Confidence regions of fixed volume based on the likelihood ratio test *

Miguel A. Arcones
Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902, USA
E-mail: arcones@math.binghamton.edu.

October 26, 2004

Abstract

Assuming that the data comes from a parametric family of d.f.'s, we build confidence regions for the unknown parameter of a fixed volume based on the likelihood ratio test. We study the asymptotics of the coverage probability of these regions in two different situations.

1 Introduction.

We consider the estimation of a finite dimensional parameter. Suppose that a random sample $X_1, \dots, X_n$ from a parametric family $\{f(x, \theta) : \theta \in \Theta\}$ of densities in $\mathbb{R}^d$, is observed, where Θ is a Borel set of $\mathbb{R}^d$. We construct confidence regions for θ with volume bounded by a constant such that the probability that the confidence region contains the unknown parameter is as large as possible. Usually, confidence regions are constructed so that the coverage probability is larger than a constant and letting the volume of the region as small as possible. Here, we proceed in opposite way.

We use standard notation. $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$. t' denotes the transpose of the vector t. X will denote a copy of X_1.

The constructed regions are based in inverting the likelihood ratio test. Constructing confidence regions inverting the likelihood ration test is a very well known elementary statistical procedure (see for example Section 9.2 Casella and Berger, 2001; and Section 7.1 in Shao,

* *Key words and phrases:* Parametric family, mle, likelihood ratio test, confidence region.

2003). In Brown, Cai and DasGupta (2003), it is argued that the likelihood ratio confidence regions are best overall for the exponential families.

Let

$$\hat{a}_n(\epsilon) := \sup\{\eta > 0 : \mu\{\theta \in \Theta : \sup_{t\in\Theta} (G_n(t) - G_n(\theta)) < \eta\} \le \epsilon\},$$

where μ is the Lebesgue measure in $\mathbb{R}^d$ and $G_n(t) = n^{-1}\sum_{j=1}^n \log f(X_j, t)$. By the monotone convergence theorem,

$$\mu\{\theta \in \Theta : \sup_{t\in\Theta} (G_n(t) - G_n(\theta)) < \hat{a}_n(\epsilon)\} \le \epsilon, \tag{1.1}$$

We take as a confidence region for θ,

$$C_{\mathrm{lrt}}(X_1, \dots, X_n, \epsilon) := \{\theta \in \Theta : \sup_{t\in\Theta} (G_n(t) - G_n(\theta)) < \hat{a}_n(\epsilon)\}. \tag{1.2}$$

This region has (Lebesgue measure) volume at most ϵ.

In this paper, we study the asymptotics of the coverage probability of the previous regions in different situations. When the volume of the region is of the order $n^{-d/2}$, the coverage probability converges to a nonzero limit, i.e. for each $\epsilon > 0$,

$$\mathbb{P}_\theta\{\theta \in C_{\mathrm{lrt}}(X_1, \dots, X_n, \epsilon n^{-d/2})\} \tag{1.3}$$

converges to a positive number, where $\mathbb{P}_\theta$ denotes the pm when the underlying r.v.'s has pdf $f(\cdot, \theta)$. Sufficient conditions to this to happen are presented in Section 2. Since mle's are asymptotically efficient, it is expected that the limit of the regions obtained using the asymptotic distribution of the mle have the most possible asymptotic coverage probability. We obtain that the limit in (1.3) agrees with the limit when the clt of the mle is used to construct the fixed volume confidence region.

In Section 3, we consider the limit of the coverage probabilities, when ϵ is a non–zero constant. In this section, we prove that, under regularity conditions,

$$n^{-1}\log(\mathbb{P}_\theta\{\theta \in C_{\mathrm{lrt}}(X_1, \dots, X_n, \epsilon)\}) \tag{1.4}$$

converges to a nonzero limit. For an regular exponential families, the limit in (1.4) agrees with the limit when the LDP of the mle is used to construct the fixed volume confidence region.

2 Asymptotics of coverage probabilities of the CLT type.

In this section, we study the asymptotic of the coverage probabilities of the confidence regions in (1.2) when the volume of the region converges to zero with the rate $n^{-d/2}$.

Theorem 2.1. *Let θ be in the interior of Θ. Suppose that the following conditions are satisfied:*

(i) There exists a function $\phi(\cdot, \theta) : \mathbb{R}^d \to \mathbb{R}^d$ such that $E_\theta[\phi(X, \theta)] = 0$, $E_\theta[|\phi(X, \theta)|^2] < \infty$ and for each $t \in \mathbb{R}^d$,

$$nE_\theta[|r_\theta(X, n^{-1/2}t)| \wedge |r_\theta(X, n^{-1/2}t)|^2] \to 0,$$

where

$$r_\theta(x,t) = \log f(x,\theta+t) - \log f(x,\theta) - t'\phi(x,\theta)$$

and E_θ indicates expectation when the underlying r.v. has pdf $f(\cdot,\theta)$.

(ii) For each $t \in \mathbb{R}^d$,

$$nE_\theta[\log f(X,\theta+n^{-1/2}t) - \log f(X,\theta)] \to -2^{-1}t'v(\theta)t,$$

where $v(\theta) = (v_{i,j}(\theta))_{1\le i,j\le \theta}$, $v_{i,j}(\theta) = E_\theta\left[\phi_i(X,\theta)\phi_j(X,\theta)\right]$ and $\phi(X,\theta) = (\phi_1(X,\theta),\dots,\phi_d(X,\theta))$.

(iii) For each $x \in \mathbb{R}^d$, $\log f(x,\cdot)$ is a strictly concave function.

(iv) Θ is a convex set.

Then,

$$\mathbb{P}_\theta\{\theta \in C_{\mathrm{lrt}}(X_1,\dots,X_n,\epsilon n^{-d/2})\} \to \mathbb{P}\{|Z_d| \le \epsilon^{1/d}c_d^{-1/d}(J((v(\theta))^{1/2}))^{1/d}\},$$

as $n\to\infty$, where Z_d is a standard $\mathbb{R}^d$–valued normal r.v., c_d is the volume of the unit ball of $\mathbb{R}^d$, and $J(A)$ denotes the Jacobian of the matrix A.

Observe that when $f(x,\theta)$ is first differentiable with respect to θ, $\phi(x,\theta)$ is the gradient with respect to θ of $\log f(x,\theta)$. The matrix $v(\theta)$ is known as the Fisher information matrix.

Hypotheses (i) in Theorem 2.1 hold under some minor smoothness and moment conditions on the function $\log f(x,\theta)$.

As it is well known, hypothesis (ii) in Theorem 2.1 holds if $f(x,\theta)$ is second differentiable with respect to θ in Θ, and it is possible to take derivatives inside the integral and to integrate by parts:

$$\tfrac{\partial}{\partial\theta^{(j)}}E_\theta[\log f(X,\theta)] = E_\theta[\tfrac{\partial}{\partial\theta^{(j)}}\log f(X,\theta)] = \textstyle\int_{\mathbb{R}^d}\tfrac{\partial}{\partial\theta^{(j)}}f(x,\theta)\,dx = \tfrac{\partial}{\partial\theta^{(j)}}\textstyle\int_{\mathbb{R}^d}f(x,\theta)\,dx = 0$$

and

$$\begin{aligned} & \tfrac{\partial^2}{\partial\theta^{(i)}\partial\theta^{(j)}}E_\theta[\log f(X,\theta)] = E_\theta[\tfrac{\partial^2\log f(X,\theta)}{\partial\theta^{(i)}\partial\theta^{(j)}}] = \textstyle\int_{\mathbb{R}^d}\tfrac{\partial^2\log f(x,\theta)}{\partial\theta^{(i)}\partial\theta^{(j)}}f(x,\theta)\,dx \\ = & -\textstyle\int_{\mathbb{R}^d}\tfrac{\partial\log f(x,\theta)}{\partial\theta^{(j)}}\tfrac{\partial f(x,\theta)}{\partial\theta^{(i)}}\,dx = -\textstyle\int_{\mathbb{R}^d}\tfrac{\partial\log f(x,\theta)}{\partial\theta^{(j)}}\tfrac{\partial\log f(x,\theta)}{\partial\theta^{(i)}}f(x,\theta)\,dx \\ = & -E_\theta[\tfrac{\partial\log f(X,\theta)}{\partial\theta^{(i)}}\tfrac{\partial\log f(X,\theta)}{\partial\theta^{(j)}}] = -v_{i,j}(\theta). \end{aligned}$$

Instead of using the likelihood ratio test, we could the confidence regions based on the asymptotics of an estimator. Let $T_n = T_n(X_1,\dots,X_n)$ be an estimator of θ. Suppose that when θ obtains, $(A(\theta))^{1/2}(T_n-\theta)$ converges in distribution to a $\mathbb{R}^d$–valued standard normal r.v., where $A(\theta)$ is a $d\times d$ matrix. Let

$$\hat{b}_n(\epsilon) := \sup\{\lambda > 0 : \mu\{\theta\in\Theta : |(A(T_n))^{1/2}(T_n-\theta)| < \lambda\} \le \epsilon\}.$$

Then,

$$\mu\{\theta\in\Theta : |(A(T_n))^{1/2}(T_n-\theta)| < \hat{b}_n(\epsilon)\} \le \epsilon.$$

We take as a confidence region for θ,

$$C_{T_n,\mathrm{clt}}(X_1,\dots,X_n,\epsilon) := \{\theta\in\Theta : |(A(T_n))^{1/2}(T_n-\theta)| < \hat{b}_n(\epsilon)\}.$$

This region has volume at most ϵ.

If T_n is in the interior of Θ and n is large enough, then $\hat{b}_n(\epsilon) = \epsilon^{1/d} c_d^{-1/d} (J((A(T_n))^{1/2}))^{1/d}$. Notice that by the change of variables $t = (A(T_n))^{1/2}(T_n - \theta)$,

$$\begin{aligned} & \int_{\mathbb{R}^d} I\{\theta \in \Theta : |(A(T_n))^{1/2}(T_n - \theta)| < \epsilon^{1/d} c_d^{-1/d} (J((A(T_n))^{1/2}))^{1/d}\} \, d\theta \\ = \; & (J((A(T_n))^{1/2}))^{-1} \int_{\mathbb{R}^d} I\{t \in \Theta : |t| < \epsilon^{1/d} c_d^{-1/d} (J((A(T_n))^{1/2}))^{1/d}\} \, dt = \epsilon \end{aligned}$$

Theorem 2.2. *Let θ be in the interior of Θ. Let T_n be an estimator of θ. Suppose that:*

(i) $A(\theta)$ is a nondegenerate matrix.

(ii) When θ obtains, $n^{1/2}(T_n - \theta)$ converges in distribution to a $\mathbb{R}^d$-valued normal r.v. with mean zero and covariance matrix $(A(\theta))^{-1}$.

(iii) $A(t)$, $t \in \Theta$, is continuous at θ.

Then, when θ obtains,

$$\mathbb{P}_\theta\{\theta \in C_{T_n,\mathrm{clt}}(X_1, \ldots, X_n, \epsilon n^{-d/2})\} \to \mathbb{P}\{|Z_d| \le \epsilon^{1/d} c_d^{-1/d} (J((A(\theta))^{1/2}))^{1/d}\},$$

as $n \to \infty$, where Z_d is a standard $\mathbb{R}^d$-valued normal r.v.

Let $\hat{\theta}_n$ be a r.v. such that

$$n^{-1} \sum_{j=1}^{n} \log f(X_j, \hat{\theta}_n) = \sup_{\theta \in \Theta} n^{-1} \sum_{j=1}^{n} \log f(X_j, \theta). \tag{2.1}$$

$\hat{\theta}_n$ is an mle of θ. Under regular conditions, the mle $\hat{\theta}_n$ satisfies the conditions in the previous theorem with $A(\theta) = v(\theta)$, where $v(\theta)$ is the Fisher information matrix (see e.g. Theorem 6.5.1 in Lehmann and Casella, 1998; or Theorem 7.12 in van der Vaart, 1998). Hence, we obtain the limits of the coverage probabilities in theorems 2.1 and 2.2 agree when the mle is used.

3 Asymptotics of coverage probabilities of the large deviations type.

Before presenting the results in this section, we recall some notation on the large deviation principle (LPD) of empirical processes. General references on the large deviation principle are Deuschel and Stroock (1989) and Dembo and Zeitouni (1998). We will use techniques from the LDP of stochastic processes in Arcones (2003a, 2003b). Given an index set T, $l_\infty(T)$ denotes the set of bounded functions in T with the norm $|z|_\infty := \sup_{t \in T} |z(t)|$. We say that a sequence of stochastic processes $\{U_n(t) : t \in T\}$ is said to follow the LDP in $l_\infty(T)$ with speed ϵ_n^{-1}, where $\{\epsilon_n\}$ is a sequence of positive numbers converging to zero, and with good rate function I if:

(i) For each $0 \le c < \infty$, $\{z \in l_\infty(T) : I(z) \le c\}$ is a compact set of $l_\infty(T)$.

(ii) For each set $A \subset l_\infty(T)$,

$$\begin{aligned} &-\inf\{I(z) : z \in A^o\} \le \liminf_{n\to\infty} \epsilon_n \ln(\Pr{}_*\{\{U_n(t) : t \in T\} \in A\}) \\ \le\ & \limsup_{n\to\infty} \epsilon_n \ln(\Pr{}^*\{\{U_n(t) : t \in T\} \in A\}) \le -\inf\{I(z) : z \in \bar{A}\}, \end{aligned}$$

where A^o (resp. $\bar{A}$) denotes the interior (resp. closure) of A in $l_\infty(T)$.

We determine the rate function of the LDP of empirical processes using Orlicz spaces theory. A reference in Orlicz spaces is Rao and Ren (1991). A function $\Phi : \mathbb{R} \to \bar{\mathbb{R}}$ is said to be a Young function if it is convex, $\Phi(0) = 0$; $\Phi(x) = \Phi(-x)$ for each $x > 0$; and $\lim_{x\to\infty} \Phi(x) = \infty$. Let X be a r.v. with values in a measurable space $(S, \mathcal{S})$. The Orlicz space $\mathcal{L}^\Phi(S, \mathcal{S})$ (abbreviated to $\mathcal{L}^\Phi$) associated with the Young function Φ is the class of measurable functions $f : (S, \mathcal{S}) \to \mathbb{R}$ such that $E[\Phi(\lambda f(X))] < \infty$ for some $\lambda > 0$. The Minkowski (or gauge) norm of the Orlicz space $\mathcal{L}^\Phi(S, \mathcal{S})$ by

$$N_\Phi(f) = \inf\{t > 0 : E[\Phi(f(X)/t)] \le 1\}.$$

It is well known that the vector space $\mathcal{L}^\Phi$ with the norm N_Φ is a Banach space. Define

$$\mathcal{L}^{\Phi_1} := \{f : S \to \mathbb{R} : E[\Phi_1(\lambda|f(X)|)] < \infty \text{ for some } \lambda > 0\},$$

where $\Phi_1(x) = e^{|x|} - |x| - 1$. Let $(\mathcal{L}^{\Phi_1})^*$ be the dual of $(\mathcal{L}^{\Phi_1}, N_{\Phi_1})$. The function $f \in \mathcal{L}^{\Phi_1} \mapsto \ln\left(E[e^{f(X)}]\right) \in \mathbb{R}$ is a convex lower semicontinuous function. The Fenchel–Legendre conjugate of the previous function is:

$$J(l) := \sup_{f\in\mathcal{L}^{\Phi_1}} \left(l(f) - \ln\left(E[e^{f(X)}]\right)\right),\ l \in (\mathcal{L}^{\Phi_1})^*. \tag{3.1}$$

J is a function with values in $[0, \infty]$. Since J is a Fenchel–Legendre conjugate, it is a nonnegative convex lower semicontinuous function. If $J(l) < \infty$, then:

(i) $l(\mathbf{1}) = 1$, where $\mathbf{1}$ denotes the function constantly 1.

(ii) l is a nonnegative definite functional: if $f(X) \ge 0$ a.s., then $l(f) \ge 0$.

Since the double Fenchel–Legendre transform of a convex lower semicontinuous function coincides with the original function (see e.g. Theorem 4.2.1 in Borwein and Lewis, 2000), we have that

$$\sup_{l\in\mathcal{L}^{\Phi_1}} (l(f) - J(l)) = \log E[e^{f(X)}].$$

The previous function J can be used to determine the rate function in the large deviation of statistics. Let $\{X_j\}_{j=1}^\infty$ be a sequence of i.i.d.r.v.'s with the distribution of X. If $f \in \mathcal{L}^{\Phi_1}$, then $\{n^{-1}\sum_{j=1}^n f(X_j)\}$ satisfies the LDP with rate function

$$I_f(t) := \sup_{\lambda\in\mathbb{R}} \left(\lambda t - \log\left(E[\exp(\lambda f(X))]\right)\right), t \in \mathbb{R} \tag{3.2}$$

(see for example Theorem 2.2.3 in Dembo and Zeitouni, 1998). By Lemma 2.2 in Arcones (2003b),

$$I_f(t) := \inf\left\{J(l) : l \in (\mathcal{L}^{\Phi_1})^*, l(f) = t\right\}. \tag{3.3}$$

It is well known that $I_f(\mu_f) = 0$, where $\mu_f = E[f(X)]$, I_f is convex, I_f is nondecreasing in $[\mu_f, \infty)$ and I is nonincreasing in $(-\infty, \mu_f]$ (see e.g. Lemma 2.2.5 in Dembo and Zeitouni, 1998). In particular, if $t \geq \mu_f$,

$$\inf\{J(l) : l \in (\mathcal{L}^{\Phi_1})^*, l(f) \geq t\} = I_f(t) \tag{3.4}$$

and for each $t \leq \mu_f$,

$$\inf\{J(l) : l \in (\mathcal{L}^{\Phi_1})^*, l(f) \leq t\} = I_f(t)$$

(see for example Corollary 2.2.19 in Dembo and Zeitouni, 1998).

Given functions $f_1, \ldots, f_m \in \mathcal{L}^{\Phi_1}$, then

$$\{(n^{-1}\sum_{j=1}^n f_1(X_j), \ldots, n^{-1}\sum_{j=1}^n f_m(X_j))\}$$

satisfies the LDP in $\mathbb{R}^m$ with speed n and rate function

$$I(u_1, \ldots, u_m) := \sup_{\lambda_1,\ldots,\lambda_m \in \mathbb{R}} \left(\sum_{j=1}^m \lambda_j u_j - \log E[\exp(\sum_{j=1}^m \lambda_j f_j(X))] \right) \tag{3.5}$$

(see for example Corollary 6.1.16 in Dembo and Zeitouni, 1998). This rate function can be written as

$$\inf\left\{J(l) : l \in (\mathcal{L}^{\Phi_1})^*, l(f_j) = u_j \text{ for each } 1 \leq j \leq m\right\}, \tag{3.6}$$

(see Lemma 2.2 in Arcones, 2003b).

When θ obtains, we denote $\mathcal{L}_\theta^{\Phi_1}$, $(\mathcal{L}_\theta^{\Phi_1})^*$ and J_θ.

Theorem 3.1. *Let $\{K_M\}_{M\geq 1}$ be a sequence of compact convex sets of $\mathbb{R}^d$ contained in Θ and containing θ. Suppose that the following conditions are satisfied:*

(i) Θ is a convex set of $\mathbb{R}^d$.

(ii) For each $t \in \Theta$, $\log f(X,t) \in \mathcal{L}_\theta^{\Phi_1}$.

(iii) For each x, $\log f(x,\cdot)$ is a strictly concave function.

(iv)

$$\lim_{M\to\infty} \sup_{t\in\partial K_M} \inf_{\lambda\in\mathbb{R}} E_\theta[\exp(\lambda(\log f(X,t) - \log f(X,\theta)))] = 0.$$

Then,

$$\begin{aligned} & -\inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, \textstyle\int_{\mathbb{R}^d} I(t \in \Theta : l(\log f(\cdot,t) - \log f(\cdot,\theta)) > 0)\, dt > \epsilon\} \\ \leq\ & \liminf_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{\theta \notin C_{\text{lrt}}(X_1,\ldots,X_n,\epsilon)\}\right) \\ \leq\ & \limsup_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{\theta \notin C_{\text{lrt}}(X_1,\ldots,X_n,\epsilon)\}\right) \\ \leq\ & -\inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, \textstyle\int_{\mathbb{R}^d} I(t \in \Theta : l(\log f(\cdot,t) - \log f(\cdot,\theta)) \geq 0)\, dt \geq \epsilon\}. \end{aligned} \tag{3.7}$$

We also can use the LDP of an estimator to find confidence regions for a parameter.

Theorem 3.2. *Let $\{T_n\}$ be a sequence of estimators of θ. Let $\theta \in \Theta$. Suppose that:*

(i) When $t \in \Theta$ obtains, T_n satisfies the LDP with speed n and rate function $I_t(\cdot)$.

(ii) $m_\theta(\cdot)$ is a continuous function, where

$$m_\theta(a) := \int_{\mathbb{R}^d} I(t \in \Theta : I_t(a) \le I_\theta(a))\, dt, a \in \Theta.$$

(iii) Given $a \in \Theta$ and $\epsilon > 0$ such that $m_\theta(a) < \epsilon$, then there exists $\theta_1 \in \Theta$ such that $I_\theta(a) < I_\theta(\theta_1)$ and

$$\int_\Theta I(t \in \Theta : I_t(a) < I_\theta(\theta_1))\, dt \le \epsilon.$$

Then,

$$\begin{aligned} & -\inf\{I_\theta(a) : m_\theta(a) > \epsilon\} \le \liminf_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{\theta \notin C_{T_n,ld}(X_1, \dots, X_n, \epsilon)\}\right) \\ \le\ & \limsup_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{\theta \notin C_{T_n,ld}(X_1, \dots, X_n, \epsilon)\}\right) \le -\inf\{I_\theta(a) : m_\theta(a) \ge \epsilon\}. \end{aligned} \tag{3.8}$$

where

$$C_{T_n,ld}(X_1, \dots, X_n, \epsilon) := \{t \in \Theta : I_t(T_n) < \hat{c}_n(\epsilon)\}$$

and

$$\hat{c}_n(\epsilon) = \sup\{\eta > 0 : \int_{\mathbb{R}^d} I(t \in \mathbb{R}^d : I_t(T_n) < \eta)\, dt \le \epsilon.$$

In Arcones (2003b), it was proved that the mle $\hat{\theta}_n$ satisfies the LDP with speed n and rate function

$$I_\theta(t) = -\inf_{\lambda \in \mathbb{R}^d} \log E_\theta[\exp(\lambda' \nabla_t \ln f(X, t))], \tag{3.9}$$

when θ obtains, where ∇_t denotes the (vector of partial derivatives) gradient of $\log f(x, t)$ with respect to the different coordinates of t.

Example 3.1. *(Exponential family). Given a measure μ on $\mathbb{R}^d$, define $\psi(t) := \ln \int_{\mathbb{R}^d} e^{t'x}\, d\mu(x)$, $t \in \mathbb{R}^d$. Let $\Theta := \{t \in \mathbb{R}^d : \psi(t) < \infty\}$. Let $f(x,t) := e^{t'x - \psi(t)}$. The family of pdf's $\{f(x,t) : t \in \Theta\}$ is a full exponential family with a canonical representation. The next theorem give manageable expressions for the rate functions in theorems 3.1 and 3.2 for an exponential family:*

Theorem 3.3. *(i) For each $\theta \in \Theta$, and each $\epsilon > 0$,*

$$\begin{aligned} & \inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, \mu\{t \in \Theta : l(\log f(\cdot, t) - \log f(\cdot, \theta)) \ge 0\} \ge \epsilon\} \\ =\ & \inf\{\sup_{t \in \mathbb{R}^d}(\psi(\theta) - \psi(t) - (\theta - t)'a) : a \in \mathbb{R}^d, \mu\{t \in \mathbb{R}^d : \psi(\theta) - \psi(t) - (\theta - t)'a \ge 0\} \ge \epsilon\}. \end{aligned}$$

(ii) For each $\theta \in \Theta$, and each $\epsilon > 0$,

$$\begin{aligned} & \inf\{I_\theta(a) : \mu\{t \in \Theta : I_t(a) \le I_\theta(a)\} \ge \epsilon\} \\ =\ & \inf\{\psi(\theta) - \psi(a) - (\theta - a)'\nabla\psi(a) : a \in \mathbb{R}^d, \\ & \quad \mu\{t \in \mathbb{R}^d : \psi(\theta) - \psi(t) - (\theta - t)'\nabla\psi(a) \ge 0\} \ge \epsilon\}. \end{aligned}$$

where $I_\theta(t)$ is as in (3.6) and $\nabla\psi$ is the gradient of ψ.

If $\{\nabla\psi(a) : a \in \mathbb{R}^d\} = \mathbb{R}^d$, *then*

$$\begin{aligned}
& \inf\{\sup_{t\in\mathbb{R}^d}(\psi(\theta)-\psi(t)-(\theta-t)'a) : a \in \mathbb{R}^d, \mu\{t \in \mathbb{R}^d : \psi(\theta)-\psi(t)-(\theta-t)'a \geq 0\} \geq \epsilon\} \\
= \ & \inf\{\sup_{t\in\mathbb{R}^d}(\psi(\theta)-\psi(t)-(\theta-t)'\nabla\psi(a)) : a \in \mathbb{R}^d, \\
& \quad \mu\{t \in \mathbb{R}^d : \psi(\theta)-\psi(t)-(\theta-t)'\nabla\psi(a) \geq 0\} \geq \epsilon\} \\
= \ & \inf\{\psi(\theta)-\psi(a)-(\theta-a)'\nabla\psi(a) : a \in \mathbb{R}^d, \\
& \quad \mu\{t \in \mathbb{R}^d : \psi(\theta)-\psi(a)-(\theta-a)'\nabla\psi(a) \geq 0\} \geq \epsilon\},
\end{aligned}$$

because by convexity

$$\begin{aligned}
& \sup_{t\in\mathbb{R}^d}(\psi(\theta)-\psi(t)-(\theta-t)'\nabla\psi(a)) \\
= \ & \psi(\theta)-\psi(a)-(\theta-a)'\nabla\psi(a)
\end{aligned}$$

Hence, if $\{\nabla\psi(a) : a \in \mathbb{R}^d\} = \mathbb{R}^d$, *the rate functions in theorem 3.1 and 3.2 agree for an exponential family.*

4 Proofs.

Proof of Theorem 2.1. If $\theta \in C_{\mathrm{lrt}}(X_1,\dots,X_n,\epsilon n^{-d/2})$, then $\sup_{t\in\Theta}(G_n(t)-G_n(\theta)) < \hat{a}_n(\epsilon n^{-m/2})$. Using the previous estimation and (1.1), we get that

$$\begin{aligned}
& C_{\mathrm{lrt}}(X_1,\dots,X_n,\epsilon n^{-d/2}) \qquad (4.1) \\
\subset \ & \{\theta \in \Theta : \textstyle\int_{\mathbb{R}^d} I(s \in \Theta : \sup_{t\in\Theta}(G_n(t)-G_n(s)) \leq \sup_{t\in\Theta}(G_n(t)-G_n(\theta))\, ds \leq \epsilon n^{-d/2}\} \\
= \ & \{\theta \in \Theta : \textstyle\int_{\mathbb{R}^d} I(s \in \Theta : G_n(s) \geq G_n(\theta))\, ds \leq \epsilon n^{-d/2}\} \\
= \ & \{\theta \in \Theta : \textstyle\int_{\mathbb{R}^d} I(t \in \mathbb{R}^d : \theta + n^{-1/2}t \in \Theta, G_n(\theta+n^{-1/2}t) \geq G_n(\theta))\, dt \leq \epsilon\}.
\end{aligned}$$

Hence, for each $0 < M < \infty$,

$$\begin{aligned}
& \mathbb{P}_\theta\{\theta \in C_{\mathrm{lrt}}(X_1,\dots,X_n,\epsilon n^{-d/2})\} \qquad (4.2) \\
\leq \ & \mathbb{P}_\theta\{\textstyle\int_{\mathbb{R}^d} I(t \in \mathbb{R}^d : \theta+n^{-1/2}t \in \Theta, G_n(\theta+n^{-1/2}t) \geq G_n(\theta))\, dt \leq \epsilon\} \\
\leq \ & \mathbb{P}_\theta\{\textstyle\int_{\mathbb{R}^d} I(t \in \mathbb{R}^d : \theta+n^{-1/2}t \in \Theta, |t| \leq M, G_n(\theta+n^{-1/2}t) \geq G_n(\theta))\, dt \leq \epsilon\} \\
& + \mathbb{P}_\theta\{G_n(\theta+n^{-1/2}t) \geq G_n(\theta) \text{ for some } t \text{ such that } \theta+n^{-1/2}t \in \Theta \text{ and } |t| > M\}
\end{aligned}$$

By hypothesis (i), for each $t \in \mathbb{R}^d$,

$$\begin{aligned}
& E_\theta[|\textstyle\sum_{i=1}^n (r_\theta(X_i,n^{-1/2}t)I(|r_\theta(X_i,n^{-1/2}t)| \leq 1) - E[r_\theta(X_i,n^{-1/2}t)I(|r_\theta(X_i,n^{-1/2}t)| \leq 1)])|] \\
\leq \ & 2nE[|r_\theta(X,n^{-1/2}t)|I(|r_\theta(X,n^{-1/2}t)| \leq 1)] \to 0.
\end{aligned}$$

and

$$\begin{aligned}
& \mathrm{Var}_\theta(\textstyle\sum_{i=1}^n (r_\theta(X_i,n^{-1/2}t)I(|r_\theta(X_i,n^{-1/2}t)| > 1) \\
& -E_\theta[r_\theta(X_i,n^{-1/2}t)I(|r_\theta(X_i,n^{-1/2}t)| > 1)])) \\
= \ & n\mathrm{Var}_\theta(r_\theta(X,n^{-1/2}t)I(|r_\theta(X,n^{-1/2}t)| > 1)) \\
\leq \ & nE_\theta[|r_\theta(X,n^{-1/2}t)|^2 I(|r_\theta(X,n^{-1/2}t)| > 1)] \to 0.
\end{aligned}$$

Hence, for each $t \in \mathbb{R}^d$,

$$\sum_{i=1}^{n}(r_\theta(X_i, n^{-1/2}t) - E_\theta[r_\theta(X_i, n^{-1/2}t)]) \xrightarrow{\mathbb{P}_\theta} 0. \tag{4.3}$$

By hypothesis (ii), for each $t \in \mathbb{R}^d$,

$$nE_\theta[r_\theta(X, n^{-1/2}t)] \to -2^{-1}t'v(\theta)t. \tag{4.4}$$

By hypothesis (i),

$$n^{-1/2}\sum_{j=1}^{n}(\phi(X_j, \theta) - E_\theta[\phi(X_j, \theta)]) \xrightarrow{\mathrm{d}} (v(\theta))^{1/2}Z_d, \tag{4.5}$$

when θ obtains. By (4.3)–(4.5), for each $t \in \mathbb{R}^d$,

$$\sum_{j=1}^{n}(\log f(X_j, \theta + n^{-1/2}t) - \log f(X_j, \theta)) \xrightarrow{\mathrm{d}} t'(v(\theta))^{1/2}Z_d - 2^{-1}t'v(\theta)t,$$

when θ obtains. By Theorem 1 in Arcones (1998), if the finite dimensional distributions of a sequence of concave stochastic processes defined in a subset of $\mathbb{R}^d$ converge, then the sequence of stochastic processes converges weakly uniformly over any compact set of its domain. Hence, for each $\theta \in \Theta$ and each $0 < M < \infty$,

$$\{\sum_{j=1}^{n}(\log f(X_j, \theta + n^{-1/2}t) - \log f(X_j, \theta)) : |t| \le M\} \xrightarrow{\mathrm{w}} \{t'(v(\theta))^{1/2}Z_d - 2^{-1}t'v(\theta)t : |t| \le M\}, \tag{4.6}$$

in $l_\infty(\{t \in \mathbb{R}^d : |t| \le M\})$, when θ obtains. Hence,

$$\begin{array}{ll} & \mathbb{P}_\theta\{\int_{\mathbb{R}^d} I(t \in \mathbb{R}^d : \theta + n^{-1/2}t \in \Theta, |t| \le M, G_n(\theta + n^{-1/2}t) \ge G_n(\theta))\, dt \le \epsilon\} \\ \to & \mathbb{P}\{\int_{\mathbb{R}^d} I(t \in \mathbb{R}^m : |t| \le M, t'(v(\theta))^{1/2}Z_d - 2^{-1}t'v(\theta)t \ge 0)\, dt \le \epsilon\}. \end{array} \tag{4.7}$$

By concavity, for n large enough,

$$\begin{array}{ll} & \mathbb{P}_\theta\{G_n(\theta + n^{-1/2}t) \ge G_n(\theta) \text{ for some } t \text{ such that } \theta + n^{-1/2}t \in \Theta \text{ and } |t| > M\} \\ = & \mathbb{P}_\theta\{G_n(\theta + n^{-1/2}t) \ge G_n(\theta) \text{ for some } t \text{ such that } \theta + n^{-1/2}t \in \Theta \text{ and } |t| = M\} \\ \to & \mathbb{P}\{t'(v(\theta))^{1/2}Z_d - 2^{-1}t'v(\theta)t \ge 0, \text{ for some } t \in \mathbb{R}^d \text{ such that } |t| = M\}. \end{array} \tag{4.8}$$

By (4.2) and (4.7)–(4.8), for each $0 < M < \infty$,

$$\begin{array}{ll} & \limsup_{n\to\infty} \mathbb{P}_\theta\{C_{\mathrm{lrt}}(X_1, \ldots, X_n, \epsilon n^{-d/2})\} \\ \le & \mathbb{P}\{\int_{\mathbb{R}^d} I(t \in \mathbb{R}^d : |t| \le M, t'(v(\theta))^{1/2}Z - 2^{-1}t'v(\theta)t \ge 0)\, dt \le \epsilon\} \\ & +\mathbb{P}\{t'(v(\theta))^{1/2}Z_d - 2^{-1}t'v(\theta)t \ge 0, \text{ for some } t \in \mathbb{R}^d \text{ such that } |t| = M\}. \end{array}$$

Letting $M \to \infty$, we get that

$$\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P}_\theta\{C_{\mathrm{lrt}}(X_1,\ldots,X_n,\epsilon n^{-d/2})\} \qquad (4.9)\\
\le\;& \mathbb{P}\{\textstyle\int_{\mathbb{R}^d} I(t\in\mathbb{R}^d : t'(v(\theta))^{1/2}Z_d - 2^{-1}t'v(\theta)t \ge 0)\,dt \le \epsilon\}\\
=\;& \mathbb{P}\{\textstyle\int_{\mathbb{R}^d} I(t\in\mathbb{R}^d : -2^{-1}|Z_d-(v(\theta))^{1/2}t|^2 + 2^{-1}|Z_d|^2 \ge 0)\,dt \le \epsilon\}\\
=\;& \mathbb{P}\{\textstyle\int_{\mathbb{R}^d} I(t\in\mathbb{R}^d : |Z_d-(v(\theta))^{1/2}t| \le |Z_d|)\,dt \le \epsilon\}\\
=\;& \mathbb{P}\{c_d|Z_d|^d(J((v(\theta))^{1/2}))^{-1} \le \epsilon\}\\
=\;& \mathbb{P}\{|Z_d| \le \epsilon^{1/d}c_d^{-1/d}(J((v(\theta))^{1/2}))^{1/d}\}.
\end{aligned}$$

Given $\theta\in\Theta$ such that $\int_{\mathbb{R}^d} I(s\in\Theta : G_n(s)\ge G_n(\theta))\,ds < \epsilon n^{-d/2}$, using concavity, there exists $\theta_1\in\Theta$ such that $G_n(\theta_1) < G_n(\theta)$ and

$$\int_{\mathbb{R}^d} I(s\in\Theta : G_n(s)\ge G_n(\theta_1))\,ds < \epsilon n^{-d/2}.$$

Hence,

$$\sup_{t\in\Theta}(G_n(t)-G_n(\theta)) < \sup_{t\in\Theta}(G_n(t)-G_n(\theta_1)) \le \hat{a}_n(\epsilon n^{-d/2}).$$

Therefore,

$$\{\theta\in\Theta : \int_{\mathbb{R}^d} I(s\in\Theta : G_n(s)\ge G_n(\theta))\,ds < \epsilon n^{-d/2}\} \subset C_{\mathrm{lrt}}(X_1,\ldots,X_n,\epsilon n^{-d/2}) \qquad (4.10)$$

Hence, for each $0<M<\infty$,

$$\begin{aligned}
&\mathbb{P}_\theta\{C_{\mathrm{lrt}}(X_1,\ldots,X_n,\epsilon n^{-d/2})\} \qquad (4.11)\\
\ge\;& \mathbb{P}_\theta\{\textstyle\int_{\mathbb{R}^d} I(t\in\Theta : G_n(t)\ge G_n(\theta))\,dt < \epsilon n^{-d/2}\}\\
=\;& \mathbb{P}_\theta\{\textstyle\int_{\mathbb{R}^d} I(t\in\mathbb{R}^d : \theta+n^{-1/2}t\in\Theta, G_n(\theta+n^{-1/2}t)\ge G_n(\theta))\,dt < \epsilon\}\\
\ge\;& \mathbb{P}_\theta\{\textstyle\int_{\mathbb{R}^d} I(t\in\mathbb{R}^d : \theta+n^{-1/2}t\in\Theta, |t|\le M, G_n(\theta+n^{-1/2}t)\ge G_n(\theta))\,dt < \epsilon\}\\
&-\mathbb{P}_\theta\{G_n(\theta+n^{-1/2}t)\ge G_n(\theta) \text{ for some } t \text{ such that } \theta+n^{-1/2}t\in\Theta \text{ and } |t|>M\}\\
\to\;& \mathbb{P}\{\textstyle\int_{\mathbb{R}^d} I(t\in\mathbb{R}^d : |t|\le M, t'(v(\theta))^{1/2}Z - 2^{-1}t'v(\theta)t\ge 0)\,dt < \epsilon\}\\
&-\mathbb{P}_\theta\{t'(v(\theta))^{1/2}Z - 2^{-1}t'v(\theta)t\ge 0 \text{ for some } t \text{ such that } \theta+n^{-1/2}t\in\Theta \text{ and } |t|>M\}.
\end{aligned}$$

Letting $M\to\infty$, we get that,

$$\begin{aligned}
&\liminf_{n\to\infty} \mathbb{P}_\theta\{C_{\mathrm{lrt}}(X_1,\ldots,X_n,\epsilon)\} \qquad (4.12)\\
\ge\;& \mathbb{P}\{\textstyle\int_{\mathbb{R}^d} I(t\in\mathbb{R}^d : t'(v(\theta))^{1/2}Z_d - 2^{-1}t'v(\theta)t\ge 0)\,dt < \epsilon\}\\
=\;& \mathbb{P}\{|Z_d| \le \epsilon^{1/d}c_d^{-1/d}(J((v(\theta))^{1/2}))^{1/d}\},
\end{aligned}$$

Finally, the claim follows from (4.9) and (4.12). $\square$

Proof of Theorem 2.2. Since $T_n \xrightarrow{\mathbb{P}_\theta} \theta$ and θ is in the interior of Θ, $\hat{b}_n(\epsilon) = \epsilon^{1/d} c_d^{-1/d} (J((A(T_n))^{1/2}))^{1/d}$ for n large enough, and with probability as close as one as wished. Hence,

$$\mathbb{P}_\theta\{\theta \in C_{\mathrm{T_n,clt}}(X_1, \ldots, X_n, \epsilon n^{-d/2})\}$$

is asymptotically equivalent to

$$\mathbb{P}_\theta\{|(A(T_n))^{1/2}(T_n - \theta)| < \epsilon^{1/d} n^{-1/2} c_d^{-1/d} (J((A(T_n))^{1/2}))^{1/d}\}$$

By hypotheses (ii) and (ii), the last quantity converges to

$$\mathbb{P}\{|Z_d| \le \epsilon^{1/d} c_d^{-1/d} (J((A(\theta))^{1/2}))^{1/d}\},$$

as $n \to \infty$. □

Proof of Theorem 3.1. Using (4.1), we get that for each $0 < M < \infty$,

$$\begin{aligned}
&\mathbb{P}_\theta\{\theta \notin C_{\mathrm{lrt}}(X_1, \ldots, X_n, \epsilon)\} \\
\ge\ & \mathbb{P}_\theta\{\textstyle\int_{\mathbb{R}^d} I(t \in \Theta : G_n(t) \ge G_n(\theta))\, dt > \epsilon\} \\
\ge\ & \mathbb{P}_\theta\{\textstyle\int_{\mathbb{R}^d} I(t \in K_M : G_n(t) \ge G_n(\theta))\, dt > \epsilon\} \\
=\ & \mathbb{P}_\theta\{\textstyle\int_{\mathbb{R}^d} I(t \in K_M : G_n(t) > G_n(\theta))\, dt > \epsilon\}
\end{aligned} \tag{4.13}$$

where the concavity of the function G_n has being used to get the last equation.

We claim that $\{G_n(t) : t \in K_M\}$ satisfies the LDP in $l_\infty(K_m)$. By (3.8) and hypothesis (ii), the finite dimensional distributions satisfy the LDP. Since the process $\{G_n(t) : t \in K_M\}$ is concave, by Corollary 3.5 in Arcones (2003c), $\{G_n(t) : t \in K_M\}$ satisfies the LDP in $l_\infty(K_m)$. Let

$$B_\infty(K_m) = \{x \in l_\infty(K_m) : x \text{ is Borel measurable}\}.$$

Since $B_\infty(K_m)$ is a closed set of $\{x \in l_\infty(K_m)$, $\{G_n(t) : t \in K_M\}$ satisfies the LDP in $l_\infty(K_m)$.

Next, we prove that the set

$$\{x \in B_\infty(K_M) : \mu\{t \in K_M : x(t) > x(\theta)\} > \epsilon\} \tag{4.14}$$

is an open set of $B_\infty(K_M)$. Given $x \in B_\infty(K_M)$ such that $\mu\{t \in K_M : x(t) > x(\theta)\} > \epsilon$, there exists an $\eta > 0$ such that $\mu\{t \in K_M : x(t) > 2\eta + x(\theta)\} > 2\eta + \epsilon$. Hence, if $\|y - x\|_{l_\infty(K_M)} < \eta$, then

$$\{t : y(t) > y(\theta)\} \supset \{t : x(t) \ge x(\theta) + \eta\}$$

and $\mu\{t \in K_M : y(t) > y(\theta)\} > \epsilon$. So, the set in (4.14) is an open set.

Using (4.13) and (4.14), we get that

$$\begin{aligned}
&-\inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, \mu\{t \in K_M : l(\log f(\cdot, t) - \log f(\cdot, \theta)) > 0\} > \epsilon\} \\
\le\ & \liminf_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{\theta \notin C_{\mathrm{lrt}}(X_1, \ldots, X_n, \epsilon)\}\right)
\end{aligned}$$

Letting $M \to \infty$, we get that

$$\begin{aligned}
& -\inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, \mu\{t \in \Theta : l(\log f(\cdot, t) - \log f(\cdot, \theta)) > 0\} > \epsilon\} \qquad (4.15) \\
\leq \quad & \liminf_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{\theta \notin C_{\mathrm{lrt}}(X_1, \ldots, X_n, \epsilon)\}\right)
\end{aligned}$$

Using (4.10),

$$\begin{aligned}
& \mathbb{P}_\theta\{\theta \notin C_{\mathrm{lrt}}(X_1, \ldots, X_n, \epsilon)\} \qquad (4.16) \\
\leq \quad & \mathbb{P}_\theta\{\mu\{t \in \Theta : G_n(t) \geq G_n(\theta)\} \geq \epsilon\} \\
\leq \quad & \mathbb{P}_\theta\{\mu\{t \in K_M : G_n(t) \geq G_n(\theta)\} \geq \epsilon\} \\
+ \quad & \mathbb{P}_\theta\{G_n(t) \geq G_n(\theta) \text{ for some } t \notin K_M\}
\end{aligned}$$

Next, we prove that the set

$$\{x \in B_\infty(K_M) : \mu\{t \in K_M : x(t) \geq x(\theta)\} \geq \epsilon\} \qquad (4.17)$$

is a closed set of $B_\infty(K_M)$. To prove that the set in (4.17) is a closed set, we show that if $\{x_n\}$ is a sequence in $B_\infty(K_M)$ such that $x_n \to x$ for some $x \in B_\infty(K_M)$, then x belongs to the set in (4.17). Let $A_n := \{t \in K_M : x_n(t) \geq x_n(\theta)\}$ and let $A := \{t \in K_M : x(t) \geq x(\theta)\}$. Since $\limsup_{n\to\infty} A_n \subset A$, using Fatou's lemma, we get that

$$\mu(A) \geq \mu(\limsup_{n\to\infty} A_n) \geq \limsup_{n\to\infty} \mu(A_n) \geq \epsilon.$$

Hence, x belongs to the set in (4.17). Therefore, the set in (4.17) is a closed.

Using that $\{G_n(t) : t \in K_M\}$ satisfies the LDP in $B_\infty(K_m)$, and the closedness of the set in (4.17), we get that

$$\begin{aligned}
& \limsup_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{\mu\{t \in K_M : G_n(t) \geq G_n(\theta)\} \geq \epsilon\}\right) \qquad (4.18) \\
\leq \quad & -\inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, \mu\{t \in K_M : l(\log f(\cdot, t) - \log f(\cdot, \theta)) \geq 0\} \geq \epsilon\}
\end{aligned}$$

Using the concavity of the function $G_n(\cdot)$,

$$\begin{aligned}
& \limsup_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{G_n(t) \geq G_n(\theta), \text{ for some } t \in K_M\}\right) \qquad (4.19) \\
= \quad & \limsup_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{G_n(t) \geq G_n(\theta), \text{ for some } t \in \partial K_M\}\right) \\
\leq \quad & -\inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, l(\log f(\cdot, t) - \log f(\cdot, \theta)) \geq 0, \text{ for some } t \in \partial K_M\}. \\
= \quad & -\inf_{t\in\partial K_M} \inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, l(\log f(\cdot, t) - \log f(\cdot, \theta)) \geq 0\}. \\
= \quad & \sup_{t\in\partial K_M} \inf_{\lambda\in\mathbb{R}}\{E_\theta[\exp(\lambda(\log f(X, t) - \log f(X, \theta)))] \\
\to \quad & -\infty, \text{ as } M \to \infty.
\end{aligned}$$

where we have used (3.4).

Using (4.16), (4.18) and (4.19), we get that

$$\begin{aligned}
& \limsup_{n\to\infty} n^{-1} \log\left(\mathbb{P}_\theta\{\theta \notin C_{\mathrm{lrt}}(X_1, \ldots, X_n, \epsilon)\}\right) \qquad (4.20) \\
\leq \quad & -\inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, \mu\{t \in \Theta : l(\log f(\cdot, t) - \log f(\cdot, \theta)) \geq 0\} \geq \epsilon\}
\end{aligned}$$

The claim of the theorem follows from (4.15) and (4.20). □

Proof of Theorem 3.2. By hypothesis (iii),

$$\begin{aligned} & \{\theta \in \Theta : m_\theta(T_n) < \epsilon\} \\ \subset & \{\theta \in C_{T_n,ld}(X_1,\dots,X_n,\epsilon)\}. \end{aligned}$$

Hence

$$\begin{aligned} & \mathbb{P}_\theta\{\theta \notin C_{T_n,ld}(X_1,\dots,X_n,\epsilon)\} \\ \leq & \mathbb{P}_\theta\{\theta \in \Theta : m_\theta(T_n) \geq \epsilon\}. \end{aligned}$$

Therefore, the upper inequality in (3.7) follows from the previous estimation and hypotheses (i) and (ii).

Since

$$\begin{aligned} & \{\theta \in C_{T_n,ld}(X_1,\dots,X_n,\epsilon)\} \\ \subset & \{\theta \in \Theta : \textstyle\int_{\mathbb{R}^d} I(t \in \Theta : I_t(T_n) \leq I_\theta(T_n))\, dt \leq \epsilon\} \\ = & \{\theta \in \Theta : m_\theta(T_n) \leq \epsilon\}, \end{aligned}$$

we have that

$$\begin{aligned} & \mathbb{P}_\theta\{\theta \notin C_{T_n,ld}(X_1,\dots,X_n,\epsilon)\} \\ \geq & \mathbb{P}_\theta\{\theta \in \Theta : m_\theta(T_n) > \epsilon\}. \end{aligned}$$

Hence, the lower inequality in (3.7) follows from the previous estimation and hypotheses (i) and (ii). □

Proof of Theorem 3.3. (i) Since

$$l(\log f(\cdot,t) - \log f(\cdot,\theta)) = (t-\theta)' l(x) - \psi(t) + \psi(\theta)$$

and

$$\begin{aligned} & \inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, l(x) = a\} = \sup_{\lambda \in \mathbb{R}^d}(\lambda' a - \log E_\theta[\exp(\lambda' X)]) \\ & = \sup_{\lambda \in \mathbb{R}^d}(\lambda' a - \psi(\lambda+\theta) + \psi(\theta)) = \sup_{t \in \mathbb{R}^d}(\psi(\theta) - \psi(t) - (\theta - t)' a), \end{aligned}$$

we have that

$$\begin{aligned} & \inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, \mu\{t \in \Theta : l(\log f(\cdot,t) - \log f(\cdot,\theta)) \geq 0\} \geq \epsilon\} \\ = & \inf\{\sup_{t \in \mathbb{R}^d}(\psi(\theta) - \phi(t) - (\theta-t)'a) : a \in \mathbb{R}^d, \mu\{t \in \mathbb{R}^d : \psi(\theta) - \psi(t) - (\theta - t)'a) \geq 0\} \geq \epsilon\}. \end{aligned}$$

(ii) For $t \in \Theta^o$ and $\theta \in \Theta$,

$$I_\theta(t) = \psi(\theta) - \psi(t) - (\theta - t)' \nabla \psi(t).$$

Hence,

$$\begin{aligned} & \inf\{I_\theta(a) : \mu\{t \in \Theta : I_t(a) \leq I_\theta(a)\} \geq \epsilon\} \\ = & \inf\{\psi(\theta) - \psi(a) - (\theta - a)' \nabla \psi(a) : a \in \mathbb{R}^d, \\ & \quad \mu\{t \in \mathbb{R}^d : \psi(\theta) - \psi(t) - (\theta - t)' \nabla \psi(a) \geq 0\} \geq \epsilon\}. \end{aligned}$$

□

References

[1] Arcones, M. A. (1998). On the weak convergence of convex stochastic processes. *Statistics and Probability Letters* **37** 171–182.

[2] Arcones, M. A. (2003a). The large deviation principle for stochastic processes I. *Theory of Probability and its Applications.* **47** 567-583.

[3] Arcones, M. A. (2003b). Large deviations for M–estimators. Preprint.

[4] Borwein, J. M. and Lewis, A. S. (2000). *Convex Analysis and Nonlinear Optimization: Theory and Examples*. Springer–Verlag, New York.

[5] Brown, L. D.; Cai, T. T.; and DasGupta, A. (2003). Interval estimation in exponential families. *Statistica Sinica* **13** 19–49.

[6] Casella, G. and Berger, R. L. (2001). *Statistical Inference.* Brooks Cole, Pacific Grove, California.

[7] Dembo, A. and Zeitouni, O. (1998). *Large Deviations Techniques and Applications*. Second Edition. Springer, New York.

[8] Deuschel, J. D. and Stroock, D. W. (1989). *Large Deviations*. Academic Press, Boston, MA.

[9] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation.* Second edition. Springer–Verlag, New York.

[10] Rao, M. M. and Ren, Z. D. (1991). *Theory of Orlicz Spaces.* Marcel Dekker, New York.

[11] Shao, J. (2003). *Mathematical Statistics, second edition*. Springer, New York.

[12] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.