Hajek - Le Cam Lower Bound in the zone of Moderate Deviation Probabilities

Mikhail Ermakov
St.Petersburg
E-mail: erm2512@gmail.com

The Hajek - Le Cam lower bound for the sharp asymptotic of moderate deviation probabilities of parameter estimators of signal and density is established. This lower bound can be interpreted as the lower bound for confidence estimation.

Small Probabilities in Statistics

I. Type I error probabilities in hypothesis testing.
$\alpha = 0.1; 0.05; 0.01.$

II. Significance levels in confidence estimation.

III. Statistical analysis of rare events.

Paradigm of asymptotic normality

sample size n several hundreds

$\Rightarrow$

Berry - Esseen inequality $\left(\frac{C}{\sqrt{n}}\right)$ does not allow to get sufficient accuracy

$\Rightarrow$

One needs to study normal approximation in large and moderate deviation zone

Bahadur efficiency

Let $X_1, \ldots, X_n$ i.i.d.r.v.'s having pm $P_\theta, \theta \in R^1$. Then for any sequence $b_n > 0, b_n \to 0, nb_n^2 \to \infty$ as $n \to \infty$ at the point $\theta_0 \in R^1$ there holds

$$\liminf_{n\to\infty} \inf_{|\theta-\theta_0|<2b_n} (\frac{1}{2} n b_n^2)^{-1} \ln P_\theta(|\widehat{\theta}_n - \theta| > b_n) \geq -I(\theta_0) \qquad (1)$$

Here we suppose that the family of pm P_θ have the finite Fisher information $I(\theta)$.

Sample mean estimation

Let $X_1, \ldots, X_n$ be i.i.d.r.v.'s. One needs to estimate $E[X]$. Let $\bar{X} = n^{-1}(X_1 + \ldots + X_n)$ is estimator of $E[X]$. Suppose that

$$E[\exp\{t|X_|\}] < C < \infty$$

i Then the Bernstein inequality holds

$$P(n^{1/2}(\bar{X} - E[X]) > x) < \exp\left\{-\frac{x^2}{2\sigma^2}\right\}(1 + o(1)), \quad x > x_0,$$

with $\sigma^2 = \mathrm{Var}[X_1]$.

Implementing the Bernstein inequality we get the following confidence interval

$$\left(\bar{X} - \frac{\sigma\sqrt{2|\ln(\alpha/2)|}}{\sqrt{n}}, \bar{X} + \frac{\sigma\sqrt{2|\ln(\alpha/2)|}}{\sqrt{n}}\right) \tag{2}$$

instead of the standard

$$\left(\bar{X} - x_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + x_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \tag{3}$$

Conclusion

The logarithmic asymptotic of moderate deviation probabilities gives rather rough description of limit distribution of statistics

Therefore it is of interest to get the sharp asymptotics for distribution of estimators and the sharp lower bounds for their efficiency

Setup. Estimation of signal parameter.

Let we observe a realization of random process $Y_\epsilon(t), t \in (0,1)$ defined by the stochastic differential equation

$$dY_\epsilon(t) = S(t,\theta)dt + \epsilon dw(t), \epsilon > 0 \tag{4}$$

Here $S \in L_2(0,1)$, $dw(t)$ is the Gaussian white noise and $\theta \in R^d$ is an unknown parameter.

Suppose that the signal S is differentiable on θ and there exists

$$\Psi(\theta) = \int_0^1 S_\theta(t,\theta) S'_\theta(t,\theta) dt < \infty \tag{5}$$

Assumptions

We fix $\lambda, 0 \leq \lambda \leq 1$.

For any $\theta_0 \in R^d$ and any $\theta \to \theta_0$ there holds

$$\int_0^1 (S(t,\theta) - S(t,\theta_0) - (\theta-\theta_0)' S_\theta(t,\theta_0))^2 dt = O(|\theta-\theta_0|^{2+\lambda}), \quad (6)$$

$$\int_0^1 (S(t,\theta) - S(t,\theta_0))^2 dt - (\theta-\theta_0)'\Psi(\theta_0)(\theta-\theta_0) = O(|\theta-\theta_0|^{2+\lambda}) \quad (7)$$

For any $v \in R^d$

$$v'\Psi(\theta)v - v'\Psi(\theta_0)v = O(|v|^2|\theta-\theta_0|^\lambda) \quad (8)$$

Main Result

Denote ζ- Gaussian random vector in R^d such that $E\zeta = 0, E[\zeta\zeta'] = I$. Here I is the unit matrix.

Theorem 1 *Assume (6)-(8). Let* $\epsilon^{-2}b_\epsilon^2 \to \infty, \epsilon^{-2}b_\epsilon^{2+\lambda} \to 0$ *as* $\epsilon \to 0$. *Let* $\Omega, \Omega \in R^d$, *be convex central-symmetric set. Then for any* $\theta_0 \in R^d$ *for any estimator* $\widehat{\theta}_\epsilon$

$$\liminf_{\epsilon \to 0} \sup_{|\theta-\theta_0|<C_\epsilon b_\epsilon} \frac{P_\theta(\Psi(\theta_0)^{-1/2}(\widehat{\theta}_\epsilon - \theta) \notin b_\epsilon\Omega)}{P(\zeta \notin \epsilon^{-1}b_\epsilon\Omega)} \geq 1, \qquad (9)$$

where $C_\epsilon \to \infty$ *as* $\epsilon \to 0$.

Main Steps of the Proof

We suppose that $\theta_0 = 0$.

1.Bayes Approach. We define a priori distribution on the grid Λ in the cube $K_\epsilon = (-v_\epsilon, v_\epsilon)^d, v_\epsilon = C_\epsilon b_\epsilon, C_\epsilon \to \infty, (C_\epsilon b_\epsilon)^{2+\lambda} \to 0$ as $\epsilon \to 0$. The step of the grid Λ equals $c_{1\epsilon}\epsilon^2 b_\epsilon^{-1}$ with $c_{1\epsilon} \to 0, c_{1\epsilon}^{-3}\epsilon^{-2} b_\epsilon^{2+\lambda} \to 0$ as $\epsilon \to 0$

2,We split the cube K_{v_ϵ} on small cubes $\Gamma_{\epsilon i} = x_{\epsilon i} + (-c_{2\epsilon}\epsilon^2 b_\epsilon^{-1}, c_{2\epsilon}\epsilon^2 b_\epsilon^{-1}]^d$ where $c_{2\epsilon} \to 0, c_{2\epsilon}c_{1\epsilon}^{-1} \to \infty, c_{1\epsilon}^{-3}\epsilon^{-2}b_\epsilon^{2+\lambda} \to 0$ as $\epsilon \to 0$

3. The fact that normalized posteriory Bayes risk go to constant as $\epsilon \to 0$, allows us to study the asymptotic behaviour of Bayes a posteriori risk independently for each event

$$\int_0^1 S_\theta(t,0)dw(t) \in \epsilon^{-1}\Gamma_{\epsilon i} \qquad (10)$$

4. For any $\theta_j, \theta_k \in \Lambda$ we show that

$$P\left(\left|\epsilon \int (S(t,\theta_k) - S(t,\theta_j) - (\theta_k - \theta_j)' S_\theta(t,\theta_j)) dw(t) \right.\right.$$
$$\left.\left. -\rho_\epsilon \int S_\theta(t,0) dw(t)\right| > \delta, \int S_\theta(t,0) dw(t) \in \epsilon^{-1}\Gamma_{\epsilon i}\right) \tag{11}$$
$$\leq C \int_{\Gamma_{\epsilon i}} \exp\{-\frac{|t|^2}{2\epsilon^2 ||S_\theta(t,0)||^2}\} dt \exp\{-c\frac{\delta^2}{|\theta_k - \theta_j|^{2+\lambda}\epsilon^2}\}$$

where

$$\rho_\epsilon = \epsilon^2 \int S_\theta(t,0)(S(t,\theta_k) - S(t,\theta_j) - (\theta_k - \theta_j) S_\theta(t,\theta_j)) dt ||S_\theta(t,0)||^{-2} \tag{12}$$

5. We show that

$$\rho'_\epsilon \int S_\theta(t,0)dw(t) < \delta_\epsilon \to 0 \tag{13}$$

as $\epsilon \to 0$ if $\int S_\theta(t,0)dw(t) \in \epsilon^{-1}\Gamma_{\epsilon i}$.

6. The statements (11),(13) together with "chaining method" allows us to write the standard local asymptotic normality approximation and to implement the technique of the standard proof of HL Theorem for obtaining moderate deviation version of this Theorem.

Main Result. Estimation of parameter of density

Let $X_1, \ldots, X_n$ be i.i.d.r.v.'s having pm P_θ, $\theta \in R^k$, defined on probability space (S, Υ).

Suppose that pms $P_\theta << \nu, \theta \in R^k$. Denote

$$f(x,\theta) = \frac{dP_\theta}{d\nu}(x), x \in S,$$

For all $\theta, t \in R^k$ denote $P^a_{\theta,t}$ and $P^s_{\theta,t}$ absolutely continuous and singular components of pm P_θ w.r.t. P_t. For all $x \in S$ denote

$$g(x,t,u) = (f(x,t+u)/f(x,t))^{1/2} - 1, u \in R^k.$$

if $f(x,t) \neq 0$.

Statistical experiment $\Psi = \{(S, \Upsilon), P_\theta, \theta \in R^k\}$ has the finite Fisher information at the point $t \in R^k$, if there exist such a function $\phi_t(x) = (\phi_{t1}(x), \ldots, \phi_{tk}(x))', x \in S$, that

$$\int_S (g(x,t,u) - u'\phi_t(x))^2 dP_t = o(|u|^2), \qquad P^s_{t+u,t}(S) = o(|u|^2).$$

as $u \to 0$.

The Fisher information at the point t equals

$$I(t) = 4 \int_S \phi_t \phi_t' dP_t.$$

For any $P_{\theta_1}, P_{\theta_2}, \theta_1, \theta_2 \in R^k$ the Hellinger distance P_{θ_1} and P_{θ_2} equals

$$\rho(P_{\theta_1}, P_{\theta_2}) = \rho(\theta_1, \theta_2) = \left(\int_S (f^{1/2}(x,\theta_1) - f^{1/2}(x,\theta_2))^2 \, d\nu \right)^{1/2}$$

Assumptions

A1. For all t in some vicinity of $\theta_0 \in R^k$ there exists the positive definite Fisher information $I(t)$.

A2. There exists $\epsilon > 0$ such that for all $u, |u| < \epsilon$ uniformly in t from some vicinity of θ_0 there holds

$$\int_S (g(x,t,u) - u'\phi_t(x))^2 \, dP_t < c|u|^{2+\lambda}, \tag{14}$$

$$|4\rho^2(t, t+u) - u'I(t)u| < C|u|^{2+\lambda}, \tag{15}$$

$$\int_S |\phi_t(x)|^{2+\lambda} \, dP_t < \infty. \tag{16}$$

A3. For any $u, h \in R^d$

$$u'I(\theta)u - u'I(\theta+h)u = O(|u|^2|h|^\lambda), \quad h \to 0. \tag{17}$$

We say that the set $\Omega \subset R^k$ is central symmetric if $x \in \Omega$ implies $-x \in \Omega$. For any set $\Omega \subset R^k$ and any $k \times k$-matrix A denote $A\Omega = \{y : y = Ax, x \in \Omega\}$.

Denote ζ- Gaussian random vector R^k having unit covariance matrix and $E\zeta = 0$.

Theorem 2 *Asume A1 - A3. Let* $nb_n^2 \to \infty, nb_n^{2+\lambda} \to 0$ *as* $n \to \infty$*. Then for any convex central-symmetric set* $\Omega \subset R^k$ *for any sequence of statistical estimators* $\widehat{\theta}_n = \widehat{\theta}_n(X_1, \ldots, X_n)$ *there holds*

$$\liminf_{n\to\infty} \sup_{|\theta-t|<C_n b_n} \frac{P_t(\widehat{\theta}_n - \theta \notin b_n\Omega)}{P(\zeta \notin n^{1/2} b_n I^{1/2}(\theta)\Omega)} \geq 1, \tag{18}$$

where $C_n \to \infty$ *as* $n \to \infty$*.*

On the problem of lower bounds in confidence estimation paid attention Wolfowitz

Wolfowitz J. Asymptotic efficiency of the maximum likelihood estimator. Theory Probab. Appl., 1965, v.10, .p.267-281.