Minimax estimators of the coverage probability of the impermissible error for a location family

by Miguel A. Arcones

Binghamton University

arcones@math.binghamton.edu

Talk based on:

Arcones, M. A. (2008). Minimax estimators of the coverage probability of the impermissible error for a location family. Statistics & Decisions. 28, 1001–1043

Available at

http://www.math.binghamton.edu/arcones/prep.html

1 Confidence regions with a constrained Lebesgue measure for a location family.

Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of $\mathbb{R}^d$–valued i.i.d.r.v.'s with a p.d.f. belonging to the family $\{f(\cdot - \theta) : \theta \in \mathbb{R}^d\}$.

Fix $L > 0$. We would like to find a translation equivariant confidence region $C_{HL,L}(X_1, \dots, X_n)$ such that:

1. For each $\vec{x} := (x_1, \dots, x_n)$,

$$\int_{\mathbb{R}^m} I(y \in C_{HL,L}(\vec{x}))\, dy \le L. \tag{1}$$

2. For any confidence region $C(X_1, \dots, X_n)$ satisfying (1),

$$\inf_{\theta \in \mathbb{R}^d} \mathbb{P}_\theta\{\theta \in C_{HL,L}(\vec{X})\} \ge \inf_{\theta \in \mathbb{R}^d} \mathbb{P}_\theta\{\theta \in C(\vec{X})\}$$

If $C(\vec{X})$ is a translation equivariant confidence region, then

$$C(\vec{X}) = X_n + C(Z_1, \ldots, Z_{n-1}, 0) = X_n + C^*(\vec{Z}).$$

We would like to find $C^*(\vec{Z})$ maximizing

$$\inf_{\theta \in \mathbb{R}^d} \mathbb{P}_\theta\{\theta \in X_n + C^*(\vec{Z})\} = \mathbb{P}_0\{-X_n \in C^*(\vec{Z})\} \quad (2)$$

$$= \int\int I(-x_n \in C^*(\vec{z})) \prod_{j=1}^{n-1} f(z_j + x_n) \times f(x_n)\, dx_n\, d\vec{z}$$

$$= \int\int I(y \in C^*(\vec{z})) \prod_{j=1}^{n-1} f(z_j - y) \times f(-y)\, dy\, d\vec{z},$$

subject that for each $\vec{z}$,

$$\int_{\mathbb{R}^m} I(y \in x_n + C^*(\vec{z}))\, dy = \int_{\mathbb{R}^m} I(y \in C^*(\vec{z}))\, dy \leq L. \quad (3)$$

Conditioning on $\vec{z}$, we get that the optimal region is

$$C^*_{HL}(\vec{z}) = \{y \in \mathbb{R}^d : \lambda_{HL}(\vec{z}) \leq \prod_{j=1}^{n-1} f(z_j - y) \times f(-y))\}.$$

where $\lambda_{HL}(\vec{z})$ is such that

$$\int_{\lambda_{HL}(\vec{z}) \leq \prod_{j=1}^{n-1} f(z_j - y) \times f(-y))} dy = L.$$

It is easy to see that

$$C_{HL,L}(x_1, \ldots, x_n) = x_n + C^*_{HL}(\vec{z})$$

$$= \{\theta \in \mathbb{R}^d : \lambda_{HL}(\vec{z}) \leq \prod_{j=1}^{n} f(x_j - \theta)\}.$$

Theorem 1. *Suppose that:*

(i) For each $0 < M < \infty$, $\inf\{f(x) : |x| \leq M\} > 0$*.*

(ii) $\lim_{|x|\to\infty} f(x) = 0$*.*

(iii) $\sup_{x\in\mathbb{R}^d} f(x) < \infty$*.*

Let $C(X_1, \ldots, X_n)$ *be a confidence region for* θ *such that for each* $x_1, \ldots, x_n$*,* $\int_{\mathbb{R}^d} I(y \in C(x_1, \ldots, x_n))\, dy \leq L$*. Then,*

$$\inf_{\theta\in\mathbb{R}^d} \mathbb{P}_\theta\{\theta \in C_{HL,L}(\vec{X})\} \geq \inf_{\theta\in\mathbb{R}^d} \mathbb{P}_\theta\{\theta \in C(\vec{X})\}.$$

2 Impermissible error

Let $T_n := T_n(\vec{X})$ be an estimator of θ. The error of the estimator T_n is $|T_n - \theta|$. Suppose that we select a number b, $b > 0$, such that any error of estimation bigger than b is impermissible. The maximum coverage probability of the impermissible error of the estimator T_n is

$$\sup_{\theta \in \Theta} \mathbb{P}_\theta\{|T_n - \theta| > b\}. \tag{4}$$

We would like to obtain an estimator $T_{HL,b}$ minimizing (4) for a location family. Finding $T_{HL,b}$ is equivalent to find a confidence interval of length $2b$ with coverage probability as large as possible.

If $d = 1$, and $C_{HL,2b}(\vec{X})$ is an interval, we take $T_{n,HL,b}(\vec{X})$ as the middle point of $C_{HL,2b}(\vec{X})$. By Theorem 1, for each estimator T_n,

$$\sup_{\theta \in \mathbb{R}} \mathbb{P}_\theta\{|T_n - \theta| > b\} \geq \sup_{\theta \in \mathbb{R}} \mathbb{P}_\theta\{|T_{n,HL,b} - \theta| > b\}.$$

3 Estimation via the m.l.e.

An mle $\hat{\theta}_n$ of θ, over a parametric family $\{f(x,\theta) : \theta \in \Theta\}$ is a r.v. such that

$$n^{-1}\sum_{j=1}^{n} \log f(X_j, \hat{\theta}_n) = \sup_{\theta\in\Theta} n^{-1}\sum_{j=1}^{n} \log f(X_j, \theta). \qquad (5)$$

Suppose that there exists a r.v. $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ satisfying (5) when $f(x,\theta) = f(x-\theta)$. We may assume that $\hat{\theta}_n$ is equivariant. This implies that $\hat{\theta}_n - \theta$ is a pivotal quantity. Suppose that $\theta - \hat{\theta}_n$ has a distribution absolutely continuous with respect to the Lebesgue measure, when θ obtains. Let $h_n(\cdot)$ be the pdf of $\theta - \hat{\theta}_n$ when θ obtains. Let $L > 0$. Given $C \in \mathcal{B}(\mathbb{R}^d)$, a confidence region for θ based on $\theta - \hat{\theta}_n$ is determined by $\{\theta \in \mathbb{R}^d : \theta - \hat{\theta}_n \in C\} = \{\theta \in \mathbb{R}^d : \theta \in \hat{\theta}_n + C\}$.

Between all sets C such that

$$\int I(\theta \in \hat{\theta}_n + C)\, d\theta = \int_C 1\, d\theta \leq L, \tag{6}$$

a set C which maximizes

$$\mathbb{P}_\theta\{\theta - \hat{\theta}_n \in C\} = \int_C h_n(\theta)\, d\theta \tag{7}$$

is

$$C_{\mathrm{mle},L} = \{x \in \mathbb{R}^d : \lambda_{\mathrm{mle},L} \leq h_n(x)\}$$

where $\lambda_{\mathrm{mle},L}$ is such that

$$\int_{\{x \in \mathbb{R}^d : \lambda_{\mathrm{mle},L} \leq h_n(x)\}} 1\, dx = L.$$

Hence, using the mle, a confidence region satisfying (6) and maximizing (7) is

$$\hat{\theta}_n + C_{\mathrm{mle},L} = \{\theta \in \mathbb{R}^d : \lambda_{\mathrm{mle},L} \leq h_n(\theta - \hat{\theta}_n)\}.$$

Notice $h_n(\theta - x)$, $x \in \mathbb{R}$, is the p.d.f. of $\hat{\theta}_n$ when θ obtains.

The obtained confidence regions maximize the coverage probability over all the regions with expected Lebesgue measure less or equal than L:

Theorem 2. *Let $C(\hat{\theta}_n)$ be a confidence region for θ such that*

$$\sup_{\theta \in \mathbb{R}^d} E_\theta \left[\int_{\mathbb{R}^d} I(y \in C(\hat{\theta}_n)) \, dy \right] \leq L. \tag{8}$$

Then,

$$\inf_{\theta \in \mathbb{R}^d} \mathbb{P}_\theta \{ \theta \in C(\hat{\theta}_n) \} \leq \inf_{\theta \in \mathbb{R}^d} \mathbb{P}_\theta \{ \theta \in \hat{\theta}_n + C_{\mathrm{mle},L} \}. \tag{9}$$

The previous theorem implies that if $C(\hat{\theta}_n)$ is a confidence region for θ such that for each $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} I(y \in C(x)) \, dy \leq L, \tag{10}$$

then, (9) holds.

Suppose that $d = 1$ and we would like to find the estimator $T(\hat{\theta}_n)$ (based on the m.l.e.) minimizing:

$$\sup_{\theta \in \mathbb{R}} \mathbb{P}_\theta\{|T(\hat{\theta}_n) - \theta| > b\}. \tag{11}$$

If $C_{\mathrm{mle},2b}$ is an interval, then the middle point $T_{n,\mathrm{mle},b}$ of $\hat{\theta}_n + C_{\mathrm{mle},2b}$ minimizes (11) between all the estimators based on the m.l.e.

4 Asymptotic results.

Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of i.i.d.r.v.'s from the family $\{f(\cdot - \theta) : \theta \in \mathbb{R}\}$.

Theorem 3. *Suppose that:*

(i) For each $x \in \mathbb{R}$, $f(x) > 0$.

(ii) $\log f(\cdot)$ *is a strictly concave function*

(iii) For each $t \in \mathbb{R}$, *there exists* $\lambda_t > 0$ *such that*

$$E_0[\exp(\lambda_t |\log f(X-t)|)] < \infty.$$

Then,

$$\lim_{n\to\infty} n^{-1} \log \mathbb{P}_\theta\{\theta \notin C_{HL,L}(X_1,\ldots,X_n)\} = -S(L),$$

where

$$S(L) = -\inf_{\lambda\in\mathbb{R}} \log \int_{\mathbb{R}} (f(x-L))^\lambda (f(x))^{1-\lambda}\, dx.$$

Besides,

(a) S *is increasing on* $[0,\infty)$ *and* S *is decreasing in* $(-\infty, 0]$.

(b) S *is a continuous function.*

(c) $\lim_{|t|\to\infty} S(t) = \infty$.

Theorem 4. *Suppose that:*

(i) For each $x \in \mathbb{R}$, $f(x) > 0$.

(ii) $\log f$ *is a strictly concave function.*

(iii) $\int_{\mathbb{R}} f'(x)\,dx = 0$.

(iv) $E[(f(X))^{-2}(f'(X))^2] < \infty$.

Then, the m.l.e. $\hat{\theta}_n$ *is well defined and* $\theta - \hat{\theta}_n$, *when* θ *obtains, has pdf*

$$h_n(t) := E_0\left[I\left(\sum_{j=1}^n (f(X_j+t))^{-1} f'(X_j+t) > 0\right)\right.$$
$$\left.\times \sum_{j=1}^n (f(X_j))^{-1} f'(X_j)\right], t \in \mathbb{R}.$$

Besides, h_n *is nonincreasing in* $[0,\infty)$ *and nondecreasing in* $(-\infty, 0]$.

Theorem 5. *Suppose that:*

(i) For each $x \in \mathbb{R}$*,* $f(x) > 0$*.*

(ii) $\log f$ *is a concave function.*

(iii) $E_0[(f(X))^{-1}f'(X)] = 0$*.*

(iv) $E_0[(f(X))^{-2}(f'(X))^2] < \infty$*.*

(v) $\hat{\theta}_n - \theta$ *satisfies the LDP with speed* n *and continuous rate function*

$$R(t) := -\inf_{\lambda \in \mathbb{R}} \log E_0[\exp(\lambda(f(X-t))^{-1}f'(X-t))], t \in \mathbb{R}.$$

Then,

$$\lim_{n\to\infty} n^{-1} \log \mathbb{P}_\theta\{\theta - \hat{\theta}_n \notin C_{\mathrm{mle},L}(\hat{\theta}_n)\} = -S_{mle}(L)$$

where $S_{mle}(L) := \inf\{u \geq 0 : \int_{x:R(x)\leq u} 1\,dx \geq L\}$*. Besides,*

(i) For each $t \in \mathbb{R}$*,* $R(t) < \infty$*.*

(ii) $R(0) = 0$

(iii) R *is increasing in* $[0,\infty)$ *and decreasing in* $(-\infty, 0]$*.*

(iv) R *is continuous in* $\mathbb{R}$*.*

(v) There exists a $t_0 \in (0, L)$ *such that* $R(t) = R(t_0 - L) = S_{mle}(L)$*.*

Example 1. *If*

$$f(x) = (2\pi)^{-1/2}\sigma^{-1}\exp(-2^{-1}\sigma^{-2}x^2), x \in \mathbb{R}, \tag{12}$$

where $\sigma > 0$, then, it is easy to see that:

(i) $T_{HL,b} = T_{\mathrm{mle},b} = \hat{\theta}_n = \bar{X} := n^{-1}\sum_{j=1}^{n} X_j$.

(ii) $C_{HL,2b}(X_1, \ldots, X_n) = C_{\mathrm{mle},2b}(\hat{\theta}_n) = [\bar{X} - b, \bar{X} + b]$.

(iii) $S_{mle}(2b) = S(2b) = R(b) = 2^{-1}\sigma^{-2}b^2$.

(iv) For each $\theta \in \mathbb{R}$, $T_{HL,b} \xrightarrow{\mathbb{P}_\theta} \theta$.

Example 2. *If*

$$f(x) = (\Gamma(\alpha))^{-1}|\gamma|\alpha^{\alpha} \exp\left(\alpha\gamma x - \alpha e^{\gamma x}\right) \tag{13}$$

where $\alpha > 0$ *and* $\gamma \neq 0$, *then, it is possible to see that:*

(i) $\hat{\theta}_n = \gamma^{-1} \log\left(n^{-1} \sum_{j=1}^{n} e^{\gamma X_j}\right)$.

$$\begin{aligned} (ii) \qquad & C_{HL,2b}(X_1, \ldots, X_n) = C_{\mathrm{mle},2b}(\hat{\theta}_n) \\ = \ & [\hat{\theta}_n + \gamma^{-1} \log\left(2^{-1}\gamma^{-1}b^{-1}(e^{2\gamma b} - 1)\right) - 2b \\ & , \hat{\theta}_n + \gamma^{-1} \log\left(2^{-1}\gamma^{-1}b^{-1}(e^{2\gamma b} - 1)\right)]. \end{aligned}$$

(iii) $T_{HL,b} = T_{\mathrm{mle},b} = \hat{\theta}_n + \gamma^{-1} \log\left(2^{-1}\gamma^{-1}b^{-1}(e^{\gamma b} - e^{-\gamma b})\right)$.

(iv) For each $b > 0$,

$$\begin{aligned} & S(2b) = S_{mle}(2b) = R(b) \\ = \ & \alpha \log(2^{-1}|\gamma|^{-1}b^{-1}(e^{2|\gamma|b} - 1)) \\ & -\alpha(e^{2|\gamma|b} - 1)^{-1}(e^{2|\gamma|b} - 1 - 2|\gamma|b). \end{aligned}$$

(v) For each $b > 0$ *and each* $\theta \in \mathbb{R}$,

$$T_{HL,b} \xrightarrow{\mathbb{P}_\theta} \theta + \gamma^{-1} \log(2^{-1}\gamma^{-1}b^{-1}(e^{\gamma b} - e^{-\gamma b})) \neq \theta.$$

By Theorem 2 in Ferguson [1] the location families in examples 1 and 2 are the only one dimensional location families, which are exponential families.

Example 3. *If*

$$f(x) = c \exp(-a_1 \exp(\tau_1 x) - a_2 \exp(-\tau_2 x)), \qquad (14)$$

where $a_1, a_2, \tau_1, \tau_2 > 0$ *and* c *makes* f *a p.d.f. Then,*

(i) $(\sum_{j=1}^n e^{\tau_1 X_j}, \sum_{j=1}^n e^{-\tau_2 X_j})$ *is a minimal sufficient stat.*

(ii) The family $\{f(\cdot - \theta) : \theta \in \mathbb{R}\}$ *is a curved exponential family.*

(iii) The m.l.e. $\hat{\theta}_n$ *of* θ *is*

$$(\tau_1 + \tau_2)^{-1} \log \left(a_2^{-1} \tau_2^{-1} a_1 \tau_1 \left(\sum_{j=1}^n e^{-\tau_2 X_j} \right)^{-1} \sum_{j=1}^n e^{\tau_1 X_j} \right).$$

(iv)

$$C_{HL,2b}(X_1, \ldots, X_n) = C_{\text{mle},2b}(\hat{\theta}_n) = \hat{\theta}_n + [t_0 - 2b, t_0],$$

where

$$t_0 = (\tau_1 + \tau_2)^{-1} \log \left(\tau_1^{-1} \tau_2 (e^{2b\tau_1} - 1)(1 - e^{-2b\tau_2})^{-1} \right).$$

(v)

$$\begin{aligned} &T_{HL,b} = T_{\text{mle},b} = \hat{\theta}_n \\ &+(\tau_1 + \tau_2)^{-1} \log \left(\tau_1^{-1} \tau_2 (e^{b\tau_1} - e^{-b\tau_1})(e^{b\tau_2} - e^{-b\tau_2})^{-1} \right). \end{aligned}$$

Example 4. *If*

$$f(x) = 2^{-1}\exp(-|x|), x \in \mathbb{R}, \tag{15}$$

then, $\hat{\theta}_n$ *is not uniquely defined. The function* $\log f(\cdot)$ *is concave, but not strictly concave. Let* $X_{(1)}, \ldots, X_{(n)}$ *be the order statistics.* $X_1, \ldots, X_n$ *are all different wit probability one. Assume that* $X_1, \ldots, X_n$ *are all different. If* n *is odd,* $\hat{\theta}_n = X_{(2^{-1}(n+1))}$*. If* n *is even, then* $\hat{\theta}_n = X_{(2^{-1}n)}$ *and* $\hat{\theta}_n = X_{(2^{-1}n+1)}$ *are both m.l.e.'s. It is easy to see that whatever choice for the m.l.e. is made theorems 1–5 apply giving that*

(i) $C_{HL,2b}(X_1, \ldots, X_n) \neq C_{\mathrm{mle},2b}(\hat{\theta}_n)$.

(ii) For each $b > 0$, $S(2b) = b - \log(1 + b)$.

(iii) For each $b > 0$, $S_{mle}(2b) = b - 2^{-1}\log\left(2e^b - 1\right)$.

(iii) For each $b > 0$, $S_{mle}(2b) < S(2b)$.

(iv) For each $\theta \in \mathbb{R}$, $T_{HL,b} \xrightarrow{\mathbb{P}_\theta} \theta$.

Theorem 6. *Suppose that:*

(i) For each $x \in \mathbb{R}$, $f(x) > 0$.

(ii) $\log f$ *is a concave function.*

(iii) f *is even.*

Then,

(i) For each $b > 0$, $R(b) \leq S(2b)$.

(ii) Assume that f *has a third derivative and that* f *is not a p.d.f. from the families of p.d.f.'s in examples 1 and 3, then for some* $b \neq 0$, $R(b) < S(2b)$.

References

[1] Ferguson, T. S. (1962). Location and scale parameters in exponential families of distributions. *Ann. Mathemat. Statist.* **33** 986–1001.