Moderate deviations for M–estimators

Miguel A. Arcones *
Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902

April 11, 2002

Abstract

We present general sufficient conditions for the moderate deviations of M–estimators. These results are applied to many different types of M-estimators such as the p–th quantile, the spatial median, the least absolute deviation estimator in linear regression, maximum likelihood estimators and other location estimators. We apply moderate deviations theorems from empirical processes.

Key Words: M–estimators, moderate deviations, maximum likelihood estimators.

AMS subject classification: 60F10, 62E20.

1 Introduction

We discuss the moderate deviations for M–estimators. Huber (1964) introduced M–estimators as a way to obtain more robust estimators. Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in a measurable space $(S, \mathcal{S})$. Let $g : S \times \Theta \to I\!R$ be a function such that $g(\cdot, \theta) : S \to I\!R$ is measurable for each $\theta \in \Theta$, where Θ be a Borel subset of $I\!R^d$. Suppose that we want to estimate a parameter $\theta_0 \in \Theta$ characterized by $E[g(X, \theta) - g(X, \theta_0)] > 0$ for each $\theta \neq \theta_0$. An M–estimator $\hat{\theta}_n$ over a kernel $g(x, \theta)$ is a random variable $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ satisfying

$$n^{-1} \sum_{i=1}^{n} g(X_i, \hat{\theta}_n) \simeq \inf_{\theta \in \Theta} n^{-1} \sum_{i=1}^{n} g(X_i, \theta). \tag{1.1}$$

We also consider M–estimators $\hat{\theta}_n$ defined by

$$n^{-1} \sum_{i=1}^{n} h(X_i, \hat{\theta}_n) \simeq 0, \tag{1.2}$$

Email: arcones@math.binghamton.edu.
Web: http://www.math.binghamton.edu/arcones/index.html.

where $h(\cdot,\theta) : S \to I\!R^d$ is a measurable function for each $\theta \in \Theta$. Here, $\hat{\theta}_n$ is estimating a value θ_0 characterized by $E[h(X,\theta_0)] = 0$.

It is well known that under regularity conditions, these estimators are asymptotically normal. In fact, there exists a function ψ such that,

$$n^{1/2}(\hat{\theta}_n - \theta_0) + n^{-1/2}\sum_{i=1}^{n}(\psi(X_i) - E[\psi(X_i)]) \stackrel{\Pr}{\to} 0,$$

(see Huber, 1964, 1981; Hampel, Ronchetti, Rousseeuw and Stahel, 1986; Serfling, 1980; and Lehmann and Casella, 1998). The function ψ is called the influence curve. The M–estimator is more robust when the influence curve is bounded (Hampel, 1974).

Given a sequence of r.v.'s $\{Y_n\}$ with values a metric space (T,d), a sequence of positive numbers $\{\epsilon_n\}$ which converges to zero and a function $I : T \to [0,\infty)$, it is said that $\{Y_n\}$ satisfies the (LDP) large deviation principle with speed ϵ_n and rate function I if for each Borel set A of T,

$$\begin{aligned} & -\inf\{I(v) : v \in A^o\} \le \liminf_{n\to\infty} \epsilon_n \log(\Pr\{Y_n \in A\}) \\ \le\ & \limsup_{n\to\infty} \epsilon_n \log(\Pr\{Y_n \in A\}) \le -\inf\{I(v) : v \in \bar{A}\}. \end{aligned}$$

We refer to page 35 in Deuschel and Stroock (1989) for more information in the large deviation principle. When dealing with stochastic processes, we will use the definition of large deviation principle in Arcones (1998b). We will obtain a large deviation principle for M–estimators under different speeds.

We study the moderate deviations of M–estimators. Given a sequence of i.i.d. random vectors $\{Y_i\}$ with values in $I\!R^d$, and a sequence of real numbers $\{a_n\}$ such that $a_n \to \infty$ and $a_n n^{-1/2} \to 0$, by Corollary 3.4 in Arcones (2001), if

$$E[|Y_1|^2] < \infty \text{ and } \lim_{n\to\infty} a_n^{-2}\log(n\Pr\{|Y_1| \ge n^{1/2}a_n\}) = -\infty, \tag{1.3}$$

then $a_n^{-1}n^{-1/2}\sum_{i=1}^{n}(Y_i - E[Y_i])$ satisfies the LDP with speed a_n^2 and the rate function

$$I(t) = 2^{-1}v\Sigma_Y^{-2}v, v \in I\!R^d$$

where $\Sigma_Y^2 := E[(Y - E[Y])(Y - E[Y])']$. The previous result is called moderate deviations, because the considered tail is smaller than the one in the large deviations set–up. In the moderate deviations set–up we are considering the tail of $|\sum_{i=1}^n (Y_i - E[Y_i])|$ in $[n^{1/2}a_n, \infty)$. The usual large deviations considers the tail of $|\sum_{i=1}^n (Y_i - E[Y_i])|$ in $[n, \infty)$. An elementary argument shows that if for some $\lambda > 0$, $E[\exp(\lambda|Y|)] < \infty$, then (1.3) holds. (1.3) is a necessary sufficient condition for the moderate deviations, in the sense that if $a_n \nearrow \infty$, $n^{-1/2}a_n \searrow 0$ and $\{a_n^{-1}n^{-1/2}\sum_{j=1}^n Y_j\}$ satisfies the LDP with speed a_n^2 and a rate $I(t)$ such that $\lim_{|t|\to\infty} t^{-1}I(t) = \infty$, then $E[Y] = 0$, $E[Y^2] < \infty$ and

$$\lim_{n\to\infty} a_n^{-2}\log(n\Pr\{|Y| \geq n^{1/2}a_n\}) = -\infty$$

(see Corollary 3.4 in Arcones, 2001).

We present very general theorems to obtain that for each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2}\log(\Pr\{|n^{1/2}(\hat{\theta}_n - \theta_0) \tag{1.4}$$

$$+n^{-1/2}\sum_{i=1}^n(\psi(X_i) - E[\psi(X_i)])| \geq a_n\tau\}) = -\infty.$$

Assuming (1.4) and that $\psi(X_1)$ satisfies (1.3), we obtain that $a_n^{-1}n^{1/2}(\hat{\theta}_n - \theta_0)$ satisfies the LDP with speed a_n^2 and the rate function

$$I_\psi(t) = 2^{-1}v\Sigma_\psi^{-2}v, v \in I\!R^d$$

where $\Sigma_\psi^2 := E[(\psi(X_1) - E[\psi(X_1)])(\psi(X_1) - E[\psi(X_1)])']$, i.e. for each Borel set $A \subset I\!R^d$,

$$\begin{aligned} &-\inf\{I_\psi(v) : v \in A^o\} \leq \liminf_{n\to\infty}\log(\Pr\{n^{1/2}(\hat{\theta}_n - \theta_0) \in A\}) \\ \leq\ &\limsup_{n\to\infty}\log(\Pr\{n^{1/2}(\hat{\theta}_n - \theta_0) \in A\}) \leq -\inf\{I_\psi(v) : v \in \bar{A}\}. \end{aligned}$$

This implies that for each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2}\log(\Pr\{a_n^{-1}n^{1/2}|\hat{\theta}_n - \theta_0| \geq \tau\} = -2^{-1}\tau^2 e_d^{-2}, \tag{1.5}$$

where e_d^2 is the biggest eigenvalue of the covariance matrix Σ_ψ^2. This result states that the tail of $n^{1/2}(\hat{\theta}_n - \theta_0)$ decreases like the tail of a Gaussian random vector with mean zero and covariance matrix Σ_ψ^2.

We present general theorems to obtain (1.4). We apply these results to several examples such as: p–th quantiles, location parameters, spatial medians, least absolute deviation estimator in linear regression and maximum likelihood estimators.

A main technique in the proofs is the moderate deviation theorems for empirical processes in Arcones (2001). In particular, we obtain the moderate deviations of M–estimators assuming weaker conditions than the existence of a moment generating function in a neighborhood of zero.

The moderate deviations of one dimensional mle's was considered by Gao (2001). By the way, it seems that some monotonicity assumption on $l_n^{(1)}$ in this paper would be necessary in order to ensure that

$$\{\underline{\theta}_n \geq \theta + \epsilon\} \subset \{l_n^{(1)}(\theta + \epsilon) \geq 0\}$$

where

$$\underline{\theta}_n = \inf\{\theta : l_n^{(1)}(\theta) \leq 0\}$$

(see Equations (1) in Gao, 2001).

Jensen and Wood (1997) considered the large deviation for M–estimators. They gave sufficient conditions so that for each $\tau > 0$,

$$\limsup_{n\to\infty} n^{-1} \log(\Pr\{|\hat{\theta}_n - \theta_0| \geq \tau\}) < 0.$$

In Section 2, we present the main results and several examples. The proofs of the theorems in Section 2 are in Section 3.

c will denote an universal constant that may vary from line to line. Let X be a copy of X_1. We will use the usual multivariate notation. For example, given $u = (u_1, \ldots, u_d)' \in \mathbb{R}^d$ and $v = (v_1, \ldots, v_d)' \in \mathbb{R}^d$, we denote $u'v = \sum_{j=1}^d u_j v_j$ and $|u| = (\sum_{j=1}^n u_j^2)^{1/2}$. We will use the notation in empirical processes in Giné and Zinn (1984). Given a function $f : S \to \mathbb{R}$, we define

$$P_n f = n^{-1} \sum_{i=1}^n f(X_i) \quad \text{and} \quad Pf = E[f(X)].$$

2 Moderate deviations for M–estimators

If the stochastic processes are convex, (1.4) can be obtained from minimal conditions:

Theorem 2.1. *With notation for the M–estimators in (1.1), let $\phi : S \to \mathbb{R}^d$ be a measurable function and let $\{a_n\}$ be a sequence of real numbers converging to infinity such that $a_n n^{-1/2} \to 0$. Suppose that:*

(i) For each $x \in I\!\!R^d$, $g(x,\cdot):\Theta \to I\!\!R$ *is a convex function.*

(ii) $\hat\theta_n = \hat\theta_n(X_1,\dots,X_n)$ *is a sequence of* $I\!\!R^d$*–valued random variables such that for each* $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2}\log(\Pr\{a_n^{-2}\sum_{j=1}^n g(X_j,\hat\theta_n) \ge \inf_{\theta\in\Theta} a_n^{-2}\sum_{j=1}^n g(X_j,\theta)+\tau\}) = -\infty.$$

(iii) There exists a positive definite symmetric matrix V *such that*

$$E[g(X,\theta)-g(X,\theta_0)] = (\theta-\theta_0)'V(\theta-\theta_0)+o(|\theta-\theta_0|^2),$$

as $\theta \to \theta_0$.

(iv) $E[|\phi(X)|^2] < \infty$

(v) $\lim_{n\to\infty} a_n^{-2}\log(n\Pr\{|\phi(X)| \ge n^{1/2}a_n\}) = -\infty$

(vi) For each $\theta \in I\!\!R$,

$$na_n^{-2}E[|r(X,n^{-1/2}a_n\theta)|I(|r(X,n^{-1/2}a_n\theta)| \ge a_n^2)] \to 0,$$

where

$$r(x,\theta) = g(x,\theta_0+\theta)-g(x,\theta_0)-\theta'\phi(x).$$

(vii) For each $\theta \in I\!\!R$,

$$a_n^{-2}\log\left(n\Pr\{|r(X,n^{-1/2}a_n\theta)| \ge a_n^2\}\right) \to 0.$$

(viii) For each $\theta \in I\!\!R$ *and each* $\lambda > 0$,

$$na_n^{-2}E[\exp(\lambda|r(X,n^{-1/2}a_n\theta)|)I(a_n^2 \ge |r(X,n^{-1/2}a_n\theta)| \ge 1)] \to 0.$$

(ix) For each $\theta \in I\!\!R$,

$$na_n^{-2}\mathrm{Var}(r(X,n^{-1/2}a_n\theta)I(|r(X,n^{-1/2}a_n\theta)| \le 1)) \to 0.$$

Then,

$$\lim_{n\to\infty} a_n^{-2}\log(\Pr\{|n^{1/2}(\hat\theta_n-\theta_0)+2^{-1}n^{1/2}(P_n-P)V^{-1}\phi| \ge a_n\tau\}) = -\infty.$$

Most the conditions above are easier to check. Hypothesis (ii) in the previous theorem is satisfied if the estimator $\hat{\theta}_n$ is close enough to maximize $\sum_{j=1}^n g(X_j,\theta)$, $\theta\in\Theta$. The previous theorem gives the moderate deviations of the sample mean assuming the minimal condition (1.3). In this case $g(x,\theta)=|x-\theta|^2$, for $x,\theta\in I\!\!R^d$, $\theta_0=E[X]$, $\phi(x)=x-\theta_0$ and $r(x,\theta)=|\theta|^2$.

The conditions in Theorem 2.1 are implied by the existence of certain moment generating function in a neighborhood of zero:

Corollary 2.1. *Assume the notation in Theorem 2.1. Suppose that:*

(i) For each $x\in I\!\!R^d$, $g(x,\cdot):\Theta\to I\!\!R$ is a convex function.

(ii) $\hat{\theta}_n=\hat{\theta}_n(X_1,\ldots,X_n)$ is a sequence of $I\!\!R^d$–valued random variables such that for each $\tau>0$,

$$\lim_{n\to\infty} a_n^{-2}\log(\Pr\{a_n^{-2}\sum_{j=1}^n g(X_j,\hat{\theta}_n)\geq \inf_{\theta\in\Theta} a_n^{-2}\sum_{j=1}^n g(X_j,\theta)+\tau\})=-\infty.$$

(iii) There exists a positive definite symmetric matrix V such that

$$E[g(X,\theta)-g(X,\theta_0)]=(\theta-\theta_0)'V(\theta-\theta_0)+o(|\theta-\theta_0|^2),$$

as $\theta\to\theta_0$.

(iii) With probability one,

$$\lim_{\theta\to 0}|\theta|^{-1}|r(X,\theta)|.$$

(iv) There are $\delta_0,\lambda_0>0$ and such that

$$E[\exp(\lambda_0|\phi(X)|)]<\infty \quad\text{and}\quad E[\exp(\lambda_0 L(X))]<\infty,$$

where

$$L(x)=\sup_{0<|\theta|\leq\delta_0}|\theta|^{-1}|r(x,\theta)|.$$

Then,

$$\lim_{n\to\infty} a_n^{-2}\log(\Pr\{|n^{1/2}(\hat{\theta}_n-\theta_0)+2^{-1}n^{1/2}(P_n-P)V^{-1}\phi|\geq a_n\tau\})=-\infty.$$

Example 2.1. (Spatial median). Let $\{X_j\}$ be a sequence of i.i.d.r.v.'s with values in $I\!R^d$. A natural extension of the median to several dimensions is the spatial median $\hat{\theta}_n$ defined as a value which minimizes

$$n^{-1}\sum_{j=1}^{n}|X_j-\theta|, \theta\in I\!R^d.$$

This estimator is equivariant by rotations, translations and dilations. This estimator was introduced by Haldane (1948) as a robust alternative to the sample mean. If X is not supported in a linear space of dimension one, then there exists a unique value θ_0 such that for each $\theta\neq\theta_0$, $E[|X-\theta|-|X-\theta_0|]>0$ (see Milasevic and Ducharme, 1987). Assuming also that $E[|X-\theta_0|^{-1}]<\infty$, Corollary 2.1 applies with $g(x,\theta)=|x-\theta|$,

$$V=2^{-1}E[|X-\theta_0|^{-1}I_{d\times d}-|X-\theta_0|^{-3}(X-\theta_0)(X-\theta_0)'],$$

and $\phi(x)=-|x-\theta_0|^{-1}(x-\theta_0)I(x\neq\theta_0)$, where $I_{d\times d}$ is the unit $d\times d$ matrix. Conditions (i) and (ii) in Corollary 2.1 are obviously satisfied. By the Taylor theorem, there exists a constant c such that for each $x,\theta\in I\!R^d$,

$$\begin{aligned}&||x-\theta|-|x-\theta_0|+|x-\theta_0|^{-1}(\theta-\theta_0)'(x-\theta_0)\\&+2^{-1}|x-\theta_0|^{-3}((\theta-\theta_0)'(x-\theta_0))^2-2^{-1}|x-\theta_0|^{-1}|\theta-\theta_0|^2|\\\leq\;&c(|x-\theta_0|^{-1}|\theta-\theta_0|^2\wedge|x-\theta_0|^{-2}|\theta-\theta_0|^3).\end{aligned}$$

This implies that

$$\begin{aligned}&|\theta-\theta_0|^{-2}|E[g(X,\theta)-g(X,\theta_0)]-(\theta-\theta_0)'V(\theta-\theta_0)|\\\leq\;&cE[(|X-\theta_0|^{-1}\wedge|X-\theta_0|^{-2}|\theta-\theta_0|)],\end{aligned}$$

which goes to zero as $\theta\to\theta_0$ by the convergence dominated theorem. So, condition (iii) in Theorem 2.1 holds. Since, $|\phi(x)|\leq 1$, the part in condition (iv) involving $\phi(\cdot)$ is satisfied. By the Taylor theorem, there exists a constant c such that for each $x,\theta\in I\!R^d$,

$$|r(x,\theta)|\leq c(|x-\theta_0|^{-1}|\theta|^2\wedge|\theta|).$$

This implies conditions (iii)–(iv) in Corollary 2.1.

Example 2.2. (Some mle's for location). Let $\{X_i\}$ be a sequence of $I\!R^d$–valued i.i.d.r.v.'s with density $p(x-\theta_0)$, where $\theta_0\in I\!R^d$ is unknown and p is a density such that $-\log p(x)$ is a convex function. Then, the mle is the M–estimator which minimizes (1.1) with $g(x, \theta) = -\log p(x-\theta)$. The previous theorem gives sufficient conditions in p for the moderate deviations of the mle. It is elementary to see that the previous theorem applies when p is either normal, or double exponential or logistic.

The following theorem applies when the observations are not i.i.d.

Theorem 2.2. *Let Θ be a subset of $\mathbb{R}^d$. Let $\{G_n(\theta) : \theta \in \Theta\}$ be a sequence of stochastic processes. Let θ_0 be a point in the interior of Θ. Let $\{M_n\}$ and let $\{V_n\}$ be two sequences of nonsingular symmetric $d \times d$ matrices. Let $\{\eta_n\}$ be a sequence of $\mathbb{R}^d$–valued r.v.'s. Let $\{\epsilon_n\}$ be a sequence of real numbers converging to zero. Suppose that:*

(i) $G_n(\theta)$ is a convex function in θ.

(ii) $\hat{\theta}_n = \hat{\theta}_n(X_1, \ldots, X_n)$ is a sequence of $\mathbb{R}^d$–valued random variables such that for each $\tau > 0$,

$$\lim_{n\to\infty} \epsilon_n \log(\Pr\{G_n(\hat{\theta}_n) \geq \inf_{\theta\in\Theta} G_n(\theta) + \tau\}) = -\infty.$$

(iii) For each $\theta \in \mathbb{R}^d$ and each $\tau > 0$,

$$\lim_{n\to\infty} \epsilon_n \log \Pr(\{|G_n(\theta_0 + M_n^{-1}\theta) - G_n(\theta_0) - \theta'\eta_n - \theta' V_n \theta| \geq \tau\}) = -\infty.$$

(iv)

$$\lim_{M\to\infty} \limsup_{n\to\infty} \epsilon_n \log(\Pr\{|\eta_n| \geq M\}) = -\infty.$$

(v) $\liminf_{n\to\infty} \inf_{|\theta|=1} \theta' V_n \theta > 0$ and $\limsup_{n\to\infty} \sup_{|\theta|=1} \theta' V_n \theta < \infty$.

Then, for each $\tau > 0$,

$$\lim_{n\to\infty} \epsilon_n \log \Pr(\{|M_n(\hat{\theta}_n - \theta_0) + 2^{-1} V_n^{-1} \eta_n| \geq \tau\}) = -\infty.$$

Moreover, if $\{V_n^{-1}\eta_n\}$ satisfies the LDP with speed ϵ_n, then so does $\{M_n(\hat{\theta}_n - \theta_0)\}$.

The previous theorem applies to the next example.

Example 2.3. (Least absolute deviation estimator in linear regression). We obtain observations $(Y_1, z_1), \ldots, (Y_n, z_n)$, where $Y_i = z_i'\theta_0 + U_i$, $1 \leq i \leq$

n, $\{U_i\}_{i=1}^n$ are i.i.d.r.v.'s; $\{z_i\}_{i=1}^n$ are $\mathbb{R}^d$ vectors and $\theta_0 \in \mathbb{R}^d$ is a parameter to be estimated. z_i is called the regressor or predictor variable. Y_i is called the response variable. U_i is an error variable. The least absolute deviation estimator $\hat{\theta}_n$ of θ_0 is a value such that

$$\sum_{i=1}^n |Y_i - z_i'\hat{\theta}_n| = \inf_{\theta\in\mathbb{R}^d} \sum_{i=1}^n |Y_i - z_i'\theta|.$$

The least absolute deviation estimator is preferred to the least squares estimator by robustness reasons. A review in his topic is in Portnoy and Koenker (1997).

Theorem 2.3. *With the above notation, suppose that:*

(i) For n large enough, $S_n = \sum_{j=1}^n z_j z_j'$ is an invertible $d \times d$ matrix.

(ii) $a_n \max_{1\le j\le n} |S_n^{-1/2} z_j| \to 0$.

(iii) $F_U(0) = 1/2$ and $F_U'(0) > 0$, where F_U is the distribution function of U.

Then, $a_n^{-1} S_n^{1/2}(\hat{\theta}_n - \theta_0)$ satisfies the LDP with speed a_n^2 and rate function $I(t) = 2^{-1}(F_U'(0))^2 |t|^2$.

Next, we give sufficient conditions for the moderate deviations of the M–estimators in (1.1) without assuming that the kernel is convex.

Theorem 2.4. *With the notation in Theorem 2.1. Let Ψ be a function form S into the set of $d \times d$ symmetric matrices. Suppose that:*

(i) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|\hat{\theta}_n - \theta_0| \ge \tau\}) = -\infty.$$

(ii) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{\sum_{i=1}^n g(X_i, \hat{\theta}_n) \ge \inf_{\theta\in\Theta} \sum_{i=1}^n g(X_i, \theta) + \tau a_n^2\}) = -\infty.$$

(iii) There is a positive definite symmetric $d \times d$ matrix V such that

$$E[g(X,\theta) - g(X,\theta_0)] = (\theta - \theta_0)' V (\theta - \theta_0) + o(|\theta - \theta_0|^2),$$

as $\theta \to \theta_0$.

(iv) $E[|\phi(X)|^2] < \infty$

(v) $\lim_{n\to\infty} a_n^{-2} \log(n \Pr\{|\phi(X)| \geq n^{1/2} a_n\}) = -\infty$.

(vi) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|n^{-1} \sum_{i=1}^n (\Psi(X_i) - E[\Psi(X_i)])| \geq \tau\}) = -\infty.$$

(vii) $\lim_{\delta\to 0} E[B_\delta(X)] = 0$, *where*

$$B_\delta(x) = \sup_{|\theta|\leq\delta} |\theta-\theta_0|^{-2} |g(x,\theta)-g(x,\theta_0)-(\theta-\theta_0)'\phi(x)-(\theta-\theta_0)'\Psi(x)(\theta-\theta_0)|.$$

(viii) For each $\tau > 0$, *there exists a* $\delta > 0$ *such that*

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{n^{-1} \sum_{i=1}^n B_\delta(X_i) \geq \tau\}) = -\infty.$$

Then,

$$\begin{aligned}&\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|n^{1/2}(\hat{\theta}_n - \theta_0)\\ &+2^{-1} a_n n^{-1} \textstyle\sum_{i=1}^n V^{-1}(\phi(X_i) - E[\phi(X_i)])| \geq a_n \tau\}) = -\infty.\end{aligned}$$

Condition (vi) in the previous theorem holds if for some $\lambda > 0$, $E[\exp(\lambda|\Psi(X)|)] < \infty$ (see the remark around (3.4) below). Similarly, condition (viii) in the previous theorem holds if for some $\lambda > 0$, $E[\exp(\lambda B_\delta(X))] < \infty$. To check hypothesis (i) in the previous theorem, it helps if there exists a unique local minimum of $G_n(\theta)$. Conditions for a function to have a unique minimum are in Mäkeläinen, Schmidt and Styan (1981). Using this condition, we have the following:

Theorem 2.5. *With the notation in the previous theorem. Suppose also that there exists a* $\tau_0 > 0$ *such that:*

(i) With probability one, $G_n(\theta)$, $\theta \in \Theta$, *has a unique minimum* $\hat{\theta}_n$ *and it does not have any other local minimum.*

(ii) With probability one, $G_n(\cdot)$ *is continuous in* Θ.

(iii) There is a positive definite symmetric $d \times d$ matrix V such that

$$E[g(X,\theta) - g(X,\theta_0)] = (\theta - \theta_0)'V(\theta - \theta_0) + o(|\theta - \theta_0|^2),$$

as $\theta \to \theta_0$.

(iv) $E[|\phi(X)|^2] < \infty$

(v) $\lim_{n\to\infty} a_n^{-2} \log(n \Pr\{|\phi(X)| \geq n^{1/2} a_n\}) = -\infty$.

(vi) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|n^{-1} \sum_{i=1}^{n} (\Psi(X_i) - E[\Psi(X_i)])| \geq \tau\}) - \infty.$$

(vii) $\lim_{\delta\to 0} E[B_\delta(X)] = 0$, *where*

$$B_\delta(x) = \sup_{|\theta|\leq\delta} |\theta-\theta_0|^{-2} |g(x,\theta) - g(x,\theta_0) - (\theta-\theta_0)'\phi(x) - (\theta-\theta_0)'\Psi(x)(\theta-\theta_0)|.$$

(viii) For each $\tau > 0$, *there exists a* $\delta > 0$ *such that*

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{n^{-1} \sum_{i=1}^{n} B_\delta(X_i) \geq \tau\}) = -\infty.$$

Then,

$$\begin{aligned} &\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|n^{1/2}(\hat{\theta}_n - \theta_0) \\ &+2^{-1} a_n n^{-1} \textstyle\sum_{i=1}^{n} V^{-1}(\phi(X_i) - E[\phi(X_i)])| \geq a_n \tau\}) = -\infty. \end{aligned}$$

The previous theorem applied to the moderate deviations of mle's gives the following:

Theorem 2.6. *Let* $\{p(x|\theta) : \theta \in \Theta\}$ *be a family of densities in* $I\!R^m$, *where* Θ *is a Borel set of* $I\!R^d$. *Let* $\{X_i\}$ *be a sequence of* $I\!R^m$*–valued i.i.d.r.v.'s with density* $p(x|\theta_0)$ *where* $\theta_0 \in \Theta^{\mathrm{o}}$. *Suppose that there exists a* $\delta_0 > 0$ *such that:*

(i) $A := \{x \in I\!R^m : p(x|\theta) > 0\}$ *does not depend on* θ.

(ii) $p(x|\theta)$ *is twice differentiable with continuity with respect to* θ *in a neighborhood of* θ_0.

(iii) With probability one, $n^{-1}\sum_{i=1}^n \log p(X_i|\theta)$, $\theta \in \Theta$ *has a unique global maximum at* $\hat{\theta}_n$ *and this is the unique local maximum.*

(iv) $0 < E[|\phi(X)|^2] < \infty$, *where*

$$\phi(x) = \left(\frac{\partial \log p(x|\theta)}{\partial \theta^{(1)}}\Big|_{\theta=\theta_0}, \ldots, \frac{\partial \log p(x|\theta)}{\partial \theta^{(d)}}\Big|_{\theta=\theta_0} \right)'$$

and $\theta = (\theta^{(1)}, \ldots, \theta^{(d)})'$.

(v) $\lim_{n\to\infty} a_n^{-2} \log(n \Pr\{|\phi(X)| \geq n^{1/2} a_n\}) = -\infty$.

(vi) The matrix $V := \left(E\left[\frac{\partial^2 \log p(X|\theta)}{\partial \theta^{(i)} \partial \theta^{(j)}}\Big|_{\theta=\theta_0} \right] \right)_{1\leq i,j\leq d}$ *is positive definite.*

(vii) There exists a $\lambda > 0$ *such that*

$$E[\exp(\lambda B_{\delta_0}(X))] < \infty,$$

where

$$B_{\delta_0}(x) = \sup_{1\leq i,j\leq d} \sup_{|\theta-\theta_0|<\delta_0} \left| \frac{\partial^2 \log p(x|\theta)}{\partial \theta^{(i)} \partial \theta^{(j)}} \right|.$$

Then,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|n^{1/2}(\hat{\theta}_n - \theta_0) + 2^{-1} a_n n^{-1} \sum_{i=1}^n (V^{-1}(\phi(X_i) - E[\phi(X_i)])| \geq a_n \tau\}) = -\infty.$$

Theorem 2.1 in Mäkeläinen, Schmidt and Styan (1981) gives sufficient conditions implying hypothesis (iii) in the previous theorem. It is easy to see that the previous theorem applies to many common mle's. Another sufficient conditions assuming less smoothness for the moderate deviations of mle's are in Theorem 2.9.

Next, we consider the moderate deviations of the M–estimators in (1.2). The next theorem considers the moderate deviations for the M–estimators over a nondecreasing kernel:

Theorem 2.7. *Let* $\{X_i\}_{i=1}^\infty$ *be a sequence of i.i.d.r.v.'s with values in a measurable space* $(S, \mathcal{S})$. *Let* $h : S \times \mathbb{R} \to \mathbb{R}$ *be function such that* $h(\cdot, \theta) : S \to \mathbb{R}$ *is measurable for each* θ *and* $h(x, \cdot) : \mathbb{R} \to \mathbb{R}$ *is nondecreasing for each* x. *Let* $\theta_0 \in \mathbb{R}$. *Let* $\{a_n\}$ *be a sequence of real numbers converging to* *infinity such that* $a_n n^{-1/2} \to 0$. *Let* $\hat{\theta}_n = \sup\{t : n^{-1}\sum_{j=1}^n h(X_j, t) \le 0\}$. *Assume that:*

(i) $H(\theta_0) = 0$ *and* $H'(\theta_0) > 0$, *where* $H(\theta) := E[h(X, \theta)]$.

(ii) $E[h^2(X, \theta_0)] < \infty$.

(iii) $\lim_{n\to\infty} a_n^{-2} \log(n \Pr\{|h(X, \theta_0)| \ge a_n n^{1/2}\}) = -\infty$.

(iv) For each $t \in \mathbb{R}$,

$$n^{1/2} a_n^{-1} E[|r(X, n^{-1/2} a_n t)| I(|r(X, n^{-1/2} a_n t)| \ge n^{1/2} a_n)] = 0,$$

where $r(X, \theta) = h(X, \theta_0 + \theta) - h(X, \theta_0)$.

(v) For each $t \in \mathbb{R}$,

$$\lim_{n\to\infty} a_n^{-2} \log(n \Pr\{|r(X, n^{-1/2} a_n t)| \ge a_n n^{1/2}\}) = -\infty.$$

(vi) For each $t \in \mathbb{R}$ *and each* $\lambda > 0$,

$$\begin{aligned}&\lim_{n\to\infty} n a_n^{-2} E\left[\exp\left(\lambda n^{-1/2} a_n |r(X, n^{-1/2} a_n t)|\right)\right.\\ &\left.\times I(a_n n^{1/2} \ge |r(X, n^{-1/2} a_n t)| \ge n^{1/2} a_n^{-1})\right] = 0.\end{aligned}$$

(vii) For each $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathrm{Var}(r(X, n^{-1/2} a_n t) I(|r(X, n^{-1/2} a_n t)| \le n^{1/2} a_n^{-1})) = 0.$$

Then, for each $\tau > 0$,

$$\begin{aligned}&\lim_{n\to\infty} a_n^{-2} \log(\Pr\{a_n^{-1} n^{1/2} |\hat{\theta}_n - \theta_0\\ &\quad + (H'(\theta_0))^{-1} n^{-1} \textstyle\sum_{j=1}^n (h(X_j, \theta_0) - E[h(X_j, \theta_0)])| \ge \tau\}) = -\infty.\end{aligned}$$

Example 2.4. (p–th quantile). Let $\{X_i\}$ be a sequence of i.i.d.r.v.'s with df F. Let $0 < p < 1$. Suppose that there exists θ_0 such that $F(\theta_0) = p$. The M-estimator over $h(x, \theta) = I(x \le \theta) - p$ is the p–th quantile. It is easy to see that if $F'(\theta_0) > 0$, then the previous theorem applies.

Example 2.5. (Parameters of location). Let ψ be a nonincreasing odd function with $\psi(0) = 0$. Take $h(x, \theta) = \psi(x - \theta)$. Several possible common choices for ψ are in Chapter 7 in Serfling (1980). Then, under minimal conditions the previous theorem hold. For example, if ψ is bounded, the previous theorem applies with θ_0 if $H(\theta_0) = 0$, $H'(\theta_0) > 0$ and $\lim_{\theta\to\theta_0} E[(h(X, \theta) - h(X, \theta_0))^2] = 0$.

The next theorem applies to the M–estimators in (1.2) in the multivariate situation:

Theorem 2.8. *Let $\{X_i\}_{i=1}^{\infty}$ be a sequence of i.i.d.r.v.'s with values in a measurable space $(S,\mathcal{S})$. Let $h : S \times I\!R^d \to I\!R^d$ be function such that $h(\cdot,\theta) : S \to I\!R$ is measurable for each $\theta \in \Theta$. Let $\theta_0 \in \Theta^o$. Let $\{a_n\}$ be a sequence of real numbers converging to infinity such that $a_n n^{-1/2} \to 0$. Suppose that:*

(i) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|\hat{\theta}_n - \theta_0| \ge \tau\}) = -\infty.$$

(ii) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{a_n^{-1} n^{1/2} |P_n h(\cdot,\hat{\theta}_n)| \ge \tau\}) = -\infty.$$

(iii) $H(\theta_0) = 0$, where $H(\theta) := E[h(X,\theta)]$

(iv) $H(\theta)$ is differentiable at θ_0 with nonsingular derivative.

(v) There exists a $\delta_0 > 0$ such that $\{a_n^{-1} n^{1/2}(P_n - P)(h(\cdot,\theta) - h(\cdot,\theta_0)) : |\theta - \theta_0| \le \delta_0\}$ satisfies the LDP with speed a_n^2 and a good rate function.

(vi) $\lim_{\theta\to\theta_0} \mathrm{Var}(h(X,\theta) - h(X,\theta_0)) = 0$.

Then, for each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{a_n^{-1} n^{1/2} |\hat{\theta}_n - \theta_0 + (H'(\theta_0))^{-1}(P_n - P)h(\cdot,\theta_0)| \ge \tau\}) = -\infty.$$

Necessary and sufficient conditions for the moderate large deviations of empirical processes were given in Theorem 3.6 in Arcones (2001). Applying these necessary and sufficient conditions to hypothesis (v) in the previous theorem, we obtain the following:

Corollary 2.2. *With the notation in the previous theorem, suppose that there exists a $\delta_0 > 0$ such that:*

(i) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|\hat{\theta}_n - \theta_0| \ge \tau\}) = -\infty.$$

(ii) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{a_n^{-1} n^{1/2} |P_n h(\cdot,\hat{\theta}_n)| \ge \tau\}) = -\infty.$$

(iii) $H(\theta_0) = 0$*, where* $H(\theta) := E[h(X, \theta)]$

(iv) $H(\theta)$ *is differentiable at* θ_0 *with nonsingular derivative.*

(v) For each $|\theta - \theta_0| \le \delta_0$*,* $E[|h(X, \theta)|^2] < \infty$*.*

(vi)

$$\lim_{n\to\infty} a_n^{-2} \log(n \Pr\{\sup_{|\theta-\theta_0|\le\delta_0} |h(X, \theta)| \ge n^{1/2} a_n\}) = -\infty.$$

(vii) $(\{\theta : |\theta - \theta_0| \le \delta_0\}, d)$ *is totally bounded, where*

$$d^2(\theta', \theta'') = \mathrm{Var}(h(X, \theta') - h(X, \theta'')).$$

(viii) $\sup_{|\theta-\theta_0|\le\delta_0} a_n^{-1} n^{-1/2} |\sum_{j=1}^n (h(X_j, \theta) - E[h(X_j, \theta)])| \xrightarrow{\Pr} 0.$

(ix) $\lim_{\theta\to\theta_0} \mathrm{Var}(h(X, \theta) - h(X, \theta_0)) = 0.$

Then, for each $\tau > 0$*,*

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{a_n^{-1} n^{1/2} |\hat{\theta}_n - \theta_0 + (H'(\theta_0))^{-1} (P_n - P) h(\cdot, \theta_0)| \ge \tau\}) = -\infty.$$

Condition (viii) in the previous theorem holds if

$$\{n^{-1/2} \sum_{j=1}^n (h(X_j, \theta) - E[h(X_j, \theta)]) : |\theta - \theta_0| \le \delta_0\}$$

converges weakly. The weak convergence of the previous stochastic process has been study by many authors (see for example Giné and Zinn, 1984; van der Vaart and Wellner, 1996; and Dudley, 1999).

The previous corollary applied to the moderate deviations of mle's gives the following:

Theorem 2.9. *Let* $\{p(x|\theta) : \theta \in \Theta\}$ *be a family of densities in* $\mathbb{R}^m$*, where* Θ *is a Borel set of* $\mathbb{R}^d$*. Let* $\{X_i\}$ *be a sequence of* $\mathbb{R}^m$*–valued i.i.d.r.v.'s with density* $p(x|\theta_0)$ *where* $\theta_0 \in \Theta^o$*. Suppose that there exists a* $\delta_0 > 0$ *such that:*

(i) $A := \{x \in \mathbb{R}^m : p(x|\theta) > 0\}$ *does not dependent on* θ*.*

(ii) $p(x|\theta)$ *is second differentiable with respect to* θ *in a neighborhood of* θ_0*.*

(iii) With probability one, $n^{-1}\sum_{i=1}^n \log p(X_i|\theta)$*,* $\theta\in\Theta$*, has a unique global maximum at* $\hat\theta_n$ *and this is the unique local maximum.*

(iv) There exists a $\lambda>0$ *such that*

$$E\left[\exp\left(\lambda \sup_{1\le i\le d}\sup_{|\theta-\theta_0|<\delta_0}\left|\frac{\partial \log p(x|\theta)}{\partial\theta^{(i)}}\right|\right)\right]<\infty.$$

(v) The matrix $V := \left(E\left[\frac{\partial^2 \log p(X|\theta)}{\partial\theta^{(i)}\partial\theta^{(j)}}\Big|_{\theta=\theta_0}\right]\right)_{1\le i,j\le d}$ *is positive definite.*

(vi)

$$E\left[\sup_{1\le i,j\le d}\sup_{|\theta-\theta_0|<\delta_0}\left|\frac{\partial^2 \log p(X|\theta)}{\partial\theta^{(i)}\partial\theta^{(j)}}\right|^2\right]<\infty.$$

Then,

$$\begin{aligned}&\lim_{n\to\infty} a_n^{-2}\log(\Pr\{|n^{1/2}(\hat\theta_n-\theta_0)\\&+2^{-1}a_n n^{-1}\textstyle\sum_{i=1}^n(V^{-1}(\phi(X_i)-E[\phi(X_i)])|\ge a_n\tau\})=-\infty,\end{aligned}$$

where

$$\phi(x)=\left(\frac{\partial\log p(x|\theta)}{\partial\theta^{(1)}}\Big|_{\theta=\theta_0},\dots,\frac{\partial\log p(x|\theta)}{\partial\theta^{(d)}}\Big|_{\theta=\theta_0}\right)'.$$

3 Proofs.

Besides of imposing conditions like (1.3), we will need to deal with some remainders. To do that we will use the following lemma:

Lemma 3.1. *Let* $\{X_{n,j}: 1\le j\le n\}$ *be a triangular array of row wise independent r.v.'s. Let* $\{\epsilon_n\}$ *be a sequence of positive numbers which converges to zero. Suppose that:*

(i)

$$\lim_{M\to\infty}\limsup_{n\to\infty}|\sum_{j=1}^n E[X_{n,j}I(|X_{n,j}|\ge M)]|=0.$$

(ii)

$$\lim_{M\to\infty}\limsup_{n\to\infty}\epsilon_n\log\left(\sum_{j=1}^n\Pr\{|X_{n,j}|\ge M\}\right)=-\infty.$$

(iii) For each $0 < M, \lambda < \infty$,

$$\lim_{a\to\infty} \limsup_{n\to\infty} \epsilon_n \sum_{j=1}^{n} \log \left(E[\exp(\lambda \epsilon_n^{-1} |X_{n,j}| I(M > |X_{n,j}| \geq a\epsilon_n))] \right) = 0.$$

(iv) For each $a > 0$,

$$\epsilon_n^{-1} \sum_{j=1}^{n} \mathrm{Var}(X_{n,j} I(|X_{n,j}| < a\epsilon_n)) = 0.$$

Then, for each $\tau > 0$,

$$\limsup_{n\to\infty} \epsilon_n \log \Pr\{|\sum_{j=1}^{n} (X_{n,j} - E[X_{n,j}])| \geq \tau\} = -\infty.$$

PROOF. We prove that for each $\lambda, \tau > 0$,

$$\limsup_{n\to\infty} \epsilon_n \log \left(\Pr\{|\sum_{j=1}^{n} (X_{n,j} - E[X_{n,j}])| \geq \tau\} \right) \leq -\lambda.$$

Take $0 < M < \infty$ such that

$$\limsup_{n\to\infty} |\sum_{j=1}^{n} E[X_{n,j} I(|X_{n,j}| \geq M)]| \leq 2^{-2}\tau \tag{3.1}$$

and

$$\limsup_{n\to\infty} \epsilon_n \log(\sum_{j=1}^{n} \Pr\{|X_{n,j}| \geq M\}) < -\lambda. \tag{3.2}$$

Take $a > 0$ such that

$$\limsup_{n\to\infty} \epsilon_n \log(E[\exp(16\lambda\tau^{-1}\epsilon_n^{-1} \sum_{j=1}^{n} |X_{n,j}| I(M > |X_{n,j}| \geq a\epsilon_n))]) < \lambda. \tag{3.3}$$

By (3.1), for n large enough,

$$\begin{aligned}
& \Pr\{|\textstyle\sum_{j=1}^{n} (X_{n,j} - E[X_{n,j}])| \geq \tau\} \\
\leq\ & \textstyle\sum_{j=1}^{n} \Pr\{|X_{n,j}| \geq M\} \\
& + \Pr\{|\textstyle\sum_{j=1}^{n} (X_{n,j} I(a\epsilon_n \leq |X_{n,j}| < M) \\
& \qquad - E[X_{n,j} I(a\epsilon_n \leq |X_{n,j}| < M)])| \geq 2^{-2}\tau\} \\
& + \Pr\{|\textstyle\sum_{j=1}^{n} (X_{n,j} I(|X_{n,j}| < a\epsilon_n) - E[X_{n,j} I(|X_{n,j}| < a\epsilon_n)])| \geq 2^{-2}\tau\} \\
& =: I + II + III.
\end{aligned}$$

By (3.2), for n large enough $I \le e^{-\lambda\epsilon_n^{-1}}$.

By symmetrization (see for example Lemma 6.3 in Ledoux and Talagrand, 1991) and (3.3), for n large enough,

$$\begin{aligned} & II \\ \le\ & e^{-2\lambda\epsilon_n^{-1}} E[\exp(8\lambda\tau^{-1}\epsilon_n^{-1}|\textstyle\sum_{j=1}^n (X_{n,j}I(a\epsilon_n \le |X_{n,j}| < M) \\ & -E[X_{n,j}I(a\epsilon_n \le |X_{n,j}| < M))|)]) \\ \le\ & e^{-2\lambda\epsilon_n^{-1}} E[\exp(16\lambda\tau^{-1}\epsilon_n^{-1}|\textstyle\sum_{j=1}^n \xi_j X_{n,j}I(a\epsilon_n \le |X_{n,j}| < M)|)] \\ \le\ & e^{-2\lambda\epsilon_n^{-1}} E[\exp(16\lambda\tau^{-1}\epsilon_n^{-1}\textstyle\sum_{j=1}^n |X_{n,j}|I(a\epsilon_n \le |X_{n,j}| < M))] \\ \le\ & e^{-\lambda\epsilon_n^{-1}}, \end{aligned}$$

where $\{\xi_j\}$ is a sequence of i.i.d. Rademacher r.v.'s independent of $\{X_{n,j} : 1 \le j \le n\}$.

By the Prokhorov inequality (Theorem 1 in Prokhorov, 1959)

$$\begin{aligned} & III \\ \le\ & 2\exp\big(-2^{-4}a^{-1}\tau\epsilon_n^{-1} \\ & \times \operatorname{arcsinh}\Big(2^{-2}a\tau\epsilon_n(\textstyle\sum_{j=1}^n \mathrm{Var}(X_{n,j}I(|X_{n,j}| < a\epsilon_n)))^{-1}\Big)\Big). \end{aligned}$$

The claim follows from all the previous estimations. □

We will also use that if $\{Y_i\}$ is a sequence of i.i.d.r.v.'s such that for some $\lambda > 0$, $E[\exp(\lambda|Y|)] < \infty$, and $\{a_n\}$ is a sequence of real numbers such that $na_n^{-2} \to \infty$, then for each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{n^{-1}|\sum_{i=1}^n (Y_i - E[Y_i])| \ge \tau\}) = -\infty. \tag{3.4}$$

This follows from the Cramer–Chernoff theorem: for each $\tau > 0$,

$$\lim_{n\to\infty} n^{-1} \log(\Pr\{n^{-1}\sum_{i=1}^n (Y_i - E[Y_i])| \ge \tau\}) = -\min(I(\tau), I(-\tau)),$$

where

$$I(z) = \sup\{\lambda z - \log(E[\exp(\lambda(Y - E[Y]))]) : \lambda \in I\!R\}$$

and $I(z) > 0$ for $z \ne 0$ (see for example Section 1.2 in Deuschel and Stroock, 1989).

We obtain Theorem 2.1 from Theorem 2.2. So, we prove Theorem 2.2 first. We need the following lemma:

Lemma 3.2. *Let $f : [-M-\tau, M+\tau]^d \to I\!R$ be a convex function, then for each $x, y \in [-M, M]^d$,*

$$|f(x) - f(y)| \le 2 \cdot 3^d |x-y|_\infty \tau^{-1} \sup_{z \in \{-M-\tau, 0, M+\tau\}^d} |f(z)|.$$

The previous lemma follows from lemmas 13 and 14 in Arcones (1998a).

Proof of Theorem 2.2. Let $U_n(\theta) = G_n(\theta_0 + M_n^{-1}\theta) - G_n(\theta_0) - \theta'\eta_n$ and let $W_n(\theta) = \theta' V_n \theta$ and let $Z_n(\theta) = U_n(\theta) - W_n(\theta)$. Let $c_1 = \liminf_{n\to\infty} \inf_{|\theta|=1} \theta' V_n \theta$ and let $c_2 := \limsup_{n\to\infty} \sup_{|\theta|=1} \theta' V_n \theta$.

First, we prove that for each $0 < \tau, M < \infty$,

$$\lim_{n\to\infty} \epsilon_n \log(\Pr\{\sup_{|\theta|\le M} |Z_n(\theta)| \ge \tau\}) = -\infty. \tag{3.5}$$

Take $0 < \epsilon < 1$, such that $4 \cdot 3^d \epsilon c_2 (M+1)^2 < \tau/2$. Then, there are $\theta_1, \dots, \theta_m \in [-M, M]^d$ and a function $\pi : [-M, M]^d \to \{\theta_1, \dots, \theta_m\}$ such that $\sup_{\theta \in [-M,M]^d} |\theta - \pi(\theta)| \le \epsilon$. By the Lemma 3.2, for n large enough, and each $\theta \in [-M, M]^d$,

$$\begin{aligned}
& |Z_n(\theta)| \\
\le\ & |Z_n(\pi(\theta))| + |U_n(\theta) - U_n(\pi(\theta))| + |W_n(\theta) - W_n(\pi(\theta)| \\
\le\ & \max_{1\le j\le m} |Z_n(\theta_j)| + 2 \cdot 3^d \epsilon \sup_{\theta \in \{-M-1,0,M+1\}^d} |U_n(\theta)| \\
& +2 \cdot 3^d \epsilon \sup_{\theta \in \{-M-1,0,M+1\}^d} |W_n(\theta)| \\
\le\ & \max_{1\le j\le m} |Z_n(\theta_j)| + 2 \cdot 3^d \epsilon \sup_{\theta \in \{-M-1,0,M+1\}^d} |Z_n(\theta)| \\
& +4 \cdot 3^d \epsilon \sup_{\theta \in \{-M-1,0,M+1\}^d} |W_n(\theta)| \\
\le\ & \max_{1\le j\le m} |Z_n(\theta_j)| + 2 \cdot 3^d \epsilon \sup_{\theta \in \{-M-1,0,M+1\}^d\}} |Z_n(\theta)| + \tau/2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \epsilon_n \log(\Pr\{\sup_{|\theta|\le M} |Z_n(\theta)| \ge \tau\}) \\
\le\ & \epsilon_n \log \big(\Pr\{\max_{1\le j\le m} |Z_n(\theta_j)| \ge 2^{-3}\tau\} \\
& + \Pr\{2 \cdot 3^d \epsilon \sup_{\theta \in \{-M-1,0,M+1\}^d} |Z_n(\theta)| \ge 2^{-3}\tau\}\big) \to -\infty
\end{aligned}$$

and (3.5) holds.

Let $\tilde{\theta}_n = -2^{-1} V_n^{-1} \eta_n$, let $\tau > 0$, let $|\theta| > \tau > 0$, let $t = \theta_0 + M_n^{-1}\tilde{\theta}_n + M_n^{-1}\theta$ and let

$$t^* := \theta_0 + M_n^{-1}\tilde{\theta}_n + \tau|\theta|^{-1} M_n^{-1}\theta = |\theta|^{-1}\tau t + |\theta|^{-1}(|\theta| - \tau)(\theta_0 + M_n^{-1}\tilde{\theta}_n).$$

By convexity,

$$\begin{array}{rl} & G_n(t^*) - G_n(\theta_0) \\ \le & |\theta|^{-1}\tau(G_n(t) - G_n(\theta_0)) + |\theta|^{-1}(|\theta| - \tau)(G_n(\theta_0 + M_n^{-1}\tilde{\theta}_n) - G_n(\theta_0)). \end{array} \tag{3.6}$$

Then, for n large enough,

$$\begin{array}{rl} & G_n(t^*) - G_n(\theta_0) \\ \ge & (\tilde{\theta}_n + \tau|\theta|^{-1}\theta)'\eta_n + (\tilde{\theta}_n + \tau|\theta|^{-1}\theta)'V_n(\tilde{\theta}_n + \tau|\theta|^{-1}\theta) \\ & - \sup_{|\theta| \le |\tilde{\theta}_n| + \tau} |Z_n(\theta)| \\ = & -2^{-2}\eta_n' V_n^{-1}\eta_n + |\theta|^{-2}\tau^2\theta' V_n\theta - \sup_{|\theta| \le |\tilde{\theta}_n| + \tau} |Z_n(\theta)| \\ \ge & G_n(\theta_0 + M_n^{-1}\tilde{\theta}_n) - G_n(\theta_0) + 2^{-1}c_1\tau^2 - 2\sup_{|\theta| \le |\tilde{\theta}_n| + \tau} |Z_n(\theta)|. \end{array} \tag{3.7}$$

By (3.6) and (3.7), for n large enough,

$$\begin{array}{rl} & G_n(\theta_0 + M_n^{-1}\tilde{\theta}_n) - G_n(\theta_0) + 2^{-1}c_1\tau^2 - 2\sup_{|\theta| \le |\tilde{\theta}_n| + \tau} |Z_n(\theta)| \\ \le & |\theta|^{-1}\tau(G_n(t) - G_n(\theta_0)) \\ & +|\theta|^{-1}(|\theta| - \tau)(G_n(\theta_0 + M_n^{-1}\tilde{\theta}_n) - G_n(\theta_0)). \end{array}$$

So, for $|\theta| > \tau$,

$$\begin{array}{rl} & G_n(\theta_0 + M_n^{-1}\tilde{\theta}_n) - G_n(\theta_0) \\ & +|\theta|\tau^{-1}(2^{-1}c_1\tau^2 - 2\sup_{|\theta| \le |\tilde{\theta}_n| + \tau} |Z_n(\theta)|) \\ \le & G_n(\theta_0 + M_n^{-1}\tilde{\theta}_n + M_n^{-1}\theta) - G_n(\theta_0). \end{array}$$

This implies that if

$$\sup_{|\theta| \le |\tilde{\theta}_n| + \tau} |Z_n(\theta)| \le 2^{-3}c_1\tau^2$$

then, for each $|\theta| > \tau$,

$$\begin{array}{rl} & G_n(\theta_0 + M_n^{-1}\tilde{\theta}_n) - G_n(\theta_0) + 2^{-2}c_1\tau^2 \\ \le & G_n(\theta_0 + M_n^{-1}\tilde{\theta}_n + M_n^{-1}\theta) - G_n(\theta_0). \end{array}$$

Hence,

$$\begin{array}{rl} & \Pr\{|M_n(\hat{\theta}_n - \theta_0) + 2^{-1}V_n^{-1}\eta_n| \ge \tau\} \\ \le & \Pr\{\sup_{|\theta| \le M + \tau} |Z_n(\theta)| \ge 2^{-3}c_1\tau^2\} + \Pr\{|\tilde{\theta}_n| \ge M\} \\ & + \Pr\{G_n(\hat{\theta}_n) \ge \inf_{\theta \in \Theta} G_n(\theta) + 2^{-2}c_1\tau^2\} \end{array}$$

and the claim follows from (3.5) and hypotheses (ii) and (iv). □

Proof of Theorem 2.1. We apply Theorem 2.2 with $M_n = a_n^{-1}n^{1/2}I_{d\times d}$, $V_n = V$, $\eta_n = a_n^{-1}n^{-1/2}\sum_{j=1}^n(\phi(X_j) - E[\phi(X_j)])$, and $G_n(\theta) = a_n^{-2}\sum_{j=1}^n g(X_j,\theta)$, where $I_{d\times d}$ is the identity $d\times d$ matrix. Conditions (i), (ii), (iv) and (v) in Theorem 2.1 are obviously satisfied. To check condition (iii), we need to prove that for each $\theta \in I\!\!R^d$ and each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2}\log(\Pr\{a_n^{-2}n|(P_n-P)r(\cdot,a_nn^{-1/2}\theta)| \ge \tau\}) = -\infty,$$

This follows from Lemma 3.1. Conditions (i), (ii) and (iv) in Lemma 3.1 are obviously satisfied. To check condition (iii) in Lemma 3.1 notice that

$$\begin{aligned} & na_n^{-2}\log(E[\exp(\lambda|r(X,n^{-1/2}a_n\theta)|)I(a_n^2 \ge |r(X,n^{-1/2}a_n\theta)| \ge 1)]) \\ \simeq\ & na_n^{-2}E[\exp(\lambda|r(X,n^{-1/2}a_n\theta)|)I(a_n^2 \ge |r(X,n^{-1/2}a_n\theta)| \ge 1) - 1] \\ =\ & na_n^{-2}E[(\exp(\lambda|r(X,n^{-1/2}a_n\theta)|) - 1)I(a_n^2 \ge |r(X,n^{-1/2}a_n\theta)| \ge 1)] \\ \le\ & na_n^{-2}E[\exp(\lambda|r(X,n^{-1/2}a_n\theta)|)I(a_n^2 \ge |r(X,n^{-1/2}a_n\theta)| \ge 1)]. \ \square \end{aligned}$$

Proof of Corollary 2.1. We apply Theorem 2.1. Conditions (i)–(iii) in Theorem 2.1 are assumed. Condition (iv) implies conditions (iv) and (v) in Theorem 2.1. For n large enough, we have that

$$\begin{aligned} & na_n^{-2}E[|r(X,n^{-1/2}a_n\theta)|I(|r(X,n^{-1/2}a_n\theta)| \ge a_n^2)] \\ \le\ & |\theta|n^{1/2}a_n^{-1}E[L(X)I(|\theta|L(X) \ge n^{1/2}a_n)] \\ \le\ & |\theta|a_n^{-2}E[(L(X))^2I(|\theta|L(X) \ge n^{1/2}a_n)] \to 0. \end{aligned}$$

So, condition (vi) in Theorem 2.1 follows.

We have that for $\theta \ne 0$,

$$\begin{aligned} & a_n^{-2}\log\left(n\Pr\{|r(X,n^{-1/2}a_n\theta)| \ge a_n^2\}\right) \\ \le\ & a_n^{-2}\log\left(n\Pr\{|\theta|L(X) \ge a_nn^{1/2}\}\right) \\ \le\ & a_n^{-2}\log\left(nce^{-ca_nn^{1/2}}\right) \to 0, \end{aligned}$$

which implies condition (vii) in Theorem 2.1.

For each $\theta \ne 0$, each $\lambda > 0$ and each n large enough,

$$\begin{aligned} & na_n^{-2}E[\exp(\lambda|r(X,n^{-1/2}a_n\theta)|)I(a_n^2 \ge |r(X,n^{-1/2}a_n\theta)| \ge 1)] \\ \le\ & na_n^{-2}E[\exp(c\lambda n^{-1/2}a_nL(X))I(cL(X)n^{-1/2}a_n \ge 1)] \\ \le\ & cE[(L(X))^2\exp(c\lambda n^{-1/2}a_nL(X))I(cL(X)n^{-1/2}a_n \ge 1)] \\ \le\ & cE[\exp(c\lambda_0L(X))I(L(X) \ge cn^{1/2}a_n^{-1})] \to 0. \end{aligned}$$

So, condition (viii) in Theorem 2.1 holds.

Given $\theta \in I\!\!R$,

$$\begin{aligned} & n a_n^{-2} \mathrm{Var}(r(X, n^{-1/2} a_n \theta) I(|r(X, n^{-1/2} a_n \theta)| \le 1)) \\ \le \; & E[(n^{1/2} a_n^{-1} r(X, n^{-1/2} a_n \theta))^2]. \end{aligned}$$

By condition (iii), $n^{1/2} a_n^{-1} r(X, n^{-1/2} a_n \theta) \to 0$ a.s. and by condition (iv), $|n^{1/2} a_n^{-1} r(X, n^{-1/2} a_n \theta)| \le cL(X)$. So, by the dominated convergence theorem,

$$E[(n^{1/2} a_n^{-1} r(X, n^{-1/2} a_n \theta))^2] \to 0.$$

This implies condition (ix) in Theorem 2.1. $\Box$

Proof of Theorem 2.3. We apply Theorem 2.2 with $M_n = a_n^{-1} S_n^{1/2}$, $V_n = F_U'(0)$, $\eta_n = -a_n^{-1} \sum_{j=1}^n S_n^{-1/2} z_j \operatorname{sign}(U_j)$ and

$$G_n(\theta) = a_n^{-2} \sum_{j=1}^n |Y_j - z_j' \theta| = a_n^{-2} \sum_{j=1}^n |U_j - z_j'(\theta - \theta_0)|.$$

Conditions (i), (ii) and (v) in Theorem 2.2 are trivially satisfied.

We have that

$$\begin{aligned} & G_n(\theta_0 + M_n^{-1}\theta) - G_n(\theta_0) - \theta' \eta_n - \theta' V_n \theta \\ = \; & a_n^{-2} \textstyle\sum_{j=1}^n \Big(|U_j - a_n z_j' S_n^{-1/2} \theta| - |U_j| \\ & + a_n z_j' S_n^{-1/2} \theta \operatorname{sign}(U_j) - (a_n z_j' S_n^{-1/2} \theta)^2 F_U'(0) \Big), \\ = \; & \textstyle\sum_{j=1}^n (X_{n,j} - E[X_{n,j}]) \\ & + \textstyle\sum_{j=1}^n \Big(E[X_{n,j}] - (z_j' S_n^{-1/2} \theta)^2 F_U'(0) \Big), \end{aligned}$$

where

$$X_{n,j} := a_n^{-2} (|U_j - a_n z_j' S_n^{-1/2} \theta| - |U_j| + a_n z_j' S_n^{-1/2} \theta \operatorname{sign}(U_j)).$$

We claim that by Lemma 3.1 for each $\tau > 0$,

$$\lim_{n \to \infty} \epsilon_n \log(\Pr\{|\sum_{j=1}^n (X_{n,j} - E[X_{n,j}])| \ge \tau\}) = -\infty. \tag{3.8}$$

where $\epsilon_n = a_n^{-2}$. We have that

$$\epsilon_n^{-1} \max_{1 \le j \le n} |X_{n,j}| \le a_n \max_{1 \le j \le n} |z_j' S_n^{-1/2}| |\theta| \to 0.$$

So, conditions (i)–(iii) in Lemma 3.1 hold. We claim that

$$\mathrm{Var}(|U-\theta|-|U|+\theta\ \mathrm{sign}(U)) = O(|\theta|^3) \text{ as } \theta \to 0. \tag{3.9}$$

If $\theta > 0$, then $|U-\theta|-|U|+\theta\ \mathrm{sign}(U) = 2(\theta-U)I(0<U<\theta)$. So,

$$\begin{aligned} & \mathrm{Var}(|U-\theta|-|U|+\theta\ \mathrm{sign}(U)) \le E[(|U-\theta|-|U|+\theta\ \mathrm{sign}(U))^2] \\ \le\ & E[4(\theta-U)^2 I(0<U<\theta)] = \textstyle\int_0^\theta 4(\theta-u)^2\, dF_U(u) \\ =\ & \textstyle\int_0^\theta 8(\theta-u)(F_U(u)-F(0))\, du = O(\theta^3). \end{aligned}$$

The case $\theta < 0$ follows similarly. So, (3.9) follows. By (3.9),

$$\begin{aligned} & \epsilon_n^{-1} \textstyle\sum_{j=1}^n \mathrm{Var}(X_{n,j}) \\ \le\ & c a_n^{-2} \textstyle\sum_{j=1}^n |a_n z_j' S_n^{-1/2}\theta|^3 \\ \le\ & c a_n \max_{1\le j\le n} |z_j' S_n^{-1/2}||\theta|^3 \to 0, \end{aligned}$$

Therefore, (3.8) holds.

We claim that

$$E[|U-\theta|-|U|+\theta\ \mathrm{sign}(U)] - \theta^2 F_U'(0) = o(|\theta|^2) \quad \text{as } \theta \to 0. \tag{3.10}$$

If $\theta > 0$, we have that

$$\begin{aligned} & E[|U-\theta|-|U|+\theta\ \mathrm{sign}(U)] - \theta^2 F'(0) \\ =\ & 2\textstyle\int_0^\theta (\theta-u)\, dF_U(u) - \theta^2 F'(0) \\ =\ & 2\textstyle\int_0^\theta (F_U(u)-F_U(0))\, du - \theta^2 F_U'(0) \\ =\ & 2\textstyle\int_0^\theta (F_U(u)-F_U(0)-uF_U'(0))\, du = o(\theta^2). \end{aligned}$$

Thus, (3.10) follows if $\theta > 0$. The case $\theta < 0$ is similar. By (3.10),

$$|\sum_{j=1}^n (E[X_{n,j}] - (z_j' S_n^{-1/2}\theta)^2 F_U'(0))| \le o(1) \sum_{j=1}^n |z_j' S_n^{-1/2}|^2 \to 0, \tag{3.11}$$

(3.8) and (3.11) imply (iii) in Theorem 2.2.

By Lemma 3.1 in Arcones (2001), $a_n^{-1}\sum_{j=1}^n S_n^{-1/2} z_j\ \mathrm{sign}(U_j)$ satisfies the LDP with speed a_n^2 and rate function $I(t) = 2^{-1}|t|^2$. This implies hypothesis (iv) in Theorem 2.2. □

To prove Theorem 2.4 we will need the following lemma:

Lemma 3.3. *Under notation for the M–estimators in (1.2), let $\phi : S \to \mathbb{R}^d$ be a measurable function and let $\{a_n\}$ be a sequence of real numbers which converges to infinity such that $n^{-1/2}a_n \to 0$. Suppose that:*

(i) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|\hat{\theta}_n - \theta_0| \geq \tau\}) = -\infty.$$

(ii) For each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{\sum_{i=1}^n g(X_i, \hat{\theta}_n) \geq \inf_{\theta\in\Theta} \sum_{i=1}^n g(X_i, \theta) + \tau a_n^2\}) = -\infty.$$

(iii) There is a positive definite symmetric $d \times d$ matrix V such that

$$E[g(X, \theta) - g(X, \theta_0)] = (\theta - \theta_0)' V(\theta - \theta_0) + o(|\theta - \theta_0|^2),$$

as $\theta \to \theta_0$.

(iv) $E[|\phi(X)|^2] < \infty$

(v) $\lim_{n\to\infty} a_n^{-2} \log(n \Pr\{|\phi(X)| \geq n^{1/2}a_n\}) = -\infty$.

(vi) For each $\tau > 0$ and each $0 < M < \infty$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{ \sup_{|\theta|\leq M a_n n^{-1/2}} n a_n^{-2} |(P_n - P) r(\cdot, \theta)| \geq \tau\}) - \infty,$$

where

$$r(x, \theta) = g(x, \theta_0 + \theta) - g(x, \theta_0) - \theta' \phi(x).$$

(vii) For each $\tau > 0$, there exists a $\delta > 0$ such that

$$\lim_{M\to\infty} \limsup_{n\to\infty} a_n^{-2} \log(\Pr\{\sup_{|\theta-\theta_0|\leq\delta} \frac{n a_n^{-2} |(P_n - P)(g(\cdot, \theta) - g(\cdot, \theta_0))|}{\tau n a_n^{-2} |\theta - \theta_0|^2 + M} \geq 1\}) = -\infty.$$

Then, for each $\tau > 0$,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{|n^{1/2}(\hat{\theta}_n - \theta_0) + 2^{-1} n^{1/2}(P_n - P)V^{-1}\phi| \geq a_n \tau\}) = -\infty.$$

PROOF. Let $G(\theta) = E[g(X,\theta)]$ and let $G_n(\theta) = n^{-1}\sum_{j=1}^n g(X_j,\theta)$. First, we prove that

$$\lim_{M\to\infty}\limsup_{n\to\infty} a_n^{-2}\log(\Pr\{a_n^{-1}n^{1/2}|\hat\theta_n-\theta_0|\ge M\}) = -\infty. \tag{3.12}$$

By condition (iii), there are $0<c,\delta$ such that if $|\theta-\theta_0|\le\delta$, then $G(\theta)-G(\theta_0)\ge c|\theta-\theta_0|^2$. So, if $0<\tau<\delta$, $|\hat\theta_n-\theta_0|\le\delta$,

$$nG_n(\hat\theta_n)\le\inf_\theta nG_n(\theta)+a_n^2$$

and

$$\sup_{|\theta-\theta_0|\le\delta}\frac{na_n^{-2}|(P_n-P)(g(\cdot,\theta)-g(\cdot,\theta_0))|}{\tau na_n^{-2}|\theta-\theta_0|^2+M}\le 1,$$

then

$$\begin{aligned} & cna_n^{-2}|\hat\theta_n-\theta_0|^2\le na_n^{-2}(G(\hat\theta_n)-G(\theta_0))\\ = \;& na_n^{-2}(G_n(\hat\theta_n)-G_n(\theta_0))-na_n^{-2}(P_n-P)(g(\cdot,\hat\theta_n)-g(\cdot,\theta_0))\\ \le\;& 1+\tau a_n^{-2}n|\hat\theta_n-\theta_0|^2+M.\end{aligned}$$

Hence, (3.12) follows.

By (3.12), (iii) and (v), we have that for each $\tau>0$,

$$\begin{aligned}&\lim_{n\to\infty} a_n^{-2}\log(\Pr\{n|G_n(\hat\theta_n)-G_n(\theta_0)\\ &-n^{-1/2}(\hat\theta_n-\theta_0)'\eta_n-(\hat\theta_n-\theta_0)'V(\hat\theta_n-\theta_0)|\ge\tau a_n^2\})=-\infty\end{aligned}$$

where $\eta_n=n^{1/2}(P_n-P)\phi$. By conditions (iv)–(vi),

$$\begin{aligned}\lim_{n\to\infty}\log(\Pr\{\;& n|G_n(\theta_0-2^{-1}n^{-1/2}V^{-1}\eta_n)-G_n(\theta_0)\\ &+2^{-2}n^{-1}\eta_nV^{-1}\eta_n|\ge\tau a_n^2\})=-\infty.\end{aligned}$$

Now, if

$$n|G_n(\hat\theta_n)-G_n(\theta_0)-n^{-1/2}(\hat\theta_n-\theta_0)'\eta_n-(\hat\theta_n-\theta_0)'V(\hat\theta_n-\theta_0)|<\tau a_n^2$$

$$n|G_n(\theta_0+2^{-1}n^{-1/2}V^{-1}\eta_n)-G_n(\theta_0)+2^{-2}n^{-1}\eta_nV^{-1}\eta_n|<\tau a_n^2$$

and

$$G_n(\hat\theta_n)<\inf_\theta G_n(\theta)+\tau n^{-1}a_n^2$$

then

$$\begin{array}{rl} & n^{-1/2}(\hat{\theta}_n-\theta_0)'\eta_n+(\hat{\theta}_n-\theta_0)'V(\hat{\theta}_n-\theta_0) \\ < & n(G_n(\hat{\theta}_n)-G_n(\theta_0))+n^{-1}a_n^2\tau \\ < & n(G_n(\theta_0-2^{-1}n^{-1/2}V^{-1}\eta_n)-G_n(\theta_0))+2n^{-1}a_n^2\tau \\ < & -2^{-2}n^{-1}\eta_n'V^{-1}\eta_n+3n^{-1}a_n^2\tau. \end{array}$$

So,

$$\begin{array}{rl} & |V^{1/2}(\hat{\theta}_n-\theta_0)+2^{-1}n^{-1/2}V^{-1/2}\eta_n|^2 \\ = & n^{-1/2}(\hat{\theta}_n-\theta_0)'\eta_n+(\hat{\theta}_n-\theta_0)'V(\hat{\theta}_n-\theta_0)+2^{-2}n^{-1}\eta_n'V^{-1}\eta_n \\ < & 3n^{-1}a_n^2\tau. \end{array}$$

We have proved that

$$\begin{array}{rl} & \{|V^{1/2}(\hat{\theta}_n-\theta_0)+2^{-1}n^{-1/2}V^{-1/2}\eta_n|^2\geq 3n^{-1}a_n^2\tau\} \\ = & \{n|G_n(\hat{\theta}_n)-G_n(\theta_0)-n^{-1/2}(\hat{\theta}_n-\theta_0)'\eta_n- \\ & \qquad (\hat{\theta}_n-\theta_0)'V(\hat{\theta}_n-\theta_0)|\geq \tau a_n^2\} \\ & \cup\{n|G_n(\theta_0+2^{-1}n^{-1/2}V^{-1}\eta_n)-G_n(\theta_0)+2^{-2}n^{-1}\eta_nV^{-1}\eta_n|\geq\tau a_n^2\} \\ & \cup\{G_n(\hat{\theta}_n)\geq\inf_\theta G_n(\theta)+\tau a_n^2n^{-1}\}, \end{array}$$

which implies the claim. □

PROOF OF THEOREM 2.4. We apply Lemma 3.3. We have that conditions (i)–(v) in the theorem hold. We have that

$$\sup_{|\theta|\leq Ma_nn^{-1/2}} na_n^{-2}|(P_n-P)r(\cdot,\theta)|\leq M^2|(P_n-P)\Psi|+M^2(P_n+P)B_{Ma_nn^{-1/2}}.$$

By conditions (vii) and (viii), we have that

$$\lim_{n\to\infty}a_n^{-2}\log(\Pr\{M^2(P_n+P)B_{Ma_nn^{-1/2}}\geq\tau\})=-\infty,$$

which implies condition (vi) in Lemma 3.3.

We also have that

$$\begin{array}{rl} & \sup_{|\theta-\theta_0|\leq\delta}\frac{na_n^{-2}|(P_n-P)(g(\cdot,\theta)-g(\cdot,\theta_0))|}{\tau na_n^{-2}|\theta-\theta_0|^2+M} \\ \leq & \sup_{|\theta-\theta_0|\leq\delta}\frac{na_n^{-2}|(P_n-P)(\theta-\theta_0)'\phi|}{\tau na_n^{-2}|\theta-\theta_0|^2+M}+\sup_{|\theta-\theta_0|\leq\delta}\frac{na_n^{-2}|\theta-\theta_0|^2|(P_n-P)\Psi|}{\tau na_n^{-2}|\theta-\theta_0|^2+M} \\ & +\sup_{|\theta-\theta_0|\leq\delta}\frac{na_n^{-2}(P_n+P)|\theta-\theta_0|^2B_\delta}{\tau na_n^{-2}|\theta-\theta_0|^2+M} \\ \leq & 2^{-1}\tau^{-1/2}M^{-1/2}n^{1/2}a_n^{-1}|(P_n-P)\phi|+\tau^{-1}|(P_n-P)\Psi| \\ & +\tau^{-1}(P_n+P)B_\delta, \end{array}$$

which implies condition (vii) in Lemma 3.3. □

Proof of Theorem 2.5. We apply Theorem 2.4. We only need to check hypothesis (i) in Theorem 2.4. Take $\tau > 0$ small enough such $\delta := \inf_{|\theta-\theta_0|=\tau} E[g(X,\theta) - g(X,\theta_0)] > 0$. By (i)

$$\begin{aligned} &\{|\hat\theta_n - \theta_0| \geq \tau\} \\ \subset\ &\{\inf_{|\theta-\theta_0|=\tau} G_n(\theta) < G_n(\theta_0) + 2^{-1}\delta\} \\ \subset\ &\{\sup_{|\theta-\theta_0|=\tau} |(P_n - P)(g(\cdot,\theta) - g(\cdot,\theta_0))| \geq 2^{-1}\delta\}. \end{aligned}$$

Since condition (vii) in Lemma 3.3 is satisfied,

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{\sup_{|\theta-\theta_0|=\tau} |(P_n - P)(g(\cdot,\theta) - g(\cdot,\theta_0))| \geq 2^{-1}\delta\}) = -\infty.$$

□

To prove Theorem 2.7 we need the following lemma:

Lemma 3.4. *Let $\{Z_n(\theta) : \theta \in I\!\!R\}$ be a sequence of stochastic processes. Let $a > 0$. Let $\theta_0 \in I\!\!R$. Let $\{\epsilon_n\}$ and let $\{b_n\}$ be two sequences of positive numbers which converge to zero. Let $\hat\theta_n = \sup\{t : Z_n(t) \leq 0\}$. Assume that:*

(i) As a function on θ, $Z_n(\theta)$ is non increasing.

(ii) For each $t \in I\!\!R$,

$$\lim_{n\to\infty} b_n E[Z_n(\theta_0 + b_n^{-1}t) - Z_n(\theta_0)] = at.$$

(iii)

$$\lim_{M\to\infty} \limsup_{n\to\infty} \epsilon_n \log(\Pr\{b_n|Z_n(\theta_0)| \geq M\}) = -\infty.$$

(iv) For each $t \in I\!\!R$ and each $\tau > 0$,

$$\begin{aligned} \lim_{n\to\infty} \epsilon_n \log(\Pr\{|b_n(Z_n(\theta_0 + b_n^{-1}t) - Z_n(\theta_0) \\ - E[Z_n(\theta_0 + b_n^{-1}t) - Z(\theta_0)])| \geq \tau\}) = -\infty. \end{aligned}$$

Then, for each $\tau > 0$,

$$\lim_{n\to\infty} \epsilon_n \log(\Pr\{|b_n(\hat\theta_n - \theta_0) + a^{-1}b_n Z_n(\theta_0)| \geq \tau\}) = -\infty.$$

PROOF. By conditions (iii) and (v), for each $t \in I\!\!R$ and each $\tau > 0$,

$$\lim_{n\to\infty} \epsilon_n \log(\Pr\{|V_n(t)| \ge \tau\}) = -\infty$$

where

$$V_n(t) = b_n(Z_n(\theta_0 + b_n^{-1}t) - Z_n(\theta_0)) - at.$$

We claim that for each $0 < M < \infty$ and each $\tau > 0$,

$$\lim_{n\to\infty} \epsilon_n \log(\Pr\{\sup_{|t|\le M} |V_n(t)| \ge \tau\}) = -\infty. \tag{3.13}$$

Take an integer $m \ge 4Ma\tau^{-1}$. Let $t_j = -M + m^{-1}j2M$, $0 \le j \le m$. If $t_{j-1} \le t \le t_j$, then

$$\begin{aligned} &V_n(t) \le b_n(Z_n(\theta_0 + b_n^{-1}t_j) - Z_n(\theta_0)) - at_{j-1} \\ &\le \max_{0\le j\le m} |V_n(t_j)| + am^{-1}2M \le \max_{0\le j\le m} |V_n(t_j)| + 2^{-1}\tau. \end{aligned}$$

Similarly we get that

$$V_n(t) \ge -\max_{0\le j\le m} |V_n(t_j)| - 2^{-1}\tau.$$

Hence, $\sup_{|t|\le M} |V_n(t)| \le \max_{0\le j\le m} |V_n(t_j)| + 2^{-1}\tau$ and (3.13) follows.

By (3.13), for each $t \in I\!\!R$ and each $\tau > 0$,

$$\lim_{n\to\infty} \epsilon_n \log(\Pr\{|V_n(-a^{-1}(t + b_nZ_n(\theta_0)))| \ge \tau\}) = -\infty.$$

Now,

$$\begin{aligned} &V_n(-a^{-1}(t + b_nZ_n(\theta_0))) \\ = \; &b_nZ_n(\theta_0 - b_n^{-1}a^{-1}(t + b_nZ_n(\theta_0))) + t \end{aligned}$$

Hence, we have that for each $t \in I\!\!R$ and each $\tau > 0$,

$$\lim_{n\to\infty} \epsilon_n \log(\Pr\{|b_nZ_n(\theta_0 - b_n^{-1}a^{-1}(t + b_nZ_n(\theta_0))) + t| \ge \tau\}) = -\infty. \tag{3.14}$$

To prove the lemma it suffices to show that for each $\tau > 0$,

$$\lim_{n\to\infty} \epsilon_n \log(\Pr\{ab_n(\hat\theta_n - \theta_0) + b_nZ_n(\theta_0) \le -\tau\}) = -\infty \tag{3.15}$$

and

$$\lim_{n\to\infty} \epsilon_n \log(\Pr\{ab_n(\hat\theta_n - \theta_0) + b_nZ_n(\theta_0) \ge \tau\}) = -\infty. \tag{3.16}$$

Given t, we have that $\{\hat{\theta}_n < t\} \subset \{Z_n(t) > 0\}$. So,

$$\begin{aligned} & \Pr\{ab_n(\hat{\theta}_n - \theta_0) + b_n Z_n(\theta_0) < -\tau\} \\ = \ & \Pr\{\hat{\theta}_n < \theta_0 - b_n^{-1}a^{-1}(\tau + b_n Z_n(\theta_0))\} \\ \le \ & \Pr\{Z_n(\theta_0 - b_n^{-1}a^{-1}(\tau + b_n Z_n(\theta_0))) > 0\}. \end{aligned}$$

This and (3.14) imply (3.15).

Finally, given t, we have that $\{Z_n(t) > 0\} \subset \{\hat{\theta}_n \le t\}$. So,

$$\begin{aligned} & \Pr\{ab_n(\hat{\theta}_n - \theta_0) + b_n Z_n(\theta_0) > \tau\} \\ \le \ & \Pr\{Z_n(\theta_0 - b_n^{-1}a^{-1}(-\tau + b_n Z_n(\theta_0))) \le 0\}. \end{aligned}$$

As before, this and (3.14) implies (3.16). □

Proof of Theorem 2.7. We apply Lemma 3.4 with $Z_n(\theta) = n^{-1}\sum_{j=1}^n h(X_j, \theta)$, $b_n = a_n^{-1}n^{1/2}$ and $\epsilon_n = a_n^{-2}$. Hypothesis (i) in Lemma 3.4 is assumed. Hypothesis (ii) in Lemma 3.4 follows from (i). By Corollary 3.4 in Arcones (2001), $a_n^{-1}n^{-1/2}\sum_{j=1}^n h(X_j, \theta_0)$ satisfies the LDP with speed a_n^2. This implies hypothesis (iii) in Lemma 3.4. To check hypothesis (iv), we need to prove that for each $\tau > 0$,

$$\begin{aligned} & \lim_{n\to\infty} a_n^{-2} \log(\Pr\{|\textstyle\sum_{i=1}^n (r(X_j, n^{-1/2}a_n t) \\ & -E[r(X_j, n^{-1/2}a_n t)])| \ge \tau a_n n^{1/2}\}) = -\infty. \end{aligned}$$

To check this, we use Lemma 3.1 with $\epsilon_n = a_n^{-2}$ and $X_{n,j} = n^{-1/2}a_n^{-1}r(X_j, n^{-1/2}a_n t)$. □

Proof of Theorem 2.8. By (iii) and (iv), there are $c, > 0$ and $\delta_0 > \delta_1 > 0$ such that if $|\theta - \theta_0| \le \delta_1$, then

$$c|\theta - \theta_0| \le |H(\theta)|.$$

If $|\hat{\theta}_n - \theta_0| \le \delta_1$,

$$\sup_{|\theta-\theta_0|\le\delta_1} a_n^{-1}n^{1/2}|(P_n - P)(h(\cdot, \theta) - h(\cdot, \theta_0))| \le M,$$

$$|a_n^{-1}n^{1/2}(P_n - P)h(\cdot, \theta_0)| \le M$$

and

$$a_n^{-1}n^{1/2}|P_n h(\cdot, \hat{\theta}_n)| \le M$$

then

$$
\begin{aligned}
& c a_n^{-1} n^{1/2} |\hat{\theta}_n - \theta_0| \le |H(\hat{\theta}_n) - H(\theta_0)| \\
\le \; & a_n^{-1} n^{1/2} |(P_n - P)(h(\cdot, \hat{\theta}_n) - h(\cdot, \theta_0))| + a_n^{-1} n^{1/2} |(P_n - P) h(\cdot, \theta_0)| \\
& + a_n^{-1} n^{1/2} |P_n h(\cdot, \hat{\theta}_n)| \le 3M.
\end{aligned}
$$

The previous estimation and conditions (i)–(v) imply that

$$
\lim_{M \to \infty} \limsup_{n \to \infty} a_n^{-2} \log(\Pr\{a_n^{-1} n^{1/2} |\hat{\theta}_n - \theta_0| \ge M\}) = -\infty. \tag{3.17}
$$

Note that condition (v) implies

$$
\begin{aligned}
\lim\nolimits_{M \to \infty} \limsup\nolimits_{n \to \infty} a_n^{-2} \log(\Pr\{\sup\nolimits_{|\theta - \theta_0| \le \delta_0} a_n^{-1} n^{1/2} & \\
\times |(P_n - P)(h(\cdot, \theta) - h(\cdot, \theta_0))| \ge M\}) = -\infty. &
\end{aligned}
$$

We also have that

$$
\begin{aligned}
& a_n^{-1} n^{1/2} |H'(\theta_0)(\hat{\theta}_n - \theta_0) + (P_n - P) h(\cdot, \theta_0)| \\
\le \; & a_n^{-1} n^{1/2} |(P_n - P)(h(\cdot, \theta_0) - h(\cdot, \hat{\theta}_n))| + a_n^{-1} n^{1/2} |P_n h(\cdot, \hat{\theta}_n)| \\
& + a_n^{-1} n^{1/2} |H(\hat{\theta}_n) - H(\theta_0) - H'(\theta_0)(\hat{\theta}_n - \theta_0)|.
\end{aligned}
$$

The previous inequality, (3.17) and conditions (ii) and (iv)–(vi) imply the claim of the theorem. Observe that conditions (v) and (vi) imply that for each $\tau > 0$,

$$
\begin{aligned}
\lim\nolimits_{\delta \to \infty} \limsup\nolimits_{n \to \infty} a_n^{-2} \log(\Pr\{\sup\nolimits_{|\theta - \theta_0| \le \delta} a_n^{-1} n^{1/2} & \\
\times |(P_n - P)(h(\cdot, \theta) - h(\cdot, \theta_0))| \ge \tau\}) = -\infty. &
\end{aligned}
$$

□

Proof of Theorem 2.9. We apply Corollary 2.2. Let $g(x, \theta) = \log p(x, \theta)$ and let

$$
h(x, \theta) = \left(\frac{\partial \log p(x|\theta)}{\partial \theta^{(1)}}, \dots, \frac{\partial \log p(x|\theta)}{\partial \theta^{(d)}} \right)'.
$$

Conditions (ii), (v)–(vii) and (ix) in Corollary 2.2 is obviously satisfied.

By the Jensen inequality for each $\theta \in \Theta$,

$$
E\left[\log \left(\frac{p(X|\theta)}{p(X|\theta_0)} \right) \right] \le \log E\left[\frac{p(X|\theta)}{p(X|\theta_0)} \right] = 0.
$$

So, each $\theta \in \Theta$,

$$E[\log p(X|\theta)] \le E[\log p(X|\theta_0)].$$

By condition (vi), $E[\log p(X|\theta)]$ is twice differentiable under the integral sign with zero first derivatives and matrix of second derivatives V. This implies conditions (iii) and (iv) in Corollary 2.2.

Take $\tau > 0$ small enough such $\delta := \inf_{|\theta-\theta_0|=\tau} E[g(X,\theta) - g(X,\theta_0)] > 0$. We have that

$$\begin{aligned} & \Pr\{|\hat{\theta}_n - \theta_0| \ge \tau\} \\ \le \;& \Pr\{\sup_{|\theta-\theta_0|=\tau} |(P_n - P)(g(\cdot,\theta) - g(\cdot,\theta_0))| \ge 2^{-1}\delta\}. \end{aligned}$$

By Theorem 2.7 in Arcones (2001), $\{(P_n - P)(g(\cdot,\theta) - g(\cdot,\theta_0)) : |\theta - \theta_0| = \tau\}$ satisfies the LPD with speed n. This implies that

$$\lim_{n\to\infty} a_n^{-2} \log(\Pr\{\sup_{|\theta-\theta_0|=\tau} |(P_n - P)(g(\cdot,\theta) - g(\cdot,\theta_0))| \ge 2^{-1}\delta\}) = -\infty.$$

So, condition (i) in Corollary 2.2 follows.

To check condition (viii) in Corollary 2.8, it suffices to show that

$$\{n^{-1/2} \sum_{j=1}^{n} (h(X_j,\theta) - E[h(X_j,\theta)]) : |\theta - \theta_0| \le \delta_0\}$$

converges weakly. This follows from the central limit theorem for empirical processes under bracketing conditions in Ossiander (1987). Observe that for each $|\theta - \theta_0| \le \delta_0$ and each $\eta > 0$,

$$\sup_{\theta' : |\theta'-\theta_0| \le \delta_0, |\theta'-\theta| \le \eta} |h(X,\theta') - h(X,\theta)| \le \eta B_{\delta_0}(X),$$

where

$$B_{\delta_0}(x) = \sup_{1 \le i,j \le d} \sup_{|\theta-\theta_0| < \delta_0} \left| \frac{\partial^2 \log p(x|\theta)}{\partial \theta^{(i)} \partial \theta^{(j)}} \right|.$$

□

References

Arcones, M. A. (1998a). Weak convergence of convex stochastic processes. *Statistics & Probability Letters*, **37**, 171–182.

Arcones, M. A. (1998b). The large deviation principle of stochastic processes. Preprint.

Arcones, M. A. (2001). The large deviation principle of empirical processes. Preprint.

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

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

Gao, F. (2001). Moderate deviations for the maximum likelihood estimator. *Statistics & Probability Letters*, **55**, 345–352.

Giné, E. and Zinn, J. (1984). Some limit theorems for empirical processes. *Annals of Probability*, **12**, 929–989.

Haldane, S. J. (1948). Note on the median of a multivariate distribution. *Biometrika*, **35**, 414–415.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383–393.

Hampel, F. R.; Ronchetti, E. M.; Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics, the Approach Based on Influence Functions.* Wiley, New York.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, **35**, 73–101.

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.

Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation.* Second edition. Springer Texts in Statistics. Springer–Verlag, New York.

Mäkeläinen, T.; Schmidt, K. and Styan, G. P. H. (1981). On the existence and uniqueness of the maximum likelihood estimate of a vector–valued parameter in fixed–size samples. *Annals of Statistics*, **9**, 758–767.

Milasevic, P. and Ducharme, G. R. (1987). Uniqueness of the spatial median. *Annals of Statistics*, **15**, 1332–1333.

Ossiander, M. (1987). A central limit theorem under entropy with L_2 bracketing. *Annals of Probability* **15** 897–919.

Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: computability of squared–error versus absolute–error estimators. *Statistical Science*, **12**, 279–300.

Prokhorov, Yu. V. (1959). An extremal problem in probability theory. *Theory of Probability and its Applications*, **4**, 211–214.

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

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and Empirical Processes with Applications to Statistics.* Springer–Verlag, New York.