_i)) K_i(\boldsymbol{y})} \right| \\ &= \prod_{i=1}^q e^{\Re(u_i) K_i(\boldsymbol{y})} \end{aligned} \tag{20}$$

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 denote a q-dimensional rectangle contained in $Q(\Theta)$, say,

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

The arguments given above then show that on the superset $R^\star$ of R 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.

9 The Cramèr-Rao bounds – Assumptions

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 $F : \mathbb{R}^k \to \mathbb{R}_+$.[1] Thus, for each θ in Θ, there exists a Borel mapping $f_\theta : \mathbb{R}^k \to \mathbb{R}_+$ such that

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

[1] The is essentially Condition (**A**) introduced earlier.

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 conditions

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

hold. 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

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

Under Conditions **(CR1)–(CR4)**, define the *Fisher information matrix* $M(\theta)$ at parameter θ in 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,

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

---

10 The Cramèr-Rao bounds

The Cramèr-Rao bounds hold for the following class of finite variance 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 generalized Cramèr-Rao bound is given first

Theorem 10.1 *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*

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

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

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

The classical Cramèr-Rao bound holds for unbiased estimators, and arises as a simple corollary of Theorem 10.1.

Theorem 10.2 *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

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

all exist and are finite on S;

CR7 For each θ in Θ,

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

CR8 For each θ in Θ, the regularity conditions

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

hold. In the same manner as in **(CR5)**, 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 10.1 *Under Conditions* **(CR1)–(CR8)**, *the Fisher information matrix takes the form*

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

Proof. Fix $i,j=1,\ldots,p$ and θ in Θ. For each $\boldsymbol{y}$ in S, under **(CR3)** and **(CR6)** we note that

$$\begin{aligned}
\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f_\theta(\boldsymbol{y}) &= \frac{\partial}{\partial\theta_i}\left(\frac{\partial}{\partial\theta_j}\log f_\theta(\boldsymbol{y})\right)\\
&= \frac{\partial}{\partial\theta_i}\left(\frac{1}{f_\theta(\boldsymbol{y})}\frac{\partial}{\partial\theta_j}f_\theta(\boldsymbol{y})\right)\\
&= -\frac{1}{f_\theta(\boldsymbol{y})^2}\cdot\frac{\partial}{\partial\theta_i}f_\theta(\boldsymbol{y})\cdot\frac{\partial}{\partial\theta_j}f_\theta(\boldsymbol{y})+\frac{1}{f_\theta(\boldsymbol{y})}\cdot\frac{\partial}{\partial\theta_i}\left(\frac{\partial}{\partial\theta_j}f_\theta(\boldsymbol{y})\right)\\
&= -\frac{1}{f_\theta(\boldsymbol{y})^2}\cdot\frac{\partial}{\partial\theta_i}f_\theta(\boldsymbol{y})\cdot\frac{\partial}{\partial\theta_j}f_\theta(\boldsymbol{y})+\frac{1}{f_\theta(\boldsymbol{y})}\cdot\frac{\partial^2}{\partial\theta_i\partial\theta_j}f_\theta(\boldsymbol{y})\\
&= -\frac{\partial}{\partial\theta_i}\log f_\theta(\boldsymbol{y})\cdot\frac{\partial}{\partial\theta_j}\log f_\theta(\boldsymbol{y})+\frac{1}{f_\theta(\boldsymbol{y})}\cdot\frac{\partial^2}{\partial\theta_i\partial\theta_j}f_\theta(\boldsymbol{y}). \qquad (21)
\end{aligned}$$

Note that

$$\mathbb{E}_\theta\left[\frac{1}{f_\theta(\boldsymbol{Y})}\cdot\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}f_\theta(\boldsymbol{Y})\right|\right] = \int_S\left|\frac{\partial^2}{\partial\theta_i\partial\theta_j}f_\theta(\boldsymbol{y})\right|dF(\boldsymbol{y})<\infty$$

by virtue of **(CR7)**. Thus, taking expectations with respect to $\mathbb{P}_\theta$ we conclude that

$$\begin{aligned}
&\mathbb{E}_\theta\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f_\theta(\boldsymbol{Y})\right]\\
&= -\mathbb{E}_\theta\left[\frac{\partial}{\partial\theta_i}\log f_\theta(\boldsymbol{Y})\cdot\frac{\partial}{\partial\theta_j}\log f_\theta(\boldsymbol{Y})\right]+\mathbb{E}_\theta\left[\frac{1}{f_\theta(\boldsymbol{Y})}\cdot\frac{\partial^2}{\partial\theta_i\partial\theta_j}f_\theta(\boldsymbol{Y})\right]\\
&= -M_{ij}(\theta)+\mathbb{E}_\theta\left[\frac{1}{f_\theta(\boldsymbol{Y})}\cdot\frac{\partial^2}{\partial\theta_i\partial\theta_j}f_\theta(\boldsymbol{Y})\right]\\
&= -M_{ij}(\theta)+\int_S\frac{\partial^2}{\partial\theta_i\partial\theta_j}f_\theta(\boldsymbol{y})dF(\boldsymbol{y})\\
&= -M_{ij}(\theta)
\end{aligned}$$

as we invoke Condition **(CR8)**. ■

11 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

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

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

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

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

$$\begin{aligned} \mathrm{Cov}_\theta[U(\theta, \boldsymbol{Y})] &= \mathbb{E}_\theta \left[U(\theta, \boldsymbol{Y}) U(\theta, \boldsymbol{Y})'\right] \\ &= \Sigma_\theta(g) - b_\theta(g) b_\theta(g)' \\ &\quad - \left(\boldsymbol{I}_p + \nabla_\theta b_\theta(g)\right) M(\theta)^{-1} \left(\boldsymbol{I}_p + \nabla_\theta b_\theta(g)\right)'. \end{aligned} \tag{22}$$

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

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

a condition equivalent to

$$g(\boldsymbol{Y}) = \theta + b_\theta(g) + \left(\boldsymbol{I}_p + \nabla_\theta b_\theta(g)\right) M(\theta)^{-1} \nabla_\theta \log f_\theta(\boldsymbol{Y}) \quad \mathbb{P}_\theta\text{-}a.s.$$

12 Efficiency

As the Cramèr-Rao bounds provide a hard limit on the performance of regular estimators, it is natural to wonder whether the bounds can be achieved. To explore this issue we introduce the notion of 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 on Θ, namely

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

Efficiency is meaningless for unbiased estimators! Efficient estimators can be given a complete characterization.

Lemma 12.1 *Assume Conditions* **(CR1)–(CR5)** *to hold. A regular estimator* $g : \mathbb{R}^k \to \mathbb{R}^p$ *that is also efficient satisfies the relations*

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

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

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

on Θ *is an efficient regular estimator.*

Proof. Assume first that the estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is both regular and efficient. Being regular, the classical Cramèr-Rao bounds hold with

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

Being efficient, we have $\Sigma_\theta(g) = M(\theta)^{-1}$ for each θ in Θ, and by the second half of Theorem 10.2 we conclude that (23) holds.

Conversely, recall the Conditions **(CR1)–(CR5)**. If the estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ satisfies (24), then for each θ in Θ, we have

(25) $$g(\boldsymbol{Y}) - \theta = M(\theta)^{-1} \nabla_\theta \log f_\theta(\boldsymbol{Y}) \quad \mathbb{P}_\theta\text{-}a.s.$$

By Condition **(CR4)**, the estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is a finite variance estimator, with

$$\mathbb{E}_\theta \left[g(\boldsymbol{Y}) \right] = \theta + \mathbb{E}_\theta \left[M(\theta)^{-1} \nabla_\theta \log f_\theta(\boldsymbol{Y}) \right] = \theta \tag{26}$$

as we invoke Condition **(CR5)** – The estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is unbiased.

Next, fix $i, j = 1, \ldots, p$. Using (24) we get

$$\begin{aligned}
& g_j(\boldsymbol{y}) \left(\frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) \right) \\
&= \left(\theta_j + \left(M(\theta)^{-1} \nabla_\theta \log f_\theta(\boldsymbol{y}) \right)_j \right) \left(\frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) \right) \\
&= \theta_i \cdot \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) + \left(M(\theta)^{-1} \nabla_\theta \log f_\theta(\boldsymbol{y}) \right)_j \cdot \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) \\
&= \theta_i \cdot \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) + \sum_{\ell=1}^p M(\theta)^{-1}_{j\ell} \frac{\partial}{\partial \theta_\ell} \log f_\theta(\boldsymbol{y}) \cdot \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) \quad F - \text{a.e. on } S
\end{aligned}$$

It follows that

$$\begin{aligned}
&= \mathbb{E}_\theta \left[g_j(\boldsymbol{Y}) \left(\frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{Y}) \right) \right] \\
&= \theta_j \cdot \mathbb{E}_\theta \left[\frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{Y}) \right] + \sum_{\ell=1}^p M(\theta)^{-1}_{j\ell} \cdot \mathbb{E}_\theta \left[\frac{\partial}{\partial \theta_\ell} \log f_\theta(\boldsymbol{Y}) \cdot \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{Y}) \right] \\
&= \sum_{\ell=1}^p M(\theta)^{-1}_{j\ell} \cdot \mathbb{E}_\theta \left[\frac{\partial}{\partial \theta_\ell} \log f_\theta(\boldsymbol{Y}) \cdot \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{Y}) \right] \\
&= \sum_{\ell=1}^p M(\theta)^{-1}_{j\ell} M(\theta)_{\ell i} \\
&= \delta_{ji}
\end{aligned}$$

by virtue of Condition **(CR5)**. It follows that

$$\mathbb{E}_\theta \left[g(\boldsymbol{Y}) \left(\frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{Y}) \right) \right] = \boldsymbol{I}_p$$

whence

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

as we have shown that the estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ is unbiased. This establishes the regularity of the estimator.

Its efficiency is now immediate since for each θ in Θ: Indeed on account of (25) we have

$$\begin{aligned} \Sigma_\theta(g) &= \mathbb{E}_\theta\left[(g(\boldsymbol{Y})-\theta)\,(g(\boldsymbol{Y})-\theta)'\right] \\ &= \mathbb{E}_\theta\left[M(\theta)^{-1}\nabla_\theta \log f_\theta(\boldsymbol{Y})\left(M(\theta)^{-1}\nabla_\theta \log f_\theta(\boldsymbol{Y})\right)'\right] \\ &= M(\theta)^{-1}\mathbb{E}_\theta\left[\nabla_\theta \log f_\theta(\boldsymbol{Y})\nabla_\theta \log f_\theta(\boldsymbol{Y})'\right] M(\theta)^{-1} \\ &= M(\theta)^{-1}M(\theta)M(\theta)^{-1} \\ &= M(\theta)^{-1} \qquad (27) \end{aligned}$$

since $M(\theta)$ is a symmetric matrix. ■

As an immediate corollary we have the following.

Corollary 12.1 *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.*

It should be pointed out that efficiency may lead to awkward estimators: For instance assume that the observation Y is given by

$$Y = \sqrt{\theta} \cdot X + N$$

where the rvs X and N are i.i.d. Gaussian rvs with zero mean and unit variance. Here $\theta > 0$ and unknown. In other words, F_θ is a Gaussian distribution with zero mean and variance $1+\theta$. It is easily checked that the estimator $g^\star : \mathbb{R} \to \mathbb{R}$ given by

$$g^\star(y) = y^2 - 1, \quad y \in \mathbb{R}$$

is a regular estimator for θ on the basis of Y.

13 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 **(CR2)–(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, upon assuming the existence of the various derivatives, we observe that

$$(28)\quad \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) = \frac{\partial}{\partial \theta_i} \log C(\theta) + \left(\frac{\partial}{\partial \theta_i} Q(\theta)\right)' K(\boldsymbol{y}), \qquad \begin{array}{c} i = 1, \ldots, p \\ \boldsymbol{y} \in S. \end{array}$$

Therefore, **(CR3)** holds when the mappings $C : \Theta \to \mathbb{R}_+$ and $Q : \Theta \to \mathbb{R}^q$ are differentiable. It follows that **(CR4)** is now equivalent to

$$(29)\qquad \mathbb{E}_\theta \left[|K_\ell(\boldsymbol{Y})|^2\right] < \infty, \quad \ell = 1, \ldots, p.$$

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

$$(30)\qquad \left(\frac{\partial}{\partial \theta_i} Q(\theta)\right)' \mathbb{E}_\theta \left[K(\boldsymbol{Y})\right] = -\frac{\partial}{\partial \theta_i} \log C(\theta), \quad \begin{array}{c} \theta \in \Theta \\ i = 1, \ldots, p. \end{array}$$

Combining (28) and (30) we get

$$(31)\quad \frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{y}) = \left(\frac{\partial}{\partial \theta_i} Q(\theta)\right)' \left(K(\boldsymbol{y}) - \mathbb{E}_\theta \left[K(\boldsymbol{Y})\right]\right), \qquad \begin{array}{c} i = 1, \ldots, p \\ \boldsymbol{y} \in S. \end{array}$$

It is now straightforward to see that the Fisher information matrix is given entrywise by the expressions

$$(32)\quad M_{ij}(\theta) = \left(\frac{\partial}{\partial \theta_i} Q(\theta)\right)' \mathrm{Cov}_\theta \left[K(\boldsymbol{Y})\right] \left(\frac{\partial}{\partial \theta_j} Q(\theta)\right), \qquad \begin{array}{c} i, j = 1, \ldots, p \\ \theta \in \Theta \end{array}$$

The regularity of a finite variance estimator $g : \mathbb{R}^k \to \mathbb{R}^p$ can be expressed as follows: Fix θ in Θ and $i = 1, \ldots, p$: Using Condition **(CR5)** and (31) we get

$$\begin{aligned}
& \mathbb{E}_\theta \left[g(\boldsymbol{Y}) \left(\frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{Y}) \right) \right] \\
= \; & \mathbb{E}_\theta \left[\left(g(\boldsymbol{Y}) - \mathbb{E}_\theta \left[g(\boldsymbol{Y}) \right] \right) \left(\frac{\partial}{\partial \theta_i} \log f_\theta(\boldsymbol{Y}) \right) \right] \\
= \; & \mathbb{E}_\theta \left[\left(g(\boldsymbol{Y}) - \mathbb{E}_\theta \left[g(\boldsymbol{Y}) \right] \right) \left(\frac{\partial}{\partial \theta_i} Q(\theta) \right)' \left(K(\boldsymbol{Y}) - \mathbb{E}_\theta \left[K(\boldsymbol{Y}) \right] \right) \right] \\
= \; & \mathbb{E}_\theta \left[\left(g(\boldsymbol{Y}) - \mathbb{E}_\theta \left[g(\boldsymbol{Y}) \right] \right) \left(K(\boldsymbol{Y}) - \mathbb{E}_\theta \left[K(\boldsymbol{Y}) \right] \right)' \right] \left(\frac{\partial}{\partial \theta_i} Q(\theta) \right)' \\
= \; & \mathrm{Cov}_\theta \left[g(\boldsymbol{Y}), K(\boldsymbol{Y}) \right] \left(\frac{\partial}{\partial \theta_i} Q(\theta) \right)'
\end{aligned}$$

and regularity can be expressed as

$$\frac{\partial}{\partial \theta_i} \mathbb{E}_\theta \left[g(\boldsymbol{Y}) \right] = \mathrm{Cov}_\theta \left[g(\boldsymbol{Y}), K(\boldsymbol{Y}) \right] \left(\frac{\partial}{\partial \theta_i} Q(\theta) \right)', \qquad \begin{array}{c} i = 1, \ldots, p \\ \theta \in \Theta \end{array}$$

As we turn to the additional conditions **(CR6)-(CR8)**, we see from (28) that

$$\begin{aligned}
& \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f_\theta(\boldsymbol{y}) \\
= \; & \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log C(\theta) + \left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} Q(\theta) \right)' K(\boldsymbol{y}), \qquad \begin{array}{c} i, j = 1, \ldots, p \\ \boldsymbol{y} \in S. \end{array}
\end{aligned} \tag{33}$$

upon assuming the existence of the various additional derivatives.

Therefore, **(CR6)** holds when the mappings $C : \Theta \to \mathbb{R}_+$ and $Q : \Theta \to \mathbb{R}^q$ are twice differentiable, and **(CR7)** holds as soon as

$$\int_S |K_\ell(\boldsymbol{y})| dF(\boldsymbol{y}) < \infty, \quad \ell = 1, \ldots, p$$

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. Using the condition for regularity given above (with $g = K$), we see that this estimator will be regular if

$$\frac{d}{d\theta} \mathbb{E}_\theta \left[K(Y) \right] = \frac{d}{d\theta} Q(\theta) \cdot \mathrm{Var}_\theta \left[K(Y) \right], \quad \theta \in \Theta. \tag{34}$$

Using the earlier calculations, we note that

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

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

and

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

It is now easy to check that

$$\left(1 + \frac{d}{d\theta} b_\theta(K)\right) M(\theta)^{-1} \cdot \frac{\partial}{\partial\theta}\log f_\theta(y) = K(y) - \mathbb{E}_\theta\left[K(Y)\right], \quad y \in S$$

so that

$$\begin{aligned} &K(y) - \theta \\ &= \; b_\theta(K) + \left(1 + \frac{d}{d\theta} b_\theta(K)\right) M(\theta)^{-1} \frac{\partial}{\partial\theta}\log f_\theta(y) \quad y \in S \end{aligned} \tag{35}$$

for every θ in Θ. Invoking Theorem 10.1 we conclude that the estimator $K : \mathbb{R} \to \mathbb{R}$ satisfies the Cramèr-Rao bound with equality.

If the estimator $K : \mathbb{R} \to \mathbb{R}$ is also *unbiased* , i.e.,

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

the condition (34) for its regularity now reads

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

whence

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

The estimator $K : \mathbb{R} \to \mathbb{R}$ obviously 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, $K : \mathbb{R} \to \mathbb{R}$ is also a complete sufficient statistic for the family $\{F_\theta,\ \theta \in \Theta\}$, then it is also MVUE (among all unbiased finite variance estimators) by by the discussion in Section 6.

14 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

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

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

$$\begin{aligned} F_\theta^{(n)}(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n) &= \mathbb{P}_\theta[\boldsymbol{Y}_1 \leq \boldsymbol{y}_1, \ldots, \boldsymbol{Y}_n \leq \boldsymbol{y}_n] \\ &= \prod_{i=1}^n \mathbb{P}_\theta[\boldsymbol{Y}_i \leq \boldsymbol{y}_i] \\ &= \prod_{i=1}^n F_\theta(\boldsymbol{y}_i), \qquad \begin{array}{c} \boldsymbol{y}_i \in \mathbb{R}^k \\ i = 1, \ldots, n \end{array} \end{aligned} \tag{36}$$

When $n \geq 2$ 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
$$M^{(n)}(\theta) = n M(\theta), \quad \theta \in \Theta.$$

15 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 $\varepsilon > 0$,

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

A stronger notion arises by requiring that convergence takes place in the almost sure sense, rather than convergence in probability.

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. In many cases, consistency is often associated with the Law of Large Numbers. This rises a natural question as to whether the underlying convergence admits rate of convergence which is characterized by a Central Limit-like Theorem. Here is one way to formalize this notion.

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 $\boldsymbol{\Sigma}(\theta)$ with the property that

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

Lack of bias in the sense that

$$\mathbb{E}_\theta \left[g_n(\boldsymbol{Y}^{(n)}) \right] = \theta, \qquad \begin{matrix} \theta \in \Theta \\ n = 1, 2, \ldots \end{matrix}$$

may not always be possible to achieve. Instead we often settle for an asymptotic notion of unbiasedness.

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.

---

16 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 $\{f_\theta,\ \theta \in \Theta\}$.

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 16.1 *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*

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

There are relationships between efficiency and ML estimators.

Theorem 16.2 *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 16.1 *Under the assumptions of Theorem 16.2, if the regular estimator* $g : \mathbb{R}^k \to \mathbb{R}^p$ *is efficient, then it must necessarily be an ML estimator.*

There exists a rich asymptotic theory for ML estimators. The result given next assumes the availability of i.i.d. samples

---

Theorem 16.3 *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.*
