# Theory of point estimation

Jana Jurečková
*Charles University in Prague*

2010

# Contents

# Chapter 1

# Basic concepts

## 1.1 Loss function and risk

Let $\mathbf{X}$ be an observable random vector, $\mathbf{X} \in \mathcal{X}$ and $\mathcal{P} = \{P_\theta, \theta \in \mathbf{\Theta}\}$ be a system of probability distributions, indexed by an unobservable parameter $\theta$. We want to estimate the function $g(\theta) : \mathbf{\Theta} \mapsto \mathbb{R}^1$. The observed values $\mathbf{x}$ of $\mathbf{X}$ create the data.

An estimator of $g(\theta)$ is a function $T(\mathbf{X}) : \mathcal{X} \mapsto \mathbb{R}^1$. The loss incurred when we estimate $g(\theta)$ by $t$ is measured by the loss function $L(\theta, t)$ which should satisfy

$$L(\theta, t) \geq 0 \qquad \forall \theta, t;$$

$$L(\theta, g(\theta)) = 0 \quad \forall \theta.$$

The quality of estimator $T$ is measured by the risk function

$$R(\theta, T) = \mathbb{E}_\theta \; L(\theta, T(\mathbf{X})).$$

We wish to get the uniformly best estimator $T$, which satisfies

$$R(\theta, T) = \min \quad \text{with respect to} \;\; T, \;\; \text{uniformly in} \;\; \theta \in \mathbf{\Theta}.$$

Such estimator exists only in special cases; if it does not exist, we minimize the risk only over a subclass of estimators, e.g.

- unbiased estimators: bias $= \mathbb{E}_\theta T(\mathbf{X}) - g(\theta) = 0$
- median unbiased estimators: $P_\theta(T(\mathbf{X}) < g(\theta)) = P_\theta(T(\mathbf{X}) > g(\theta))$.
- If $X \sim F(x - \theta)$ (shift parameter) and $L(\theta, t) = L(|\theta - t|)$, then we consider the equivariant estimators satisfying $T(X_1 + c, \ldots, X_n + c) = T(\mathbf{X}) + c$.

Other possibilities:

- Instead of minimizing the risk uniformly over $\theta \in \mathbf{\Theta}$, we can minimize

  $$\int_{\mathbf{\Theta}} R(\theta, T) w(\theta) d\theta = \min \quad \text{over } T \;\; \text{with respect to the weight function} \;\; w.$$

  Such estimator is called the formal Bayesian estimator with the (generalized) prior density $w(\theta)$
- or $\sup_{\theta \in \mathbf{\Theta}} R(\theta, T) = min$ (minimax estimator).

## 1.2 Convex loss function

Convex function: $\phi(\lambda x + (1-\lambda)y) \le \lambda\phi(x) + (1-\lambda)\phi(y)$
strictly convex: $\phi(\lambda x + (1-\lambda)y) < \lambda\phi(x) + (1-\lambda)\phi(y), x \ne y$.
If $\phi$ is convex in $I = (a,b)$ and $t_0 \in I$ is fixed, then there exists a straight line $y = L(x) = c(t - t_0) + \phi(t_0)$ coming through the point $[t_0, \phi(t_0)]$ such that $L(x) \le \phi(x) \ \ \forall x \in I$.

**Theorem 1.2.1** *Jensen inequality. If $\phi$ is convex on an open interval $I$ and the random variable $X$ satisfies*

$$P(X \in I) = 1 \quad \text{and} \quad |\mathbb{E}\ X| < \infty,$$

*then*

$$\phi(\mathbb{E}\ X) \le \mathbb{E}\ \phi(X). \tag{1.2.1}$$

*If $\phi$ is strictly convex and $X$ is not constant with probability 1, then (1.2.1) holds as a sharp inequality.*

**Proof.** Put $t_0 = \mathbb{E}\ X$. Let $L(x)$ is the straight line such $L(\mathbf{x}) \le \phi(x)$ coming through point $[t_0, \phi(t_0)]$. Then

$$E\ \phi(X) \ge \mathbb{E}\ L(X) = \mathbb{E}[c(X - \mathbb{E}\ X)] + \phi(\mathbb{E}\ X) = \phi(X).$$

If $\phi$ is strictly convex, then the line touches $\phi$ at $t_0$ only, otherwise $L(x) < \phi(x)$. □

**Definition 1.2.1** *The statistic $S : \mathcal{X} \mapsto \mathcal{S}$ is called sufficient for the system $\mathcal{P}$ if there is a version of the conditional distribution $P_\theta(\mathbf{X} \in \mathbf{A} \mid S = s) = \lambda_s(\mathbf{A})$ independent of $\theta$.*

**Theorem 1.2.2** *Rao-Blackwell. Let $\mathbf{X}$ be an observable random vector with distribution $P_\theta \in \mathcal{P} = \{P_\vartheta : \ \vartheta \in \mathbf{\Theta}\}$. Let $S$ be a sufficient statistic for $\mathcal{P}$. Consider the estimation problem with a strictly convex loss function $L(\theta, t)$. Let $T$ be an estimator of $g(\theta)$ with finite expectation and risk, i.e. $R(\theta, T) = \mathbb{E}_\theta\ L(\theta, T(\mathbf{X})) < \infty \ \ \forall\theta$. Denote*

$$T^*(S) = \mathbb{E}\{T(\mathbf{X})|S(\mathbf{X}) = s\}.$$

*Then $T^*(S(\mathbf{X}))$ is an estimator satisfying $R(\theta, T^*) < R(\theta, T)$, unless $T(\mathbf{X}) = T^*(S(\mathbf{X}))$ with probability 1.*

**Proof.** Because $S$ is sufficient, $T^*(S)$ does not depend on $\theta$ and thus it is an estimator. Put $\phi(t) = L(\theta, t)$. Then

$$\phi(T^*(s)) = L(\theta, T^*(s)) = L[\theta, \mathbb{E}(T|S = s)] = \phi(\mathbb{E}(T|s))$$

$$< \mathbb{E}(\phi(T)|S = s) = \mathbb{E}[L(\theta, T(\mathbf{X}))|S = s]$$

unless $T(\mathbf{X}) = T^*(S(\mathbf{X}))$ with probability 1; hence

$$R(\theta, T^*) = \mathbb{E}_\theta L[\theta, T^*(S(\mathbf{X}))] < \mathbb{E}_\theta L(\theta, T(\mathbf{X})).$$

□

**Remark 1.2.1** *If $L(\theta, t)$ is convex, but not strictly, then Theorem 1.2.2 holds with an unsharp inequality.*

**Definition 1.2.2** *(Admissibility). The estimator $T$ is called inadmissible if there exists another estimator $T'$ dominating $T$, i.e. such that*

$$R(\theta, T') \leq R(\theta, T), \quad \forall\theta, \quad \text{with a sharp inequality at least for one } \theta. \tag{1.2.2}$$

*The estimator $T$ is called admissible with respect to the loss $L(\theta, t)$, if there is no estimator $T'$ satisfying (1.2.2).*

If $L(\theta, t)$ is strictly convex, and an admissible estimator exists, then it is uniquely determined. More precisely,

**Theorem 1.2.3** *Let $L(\theta, t)$ be strictly convex and $T$ be an admissible estimator of $g(\theta)$. If $T'$ is another estimator with the same risk and $T$, i.e. $R(\theta, T) = R(\theta, T')$ $\forall\theta$, then $T(\mathbf{X}) = T'(\mathbf{X})$ with probability 1.*

**Proof.** Put $T^* = \frac{1}{2}(T + T')$. Then

$$R(\theta, T^*) < \frac{1}{2}[R(\theta, T) + R(\theta, T')] = R(\theta, T) \ \ \forall\theta,$$

unless $T = T'$ with probability 1. But this contradicts with the admissibility of $T$. □

## 1.3 Estimation of vector function

The situation is analogous for the estimation of the vector function $\mathbf{g}(\theta) = (g_1(\theta), \ldots, g_k(\theta))$. Its estimator $\mathbf{T}(\mathbf{X})$ is also a $k$-dimensional vector. The function $\phi : E \mapsto \mathbb{R}^1$ where $E$ is a convex set, is called convex, if

$$\phi(\lambda\mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda\phi(\mathbf{x}_1) + (1-\lambda)\phi(\mathbf{x}_2) \ \ \forall\mathbf{x}_1, \mathbf{x}_2 \in E \ \text{ and } \ 0 < \lambda < 1.$$

If $\phi$ is twice differentiable in $E$, then $\phi$ is convex [strictly convex] iff the Hessian matrix

$$\mathbf{H} = \left[\frac{\partial^2\phi(x_1, \ldots, x_k)}{\partial x_i \partial x_j}\right]_{i,j=1,\ldots,k}$$

is positively semidefinite [positively definite].

If $\phi$ is convex defined on an open convex set $E \subset \mathbb{R}^k$, then to every fixed point $\mathbf{t}^0$ there exists a hyperplane

$$y = L(\mathbf{x}) = \phi(\mathbf{t}^0) + \sum_{i=1}^{k} c_i(x_i - t_i^0)$$

going through the point $(\mathbf{t}^0, \phi(\mathbf{t}^0))$ and satisfying $L(\mathbf{x}) \leq \phi(\mathbf{x})$ $\forall\mathbf{x} \in E$.

If $\mathbf{X}$ is a random vector such that $P(\mathbf{X} \in A) = 1$ for an open convex set $A$ and $\mathbb{E}\mathbf{X}$ exists, then $\mathbb{E}\mathbf{X} \in A$.

# Chapter 2

# Unbiased estimates

## 2.1 Uniformly best unbiased estimate

$T(\mathbf{X})$ is unbiased estimate of $g(\theta)$, if $\mathbb{E}(T(\mathbf{X})) = g(\theta) \ \ \forall\theta \in \mathbf{\Theta}$.

**Example:** *Unbised estimates need not exist. Let $X$ have binomial distribution $B(n, p)$ and let $g(p) = \frac{1}{p}$. If $T$ is an unbiased estimate of $g(p)$, then*

$$\sum_{i=1}^{n} T(i) \binom{n}{i} p^i (1-p)^{n-i} = \frac{1}{p} \quad \forall p \in (0, 1). \tag{2.1.1}$$

*But if $p \downarrow 0$, then the left-hand side of (2.1.1) $\to T(0)$, while the right-hand side $\to \infty$, what is a contradiction with the unbiasedness.*

The function $g(\theta)$ is called estimable, if there exists at least one unbiased estimate of $g(\theta)$.

**Lemma 2.1.1** *[Structure of the class of unbiased estimates.] If $T_0$ is an unbiased estimate of $g(\theta)$, then every unbiased estimate $T$ of $g(\theta)$ can be written in the form $T = T_0 - U$, where $U$ is an unbiased estimate of zero, i.e. such that $\mathbb{E}_\theta U = 0 \ \ \forall\theta \in \mathbf{\Theta}$.*

**Proof.** If $T_0$ is unbiased, then $T_0 - U$ is unbiased $\forall U$. If $T$ is any unbiased estimate, then $T_0 - T = U$ is an unbiased estimate of zero and $T = T_0 - U$. □

Specifically, consider the quadratic loss $L(\theta, t) = (t - g(\theta))^2$. Then the risk of an unbiased estimate $T$ is its variance:

$$R(\theta, T) = \mathbb{E}_\theta(T - g(\theta))^2 = \mathrm{var}_\theta \mathrm{T}(\mathbf{X}).$$

If $T^0$ minimizes $\mathrm{var}_\theta \mathrm{T}(\mathbf{X}) \ \ \forall\theta$ among all unbiased estimates of $g(\theta)$, then it is called best minimum variance estimate (BMVE) of $g(\theta)$.

Denote $\mathbf{\Delta}$ the set of all unbiased estimates of $g(\theta)$ satisfying $\mathbb{E}_\theta T^2 < \infty \ \ \forall\theta \in \mathbf{\Theta}$. Let $\mathcal{U}$ be the set of all unbiased estimates of 0 which belong to $\mathbf{\Delta}$.

**Theorem 2.1.1** *Let $\mathbf{X}$ have distribution $P_\theta$, $\theta \in \mathbf{\Theta}$, and let $T \in \mathbf{\Delta}$. Then $T$ is BMVE of its expected value $g(\theta)$ if and only if*

$$\mathbb{E}_\theta[T(\mathbf{X}).U(\mathbf{X})] = 0 \ \ \forall U \in \mathcal{U} \ \ \textit{and} \ \ \ \forall\theta \in \mathbf{\Theta}.$$

**Proof.**

**(i)** Necessity: Let $T$ be BMVE, $E_\theta T = g(\theta)$. Let $U \in \mathcal{U},\ U \neq 0$. Put $T' = T + \lambda U,\ \lambda \in \mathbb{R}^1$. Then

$$E_\theta T'(\mathbf{X}) = E_\theta T(\mathbf{X}) = g(\theta)\ \ \forall\theta$$

$$\Rightarrow \mathrm{var}_\theta \mathrm{T}'(\mathbf{X}) \geq \mathrm{var}_\theta \mathrm{T}(\mathbf{X})\ \ \forall\lambda \Rightarrow \mathrm{E}_\theta(\mathrm{T}')^2 \geq \mathrm{E}_\theta \mathrm{T}^2,$$

$$\text{thus}\ \ E_\theta T^2 + \lambda^2 E_\theta U^2 + 2\lambda E_\theta(T.U) \geq E_\theta T^2,\ \ \text{and}$$

$$\lambda^2 E_\theta U^2 + 2\lambda E_\theta(T.U) \geq 0 \quad \forall\lambda. \tag{2.1.2}$$

The roots of the quadratic equation $\lambda^2 E_\theta U^2 + 2\lambda E_\theta(T.U) = 0$ are

$$\lambda = 0\ \ \text{and}\quad \lambda = -\frac{2\mathrm{cov}_\theta(\mathrm{T},\mathrm{U})}{\mathrm{varU}},$$

hence the quadratic function can be negative unless $\mathrm{cov}_\theta(\mathrm{T},\mathrm{U}) = 0$.

**(ii)** Sufficiency: Let $E_\theta(T.U) = 0\quad \forall U \in \mathcal{U}$ and let $T'$ be an unbiased estimate of $g(\theta)$. If $\mathrm{var}_\theta \mathrm{T}' = \infty$, the $T'$ cannot be better than $T$. Let $\mathrm{var}_\theta \mathrm{T}' < \infty$. Then $\mathrm{var}_\theta(\mathrm{T}-\mathrm{T}') < \infty$ and $T - T' \in \mathcal{U}$, thus

$$E_\theta(T(T-T')) = 0 \Rightarrow E_\theta T^2 = E_\theta(T.T' \Rightarrow E_\theta T^2 - g^2(\theta) = E_\theta(T.T') - g^2(\theta)$$

$$\Rightarrow \mathrm{var}_\theta \mathrm{T} = \mathrm{cov}_\theta(\mathrm{T},\mathrm{T}') \Rightarrow 0 \leq \mathrm{var}_\theta(\mathrm{T}-\mathrm{T}') = \mathrm{var}_\theta \mathrm{T} + \mathrm{var}_\theta \mathrm{T}' - 2\mathrm{var}_\theta \mathrm{T}$$

$$\Rightarrow \mathrm{var}_\theta \mathrm{T} \leq \mathrm{var}_\theta \mathrm{T}'.$$

□

**Definition 2.1.1** *The statistic $S(\mathbf{X})$ is called complete for the system of distributions $\mathcal{P} = \{P_\theta,\ \theta \in \mathbf{\Theta}\}$ if, for any function $h(S)$*

$$E_\theta h(S(\mathbf{X})) = 0 \quad \forall\theta\ \ \Rightarrow h(S(\mathbf{X})) = 0\ a.s.[P_\theta],\ \forall\theta \in \mathbf{\Theta}.$$

**Theorem 2.1.2** *Let $X$ follow distribution $P_\theta \in \mathcal{P}$ and let $S$ be a complete and sufficient statistic for $\mathcal{P}$. Then every estimable function $g(\theta)$ has one and only one unbiased estimate, which is a function of $S$.*

**Proof.** Let $T$ be an unbiased estimate of $g(\theta)$. Then $T^*(S(\mathbf{X})) = E(T(\mathbf{X})|S(\mathbf{X}))$ is an unbiased estimate which is a function of $S$.

Let $T_1(S)$ and $T_2(S)$ be two unbiased estimates of $g(\theta)$. Then $E_\theta(T_1 - T_2) = 0\quad \forall\theta$, and because $S$ is complete, it implies that $T_1 - T_2 = 0$ s.j. $[P_\theta],\ \theta \in \mathbf{\Theta}$. □

**Theorem 2.1.3** *(Lehmann-Scheffé). Let $X$ follow distribution $P_\theta \in \mathcal{P}$ and let $S$ be a complete and sufficient statistic for $\mathcal{P}$. Then*

**(i)** *For every estimable function $g(\theta)$ and every loss function $L(\theta, t)$ convex in $t$, there exists an unbiased estimate $T$ of $g(\theta)$ which uniformly minimizes the risk $R(\theta, T)$.*

**(ii)** *$T$ is the only unbiased estimate which is a function of $S$. If $L$ is strictly convex in $t$, then $T$ is the only unbiased estimate with minimum risk.*

**Proof.** The Rao-Blackwell theorem holds for $S$ and convex loss function. By Theorem 2.1.2 the estimate $T^*(S(\mathbf{X})) = \mathbb{E}(T(\mathbf{X})|S(\mathbf{X}))$ is unique, and because $S$ is complete, it cannot be further improved. □

### 2.1.1 How to find the best unbiased estimate

Let $S(\mathbf{X})$ be a complete and sufficient statistic.

**Method 1:** The best unbiased estimate of an estimable function $g(\theta)$ is any function $T(S)$ such that

$$\mathbb{E}_\theta T(S) = g(\theta) \quad \forall \theta \in \mathbf{\Theta}.$$

**Method 2:** Let us start with any unbiased estimate $T(\mathbf{X})$. Then

$$T'(\mathbf{X}) = \mathbb{E}(T(\mathbf{X})|S) \quad \text{is the best unbiased estimate.}$$

# Chapter 3

# Equivariant estimators

## 3.1 Estimation of the shift parameter

Let $X_1, \dots, X_n$ be a sample from a distribution with distribution function $F(x-\theta)$ and density $f(x-\theta)$. The problem is to estimate $\theta \in \mathbb{R}^1$ with respect to the loss $L(\theta, t)$.

Consider the loss which is invariant to the shift, i.e. $L(\theta, t) = L(\theta + c, t + c) \quad \forall c \in \mathbb{R}^1$. Then $L(\theta, t) = L(0, t - \theta)$, hence the loss depends only on the difference $t - \theta$. If the loss is invariant, then the whole problem is invariant to the shift. If we estimate $\theta$ by $T(\mathbf{X})$, then a natural estimate of $\theta + c$ is $T(\mathbf{X}) + c$.

**Definition 3.1.1** *The estimator $T(\mathbf{X})$ is called equivariant (with respect to the shift) if it satisfies*

$$T(X_1 + c, \dots, X_n + c) = T(X_1 \dots, X_n) + c \quad \forall c \in \mathbb{R} \quad \textit{and} \quad \forall \mathbf{X} \in \mathbb{R}^n.$$

**Lemma 3.1.1** *The bias, risk and variance of an equivariant estimate $T(\mathbf{X})$ do not depend on value of $\theta$, and hence are constant with respect to $\theta$.*

**Proof.** If $X_1$ has d.f. $F(x-\theta)$, then $P(X_1 - \theta \le z) = P(X_1 \le z + \theta) = F(z)$, thus $X_1 - \theta$ has distribution function $F(\cdot)$. Then

$$\begin{aligned}
\text{bias} &= b(\theta) = \mathbb{E}_\theta(T(\mathbf{X})) - \theta = \mathbb{E}_\theta(T(X_1 - \theta, \dots, X_n - \theta)) = \mathbb{E}_0(T(\mathbf{X})) = b, \\
\mathrm{var}_\theta \mathrm{T}(\mathbf{X}) &= \mathrm{E}_\theta(\mathrm{T}(\mathbf{X}) - \mathrm{E}_\theta \mathrm{T}(\mathbf{X}))^2 = \mathrm{E}_\theta \left[\mathrm{T}(\mathbf{X} - \theta) + \theta - \mathrm{E}_\theta \mathrm{T}(\mathbf{X})\right]^2 \\
&= \mathbb{E}_\theta \left(T(\mathbf{X} - \theta) - b\right)^2 = \mathbb{E}_0 (T(\mathbf{X}) - b)^2, \\
R(T, \theta) &= \mathbb{E}_\theta[L(T(\mathbf{X}) - \theta)] = \mathbb{E}_\theta[L(T(\mathbf{X} - \theta))] = \mathbb{E}_0[L(T(\mathbf{X}))] = R(T).
\end{aligned}$$

□

We shall look for an equivariant estimate with minimal risk (MRE), i.e. $T^*$ such that

$$R(T^*) < R(T) \quad \text{for any equivariant estimator} \quad T \neq T^*.$$

First we should investigate the structure of the class of equivariant estimators.

**Lemma 3.1.2** *Let $T_0(\mathbf{X})$ be an equivariant estimate. Then the estimate $T(\mathbf{X})$ is equivariant if and only if there exists a statistic $U(\mathbf{X})$, invariant to the shift, i.e. satisfying*

$$U(X_1 + c, \dots, X_n + c) = U(X_1, \dots, X_n) \quad \forall c \in \mathbb{R}^1, \ \forall \mathbf{X}, \tag{3.1.1}$$

*such that*

$$T(\mathbf{X}) = T_0(\mathbf{X}) + U(\mathbf{X}) \qquad \forall \mathbf{X}. \tag{3.1.2}$$

**Proof.**

- Let $T$ satisfy (3.1.1) and (3.1.2). Then

$$T(\mathbf{X} + c) = T_0(\mathbf{X} + c) + U(\mathbf{X} + c) = T_0(\mathbf{X}) + c + U(\mathbf{X}) = T(\mathbf{X}) + c,$$

  thus $T$ is equivariant.

- Let $T$ be equivariant, and let $T_0$ be any equivariant estimator.
  Put $U(\mathbf{X}) = T(\mathbf{X}) - T_0(\mathbf{X})$. Then $U$ is invariant and $T = T_0 + U$. □

**Lemma 3.1.3** *The function $U(\mathbf{x})$ is invariant if and only if it depends only on differences $y_i = x_i - x_1, \ i = 2, \dots, n$ in case that $n \geq 2$. If $n = 1$, then the only invariant are the constant functions.*

**Proof.** If $n = 1$, then $U(x + c) \equiv U(x)$ iff $U(x)$ is a constant.
Let $n \geq 2$ and $U(\mathbf{x} + c) \equiv U(\mathbf{x})$. Then
$U(x_1, \dots, x_n) = U(x_1 - x_1, x_2 - x_1, \dots, x_n - x_1) = U(0, y_2, \dots, y_n) = \tilde{U}(y_2, \dots, y_n)$. □

**Corollary 3.1.1** *Let $T_0$ be an equivariant estimate and $n \geq 2$. Then the estimator $T$ is equivariant if and only if there exists a function $\tilde{U}(Y_2, \dots, Y_n)$ of $\mathbf{Y} = (Y_2, \dots, Y_n)$ such that $T(\mathbf{X}) \equiv T_0(\mathbf{X}) - \tilde{U}(\mathbf{Y})$.*

**Remark 3.1.1** *The differences $Y_2 = X_2 - X_1, \dots, Y_n = X_n - X_1$ determine all differences $X_i - X_j, \ i \neq j$. Instead of $\mathbf{Y}$ we may take e.g. $X_1 - \bar{X}, \dots, X_n - \bar{X}$.*

**Definition 3.1.2** *The statistic $S(\mathbf{X})$ is called maximal invariant with respect to the shift, if it is invariant and if*

$$S(\mathbf{X}_1) = S(\mathbf{X}_2) \quad \textit{if and only if} \quad \mathbf{X}_2 = \mathbf{X}_1 + c \textit{ for some } c \in \mathbb{R}^1.$$

We see that $Y_2, \dots, Y_n$ or $X_1 - \bar{X}, \dots, X_n - \bar{X}$ are maximal invariants. Maximal invariants are important, because of the following property:

**Lemma 3.1.4** *The function $U(\mathbf{x})$ is invariant if and only if it is a function of a maximal invariant.*

**Proof.** If $U$ is a function of $S$, i.e. $U(\mathbf{x}) = h(S(\mathbf{x}))$, then it is invariant.

Let $U$ be invariant and let $S(\mathbf{x}_1) = S(\mathbf{x}_2)$. Then $\mathbf{x}_2 = \mathbf{x}_1 + c$, hence $U(x_2) = U(x_1)$. □

**Theorem 3.1.1** *(Minimum risk estimate). Let $T_0$ be an equivariant estimate with a finite risk. If for any value of differences $\mathbf{y}$ there exists $v^*(\mathbf{y})$ which minimizes*

$$\mathbb{E}_0 \left\{ L[T_0(\mathbf{X}) - v(\mathbf{Y})] \Big| \; \mathbf{Y} = \mathbf{y} \right\}$$

*with respect to functions of $\mathbf{y}$, then there exists a minimum risk estimate and is equal to*

$$T^*(\mathbf{X}) = T_0(\mathbf{X}) - v^*(\mathbf{Y}).$$

**Proof.** Let $T(\mathbf{X}) = T_0(\mathbf{X}) - v(\mathbf{Y})$. Then

$$R_\theta(T(\mathbf{X}), \theta) = \mathbb{E}_\theta[L(T_0(\mathbf{X}) - v(\mathbf{Y}) - \theta)] = \mathbb{E}_0\{L[T_0(\mathbf{X}) - v(\mathbf{Y})]\}$$

$$\mathbb{E}_0 \left[ E_0 \left\{ L(T_0(\mathbf{X}) - v(\mathbf{Y})) \Big| \mathbf{Y} \right\} \right] = \int \mathbb{E}_0 \left\{ L\left[T_0(\mathbf{X}) - v(\mathbf{y})\right] \Big| \mathbf{y} \right\} dP_0(\mathbf{y})$$

should be minimized with respect to $v(\cdot)$. But this is minimized if the integrand in minimized for every $\mathbf{y}$. □

**Corollary 3.1.2** **(a)** *If $L(t-\theta) = (t-\theta)^2$, then $v^*(\mathbf{y}) = \mathbb{E}_0 \left\{ T_0(\mathbf{X}) \Big| \; \mathbf{Y} = \mathbf{y} \right\}$.*

**(b)** *If $L(t-\theta) = |t-\theta|$, then $v^*(\mathbf{y})$ is the conditional median of $T_0(\mathbf{X})$ with respect to the conditional distribution of $\mathbf{X}$ given $\mathbf{Y} = \mathbf{y}$.*

**Example 3.1.1** Let $X_1, \ldots, X_n$ be a sample from the normal distribution $N(\xi, \sigma^2)$ with $\sigma$ known. Put $T_0(\mathbf{X}) = \bar{X}$. Then $\bar{X}$ and $\mathbf{Y} = (X_2 - X_1, \ldots, X_n - X_1)$ are independent, hence when we consider $\mathbb{E}_0[L(\bar{X} - v(\mathbf{Y})|\mathbf{Y} = \mathbf{y}]$, we conclude that $v(\mathbf{y}) = const$ and is determined by the condition that $\mathbb{E}_0[L(\mathbf{X} - v)] = \min$. Thus, if $L$ is a convex and odd function, then $v = 0$ and $\bar{X}$ is the MRE (minimum risk estimator).

**Theorem 3.1.2** *Let $\mathcal{F}$ be the class of all distribution functions with Lebesgue densities $f$, which have a finite fixed variance, say $\sigma = 1$. Let $X_1, \ldots, X_n$ be a sample from the distribution with density $f(x - \xi)$, where $\xi = \mathbb{E}\, X$. Let $r_n(f)$ be the risk of the MRE of $\xi$ with respect to the quadratic loss function. Then $r_n(f)$ is maximal over $\mathcal{F}$ for the normal $f$.*

**Proof.** If $F$ is normal, then $\bar{X}$ is the MRE and $\mathbb{E}(\bar{X} - \xi)^2 = 1/n$. Because $1/n$ is also the quadratic risk of $\bar{X}$ for every $F \in \mathcal{F}$, the risk of the MRE $\leq 1/n$. □

**Remark 3.1.2** *It follows from Corrollary 3.1.2 that the MRE should satisfy*

$$T^*(\mathbf{X}) = \bar{X} - \mathbb{E}_0(\bar{X}|\mathbf{Y}), \quad \textit{hence}$$

$$T^*(\mathbf{X}) = \bar{X} \iff \mathbb{E}_0(\bar{X}|\mathbf{Y}) = 0.$$

*But by Theorem of Kagan-Linnik-Rao (1967), $\mathbb{E}_0(\bar{X}|\mathbf{Y}) = 0$ is true if and only if the distribution of $X_1, \ldots, X_n$ is normal.*

**Example 3.1.2** *Exponential distribution.* Let $X_1,\dots,X_n$ have the distribution function

$$F(x-\theta)=\begin{cases}1-\exp\{x-\theta\} & \dots \quad x\geq\theta\\ 0 & \dots \quad x<\theta.\end{cases}$$

Put $T_0(\mathbf{X})=X_{(1)}$, where $X_{(1)}\leq X_{(2)}\leq\dots\leq X_{(n)}$ are order statistics. Then

$$P(X_{(1)}>x)=\prod_{i=1}^{n}P(X_i>x)=\exp\{-n(x-\theta)\},$$

hence the density of $X_{(1)}$ is $n\exp\{-n(x-\theta)\}$.

Because $X_{(1)}$ and $\mathbf{Y}$ are independent, the invariant function $v(\mathbf{Y})=const$, similarly as in Example 3.1.1. We look for $v$ such that $I\!E_0[L(X_{(1)}-v)]=\min$.

If $L(t-\theta)=(t-\theta)^2$, then $I\!E_0(X_{(1)}-v)^2=\min$ for

$$v=I\!EX_{(1)}=n\int_0^\infty x\exp\{-nx\}dx=\frac{1}{n}\int_0^\infty y\exp\{y\}dy=\frac{1}{n}$$

and the MRE is $T^*(\mathbf{X})=X_{(1)}-\frac{1}{n}$ .

### 3.1.1 The form of Pitman (MRE) estimator

Let $X_1,\dots,X_n$ be a sample form a distribution with density $f(x-\theta)$. Then the Pitman (MRE) estimator with respect to quadratic loss is $T^*(\mathbf{X})=T_0(\mathbf{X})-I\!E_0[T_0(\mathbf{X})|\mathbf{Y}]$, where $T_0$ is an initial equivariant estimator with a finite risk. Then $T^*(\mathbf{X})$ can be also written in the following form:

$$T^*(\mathbf{X})=\frac{\int_{-\infty}^{\infty}t\cdot f(X_1-t)\dots f(X_n-t)dt}{\int_{-\infty}^{\infty}f(X_1-t)\dots f(X_n-t)dt}\ .$$

**Proof.** Put $T_0(\mathbf{X})=X_1$. We shall look for the conditional density of $X_1$ given $\mathbf{Y}=\mathbf{y}$ under $\theta=0$. Make the substitution

$$\begin{aligned}y_i&=x_i-x_1,\quad i=2,\dots,n\\ x_1&=x_1.\end{aligned}$$

Then the density of $\mathbf{Y}^*=(X_1,Y_2,\dots,Y_n)$ is

$$p(\mathbf{y}^*)=f(x_1,x_1+y_2,\dots,x_1+y_n)$$

and the conditional density of $X_1$ given $\mathbf{y}=(y_2,\dots,y_n)$ is

$$\frac{f(x_1,x_1+y_2,\dots,x_1+y_n)}{\int_{-\infty}^{\infty}f(u,u+y_2,\dots,u+y_n)du}\ .$$

Hence,

$$\mathbb{E}(X_1|\mathbf{Y}=\mathbf{y}) = \frac{\int_{-\infty}^{\infty} uf(u, u+y_2, \ldots, u+y_n)du}{\int_{-\infty}^{\infty} f(u, u+y_2, \ldots, u+y_n)du}$$
$$= \frac{\int_{-\infty}^{\infty} (X_1-t)f(X_1-t, X_2-t, \ldots, X_n-t)dt}{\int_{-\infty}^{\infty} f(X_1-t, \ldots, X_n-t)dt},$$

where we inserted $t = X_1 - u,\ y_i = X_i - X_1,\ i = 2, \ldots, n$. Then

$$T^*(\mathbf{X}) = X_1 - \mathbb{E}(X_1|\mathbf{Y}=\mathbf{y}) = X_1 - \frac{\int_{-\infty}^{\infty} (X_1-t)f(X_1-t, X_2-t, \ldots, X_n-t)dt}{\int_{-\infty}^{\infty} f(X_1-t, \ldots, X_n-t)dt}$$
$$= \frac{\int_{-\infty}^{\infty} t \cdot f(X_1-t) \ldots f(X_n-t)dt}{\int_{-\infty}^{\infty} f(X_1-t) \ldots f(X_n-t)dt}\ .$$

□

**Example 3.1.3** Let $X_1, \ldots, X_n$ be a sample from the uniform distribution $R(\theta-\frac{1}{2}, \theta+\frac{1}{2})$ and let $L(t-\theta) = (t-\theta)^2$. Then

$$f(x_1, \ldots, x_n) = \begin{cases} 1 & \ldots \quad \theta - \frac{1}{2} \le X_{(1)} \le X_{(n)} \le \theta + \frac{1}{2}\ , \\ 0 & \ldots \quad \text{otherwise} \end{cases}$$

Then, under $\theta = 0$,

$$f(x_1-t, \ldots, x_n-t) = \begin{cases} 1 & \ldots \quad X_{(n)} - \frac{1}{2} \le t \le X_{(1)} + \frac{1}{2}\ , \\ 0 & \ldots \quad \text{otherwise} \end{cases}$$

Put $T_0 = \frac{1}{2}(X_{(1)} + X_{(n)})$. Then

$$\int tf(x_1-t, \ldots, x_n-t)dt = \int_{X_{(n)}-\frac{1}{2}}^{X_{(1)}+\frac{1}{2}} tdt = \tfrac{1}{2}\left[(X_{(1)}+\tfrac{1}{2})^2 - (X_{(n)}-\tfrac{1}{2})^2\right]$$

and

$$\int f(x_1-t, \ldots, x_n-t)dt = \int_{X_{(n)}-\frac{1}{2}}^{X_{(1)}+\frac{1}{2}} dt = 1 - (X_{(n)} - X_{(1)}).$$

Finally,

$$T^*(\mathbf{X}) = \frac{1}{2}\frac{(X_{(1)}+X_{(n)})(1-(X_{(n)}-X_{(1)}))}{(1-(X_{(n)}-X_{(1)}))} = \tfrac{1}{2}(X_{(1)}+X_{(n)}).$$

## 3.2 Relation of equivariance and unbiasedness

**Lemma 3.2.1** *Let $L(t-\theta) = (t-\theta)^2$.*

**(i)** *If $T(\mathbf{X})$ is equivariant and has constant bias $\mathbb{E}_\theta T(\mathbf{X}) - \theta = b$ (a non-null constant), then $T(\mathbf{X}) - b$ is an equivariant and unbiased estimator, whose risk is less that the risk of $T(\mathbf{X})$.*

**(ii)** *If the MRE is uniquely determined, then it is unbiased.*

**(iii)** *If there exists a uniformly best unbiased estimate which is equivariant, then it is the MRE.*

**Proof.**

**(i)** Let $T_1(\mathbf{X}) = T(\mathbf{X}) - b$. Then it is equivariant and $\mathbb{E}_\theta(T_1(\mathbf{X})) = \theta + b - b = \theta$, and

$$\mathbb{E}_0(T_1(\mathbf{X}))^2 = \mathbb{E}_0(T(\mathbf{X}) - b)^2 = \mathbb{E}_0 T^2(\mathbf{X}) - b^2 < \mathbb{E}_0 T^2(\mathbf{X}).$$

**(ii)** Let $T^*$ be the MRE and $T$ be any equivariant estimate with finite risk. Then

$$T^*(\mathbf{X}) = T(\mathbf{X}) - \mathbb{E}(T|\mathbf{Y}), \quad \mathbb{E}_0(T^*)^2 < \mathbb{E}_0 T^2, \quad \mathbb{E}_0 T^* = 0.$$

**(iii)** Let $T$ be uniformly best unbiased and also equivariant. Let $T_1$ be equivariant. Then $\mathbb{E}_\theta T_1 = \theta + b$ and if $b \neq 0$, then $\mathbb{E}_0(T_1 - b)^2 < \mathbb{E}_0 T_1^2$. This implies that

$$\mathbb{E}_\theta(T-\theta)^2 = \mathbb{E}_0(T)^2 \leq \mathbb{E}(T_1 - b)^2 \leq \mathbb{E}_0 T_1^2.$$

□

**Definition 3.2.1** *Estimator $T$ of $g(\theta)$ is called risk unbiased with respect to the loss $L$, if*

$$E_\theta L(\theta, T) \leq \mathbb{E}_\theta L(\theta', T) \quad \forall \theta' \neq \theta.$$

The following theorem shows that the MRE is risk unbiased:

**Theorem 3.2.1** *Let $X_1, \ldots, X_n$ be a sample from a distribution with the density $f(x-\theta)$. Then the MRE with respect to the loss $L(\theta, t) = L(t-\theta)$ is risk unbiased.*

**Proof.** The risk unbiasedness means that

$$\mathbb{E}_\theta L(T(\mathbf{X}) - \theta') \geq \mathbb{E}_\theta L(T(\mathbf{X}) - \theta) \qquad \forall \theta' \neq \theta,$$

otherwise speaking,

$$\mathbb{E}_0 L(T(\mathbf{X}) - a) \geq \mathbb{E}_0 L(T(\mathbf{X})) \qquad \forall a \neq 0.$$

Let $T^*$ be the MRE. Then $T^*(\mathbf{X}) = T_0(\mathbf{X}) - v^*(\mathbf{Y})$ where

$$\mathbb{E}_0[L(T_0(\mathbf{X}) - v^*(\mathbf{Y})|\mathbf{Y} = \mathbf{y}] = \min .$$

Then

$$\begin{aligned}
&\mathbb{E}_0[L(T(\mathbf{X}) - a)] = \mathbb{E}_0[L(T_0(\mathbf{X}) - v^*(\mathbf{Y}) - a)] \\
&= \mathbb{E}_0\{\mathbb{E}_0[L(T_0(\mathbf{X}) - v^*(\mathbf{Y}) - a)|\mathbf{Y}]\} \geq \mathbb{E}_0\{\mathbb{E}_0[L(T_0(\mathbf{X}) - v^*(\mathbf{Y})|\mathbf{Y}]\} \\
&= \mathbb{E}_0[L(T(\mathbf{X}))]
\end{aligned}$$

where we used the fact that $v^*(\mathbf{Y}) + a$ is also an invariant function. □

# Chapter 4

# Asymptotic behavior of estimates

## 4.1 Consistency

Let $X_1, \dots, X_n$ be independent observations with distribution $P_\theta$, $\theta \in \boldsymbol{\Theta}$. We want to estimate the function $g(\theta)$. Then the estimator $T_n$ is called

weakly consistent estimate of $g(\theta)$ if $T_n \xrightarrow{p} g(\theta) \qquad \forall \theta \in \boldsymbol{\Theta}$ as $n \to \infty$

strongly consistent estimate of $g(\theta)$ if $T_n \to g(\theta)\ a.s.[P_\theta] \quad \forall \theta \in \boldsymbol{\Theta}$ as $n \to \infty$.

Let $R(\theta, T_n) = \mathbb{E}_\theta (T(\mathbf{X}) - g(\theta))^2$ be the quadratic risk. Then

**Theorem 4.1.1** **(i)** *If* $\lim_{n\to\infty} R(\theta, T_n) = 0 \qquad \forall \theta \in \boldsymbol{\Theta}$, *then* $T_n$ *is weakly consistent.*

**(ii)** *If*

$$\lim_{n\to\infty} \mathbb{E}_\theta T_n(\mathbf{X}) = g(\theta) \qquad \forall \theta \in \boldsymbol{\Theta},$$
$$\lim_{n\to\infty} \mathrm{var}_\theta \mathrm{T_n}(\mathbf{X}) = 0 \qquad \forall \theta \in \boldsymbol{\Theta}$$

*then* $T_n$ *is weakly consistent.*

**(iii)** *Especially, if* $T_n$ *is unbiased* $\forall n$ *and* $\lim_{n\to\infty} \mathrm{var}_\theta \mathrm{T_n}(\mathbf{X}) = 0 \qquad \forall \theta \in \boldsymbol{\Theta}$, *then* $T_n$ *is weakly consistent.*

**Proof.**

**(i)** By Chebyshev inequality,

$$P_\theta(|T_n(\mathbf{X})) - g(\theta)| > \varepsilon) \le \frac{1}{\varepsilon^2} \mathbb{E}_\theta \left( |T_n(\mathbf{X})) - g(\theta)|^2 \right) = \frac{1}{\varepsilon^2} R(\theta, T_n) \to 0.$$

**(ii)**

$$P_\theta(|T_n(\mathbf{X})) - g(\theta)| > \varepsilon) \le \frac{1}{\varepsilon^2} \mathbb{E}_\theta \left( |T_n(\mathbf{X})) - g(\theta)|^2 \right) \le \frac{1}{\varepsilon^2}$$
$$\frac{1}{\varepsilon^2} \mathbb{E}_\theta \left[ T_n - \mathbb{E}_\theta T_n + \mathbb{E}_\theta T_n - g(\theta) \right]^2 \le \frac{2}{\varepsilon^2} \left( \mathrm{var}_\theta \mathrm{T_n} + (\mathrm{b_n}(\theta))^2 \right) \to 0.$$

□

The parameter $\theta$ or the function $g(\theta)$ can be estimated by a consistent estimate only if $\theta$ is identifiable, i.e. if $[\theta_1 \neq \theta_2] \Longrightarrow [P_{\theta_1} \neq P_{\theta_2}]$.

## 4.2 Efficiency

**Definition 4.2.1** *(Limiting risk efficiency of $T_n$ to $T_n^*$). Assume that two sequences $\{T_n\}$, $\{T_n^*\}$ of estimates satisfy*

$$\lim_{n\to\infty} n^r R(T_n, \theta) = \lim_{n\to\infty} n^r R(T_{m_n}^*, \theta) \tag{4.2.1}$$

*for some sequence $\{m_n\}_{n=1}^{\infty}$ and a fixed $r > 0$. Then the limit*

$$\lim_{n\to\infty} \frac{m_n}{n},$$

*if it exists and is independent of the special choice of $\{m_n\}$, is called the limiting risk efficiency of $T_n$ with respect to $T_n^*$.*

**Definition 4.2.2** *(Relative asymptotic efficiency of $T_n$ to $T_n^*$). Let*

$$\begin{aligned} \sqrt{n}(T_n - g(\theta)) &\xrightarrow{\mathcal{D}} N(0, \sigma^2) \quad as \quad n \to \infty, \\ \sqrt{n}(T_{m_n}^* - g(\theta)) &\xrightarrow{\mathcal{D}} N(0, \sigma^2) \quad as \quad n \to \infty. \end{aligned} \tag{4.2.2}$$

*Then the limit*

$$e_{T,T^*} = \lim_{n\to\infty} \frac{m_n}{n},$$

*if it exists and is independent of the special choice of $\{m_n\}$, is called the relative asymptotic efficiency (ARE) of $T_n$ to $T_n^*$.*

**Theorem 4.2.1** *Let*

$$\begin{aligned} \sqrt{n}(T_n - g(\theta)) &\xrightarrow{\mathcal{D}} N(0, \sigma^2) \quad as \quad n \to \infty, \\ \sqrt{n}(T_n^* - g(\theta)) &\xrightarrow{\mathcal{D}} N(0, \sigma_*^2) \quad as \quad n \to \infty. \end{aligned} \tag{4.2.3}$$

*Then*

$$e_{T,T^*} = \frac{\sigma_*^2}{\sigma^2}.$$

**Proof.** Assume (4.2.3). Then

$$\sqrt{n}(T^*_{m_n} - g(\theta)) = \sqrt{\frac{n}{m_n}}\sqrt{m_n}(T^*_{m_n} - g(\theta))$$

and

$$\sqrt{n}(T_n - g(\theta)) \xrightarrow{\mathcal{D}} N(0,\sigma^2), \qquad \sqrt{\frac{n}{m_n}} \to \frac{1}{e_{T,T^*}},$$
$$\sqrt{m_n}(T^*_{m_n} - g(\theta)) \xrightarrow{\mathcal{D}} N(0,\sigma^2_*),$$

thus $e_{T,T^*} = \sigma^2_*/\sigma^2$. □

Consider the system of distributions $\mathcal{P} = \{P_\theta;\ \theta \in \mathbf{\Theta}\}$ with densities $f(x,\theta)$ satisfying

$(A_0)$ $P_{\theta_1} \neq P_{\theta_2}$ for $\theta_1 \neq \theta_2$.

$(A_1)$ $B = \{x:\ f(x,\theta) > 0\}$ is independent of $\theta$.

$(A_2)$ Let $X_1,\ldots,X_n$ be a sample from a distribution with density $f(x,\theta_0)$, where $\theta_0 \in \mathcal{I} \subset \mathbf{\Theta}$ for an open interval $\mathcal{I}$.

**Theorem 4.2.2** *Under conditions $(A_0)$–$(A_2)$, it holds for any $\theta \neq \theta_0,\ \theta \in \mathbf{\Theta}$*

$$\lim_{n\to\infty} P_{\theta_0}\left\{\prod_{i=1}^n f(X_i,\theta_0) > \prod_{i=1}^n f(X_i,\theta)\right\} = 1. \tag{4.2.4}$$

**Proof.** By the law of large numbers and Jenssen inequality, as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^n \log\frac{f(X_i,\theta)}{f(X_i,\theta_0)} \xrightarrow{P_{\theta_0}} \mathbb{E}_{\theta_0}\ \log\frac{f(X,\theta)}{f(X,\theta_0)} < \log\ \mathbb{E}_{\theta_0}\frac{f(X,\theta)}{f(X,\theta_0)} = 0.$$

This implies

$$P_{\theta_0}\left\{\frac{1}{n}\sum_{i=1}^n \log\frac{f(X_i,\theta)}{f(X_i,\theta_0)} > 0\right\} \to 0$$

and that gives (4.2.4). □

Denote

$$L(\theta,\mathbf{X}) = \log\prod_{i=1}^n f(X_i,\theta) \qquad \text{(the likelihood).}$$

The maximum likelihood estimate (MLE) of $\theta$ is defined as a solution of the maximization

$$L(\theta,\mathbf{X}) = \max,\ \ \theta \in \mathbf{\Theta}.$$

It is one of the solutions of the likelihood equation

$$\frac{\partial L(\theta,\mathbf{X})}{\partial\theta} = \sum_{i=1}^n \frac{\dot{f}(X_i,\theta)}{f(X_i,\theta)} = 0. \tag{4.2.5}$$

Assume that conditions $A_0$–$A_2$ are satisfied and that $f(x,\theta)$ is differentiable in $\theta \in \mathcal{I} \subset \boldsymbol{\Theta}$, where $\mathcal{I} \ni \theta_0$. Then

**Theorem 4.2.3** *Under the above conditions, there exists a root $\hat{\theta}_n$ of the likelihood equation (4.2.5) such that*

$$\hat{\theta}_n \xrightarrow{P_{\theta_0}} \theta_0$$

*as $n \to \infty$.*

**Proof.** Let $a > 0$ is such that $(\theta_0 - a, \theta_0 + a) \subset \mathcal{I}$. Let

$$S_n = \{\mathbf{x} : L(\theta_0, \mathbf{x}) > L(\theta_0 - a, \mathbf{x}) \ \text{ and } \ L(\theta_0, \mathbf{x}) > L(\theta_0 + a, \mathbf{x})\}.$$

By Theorem 4.2.2 is $\lim_{n\to\infty} P_{\theta_0}(S_n) = 1$. There is a local maximum $\hat{\theta}_n$ between $\theta_0 - a$ and $\theta_0 + a$ and it satisfies $L'(\hat{\theta}_n) = 0$. Let $\theta_n^*$ be the root of $L'(\theta) = 0$ the closest to $\theta_0$. Then

$$\lim_{n\to\infty} P_{\theta_0}(|\theta_n^* - \theta_0| < a) = 1 \qquad \forall a > 0.$$

□

**Remark 4.2.1** *We know that $\Theta_n^*$ exists as the root the closest to $\theta_0$, but we are not able to find it, because $\theta_0$ is unknown.*

*Everything holds only with probability tending to 1.*

*If the likelihood equation has only one root $T_n$ $\forall n$ and $\forall \mathbf{x}$, then $T_n$ is consistent estimate of $\theta_0$.*

**Theorem 4.2.4** *Let the conditions $(A_0)$–$(A_2)$ be satisfied, and let it further hold*

$$(A_3) \qquad \left| \frac{\partial^3 \log f(x,\theta)}{\partial \theta^3} \right| \le M(\mathbf{x})$$

*for $\mathbf{x} \in B$ and for $|\theta - \theta_0| < C$, where $M(\mathbf{x})$ is such that $\mathbb{E}_{\theta_0} M(\mathbf{X}) < \infty$. Then every consistent sequence $\hat{\theta}_n = \hat{\theta}_n(\mathbf{X})$ of roots of the likelihood equation is asymptotically normally distributed, i.e.*

$$\mathcal{L}(\sqrt{n}(\hat{\theta}_n - \theta_0)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{\mathcal{I}(\theta_0)}\right)$$

*where $\mathcal{I}(\theta) = \int \left(\frac{\partial \log f(x,\theta)}{\partial \theta}\right)^2 f(x,\theta) d\mu$ is the Fisher information.*

**Some steps of the proof.**

$$0 = n^{-1/2} L_n'(\hat{\theta}_n) = n^{-1/2} \sum_{i=1}^{n} \frac{\dot{f}(X_i, \hat{\theta}_n)}{f(X_i, \hat{\theta}_n)}$$

$$= n^{-1/2} L_n'(\theta_0) + n^{1/2}(\hat{\theta}_n - \theta_0) \cdot \tfrac{1}{n} L_n''(\theta_0) + \tfrac{1}{2} n^{-1/2} [n^{1/2}(\hat{\theta}_n - \theta_0)]^2 \tfrac{1}{n} L_n'''(\theta_n^*)$$

with $\theta_n^*$ between $\theta_0$ and $\hat{\theta}_n$. Then

$$n^{1/2}(\hat{\theta}_n - \theta_0) \approx -\frac{n^{-1/2}L_n'(\theta_0)}{n^{-1}L_n''(\theta_0)} - \frac{1}{2}(\hat{\theta}_n - \theta_0)\frac{n^{-1}L_n'''(\theta_n^*)}{n^{-1}L_n''(\theta_0)}.$$

We should show that

$$n^{-1/2}L_n'(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\ \mathcal{I}(\theta_0)) \tag{4.2.6}$$

$$-\frac{1}{n}L_n''(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\frac{\ddot{f}(X_i,\theta_0)}{f(X_i,\theta_0)} - \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\dot{f}(X_i,\theta_0)}{f(X_i,\theta_0)}\right)^2 \xrightarrow{p} \mathcal{I}(\theta_0) \tag{4.2.7}$$

$$\frac{1}{n}L_n'''(\theta_n^*) = O_p(1). \tag{4.2.8}$$

(4.2.7) follows from the central limit theorem, (4.2.8) from the law of large numbers, (4.2.8) from the consistency of $\hat{\theta}_n$ and from $(A_3)$. Then we obtain

$$0 = n^{-1/2}L_n'(\hat{\theta}_n) \approx \mathcal{N}(0,\mathcal{I}(\theta_0)) - n^{1/2}(\hat{\theta}_n - \theta_0)\mathcal{I}(\theta_0) + \frac{1}{2\sqrt{n}}\left(\sqrt{n}(\hat{\theta}_n - \theta_0)\right)^2\frac{1}{n}L_n'''(\theta_n^*),$$

thus

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \approx -\frac{n^{-1/2}L_n'(\theta_0)}{n^{-1}L_n''(\theta_0)} - \frac{1}{2}(\hat{\theta}_n - \theta_0)\frac{n^{-1}L_n'''(\theta_n^*)}{n^{-1}L_n''(\theta_0)} \approx \mathcal{N}\left(0, \frac{1}{\mathcal{I}(\theta_0)}\right) + o_p(1)$$

□

**Remark 4.2.2** *Such estimator is called the efficient likelihood estimator. It is usually the maximal likelihood estimator, but not neccessary.*

**Corollary 4.2.1** *If the likelihood equation has only one root, or if it has a multiple root with probability tending to 0 as $n \to \infty$, then, under the conditions of Theorem 4.2.4, the maximal likelihood estimator is asymptotically efficient.*

**Example 4.2.1** One-parameter exponential family.

$$f(x,\theta) = \exp\{\theta T(\mathbf{x}) + A(\theta),$$

$$\sum_{i=1}^{n}\log f(X_i,\theta) = \theta\sum_{i=1}^{n}T(X_i) + nA(\theta) = \max$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}T(X_i) = -A'(\theta) = \mathbb{E}_\theta\left(\frac{1}{n}\sum_{i=1}^{n}T(X_i)\right) \quad \text{(likelihood equation).} \tag{4.2.9}$$

On the other hand, because $\int f(x,\theta)d\mu = 1$,

$$0 = \int (A'(\theta) + T(x))\exp\{\theta T(x) + A(\theta)\}d\mu \quad \Longrightarrow \quad A'(\theta) = -\mathbb{E}_\theta T(X).$$

We can show that $\mathbb{E}_\theta T(X)$ is increasing in $\theta$ : Indeed,

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta T(X) = \int T(x)(A'(\theta)+T(x))\exp\{\theta T(\mathbf{x})+A(\theta)d\mu = \mathbb{E}_\theta T^2(X)-(\mathbb{E}_\theta T(X))^2 = \mathrm{var}_\theta \mathrm{T(X)} > 0.$$

Thus the likelihood equation

$$\mathbb{E}_\theta T(X) = \frac{1}{n}\sum_{i=1}^{n} T(X_i)$$

has at most one solution, and the conditions of Theorem 4.2.4 are satisfied. Thus, with probability tending to 0 the likelihood equation has one root $\hat{\theta}_n$, which is consistent, asymptotically efficient and asymptotically normal

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{\mathrm{var}_\theta \mathrm{T}}\right),$$

because

$$\mathcal{I}(\theta) = \mathbb{E}_\theta\left[\frac{\partial \log f(X,\theta)}{\partial\theta}\right]^2 = \mathbb{E}_\theta[T(X) + A'(\theta)]^2 = \mathrm{var}_\theta \mathrm{T(X)}.$$

**Example 4.2.2** *Truncated normal distribution.* Let $X_1, \ldots, X_n$ have normal distribution $\mathcal{N}(\theta, 1)$ truncated at $(a, b)$, $a < b$, with the density

$$p(x,\theta) = \begin{cases} \frac{1}{\sqrt{2\pi}}\exp\{-\frac{(x-\theta)^2}{2}\}\Big/[\Phi(b-\theta)-\Phi(a-\theta)] & \ldots\quad a < x < b \\ 0 & \ldots\quad \textit{otherwise.}\end{cases}$$

Thus

$$p(x,\theta) = \exp\{\theta x - \frac{\theta^2}{2} - \frac{x^2}{2} + \log\frac{1}{\sqrt{2\pi}} - \log[\Phi(b-\theta)-\Phi(a-\theta)] \ \Rightarrow\ T(x) = x$$

and the likelihood equation has the form

$$\bar{X}_n = \mathbb{E}_\theta X.$$

If $\theta \to \pm\infty$, then $X \xrightarrow{p} a$ or $b$, thus also $\mathbb{E}_\theta X \to a$ or $b$ and $\mathbb{E}_\theta X$ is continuous, hence the likelihood equation has a root.

### 4.2.1 Shift parameter

Let $X_1, \ldots, X_n$ be a sample from the population with density $f(x-\theta)$. The MLE $\hat{\theta}_n$ is a solution of

$$\prod_{i=1}^{n} f(X_i - \theta) := \max$$

and it is equivariant. The Pitman estimate $T_n^*$ is asymptotically equivalent to $\hat{\theta}_n$ in the sense that $\sqrt{n}(\hat{\theta}_n - T_n^*) \xrightarrow{p} 0$ as $n \to \infty$ (Stone 1974). The likelihood equation can be rewritten as

$$\sum_{i=1}^{n} \frac{f'(X_i - \theta)}{f(X_i - \theta)} = 0. \tag{4.2.10}$$

If $f$ is strongly unimodal, i.e. $-\frac{f'}{f}$ is strictly increasing, then (4.2.10) has at most one root. Because $\prod_{i=1}^{n} f(x_i - \theta) \to 0$ as $\theta \to \pm\infty$, then $\prod_{i=1}^{n} f(X_i - \theta)$ has the maximum inside the real line, hence the root of (4.2.10) exists and is asymptotically efficient.

### 4.2.2 Multiple root

Let $L(\theta, \mathbf{x}) = \log \prod_{i=1}^{n} f(x_i, \theta)$. Assume that the equation

$$L'(\theta) = \sum_{i=1}^{n} \frac{f'(X_i, \theta)}{f(X_i, \theta)} = 0 \tag{4.2.11}$$

can have a multiple root, but that there exists a consistent estimate $\tilde{\theta}_n^0$.

**Theorem 4.2.5** **(i)** *Let $\tilde{\theta}_n^0$ be a consistent estimate and the conditions $(A_0)$−−$(A_2)$ hold. Then the root of equation (4.2.11), the closest to $\tilde{\theta}_n^0$ is also consistent, and hence it is asymptotically efficient.*

**(ii)** *Let $\tilde{\theta}_n$ be a consistent initial estimate satisfying*

$$\sqrt{n}(\tilde{\theta}_n - \theta) = O_p(1) \quad as \quad n \to \infty.$$

*Put*

$$T_n = \tilde{\theta}_n - \frac{L'(\tilde{\theta}_n)}{L''(\tilde{\theta}_n)}.$$

*Then $T_n$ is an asymptotically efficient estimate of $\theta$, i.e.*

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1/\mathcal{I}(\theta)).$$

**Proof** is similar to the proof of Theorem 4.2.4.