Bahadur efficiency of the likelihood ratio test*

Miguel A. Arcones

Department of Mathematical Sciences

Binghamton University

Binghamton, NY 13902. USA.

E-mail: arcones@math.binghamton.edu.

June 6, 2005

Abstract

We present a new approach to study the Bahadur efficiency of likelihood tests. Our approach is based on the large deviation principle for empirical processes. We prove that the likelihood ratio test is Bahadur asymptotically optimal under mild sufficient conditions. Our results apply to common families of distributions such as location and scale families.

Running Title: likelihood ratio tests

1 Introduction

A natural definition of efficiency of tests was given by Bahadur (1965, 1967, 1971). This definition is as follows. Let $\{f(\cdot,\theta) : \theta \in \Theta\}$ be a family of pdf's on a measurable space $(S, \mathcal{S})$ with respect to a measure μ, where Θ is a Borel subset of $\mathbb{R}^d$. Let $X_1, \dots, X_n$ be i.i.d.r.v.'s with values in $(S, \mathcal{S})$ and pdf $f(\cdot, \theta)$, for some unkonwn value of $\theta \in \Theta$. Let $\Theta_0 \subset \Theta$. Consider the hypothesis testing problem $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \notin \Theta_1$, where $\Theta_1 := \Theta - \Theta_0$. The p–value of a test is the smallest significance level at which the null hypothesis can be rejected.

* *AMS 2000 subject classifications.* Primary 62F05; secondary 60F10.

Key words and phrases. Bahadur efficiency, likelihood ratio test, large deviation principle, empirical processes.

Suppose that a test rejects H_0 if $T_n \geq c$, where $T_n := T_n(X_1, \ldots, X_n)$ is a statistic and c is a constant. Then, the p-value of the test is $H_n(T_n)$, where

$$H_n(t) := \sup_{\theta_0 \in \Theta_0} \mathbb{P}_{\theta_0}(T_n \geq t), \tag{1.1}$$

where $\mathbb{P}_\theta$ denotes the probability measure for which the data has pdf $f(\cdot, \theta)$. For $\theta_1 \in \Theta_1$ we would like that the p–value converges to zero as fast as possible. Bahadur (1967) (see also Raghavachari, 1970) proved that for any test

$$\liminf_{n \to \infty} n^{-1} \ln H_n(T_n) \geq -\inf\{K(f(\cdot, \theta_1), f(\cdot, \theta_0)) : \theta_0 \in \Theta_0\} \text{ a.s.} \tag{1.2}$$

when θ_1 obtains (the alternative θ_1 holds), where $K(f(\cdot, \theta_1), f(\cdot, \theta_0)$ is the Kullback–Leibler information of the densities $f(\cdot, \theta_1)$ and $f(\cdot, \theta)$. Given densities f and g with respect to a probability measure μ, the Kullback–Leibler information of the densities f and g is defined by

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

A test is said to be Bahadur efficient if for each $\theta_1 \notin \Theta_0$,

$$\lim_{n \to \infty} n^{-1} \ln H_n(T_n) = -\inf\{K(f(\cdot, \theta_1), f(\cdot, \theta_0)) : \theta_0 \in \Theta_0\} \text{ a.s.} \tag{1.3}$$

when θ_1 obtains. For a review on Bahadur asymptotic optimality see Serfling (1980) and Nikitin (1995).

We consider the Bahadur efficiency of the likelihood ratio test. The likelihood ratio statistic is

$$\sup_{\theta \in \Theta} \prod_{j=1}^{n} f(X_j, \theta) / \sup_{\theta \in \Theta_0} \prod_{j=1}^{n} f(X_j, \theta). \tag{1.4}$$

We will prove that this statistic satisfies (1.3) for common families of distributions. Bahadur proved that the likelihood ratio test satisfies (1.3) assuming among other conditions that Θ_0 and Θ_1 are relatively compact sets with respect to the topology determined by the convergence in distribution. This condition is too stringent. In this paper, we present new sufficient conditions for the Bahadur efficiency of the likelihood ratio test which apply to common families of distributions.

The efficiency of the likelihood ratio test has being studied by several authors. Bahadur and Raghavachari (1972) consider the case of Markov chains. Oosterhoff and van Zwet (1970) consider the case of the multinomial distribution. Herr (1969) considers the case of the multinomial distribution. Efron and Truax (1968), Kallenberg (1978) and Kourouklis (1984) consider exponential families. Rublik (1989a, 1989b) considers several types of distributions..

Hsieh (1979a, 1979b) considers the Bahadur efficiency of several statistics for multivariate data. Besides the considered situation, the likelihood ratio test is optimal according with another criteria (see e.g Brown, 1971; Kolomiec, 1989; and Rublik, 1995, 1996).

Our results apply to common families of distribution such as location and scale families. Our results also apply to exponential families.

Given a set A of $\mathbb{R}^d$, $\mathrm{Int}(A)$ will denote the interior of A. By an abuse of notation, given $\theta_1, \theta_2 \in \Theta$, we define

$$K(\theta_1, \theta_2) = K(f(\cdot, \theta_1), f(\cdot, \theta_2)).$$

Our techniques are based on the (LDP) large deviation principle for empirical processes in Arcones (2003a, 2003b). In Section 2, we review the notation on the LDP of empirical processes. Section 3 contains the main results. The proofs are in Section 4.

2 Large deviations via empirical processes

In this section we review some results on the LDP for empirical processes. We refer to the large deviation principle to Deuschel and Stroock (1989) and Dembo and Zeitouni (1998). 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 $\mathcal{L}^{\Phi}(S, \mathcal{S})$ is defined as

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

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:

$$J(l) := \sup_{f\in\mathcal{L}^{\Phi_1}} \left(l(f) - \ln\left(E[e^{f(X)}]\right)\right),\ l \in (\mathcal{L}^{\Phi_1})^*. \tag{2.1}$$

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 $f(X) \geq 0$ a.s., then $l(f) \geq 0$.

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

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

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

$$\Psi_2(x) = x \ln\left(\frac{x}{e}\right) + 1, \text{ if } x > 0; \Psi_2(0) = 1; \quad \text{and } \Psi_2(x) = \infty, \quad \text{if } x < 0. \tag{2.3}$$

Besides, we have that

$$J(l_\gamma) = \sup_{f \in \mathcal{L}^{\Phi_1}} (E[f(X)\gamma(X) - \Phi_2(f(X))]) = E[\Psi_2(\gamma(X))], \tag{2.4}$$

where $\Phi_2(x) = e^x - 1$ (see (2.5) in Arcones ,2003b). The Fenchel–Legendre conjugate of Φ_2 is the function Ψ_2. We also have that if $l \in (\mathcal{L}^{\Phi_1})^*$ and $l(\mathbf{1}) = 1$, then

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

(see Lemma 2.1 in Arcones, 2003b).

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

$$I_f(t) := \sup_{\lambda \in \mathbb{R}} (\lambda t - \ln(E[\exp(\lambda f(X))])) \,, t \in \mathbb{R}$$

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

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

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 $t \geq \mu_f$,

$$\lim_{n\to\infty} n^{-1} \ln\left(\Pr\left\{n^{-1}\sum_{j=1}^n f(X_j) \geq t\right\}\right) = -I_f(t) \tag{2.6}$$

and for each $t \leq \mu_f$,

$$\lim_{n\to\infty} n^{-1} \ln \left(\Pr \left\{ n^{-1} \sum_{j=1}^{n} f(X_j) \leq t \right\} \right) = -I_f(t) \tag{2.7}$$

(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

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

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

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

(see Lemma 2.2 in Arcones, 2003b).

To deal with empirical processes, we will use the following theorem:

Theorem 2.1. *(Theorem 2.1 in Arcones, 2003b). Let T be a compact subset of $\mathbb{R}^d$. Let $\{f(\cdot, t) : t \in T\}$ be a collection of measurable functions on $(S, \mathcal{S})$. 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

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

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

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

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

$$I(z) = \inf\{J(l) : l \in (\mathcal{L}^{\Phi_1})^*, l(f(\cdot, t)) = z(t), \text{ for each } t \in T\}, \ z \in l_\infty(T). \tag{2.8}$$

3 Main Results

Let $(S, \mathcal{S})$ be a measurable space, let μ be a measure on $(S, \mathcal{S})$ and let $\{f(\cdot, \theta) : \theta \in \Theta\}$ be a family of pdf's with respect to the measure μ, 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 values in $(S, \mathcal{S})$ with pdf $f(\cdot, \theta)$ for some unknown value θ of Θ. Since the distribution of the r.v.'s depends on θ, E_θ will denote the expectation when the data comes from the pdf $f(\cdot, \theta)$. Similarly, we define $\mathbb{P}_\theta$, $\mathcal{L}_\theta^{\Phi_1}$, $(\mathcal{L}_\theta^{\Phi_1})^*$ and J_θ.

In the case of a simple null hypothesis, we present the following theorem:

Theorem 3.1. *With the notation above, let $\theta_0, \theta_1 \in \Theta$, let $d : \Theta \to (0, \infty)$ be a measurable function. Let $\Theta_m \subset \Theta$, $m \geq 1$. Suppose that:*

(i) $K(\theta_1, \theta_0) < \infty$.

(ii) $\{x \in S : f(x, \theta) > 0\}$ does not depend on θ.

(iii) $\{n^{-1} \sum_{j=1}^n \ln(f(X_j, \theta)/f(X_j, \theta_0)) : \theta \in \Theta_m\}$ satisfies the LDP in $l_\infty(\Theta_m)$ with speed n.

(iv)

$$\lim_{m\to\infty} \inf_{\lambda>0} E_{\theta_0}\left[\exp\left(\lambda \sup_{\theta\notin\Theta_m} (d(\theta))^{-1}(\ln f(X, \theta) - \ln f(X, \theta_0))\right)\right] = 0.$$

Then,

$$\lim_{n\to\infty} n^{-1} \ln H_n(T_n) = -K(\theta_1, \theta_0) \text{ a.s.} \tag{3.1}$$

when θ_1 obtains, where

$$T_n := \sup_{\theta\in\Theta} n^{-1} \sum_{j=1}^n (\ln(f(X_j, \theta)/f(X_j, \theta_0)),$$

and $H_n(t) := \mathbb{P}_{\theta_0}(T_n \geq t)$.

If $\{n^{-1} \sum_{j=1}^n \ln(f(X_j, t)/f(X_j, \theta_0)) : t \in \Theta\}$ satisfies the LDP in $l_\infty(\Theta)$ with speed n, then we can take $\Theta_m = \Theta$. The supremum over the empty set in (iv) is interpreted as $-\infty$.

Corollary 3.1. *With the notation above, let $\theta_0, \theta_1 \in \Theta$, let $d : \Theta \to (0, \infty)$ be a measurable function. Suppose that:*

(i) $K(\theta_1, \theta_0) < \infty$.

(ii) $\{x \in S : f(x, \theta) > 0\}$ does not depend on θ.

(iii) For each $\lambda > 0$, each $t \in \Theta_m$ and each $m \geq 1$, there exists a $\eta > 0$ such that

$$E_{\theta_0}\left[\exp\left(\lambda \sup_{\substack{s\in\Theta_m \\ |s-t|\leq\eta}} |\ln(f(X, s)/f(X, t))|\right)\right] < \infty,$$

where

$$\Theta_m := \{t \in \Theta : d(t, \Theta^c) \geq m^{-1} \text{ and } |t - \theta_0| \leq m\}.$$

(iv) For each $t \in \mathrm{Int}(\Theta)$,

$$\lim_{\epsilon \to 0} \sup_{\substack{s \in \Theta \\ |s-t| \leq \epsilon}} |\ln(f(X,s)/f(X,t))| = 0 \quad \mathbb{P}_{\theta_0} - \text{a.s.}$$

(v)

$$\lim_{m \to \infty} \inf_{\lambda > 0} E_{\theta_0} \left[\exp\left(\lambda \sup_{\theta \notin \Theta_m} (d(\theta))^{-1} (\ln(f(X,\theta)/f(X,\theta_0)) \right) \right] = 0.$$

Then, (3.1) holds.

Condition (ii) in the previous theorem is needed to avoid the case of families of pdf's whose support depends on the parameter such as the uniform distribution. Conditions (iii) and (iv) in the previous theorem are used to get the LDP of $\{n^{-1} \sum_{j=1}^n \ln(f(X_j,\theta)/f(X_j,\theta_0)) : \theta \in \Theta_m\}$. Condition (v) in the previous theorem is used to deal with the large deviations of the part outside the compact set Θ_m.

The previous corollary gives the Bahadur efficiency of the likelihood ratio test under different conditions than those in the literature. The previous corollary can be used to obtain the optimality of the likelihood ratio statistic for common families of pdf's. For a location family the previous corollary gives the following:

Corollary 3.2. *Let* μ *be a measure in* $\mathbb{R}^d$. *Let* f *be a positive continuous function in* $\mathbb{R}^d$ *with* $\int_{\mathbb{R}^d} f(x) d\,\mu(x) = 1$ *and* $\lim_{|x| \to \infty} f(x) = 0$. *Let* $d : \mathbb{R}^d \to (0, \infty)$. *Consider the location family of pdf's* $\{f(x - \theta) : \theta \in \mathbb{R}^d\}$. *Suppose that:*

(i) $K(\theta_1, \theta_0) < \infty$.

(ii) For each $\lambda > 0$ *and each* $\theta \in \mathbb{R}^d$, *there exists* $\eta > 0$ *such that*

$$\int_{\mathbb{R}^d} \exp\left(\lambda \sup_{|t| \leq \eta} |\ln(f(x-t)/f(x))| \right) f(x - \theta)\, d\mu(x) < \infty.$$

(iii) For each $\lambda > 0$ *there exists* $m \geq 1$ *such that*

$$\int_{\mathbb{R}^d} \exp\left(\lambda \sup_{|t| \geq m} \left((d(t))^{-1} \ln(f(x-t)/f(x)) \right) \right) f(x)\, d\mu(x) < \infty.$$

Then, (3.1) holds for each $\theta_0, \theta_1 \in \Theta$ *with* $\theta_0 \neq \theta_1$.

From the previous corollary, we obtain the following:

Corollary 3.3. *Consider the location family of pdf's* $\{a^{-1}\lambda^d e^{-\lambda^p|x-\theta|^p} : \theta \in \mathbb{R}^d\}$ *with respect to the Lebesgue measure, where* $\lambda, p > 0$ *and* $a = \int_{\mathbb{R}^d} e^{-|x|^p}\, dx$. *Then, (3.1) holds for each* $\theta_0, \theta_1 \in \mathbb{R}^d$ *with* $\theta_0 \neq \theta_1$.

For a scale family, Theorem 3.1 gives the following:

Corollary 3.4. *Let* μ *be a measure in* $\mathbb{R}^d$. *Let* f *be a nonnegative continuous function in* $\mathbb{R}^d$ *with* $\int_{\mathbb{R}^d} f(x) d\,\mu(x) = 1$ *and* $\lim_{|x|\to\infty} f(x) = 0$. *Consider the scale family of pdf's* $\{\theta^{-1} f(\theta^{-1}x) : \theta \in \Theta\}$, *where* $\Theta = (0, \infty)$. *Let* $\theta_0, \theta_1 > 0$. *Suppose that:*

(i) $K(\theta_1, \theta_0) < \infty$.

(ii) For each x *with* $f(x) > 0$ *and each* $\theta > 0$, $f(\theta x) > 0$.

(iii) For each $\lambda > 0$ *and each* $\theta > 0$, *there exists* $\eta > 0$ *such that*

$$\int_{\mathbb{R}^d} \exp\left(\lambda \sup_{|t-1|\le\eta} |\ln(t^{-1} f(t^{-1}x)/f(x))|\right) \theta^{-1} f(\theta^{-1}x)\, d\mu(x) < \infty.$$

(iv) For each $\lambda > 0$ *there exists* $m \geq 1$ *such that*

$$\int_{\mathbb{R}^d} \exp\left(\lambda \sup_{t\ge m} \ln(t^{-1} f(t^{-1}x)/f(x))\right) f(x)\, d\mu(x) < \infty.$$

(v) For each $\lambda > 0$ *there exists* $\tau > 0$ *such that*

$$\int_{\mathbb{R}^d} \exp\left(\lambda \sup_{0<t\le\tau} \ln(t^{-1} f(t^{-1}x)/f(x))\right) f(x)\, d\mu(x) < \infty.$$

Then, (3.1) holds for each $\theta_0, \theta_1 > 0$ *with* $\theta_0 \neq \theta_1$.

Common families of pdf's are exponential families. We refer to Brown (1986) for a review on exponential families. Let μ be a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, where $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ–field in $\mathbb{R}^d$. 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). We present the following theorem for exponential families

Theorem 3.2. *With the notation above, let* $\theta_0 \in \Theta^o$ *and let* $\theta_1 \in \Theta$. *Suppose that:*

(i) For each $a \in \mathbb{R}$, $\mu(\mathbb{R}^d - \{a\}) > 0$.

(ii) $K(\theta_1, \theta_0) < \infty$.

(iii)

$$\lim_{\tau\to 0+} \inf\{I_{\theta_0}(a) : a \in \overline{\{b \in \mathbb{R}^d : I_{\theta_0}(b) \geq K(\theta_1, \theta_0) - \tau\}}\ \} \geq K(\theta_1, \theta_0),$$

where

$$I_{\theta_0}(a) = \sup_{\theta \in \Theta}((\theta - \theta_0)'a - \psi(\theta) + \psi(\theta_0)).$$

Then,

$$\lim_{n \to \infty} n^{-1} \ln H_n(T_n) = -K(\theta_1, \theta_0) \text{ a.s.} \tag{3.2}$$

when θ_1 obtains, where

$$T_n := \sup_{\theta \in \Theta} n^{-1} \sum_{j=1}^{n} (\ln(f(X_j, \theta)/f(X_j, \theta_0)),$$

and $H_n(t) := \mathbb{P}_{\theta_0}(T_n \geq t)$.

Condition (iii) in the previous theorem holds under minor assumptions. Since $I_{\theta_0}(\cdot)$ is a convex function, if $I_{\theta_0}(\theta) < \infty$, for each $\theta \in \mathbb{R}^d$, then $I_{\theta_0}(\cdot)$ is continuous in $\mathbb{R}^d$ and condition (iii) in Theorem 3.2 holds.

Suppose that there exists a sequence $\{K_m\}_{m=1}^{\infty}$ of compacts sets contained in $\{\theta \in \mathbb{R}^d : I_{\theta_0}(\theta) < \infty\}$ and

$$\lim_{m \to \infty} \inf_{\theta \notin K_m} I_{\theta_0}(\theta) = \infty,$$

then condition (iii) in Theorem 3.2 holds. Notice that since $I_{\theta_0}(\cdot)$ is a convex function, $I_{\theta_0}(\cdot)$ continuous in K_m. Hence, if $m > K(\theta_1, \theta_0)$, then

$$\begin{aligned} & \overline{\{b \in \mathbb{R}^d : I_{\theta_0}(b) \geq K(\theta_1, \theta_0) - \tau\}} \\ \subset \quad & \{b \in \mathbb{R}^d : I_{\theta_0}(b) \geq K(\theta_1, \theta_0) - \tau\} \cup K_m^c \end{aligned}$$

and

$$\begin{aligned} & \inf\{I_{\theta_0}(a) : a \in \overline{\{b \in \mathbb{R}^d : I_{\theta_0}(b) \geq K(\theta_1, \theta_0) - \tau\}}\ \} \\ \geq \quad & \min\left(\inf\{I_{\theta_0}(a) : a \in \{b \in \mathbb{R}^d : I_{\theta_0}(b) \geq K(\theta_1, \theta_0) - \tau\}, \inf\{I_{\theta_0}(a) : a \in K_m^c\}\right) \\ \geq \quad & K(\theta_1, \theta_0) - \tau. \end{aligned}$$

The following theorem deals with the case of a composite null hypothesis.

Theorem 3.3. *With the notation above, let $\Theta_0 \subset \Theta$ and let $\theta_1 \in \Theta - \Theta_0$. For each $\theta_0 \in \Theta_0$, let $d_{\theta_0} : \Theta \to (0, \infty)$ be a measurable function. Suppose that:*

(i) $\{x \in S : f(x, \theta) > 0\}$ does not depend on θ.

(ii) For each $\tau > 0$ and each $m \geq 1$, there exists a $\eta > 0$ such that

$$\lim_{\eta \to 0} \sup_{\theta_0 \in \Theta_0} \sup_{\lambda > 0} \left(\lambda \tau - E_{\theta_0}[\exp(\lambda G_{\theta_0, m}^{(\eta)}(X))]\right) = \infty,$$

where
$$G^{(\eta)}_{\theta_0,m}(x) := \sup_{\substack{s\in\Theta_m \\ |s-t|\le\eta}} |\ln(f(x,s)/f(x,t))|$$
and
$$\Theta_{\theta_0,m} := \{t\in\Theta : d(t,\Theta^c)\ge m^{-1} \text{ and } |t-\theta_0|\le m\}.$$

(iii)
$$\lim_{m\to\infty}\inf_{\lambda>0}\sup_{\theta_0\in\Theta_0} E_{\theta_0}\left[\exp\left(\lambda\sup_{\theta\notin\Theta_m}(d_{\theta_0}(\theta))^{-1}(\ln(f(X,\theta)/f(X,\theta_0))\right)\right]=0.$$

(iv) For each $\theta_0\in\Theta_0$,
$$E_{\theta_1}\left[|\ln(f(X,\theta_0)/f(X,\theta_1))|\right]<\infty.$$

(v) For each $\epsilon>0$ *there exists a compact set* K *of* $\mathbb{R}^d$ *such that* $K\subset \mathrm{Int}(\Theta_0)$ *such that* $\inf_{\theta\in\Theta_0-K} b(\theta)>0$ *and*
$$\begin{aligned} & E_{\theta_1}[\sup_{\theta\in\Theta_0-K}(b(\theta))^{-1}\ln(f(X,\theta)/f(X,\theta_1))] \\ \le\ & -(\inf_{\theta\in\Theta_0-K} b(\theta))^{-1}\left(\sup_{\theta\in\Theta_0}E_{\theta_1}[\ln(f(X,\theta)/f(X,\theta_1))]-\epsilon\right). \end{aligned}$$

(vi) For each $t\in\mathrm{Int}(\Theta_0)$,
$$\lim_{\epsilon\to 0}E_{\theta_1}[\sup_{\substack{s\in\Theta \\ |s-t|\le\epsilon}}|\ln(f(X,s)/f(X,t))|]=0.$$

Then,
$$\lim_{n\to\infty} n^{-1}\ln H_n(T_n) = -\inf_{\theta_0\in\Theta_0} K(f(\cdot,\theta_1),f(\cdot,\theta_0)) \text{ a.s.} \tag{3.3}$$
when θ_1 *obtains, where*
$$T_n := \sup_{\theta\in\Theta} n^{-1}\sum_{j=1}^n \ln(f(X_j,\theta)) - \sup_{\theta_0\in\Theta_0} n^{-1}\sum_{j=1}^n \ln(f(X_j,\theta_0)),$$
and $H_n(t) := \sup_{\theta_0\in\Theta_0}\mathbb{P}_{\theta_0}(T_n\ge t)$.

For a location family the previous theorem gives the following:

Corollary 3.5. *Let* μ *be a measure in* $\mathbb{R}^d$. *Let* f *be a nonnegative continuous function in* $\mathbb{R}^d$ *with* $\int_{\mathbb{R}^d} f(x)d\,\mu(x)=1$ *and* $\lim_{|x|\to\infty} f(x)=0$. *Consider the location family of pdf's* $\{f(x-\theta):\theta\in\Theta\}$, *where* $\Theta:=\mathbb{R}^d$. *Suppose that:*

(i) $K(f(\cdot-\theta_1),f(\cdot-\theta_0)<\infty$.

(ii) For each $\lambda > 0$ and each $\theta \in \mathbb{R}^d$, there exists $\eta > 0$ such that

$$\int_{\mathbb{R}^d} \exp\left(\lambda \sup_{|t|\le\eta} |\ln(f(x-t)/f(x))|\right) f(x-\theta)\, d\mu(x) < \infty.$$

(iii) For each $\lambda > 0$ there exists $m \ge 1$ such that

$$\int_{\mathbb{R}^d} \exp\left(\lambda \sup_{|t|\ge m} \ln(f(x-t)/f(x))\right) f(x)\, d\mu(x) < \infty.$$

Then, (3.2) holds for each $\Theta_0 \subset \Theta$ and each $\theta_1 \notin \Theta_0$.

4 Proofs

Proof of Theorem 3.1. By (1.2), it suffices to prove that

$$\limsup_{n\to\infty} n^{-1} \ln H_n(T_n) \le -K(\theta_1, \theta_0) \text{ a.s.} \tag{4.1}$$

when θ_1 obtains. Since

$$\liminf_{n\to\infty} T_n \ge \liminf_{n\to\infty} n^{-1} \sum_{j=1}^n \ln(f(X_j, \theta_1)/f(X_j, \theta_0)) = K(\theta_1, \theta_0) \text{ a.s.}$$

when θ_1 obtains, to prove (4.1), it suffices to obtain that

$$\lim_{\tau\to 0+} \limsup_{n\to\infty} n^{-1} \ln H_n(K(\theta_1, \theta_0) - \tau) \le -K(\theta_1, \theta_0). \tag{4.2}$$

Let $U_n(\theta) := n^{-1} \sum_{j=1}^n \ln f(X_j, \theta)$. We have that

$$\begin{aligned} & \mathbb{P}_{\theta_0}\{\sup_{\theta\in\Theta} U_n(\theta) - U_n(\theta_0) \ge K(\theta_1, \theta_0) - \tau\} \\ \le\ & \mathbb{P}_{\theta_0}\{\sup_{\theta\in\Theta_m} U_n(\theta) - U_n(\theta_0) \ge K(\theta_1, \theta_0) - \tau\} + \mathbb{P}_{\theta_0}\{\sup_{\theta\in\Theta} U_n(\theta) - \sup_{\theta\in\Theta_m} U_n(\theta) > 0\} \\ \le\ & \mathbb{P}_{\theta_0}\{\sup_{\theta\in\Theta_m} U_n(\theta) - U_n(\theta_0) \ge K(\theta_1, \theta_0) - \tau\} + \mathbb{P}_{\theta_0}\{\sup_{\theta\notin\Theta_m} U_n(\theta) - U_n(\theta_0) > 0\} \\ =:\ & I + II \end{aligned} \tag{4.3}$$

Since $\{U_n(\theta) - U_n(\theta_0) : \theta \in \Theta_m\}$ satisfies the LDP with speed n,

$$\begin{aligned} & \limsup_{n\to\infty} n^{-1} \ln \mathbb{P}_{\theta_0}\{\sup_{\theta\in\Theta_m} U_n(\theta) - U_n(\theta_0) \ge K(\theta_1, \theta_0) - \tau\} \\ \le\ & -\inf\{J_{\theta_0}(l) : l \in (\mathcal{L}_{\theta_0}^{\Phi_1})^*, \sup_{\theta\in\Theta_m} l(\ln(f(\cdot, \theta)/f(\cdot, \theta_0)) \ge K(\theta_1, \theta_0) - \tau\} \end{aligned} \tag{4.4}$$

We also have that

$$
\begin{array}{ll}
& n^{-1}\ln\left(\mathbb{P}_{\theta_0}\{\sup_{\theta\notin\Theta_m} n^{-1}\sum_{j=1}^{n}(\ln f(X_j,\theta)-\ln f(X_j,\theta_0))>0\}\right) \qquad (4.5)\\
\leq & n^{-1}\ln\left(\mathbb{P}_{\theta_0}\{n^{-1}\sum_{j=1}^{n}\sup_{\theta\notin\Theta_m}(d(\theta))^{-1}(\ln f(X_j,\theta)-\ln f(X_j,\theta_0))>0\}\right)\\
\leq & \inf_{\lambda>0} n^{-1}\ln\left(E_{\theta_0}\left[\exp\left(\lambda\sum_{j=1}^{n}\sup_{\theta\notin\Theta_m}(d(\theta))^{-1}(\ln f(X_j,\theta)-\ln f(X_j,\theta_0))\right)\right]\right)\\
= & \inf_{\lambda>0}\ln\left(E_{\theta_0}\left[\exp\left(\lambda\sup_{\theta\notin\Theta_m}(d(\theta))^{-1}(\ln f(X,\theta)-\ln f(X,\theta_0))\right)\right]\right).
\end{array}
$$

Thus,

$$
\lim_{m\to\infty}\limsup_{n\to\infty} n^{-1}\ln II=-\infty. \tag{4.6}
$$

Combining (4.3)–(4.6), we obtain that

$$
\begin{array}{ll}
& \lim_{\tau\to 0+}\limsup_{n\to\infty} n^{-1}\ln\left(\mathbb{P}_{\theta_0}\{\sup_{\theta\in\Theta}U_n(\theta)-U_n(\theta_0)\geq K(\theta_1,\theta_0)-\tau\}\right)\\
\leq & \limsup_{\tau\to 0+}\limsup_{m\to\infty} -\inf\{J_{\theta_0}(l):\sup_{\theta\in\Theta_m} l(\ln(f(\cdot,\theta)/f(\cdot,\theta_0)))\geq K(\theta_1,\theta_0)-\tau\}.
\end{array}
$$

By the definition of J_{θ_0}, for each $\theta\in\Theta$ and each $l\in(\mathcal{L}_{\theta_0}^{\Phi_1})^*$,

$$
l(\ln(f(\cdot,\theta)/f(\cdot,\theta_0)))\leq J_{\theta_0}(l)+\ln E_{\theta_0}[\exp(\ln(f(X,\theta)/f(X,\theta_0)))]=J_{\theta_0}(l). \tag{4.7}
$$

Hence, if

$$
\sup_{\theta\in\Theta_m} l(\ln(f(\cdot,\theta)/f(\cdot,\theta_0)))\geq K(\theta_1,\theta_0)-\tau,
$$

then $K(\theta_1,\theta_0)-\tau\leq J_{\theta_0}(l)$. Therefore,

$$
K(\theta_1,\theta_0)-\tau\leq\inf\{J_{\theta_0}(l):\sup_{\theta\in\Theta_m} l(\ln(f(\cdot,\theta)/f(\cdot,\theta_0))\geq K(\theta_1,\theta_0)-\tau\} \tag{4.8}
$$

and

$$
\begin{array}{l}
\limsup_{\tau\to 0+}\limsup_{m\to\infty} -\inf\{J_{\theta_0}(l):\sup_{\theta\in\Theta_m} l(\ln(f(\cdot,\theta)/f(\cdot,\theta_0))\geq K(\theta_1,\theta_0)-\tau\}\\
\leq -K(\theta_1,\theta_0),
\end{array}
$$

which implies (4.2). $\square$

Proof of Corollary 3.1. We apply Theorem 3.1. Conditions (i), (ii) and (iv) in Theorem 3.1 are assumed. To get condition (iii) in Theorem 3.1 we apply Theorem 2.1. By the Cauchy–Schwartz inequality, for each $\theta\in\Theta$,

$$
\begin{array}{l}
E_{\theta_0}[\exp(2^{-1}\ln(f(X,\theta)/f(X,\theta_0))]=\int(f(x,\theta)f(x,\theta_0))^{1/2}\,\mu(x)\\
\leq(\int f(x,\theta)\,\mu(x))^{1/2}(\int(f(x,\theta)\,\mu(x))^{1/2}<\infty.
\end{array}
$$

Hence, $\ln(f(X,\theta)/f(X,\theta_0))\in\mathcal{L}_{\theta_0}^{\Phi}$, i.e. condition (i) in Theorem 2.1 holds. Conditions (iii) and (iv) imply conditions (ii) and (iii) in Theorem 2.1. $\square$

The proof of Corollaries 3.2 and 3.4 are omitted. Since they follow directy from Corollary 3.1.

Proof of Corollary 3.3. We apply Corollary 3.2. It is obvious that $K(\theta_1, \theta_0), \infty$. First, we consider the case $p \geq 1$. By the Taylor theorem, there exists a constant $c > 0$ such that for each $x, \theta \in \mathbb{R}^d$,

$$||x-\theta|^p - |x|^p| \leq c(|x|^{p-1}|\theta| + |\theta|^p).$$

This implies that conditions (ii) and (iii) in Corollary 3.2 hold with $d(\theta) = 1 + |\theta|^{p+1}$.

When $0 < p < 1$, there exists a constant $c > 0$ such that for each $x, \theta \in \mathbb{R}^d$,

$$||x-\theta|^p - |x|^p| \leq c(|x|^{p-1}|\theta| \wedge |\theta|^p).$$

Conditions (ii) and (iii) in Corollary 3.2 hold with $d(\theta) = 1 + |\theta|^{p+1}$. □

Proof of Theorem 3.2. As in the proof of Theorem 3.1, it suffices to prove (4.2). We have that

$$T_n = \sup_{\theta \in \Theta}(\theta' \bar{X}_n - \psi(\theta) - \theta_0' \bar{X}_n + \psi(\theta_0)) = I_{\theta_0}(\bar{X}_n)$$

Since $\theta_0 \in \Theta^o$, there exists a $\delta > 0$ such that $E_{\theta_0}[\exp(\lambda' X)] < \infty$, for each $|\lambda| \leq \delta$. This implies that $\bar{X}_n$ satisfies the LDP with speed and rate function

$$\begin{aligned} I(t) &= \sup_{\lambda \in \mathbb{R}^d}(\lambda' t - \ln E_{\theta_0}[\exp(\lambda' X]] \\ &= \sup_{\lambda \in \mathbb{R}^d}(\lambda' t - \psi(\lambda + \theta_0) - \psi(\theta_0)) \\ &= I_{\theta_0}(t). \end{aligned}$$

Hence,

$$\begin{aligned} & \limsup_{n\to\infty} n^{-1} \ln H_n(K(\theta_1, \theta_0) - \tau) \\ = \; & \limsup_{n\to\infty} n^{-1} \ln \mathbb{P}_{\theta_0}\{I_{\theta_0}(\bar{X}_n) \geq K(\theta_1, \theta_0) - \tau) \\ \leq \; & -\inf\{I_{\theta_0}(a) : a \in \overline{\{b \in \mathbb{R}^d : I_{\theta_0}(b) \geq K(\theta_1, \theta_0) - \tau\}}\ \} \end{aligned}$$

and the claim follows. □

We will need the following lemma:

Lemma 4.1. *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 Θ be a Borel subset of $\mathbb{R}^d$. Let $g : S \times \Theta \to \mathbb{R}$ be a measurable function. Let $b : S \to (0, \infty)$ be a measurable function. Suppose that:*

(i) For each $\theta \in \Theta$, $E[|g(X, \theta)|] < \infty$.

(ii) $\sup_{\theta \in \Theta} E[g(X, \theta)] \leq 0$.

(iii) For each $\epsilon > 0$, there exists a compact set K such that $K \subset \mathrm{Int}(\Theta)$, $\inf_{\theta \notin K} b(\theta) > 0$ and

$$E[\sup_{\theta \notin K}(b(\theta))^{-1} g(X, \theta)] \leq (\inf_{\theta \notin K} b(\theta))^{-1} \left(\sup_{\theta \in \Theta} E[g(X, \theta)] - \epsilon \right).$$

(iv) For each $\epsilon > 0$ and each $\theta_1 \in \mathrm{Int}(\Theta)$, there exists a $\delta > 0$ such that

$$E[\sup_{\theta:|\theta-\theta_1|\le\delta} |g(X,\theta_1) - g(X,\theta)|] \le \epsilon.$$

Then,

$$\sup_{\theta\in\Theta} n^{-1} \sum_{j=1}^{n} g(X_j,\theta) \stackrel{\text{a.s.}}{\to} \sup_{\theta\in\Theta} E[g(X,\theta)].$$

Proof. Given $\epsilon > 0$, there exists a compact set K satisfying condition (iii). Then, for each $\theta \notin K$,

$$E[(b(\theta))^{-1} g(X,\theta)] \le (\inf_{\theta\notin K} b(\theta))^{-1} \left(\sup_{\theta\in\Theta} E[g(X,\theta)] - \epsilon \right) \le (b(\theta))^{-1} \left(\sup_{\theta\in\Theta} E[g(X,\theta)] - \epsilon \right).$$

Hence, for each $\theta \notin K$,

$$E[g(X,\theta)] \le \sup_{\theta\in\Theta} E[g(X,\theta)] - \epsilon.$$

So,

$$\sup_{\theta\notin K} E[g(X,\theta)] \le \sup_{\theta\in\Theta} E[g(X,\theta)] - \epsilon. \tag{4.9}$$

By the strong law of the large numbers, there exists a set Ω_0 with probability one such that for each $\omega \in \Omega_0$,

$$n^{-1} \sum_{j=1}^{n} \sup_{\theta\notin K} (b(\theta))^{-1} g(X_j(\omega),\theta) \to E[\sup_{\theta\notin K} (b(\theta))^{-1} g(X,\theta)].$$

Given $\eta > 0$ such that $\eta \inf_{\theta\in K} b(\theta) \le 2^{-1}\epsilon$, and $\omega \in \Omega_0$, for n large enough,

$$\begin{aligned} & \sup_{\theta\notin K} n^{-1} \textstyle\sum_{j=1}^{n} (b(\theta))^{-1} g(X_j(\omega),\theta) \le n^{-1} \sum_{j=1}^{n} \sup_{\theta\notin K} (b(\theta))^{-1} g(X_j(\omega),\theta) \\ \le\ & E[\sup_{\theta\notin K} (b(\theta))^{-1} g(X,\theta)] + \eta \le (\inf_{\theta\notin K} b(\theta))^{-1} (\sup_{\theta\in\Theta} E[g(X,\theta)] - \epsilon) + \eta. \end{aligned}$$

Hence, for each $\omega \in \Omega_0$, each $\theta \notin K$ and each n large enough,

$$\begin{aligned} & n^{-1} \textstyle\sum_{j=1}^{n} g(X_j(\omega),\theta) \le n^{-1} \sum_{j=1}^{n} (b(\theta))^{-1} g(X_j(\omega),\theta) b(\theta) \\ \le\ & ((\inf_{\theta\notin K} b(\theta))^{-1} (\sup_{\theta\in\Theta} E[g(X,\theta)] - \epsilon) + \eta)\, b(\theta) \\ \le\ & ((\inf_{\theta\notin K} b(\theta))^{-1} (\sup_{\theta\in\Theta} E[g(X,\theta)] - \epsilon) + \eta) \inf_{\theta\notin K} b(\theta) \\ =\ & \sup_{\theta\in\Theta} E[g(X,\theta)] - \epsilon + \eta \inf_{\theta\notin K} b(\theta) \\ \le\ & \sup_{\theta\in\Theta} E[g(X,\theta)] - 2^{-1}\epsilon, \end{aligned}$$

where we have used that $b(\theta) > 0$ and $(\inf_{\theta\notin K} b(\theta))^{-1} (\sup_{\theta\in\Theta} E[g(X,\theta)] - \epsilon) + \eta < 0$. Hence, for each $\omega \in \Omega_0$ and each n large enough,

$$\sup_{\theta\notin K} n^{-1} \sum_{j=1}^{n} g(X_j(\omega),\theta) \le \sup_{\theta\in\Theta} E[g(X,\theta)] - 2^{-1}\epsilon. \tag{4.10}$$

By condition (iv), the class of functions $\{g(X,\theta) : \theta \in K\}$ is a Glivenko–Cantelli class (see either Lemma 2 in Dehardt, 1971; or Section 7 in Dudley, 1999). So,

$$\sup_{\theta \in K} |n^{-1} \sum_{j=1}^{n} g(X_j,\theta) - E[g(X,\theta)]| \to 0 \text{ a.s.}$$

Therefore,

$$\sup_{\theta \in K} n^{-1} \sum_{j=1}^{n} g(X_j,\theta) - \sup_{\theta \in K} E[g(X,\theta)] \to 0 \text{ a.s.} \tag{4.11}$$

The claim follows from (4.9)–(4.11). □

Proof of Theorem 3.3. By (1.2), it suffices to prove that

$$\sup_{\theta_0 \in \Theta_0} n^{-1} \sum_{j=1}^{n} \ln(f(X_j,\theta_0))/f(X_j,\theta_1)) \xrightarrow{P_{\theta_1}} \sup_{\theta \in \Theta_0} E_{\theta_1}[\ln(f(X,\theta)/f(X,\theta_1))] \text{ a.s.} \tag{4.12}$$

$$\lim_{\tau \to 0+} \limsup_{n \to \infty} n^{-1} \ln\left(\sup_{\theta_0 \in \Theta_0} \mathbb{P}_{\theta_0}\{T_n \geq L_1 - \tau\}\right) \leq -L_1 \tag{4.13}$$

where

$$L_1 := \inf_{\theta_0 \in \Theta_0} K(\theta_1,\theta_0).$$

(4.12) follows from Lemma 4.1 with $g(x,\theta) = \ln(f(x,\theta)/f(x,\theta_1))$. Hypothesis (i), (iii)-(iv) in Lemma 4.1 follow from (iv)–(vi). By the concavity of the logarithmic function and the Jensen inequality, for each $\theta \in \Theta$

$$E_{\theta_1}[\ln(f(X,\theta)/f(X,\theta_1))] \leq 0, \tag{4.14}$$

which implies hypothesis (ii) in Lemma 4.1.

As to (4.13), we may assume that $L_1 > 0$. Take $0 < \tau < 2^{-1}L_1$. For each $\theta_0 \in \Theta_0$, we have that

$$\begin{aligned} & \mathbb{P}_{\theta_0}\{T_n \geq L_1 - \tau\} \qquad (4.15)\\ \leq\; & \mathbb{P}_{\theta_0}\{\sup_{\theta \in \Theta}(U_n(\theta) - U_n(\theta_0)) \geq L_1 - \tau\} \\ \leq\; & \mathbb{P}_{\theta_0}\{\sup_{\theta \in \Theta_{\theta_0,m}}(U_n(\theta) - U_n(\theta_0)) \geq L_1 - \tau\} + \mathbb{P}_{\theta_0}\{\sup_{\theta \in \Theta} U_n(\theta) > \sup_{\theta \in \Theta_{\theta_0,m}} U_n(\theta)\} \\ \leq\; & \mathbb{P}_{\theta_0}\{\sup_{\theta \in \Theta_{\theta_0,m}}(U_n(\theta) - U_n(\theta_0)) \geq L_1 - \tau\} + \mathbb{P}_{\theta_0}\{\sup_{\theta \in \Theta_0 - \Theta_{\theta_0,m}}(U_n(\theta) - U_n(\theta_0)) > 0\}. \end{aligned}$$

Given $\eta > 0$, since $\Theta_{\theta_0,m}$ is a compact set, there are $t_{\theta_0,1},\ldots,t_{\theta_0,k_{\theta_0}} \subset \Theta_{\theta_0,m}$ such that the union of the balls with centers $t_{\theta_0,1},\ldots,t_{\theta_0,k_{\theta_0}}$ and radius η cover $\Theta_{\theta_0,m}$. Although, $t_{\theta_0,1},\ldots,t_{\theta_0,k_{\theta_0}}$ depend on m and η, we do not use the subscripts m and η to simplify notation. Since

$\Theta_{\theta_0,m} \subset \{\theta : |\theta - \theta_0| \le m\}$, k_{θ_0} is bounded uniformly on θ_0. Let $\bar{k} = \sup\{k_{\theta_0} : \theta_0 \in \Theta_0\}$. Note that $\bar{k}$ depends on m and η. We have that

$$\begin{aligned} & \mathbb{P}_{\theta_0}\{\sup_{\theta\in\Theta_{\theta_0,m}}(U_n(\theta) - U_n(\theta_0)) \ge L_1 - \tau\} \qquad (4.16) \\ \le\ & \mathbb{P}_{\theta_0}\{\max_{1\le j\le k_{\theta_0}}(U_n(t_{\theta_0,j}) - U_n(\theta_0)) \ge L_1 - 2\tau\} + \mathbb{P}_{\theta_0}\{n^{-1}\sum_{j=1}^n G_m^{(\eta)}(X_j) \ge \tau\} \\ \le\ & \bar{k}\max_{1\le j\le k_{\theta_0}} \mathbb{P}_{\theta_0}\{U_n(t_{\theta_0,j}) - U_n(\theta_0) \ge L_1 - 2\tau\} + \mathbb{P}_{\theta_0}\{n^{-1}\sum_{j=1}^n G_m^{(\eta)}(X_j) \ge \tau\}. \end{aligned}$$

For each j and each θ_0,

$$\begin{aligned} & \mathbb{P}_{\theta_0}\{U_n(t_{\theta_0,j}) - U_n(\theta_0) \ge L_1 - 2\tau\} \\ & \le e^{-n(L_1-2\tau)} E_{\theta_0}[\exp\left(\sum_{j=1}^n \ln(f(X_j, t_{\theta_0,j})/f(X_j,\theta_0))\right)] = e^{-n(L_1-2\tau)}. \end{aligned}$$

So, for each j and each θ_0,

$$\bar{k}\max_{1\le j\le k_{\theta_0}} \mathbb{P}_{\theta_0}\{U_n(t_{\theta_0,j}) - U_n(\theta_0) \ge L_1 - 2\tau\} \le \bar{k}e^{-n(L_1-2\tau)}. \qquad (4.17)$$

We also have that

$$\begin{aligned} & n^{-1}\ln \mathbb{P}_{\theta_0}\{n^{-1}\sum_{j=1}^n G_m^{(\eta)}(X_j) \ge \tau\} \qquad (4.18) \\ \le\ & \sup_{\lambda>0}\left(\lambda\tau - \ln E_{\theta_0}[\exp(\lambda G_m^{(\eta)}(X))]\right) \end{aligned}$$

and

$$\begin{aligned} & n^{-1}\ln(\mathbb{P}_{\theta_0}\{\sup_{\theta\in\Theta_0-\Theta_{\theta_0,m}}(U_n(\theta) - U_n(\theta_0)) > 0\}) \qquad (4.19) \\ \le\ & \inf_{\lambda>0} E_{\theta_0}\left[\exp\left(\lambda \sup_{\theta\in\theta\in\Theta_0-\Theta_{\theta_0,m}}(d_{\theta_0}(\theta))^{-1}(\ln f(X,\theta)/f(X,\theta_0))\right)\right] \end{aligned}$$

We get from (4.15)–(4.19) that for each $\tau > 0$,

$$\limsup_{n\to\infty} n^{-1}\ln \sup_{\theta_0\in\Theta_0}(\mathbb{P}_{\theta_0}\{T_n \ge L_1 - \tau\}) \le -(L_1 - 2\tau).$$

Therefore, (4.13) holds. $\square$

References

[1] M. A. Arcones, *The large deviation principle for empirical processes.* In "High Dimensional Probability III" (2003a), Edited by J. Hoffmann–Jorgensen, M. B. Marcus and J. A. Wellner. Birkhauser, Boston, pp. 205–223.

[2] M. A. Arcones, *Large deviations for M-estimators.* To appear in Ann. Inst. Statist. Mathem., 2003b.

[3] R. R. Bahadur, *An optimal property of the likelihood ratio statistic*, Proc. Fifth Berkeley Symp. Math. Statist. Prob. 1 (1965), 13–26.

[4] R. R. Bahadur, *Rates of convergence of estimates and test statistics.* Annals Mathem. Statist., 38 (1967), pp. 303–324.

[5] R. R. Bahadur, *Some Limit Theorems in Statistics.* SIAM, Philadelphia, PA, 1971.

[6] R. R. Bahadur and M. Raghavachari, Some asymptotic properties of likelihood ratios on general sample spaces. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability* (Univ. California, Berkeley, Calif., 1970/1971), Vol. I, Univ. California Press, Berkeley, California, 1972, pp. 129–152.

[7] L. D. Brown, *Non–local asymptotic optimality of appropriate likelihood ratio tests.* Ann. Math. Stat., 42 (1971), 1206–1240

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

[9] J. Dehardt, *Generalizations of the Glivenko–Cantelli theorem*, Annals of Mathematical Statistics 42 (1971), pp. 2050–2055

[10] A. Dembo, and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer, New York, 1998.

[11] J. D. Deuschel, and D. W. Stroock, *Large Deviations*, Academic Press, Inc. Boston, MA, 1989.

[12] R. M.Dudley, *Uniform Central Limit Theorems*, Cambridge University Press, Cambridge, UK, 1999.

[13] B. Efron and D. Truax, *Large deviation theory in exponential families*, Ann. Mathem. Statist. 39 (1968), pp. 1403–1424.

[14] D. G. Herr, *Asymptotical optimal tests for multivariate normal distribution.* Ann. Mathem. Statist. 38 (1967), pp. 1829–1844.

[15] H. K. Hsieh, *On asymptotic optimality of likelihood ratio tests for multivariate normal distribution*, Ann. Stat. 7 (1979a), pp. 592–598.

[16] H. K. Hsieh, *Exact Bahadur efficiencies for tests of the multivariate linear hypothesis*, Ann. Stat. 7 (1979b), pp. 1231–1245.

[17] W. C. M. Kallenberg, *Asymptotic Optimal Tests of Likelihood Ratio Tests in Exponential Families*. Mathematical Centre Tracts, Amsterdam, 1978.

[18] E.I. Kolomiec, *On asymptotic optimality of the likelihood ratio test*, Proceedings of the Fourth Prague Symposium on Asymptotic Statistics (Prague, 1988), (1989), pp. 337–348.

[19] S. Kourouklis, *A large deviation results for the likelihood ratio statistic in exponential families*, Ann. Statist. 12 (1984), pp. 1510–1521.

[20] Y. Nikitin, *Asymptotic Efficiency of Nonparametric Tests*, Cambridge University Press, Cambridge, UK, 1995.

[21] J. Oosterhoff and W. R. van Zwet, *The likelihood ratio test for the multinomial distribution*, *Proc. Sixth Berkeley Symp. Math. Stat. Prob* **1** (1970), pp.31–50.

[22] M. Raghavachari, *On a theorem of Bahadur on the rate of convergence of test statistics*, Ann. Mathem. Statist. 41 (1970), pp. 1695–1699.

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

[24] F. Rublik, *On optimality of the LR tests in the sense of exact slopes. I. General case*, Kybernetika, 25 (1989a), pp. 13–25.

[25] F. Rublik, *On optimality of the LR tests in the sense of exact slopes. II Application to individual distributions*, Kybernetika 25 (1989b), pp. 117–135.

[26] F. Rublik, *On d–optimality of the LR tests*, Kybernetika 31 (1995), pp. 189-206.

[27] F. Rublik, *A large deviation theorem for the q–sample likelihood ratio statistic*, Ann. Stat., 24 (1996), pp. 2280–2287

[28] R. J. Serfling, *Approximation Theorems of Mathematical Statistics.* Wiley, New York, 1980.