Ann. Inst. Statist. Math.
Vol. 00, No. 0, 1–000 (0000)

LARGE DEVIATIONS FOR M–ESTIMATORS

Miguel A. Arcones

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

(Received August 9, 2004; revised November 29, 2004)

Abstract. We study the large deviation principle for M–estimators (and maximum likelihood estimators in particular). We obtain the rate function of the large deviation principle for M–estimators. For exponential families, this rate function agrees with the Kullback–Leibler information number. However, for location or scale families this rate function is smaller than the Kullback–Leibler information number. We apply our results to obtain confidence regions of minimum size whose coverage probability converges to one exponentially. In the case of full exponential families, the constructed confidence regions agree with the ones obtained by inverting the likelihood ratio test with a simple null hypothesis.

Key words and phrases: M–estimators, maximum likelihood estimators, large deviations, empirical processes, Kullback–Leibler information.

1. Introduction

We discuss the (LDP) large deviation principle for M–estimators. M-estimators have many good properties and they are used in many different situations. Their main property is that they are robust statistics. As an application, we obtain new results on the large deviations of (mle's) maximum likelihood estimators.

The large deviations of mle's have being considered by many authors. Let $\{f(\cdot,\theta):\theta\in\Theta\}$ be a family of pdf's, where Θ is a Borel subset of $\mathbb{R}^d$. Let $\{X_j\}_{j=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with a pdf belonging to $\{f(\cdot,\theta):\theta\in\Theta\}$. An mle $\hat\theta_n=\hat\theta_n(X_1,\dots,X_n)$ of θ is a value such that

$$\prod_{j=1}^n f(X_j,\hat\theta_n)=\sup_{\theta\in\Theta}\prod_{j=1}^n f(X_j,\theta).$$

Given an estimator T_n of a parameter θ, the error of the estimation is $|T_n-\theta|$. The probability that the error of estimation is bigger than ϵ is $\mathbb{P}_\theta\{|T_n-\theta|\ge\epsilon\}$. The limit

$$J(T_n,\epsilon,\theta):=\liminf_{n\to\infty} n^{-1}\ln\left(\mathbb{P}_\theta\{|T_n-\theta|>\epsilon\}\right). \tag{1.1}$$

is called the inaccuracy rate of the estimator T_n. In the one dimensional case, Bahadur (1967, 1971) proved that, if T_n is a consistent estimator of θ, then, for each $\theta\in\Theta$,

$$\lim_{\epsilon\to 0}\liminf_{n\to\infty}\epsilon^{-2}n^{-1}\ln\left(\mathbb{P}_\theta\{|T_n-\theta|\ge\epsilon\}\right)\ge -2^{-1}v(\theta), \tag{1.2}$$

where $v(\theta)$ is the Fisher information at θ and $\mathbb{P}_\theta$ is the probability when θ obtains, i.e.

$$v(\theta)=E_\theta\left[\left(\frac{\partial\ln f(X,\theta)}{\partial\theta}\right)^2\right]=-E_\theta\left[\frac{\partial^2\ln f(X,\theta)}{\partial\theta^2}\right].$$

Bahadur also proved that, under regularity conditions, for each θ

$$\lim_{\epsilon\to 0}\liminf_{n\to\infty} \epsilon^{-2}n^{-1}\ln\left(\mathbb{P}_\theta\{|\hat\theta_n-\theta|\ge\epsilon\}\right)=-2^{-1}v(\theta), \tag{1.3}$$

This shows that mle's are asymptotically efficient in the sense that they minimize the former limit.

Bahadur et al. (1979) showed that if T_n is a consistent estimator of θ, for each $\theta\in\Theta$, then, for each $\theta\in\Theta$,

$$\begin{aligned}&\liminf_{n\to\infty} n^{-1}\ln\left(\mathbb{P}_\theta\{|T_n-\theta|>\epsilon\}\right)\\ &\ge -\inf\{K(f(\cdot,\theta_1),f(\cdot,\theta)):\theta_1 \text{ satisfying } |\theta_1-\theta|>\epsilon\},\end{aligned} \tag{1.4}$$

where K is the Kullback–Leibler information of the densities $f(\cdot,\theta_1)$ and $f(\cdot,\theta)$, i.e. for densities f and g with respect to a probability measure μ,

$$K(f,g)=\int\ln(f(t)/g(t))f(t)\,d\mu(t).$$

In this situation, mle's are not optimal estimators. Kester and Kallenberg (1986) gave examples of mle's satisfying and not satisfying

$$\begin{aligned}&\liminf_{n\to\infty} n^{-1}\ln\left(\mathbb{P}_\theta\{|\hat\theta_n-\theta|>\epsilon\}\right)\\ &= -\inf\{K(f(\cdot,\theta_1),f(\cdot,\theta)):\theta_1 \text{ satisfying } |\theta_1-\theta|>\epsilon\}.\end{aligned} \tag{1.5}$$

For exponential families, there exists equality in the previous expression. We will prove that for location families which are not member of an exponential family, the previous equality does not hold.

Our techniques are based on the (LPD) large deviation principle of empirical processes. In Section 2, we present new results on the LPD for empirical processes with values $l_\infty(T)$, where T is an index set and $l_\infty(T)$ is the set of bounded functions in T with the norm $|z|_\infty:=\sup_{t\in T}|z(t)|$. 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 $A\subset l_\infty(T)$,

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

where A^o (resp. $\bar A$) denotes the interior (resp. closure) of A in $l_\infty(T)$ and $\mathbb{P}_*$ ($\mathbb{P}^*$) denotes the inner (outer) probability. General references on the LDP are Deuschel and Stroock (1989) and Dembo and Zeitouni (1998). The main property of the LDP is that it is closed by continuous functions: if $\{U_n(t):t\in T\}$ satisfies the LDP with speed ϵ_n^{-1} and good rate function I and $F:l_\infty(T)\to\mathbb{R}^d$ is a continuous function, then $F(\{U_n(t):t\in T\})$ satisfies the LDP with speed ϵ_n^{-1} and with good rate function

$$I_F(t)=\inf\{I(z):z\in l_\infty(T),F(z)=t\}$$

(see for example Lemma 2.1.4 in Deuschel and Stroock, 1989).

From the LDP of estimators, it is possible to obtain the inaccuracy rates of estimators. Suppose that a sequence of estimators $\{T_n\}$ satisfies the LDP with rate function $I_\theta(t)$, when θ obtains. Assuming that

$$\inf\{I_\theta(t) : t \text{ satisfying } |t-\theta| > \epsilon\} = \inf\{I_\theta(t) : t \text{ satisfying } |t-\theta| \geq \epsilon\},$$

we have that

$$J(T_n, \epsilon, \theta) = \inf\{I_\theta(t) : t \text{ satisfying } |t-\theta| \geq \epsilon\}. \tag{1.6}$$

In Section 3, we present sufficient condition to obtain the LDP for M–estimators. Let $g : S \times \Theta \to \mathbb{R}$ be a function such that $g(\cdot, t) : S \to \mathbb{R}$ is measurable for each $t \in \Theta$, where Θ be a Borel subset of $\mathbb{R}^d$. A natural estimator of a parameter $\theta \in \Theta$ such that $E[g(X,t) - g(X,\theta)] > 0$ for each $t \neq \theta$, is the estimator $\hat\theta_n$ such that

$$n^{-1}\sum_{j=1}^n g(X_j, \hat\theta_n) = \inf_{t\in\Theta} n^{-1}\sum_{j=1}^n g(X_j, t). \tag{1.7}$$

Since the estimator $\hat\theta_n$ is minimizing the stochastic process

$$\{n^{-1}\sum_{j=1}^n g(X_j, t) : t \in \Theta\}, \tag{1.8}$$

it is expected that, under certain conditions, $\hat\theta_n$ satisfies the LDP with the rate function

$$I_{\hat\theta}(t) = \inf\{I_g(z) : z \in l_\infty(\Theta), z(\cdot) \text{ is minimized at } t\},$$

where I_g is the rate function of the LDP of the sequence of stochastic processes in (1.8). Heuristically, this is true because in some sense, the function which assigns to a function the value where the minimum of the function is attained is a continuous function.

We also consider M–estimators $\hat\theta_n$ defined by

$$n^{-1}\sum_{j=1}^n h(X_j, \hat\theta_n) = 0, \tag{1.9}$$

where $h(\cdot, t) : S \to \mathbb{R}^d$ is a measurable function for each $t \in \Theta$. Here, $\hat\theta_n$ is estimating a value θ characterized by $E[h(X,\theta)] = 0$. In this case, it is expected that, under certain conditions, $\hat\theta_n$ satisfies the LDP with the rate function

$$I_{\hat\theta}(t) = \inf\{I_h(z) : z \in l_\infty(\Theta), z(t) = 0\},$$

where I_h is the rate function of the LDP of $\{n^{-1}\sum_{j=1}^n h(X_j, t) : t \in \Theta\}$. We will show that

$$I_{\hat\theta}(t) = -\inf_{\lambda\in\mathbb{R}^d} \ln E[\exp(\lambda' h(X,t))]. \tag{1.10}$$

For some one dimensional M–estimators, Sievers (1978) and Rubin and Rukhin (1983) obtained that the rate of certain M-estimators is given by (1.10). Fu et al. (1993) obtained much more general results for the large deviation of one dimensional mle's. Kester (1985) and Kester and Kallenberg (1986) found the inaccuracy rates of mle's from an exponential family. Borovkov and Mogulskii (1992) gave upper and lower bounds for the large deviations of M-estimators parameterized by a compact set. Joutard (2004) considered the large deviations of M–estimators over a sequence of necessarily identically distributed sequence of r.v.'s when the parameter set is compact. Our results apply non necessarily compact parameter sets.

When applied to the mle's, we obtain that, under certain conditions, when θ obtains, the mle $\hat{\theta}_n$ satisfies the LDP with speed n and rate function

$$I_\theta(t) := -\inf_{\lambda\in\mathbb{R}^d} \ln E_\theta[\exp(\lambda'\nabla_t \ln f(X,t))], \tag{1.11}$$

where ∇_t denotes the (vector of partial derivatives) gradient of $\ln f(x,t)$. We prove that for each $t,\theta\in\Theta$, $I_\theta(t)\le K(f(\cdot,t),f(\cdot,\theta))$. If for each $t,\theta\in\Theta$, $I_\theta(t)=K(f(\cdot,t),f(\cdot,\theta))$, then the mle minimizes the limit in (1.4) among all possible estimators. However, in general $I_\theta(t)<K(f(\cdot,t),f(\cdot,\theta))$, for $t\ne\theta$. Theorem 3.2 determines when $I_\theta(t)=K(f(\cdot,t),f(\cdot,\theta))$. For an exponential family, we have that $I_\theta(t)=K(f(\cdot,t),f(\cdot,\theta))$. The only location families for which $I_\theta(t)=K(f(\cdot,t),f(\cdot,\theta))$, for each $t,\theta\in\Theta$, are the ones which are exponential families.

In Section 4, we apply the results in Section 3 to obtain confidence regions whose coverage probability approaches to 1 exponentially. Suppose that, when θ obtains, $\hat{\theta}_n$ satisfies the LDP with speed n and rate function $I_\theta(\cdot)$. Given $0<\alpha<\infty$, define $U_{\theta,\alpha}:=\{t\in\mathbb{R}^d : I_\theta(t)<\alpha\}$ and

$$C_\alpha(X_1,\dots,X_n) := \{\theta\in\Theta : \hat{\theta}_n(X_1,\dots,X_n)\in U_{\theta,\alpha}\}. \tag{1.12}$$

Assuming that $\{t\in\mathbb{R}^d : I_\theta(t)<\alpha\}$ is an open set, then $C_\alpha(X_1,\dots,X_n)$ is a confidence region for Θ such that for each $\theta\in\Theta$,

$$\begin{aligned}(1.13)\qquad &\limsup_{n\to\infty} n^{-1}\ln\left(P_\theta\{\theta\notin C_\alpha(X_1,\dots,X_n)\}\right)\\ &= \limsup_{n\to\infty} n^{-1}\ln\left(P_\theta\{I_\theta(\hat{\theta}_n)\ge\alpha\}\right)\le -\inf\{I_\theta(t): I_\theta(t)\ge\alpha\}\le -\alpha.\end{aligned}$$

The confidence regions obtained in this way have some minimality properties. In some sense, they are the smallest regions based on $\hat{\theta}_n$ satisfying (1.13). Suppose that given θ, $G_{\theta,\alpha}$ is a set such that $\{t\in\mathbb{R}^d : I_\theta(t)<\alpha\}\not\subset\overline{G_{\theta,\alpha}}$, then,

$$\liminf_{n\to\infty} n^{-1}\ln\left(P_\theta\{\hat{\theta}_n(X_1,\dots,X_n)\notin\overline{G_{\theta,\alpha}}\ \}\right)\ge -\inf\{I_\theta(t): t\notin\overline{G_{\theta,\alpha}}\ \}>-\alpha.$$

Hence, if

$$\liminf_{n\to\infty} n^{-1}\ln\left(P_\theta\{\hat{\theta}_n(X_1,\dots,X_n)\notin G_{\theta,\alpha}\}\right)\le -\alpha,$$

then $\{t\in\mathbb{R}^d : I_\theta(t)<\alpha\}\subset\overline{G_{\theta,\alpha}}$. Assuming that $\{t\in\mathbb{R}^d : I_\theta(t)<\alpha\}$ is an open set, we have that $\{t\in\mathbb{R}^d : I_\theta(t)<\alpha\}\subset(\overline{G_{\theta,\alpha}})^o$. When the mle's are sufficient statistics, once should expect that the regions in (1.12) are the smallest regions over all the regions satisfying (1.13).

The classical asymptotic confidence intervals are constructed fixing the coverage probability to a fixed number less than one and letting the size of the region go to zero as $n\to\infty$. The procedure here is opposite. We allow the size of the confidence region do not go to zero, but the coverage probability goes to one as the sample size goes to infinity. In the case of full exponential families, the constructed confidence regions agree with the ones obtained by inverting the likelihood ratio test for a simple null hypothesis.

In Brown et al. (2003), it is argued that for a big group of exponential families the confidence intervals obtained by inverting the likelihood ratio test are best overall. Our results complement the results of these authors. These authors study the size of the confidence regions when the coverage probability is constant.

Large deviations have many applications in statistics. Large deviations are used in some definitions of efficiency (see Bahadur 1971; Serfling, 1980; Nikitin, 1995). Often in sequential analysis, it is of interest to use confidence intervals of fixed length. Fu (1975) proved that the limits of the density of a sequence of estimators is related with their large deviations. Jensen and Wood (1998) have used the large deviations of mle's to study the density of mle's (see also Skovgaard, 1990).

The proofs of the theorems in sections 2–4 are in Section 5. We will use usual multivariate notation. For example, given $u = (u_1, \dots, u_d)' \in \mathbb{R}^d$ and $v = (v_1, \dots, v_d)' \in \mathbb{R}^d$, $u'v = \sum_{j=1}^d u_j v_j$ and $|u| = (\sum_{j=1}^n u_j^2)^{1/2}$. Given $\theta \in \mathbb{R}^d$ and $\epsilon > 0$, $B(\theta, \epsilon) = \{x \in \mathbb{R}^d : |x - \theta| < \epsilon\}$. Given a $d \times d$ matrix A, $\|A\| = \sup_{v_1, v_2 \in \mathbb{R}^d, |v_1|, |v_2| = 1} v_1' A v_2$. Given a rate function I and a set A, we denote $I(A) = \inf\{I(x) : x \in A\}$.

2. LDP of empirical processes

In this section, we study the LDP of empirical processes. Let $\{X_j\}_{j=1}^\infty$ be a sequence of i.i.d.r.v.'s with values in a measurable space $(S, \mathcal{S})$. Let $\{f(\cdot, t) : t \in T\}$ be a collection of measurable functions on $(S, \mathcal{S})$, where T is an index set. Let X be a copy of X_1. Necessary and sufficient conditions for the LDP of empirical processes $\{n^{-1} \sum_{j=1}^n f(X_j, t) : t \in T\}$ with speed n were given in Arcones (2003a). However, we need to represent the rate function in a convenient way. Our method, using the dual (vector space of continuous linear functionals) of certain Orlicz space, is a variation of the method used by Léonard and Najim (2002) to determine the rate function of the LDP of empirical measures. We refer to the theory in Orlicz spaces to 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 $\lim_{x \to \infty} \Phi(x) = \infty$. 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 $\lambda > 0$. Let Ψ be the Fenchel–Legendre conjugate of Φ, i.e. $\Psi(x) = \sup_{y \in \mathbb{R}} (xy - \Phi(y))$. The Minkowski (or gauge) norm of the Orlicz space $\mathcal{L}^\Phi(S, \mathcal{S})$ is defined 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.

In the case of large deviations, we have that given functions $f_1, \dots, f_m$ such that for some $\lambda > 0$ and each $1 \le k \le m$, $E[\exp(\lambda |f_k(X)|)] < \infty$, then

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

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

$$I(u_1, \dots, u_m) = \sup_{\lambda_1, \dots, \lambda_m \in \mathbb{R}} \left(\sum_{j=1}^m \lambda_j u_j - \ln E[\exp(\sum_{j=1}^m \lambda_j f_j(X))] \right) \tag{2.1}$$

(see Corollary 6.1.6 in Dembo and Zeitouni, 1998). We will work in the space

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

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. Observe that if $f_n \overset{\mathcal{L}^{\Phi_1}}{\to} f$, then $f_n(X) \overset{\mathbb{P}}{\to} f(X)$, which, by the Fatou's lemma, implies that $\ln\left(E[e^{f(X)}]\right) \le \liminf_{n\to\infty} \ln\left(E[e^{f_n(X)}]\right)$. The Fenchel–Legendre conjugate of the previous function is:

$$J(l) := \sup_{f \in \mathcal{L}^{\Phi_1}} \left(l(f) - \ln\left(E[e^{f(X)}]\right)\right),\ l \in (\mathcal{L}^{\Phi_1})^*. \tag{2.2}$$

J is a function with values in $[0, \infty]$. Since J is a Fenchel–Legendre conjugate, it is a nonnegative convex lower semicontinuous function. It is easy to see that 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. Lemma 4.5.8 in Dembo and Zeitouni, 1998), we have that

$$\sup_{l \in (\mathcal{L}^{\Phi_1})^*} (l(f) - J(l)) = \ln E[e^{f(X)}]. \tag{2.3}$$

We also will consider the convex function $\Phi_2(x) = e^x - 1$. The Fenchel–Legendre conjugate of Φ_2 is

$$\Psi_2(x) = x \ln\left(\frac{x}{e}\right) + 1, \text{ if } x > 0; \Psi_2(0) = 1; \quad \text{and } \Psi_2(x) = \infty, \quad \text{if } x < 0. \tag{2.4}$$

We also have the following:

Lemma 2.1. *If $l \in (\mathcal{L}^{\Phi_1})^*$ and $l(\mathbf{1}) = 1$, then*

$$J(l) = \sup_{f \in \mathcal{L}^{\Phi_1}} \left(l(f) - E[e^{f(X)} - 1]\right).$$

Given a nonnegative function γ on S such that $E[\gamma(X)] = 1$ and $E[\Psi_2(\gamma(X))] < \infty$, then $l_\gamma(f) = E[f(X)\gamma(X)]$ defines a continuous linear functional in $\mathcal{L}^{\Phi_1}$. Besides, it is easy to see that

$$J(l_\gamma) = \sup_{f \in \mathcal{L}^{\Phi_1}} E[f(X)\gamma(X) - \Phi_2(f(X))] = E[\Psi_2(\gamma(X))]. \tag{2.5}$$

Observe that by the Fenchel–Young inequality,

$$\sup_{f \in \mathcal{L}^{\Phi_1}} E[f(X)\gamma(X) - \Phi_2(f(X))] \le E[\Psi_2(\gamma(X))].$$

Given $1 < M < \infty$, taking $f(x) = \ln(\gamma(x)) I(M^{-1} \le \gamma(x) \le M)$ and letting $M \to \infty$, we get that

$$\sup_{f \in \mathcal{L}^{\Phi_1}} E[f(X)\gamma(X) - \Phi_2(f(X))] \ge E[\Psi_2(\gamma(X))].$$

Hence, (2.5) follows. $E[\Psi_2(\gamma(X))]=E[\gamma(X)\ln(\gamma(X))]$ is the Kullback-Leibler information number of the probability measures $\gamma(\cdot)\,d\mu(\cdot)$ and $d\mu(\cdot)$, where $\mu(\cdot)$ is the distribution of X. But, the set $\{l_\gamma\in(\mathcal{L}^{\Phi_1})^*:E[\gamma(X)]=1,E[\Psi_2(\gamma(X))]<\infty\}$ does not have the compactness properties that $\{l\in(\mathcal{L}^{\Phi_1})^*:J(l)<\infty\}$ has.

We may express, the rate function in (2.1) using the function J:

Lemma 2.2. *Let $f_1,\dots,f_m\in\mathcal{L}^{\Phi_1}$. Then, for each $u_1,\dots,u_m\in\mathbb{R}$,*

$$\begin{aligned}&\sup\left\{\textstyle\sum_{j=1}^m\lambda_j u_j-\ln\left(E\left[\exp\left(\sum_{j=1}^m\lambda_j f_j(X)\right)\right]\right):\lambda_1,\dots,\lambda_m\in\mathbb{R}\right\}\\ =\ &\inf\left\{J(l):l\in(\mathcal{L}^{\Phi_1})^*,l(f_j)=u_j\text{ for each }1\le j\le m\right\}.\end{aligned}$$

In the case of one function f, the rate function is

$$I_f(t):=\inf\left\{J(l):l\in(\mathcal{L}^{\Phi_1})^*,l(f)=t\right\}=\sup\{\lambda t-\ln(E[\exp(\lambda f(X))]):\lambda\in\mathbb{R}\}.$$

Let $\mu_f=E[f(X)]$. Let $b_f=\inf\{t\in\mathbb{R}:P(f(X)>t)=0\}$ be the least upper a.s. bound of $f(X)$, where $\inf\emptyset$ is interpreted as ∞. Let $a_f=\sup\{t\in\mathbb{R}:P(f(X)<t)=0\}$ be the most lower a.s. bound of $f(X)$, where $\sup\emptyset$ is interpreted as $-\infty$. It is well known that I_f is convex in $\mathbb{R}$, I_f is continuous in $[a_f,b_f]$, I_f is infinity in $\mathbb{R}-[a_f,b_f]$, $I_f(\mu_f)=0$, $I_f(a_f)=-\ln P(f(X)=a_f)$, $I_f(b_f)=-\ln P(f(X)=b_f)$, I_f is nondecreasing in $[\mu_f,\infty)$ and I_f is nonincreasing in $(-\infty,\mu_f]$ (see Lemma 6 in Chernoff, 1952). This implies that for $t\ge\mu_f$,

$$\inf\{J(l):l\in\left(\mathcal{L}^{\Phi_1}\right)^*,l(f)\ge t\}=\inf\{J(l):l\in\left(\mathcal{L}^{\Phi_1}\right)^*,l(f)=t\}\tag{2.6}$$

and for $t\le\mu_f$,

$$\inf\{J(l):l\in\left(\mathcal{L}^{\Phi_1}\right)^*,l(f)\le t\}=\inf\{J(l):l\in\left(\mathcal{L}^{\Phi_1}\right)^*,l(f)=t\}\tag{2.7}$$

By Theorem 1 in Chernoff (1952), for each $t\ge\mu_f$,

$$\lim_{n\to\infty}n^{-1}\ln\left(\mathbb{P}\left\{n^{-1}\sum_{j=1}^n f(X_j)\ge t\right\}\right)=-I_f(t)\tag{2.8}$$

and for each $t\le\mu_f$,

$$\lim_{n\to\infty}n^{-1}\ln\left(\mathbb{P}\left\{n^{-1}\sum_{j=1}^n f(X_j)\le t\right\}\right)=-I_f(t).\tag{2.9}$$

The previous limits and the continuity of the function I_f imply that for each $b_f>t\ge\mu_f$,

$$\lim_{n\to\infty}n^{-1}\ln\left(\mathbb{P}\left\{n^{-1}\sum_{j=1}^n f(X_j)>t\right\}\right)=-I_f(t)\tag{2.10}$$

and for each $a_f<t\le\mu_f$,

$$\lim_{n\to\infty}n^{-1}\ln\left(\mathbb{P}\left\{n^{-1}\sum_{j=1}^n f(X_j)<t\right\}\right)=-I_f(t).\tag{2.11}$$

By Lemma 1 in Chernoff (1952), if $P(f(X)<0)>0$ and $P(f(X)>0)>0$, then there exists $\lambda_0 \in \mathbb{R}$ such that

$$-\ln E[\exp(\lambda_0 f(X))] = \inf_{\lambda\in\mathbb{R}} \left(-\ln E[\exp(\lambda f(X))]\right). \tag{2.12}$$

We also will use that by the Chebyshev inequality, we also have that for each $t \in \mathbb{R}$,

$$n^{-1}\ln\left(\mathbb{P}\{n^{-1}\sum_{j=1}^n f(X_j)\ge t\}\right) \le -\sup_{\lambda>0}(\lambda t - \ln\left(E[\exp(\lambda f(X))]\right)). \tag{2.13}$$

As to the LDP of empirical processes. If $\{n^{-1}\sum_{j=1}^n f(X_j,t) : t\in T\}$ satisfies the LDP in $l_\infty(T)$ with speed n, then the rate function is

$$I(z) := \sup\{I_{t_1,\dots,t_m}(z(t_1),\dots,z(t_m)) : t_1,\dots,t_m \in T, m\ge 1\}, z\in l_\infty(T), \tag{2.14}$$

where

$$I_{t_1,\dots,t_m}(u_1,\dots,u_m) = \sup_{\lambda_1,\dots,\lambda_m\in\mathbb{R}}\left(\sum_{j=1}^m \lambda_j u_j - \ln E[\exp(\sum_{j=1}^m \lambda_j f(X,t_j))]\right) \tag{2.15}$$

(see Arcones, 2003a). The next lemma shows that this rate can be represented using the function J:

Lemma 2.3. *Let I and let $I_{t_1,\dots,t_m}$ be as in (2.14) and (2.15). If $\{f(\cdot,t) : t\in T\}$ is a totally bounded set of $(\mathcal{L}^{\Phi_1}, N_{\Phi_1})$, then:*

(i) For each $z\in l_\infty(T)$,

$$\begin{aligned}&\sup\{I_{t_1,\dots,t_m}(z(t_1),\dots,z(t_m)) : t_1,\dots,t_m\in T, m\ge 1\}\\ =\ &\inf\{J(l) : l\in(\mathcal{L}^{\Phi_1})^*, l(f(\cdot,t)) = z(t), \text{ for each } t\in T\}.\end{aligned}$$

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

(iii) For each $t_1,\dots,t_m\in T$ and each $u_1,\dots,u_m\in\mathbb{R}$,

$$I_{t_1,\dots,t_m}(u_1,\dots,u_m) = \inf\{I(z) : z(t_j) = u_j \text{ for each } 1\le j\le m\}.$$

The total boundedness condition in the previous lemma is best in the following sense:

Lemma 2.4. *Let $\{f(\cdot,t) : t\in T\}$ be a collection of functions of $(\mathcal{L}^{\Phi_1}, N_{\Phi_1})$. Let I and let $I_{t_1,\dots,t_m}$ be as in (2.14) and (2.15). Suppose that:*

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

(ii) For each $t_1,\dots,t_m\in T$, and each $u_1,\dots,u_m\in\mathbb{R}$,

$$I_{t_1,\dots,t_m}(u_1,\dots,u_m) = \inf\{I(z) : z\in l_\infty(T), z(t_1) = u_1,\dots,z(t_m) = u_m\}.$$

Then, $\{f(\cdot,t) : t\in T\}$ is a totally bounded set of $(\mathcal{L}^{\Phi_1}, N_{\Phi_1})$.

Finally, we present the main result to be used:

THEOREM 2.1. *Let $\{f(\cdot,t) : t \in T\}$ be a collection of measurable functions, where T is a compact subset of $\mathbb{R}^d$. Suppose that:*

(i) For each $t \in T$, $f(\cdot,t) \in \mathcal{L}^{\Phi_1}$.

(ii) For each $\lambda > 0$, and each $t \in T$ there exists a $\eta > 0$, such that

$$E[\exp(\lambda \sup_{s \in T, |s-t| \le \eta} |f(X,s) - f(X,t)|)] < \infty.$$

(iii) For each $t \in T$,

$$\lim_{\epsilon \to 0} \sup_{s \in T, |s-t| \le \epsilon} |f(X,s) - f(X,t)| = 0 \ a.s.$$

Then, $\{n^{-1} \sum_{j=1}^n f(X_j,t) : t \in T\}$ satisfies the LDP in $l_\infty(T)$ with speed n and rate function

$$I(z) = \inf\{J(l) : l \in (\mathcal{L}^{\Phi_1})^*, l(f(\cdot,t)) = z(t), \text{ for each } t \in T\}, z \in l_\infty(T).$$

3. Large deviations for M–estimators

In this section, we present several results on the large deviations for M-estimators. First, we consider the LDP for the M–estimators defined in (1.7).

THEOREM 3.1. *Let Θ be a convex set of $\mathbb{R}^d$. Let $g : S \times \Theta \to \mathbb{R}$ be a function such that for each $x \in S$, $g(x,\cdot)$ is a convex function. Let $\theta \in \Theta$. Let $\{K_m\}_{m \ge 1}$ be a sequence of compact convex sets of $\mathbb{R}^d$ contained in Θ and containing θ. Suppose that:*

(i) There exists a sequence of r.v.'s $\hat{\theta}_n = \hat{\theta}_n(X_1,\dots,X_n)$ such that $G_n(\hat{\theta}_n) = \inf_{t \in \Theta} G_n(t)$, where $G_n(t) = n^{-1}\sum_{j=1}^n g(X_j,t)$.

(ii) For each $t \in \Theta$, $E[g(X,t)] \ge E[g(X,\theta)]$.

(iii) $\{g(\cdot,t) : t \in \Theta\} \subset \mathcal{L}^{\Phi_1}$.

(iv) $\lim_{m\to\infty} \sup_{t \notin \partial K_m} \inf_{\lambda \in \mathbb{R}} E[\exp(\lambda(g(X,t) - g(X,\theta)))] = 0$.

(v) For each $t \in \Theta^o$, there exists a function $h(\cdot,t) : S \to \mathbb{R}^d$ such that

$$\lim_{v \to 0} |v|^{-1} N_{\Phi_1}\left(g(\cdot,t+v) - g(\cdot,t) - v'h(\cdot,t)\right) = 0.$$

(vi) For each $t \in \Theta^o$ such that $-\inf_{\lambda \in \mathbb{R}^d} E[\exp(\lambda' h(X,t))] < 0$, there exists $\epsilon_t > 0$ such that for each $\epsilon_t > \epsilon > 0$,

$$-\inf_{\lambda \in \mathbb{R}^d} E[\exp(\lambda' h(X,t))] < \inf_{t_1 : |t_1 - t| = \epsilon} \left(-\inf_{\lambda_1,\lambda_2 \in \mathbb{R}^d} E[\exp(\lambda_1' h(X,t) + \lambda_2' h(X,t_1))] \right).$$

Then, $\{\hat{\theta}_n\}$ satisfies the LDP with speed n and rate function

$$I(t) = \begin{cases} -\inf\{\ln(E[\exp(\lambda' h(X,t))]) : \lambda \in \mathbb{R}^d\} & \text{if} \quad t \in \Theta^o, \\ \infty & \text{if} \quad t \in \partial\Theta. \end{cases} \tag{3.1}$$

Observe that this is the natural rate. Under regularity conditions, $\{G_n(t) : t \in \Theta\}$ satisfies the LPD in $l_\infty(\Theta)$ with rate function

$$\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(g(\cdot,t)) = z(t), \text{ for each } t \in \Theta\}, z \in l_\infty(\Theta).$$

In the proof of Theorem 3.1, it is shown that for each $l \in \left(\mathcal{L}^{\Phi_1}\right)^*$ with $J(l) < \infty$, the function $l(g(\cdot,t)), t \in \Theta$, is differentiable in Θ with derivative is $l(h(\cdot,t)), t \in \Theta$. Besides,

$$\begin{aligned}(3.2)\qquad & \inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(g(\cdot,t)) = \sup_{t_1\in\Theta} l(g(\cdot,t_1))\} \\ = \quad & \inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(h(\cdot,t)) = 0\}.\end{aligned}$$

By Lemma 2.2, for each $t \in T$,

$$\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(h(\cdot,t)) = 0\} = -\inf_{\lambda\in\mathbb{R}^d} \ln\left(E\left[\exp(\lambda' h(X,t))\right]\right). \tag{3.3}$$

The conditions assumed on Theorem 3.1 are minimal conditions. Example 3 shows that condition (i) in Theorem 3.1 is needed. Condition (iv) in Theorem 3.1 is used to show that the M–estimator is eventually inside a compact set. Condition (vi) in Theorem 3.1 is used to get (3.2).

It follows from the previous theorem that for each $\epsilon > 0$,

$$\begin{aligned}&-\inf\{I(t) : |t-\theta| > \epsilon\} \le \liminf_{n\to\infty} n^{-1} \ln\left(\mathbb{P}\{|\hat\theta_n - \theta| > \epsilon\}\right) \\ &\le \limsup_{n\to\infty} n^{-1} \ln\left(\mathbb{P}\{|\hat\theta_n - \theta| \ge \epsilon\}\right) \le -\inf\{I(t) : |t-\theta| \ge \epsilon\}.\end{aligned}$$

A possible choice for the sequence of compact convex sets in Theorem 3.1 is

$$K_m = \{t \in \Theta : |t-\theta| \le m \text{ and } d(t,\Theta^c) \ge m^{-1}\}, m \ge 1.$$

Example 1. Let $\{f(\cdot,t) : t \in \Theta\}$ be a family of pdf's with respect to a measure μ defined on a measurable space $(S,\mathcal{S})$. We will assume that the support of $f(\cdot,t)$ does not depend on t and that for each $t \in \Theta$, $\nabla_t \ln f(x,t)$ exists. The M–estimator with respect to the kernel $g(x,t) = -\ln f(x,t)$ is the mle. It is well known that, by the concavity of the logarithmic function and the Jensen inequality, for any densities f and g with respect to the measure μ,

$$\int_S \ln(f(x)/g(x))g(x)\, d\mu(x) \le 0. \tag{3.4}$$

Hence, for each $t,\theta \in \Theta$

$$E_\theta[\ln f(X,t)] \le E_\theta[\ln f(X,\theta)],$$

where E_θ is the expectation when X has pdf $f(x,\theta)$. If the conditions in the previous theorem apply when θ obtains (the data comes from the pdf $f(\cdot,\theta)$), $\hat\theta_n$ satisfies the LDP with rate function

$$I_\theta(t) = -\inf_{\lambda\in\mathbb{R}^d} \ln\left(E_\theta\left[\exp(\lambda'\nabla_t \ln f(X,t))\right]\right) = \inf\{J_\theta(l) : l \in (\mathcal{L}_\theta^{\Phi_1})^*, l(\nabla_t \ln f(X,t)) = 0\},$$

where $\mathcal{L}_\theta^{\Phi_1}$, $(\mathcal{L}_\theta^{\Phi_1})^*$ and J_θ are defined when θ obtains. We claim that for each $\theta, t \in \Theta$, such that $K(f(\cdot,t), f(\cdot,\theta)) < \infty$ and $\int \nabla_t f(x,t)\, \mu(x) = 0$, then

$$I_\theta(t) \le K(f(\cdot,t), f(\cdot,\theta)). \tag{3.5}$$

Observe that we may define $l_t \in (\mathcal{L}_\theta^{\Phi_1})^*$ by $l_t(g) = E_t[g(X)] = E_\theta[g(X)\gamma_t(X)]$, $g \in \mathcal{L}_\theta^{\Phi_1}$, where $\gamma_t(x) = f(x,t)/f(x,\theta)$. We have that $E_\theta[\gamma_t(X)] = 1$ and

$$l_t(\nabla_t \ln f(X,t)) = \int \nabla_t f(x,t)\, d\mu(x) = 0.$$

So, by (2.5),

$$I_\theta(t) \le J_\theta(l_t) = E_\theta[\Psi_2(\gamma_t(X))] = K(f(\cdot,t), f(\cdot,\theta)).$$

The next theorem discerns when there exists equality in (3.5):

THEOREM 3.2. *Suppose that $\int \nabla_t f(x,t) d\mu(x) = 0$. Then,*
(i) If there exist $\lambda_{t,\theta} \in \mathbb{R}^d$ and $c_{t,\theta} \in \mathbb{R}$ such that

$$\lambda'_{t,\theta} \nabla_t \ln f(x,t) + c_{t,\theta} = \ln f(x,t) - \ln f(x,\theta), \quad P_t - \text{a.s.} \tag{3.6}$$

then,

$$-\inf_{\lambda \in \mathbb{R}^d} \ln\left(E_\theta\left[\exp(\lambda' \nabla_t \ln(X,t))\right]\right) = K(f(\cdot,t), f(\cdot,\theta)) = c_{t,\theta}.$$

(ii) If there exists $\lambda_{t,\theta} \in \mathbb{R}^d$ such that

$$\begin{aligned} &-\inf_{\lambda \in \mathbb{R}^d} \ln\left(E_\theta\left[\exp(\lambda' \nabla_t \ln f(X,t))\right]\right) \\ &= -\ln\left(E_\theta\left[\exp(\lambda'_{t,\theta} \nabla_t \ln f(X,t))\right]\right) = K(f(\cdot,t), f(\cdot,\theta)), \end{aligned}$$

then

$$\ln f(x,t) - \ln f(x,\theta) - \lambda'_{t,\theta} \nabla_t \ln f(x,t) = K(f(\cdot,t), f(\cdot,\theta)), \; P_t - \text{a.s.}$$

Example 2. Let $\{f(x-\theta) : \theta \in \mathbb{R}\}$ be a one dimensional location family. Assume that $f(x) > 0$, for each $x \in \mathbb{R}$. Then, condition (3.6) holds for each $t, \theta \in \mathbb{R}$ if and only if for each $t \in \mathbb{R}$, there exists $\lambda(t)$ and $c(t)$ such that

$$\lambda(t) \frac{f'(x)}{f(x)} + c(t) = \ln f(x) - \ln f(x-t), \text{for each } x \in \mathbb{R}. \tag{3.7}$$

It is easy to see that the normal pdf $f(x) = (2\pi)^{-1/2} \sigma^{-1} \exp\left(-2^{-1}\sigma^{-2}(x-\mu)^2\right)$, where $\mu \in \mathbb{R}$, and $\sigma > 0$, satisfies (3.7) with $\lambda(t) = t$ and $c(t) = 2^{-1}\sigma^{-2}t^2$. We also have that the pdf

$$f(x) = (\Gamma(\alpha))^{-1} |\gamma| \alpha^\alpha \exp\left(-\alpha e^{\gamma(x-\theta)} + \alpha\gamma(x-\theta)\right) \tag{3.8}$$

where $\alpha > 0$, $\gamma \ne 0$ and $\theta \in \mathbb{R}$, satisfies (3.7) with $\lambda(t) = \gamma^{-1}\,(1 - e^{-\gamma t})$ and $c(t) = \alpha(e^{-\gamma t} - 1 + \gamma t)$. If X has the pdf in (3.8) and $Y = \gamma^{-1} \ln(X\alpha) - \theta$, then Y has a Gamma$(\alpha, 1)$ distribution.

The following theorem determines the one dimensional location families for which the rate function of the large deviations of the mle coincides with the Kullback–Leibler information number.

THEOREM 3.3. *Suppose that f is a second differentiable pdf satisfying (3.7), then either f is a normal pdf or f is as in (3.8), for some $\alpha > 0$, $\gamma \ne 0$ and $\theta \in \mathbb{R}$.*

By Theorem 2 in Ferguson (1962), the normal family, with a fixed σ^2, and the family in (3.8), with some fixed $\alpha > 0$ and $\gamma \ne 0$, are the only one dimensional location families, which are exponential families.

Theorem 3.1 gives the following for a one dimensional location family:

Theorem 3.4. *Let $\{X_j\}$ be a sequence of i.i.d.r.v.'s with a pdf belonging to $\{f(\cdot - t) : t \in \mathbb{R}\}$ where f is a pdf. Suppose that the following conditions are satisfied:*

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

(ii) f has a continuous first derivative.

(iii) $-\ln f(\cdot)$ is a strictly convex function.

(iv) $\lim_{x\to\pm\infty} f(x) = 0$.

(v) $\lim_{t\to\pm\infty} \inf_{\lambda\in\mathbb{R}} \int_{-\infty}^{\infty} (f(x-t))^{\lambda}(f(x))^{1-\lambda}\, dx = 0$.

(vi) For each $t, \lambda \in \mathbb{R}$, $\int_{-\infty}^{\infty} \exp(\lambda f'(x-t)/f(x-t)) f(x)\, dx < \infty$.

(vii) The function $t \in \mathbb{R} \mapsto f'(X-t)/f(X-t) \in \mathcal{L}^{\Phi_1}$ is continuous.

Then, there exists a sequence of r.v.'s $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ such that $G_n(\hat{\theta}_n) = \inf_{t\in\Theta} G_n(t)$, where $G_n(t) = -n^{-1}\sum_{j=1}^n \ln f(X_j - t)$. Besides, when θ obtains, $\{\hat{\theta}_n\}$ satisfies the LDP with speed n and rate function $I(t-\theta)$, $t \in \mathbb{R}$, where

$$I(t) = -\inf_{\lambda\in\mathbb{R}} \ln(E_0[\exp(\lambda f'(X-t)/f(X-t))]). \tag{3.9}$$

Condition (v) in the previous theorem follows if for some $0 < \lambda < 1$, $\int_{-\infty}^{\infty} (f(x))^{1-\lambda}\, dx < \infty$ (using conditions (ii) and (iv)).

For a scale family of pdf's, we have results similar to the ones for the location family. For example, it is easy to see that if $\{\lambda^{-1} f(\lambda^{-1}x) : \lambda > 0\}$ is a scale family, then this family satisfies (3.6) if and only if $f(x) = \alpha_1 |x|^{-1} g_1(\ln(-x))$, if $x < 0$ and $f(x) = \alpha_1 x^{-1} g_2(\ln x)$, if $x > 0$, where $\alpha_1, \alpha_2 > 0$, $\alpha_1 + \alpha_2 = 1$ and g_1 and g_2 are two pdf's satisfying condition (3.7). Hence, g_1 and g_2 are either normal or as in (3.8). This implies that $f(x) = \alpha_1 f_1(-x)$, if $x < 0$ and $f(x) = \alpha_1 f_2(x)$, if $x > 0$, where f_1 and f_2 are pdfs on $(0, \infty)$ for the form either $f(x) = (2\pi)^{-1/2}\sigma^{-1}x^{-1}\exp\left(-2^{-1}\sigma^{-2}(\ln x - \mu)^2\right)$, where $\mu \in \mathbb{R}$, and $\sigma > 0$, or $f(x) = (\Gamma(\alpha))^{-1} e^{-cx^r} r c^{\alpha} x^{\alpha r - 1}$, where $\alpha, r, c > 0$.

A common family of pdf's is the exponential family. Let μ be a measure on $\mathbb{R}^d$. Define $\psi(t) := \ln \int_{\mathbb{R}^d} e^{t'x}\, d\mu(x)$. 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. By a change of parameter, any full exponential family of distribution can have this representation (see Brown, 1986). It is easy to see that a full exponential family of pdf's satisfies (3.6). If $t \in \Theta^o$, then the Kullback–Leibler information of $f(\cdot, t)$ and $f(\cdot, \theta)$ is

$$\begin{aligned} K(f(\cdot,t), f(\cdot,\theta)) &= \psi(\theta) - \psi(t) + \int (t-\theta)' x \exp(t'x - \psi(t))\, d\mu(x) \\ &= \psi(\theta) - \psi(t) + (t-\theta)'\nabla\psi(t), \end{aligned} \tag{3.10}$$

because by taking derivates inside the integral,

$$\nabla\psi(t) = \int x \exp(t'x)\, d\mu(x) \left(\int \exp(t'x)\, d\mu(x) \right)^{-1}.$$

Theorem 3.1 gives the following for a full exponential family:

Theorem 3.5. *With the notation above, let $\theta \in \Theta^o$ and 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:*

(i) There exists a sequence of r.v.'s $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ such that

$$\psi(\hat{\theta}_n) - \hat{\theta}_n' \bar{X}_n = \inf_{t\in\Theta} (\psi(t) - t'\bar{X}_n),$$

where $\bar{X}_n := n^{-1}\sum_{j=1}^n X_j$,

(ii) $\lim_{m\to\infty}\sup_{t\notin K_m}\inf_{\lambda\in\mathbb{R}}(\psi(\theta+\lambda(t-\theta))-\psi(\theta)-\lambda(\psi(t)-\psi(\theta)))=-\infty$.

(iii) For each $t_1,t_2\in\Theta^o$, $\nabla\psi(t_1)\neq\nabla\psi(t_2)$.

Then, when θ *obtains,* $\hat{\theta}_n$ *satisfies the LDP with speed* n *rate function*

$$I(t)=\begin{cases} K(f(\cdot,t),f(\cdot,\theta)) & \text{if} \quad t\in\Theta^o,\\ \infty & \text{if} \quad t\in\partial\Theta.\end{cases} \tag{3.11}$$

Condition (ii) in Theorem 3.5 holds if $\lim_{m\to\infty}\inf_{t\in\Theta_m^c}\frac{\psi(t)-\psi(\theta)}{|t-\theta|}=\infty$, where

$$\Theta_m=\{t\in\Theta: d(t,\partial\Theta)\geq m^{-1}, |t-\theta|\leq m\}.$$

Observe that taking $\lambda=\tau_0|t-\theta|^{-1}$, where $\tau_0<d(\theta,\Theta^c)$, we get that

$$\begin{aligned}&\inf_{\lambda\in\mathbb{R}}(\psi(\theta+\lambda(t-\theta))-\psi(\theta)-\lambda(\psi(t)-\psi(\theta)))\\ \geq\ &\psi(\theta+\tau_0|t-\theta|^{-1}(t-\theta))-\psi(\theta)-\tau_0|t-\theta|^{-1}(\psi(t)-\psi(\theta)).\end{aligned}$$

Example 3. Consider the measure μ in $\mathbb{R}$ defined by $\mu\{0\}=\mu\{1\}=1$ and $\mu(\mathbb{R}-\{0,1\})=0$. Let $\psi(t):=\ln\int_{\mathbb{R}^d}e^{t'x}\,d\mu(x)=\ln(1+e^t)$ and let $f(x,t):=e^{tx-\psi(t)}=e^{tx}/(1+e^t)$, for $x\in\{0,1\}$. The family of pdf's $\{f(x,t):t\in\mathbb{R}\}$ is a reparametrization of the Bernoulli distribution. In this situation, the mle is not defined: $\sup_{t\in\mathbb{R}}(t\bar{X}_n-\ln(1+e^t))$ is not attained if either $\bar{X}_n=0$ or $\bar{X}_n=1$. Theorem 3.5 does not apply to this example. The mle exists as a random element with values in $[-\infty,\infty]$. It is easy to see that the mle (when defined in $[-\infty,\infty]$) when θ obtains satisfies the LPD in $[-\infty,\infty]$ with rate function

$$I_\theta(t)=\begin{cases}\ln(1/p) & \text{if } t=-\infty\\ u\ln(u/p)+(1-u)\ln((1-u)/(1-p)) & \text{if } t\in\mathbb{R}\\ \ln(1/(1-p)) & \text{if } t=\infty\end{cases}$$

where $p=e^\theta/(1+e^\theta)$ and $u=e^t/(1+e^t)$. This example shows that condition (i) in Theorem 3.5 is needed.

In Theorem 3.1, we assumed that for each x, the function $g(x,\cdot)$ is convex. Next, we consider theorems which apply to other situations. Next, we consider the one dimensional case.

THEOREM 3.6. *Let* $h:S\times\mathbb{R}\to\mathbb{R}$ *be a function such that for each* $x\in S$, $h(x,\cdot):\mathbb{R}\to\mathbb{R}$ *is a nondecreasing function. Let* $\hat{\theta}_n^{(1)}=\inf\{t:H_n(t)\geq 0\}$ *and let* $\hat{\theta}_n^{(2)}=\sup\{t:H_n(t)\leq 0\}$, *where* $H_n(t)=n^{-1}\sum_{j=1}^n h(X_j,t)$. *Let* $\hat{\theta}_n$ *be a sequence of* $[-\infty,\infty]$*–valued r.v.'s such that* $\hat{\theta}_n^{(1)}\leq\hat{\theta}_n\leq\hat{\theta}_n^{(2)}$. *Let* $\theta\in\mathbb{R}$. *Suppose that:*

(i) $\{h(\cdot,t):t\in\mathbb{R}\}\subset\mathcal{L}^{\Phi_1}$.

(ii) $E[h(X,\theta)]=0$.

(iii) For each $t>\theta$, $\mathbb{P}(h(X,t)<0)>0$, *and for each* $t<\theta$, $\mathbb{P}(h(X,t)>0)>0$.

Then, for each $t>\theta$,

$$\lim_{n\to\infty}n^{-1}\ln\left(\mathbb{P}\{\hat{\theta}_n\geq t\}\right)=\inf_{\lambda\in\mathbb{R}}\ln(E[\exp(\lambda h(X,t-))]);$$

for each $t \geq \theta$,

$$\lim_{n\to\infty} n^{-1} \ln\left(\mathbb{P}\{\hat{\theta}_n > t\}\right) = \inf_{\lambda\in\mathbb{R}} \ln(E[\exp(\lambda h(X, t+))]);$$

for each $t < \theta$,

$$\lim_{n\to\infty} n^{-1} \ln\left(\mathbb{P}\{\hat{\theta}_n \leq t\}\right) = \inf_{\lambda\in\mathbb{R}} \ln(E[\exp(\lambda h(X, t+))]);$$

and for each $t \leq \theta$,

$$\lim_{n\to\infty} n^{-1} \ln\left(\mathbb{P}\{\hat{\theta}_n < t\}\right) = \inf_{\lambda\in\mathbb{R}} \ln(E[\exp(\lambda h(X, t-))]),$$

where $h(x, t-) = \lim_{s\to t-} h(x, s)$ and $h(x, t+) = \lim_{s\to t+} h(x, s)$.

The previous theorem is very close to Theorem 2 in Rubin and Kukhin (1983). However, there is an error in Theorem 2 in Rubin and Kukhin (1983).

Example 4. Given a sequence of i.i.d.r.v's $\{X_j\}_{j=1}^\infty$ and $0 < p < 1$, a sample p–quantile $\hat{\theta}_n$ is defined as in the previous theorem with $h(x, t) = I(x \leq t) - p$. Suppose that $P(X \leq \theta) = p$. Theorem 3.6 gives that for each $t > \theta$, such that $\mathbb{P}\{X > t\} > 0$,

$$\lim_{n\to\infty} n^{-1} \ln(\mathbb{P}\{\hat{\theta}_n \geq t\}) = -\ln\left(\frac{p^p(1-p)^{1-p}}{(F(t-))^p(1-F(t-))^{1-p}}\right),$$

and

$$\lim_{n\to\infty} n^{-1} \ln(\mathbb{P}\{\hat{\theta}_n > t\}) = -\ln\left(\frac{p^p(1-p)^{1-p}}{(F(t))^p(1-F(t))^{1-p}}\right);$$

and for each $t < \theta$, such that $\mathbb{P}\{X < t\} > 0$,

$$\lim_{n\to\infty} n^{-1} \ln(\mathbb{P}\{\hat{\theta}_n \leq t\}) = -\ln\left(\frac{p^p(1-p)^{1-p}}{(F(t))^p(1-F(t))^{1-p}}\right)$$

and

$$\lim_{n\to\infty} n^{-1} \ln(\mathbb{P}\{\hat{\theta}_n < t\}) = -\ln\left(\frac{p^p(1-p)^{1-p}}{(F(t-))^p(1-F(t-))^{1-p}}\right).$$

Example 5. Condition (iii) in Theorem 3.6 is needed. Suppose that $h(x, t) = (t - x)I(0 < x < t)$, if $t > 0$; $h(x, 0) = 0$; and $h(x, t) = (t - x)I(t < x < 0)$, if $t < 0$. Suppose that X has a nondegenerate distribution symmetric about 0. Then, the rate functions of the large deviations of $\hat{\theta}_n^{(1)}$ and $\hat{\theta}_n^{(2)}$ are different. We have that $\hat{\theta}_n^{(1)} = -\infty$, if $X_i \geq 0$, for each $1 \leq i \leq n$; $\hat{\theta}_n^{(1)} = \max\{X_i : X_i < 0\}$, if $X_i < 0$, for some $1 \leq i \leq n$; $\hat{\theta}_n^{(2)} = \infty$, if $X_i \leq 0$, for each $1 \leq i \leq n$; $\hat{\theta}_n^{(2)} = \min\{X_i : X_i > 0\}$, if $X_i > 0$, for some $1 \leq i \leq n$. We have that for each $t \geq 0$, $\lim_{n\to\infty} n^{-1} \ln(\mathbb{P}\{\hat{\theta}_n^{(1)} \geq t\}) = -\infty$ and $\lim_{n\to\infty} n^{-1} \ln(\mathbb{P}\{\hat{\theta}_n^{(2)} \geq t\}) = -\ln 2$; and for each $t \leq 0$, $\lim_{n\to\infty} n^{-1} \ln(\mathbb{P}\{\hat{\theta}_n^{(1)} \leq t\}) = -\ln 2$ and $\lim_{n\to\infty} n^{-1} \ln(\mathbb{P}\{\hat{\theta}_n^{(2)} \leq t\}) = -\infty$.

Theorem 3.6 gives the large deviations for the mle over an one-dimensional exponential family under minimal conditions:

THEOREM 3.7. *Le μ be a measure in $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ such that:*
(i) For each $a\in\mathbb{R}$, $\mu(\mathbb{R}-\{a\})>0$.
(ii) $\Theta:=\{t\in\mathbb{R}:\int e^{tx}\,d\mu(x)<\infty\}$ has nonempty interior.
Let $\psi(t):=\ln\int e^{tx}\,d\mu(x)$, $t\in\Theta$, and let $f(x,t):=\exp(tx-\psi(t))$, $x\in\mathbb{R}$, $t\in\Theta$. Let $a_\psi=\inf\{t\in\mathbb{R}:\int e^{tx}\,d\mu(x)\}$ and let $b_\psi=\sup\{t\in\mathbb{R}:\int e^{tx}\,d\mu(x)\}$. Let $\{X_j\}$ be a sequence of i.i.d.r.v.s from the pdf $f(\cdot,\theta)$, where $\theta\in\Theta^o$. Let $\hat\theta_n=\inf\{t\in\Theta^o: n^{-1}\sum_{j=1}^n(\psi'(t)-X_j)\ge 0\}$, where $\inf(\emptyset)=a_\psi$.

Then, for each $b_\psi>t\ge\theta$,

$$\lim_{n\to\infty}n^{-1}\ln\left(\mathbb{P}\{\hat\theta_n\ge t\}\right)=-K(f(\cdot,t),f(\cdot,\theta))$$

and for each $a_\psi<t\le\theta$,

$$\lim_{n\to\infty}n^{-1}\ln\left(\mathbb{P}\{\hat\theta_n\le t\}\right)=-K(f(\cdot,t),f(\cdot,\theta)).$$

The following theorem deals with the multivariate case.

THEOREM 3.8. *Let Θ be a subset of $\mathbb{R}^d$. Let $h:S\times\Theta\to\mathbb{R}^d$ be a function. Let $\{K_m\}_{m\ge1}$ be a nondecreasing sequence of compact sets of $\mathbb{R}^d$ contained in Θ. Suppose that:*

(i) There exists a sequence of r.v.'s $\hat\theta_n=\hat\theta_n(X_1,\dots,X_n)$ such that $H_n(\hat\theta_n)=0$, where $H_n(t):=n^{-1}\sum_{j=1}^n h(X_j,t)$.

(ii) $\{h(\cdot,t):t\in\Theta\}\subset\mathcal{L}^{\Phi_1}$.

(iii) $t\in\Theta\mapsto h(\cdot,t)\in\mathcal{L}^{\Phi_1}$ is a continuous function.

(iv) $\inf_{t\in\Theta-K_1}|H(t)|>0$, where $H(t)=E[h(X,t)]$.

(v) For each $t>0$, $\lim_{m\to\infty}\sup_{\lambda>0}(\lambda t-\ln(E[\exp(\lambda R_m(X))]))=\infty$, where $R_m(x):=\sup_{t\in\Theta-K_m}\left|\frac{h(x,t)-H(t)}{H(t)}\right|$.

(vi) For each $m\ge1$, $\{n^{-1}\sum_{j=1}^n h(X_j,t):t\in K_m\}$ satisfies the LDP in $l_\infty(K_m)$ with speed n.

(vii) For each $t\in\Theta^o$ such that $-\inf_{\lambda\in\mathbb{R}^d}E[\exp(\lambda'h(X,t))]<0$, and each $t_1\in\Theta$,

$$-\inf_{\lambda\in\mathbb{R}^d}E[\exp(\lambda'h(X,t))]<-\inf_{\lambda_1,\lambda_2\in\mathbb{R}^d}E[\exp(\lambda_1'h(X,t)+\lambda_2'h(X,t_1))].$$

Then, $\hat\theta_n$ satisfies the LDP with speed n and rate function

$$I(t)=-\inf\{\ln(E[\exp(\lambda'h(X,t))]):\lambda\in\mathbb{R}^d\}.$$

Condition (v) in the previous theorem can be checked using Theorem 2.1.

Example 6. The previous theorem applies to many common parametric families of pdf's. For example, consider the mle over the family of pdf's $\{f(x,t):=\frac{1}{\sqrt{2\pi}t}e^{-\frac{(x-t)^2}{2t^2}}: t>0\}$, i.e. $t^{-1}(X-t)$ has a standard normal distribution. If θ obtains, $\hat\theta_n$ satisfies the LDP with speed n and rate function $I_\theta(t)=r(t/\theta)$, where

$$r(a):=2^{-1}\ln(2+(4+5a^2(2-a)^2)^{1/2})-2^{-1}\ln(5a^2)+2^{-1}-2^{-1}a+(3/4)a^2-2^{-2}(4+5a^2(2-a)^2)^{1/2}.$$

It is easy to see that if $t \neq \theta$, (3.6) does not hold. The Kullback–Leibler information is

$$K(f(\cdot,t),f(\cdot,\theta)) = (t/\theta)^2 - (t/\theta) - \ln(t/\theta).$$

For each $t \neq \theta$, $K(f(\cdot,t),f(\cdot,\theta)) > I_\theta(t)$, i.e. for $a \neq 1$, $a^2 - a - \ln a > r(a)$.

4. Confidence regions

As mentioned in the introduction, the LDP of statistics can be used to obtain confidence regions of non vanishing size such that their coverage probability goes to one exponentially fast. These confidence regions have a certain minimality properties. Let Θ be a parameter set. Suppose that, when θ obtains, $\hat\theta_n$ satisfies the LDP with speed n and continuous rate function $I_\theta(\cdot)$. Given $0 < \alpha < \infty$, let

$$C_\alpha(X_1,\dots,X_n) := \{\theta \in \Theta : I_\theta(\hat\theta_n(X_1,\dots,X_n)) < \alpha\}. \tag{4.1}$$

Assuming that $U_{\theta,\alpha} := \{t \in \mathbb{R}^d : I_\theta(t) < \alpha\}$ is an open set, then $C_\alpha(X_1,\dots,X_n)$ is a confidence region for Θ such that

$$\limsup_{n\to\infty} n^{-1} \ln\left(P_\theta\{\theta \notin C_\alpha(X_1,\dots,X_n)\}\right) \le -\alpha. \tag{4.2}$$

By the results in Section 3, the rate function for the LDP of mle's is

$$I_\theta(t) = -\inf_{\lambda\in\mathbb{R}^d} \ln E_\theta[\exp\left(\lambda'\nabla_t \ln f(X,t)\right)]. \tag{4.3}$$

Because of the equivariance properties of the mle, the constructed confidence regions satisfy the usual equivariance properties. For a location family of pdf's, i.e. $\Theta = \mathbb{R}^d$, $f(x,t) = f(x-t)$, $\theta \in \Theta$, where f is a fixed pdf, then

$$-\inf_{\lambda\in\mathbb{R}^d} \ln E_\theta[\exp\left(\lambda'(\nabla \ln f)(X-t)\right)] = I(t-\theta),$$

where

$$I(t) = -\inf_{\lambda\in\mathbb{R}^d} \ln E_0[\exp\left(\lambda'(\nabla \ln f)(X-t)\right)].$$

Hence, the confidence region in (4.1) is $C_\alpha(X_1,\dots,X_n) := \{\theta \in \Theta : I(\hat\theta_n - \theta) < \alpha\}$. Similarly, for a scale family ($\Theta = (0,\infty)$, $f(x,t) = t^{-1}f(t^{-1}x)$, where f is a fixed pdf), $I_\theta(t) = I(t/\theta)$, where

$$I(t) = -\inf_{\lambda\in\mathbb{R}} \ln E_1\left[\exp\left(\frac{\partial}{\partial t}\ln(t^{-1}f(t^{-1}X))\right)\right].$$

Example 7. Let $X_1,\dots,X_n$ be a i.i.d.r.v.'s from an exponential distribution with mean $\theta > 0$. The mle of θ is $\hat\theta_n = \bar X_n$. From the results in Section 3, $\hat\theta_n$ satisfies the LDP with rate function

$$I_\theta(t) = -\inf_\lambda \ln E_\theta[\exp(\lambda \frac{\partial}{\partial t}\ln f(X,t))] = (t/\theta) - 1 - \ln(t/\theta),$$

when θ obtains. Given $\alpha > 0$, take $a_\alpha < 1 < b_\alpha$ such that

$$a_\alpha - 1 - \ln(a_\alpha) = b_\alpha - 1 - \ln(b_\alpha) = \alpha. \tag{4.4}$$

Then, the confidence region in (4.1) is $(b_\alpha^{-1}\hat{\theta}, a_\alpha^{-1}, \hat{\theta})$. By Example 9.3.4 in Casella and Berger (2002), the shortest confidence interval based on the pivotal quantity $\theta^{-1}\bar{X}_n$ has the form $[b^{-1}\bar{X}_n, a^{-1}\bar{X}_n]$, where $a^{1+n^{-1}}e^{-a} = b^{1+n^{-1}}e^{-b}$. As $n \to \infty$, this condition goes to (4.4). Simulations show that the two confidence intervals are very close.

For mle's from a full exponential family the constructed confidence regions agree with the confidence regions obtained by inverting the acceptance region of the likelihood ration test with the null hypothesis $H_0 : t = \theta$.

Theorem 4.1. *Consider the parametric family $\{f(x,t) := e^{t'x-\psi(t)} : t \in \Theta\}$ in Theorem 3.2. Assuming that $\sup_{t\in\Theta} L(t)$ is attained in the interior of Θ, then*

$$-n^{-1}\ln\left(\frac{L(\theta)}{\sup_{t\in\Theta} L(t)}\right) = \psi(\theta) - \psi(\hat{\theta}_n) - (\theta - \hat{\theta})'\nabla\psi(\hat{\theta}_n) = K(f(\cdot,\hat{\theta}_n), f(\cdot,\theta)),$$

where $L(t) := \prod_{j=1}^n f(X_j, t)$.

5. Proofs

Proof of Lemma 2.1. If $l(\mathbf{1}) = 1$, then

$$\begin{aligned} &\sup_{f\in\mathcal{L}^{\Phi_1}}\left(l(f) - E[e^{f(X)} - 1]\right) = \sup_{f\in\mathcal{L}^{\Phi_1},\lambda\in\mathbb{R}}\left(l(\lambda\mathbf{1} + f) - E[e^{\lambda+f(X)} - 1]\right) \\ &= \sup_{f\in\mathcal{L}^{\Phi_1},\lambda\in\mathbb{R}}\left(\lambda + l(f) - e^\lambda E[e^{f(X)}] + 1\right) = \sup_{f\in\mathcal{L}^{\Phi_1}}\left(l(f) - \ln E[e^{f(X)}]\right), \end{aligned}$$

(the maximum over λ is attained when $1 = e^\lambda E[e^{f(X)}]$). □

Proof of Lemma 2.2. Define

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

and

$$I^{(2)}(u_1,\dots,u_m) = \inf\left\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(f_j) = u_j \text{ for each } 1 \le j \le m\right\}.$$

Then, $I^{(1)}$ and $I^{(2)}$ are convex lower semicontinuous functions. To prove that the two functions are equal, it suffices to prove that their Fenchel conjugates agree. Using (2.3), we have that for each $\lambda_1,\dots,\lambda_m \in \mathbb{R}$,

$$\begin{aligned} &\sup\{\textstyle\sum_{j=1}^m \lambda_j u_j - I^{(2)}(u_1,\dots,u_m) : u_1,\dots,u_m \in \mathbb{R}\} \\ &= \sup_{u_1,\dots,u_m\in\mathbb{R}}\sup\left\{\textstyle\sum_{j=1}^m \lambda_j u_j - J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(f_j) = u_j \text{ for each } 1 \le j \le m\right\} \\ &= \sup_{u_1,\dots,u_m\in\mathbb{R}}\sup\left\{l(\textstyle\sum_{j=1}^m \lambda_j f_j) - J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(f_j) = u_j \text{ for each } 1 \le j \le m\right\} \\ &= \sup\left\{l(\textstyle\sum_{j=1}^m \lambda_j f_j) - J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*\right\} \\ &= \ln\left(E\left[\exp\left(\textstyle\sum_{j=1}^m \lambda_j f_j(X)\right)\right]\right). \end{aligned}$$
□

We will need the following lemma:

Lemma 5.1. *(i) For each $k \ge 0$ and each function $f \in \mathcal{L}^{\Phi_1}$,*

$$\sup\{|l(f)| : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, J(l) \le k\} \le (k + 1 + 2^{1/2})N_{\Phi_1}(f).$$

(ii) For each function $f \in \mathcal{L}^{\Phi_1}$,

$$\left(\frac{\ln(2)}{\ln(8e^2)}\right) N_{\Phi_1}(f) \le \sup\{|l(f)| : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, J(l) \le 1\}.$$

Proof. First, we prove (i). Let $\lambda := N_{\Phi_1}(f)$ (so that $E[e^{\lambda^{-1}|f(X)|} - 1 - \lambda^{-1}|f(X)|] \le 1$). By (2.2), for each $l \in \mathcal{L}^{\Phi_1}$,

$$\begin{aligned} l(\lambda^{-1} f) &\le J(l) + \ln(E[e^{\lambda^{-1} f(X)}]) \le J(l) + E[e^{\lambda^{-1} f(X)} - 1], \\ l(-\lambda^{-1} f) &\le J(l) + \ln(E[e^{-\lambda^{-1} f(X)}]) \le J(l) + E[e^{-\lambda^{-1} f(X)} - 1]. \end{aligned}$$

So, for J with $J(l) \le k$,

$$\begin{aligned} &|l(f)| \le \lambda J(l) + \lambda E[e^{\lambda^{-1}|f(X)|} - 1] \\ \le\ & \lambda k + E[|f(X)|] + \lambda E[e^{\lambda^{-1}|f(X)|} - 1 - \lambda^{-1}|f(X)|] \le \lambda(k+1) + E[|f(X)|]. \end{aligned}$$

We also have that

$$(\lambda^{-1} E[|f(X)|])^2 \le E[\lambda^{-2}|f(X)|^2] \le 2E[e^{\lambda^{-1}|f(X)|} - 1 - \lambda^{-1}|f(X)|] \le 2.$$

From these estimations (i) follows.

As to (ii). Define

$$\|f\|_K := \sup\{|E[f(X)\gamma(X)]| : E[\gamma(X)] = 1, E[\Psi_2(\gamma(X))] \le 1\}.$$

First, we prove that for each function f with $E[f(X)] = 0$ and each nonnegative function γ, with $E[\gamma(X)] = 1$,

$$|E[f(X)\gamma(X)]| \le (1 + E[\Psi_2(\gamma(X))])\|f\|_K \tag{5.1}$$

(5.1) is obviously true if $E[\Psi_2(\gamma(X))] \le 1$. If $E[\Psi_2(\gamma(X))] > 1$, take $t = (E[\Psi_2(\gamma(X))])^{-1}$. By convexity $E[\Psi_2(t\gamma(X) + 1 - t)] \le tE[\Psi_2(\gamma(X))] = 1$. So, $|E[f(X)(t\gamma(X) + 1 - t)]| \le \|f\|_K$ and $|E[f(X)\gamma(X)]| \le E[\Psi_2(\gamma(X))]\|f(X)\|_K$. Hence, (5.1) holds.

Next, we prove that for each function f with $E[f(X)] = 0$,

$$N_{\Phi_1}(f) \le \frac{\ln(2e)}{\ln(2)}\|f\|_K. \tag{5.2}$$

Given a function f with $\|f\|_K \le 1$ and $E[f(X)] = 0$, we get from (5.1) with $\gamma(x) = e^{f(X)} I(f(X) \ge 0)(E[e^{f(X)} I(f(X) \ge 0)])^{-1}$ that

$$\begin{aligned} &E[f(X)e^{f(X)} I(f(X) \ge 0)(E[e^{f(X)} I(f(X) \ge 0)])^{-1}] \le 1 + E[\Psi_2(\gamma(X))] \\ =\ & 1 + E[e^{f(X)} I(f(X) \ge 0)(E[e^{f(X)} I(f(X) \ge 0)])^{-1} \ln\left(e^{f(X)}(E[e^{f(X)} I(f(X) \ge 0)])^{-1}\right)] \\ =\ & 1 + E[f(X) I(f(X) \ge 0)e^{f(X)}(E[e^{f(X)} I(f(X) \ge 0)])^{-1}] - \ln(E[e^{f(X)} I(f(X) \ge 0)]). \end{aligned}$$

So, $E[e^{f(X)} I(f(X) \ge 0)] \le e$. Similarly, we get that $E[e^{-f(X)} I(f(X) \le 0)] \le e$. Hence, $E[e^{|f(X)|}] \le 2e$. Finally, we have that

$$\begin{aligned} &E[e^{\ln(2)(\ln(2e))^{-1}|f(X)|} - 1 - \ln(2)(\ln(2e))^{-1}|f(X)|] \le E[e^{\ln(2)(\ln(2e))^{-1}|f(X)|} - 1] \\ \le\ & \left(E[e^{|f(X)|}]\right)^{\ln(2)(\ln(2e))^{-1}} - 1 \le 1. \end{aligned}$$

Hence, if $\|f\|_K \le 1$, then $N_{\Phi_1}(f) \le \frac{\ln(2e)}{\ln(2)}$ and (5.2) follows.

Using that $|E[f(X)]| \le \|f\|_K$ and (5.2),

$$N_{\Phi_1}(f(X)) \le N_{\Phi_1}(f - E[f(X)]) + N_{\Phi_1}(E[f(X)]) \le \tfrac{\ln(2e)}{\ln(2)} \|f - E[f(X)]\|_K + |E[f(X)]|$$
$$\le \left(1 + 2\tfrac{\ln(2e)}{\ln(2)}\right) \|f(X)\|_K = \left(\tfrac{\ln(8e^2)}{\ln(2)}\right) \|f(X)\|_K.$$

Hence, (ii) follows. □

By the previous lemma, given a class of functions $\{f(\cdot,t) : t \in T\}$ in $\mathcal{L}^{\Phi_1}$ and $l \in \left(\mathcal{L}^{\Phi_1}\right)^*$ with $J(l) < \infty$, then $t \in (T, d_{\Phi_1}) \mapsto l(f(\cdot,t)) \in \mathbb{R}$ is a Lipschitz function, where $d_{\Phi_1}(s,t) = N_{\Phi_1}(f(\cdot,s) - f(\cdot,t))$. We also have that if $z \in l_\infty(T)$ satisfies $\sup_{s,t\in T} I_{s,t}(z(s), z(t)) < \infty$, then $t \in (T, d_{\Phi_1}) \mapsto z(t) \in \mathbb{R}$ is a Lipschitz function. Observe that if $\sup_{s,t\in T} I_{s,t}(z(s), z(t)) < c$, then for each $s, t \in T$, there exists $l \in \left(\mathcal{L}^{\Phi_1}\right)^*$ with $J(l) < c$, $l(f(\cdot,s)) = z(s)$ and $l(f(\cdot,t)) = z(t)$. So, by Lemma 5.1 (i)

$$|z(s) - z(t)| = |l(f(\cdot,s) - f(\cdot,t))| \le (c + 1 + 2^{1/2}) d_{\Phi_1}(s,t).$$

Proof of Lemma 2.3. First, we prove (i). Let

$$I^{(1)}(z) = \sup\{I_{t_1,\dots,t_m}(z(t_1), \dots, z(t_m)) : t_1, \dots, t_m \in T, m \ge 1\}.$$

and let

$$I^{(2)}(z) = \inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(f(\cdot,t)) = z(t) \text{ for each } t \in T\}.$$

By Lemma 2.2, for each $z \in l_\infty(T)$, we have that $I^{(1)}(z) \le I^{(2)}(z)$. To prove the reverse inequality, we may assume that $I^{(1)}(z) < \infty$. Since for each $r_1, \dots, r_m, s_1, \dots, s_p \in T$ and each $u_1, \dots, u_m, v_1, \dots, v_p \in \mathbb{R}$,

$$I_{r_1,\dots,r_m}(u_1, \dots, u_m) \le I_{r_1,\dots,r_m,s_1,\dots,s_p}(u_1, \dots, u_m, v_1, \dots, v_p),$$

we can find a sequence $\{s_n\}$ of T such that

$$\lim_{n\to\infty} I_{s_1,\dots,s_m}(z(s_1), \dots, z(s_m)) = I^{(1)}(z)$$

and $\{f(\cdot,s_n)\}_{n=1}^\infty$ is a dense set of $\{f(\cdot,t) : t \in T\}$ with respect to the norm N_{Φ_1}. Take $l_n \in \left(\mathcal{L}^{\Phi_1}\right)^*$ such that $l_n(f(\cdot,s_j)) = z(s_j)$ for each $1 \le j \le n$ and

$$J(l_n) \le I_{s_1,\dots,s_n}(z(s_1), \dots, z(s_n)) + n^{-1}.$$

Let $k := \sup_{n\ge1} J(l_n) < \infty$. By Lemma 5.1, $\{l_n\}$ is a bounded set of $\left(\mathcal{L}^{\Phi_1}\right)^*$. By the Alouglu theorem, $\{l_n\}$ is compact in the $\sigma(\left(\mathcal{L}^{\Phi_1}\right)^*, \mathcal{L}^{\Phi_1})$ topology. Hence, there exists a subnet $\{l_{n_\alpha}\}$ of $\{l_n\}$ which converges in the weak* topology. Let l be the limit of this subnet. We have that for each $j \ge 1$, $l(f(\cdot,s_j)) = z(s_j)$. Since the functions $t \in (T, d_{\Phi_1}) \mapsto l(f(\cdot,t)) \in \mathbb{R}$ and $t \in (T, d_{\Phi_1}) \mapsto z(t) \in \mathbb{R}$ are continuous, we get that $l(f(\cdot,t)) = z(t)$ for each $t \in T$. Hence, $I^{(2)}(z) \le J(l) \le I^{(1)}(z)$.

To prove (ii), we show that each sequence $\{z_n\}$ in $l_\infty(T)$, such that $I(z_n) \le k$, has a converging subnet. Take $l_n \in \left(\mathcal{L}^{\Phi_1}\right)^*$ such that $J(l_n) \le k + 1$ and $l_n(f(\cdot,t)) = z_n(t)$ for each $t \in T$. Since $\sup_{n\ge1} J(l_n) < \infty$, there exists a subnet l_{n_α} and $l \in \left(\mathcal{L}^{\Phi_1}\right)^*$ such that $l_{n_\alpha} \to l$ in the weak* topology. Hence, for each $t \in T$, $l_{n_\alpha}(f(\cdot,t)) \to l(f(\cdot,t))$. Since

$\{l_n\}$ and l are uniformly Lipschitz functions from (T, d_{Φ_1}) into $\mathbb{R}$, $\sup_{t\in T} |l_{n_\alpha}(f(\cdot,t)) - l(f(\cdot,t))| \to 0$.

Part (iii) follows from (i) and Lemma 2.2. □

Proof of Lemma 2.4. Fix $k > 1$. Since $K := \{z \in l_\infty(T) : I(z) \le k\}$ is a compact set of $l_\infty(T)$, K is totally bounded. Hence, (T, e) is a totally bounded set, where $e(s,t) = \sup_{z\in K} |z(s) - z(t)|$. We have that

$$\begin{aligned} e(s,t) &= \sup\{|z(s) - z(t)| : I(z) \le k\} \ge \sup\{|v-u| : I_{s,t}(u,v) < k\} \\ &= \sup\{|l(f(\cdot,s) - f(\cdot,t))| : l \in (\mathcal{L}_{\Phi_1})^*, J(l) < k\} \ge \left(\tfrac{\ln(2)}{\ln(8e^2)}\right) N_{\Phi_1}(f(\cdot,s) - f(\cdot,t)), \end{aligned}$$

by Lemma 5.1. So, the claim follows. □

Proof of Theorem 2.1. We apply Theorem 2.8 in Arcones (2003a). Let $d(s,t) = |t-s|$. By conditions (ii) and (iii), given $\epsilon > 0$ and $t \in T$, there exists a $\delta > 0$ such that

$$E[\sup_{s\in T, |s-t|\le\delta} |f(X,s) - f(X,t)|] \le \epsilon. \tag{5.3}$$

This implies that $\{f(X,t) : t \in T\}$ is a totally bounded set of L_1, i.e. condition (a.1) in Theorem 2.8 in Arcones (2003a) holds. Conditions (i) and (ii) imply that there exists a $\lambda > 0$ such that $E[\exp(\lambda F(X))] < \infty$, where $F(x) := \sup_{t\in T} |f(x,t)|$, i.e. condition (a.2) in Theorem 2.8 holds. Conditions (i) and (ii) and the compactness of T imply that given $\lambda > 0$, there exists a $\eta > 0$, such that

$$E[\exp(\lambda \sup_{d(s,t)\le\eta} |f(X,s) - f(X,t)|)] < \infty.$$

i.e. condition (a.3) in Theorem 2.8 in Arcones (2003a) holds. Since T is a compact set of $\mathbb{R}^d$ and (5.3) holds, given $\epsilon > 0$, there exists $t_1, \dots, t_m \in T$ and $\delta > 0$ such that for each $1 \le j \le m$,

$$E[\sup_{t\in T, |t-t_j|\le\delta} |f(X,t) - f(X,t_j)|] \le \epsilon$$

and $T \subset \cup_{j=1}^m \{t \in \mathbb{R}^d : |t - t_j| \le \delta\}$. Hence, by the Blum–DeHardt theorem (see for example Theorem 7.1.5 in Dudley, 1999), condition (a.4) in Theorem 2.8 in Arcones (2003a) holds. □

In the proof of Theorem 3.1, we will use the following lemma:

Lemma 5.2. *Let Θ be a convex set of $\mathbb{R}^d$, let K be a compact convex set contained in Θ, let $t_0 \in K$ and let $g : \Theta \to \mathbb{R}$ be a convex function. If $g(t_0) < \inf_{t\in\partial K} g(t)$, where $t_0 \in K$, then $\inf_{t\in\partial K} g(t) \le \inf_{t\notin K} g(t)$.*

Proof. Take $t \notin K$. Let $C_t := \{u \in \mathbb{R} : t_0 + u(t-t_0) \in \Theta\}$ and let $r_t : C_t \to \mathbb{R}$ defined by $r_t(u) = g(t_0 + u(t-t_0))$, $u \in C_t$. Let $a = \sup\{u \in \mathbb{R} : t_0 + u(t-t_0) \in K\}$. Since K is a compact set, $t_0 \in K^o$ and $t \notin K$, $0 < a < 1$ and $t_0 + a(t-t_0) \in \partial K$. By convexity of the function r_t, $a^{-1}(r_t(a) - r_t(0)) \le (1-a)^{-1}(r_t(1) - r_t(a))$. Using that $0 < \inf_{t\in\partial K} g(t) - g(t_0) \le r_t(a) - r_t(0)$, we get that $0 < r_t(1) - r_t(a) = g(t) - g(t_0 + a(t-t_0))$. Hence, $\inf_{s\in\partial K} g(s) < g(t)$. □

Proof of Theorem 3.1. First, we prove that

$$\lim_{m\to\infty} \limsup_{n\to\infty} n^{-1} \ln\left(\mathbb{P}\{\hat\theta_n \notin K_m\}\right) = -\infty. \tag{5.4}$$

By Lemma 5.2,

$$\mathbb{P}\{\hat{\theta}_n \notin K_m\} \le \mathbb{P}\{G_n(\theta) \ge \inf_{t\in\partial K_m} G_n(t)\}.$$

By Corollary 3.5 in Arcones (2003b), for each compact set $T \subset \Theta$, $\{G_n(t) : t \in T\}$ satisfies the LPD. In particular, we may take $T = \{\theta\} \cup \partial K_m$. Since the set $\{z \in l_\infty(T) : z(t) \ge \inf_{t\in\partial K_m} z(t)\}$ is a closed set of $l_\infty(T)$, we have that

$$\begin{aligned}&\limsup_{n\to\infty} n^{-1}\ln\left(\mathbb{P}\{G_n(\theta) \ge \inf_{t\in\partial K_m} G_n(t)\}\right)\\ &\le -\inf\{J(l) : (\mathcal{L}^{\Phi_1})^*, \quad \text{and} \quad l(g(\cdot,\theta)) \ge \inf_{t\in\partial K_m} l(g(\cdot,t))\}.\end{aligned}$$

Using that if $J(l) < \infty$, then $l(g(\cdot,t)), t \in \Theta$, is a continuous function, (2.6) and Lemma 2.2, we have that

$$\begin{aligned}(5.5)\qquad &\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(g(\cdot,\theta)) \ge \inf_{t\in\partial K_m} l(g(\cdot,t))\}\\ &= \inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(g(\cdot,\theta)) \ge l(g(\cdot,t)) \text{ for some } t \in \partial K_m\}\\ &= \inf_{t\in\partial K_m}\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(g(\cdot,\theta)) \ge l(g(\cdot,t))\}\\ &= \inf_{t\in\partial K_m}\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(g(\cdot,\theta)) = l(g(\cdot,t))\}\\ &= \inf_{t\in\partial K_m}\left(-\inf_{\lambda\in\mathbb{R}}\ln\left(E[\exp(\lambda(g(X,t)-g(X,\theta)))]\right)\right)\\ &= -\sup_{t\in\partial K_m}\inf_{\lambda\in\mathbb{R}}\ln\left(E[\exp(\lambda(g(X,t)-g(X,\theta)))]\right).\end{aligned}$$

Hence, by condition (iv), (5.4) holds.

Next, we prove that if $J(l) < \infty$, where $l \in \left(\mathcal{L}^{\Phi_1}\right)^*$, then the convex function $l(g(\cdot,t)), t \in \Theta$, has a minimum on Θ. By (iv) and (5.5), for m large enough, $l(g(\cdot,\theta)) < \inf_{t\in\partial K_m} l(g(\cdot,t))$. Hence, Lemma 5.2 implies that $l(g(\cdot,\theta)) < \inf_{t\notin K_m} l(g(\cdot,t))$. Therefore, the function $l(g(\cdot,t)), t \in \Theta$, has a minimum on Θ.

To prove that for each open set U,

$$(5.6)\qquad \liminf_{n\to\infty} n^{-1}\ln\mathbb{P}\{\hat{\theta}_n \in U\} \ge -I(U)$$

it suffices to prove that for each $t \in \mathbb{R}^d$ and each $\epsilon > 0$,

$$(5.7)\qquad \liminf_{n\to\infty} n^{-1}\ln(\mathbb{P}\{\hat{\theta}_n \in B(t,\epsilon)\}) \ge -I(t).$$

If $I(t) = \infty$, (5.7) is obviously satisfied. Assume that $I(t) < \infty$. By Lemma 5.2,

$$\mathbb{P}\{\hat{\theta}_n \in B(t,\epsilon)\} \ge \mathbb{P}\{G_n(t) < \inf_{t_1\in\Theta:|t_1-t|=\epsilon} G_n(t_1)\}.$$

Since the set $\{z \in l_\infty(T) : z(t) < \inf_{t_1\in\Theta:|t_1-\theta|=\epsilon} z(t_1)\}$, where $T = \{t\} \cup \{t_1 \in \Theta : |t_1 - t| = \epsilon\}$, is an open set of $l_\infty(T)$,

$$\begin{aligned}(5.8)\qquad &\liminf_{n\to\infty} n^{-1}\ln\mathbb{P}\{\hat{\theta}_n \in B(t,\epsilon)\}\\ &\ge -\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(g(\cdot,t)) < \inf_{t_1\in\Theta:|t_1-t|=\epsilon} l(g(\cdot,t_1))\}.\end{aligned}$$

Using that if $J(l) < \infty$, and $l(h(\cdot,t)) = 0$, then the function $l(g(\cdot,t)), t \in \Theta$, has a minimum at t, and condition (vi), we get that

$$\begin{aligned}&\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(h(\cdot,t)) = 0, l(g(\cdot,t)) = \inf_{t_1:|t_1-t|=\epsilon} l(g(\cdot,t_1))\}\\ &= \inf_{t_1\in\Theta:|t_1-t|=\epsilon}\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(h(\cdot,t)) = 0, l(g(\cdot,t)) = l(g(\cdot,t_1))\}\\ &= \inf_{t_1\in\Theta:|t_1-t|=\epsilon}\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(h(\cdot,t)) = l(h(\cdot,t_1)) = 0\} > I(t).\end{aligned}$$

So,

$$I(t) = \inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(h(\cdot,t)) = 0, l(g(\cdot,t)) < \inf_{t_1:|t_1-t|<\epsilon} l(g(\cdot,t_1))\} \tag{5.9}$$
$$\geq \inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(g(\cdot,t)) < \inf_{t_1\in\Theta:|t_1-t|=\epsilon} l(g(\cdot,t_1))\}.$$

(5.8) and (5.9) imply (5.7).

We claim that for each closed set F,

$$\limsup_{n\to\infty} n^{-1}\ln \mathbb{P}\{\hat{\theta}_n \in F\} \leq -I(F). \tag{5.10}$$

We may assume that $I(F) < \infty$. The case $I(F) = \infty$ is similar. Take m such that

$$\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(g(\cdot,\theta)) \geq \inf_{t\in\partial K_m} l(g(\cdot,t))\} > I(F). \tag{5.11}$$

We have that

$$\mathbb{P}\{\hat{\theta}_n \in F\} \leq \mathbb{P}\{\hat{\theta}_n \in F\cap K_m\} + \mathbb{P}\{\hat{\theta}_n \notin K_m\}.$$

Using that the set $\{z \in l_\infty(K_m) : \inf_{t\in K_m\cap F} z(t) = \inf_{t\in K_m} z(t)\}$ is a closed set of $l_\infty(K_m)$, we get that

$$\begin{aligned}
&\limsup_{n\to\infty} n^{-1}\ln\left(\mathbb{P}\{\hat{\theta}_n \in F\cap K_m\}\right)\\
&\leq \limsup_{n\to\infty} n^{-1}\ln\left(\mathbb{P}\{\inf_{t\in F\cap K_m} G_n(\theta) = \inf_{t\in K_m} G_n(t)\}\right)\\
&\leq -\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, \inf_{t\in F\cap K_m} l(g(\cdot,t)) = \inf_{t\in K_m} l(g(\cdot,t))\}.
\end{aligned}$$

Let $l \in \left(\mathcal{L}^{\Phi_1}\right)^*$ such that $J(l) < \infty$ and $\inf_{t\in F\cap K_m} l(g(\cdot,t)) = \inf_{\theta\in K_m} l(g(\cdot,t))$. If $l(g(\cdot,\theta)) \geq \inf_{t\in\partial K_m} l(g(\cdot,t))$, then, by (5.11), $J(l) \geq I(F)$. If $l(g(\cdot,\theta)) < \inf_{\theta\in\partial K_m} l(g(\cdot,t))$, then, by Lemma 5.2, $\inf_{t\in\Theta} l(g(\cdot,t)) = \inf_{t\in K_m} l(g(\cdot,t)) = \inf_{t\in F\cap K_m} l(g(\cdot,t))$. Thus, there exists $t_l \in F\cap K_m$ be such that $\inf_{t\in\Theta} l(g(\cdot,t)) = l(g(\cdot,t_l))$. By Lemma 5.1 (ii) and hypothesis (v), $l(g(\cdot,t))$, $t\in\Theta$, is differentiable in Θ^o and $\nabla(l(g(\cdot,t))) = l(h(\cdot,t))$. Hence, $l(h(\cdot,t_l)) = 0$,

$$J(l) \geq \inf_{t\in F}\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(h(\cdot,t)) = 0\}.$$

and (5.10) holds. □

Proof of Theorem 3.2. (i) By (3.5), it suffices to prove that $I_\theta(t) \geq K(f(\cdot,t), f(\cdot,\theta))$. We have that

$$\begin{aligned}
&I_\theta(t) = -\inf_{\lambda\in\mathbb{R}^d}\ln\left(E_\theta\left[\exp(\lambda'\nabla_t\ln(X,t))\right]\right) \geq -\ln\left(E_\theta\left[\exp(\lambda'_{t,\theta}\nabla_t\ln(X,t))\right]\right)\\
&= -\ln\left(E_\theta\left[\exp(\ln(f(X,t)/f(X,\theta)) - c_{t,\theta})\right]\right) = c_{t,\theta}
\end{aligned}$$

and

$$\begin{aligned}
&K(f(\cdot,t), f(\cdot,\theta)) = E_t\left[\ln(f(X,t)/f(X,\theta))\right]\\
&= E_t\left[\lambda'_{t,\theta}\nabla_t\ln f(X,t) + c_{t,\theta}\right] = \int(\lambda'_{t,\theta}\nabla_t f(x,t) + c_{t,\theta}f(x,t))\,d\mu(x) = c_{t,\theta}.
\end{aligned}$$

(ii) Let $g(x) = \ln(f(x,t)/f(x,\theta)) - \lambda'_{t,\theta}\nabla_t\ln f(x,t)$. Then,

$$-\ln\left(E_\theta\left[\exp(\lambda'_{t,\theta}\nabla_t\ln(X,t))\right]\right) = -\ln\int e^{-g(x)}f(x,t)\,d\mu(x)$$

and

$$K(f(\cdot,t),f(\cdot,\theta)) = \int \ln(f(x,t)/f(x,\theta))f(x,t)\,d\mu(x)$$
$$= \int (\lambda'_{t,\theta}\nabla_t \ln f(x,t) + g(x))f(x,t)\,d\mu(x) = \int g(x)f(x,t)\,d\mu(x).$$

Hence,

$$\exp\left(\int -g(x)f(x,t)\,d\mu(x)\right) = \int \exp(-g(x))f(x,t)\,d\mu(x)$$

which implies that g is a constant P_t a.s. $\square$

PROOF OF THEOREM 3.3. Let $g(x) = \ln f(x)$. We have that $\lambda(t)g'(x) + c(t) = g(x) - g(x-t)$, for each $x,t \in \mathbb{R}$. Taking derivatives with respect to t and with respect to x, we get that for each $x,t \in \mathbb{R}$,

$$\lambda'(t)g''(x) = g''(x-t). \tag{5.12}$$

If $g''(x) = 0$, for each $x \in \mathbb{R}$, then g is a linear function, which contradicts the fact that f is a pdf. So, there exists $x_0 \in \mathbb{R}$, such that $g''(x_0) \neq 0$. Using (5.12), we get that $\lambda'(t) = g''(x_0 - t)/g''(x_0)$. So, from (5.12), we get that $g''(x_0 - t)g''(x) = g''(x_0)g''(x-t)$, for each $x,t \in \mathbb{R}$. Hence, $h(x+y) = h(x)h(y)$, for each $x,y \in \mathbb{R}$, where $h(x) = g''(x+x_0)/g''(x_0)$. This means that h satisfies the Cauchy's exponential equation, $h(0) = 1$ and it is measurable. So, by Theorem 5 in Aczél and Dhombres (1989), $h(x) = e^{ax}$ for some $a \in \mathbb{R}$. Hence, $g''(x) = be^{ax}$ for some $a \in \mathbb{R}$ and some $b \neq 0$. If $a = 0$, then $g(x) = 2^{-1}bx^2 + cx + d$, for some $b \neq 0$ and some $c,d \in \mathbb{R}$, and f has a normal pdf. If $a \neq 0$, then $g(x) = a^{-2}be^{ax} + cx + d$, for some $a \neq 0$, $b \neq 0$ and $c,d \in \mathbb{R}$. Since $e^{g(x)}$ is pdf, $b < 0$ and $ac > 0$. Taking $\alpha = a^{-1}c$, $\gamma = a$ and $\theta = -a^{-1}\ln\left(-a^{-1}c^{-1}b\right)$, we get that $\ln f(x) = -\alpha e^{\gamma(x-\theta)} + \alpha\gamma(x-\theta) + d + \alpha\gamma\theta$, where $\alpha > 0$, $\gamma \neq 0$ and $\theta \in \mathbb{R}$. Hence, f has the form in (3.8). $\square$

We will need the following lemma:

LEMMA 5.3. *Let X be a r.v. defined in a measurable space $(S,\mathcal{S})$. Let $h : S\times T \to \mathbb{R}$ be a function such that $h(\cdot,t)$ is measurable for each $t \in T$, where T is an index set. Let $t_0 \in T$. Suppose that for each $t \in T$, $h(X,t) \in \mathcal{L}^{\Phi_1}$. Then, $M(t) = \inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, l(h(\cdot,t_0)) = l(h(\cdot,t)) = 0\}$, $t \in T$, defines a lower semicontinuous function in (T, d_{Φ_1}), where $d_{\Phi_1}(s,t) = N_{\Phi_1}(f(X,s) - f(X,t))$.*

PROOF. We need to prove that if $d_{\Phi_1}(t_n,t) \to 0$, then, $\liminf_{n\to\infty} M(t_n) \geq M(t)$. We may assume that $c := \liminf_{n\to\infty} M(t_n) < \infty$. There exists $l_n \in \left(\mathcal{L}^{\Phi_1}\right)^*$ such that $J(l_n) \leq M(t_n) + n^{-1}$ and $l_n(h(\cdot,t_0)) = l_n(h(\cdot,t_n)) = 0$. Since $\sup_{n\geq 1} J(l_n) < \infty$, there exists a subnet l_{n_α} and $l \in \left(\mathcal{L}^{\Phi_1}\right)^*$ such that $l_{n_\alpha} \to l$ in the weak* topology. This implies that $J(l) \leq c$ and $l(h(\cdot,t_0)) = 0$. By Lemma 5.1, $l_{n_\alpha}(h(\cdot,t_{n_\alpha})) \to l(h(\cdot,t))$. Hence, $J(l) \geq M(t)$ and the claim follows. $\square$

PROOF OF THEOREM 3.4. Without loss of generality, we may assume that $\theta = 0$. We apply Theorem 3.1 to $g(x,t) = \ln(f(x-t)/f(x))$, $\Theta = \mathbb{R}$ and $K_m = [-m,m]$. Since $G_n(\cdot)$, is a continuous function and $\lim_{t\to\pm\infty} G_n(t) = \infty$, there exists a $\hat{\theta}_n$ such that $G_n(\hat{\theta}_n) = \inf_{t\in\Theta} G_n(t)$. Hence, (i) in Theorem 3.1 holds. Condition (ii) in Theorem 3.1 follows from (3.4). Using that $f'(\cdot)/f(\cdot)$ is a decreasing function, for each $t > 0$ and each $x \in \mathbb{R}$,

$$tf'(x)/f(x) \leq -\ln(f(x-t)/f(x)) \leq tf'(x-t)/f(x-t).$$

Hence, for each $\lambda \in \mathbb{R}$, $E_0[\exp(\lambda \ln(f(X-t)/f(X)))] < \infty$. A similar argument holds for $t < 0$. Hence, condition (iii) in Theorem 3.1 follows. Condition (iv) in Theorem 3.1 follows from (v). Again using that $f'(\cdot)/f(\cdot)$ is a decreasing function, for each $v > 0$, each $t \in \mathbb{R}$ and each $x \in \mathbb{R}$,

$$\begin{aligned} &0 \le -v^{-1}\ln(f(x-t-v)/f(x-t)) + (f'(x-t)/f(x-t)) \\ &\le -f'(x-t-v)/f(x-t-v) + (f'(x-t)/f(x-t)). \end{aligned}$$

By the monotone convergence theorem,

$$N_{\Phi_1}(-(f'(X-t-v)/f(X-t-v)) + (f'(X-t)/f(X-t))) \to 0, \text{ as } v \to 0+.$$

A similar argument holds if $v \to 0-$. Hence, condition (v) in Theorem 3.1 follows. By Lemma 5.3, to prove condition (vi) in Theorem 3.1, it suffices to prove that if $t_1 \ne t_2$, then

$$-\inf_{\lambda \in \mathbb{R}} M(\lambda, 0) < -\inf_{\lambda_1,\lambda_2 \in \mathbb{R}} M(\lambda_1, \lambda_2), \tag{5.13}$$

where

$$M(\lambda_1, \lambda_2) := E_0[\exp(\lambda_1(f'(X-t_1)/f(X-t_1)) + \lambda_2(f'(X-t_2)/f(X-t_2)))].$$

Since $\lim_{x\to\pm\infty} f(x) = 0$, $\lim_{x\to-\infty} f'(x)/f(x) > 0$ and $\lim_{x\to\infty} f'(x)/f(x) < 0$. Hence

$$\mathbb{P}_0\{f'(X-t_1)/f(X-t_1) < 0\} > 0 \text{ and } \mathbb{P}_0\{f'(X-t_1)/f(X-t_1) > 0\} > 0.$$

Hence, by (2.12), there exists $\lambda_1^* \in \mathbb{R}$, such that

$$M(\lambda_1^*, 0) = -\inf_{\lambda \in \mathbb{R}} M(\lambda, 0) \tag{5.14}$$

Since $-\inf_{\lambda\in\mathbb{R}} M(\lambda, 0) \le -\inf_{\lambda_1,\lambda_2\in\mathbb{R}} M(\lambda_1, \lambda_2)$, to prove (5.13), it suffices to prove that

$$-M(\lambda_1^*, 0) < -\inf_{\lambda_2 \in \mathbb{R}} M(\lambda_1^*, \lambda_2). \tag{5.15}$$

The derivative at zero of the function $M(\lambda_1^*, \lambda_2)$, $\lambda_2 \in \mathbb{R}$, is

$$E_0[(f'(X-t_2)/f(X-t_2))\exp(\lambda_1^* f'(X-t_1)/f(X-t_1))]$$

If we show that the previous number is different from zero, then (5.15) will follow. Since λ_1^* satisfies (5.14),

$$E_0[(f'(X-t_1)/f(X-t_1))\exp(\lambda_1^* f'(X-t_1)/f(X-t_1))] = 0.$$

Since $f'(\cdot)/f(\cdot)$ is a decreasing function and $t_1 \ne t_2$,

$$E_0[(f'(X-t_2)/f(X-t_2))\exp(\lambda_{t_1} f'(X-t_1)/f(X-t_1))] \ne 0.$$

Therefore, condition (vi) in Theorem 3.1 follows. □

Proof of Theorem 3.5. We apply Theorem 3.1 with $g(x,t) = -t'x + \psi(t)$ and $h(x,t) = -x + \nabla\psi(t)$. Condition (i) in Theorem 3.1 is assumed. Condition (ii) follows from (3.4). Since $\theta \in \Theta^o$, there exists a $\lambda_0 > 0$, such that $\int e^{\lambda_0 |x| + \theta' x} \mu(x) < \infty$. This implies condition (iii) in Theorem 3.1. We have that

$$\begin{aligned} &E_\theta[\exp(\lambda(g(X,\theta) - g(X,t)))] = \int \exp(\lambda(-\theta' x + \psi(\theta) + t' x - \psi(t))) e^{\theta' x - \psi(\theta)} \, d\mu(x) \\ &= \exp\left(\psi(\theta + \lambda(t-\theta)) - \psi(\theta) - \lambda(\psi(t) - \psi(\theta))\right). \end{aligned}$$

Hence, (ii) implies condition (iv) in Theorem 3.1. Condition (v) in Theorem 3.1 holds because for each $t \in \Theta^o$, $\nabla\psi(t)$ exists. For each $t_1, t_2 \in \Theta^o$ with $t_1 \neq t_2$, taking $\lambda_1 = -\lambda_2 = u(\nabla\psi(t_1) - \nabla\psi(t_2))$, we get that

$$\inf_{\lambda_1, \lambda_2 \in \mathbb{R}} E[\exp(\lambda_1' h(X, t_1) + \lambda_2' h(X, t_2))] \leq \inf_{u \in \mathbb{R}} \exp(u |\nabla\psi(t_1) - \nabla\psi(t_2)|^2) = 0.$$

Hence, condition (vi) in Theorem 3.1 holds. □

Proof of Theorem 3.6. We only prove the case $t > \theta$. The cases $t = \theta$ and $t < \theta$ is similar. Let $H_n(t) = n^{-1} \sum_{j=1}^n h(X_j, t)$ and let $H_n(t-) = n^{-1} \sum_{j=1}^n h(X_j, t-)$. We have that for each $t > \theta$,

$$\{H_n(t-) < 0\} \subset \{\hat\theta_n^{(1)} \geq t\} \subset \{\hat\theta_n \geq t\} \subset \{\hat\theta_n^{(2)} \geq t\} = \{H_n(t-) \leq 0\}.$$

and

$$\{H_n(t+) < 0\} = \{\hat\theta_n^{(1)} > t\} \subset \{\hat\theta_n > t\} \subset \{\hat\theta_n^{(2)} > t\} \subset \{H_n(t+) \leq 0\}.$$

By hypotheses (ii) and (iii), we have that $0 = E[h(X,\theta)] \leq E[h(X,t-)] \leq E[h(X,t+)]$ and if $\theta < t < s$, then

$$\begin{aligned} &\sup\{u : \mathbb{P}\{h(X,t-) < u\} = 0\} \leq \sup\{u : \mathbb{P}\{h(X,t+) < u\} = 0\} \\ &\leq \sup\{u : \mathbb{P}\{h(X,s) < u\} = 0\} < 0. \end{aligned}$$

Hence, by (2.9) and (2.11),

$$\lim_{n\to\infty} n^{-1} \ln\left(\mathbb{P}\{\hat\theta_n \geq t\}\right) = \inf_{\lambda \in \mathbb{R}} \ln(E[\exp(\lambda h(X, t-))]).$$

and

$$\lim_{n\to\infty} n^{-1} \ln\left(\mathbb{P}\{\hat\theta_n > t\}\right) = \inf_{\lambda \in \mathbb{R}} \ln(E[\exp(\lambda h(X, t+))]). \square$$

Proof of Theorem 3.7. We apply Theorem 3.6 to $h(x,t) = \psi'(t) - x$, $t \in \Theta^o$. Note that Theorem 3.6 holds true if the range of t is restricted to Θ^o. Since μ is nondegenerate, ψ is a striclty convex function. Hence, $h(x,t) = \psi'(t) - x$ is an increasing function on t, for each fixed x. It is easy to see that hypotheses (i) and (ii) in Theorem 3.6 hold. Let $a_\mu = \sup\{t \in \mathbb{R} : \mu(-\infty, t) = 0\}$ and $b_\mu = \inf\{t \in \mathbb{R} : \mu(t, \infty) = 0\}$. Then, for each $t \in \Theta^0$, $a_\mu < \psi'(t) < b_\mu$. The support of X contains (a_μ, b_μ). Hence, hypothesis (iii) in Theorem 3.6 hold. We have that

$$\begin{aligned} &\inf_{\lambda \in \mathbb{R}} \ln(E_\theta[\exp(\lambda h(X,t))]) = \inf_{\lambda \in \mathbb{R}} \ln \int \exp(\lambda(\psi'(t) - x)) \exp(x\theta - \psi(\theta)) \, d\mu(x) \\ &= \inf_{\lambda \in \mathbb{R}} (\lambda \psi'(t) + \psi(\theta - \lambda) - \psi(\theta)) = \psi(t) - \psi(\theta) - (t-\theta)\psi'(t) = -K(f(\cdot, t), f(\cdot, \theta)), \end{aligned}$$

because of convexity of the function ψ,

$$\begin{aligned} &\inf_{\lambda \in \mathbb{R}} (\lambda \psi'(t) + \psi(\theta - \lambda) - \psi(\theta)) - (\psi(t) - \psi(\theta) - (t-\theta)\psi'(t)) \\ &= \inf_{\lambda \in \mathbb{R}} (\psi(\theta - \lambda) - \psi(t) - (\theta - \lambda - t)\psi'(t)) \\ &= \inf_{u \in \mathbb{R}} (\psi(u) - \psi(t) - (u - t)\psi'(t)) = 0. \qquad \square \end{aligned}$$

Proof of Theorem 3.8. First, we prove that

$$\lim_{m\to\infty}\limsup_{n\to\infty} n^{-1}\ln\left(\mathbb{P}\{\hat{\theta}_n\notin K_m\}\right)=-\infty \tag{5.16}$$

Given $1>\epsilon>0$, we claim that for each $m\ge 1$,

$$\{n^{-1}\sum_{j=1}^n R_m(X_j)\le 1-\epsilon\}\subset\{\hat{\theta}_n\in K_m\}. \tag{5.17}$$

If $n^{-1}\sum_{j=1}^n R_m(X_j)\le 1-\epsilon$, then for $t\in\Theta-K_m$,

$$\begin{aligned}&|H(t)|-|H_n(t)|\le|H_n(t)-H(t)|\le n^{-1}\textstyle\sum_{j=1}^n|h(X_j,t)-H(t)|\\ &\le |H(t)|n^{-1}\textstyle\sum_{j=1}^n R_m(X_j)\le(1-\epsilon)|H(t)|.\end{aligned}$$

So, for $t\in\Theta-K_m$, $|H_n(t)|\ge\epsilon|H(t)|\ge\epsilon\inf_{t\notin K_m}|H(t)|\ge\inf_{t\notin K_1}|H(t)|>0$. Therefore, (5.17) holds. By (2.13) and (5.17),

$$\begin{aligned}&n^{-1}\ln\left(\mathbb{P}\{\hat{\theta}_n\notin K_m\}\right)\le n^{-1}\ln\left(\mathbb{P}\{n^{-1}\textstyle\sum_{j=1}^n R_m(X_j)\ge 1-\epsilon\}\right)\\ &\le -\sup_{\lambda>0}(\lambda(1-\epsilon)-\ln(E[\exp(\lambda R_m(X))])).\end{aligned}$$

Letting $m\to\infty$, using (v), (5.16) follows.

Next, we prove that for each $t\in\mathbb{R}^d$ and each $\epsilon>0$,

$$\liminf_{n\to\infty} n^{-1}\ln(\mathbb{P}\{\hat{\theta}_n\in B(t,\epsilon)\})\ge -I(t). \tag{5.18}$$

We may assume that $I(t)<\infty$. Take an integer $m\ge 1$ such that

$$\limsup_{n\to\infty} n^{-1}\ln\left(\mathbb{P}\{\hat{\theta}_n\notin K_m\}\right)<-I(t).$$

We have that

$$\mathbb{P}\{\hat{\theta}_n\in B(t,\epsilon)\}\ge\mathbb{P}\{\inf_{t_1\in K_m,|t_1-t|\ge\epsilon}|H_n(t_1)|>0\}-\mathbb{P}\{\hat{\theta}_n\notin K_m\}.$$

Since the set $\{z\in l_\infty(T):\inf_{t_1\in T}|z(t_1)|>0\}$ is an open set of $l_\infty(T)$, where $T=\{t_1\in K_m:|t_1-t|\ge\epsilon\}$,

$$\begin{aligned}&\liminf_{n\to\infty} n^{-1}\ln\mathbb{P}\{\inf_{t_1\in K_m,|t_1-t|\ge\epsilon}|H_n(t_1)|>0\}\\ &\ge -\inf\{J(l): l\in\left(\mathcal{L}^{\Phi_1}\right)^*,\inf_{t_1\in K_m,|t_1-t|\ge\epsilon}|l(h(\cdot,t_1))|>0\}.\end{aligned}$$

From the previous estimations, to finish the proof of (5.18), we need to get that

$$\inf\{J(l): l\in\left(\mathcal{L}^{\Phi_1}\right)^*,\inf_{t_1\in K_m,|t_1-t|\ge\epsilon}|l(h(\cdot,t_1))|>0\}\le I(t). \tag{5.19}$$

By condition (vii), $I(t)<q$, where

$$q:=\inf\{J(l): l\in\left(\mathcal{L}^{\Phi_1}\right)^*, l(h(\cdot,t))=l(h(\cdot,t_1))=0,\ \text{for some } t_1\in K_m,|t_1-t|\ge\epsilon\}.$$

Thus, for each $\tau > 0$, there exists $l_0 \in \left(\mathcal{L}^{\Phi_1}\right)^*$, with $l(h(\cdot, t)) = 0$, $J(l_0) < I(t) + \tau$ and $J(l_0) < q$. Since $l_0(h(\cdot, t))$, $t \in \Theta$, is continuous, $\inf_{t_1 \in K_m, |t_1 - t| \geq \epsilon} |l_0(h(\cdot, t_1))| > 0$ and

$$\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, \inf_{t_1 \in K_m, |t_1 - t| \geq \epsilon} |l(h(\cdot, t_1))| > 0\} \leq J(l_0) < I(t) + \tau.$$

Since $\tau > 0$ is arbitrary, (5.19) holds.

We claim that for each closed set F, we have that

$$\limsup_{n \to \infty} n^{-1} \ln \mathbb{P}\{\hat{\theta}_n \in F\} \leq -I(F). \tag{5.20}$$

Assume that $I(F) < \infty$. The case $I(F) = \infty$ is similar. Take $m \geq 1$ such that

$$\limsup_{n \to \infty} n^{-1} \ln \left(\mathbb{P}\{\hat{\theta}_n \notin K_m\}\right) < -I(F).$$

We have that

$$\mathbb{P}\{\hat{\theta}_n \in F\} \leq \mathbb{P}\{\hat{\theta}_n \notin K_m\} + \mathbb{P}\{\inf_{t \in F \cap K_m} |H_n(t)| = 0\}.$$

Since the set $\{z \in l_\infty(F \cap K_m) : \inf_{t \in F \cap K_m} |z(t)| = 0\}$ is a closed set of $l_\infty(F \cap K_m)$,

$$\begin{aligned} &\limsup\nolimits_{n \to \infty} n^{-1} \ln(\mathbb{P}\{\inf\nolimits_{t \in F \cap K_m} |H_n(t)| = 0\}) \\ &\leq -\inf\{J(l) : l \in \left(\mathcal{L}^{\Phi_1}\right)^*, \inf\nolimits_{t \in F \cap K_m} |l(h(\cdot, t))| = 0\} = -I(F \cap K_m) \leq -I(F). \square \end{aligned}$$

Proof of Theorem 4.1. By algebra

$$-n^{-1} \ln \left(\frac{L(\theta)}{\sup_{t \in \Theta} L(t)}\right) = (\hat{\theta}_n - \theta)' \bar{X}_n - \psi(\hat{\theta}_n) + \psi(\theta).$$

By (3.10),

$$K(f(\cdot, \hat{\theta}_n), f(\cdot, \theta)) = (\hat{\theta}_n - \theta)' \nabla \psi(\hat{\theta}_n) - \psi(\hat{\theta}_n) + \psi(\theta).$$

Since the mle maximizes likelihood function, $\nabla \psi(\hat{\theta}_n) = \bar{X}_n$. $\square$

References

Arcones, M. A. (2003a). The large deviation principle for empirical processes. In *"High Dimensional Probability III"* (eds. J. Hoffmann-Jorgensen, M. B. Marcus and J. A. Wellner), 205–223. Birkhäuser, Boston.

Arcones, M. A. (2003b). The large deviation principle for stochastic processes I, *Theory of Probability & Its Applications*, **47**, 567-583.

Aczél, J. and Dhombres, J. (1989). *Functional Equations in Several Variables.* Cambridge University Press, Cambridge, UK.

Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics, *Annals of Mathematical Statistics*, **38**, 303–324.

Bahadur, R. R. (1971). *Some Limit Theorems in Statistics.* SIAM, Philadelphia, PA.

Bahadur, R. R.; Zabell, S. L. and Gupta, J. C. (1979). Large deviations, tests, and estimates. In *Asymptotic Theory of Statistical Tests and Estimation*, 33–64. Academic Press, New York.

Borovkov, A. A. and Mogulskii, A. A. (1992). Large deviations and testing statistical hypotheses. II: Large deviations of maximum points of random fields, *Siberian Advances in Mathematics*, vol. **2**, iss. **4**, 43-72.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, 9. Hayward, CA.

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

Casella, G. and Berger, R. L. (2002). *Statistical Inference.* Duxbury, Pacific Grove, CA.

Chernoff, H. (1952). A measure of asymptotic efficiency for tests of hypothesis based on the sum of observations, *Annals of Mathematical Statistics*, **23**, 493–507.

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

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

Dudley, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge.

Ferguson, T. S. (1962). Location and scale parameters in exponential families of distributions, *Annals of Mathematical Statistics*, **33**, 986–1001.

Fu, J. C. (1975). The rate of convergence of consistent point estimators, *Annals of Statistics*, **3**, 234–240.

Fu, J. C.; Li, G.; and Zhao, L. C. (1993). On large deviation expansion of distribution of maximum likelihood estimator and its application in large sample estimation. *Annals of the Institute of Statistical Mathematics*, **45**, 477–498

Jensen, J. L. and Wood, A. T. A. (1998). Large deviation and other results for minimum contrast estimators, *Annals of the Institute of Statistical Mathematics*, **50**, 673–695.

Joutard, C. (2004). Large deviations for M-estimators, *Mathematical Methods of Statistics*, **13**, 179–200.

Kester, A. D. M. and Kallenberg, W. C. M. (1986). Large deviations of estimators, *Annals of Statistics*, **14**, 648–664.

Kester, A. D. M. (1985). *Some Large Deviation Results in Statistics.* Centre for Mathematics and Computer Science, Amsterdam, The Netherlands.

Léonard, C. and Najim, J. (2002). An extension of Sanov's theorem. Application to the Gibbs conditioning principle, *Bernoulli*, **18**, 721–743.

Nikitin, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests.* Cambridge University Press, Cambridge, UK.

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

Rubin, H. and Rukhin, A. L. (1983). Convergence rates of large deviations probabilities for point estimators, *Statistics and Probability Letters*, **1**, 197–202.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Sievers, G. L. (1978). Estimates of location: a large deviation comparison, *Annals of Statistics*, **6**, 610–618.

Skovgaard, I, M. (1990). On the density of minimum contrast estimators, *Annals of Statistics*, **18**, 779–789.