The Annals of Statistics
2011, Vol. 39, No. 2, 1211–1240
DOI: 10.1214/10-AOS865

DELTA METHOD IN LARGE DEVIATIONS AND MODERATE DEVIATIONS FOR ESTIMATORS

BY FUQING GAO[1] AND XINGQIU ZHAO[2]

Wuhan University and Hong Kong Polytechnic University and Zhongnan University of Economics and Law

The delta method is a popular and elementary tool for deriving limiting distributions of transformed statistics, while applications of asymptotic distributions do not allow one to obtain desirable accuracy of approximation for tail probabilities. The large and moderate deviation theory can achieve this goal. Motivated by the delta method in weak convergence, a general delta method in large deviations is proposed. The new method can be widely applied to driving the moderate deviations of estimators and is illustrated by examples including the Wilcoxon statistic, the Kaplan–Meier estimator, the empirical quantile processes and the empirical copula function. We also improve the existing moderate deviations results for M-estimators and L-statistics by the new method. Some applications of moderate deviations to statistical hypothesis testing are provided.

1. Introduction. Consider a family of random variables $\{Y_n, n \geq 1\}$ such as the sample mean. Assume that it satisfies a law of large numbers and a fluctuation theorem such as central limit theorem, that is, $Y_n \to \theta$ in law and there exists a sequence $b_n \to \infty$ such that $b_n(Y_n - \theta) \to Y$ in law, where θ is a constant and Y is a nontrivial random variable. A large deviation result is concerned with estimation of large deviation probabilities $P(|Y_n - \theta| \geq \varepsilon)$ for $\varepsilon > 0$. A moderate deviation result is concerned with estimation of large deviation probabilities $P(r_n|Y_n - \theta| \geq \varepsilon)$ for $\varepsilon > 0$, where r_n is an intermediate scale between 1 and b_n, that is, $r_n \to \infty$ and $b_n/r_n \to \infty$. In particular, if $b_n = \sqrt{n}$, then $r_n = n^{1/2-\delta}$ with $0 < \delta < 1$.

The large deviation and moderate deviation problems arise in the theory of statistical inference quite naturally. For estimation of unknown parameters and functions, it is first of all important to minimize the risk of wrong decisions implied by deviations of the observed values of estimators from the true values of parameters or functions to be estimated. Such gross errors are precisely the subject of large deviation theory. The large deviation and moderate deviation results of estimators can

Received April 2010; revised August 2010.

[1]Supported in part by the National Natural Science Foundation of China (NSFC) Grant 10871153 and the Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) of China Grant 200804860048.

[2]Supported in part by the Research Grant Council of Hong Kong (PolyU 5032/09P).

MSC2010 subject classifications. Primary 60F10, 62G20; secondary 62F12.

Key words and phrases. Delta method, hypothesis testing, Kaplan–Meier estimator, large deviations, L-statistics, M-estimator, moderate deviations.

provide us with the rates of convergence and a useful method for constructing asymptotic confidence intervals. For the classical large deviation theory with the empirical measures and sample means, one can refer to Sanov (1957), Groeneboom, Oosterhoff and Ruymgaart (1979) and Bahadur and Zabell (1979). The large deviations for linear combinations of order statistics (L-estimators) were also investigated in Groeneboom, Oosterhoff and Ruymgaart (1979). Bahadur and Zabell (1979) developed a subadditive method to study the large deviations for general sample means. For some developments of large deviations and moderate deviations in statistics, see Fu (1982), Kester and Kallenberg (1986), Sieders and Dzhaparidze (1987), Inglot and Ledwina (1990), Borovkov and Mogul′skii (1992), Puhalskii and Spokoiny (1998), Bercu (2001), Joutard (2004) and Arcones (2006) for large deviations of estimators; Kallenberg (1983), Gao (2001), Arcones (2002), Inglot and Kallenberg (2003), Djellout, Guillin and Wu (2006) and Ermakov (2008) for moderate deviations of estimators; Louani (1998), Worms (2001), Gao (2003), Lei and Wu (2005) for large deviations and moderate deviations of kernel density estimators, and references therein. On the other hand, large deviations of estimators can be applied to Bahadur efficiency to determine the Bahadur slope [Bahadur (1967), Nikitin (1995), He and Shao (1996)] and hypothesis testing [see Dembo and Zeitouni (1998), Sections 3.5 and 7.1].

In statistics, many important estimators are functionals $\Phi(L_n)$ of the empirical processes L_n, and so deriving limiting distribution of $r_n(\Phi(L_n) - \Phi(\mu))$ from limiting distribution of $r_n(L_n - \mu)$ is a fundamental problem, where r_n is a sequence of positive numbers and μ is the mean of L_n. It is well known that the delta method is a popular and elementary tool for solving the problem. The method tells us that the weak convergence of $r_n(X_n - \theta)$ yields the weak convergence of $r_n(\Phi(X_n) - \Phi(\theta))$ if Φ is Hadamard differentiable (see Section 3), where X_n is a sequence of random variables, θ is a constant and $r_n \to \infty$. For some developments and applications of the delta method, one can refer to Gill (1989), Kosorok (2008), Reeds (1976), and van der Vaart and Wellner (1996) among others. For example, Reeds (1976) systematically developed the use of Hadamard instead of Fréchet differentiability to derive asymptotic distributions of transformed processes. Andersen et al. (1993) also described some applications of the delta method in survival analysis. More recently, van der Vaart and Wellner (1996) and Kosorok (2008) provided an excellent summary of the functional delta method in terms of a weak convergence.

A natural problem is whether the large deviations of $r_n(\Phi(X_n) - \Phi(\theta))$ can be obtained from the large deviations of $r_n(X_n - \theta)$ if the function Φ defined on a set $\mathcal{D}_\Phi$ is Hadamard differentiable. When $r_n = r$ for all n with a constant r, the problem can be solved by the contraction principle [see Dembo and Zeitouni (1998)]. When $r_n \to \infty$, for each $n \geq 1$, define $\mathcal{D}_n = \{h; \theta + h/r_n \in \mathcal{D}_\Phi\}$ and $f_n(h) = r_n(\Phi(\theta + h/r_n) - \Phi(\theta))$ for all $h \in \mathcal{D}_n$. Then by Hadamard differentiability, for every sequence $h_n \in \mathcal{D}_n$ converging to h, the sequence f_n satisfies

$f_n(h_n) \to \Phi'_\theta(h)$. Note that $f_n(r_n(X_n - \theta)) = r_n(\Phi(X_n) - \Phi(\theta))$. Motivated by this, we can also consider to use a contraction principle for establishing the large deviations of $r_n(\Phi(X_n) - \Phi(\theta))$. However, the existing contraction principles cannot be applicable to these situations as addressed in Remark 2.1 of next section. For this reason, we need to extend the contraction principle in large deviations.

The objective of this paper is to develop a general delta method in large deviations similar to that in week convergence and applies the method to solve some moderate deviation problems in statistics. The remainder of the paper is organized as follows. In Section 2, we present an extended contraction principle, while its proof will be given in the Appendix. Then a general delta method in large deviations is established by using the extended contraction principle in Section 3. In Section 4, we apply the proposed delta method in large deviations to some statistical models including censored data, empirical quantile process, copula function, M-estimators and L-statistics. The moderate deviation principles for the Wilcoxon statistic, the Kaplan–Meier estimator, the empirical quantile estimator and the empirical copula estimator are established. We also improve the existing moderate deviation results for M-estimators and L-statistics in Section 4, where our proofs are different from others but more simple by the new method. Section 5 presents some applications of the moderate deviation results to statistical hypothesis testing. Some concluding remarks are made in Section 6.

2. An extended contraction principle. As explained in previous section, to establish a delta method in large deviation, we first need to generalize the contraction principle in large deviation theory. In this section, we present an extension of the contraction principle which plays an important role.

First, let us introduce some notation in large deviations [Dembo and Zeitouni (1998), Deuschel and Stroock (1989)]. For a metric space $\mathcal{X}$, $\mathcal{B}(\mathcal{X})$ is the Borel σ-algebra of $\mathcal{X}$. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let T be an arbitrary map from Ω to $\overline{\mathbb{R}}$, where $\overline{\mathbb{R}} = [-\infty, \infty]$ is the space of extended real numbers. The outer integral of T with respect to P is defined by van der Vaart and Wellner (1996)

$$E^*(T) = \inf\{E(U); U \geq T, U : \Omega \mapsto \overline{\mathbb{R}} \text{ measurable and } E(U) \text{ exists}\}.$$

The outer probability of an arbitrary subset B of Ω is

$$P^*(B) = \inf\{P(A); A \supset B, A \in \mathcal{F}\}.$$

Inner integral and inner probability are defined by

$$E_*(T) = -E^*(-T) \quad \text{and} \quad P_*(B) = 1 - P^*(\Omega \setminus B),$$

respectively.

Let $\{(\Omega_n, \mathcal{F}_n, P_n), n \geq 1\}$ be a sequence of probability spaces and let $\{X_n, n \geq 1\}$ be a sequence of maps from Ω_n to $\mathcal{X}$. Let $\{\lambda(n), n \geq 1\}$ be a sequence of positive numbers tending to $+\infty$ and let $I : \mathcal{X} \to [0, +\infty]$ be inf-compact; that is, $[I \le L]$ is compact for any $L \in \mathbb{R}$. Then $\{X_n, n \ge 1\}$ is said to satisfy the lower bound of large deviation (LLD) with speed $\lambda(n)$ and rate function I, if for any open measurable subset G of $\mathcal{X}$,

$$l(G) := \liminf_{n\to\infty} \frac{1}{\lambda(n)} \log P_{n*}(X_n \in G) \ge -\inf_{x\in G} I(x). \tag{2.1}$$

$\{X_n, n \ge 1\}$ is said to satisfy the upper bound of large deviation (ULD) with speed $\lambda(n)$ and rate function I, if for any closed measurable subset F of $\mathcal{X}$,

$$\mathcal{U}(F) := \limsup_{n\to\infty} \frac{1}{\lambda(n)} \log P_n{}^*(X_n \in F) \le -\inf_{x\in F} I(x). \tag{2.2}$$

We say that $\{X_n, n \ge 1\}$ satisfies the large deviation principle (LDP) with speed $\lambda(n)$ and rate function I, if both LLD and ULD hold.

Now, we present the extended contraction principle.

THEOREM 2.1 (Extended contraction principle). *Let $(\mathcal{X}, d)$ and $(\mathcal{Y}, \rho)$ be two metric spaces. Let $\{\mathcal{D}_n, n \ge 1\}$ be a sequences of subsets in $(\mathcal{X}, d)$, and let $\{f_n : \mathcal{D}_n \mapsto \mathcal{Y}; n \in \mathbb{N}\}$ be a family of mappings. Also for each $n \ge 1$, let X_n be a map from probability space $(\Omega_n, \mathcal{F}_n, P_n)$ to $\mathcal{D}_n$. Suppose that*:

(i) *$\{X_n, n \ge 1\}$ satisfies the large deviation principle with speed $\lambda(n)$ and rate function I*;

(ii) *there exists a mapping $f : \{I < \infty\} \mapsto \mathcal{Y}$ such that if for a sequence $\{x_n \in \mathcal{D}_n, n \ge 1\}$, $x_n \to x \in \{I < \infty\}$ as $n \to \infty$, then $f_n(x_n) \to f(x)$ as $n \to \infty$.*

Then $\{f_n(X_n), n \ge 1\}$ satisfies the large deviation principle with speed $\lambda(n)$ and rate function I_f, where

$$I_f(y) = \inf\{I(x); f(x) = y\}, \qquad y \in \mathcal{Y}. \tag{2.3}$$

The proof of the theorem is given in the Appendix.

REMARK 2.1. (1) If $\mathcal{D}_n = \mathcal{X}$ for all $n \ge 1$, then Theorem 2.1 yields Theorem 2.1 in Arcones (2003b). Another popular contraction principle was given in Theorem 4.3.23 of Dembo and Zeitouni (1998), in which $\mathcal{D}_n = \mathcal{X}$ for all $n \ge 1$, f_n is continuous for all $n \ge 1$ and for any $L \in (0, \infty)$,

$$\lim_{n\to\infty} \sup_{x : I(x) \le L} \rho(f_n(x), f(x)) = 0. \tag{2.4}$$

This condition cannot be compared to condition (ii) in Theorem 2.1.

(2) It is necessary for proving Theorem 3.1 to introduce the sequence of subsets $\mathcal{D}_n$ in Theorem 2.1, because subsets $\{h \in \mathcal{X}; \theta + h/r_n \in \mathcal{D}_\Phi\}, n \ge 1$ are not equal,

generally, for $\theta \in \mathcal{X}$ and a subset $\mathcal{D}_\Phi$ of a topological linear spaces $\mathcal{X}$. In fact, $\mathcal{D}_\Phi$ is usually a subset of $\mathcal{X}$ in applications (see Section 4).

3. Delta method in large deviations. In this section, we establish a delta method in large deviations by using the extended contraction principle presented in Section 2.

Let us first recall some conceptions of Hadamard differentiability [Gill (1989), van der Vaart and Wellner (1996), Kosorok (2008), Römisch (2005)]. Let $\mathcal{X}$ and $\mathcal{Y}$ be two metrizable topological linear spaces. A map Φ defined on a subset $\mathcal{D}_\Phi$ of $\mathcal{X}$ with values in $\mathcal{Y}$ is called Hadamard differentiable at x if there exists a continuous mapping $\Phi'_x : \mathcal{X} \mapsto \mathcal{Y}$ such that

$$\lim_{n\to\infty} \frac{\Phi(x + t_n h_n) - \Phi(x)}{t_n} = \Phi'_x(h) \tag{3.1}$$

holds for all sequences t_n converging to $0+$ and h_n converging to h in $\mathcal{X}$ such that $x + t_n h_n \in \mathcal{D}_\Phi$ for every n.

REMARK 3.1. Linearity of the Hadamard directional derivative $\Phi'_x(\cdot)$ is not required. In fact, $\Phi'_x(\cdot)$ is often not linear if Φ is given by inequality constraints. However, by the definition, we can see that $\Phi'_x(\cdot)$ is positively homogenous; that is, $\Phi'_x(th) = t\Phi'_x(h)$ for all $t \geq 0$ and $h \in \mathcal{X}$.

The definition of the Hadamard differentiable may be refined to Hadamard differentiable tangentially to a set $\mathcal{D}_0 \subset \mathcal{X}$. For a subset $\mathcal{D}_0$ of $\mathcal{X}$, the map Φ is said to be Hadamard differentiable at $x \in \mathcal{D}_\Phi$ tangentially to $\mathcal{D}_0$ if the limit (3.1) exists for all sequences t_n converging to $0+$ and h_n converging to h in $\mathcal{D}_0$ such that $x + t_n h_n \in \mathcal{D}_\Phi$ for every n. In this case, the Hadamard derivative $\Phi'_x(\cdot)$ is a continuous mapping on $\mathcal{D}_0$. If $\mathcal{D}_0$ is a cone, then $\Phi'_x(\cdot)$ is again positively homogenous.

THEOREM 3.1 (Delta method in large deviation). *Let $\mathcal{X}$ and $\mathcal{Y}$ be two metrizable linear topological spaces and let d and ρ be compatible metrics on $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let $\Phi : \mathcal{D}_\Phi \subset \mathcal{X} \mapsto \mathcal{Y}$ be Hadamard-differentiable at θ tangentially to $\mathcal{D}_0$, where $\mathcal{D}_\Phi$ and $\mathcal{D}_0$ are two subsets of $\mathcal{X}$. Let $X_n : \Omega_n \mapsto \mathcal{D}_\Phi, n \geq 1$ be a sequence of maps and let $r_n, n \geq 1$, be a sequence of positive real numbers satisfying $r_n \to +\infty$.*

If $\{r_n(X_n - \theta), n \geq 1\}$ satisfies the large deviation principle with speed $\lambda(n)$ and rate function I and $\{I < \infty\} \subset \mathcal{D}_0$, then $\{r_n(\Phi(X_n) - \Phi(\theta)), n \geq 1\}$ satisfies the large deviation principle with speed $\lambda(n)$ and rate function $I_{\Phi'_\theta}$, where

$$I_{\Phi'_\theta}(y) = \inf\{I(x); \Phi'_\theta(x) = y\}, \qquad y \in \mathcal{Y}. \tag{3.2}$$

Furthermore, if Φ'_θ *is defined and continuous on the whole space of* $\mathcal{X}$, *then* $\{r_n(\Phi(X_n) - \Phi(\theta)) - \Phi'_\theta(r_n(X_n - \theta)), n \geq 1\}$ *satisfies the large deviation principle with speed* $\lambda(n)$ *and rate function*

$$I_{\Phi,\theta}(z) = \begin{cases} 0, & z = 0, \\ +\infty, & \textit{otherwise}. \end{cases} \tag{3.3}$$

In particular, for any $\delta > 0$,

$$\begin{aligned} &\limsup_{n\to\infty} \frac{1}{\lambda(n)} \log P_n^*\bigl(\rho\bigl(r_n\bigl(\Phi(X_n) - \Phi(\theta)\bigr) - \Phi'_\theta\bigl(r_n(X_n - \theta)\bigr), 0\bigr) \geq \delta\bigr) \\ &\quad = -\infty. \end{aligned} \tag{3.4}$$

PROOF. For each $n \geq 1$, define $\mathcal{D}_n = \{h \in \mathcal{X}; \theta + h/r_n \in \mathcal{D}_\Phi\}$ and

$$f_n : \mathcal{D}_n \mapsto \mathcal{Y}, \qquad f_n(h) = r_n\bigl(\Phi(\theta + h/r_n) - \Phi(\theta)\bigr) \qquad \text{for all } h \in \mathcal{D}_n.$$

Then for every sequence $h_n \in \mathcal{D}_n$ converging to $h \in \mathcal{D}_0$, the sequence f_n satisfies $f_n(h_n) \to \Phi'_\theta(h)$. In addition, $\Phi'_\theta(\cdot)$ is continuous on $\mathcal{D}_0$. Therefore, Theorem 2.1 implies that

$$\bigl\{r_n\bigl(\Phi(X_n) - \Phi(\theta)\bigr), n \geq 1\bigr\} = \bigl\{f_n\bigl(r_n(X_n - \theta)\bigr), n \geq 1\bigr\}$$

satisfies the large deviation principle with speed $\lambda(n)$ and rate function $I_{\Phi'_\theta}$.

Now, we consider the mapping $\varphi_n : \mathcal{D}_n \mapsto \mathcal{Y} \times \mathcal{Y}$, where $\varphi_n(h) = (f_n(h), \Phi'_\theta(h))$ for all $h \in \mathcal{D}_n$. If $\Phi'_\theta(\cdot)$ is continuous on $\mathcal{X}$, then for every subsequence $h_{n'} \in \mathcal{D}_{n'}$ converging to $h \in \mathcal{X}$, $\varphi_{n'}(h_{n'})$ converges to $(\Phi'_\theta(h), \Phi'_\theta(h))$. Hence, Theorem 2.1 implies $\{\varphi_n(r_n(X_n - \theta)), n \geq 1\}$ satisfies the large deviation principle with speed $\lambda(n)$ and rate function

$$J_{\Phi,\theta}(y_1, y_2) = \inf\{I(x); \Phi'_\theta(x) = y_1 = y_2\}, \qquad (y_1, y_2) \in \mathcal{Y} \times \mathcal{Y}.$$

Therefore, by the classical contraction principle [see Dembo and Zeitouni (1998), Theorem 4.2.1], we conclude that the difference

$$\bigl\{r_n\bigl(\Phi(X_n) - \Phi(\theta)\bigr) - \Phi'_\theta\bigl(r_n(X_n - \theta)\bigr), n \geq 1\bigr\}$$

satisfies the large deviation principle with speed $\lambda(n)$ and rate function

$$\inf\{J_{\Phi,\theta} = (y_1, y_2); y_1 - y_2 = z\} = I_{\Phi,\theta}(z) \qquad \text{for } z \in \mathcal{Y}. \qquad \square$$

4. Moderate deviations of estimators. In this section, moderate deviation principles for some estimators will be established by applying the delta method in large deviation to Wilcoxon statistic, Kaplan–Meier estimator, the empirical quantile processes, M-estimators and L-statistics.

Let us introduce some notation. Given an arbitrary set T and a Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$, the Banach space $l_\infty(T, \mathbb{B})$ is the set of all maps $z : T \mapsto \mathbb{B}$ that are uniformly norm-bounded equipped with the norm $\|z\| = \sup_{t\in T} \|z(t)\|_{\mathbb{B}}$. Let $l_\infty(T)$

be the Banach space of all bounded real functions x on T, equipped with the sup-norm $\|x\| = \sup_{t\in T} |x(t)|$. It is a nonseparable Banach space if T is infinite. On $l_\infty(T)$, we will consider the σ-field $\mathcal{B}$ generated by all balls and all coordinates $x(t), t \in T$.

Let (S, d) be a complete separable and measurable metric space and let $b\mathcal{S}$ be the space of all bounded real measurable functions on $(S, \mathcal{S})$ where $\mathcal{S}$ is the Borel σ-algebra of S. Let $\{X, X_n, n \geq 1\}$ be a sequence of i.i.d. random variables with values in S on a probability space $(\Omega, \mathcal{F}, P)$, of law μ. Let L_n denote the empirical measures; that is,

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}, \qquad n \geq 1.$$

For given a class of functions $\mathfrak{F} \subset b\mathcal{S}$, let $l_\infty(\mathfrak{F})$ be the space of all bounded real functions on $\mathfrak{F}$ with sup-norm $\|F\|_{\mathfrak{F}} = \sup_{f\in\mathfrak{F}} |F(f)|$. This is a Banach space. Every $\nu \in M_b(S)$ [the space of signed measures of finite variation on $(S, \mathcal{S})$] corresponds to an element $\nu^{\mathfrak{F}} = \nu(f) = \int f\,d\nu$ for all $f \in \mathfrak{F}$.

Let $D[a, b]$ denote the Banach space of all right continuous with left-hand limits functions $z : [a, b] \mapsto \mathbb{R}$ on an interval $[a, b] \subset \overline{\mathbb{R}}$ equipped with the uniform norm. Let $BV[a, b]$ denote the set of all cadlag functions with finite total variation and set $BV_M[a, b] = \{A \in BV[a, b]; \int |dA| \leq M\}$, where the notation $\int |dA|$ denotes the total variation of the function A. In this article, we also let $\{a_n = a(n), n \geq 1\}$ be a sequence of real numbers such that as $n \to \infty$,

$$a_n \to \infty \quad \text{and} \quad a_n/\sqrt{n} \to 0.$$

4.1. *Moderate deviations for Wilcoxon statistic*. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from distribution functions F and G on $\mathbb{R}$, respectively. If F_m and G_n are the empirical distribution functions of the two samples; that is,

$$F_m(x) = \frac{1}{m}\sum_{i=1}^{m} \delta_{X_i}((-\infty, x]) \quad \text{and} \quad G_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{Y_i}((-\infty, x]),$$

then the Wilcoxon statistic is defined by $W_{m,n} = \int F_m\,dG_n$. It is an estimator of $P(X \leq Y)$.

THEOREM 4.1. *Assume that $m/(m+n) \to \lambda \in (0, 1)$ as $m, n \to \infty$. Then*

$$\left\{\frac{\sqrt{mn/(m+n)}}{a(mn/(m+n))}\left(\int F_m\,dG_n - \int F\,dG\right), n \geq 1\right\} \tag{4.1}$$

satisfies the LDP in $\mathbb{R}$ with speed $a^2(mn/(m+n))$ and rate function I^W defined by

$$I^W(x) = \frac{x^2}{2(\lambda \operatorname{Var}(F(Y)) + (1-\lambda)\operatorname{Var}(G(X)))}. \tag{4.2}$$

PROOF. Applying Theorem 2 of Wu (1994) to $L_n^X = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, $\mathfrak{F}_1 = \{(-\infty, x]; x \in \mathbb{R}\}$, and $L_n^Y = \frac{1}{n}\sum_{i=1}^n \delta_{Y_i}$, $\mathfrak{F}_2 = \{(-\infty, y]; y \in \mathbb{R}\}$, respectively, and using the product principle in large deviations [Dembo and Zeitouni (1998)], we obtain that $\{\frac{\sqrt{n}}{a_n}(F_n - F, G_n - G), n \geq 1\}$ satisfies the LDP in $l_\infty(\mathbb{R}) \times l_\infty(\mathbb{R})$ with speed a_n^2 and rate function $\{I_F(\alpha) + I_G(\beta)\}$, where

$$I_F(\alpha) = \inf\biggl\{\frac{1}{2}\int \gamma^2(x)\,dF(x);\ \int \gamma(x)\,dF(x) = 0, \alpha(t) = \int_{(-\infty,t]} \gamma(x)\,dF(x)$$

$$\text{for each } t \in \mathbb{R},\ \gamma : \mathbb{R} \to \mathbb{R} \text{ is measurable}\biggr\}$$

$$= \begin{cases} \dfrac{1}{2}\displaystyle\int |\alpha_F'(x)|^2\,dF(x), & \text{if } \alpha \ll F \text{ and } \displaystyle\lim_{|t|\to\infty} |\alpha(t)| = 0, \\ \infty, & \text{otherwise} \end{cases}$$

and $\alpha_F' = d\alpha/dF$. Since $\frac{\sqrt{m}}{\sqrt{mn/(m+n)}} \to (1-\lambda)^{1/2}$ and $\frac{\sqrt{n}}{\sqrt{mn/(m+n)}} \to \lambda^{1/2}$, then

$$\left\{\frac{\sqrt{mn/(m+n)}}{a(mn/(m+n))}(F_m - F, G_n - G), n \geq 1\right\}$$

satisfies the LDP in $l_\infty(\mathbb{R}) \times l_\infty(\mathbb{R})$ with speed $a^2(mn/(m+n))$ and rate function given by

$$I_{F,G}(\alpha, \beta) = \frac{1}{1-\lambda} I_F(\alpha) + \frac{1}{\lambda} I_G(\beta).$$

Note that $\{I_{F,G}(\alpha, \beta) < \infty\} \subset BV(\mathbb{R}) \times BV(\mathbb{R})$ and $(F_m, G_n) \in BV_1(\mathbb{R}) \times BV_1(\mathbb{R})$. For each $M \geq 1$, we consider the map $\Phi : D(\mathbb{R}) \times BV_M(\mathbb{R}) \mapsto \mathbb{R}$ defined as

$$\Phi(A, B) = \int_{\mathbb{R}} A(s)\,dB(s).$$

Then $\Phi(F_m, G_n) = \int F_m\,dG_n$, and by Lemma 3.9.17 of van der Vaart and Wellner (1996), Φ is Hadamard differentiable at each $(A, B) \in \mathcal{D}_\Phi = \{\int |dA| < \infty\}$ and the derivative is given by

$$\Phi_{A,B}'(\alpha, \beta) = \int_{\mathbb{R}} A(s)\,d\beta(s) + \int_{\mathbb{R}} \alpha(s)\,dB(s),$$

where $\int_{(a,b]} A(s)\,d\beta(s)$ is defined via integration by parts if β is not of bounded variation; that is,

$$\int_{(a,b]} A(s)\,d\beta(s) = A(b)\beta(b) - A(a)\beta(a) - \int_{(a,b]} \beta(s-)A(s).$$

Thus, by Theorem 3.1 with $\mathcal{D}_0 = \{(\alpha, \beta); I_F(\alpha) < \infty, I_G(\beta) < \infty\}$, we conclude that

$$\left\{\frac{\sqrt{mn/(m+n)}}{a(mn/(m+n))}\left(\int F_m\, dG_n - \int F\, dG\right), n \geq 1\right\}$$

satisfies the LDP on $\mathbb{R}$ with speed $a^2(mn/(m+n))$ and rate function given by

$$\begin{aligned}
I^W(x) &= \inf\left\{\frac{1}{1-\lambda} I_F(\alpha) + \frac{1}{\lambda} I_G(\beta), \int F(s)\, d\beta(s) + \int \alpha(s)\, dG(s) = x\right\} \\
&= \inf\left\{\frac{1}{2(1-\lambda)} \int (\alpha_F')^2\, dF + \frac{1}{2\lambda} \int (\beta_G')^2\, dG,\right. \\
&\qquad \int F\beta_G'\, dG - \int G\alpha_F'\, dF = x, \alpha \ll F, \\
&\qquad \left.\beta \ll G, \lim_{|t|\to\infty} |\alpha(t)| = 0, \lim_{|t|\to\infty} |\beta(t)| = 0\right\} \\
&= \frac{x^2}{2(\lambda \operatorname{Var}(F(Y)) + (1-\lambda) \operatorname{Var}(G(X)))}. \qquad \square
\end{aligned}$$

4.2. *Moderate deviations for Kaplan–Meier estimator.* Let X and C be independent, nonnegative random variables with distribution functions F and G. Let $X_1, \ldots, X_n$ be i.i.d. random variables distributed according to the distribution function F and let $C_1, \ldots, C_n$ be i.i.d. random variables distributed according to the distribution function G. $X_1, \ldots, X_n$ and $C_1, \ldots, C_n$ are assumed to be independent. Observed data are the pairs $(Z_1, \Delta_1), \ldots, (Z_n, \Delta_n)$, where $Z_i = X_i \wedge C_i$, and $\Delta_i = 1_{\{X_i \leq C_i\}}$. The cumulative hazard function is defined by

$$\Lambda(t) = \int_{[0,t]} \frac{1}{\overline{F}(s)}\, dF(s) = \int_{[0,t]} \frac{1}{\overline{H}(s)}\, dH^{uc}(s), \tag{4.3}$$

where $\overline{F}(t) = P(X \geq t)$ and $\overline{H}(t) = P(Z \geq t)$ are (left-continuous) survival distributions, and $H^{uc}(t) = P(Z \leq t, \Delta = 1)$ is a subdistribution function of the uncensored observations, where $\Delta = 1_{\{X \leq C\}}$. We also denote $H^c(t) = P(Z \leq t, \Delta = 0)$. The Nelson–Aalen estimator is defined by

$$\Lambda_n(t) = \int_{[0,t]} \frac{1}{\overline{H}_n(s)}\, dH_n^{uc}(s), \tag{4.4}$$

where

$$H_n^{uc}(t) = \frac{1}{n}\sum_{i=1}^n 1_{\{Z_i \leq t, \Delta_i = 1\}} \quad \text{and} \quad \overline{H}_n(t) = \frac{1}{n}\sum_{i=1}^n 1_{\{Z_i \geq t\}} \tag{4.5}$$

are the empirical subdistribution functions of the uncensored failure time and the survival function of the observation times, respectively.

The distribution function $F(t)$ can be rewritten as

$$1 - F(t) = \prod_{0<s\le t} \bigl(1 - d\Lambda(s)\bigr).$$

The Kaplan–Meier estimator $\hat{F}_n(t)$ for the distribution function $F(t)$ is defined by

$$1 - \hat{F}_n(t) = \prod_{0<s\le t} \bigl(1 - d\Lambda_n(s)\bigr). \tag{4.6}$$

The Kaplan–Meier estimator $\hat{F}_n$ is the nonparametric maximum likelihood estimator of F in the right censored data model, proposed by Kaplan and Meier (1958). Dinwoodie (1993) studied large deviations for censored data and established a large deviation principle for $\sup_{x\in\tau} |\hat{F}_n(x) - F(x)|$ where τ is a fixed time satisfying $\{1 - F(\tau)\}\{1 - G(\tau)\} > 0$. Bitouzé, Laurent and Massart (1999) obtained an exponential inequality for $\sup_{x\in\mathbb{R}}\{(1 - G(x))|\hat{F}_n(x) - F(x)|\}$. Wellner (2007) provided a bound for the constant in the inequality. In this subsection, we establish its moderate deviation principle.

THEOREM 4.2. *Let $\tau > 0$ such that $H(\tau) < 1$. Then $\{\frac{\sqrt{n}}{a(n)}(\Lambda_n - \Lambda), n \ge 1\}$ satisfies the LDP in $D[0, \tau]$ with speed $a^2(n)$ and rate function I^Λ given by*

$$I^\Lambda(\phi) = \inf\biggl\{ I_{F,G}(\alpha, \beta); \int_{[0,t]} \frac{1}{\overline{H}(s)}\, d\alpha(s) - \int_{[0,t]} \frac{\beta(s)}{\overline{H}^2(s)}\, dH^{uc}(s) = \phi(t), \text{ for any } t \in [0, \tau] \biggr\}, \tag{4.7}$$

where

$$I_{F,G}(\alpha,\beta) = \begin{cases} \dfrac{1}{2}\Bigl(\displaystyle\int |\alpha'_{H^{uc}}(u)|^2\, dH^{uc}(u) + \int |(\alpha+\beta)'_{H^c}(u)|^2\, dH^c(u) \Bigr), \\ \qquad \text{if } \alpha \ll H^{uc},\ \alpha+\beta \ll H^c \text{ and } \lim_{t\to\infty} |\beta(t)| = 0, \\ \infty, \qquad \text{otherwise}. \end{cases} \tag{4.8}$$

PROOF. The pair $(H_n^{uc}, \overline{H}_n)$ can be identified with the empirical distribution of the observations indexed by the functions $\mathfrak{F}_1 = \{I_{\{z\le t, \Delta=1\}}, t \in \mathbb{R}\}$ and $\mathfrak{F}_2 = \{I_{\{z\ge t\}}, t \in \mathbb{R}\}$. It is easy to verify that the two classes $\mathfrak{F}_1$ and $\mathfrak{F}_2$ are Donsker classes and the mapping $\Psi : l_\infty(\mathfrak{F}) \mapsto l_\infty(\mathfrak{F}_1) \times l_\infty(\mathfrak{F}_2)$ defined by $\phi \longrightarrow (\phi|_{\mathfrak{F}_1}, \phi|_{\mathfrak{F}_2})$ is continuous, where $\mathfrak{F} = \bigcup_{j=1}^2 \mathfrak{F}_j$. Applying Theorem 2 of Wu (1994) to $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{(Z_i,\Delta_i)}$ and $\mathfrak{F}$, and the classical contraction principle [see Dembo and Zeitouni (1998), Theorem 4.2.1] to Ψ, we can get that

$$\biggl\{ \frac{\sqrt{n}}{a(n)} (H_n^{uc} - H^{uc}, \overline{H}_n - \overline{H}), n \ge 1 \biggr\}$$

satisfies the LDP on $D([0,\tau])\times D([0,\tau])$ with speed $a^2(n)$ and rate function

$$
\begin{aligned}
I_{F,G}(\alpha,\beta) = \inf\bigg\{ &\frac{1}{2}\bigg(\int \gamma_1^2(u)\,dH^{uc}(u) + \int \gamma_0^2(u)\,dH^c(u)\bigg);\\
&\int \gamma_1(u)\,dH^{uc}(u) + \int \gamma_0(u)\,dH^c(u) = 0,\\
&\text{and for any } t\in[0,\infty), \int_{[0,t]} \gamma_1(u)\,dH^{uc}(u) = \alpha(t),\\
&\int_{[t,\infty)} \gamma_1(u)\,dH^{uc}(u) + \int_{[t,\infty)} \gamma_0(u)\,dH^c(u) = \beta(t)\bigg\}\\
= &\begin{cases} \frac{1}{2}\bigg(\int |\alpha'_{H^{uc}}(u)|^2\,dH^{uc}(u) + \int |(\alpha+\beta)'_{H^c}(u)|^2\,dH^c(u)\bigg), \\ \qquad \text{if } \alpha \ll H^{uc},\ \alpha+\beta \ll H^c \text{ and } \lim_{t\to\infty}|\beta(t)| = 0, \\ \infty, \qquad \text{otherwise.} \end{cases}
\end{aligned}
$$

Set $\mathcal{D}_\Phi = \{(A,B)\in BV_1([0,\tau])\times D([0,\tau]);\, B\ge \overline{H}(\tau)/2\}$. By the Dvoretzky–Kiefer–Wolfowitz inequality [cf. Massart (1990)], for any $\varepsilon>0$,

$$P\Big(\sup_{t\in[0,\tau]} |\overline{H}_n(t) - \overline{H}(t)| > \varepsilon\Big) \le 2\exp\{-2n\varepsilon^2\}.$$

In particular, take $\varepsilon = \overline{H}(\tau)/2$, then we have

$$
\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{a^2(n)} \log P^*\big((H_n^{uc},\overline{H}_n)\notin \mathcal{D}_\Phi\big)\\
&\qquad \le \limsup_{n\to\infty} \frac{1}{a^2(n)} \log P^*\big(\overline{H}_n(t) \le \overline{H(\tau)}/2\big) = -\infty.
\end{aligned}
\tag{4.9}
$$

Consider the maps $\Phi_1:\mathcal{D}_\Phi \subset BV_1([0,\tau])\times D([0,\tau]) \mapsto BV_1([0,\tau])\times D([0,\tau])$ and $\Phi_2: BV([0,\tau])\times D([0,\tau]) \mapsto D([0,\tau])$ defined as

$$\Phi_1(A,B) = (A, 1/B) \quad \text{and} \quad \Phi_2:(A,B)\mapsto \int_{[0,\cdot]} B\,dA.$$

Define $\Phi(A,B) = \Phi_2(\Phi_1(A,B))$. Then $\Phi(H_n^{uc},\overline{H}_n) = \Lambda_n$, $\Phi(H^{uc},\overline{H}) = \Lambda$ and by Lemma 3.9.17 of van der Vaart and Wellner (1996), Φ is Hadamard differentiable at each $(A,B)\in\mathcal{D}_\Phi$. The derivative is given by

$$\Phi'_{A,B}(\alpha,\beta)(t) = \int_{[0,t]} \frac{1}{B(s)}\,d\alpha(s) - \int_{[0,t]} \frac{\beta(s)}{B^2(s)}\,dA(s).$$

Applying Theorem 3.1 to $\Omega_n = \{(H_n^{uc},\overline{H}_n)\in\mathcal{D}_\Phi\}$, $P_n(\cdot) = P(\cdot|\Omega_n)$ and $\mathcal{D}_0 = \mathcal{D}_\Phi$ together with (4.9), we conclude that $\{\frac{\sqrt{n}}{a(n)}(\Lambda_n - \Lambda), n\ge 1\}$ satisfies the LDP in $D[0,\tau]$ with speed $a^2(n)$ and rate function I^Λ given by

$$I^\Lambda(\phi) = \inf\biggl\{ I_{F,G}(\alpha,\beta);\ \int_{[0,t]} \frac{1}{\overline{H}(s)}\, d\alpha(s) - \int_{[0,t]} \frac{\beta(s)}{\overline{H}^2(s)}\, dH^{uc}(s) = \phi(t), \text{ for any } t \in [0,\tau] \biggr\}. \qquad \square$$

Next, we give some other representations. Let $\{(G^{uc}(t), \overline{G}(t)), t \in [0,\tau]\}$ be a zero-mean Gaussian process with covariance structure

$$E(G^{uc}(s)G^{uc}(t)) = H^{uc}(s \wedge t) - H^{uc}(s)H^{uc}(t),$$
$$E(\overline{G}(s)\overline{G}(t)) = \overline{H}(s \vee t) - \overline{H}(s)\overline{H}(t),$$

and

$$E(G^{uc}(s)\overline{G}(t)) = \bigl(H^{uc}(s) - H^{uc}(t-)\bigr) I_{(-\infty,s]}(t) - H^{uc}(s)\overline{H}(t).$$

Set $\tilde{T} = \{(j,t), j = 1,2, t \in [0,\tau]\}$ and

$$\tilde{Z} = \{\tilde{Z}_{(j,t)};\ j = 1,2, t \in [0,\tau], \tilde{Z}_{(1,t)} = G^{uc}(t), \tilde{Z}_{(2,t)} = \overline{G}(t)\}.$$

Then by Theorem 5.2 of Arcones (2004), $\{\{\tilde{Z}_{(j,t)}/\sqrt{\lambda(n)}, (j,t) \in \tilde{T}\}, n \geq 1\}$ satisfies LDP on $l_\infty(\tilde{T})$ with speed $\lambda(n)$ and rate function given by

$$\tilde{I}(x) = \inf\{\tfrac{1}{2}E(\gamma^2);\ \gamma \in \mathcal{L},\ E(\gamma \tilde{Z}_{(j,t)}) = x_{(j,t)} \text{ for all } (j,t) \in \tilde{T}\},$$

where $\mathcal{L}$ is the closed vector space of $L^2(P)$ generated by $\{\tilde{Z}_{(j,t)}, (j,t) \in \tilde{T}\}$. Since the mapping $\Psi : l_\infty(\tilde{T}) \mapsto l_\infty([0,\tau], \mathbb{R}^2)$ defined by

$$\{\phi_{(j,t)}, (j,t) \in \tilde{T}\} \longrightarrow \{(\phi_{(1,t)}, \phi_{(2,t)}), t \in [0,\tau]\}$$

is continuous, then by the classical contraction principle [see Dembo and Zeitouni (1998), Theorem 4.2.1], we know that $\{\frac{1}{\sqrt{\lambda(n)}}(G^{uc}, \overline{G}), n \geq 1\}$ satisfies the LDP on $D([0,\tau]) \times D([0,\tau])$ with speed $\lambda(n)$ and rate function $I_{F,G}(\alpha,\beta)$, where $\lambda(n) \to \infty$ as $n \to \infty$.

Define $M^{uc}(t) = G^{uc}(t) - \int_{[0,t]} \overline{G}(u)\, d\Lambda(u)$ and

$$Z(t) = \int_{[0,t]} \frac{1}{\overline{H}(s)}\, dG^{uc}(s) - \int_{[0,t]} \frac{\overline{G}(s)}{\overline{H}^2(s)}\, dH^{uc}(s), \tag{4.10}$$

where the first term on the right-hand side is to be understood via integration by parts. Then M^{uc} is a zero-mean Gaussian martingale with covariance function [van der Vaart and Wellner (1996), page 384]

$$E(M^{uc}(s)M^{uc}(t)) = \int_{[0,s\wedge t]} \overline{H}(u)\bigl(1 - \Delta\Lambda(u)\bigr)\, d\Lambda(u),$$

where $\Delta\Lambda(u) = \Lambda(u) - \Lambda(u-)$ and $Z(t) = \int_{[0,t]} \frac{1}{\overline{H}(s)}\, dM^{uc}(s)$ is a zero-mean Gaussian process with covariance function

$$E(Z(s)Z(t)) = \int_{[0,s\wedge t]} \frac{1-\Delta\Lambda(u)}{\overline{H}(u)}\, d\Lambda(u).$$

Therefore, by Theorem 2.1, we conclude that $\{\{Z(t)/\sqrt{\lambda(n)}, t \in [0,\tau]\}, n \geq 1\}$ satisfies the LDP on $D([0,\tau])$ with speed $\lambda(n)$ and rate function $I^{\Lambda}(\phi)$. Furthermore, from Theorem 5.2 of Arcones (2004) and Theorem 3.1 of Arcones (2003b), we have the following result.

THEOREM 4.3. *Let $\tau > 0$ such that $H(\tau) < 1$. Then $\{\frac{\sqrt{n}}{a(n)}(\Lambda_n - \Lambda), n \geq 1\}$ satisfies the LDP in $D[0,\tau]$ with speed $a^2(n)$ and rate function I^{Λ} given by*

$$\begin{aligned} I^{\Lambda}(\phi) = \sup_{m\geq 1, t_1,\dots,t_m\in[0,\tau]} \sup_{\alpha_1,\dots,\alpha_m\in\mathbb{R}} \Bigg\{ &\sum_{i=1}^m \phi_{t_i}\alpha_i \\ &- \frac{1}{2}\sum_{k,j=1}^m \alpha_k\alpha_j \\ &\times \int_{[0,t_k\wedge t_j]} \frac{1-\Delta\Lambda(u)}{\overline{H}(u)}\, d\Lambda(u)\Bigg\}. \end{aligned} \tag{4.11}$$

In particular, for any $r > 0$,

$$\lim_{n\to\infty} \frac{1}{a^2(n)} \log P\left(\frac{\sqrt{n}}{a(n)} \sup_{x\in[0,\tau]} |\Lambda_n(x) - \Lambda(x)| \geq r\right) = -\frac{r^2}{2\sigma_{\Lambda}^2}, \tag{4.12}$$

where $\sigma_{\Lambda}^2 = \int_{[0,\tau]} \frac{1-\Delta\Lambda(u)}{\overline{H}(u)}\, d\Lambda(u)$.

Now we present the moderate deviations for the Kaplan–Meier estimator $\hat{F}_n(t)$.

THEOREM 4.4. *Let $\tau > 0$ such that $H(\tau) < 1$. Then $\{\frac{\sqrt{n}}{a(n)}(\hat{F}_n - F), n \geq 1\}$ satisfies the LDP in $D[0,\tau]$ with speed $a^2(n)$ and rate function I^{KM} given by*

$$\begin{aligned} I^{KM}(\phi) = \sup_{m\geq 1, t_1,\dots,t_m\in[0,\tau]} \sup_{\alpha_1,\dots,\alpha_m\in\mathbb{R}} \Bigg\{ &\sum_{i=1}^m \phi_{t_i}\alpha_i \\ &- \frac{1}{2}\sum_{k,j=1}^m \alpha_k\alpha_j \\ &\times \int_{[0,t_k\wedge t_j]} \frac{(1-F(t_k))(1-F(t_j))}{(1-\Delta\Lambda(u))\overline{H}(u)}\, d\Lambda(u)\Bigg\}. \end{aligned} \tag{4.13}$$

In particular, for any $r > 0$,

$$\lim_{n\to\infty}\frac{1}{a^2(n)}\log P\left(\frac{\sqrt{n}}{a(n)}\sup_{x\in[0,\tau]}|\hat{F}_n(x)-F(x)|\ge r\right)=-\frac{r^2}{2\sigma_{KM}^2}, \tag{4.14}$$

where

$$\sigma_{KM}^2=\sup_{t\in[0,\tau]}\big(1-F(t)\big)^2\int_{(0,t]}\frac{1}{(1-\Delta\Lambda(u))\overline{H}(u)}\,d\Lambda(u).$$

PROOF. The map $\Phi: BV[0,\tau]\subset D[0,\tau]\mapsto D[0,\tau]$ is defined as

$$\Phi(A)(t)=\prod_{0<s\le t}\big(1+dA(s)\big).$$

Then, $1-F(x)=\Phi(-\Lambda)(x)$ and $1-\hat{F}_n(x)=\Phi(-\Lambda_n)(x)$. Since $H(\tau)<1$, there exists some $M\in(0,\infty)$ such that $\Lambda\in BV_M[0,\tau]$. From (4.12), we have

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{a^2(n)}\log P^*(\Lambda_n\notin BV_{M+1}[0,\tau])\\&\qquad\le\lim_{n\to\infty}\frac{1}{a^2(n)}\log P\Big(\sup_{x\in[0,\tau]}|\Lambda_n(x)-\Lambda(x)|\ge 1\Big)=-\infty.\end{aligned}$$

By Lemma 3.9.30 of van der Vaart and Wellner (1996), we know that Φ is Hadamard differentiable in $BV_{M+1}[0,\tau]$ with derivative

$$\Phi'_A(\alpha)(t)=\int_{(0,t]}\Phi(A)(0,u)\Phi(A)(u,t]\,d\alpha(u),$$

where $\Phi(A)(u,t]=\prod_{u<s\le t}(1+dA(s))$. Applying Theorem 3.1 to $\Omega_n=\{\Lambda_n\in BV_{M+1}[0,\tau]\}$, $P_n(\cdot)=P(\cdot|\Omega_n)$ and $\mathcal{D}_0=BV_{M+1}[0,\tau]$, we obtain from Theorem 4.2 that $\{\frac{\sqrt{n}}{a(n)}(\hat{F}_n-F), n\ge 1\}$ satisfies the LDP in $D[0,\tau]$ with speed $a^2(n)$ and rate function $\tilde{I}^{KM}$ given by

$$\begin{aligned}\tilde{I}^{KM}=\inf\bigg\{&I^{\Lambda}(\alpha);\\&\int_{(0,t]}\Phi(F)(0,u)\Phi(F)(u,t]\,d\alpha(u)=\phi(t),\ \text{for any } t\in[0,\tau]\bigg\}.\end{aligned}$$

On the other hand, we consider the process $\Phi'_{-\Lambda}(Z)(t)$, where Z is defined by (4.10). Since

$$\begin{aligned}\Phi'_{-\Lambda}(Z)(t)&=\int_{(0,t]}\frac{(1-F(u-))((1-F(t))}{1-F(u)}\,dZ(u)\\&=\big(1-F(t)\big)\int_{(0,t]}\frac{1}{1-\Delta\Lambda(u)}\,dZ(u),\end{aligned}$$

which is a zero-mean Gaussian process with covariance function

$$(1 - F(s))(1 - F(t)) \int_{(0,s\wedge t]} \frac{1}{(1 - \Delta\Lambda(u))\overline{H}(u)} \, d\Lambda(u),$$

then, by Theorem 5.2 of Arcones (2004) and Theorem 3.1 of Arcones (2003b), we obtain the conclusion of the theorem. □

4.3. *Moderate deviations for the empirical quantile processes.* For a nondecreasing function $G \in D[a, b]$ and any $p \in \mathbb{R}$, define $G^{-1}(p) = \inf\{x; G(x) \geq p\}$. Let $D_1[a, b]$ denote the set of all restrictions of distribution functions on $\mathbb{R}$ to $[a, b]$ and let $D_2[a, b]$ denote the set of distribution functions of measures that concentrate on $(a, b]$.

THEOREM 4.5. *Let $0 < p < q < 1$ be fixed and let F be a distribution function with continuous and positive derivative f on the interval $[F^{-1}(p) - \varepsilon, F^{-1}(q) + \varepsilon]$ for some $\varepsilon > 0$. Let F_n be the empirical distribution function of an i.i.d. sample $X_1, \dots, X_n$ of size n from F. Then $\{\frac{\sqrt{n}}{a(n)}(F_n^{-1} - F^{-1}), n \geq 1\}$ satisfies the LDP in $l_\infty[p, q]$ with speed $a^2(n)$ and rate function I^{EQ} given by*

$$I^{EQ}(\phi) = \inf\left\{ I_F(\alpha); -\frac{\alpha(F^{-1}(x))}{f(F^{-1}(x))} = \phi(x) \text{ for all } x \in [p, q] \right\},$$

where

$$I_F(\alpha) = \begin{cases} \dfrac{1}{2} \displaystyle\int |\alpha'_F(x)|^2 \, dF(x), & \text{if } \alpha \ll F \text{ and } \lim\limits_{|t| \to \infty} |\alpha(t)| = 0, \\ \infty, & \text{otherwise.} \end{cases}$$

PROOF. Applying Theorem 2 of Wu (1994) to $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$, and $\mathfrak{F} = \{(-\infty, x]; x \in \mathbb{R}\}$, we know that $\{\frac{\sqrt{n}}{a(n)}(F_n - F), n \geq 1\}$ satisfies the LDP on $D(\mathbb{R})$ with speed $a^2(n)$ and rate function I_F. By Lemma 3.9.23 of van der Vaart and Wellner (1996), it follows that the inverse map $\Phi : G \mapsto G^{-1}$ as a map $D_1[F^{-1}(p) - \varepsilon, F^{-1}(q) + \varepsilon] \mapsto l_\infty[p, q]$ is Hadamard differentiable at F tangentially to $C[F^{-1}(p) - \varepsilon, F^{-1}(q) + \varepsilon]$, and the derivative is the map $\alpha \mapsto -\alpha(F^{-1})/f(F^{-1})$. Therefore, by Theorem 3.1, we conclude that $\{\frac{\sqrt{n}}{a(n)}(F_n^{-1} - F^{-1}), n \geq 1\}$ satisfies the LDP in $l_\infty[p, q]$ with speed $a^2(n)$ and the rate function I^{EQ}. □

4.4. *Moderate deviations for the empirical copula processes.* Let $BV_1^+(\mathbb{R}^2)$ denote the space of bivariate distribution functions on $\mathbb{R}^2$. For $H \in BV_1^+(\mathbb{R}^2)$, set $F(x) = H(x, \infty)$ and $G(y) = H(\infty, y)$.

Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. vectors with distribution function H. The empirical estimator for the copula function $C(u, v) = H(F^{-1}(u), G^{-1}(v))$ is defined by $C_n(u, v) = H_n(F_n^{-1}(u), G_n^{-1}(v))$, where H_n, F_n and G_n are the joint and marginal empirical distributions of the observations.

THEOREM 4.6. *Let $0 < p < q < 1$ be fixed. Suppose that F and G are continuously differentiable on the intervals $[F^{-1}(p) - \varepsilon, F^{-1}(q) + \varepsilon]$ and $[G^{-1}(p) - \varepsilon, G^{-1}(q) + \varepsilon]$ with strictly positive derivatives f and g, respectively, for some $\varepsilon > 0$. Furthermore, assume that $\partial H/\partial x$ and $\partial H/\partial y$ exist and are continuous on the product intervals. Then $\{\frac{\sqrt{n}}{a(n)}(C_n - C), n \geq 1\}$ satisfies the LDP in $l_\infty([p,q]^2)$ with speed $a^2(n)$ and rate function I^C defined by*

$$I^C(\phi) = \inf\{I_H(\alpha); \Phi'_H(\alpha) = \phi\},$$

where

$$\Phi'_H(\alpha)(u,v) = \alpha(F^{-1}(u), G^{-1}(v)) - \frac{\partial H}{\partial x}(F^{-1}(u), G^{-1}(v))\frac{\alpha(F^{-1}(u), \infty)}{f(F^{-1}(u))} - \frac{\partial H}{\partial y}(F^{-1}(u), G^{-1}(v))\frac{\alpha(\infty, G^{-1}(u))}{g(G^{-1}(u))}.$$

PROOF. By Theorem 2 of Wu (1994), we know that

$$P\left(\frac{\sqrt{n}}{a(n)}\left(\sum_{k=1}^n \delta_{(X_k,Y_k)}\big((-\infty, x] \times (-\infty, y]\big) - H(x,y)\right) \in \cdot\right)$$

satisfies the LDP on $D(\mathbb{R}^2)$ with speed $a^2(n)$ and rate function defined as

$$\begin{aligned} I_H(\alpha) &= \inf\left\{\frac{1}{2}\int \gamma^2(x,y)H(dx,dy);\ \alpha(s,t) = \int \gamma(x,y) I_{\{x \leq s, y \leq t\}} H(dx,dy) \right. \\ &\qquad\qquad \left. \text{for each } (s,t) \in \mathbb{R}^2, \text{ and } \int \gamma\, dH = 0\right\} \\ &= \begin{cases} \frac{1}{2}\int (\alpha'_H)^2(x,y)H(dx,dy), & \text{if } \alpha \ll H \text{ and } \lim_{|s|,|t|\to\infty} |\alpha(s,t)| = 0, \\ \infty, & \text{otherwise.} \end{cases} \end{aligned}$$

Then, by Lemma 3.9.28 of van der Vaart and Wellner (1996), we conclude that the map $\Phi: H \mapsto H(F^{-1}, G^{-1})$ as a map $BV_1^+(\mathbb{R}^2) \subset D(\overline{\mathbb{R}}^2) \mapsto l_\infty([p,q]^2)$ is Hadamard differentiable at H tangentially to $C(\overline{\mathbb{R}}^2)$, and the derivative is Φ'_H. Therefore, it follows from Theorem 3.1 that, $\{\frac{\sqrt{n}}{a(n)}(C_n - C), n \geq 1\}$ satisfies the LDP in $l_\infty([p,q]^2)$ with speed $a^2(n)$ and rate function I^C as defined in the theorem. $\square$

4.5. *Moderate deviations for M-estimators.* M-estimators were first introduced by Huber (1964). Let X be a random variable taking its values in a measurable space $(S, \mathcal{S})$ with distribution F, let $X_1, \ldots, X_n$ be a random sample of X, and let F_n denote the empirical distribution function of X. Let Θ be a Borel subset of $\mathbb{R}^d$. A M-estimator $\theta_n(X_1, \dots, X_n)$ over the function g is a solution of

$$\int g(x, \theta_n)\, dF_n(x) = \inf_{\theta \in \Theta} \int g(x, \theta)\, dF_n(x).$$

If $g(x, \theta)$ is differentiable with respect to θ, then the M-estimator $\theta_n(X_1, \dots, X_n)$ may be defined as a solution of the equation

$$\int \nabla_\theta g(x, \theta_n)\, dF_n(x) = 0,$$

where $\nabla_\theta g(x, \theta) = (\frac{\partial g(x,\theta)}{\partial \theta^1}, \dots, \frac{\partial g(x,\theta)}{\partial \theta^d})$. The detailed description on M-estimators can be found in Serfling (1980).

Jurečková, Kallenberg and Veraverbeke (1988), Arcones (2002) and Inglot and Kallenberg (2003) studied moderate deviations for M-estimators. In this subsection, we study the problem by the delta method. Let $\psi(x, \theta) = (\psi^1(x, \theta), \dots, \psi^d(x, \theta)) : S \times \Theta \mapsto \mathbb{R}^d$. We also need the following conditions.

(C1) $\psi(x, \theta)$ is continuous in θ for each $x \in S$, and $\psi(x, \theta)$ is measurable in x for each $\theta \in \Theta$.

Define

$$\Psi(\theta) = (\Psi^1(\theta), \dots, \Psi^d(\theta)) = E(\psi(X, \theta)) = \int \psi(x, \theta)\, dF(x), \qquad \theta \in \Theta,$$

and

$$\Psi_n(\theta) = (\Psi_n^1(\theta), \dots, \Psi_n^d(\theta)) = \frac{1}{n} \sum_{i=1}^n \psi(X_i, \theta) = \int \psi(x, \theta)\, dF_n(x), \qquad \theta \in \Theta.$$

(C2) Ψ has a unique zero at θ_0; there exists some $\eta > 0$ such that $\overline{B}(\theta_0, \eta) := \{\theta \in \mathbb{R}^d; |\theta - \theta_0| \le \eta\} \subset \Theta$ and Ψ is homeomorphism on $\overline{B}(\theta_0, \eta)$; Ψ is differentiable at θ_0 with nonsingular derivative $A : \mathbb{R}^d \mapsto \mathbb{R}^d$; and $E(|\psi(X, \theta)|^2) < \infty$.

Let $C(\overline{B}(\theta_0, \eta))$ denote the space of continuous $\mathbb{R}^d$-valued functions on $\overline{B}(\theta_0, \eta)$ and define $\|f\| = \sup_{\theta \in \overline{B}(\theta_0, \eta)} |f(\theta)|$ for $f \in C(\overline{B}(\theta_0, \eta))$. Let $\Psi_0(\theta)$ and Ψ_{0n} be the restrictions of Ψ and Ψ_n on $\overline{B}(\theta_0, \eta)$, respectively.

(C3) $\{a(n), n \ge 1\}$ satisfies

$$a(n) \nearrow \infty \quad \text{and} \quad \frac{a(n)}{\sqrt{n}} \searrow 0 \tag{4.15}$$

and $\{\psi(X_i, \theta), i \ge 1\}$ satisfies

$$\frac{\sqrt{n}}{a(n)} \sup_{\theta \in \overline{B}(\theta_0, \eta)} |\Psi_n(\theta) - \Psi(\theta)| \xrightarrow{P} 0 \tag{4.16}$$

and

$$\limsup_{n \to \infty} \frac{1}{a^2(n)} \log\Big(nP\Big(\sup_{\theta \in \overline{B}(\theta_0, \eta)} |\psi(X, \theta)| \ge \sqrt{n} a(n)\Big)\Big) = -\infty. \tag{4.17}$$

REMARK 4.1. Let Y be a random variable taking its values in a Banach space and $E(Y)=0$. If there exists a sequence of increasing nonnegative functions $\{H_k, k \geq 1\}$ on $(0,+\infty)$ satisfying

$$\lim_{u\to\infty} u^{-2}H_k(u)=+\infty, \qquad \lim_{k\to\infty}\lim_{n\to\infty}\frac{1}{a^2(n)}\log\frac{H_k(\sqrt{n}a(n))}{n}=+\infty, \tag{4.18}$$

and

$$E(H_k(\|Y\|))<\infty \qquad \text{for any } k\geq 1, \tag{4.19}$$

then

$$\limsup_{n\to\infty}\frac{1}{a^2(n)}\log\big(nP\big(\|Y\|\geq\sqrt{n}a(n)\big)\big)=-\infty. \tag{4.20}$$

In particular [cf. Chen (1991), Ledoux (1992)], if for each $k \geq 1$,

$$E(\|Y\|^2(\log\|Y\|)^k)<+\infty,$$

then (4.20) holds for $a(n)=\sqrt{\log\log n}$; if for each $k\geq 1$,

$$E(\|Y\|^k)<+\infty,$$

then (4.20) holds for $a(n)=\sqrt{\log n}$; if for some $1\leq p<2$, there exists some $\delta>0$ such that

$$E(\exp\{\delta\|Y\|^{2-p}\})<+\infty, \tag{4.21}$$

then (4.20) holds for $a(n)=o(n^{(2-p)/2p})$; if for some $1<p<2$, and

$$E(\exp\{\delta\|Y\|^{2-p}\})<+\infty \qquad \text{for all } \delta>0, \tag{4.22}$$

then (4.20) holds for $a(n)=O(n^{(2-p)/2p})$.

In fact, by Chebychev's inequality,

$$P\big(\|Y\|>\sqrt{n}a(n)\big)\leq\frac{E(H_k(\|Y\|))}{H_k(\sqrt{n}a(n))}.$$

Hence, (4.18) and (4.19) yield (4.20).

LEMMA 4.1 [See Lemma 4.3 in Heesterman and Gill (1992)]. *Assume that* (C1) *and* (C2) *hold. Then there exists a neighborhood V of Ψ_0 in $C(\overline{B}(\theta_0,\eta))$ and a functional $\Phi: C(\overline{B}(\theta_0,\eta))\mapsto\overline{B}(\theta_0,\eta)$ such that $f(\Phi(f))=0$, $for any f\in V$, and Φ is Hadamard differentiable at Ψ_η with derivative $\Phi'_{\Psi_0}(f)=-A^{-1}f(\theta_0)$.*

THEOREM 4.7. *Suppose that* (C1), (C2) *and* (C3) *hold. Define*

$$\theta_n=\Phi(\Psi_{0n}). \tag{4.23}$$

Then $\{\frac{\sqrt{n}}{a(n)}(\theta_n-\theta_0), n\geq 1\}$ satisfies the LDP with speed $a^2(n)$ and rate function

$$I^M(z)=\tfrac{1}{2}\langle Az,\Gamma^{-1}Az\rangle, \tag{4.24}$$

where Γ *is the covariance of* $\psi(X,\theta_0)-\Phi(\theta_0)$, *and*

$$\limsup_{n\to\infty}\frac{1}{a^2(n)}\log P\big(\Psi_n(\theta_n)\neq 0\big)=-\infty. \tag{4.25}$$

PROOF. Set $T=\{1,\dots,d\}\times\overline{B}(\theta_0,\eta)$. Since

$$T\times T\ni((i,s),(j,t))\mapsto d((i,s),(j,t)):=\big(\mathrm{Var}\big(\psi^i(X_1,t)-\psi^j(X_1,s)\big)\big)^{1/2}$$

is continuous on $T\times T$ and $d((i,t),(i,t))=0$, then (T,d) is totally bounded. Hence, under (C3), Theorem 2.8 in Arcones (2003a) yields that $\{\{\frac{\sqrt{n}}{a(n)}(\Psi_{0n}^i(\theta)-\Psi^i(\theta)),(i,\theta)\in T\},n\geq 1\}$ satisfies the LDP in $l_\infty(T)$ with speed $a^2(n)$ and rate function

$$\hat{I}(f)=\tfrac{1}{2}\inf\big\{E(\alpha^2(X));\ f(i,\theta)=E\big(\alpha(X)\big(\psi^i(X,\theta)-\Psi^i(\theta)\big)\big)\big\}$$

satisfying

$$\limsup_{\lambda\to\infty}\frac{1}{\lambda}\inf\{\hat{I}(f);\ \|f\|\geq\lambda\}=-\infty.$$

Then, applying the classical contraction principle [see Dembo and Zeitouni (1998), Theorem 4.2.1] to $l_\infty(\tilde{T})\ni f\to(f(1,\theta),\dots,f(d,\theta))\in l_\infty(\overline{B}(\theta_0,\eta),\mathbb{R}^d)$, we obtain that $\{\{\frac{\sqrt{n}}{a(n)}(\Psi_{0n}(\theta)-\Psi(\theta)),\theta\in\overline{B}(\theta_0,\eta)\},n\geq 1\}$ satisfies the LDP in $C(\overline{B}(\theta_0,\eta))$ with speed $a^2(n)$ and rate function

$$I(f)=\tfrac{1}{2}\inf\big\{E(\alpha^2(X));\ f(\theta)=E\big(\alpha(X)\big(\psi(X,\theta)-\Phi(\theta)\big)\big)\big\}.$$

Therefore, we have

$$\limsup_{n\to\infty}\frac{1}{a^2(n)}\log P(\Psi_{0n}\notin V)=-\infty,$$

and so (4.25) holds. Then, by Theorem 3.1, we conclude that $\{\frac{\sqrt{n}}{a(n)}(\theta_n-\theta),n\geq 1\}$ satisfies the LDP with speed $a^2(n)$ and rate function

$$\begin{aligned}I^M(z)&=\tfrac{1}{2}\inf\big\{E(\alpha^2(X)),\ E\big(\alpha(X)\big(\psi(X,\theta_0)-\Phi(\theta_0)\big)\big)=-Az\big\}\\&=\tfrac{1}{2}\langle Az,\Gamma^{-1}Az\rangle.\end{aligned}$$

□

REMARK 4.2. Comparing with Theorem 2.8 in Arcones (2002), in Theorem 4.7, we remove the condition

$$\limsup_{n\to\infty}\frac{1}{a^2(n)}\log P(|\theta_n-\theta_0|>\varepsilon)=-\infty,$$

which is required by Arcones (2002).

4.6. *Moderate deviations for L-statistics.* Let $X_{1n} \le X_{2n} \le \cdots \le X_{nn}$ be the order statistics of a random sample $X_1, \dots, X_n$ from a random variable X with distribution function $F(x)$ and let J be a fixed score function on (0, 1). Also let F_n be the empirical distribution function of the sample. We consider the L-statistics of the form

$$L_n := \sum_{i=1}^{n} X_{in} \int_{(i-1)/n}^{i/n} J(u)\,du = \int_0^1 F_n^{-1}(s) J(s)\,ds.$$

Groeneboom, Oosterhoff and Ruymgaart (1979) had obtained some large deviations for L-statistics. The Cramér type moderate deviations for L-statistics had been studied in Vandemaele and Veraverbeke (1982), Bentkus and Zitikis (1990) and Aleskeviciene (1991). In this subsection, we study the moderate deviation principle for L-statistics by the delta method.

Take $\mathcal{X} = l_\infty(\mathbb{R})$ and $\mathcal{Y} = \mathbb{R}$. Let $\mathcal{D}_\Phi$ be the set of all distribution functions on $\mathbb{R}$, and set $\mathcal{D}_0 = \{a(G - F); G \in \mathcal{D}_\Phi, a \in \mathbb{R}\}$. Define $\Phi : \mathcal{D}_\Phi \mapsto \mathbb{R}$ as follows:

$$\Phi(G) = \int_0^1 G^{-1}(s) J(s)\,ds = \int_{-\infty}^{\infty} x J(G(x))\,dG(x).$$

Assume that $E(X^2) < \infty$. Set $m(J, F) = \int_{-\infty}^{\infty} x J(F(x))\,dF(x)$, and

$$\sigma^2(J, F) = \int_{\mathbb{R}^2} J(F(x)) J(F(y)) \big(F(x \wedge y) - F(x)F(y)\big)\,dx\,dy,$$

where $x \wedge y = \min\{x, y\}$. We also assume $\sigma^2(J, F) > 0$.

THEOREM 4.8. *Suppose that the score function J is trimmed near* 0 *and* 1, *that is*, $J(u) = 0, u \in [0, t_1) \cup (t_2, 1]$ *where* $0 < t_1 < t_2 < 1$. *If J is bounded and continuous a.e. Lebesgue measure and a.e.* F^{-1}, *then* $\{\frac{\sqrt{n}}{a(n)}(L_n - m(J, F)), n \ge 1\}$ *satisfies the LDP in $\mathbb{R}$ with speed $a^2(n)$ and rate function* $I^L(x) = \frac{x^2}{2\sigma^2(J,F)}$.

PROOF. By Theorem 1 in Boos (1979), we have

$$\lim_{\|G-F\|\to 0} \frac{|\Phi(G) - \Phi(F) - \int (F(x) - G(x)) J(F(x))\,dx|}{\|G - F\|} = 0.$$

Therefore, for any $t_n \to 0+$ and $H_n \to \alpha \in \mathcal{D}_0$ with $F + t_n H_n \in \mathcal{D}_\Phi$,

$$\lim_{n\to\infty} \left| \frac{|\Phi(F + t_n H_n) - \Phi(F)}{t_n} + \int H_n(x) J(F(x))\,dx \right| = 0,$$

and so, $\Phi : \mathcal{D}_\Phi \mapsto \mathbb{R}$ is Hadamard-differentiable at F tangentially to $\mathcal{D}_0$ with respect to the uniform convergence, and $\Phi'_F(\alpha) = -\int_{\mathbb{R}} \alpha(x) J(F(x))\,dx, \alpha \in \mathcal{D}_0$. By Theorem 3.1, we conclude that $\{\frac{\sqrt{n}}{a(n)}(L_n - m(J, F)), n \ge 1\}$ satisfies the LDP in $\mathbb{R}$ with speed $a^2(n)$ and rate function I^L given by

$$I^L(y) = \inf\left\{ I_F(\alpha); -\int_{\mathbb{R}} \alpha(x) J(F(x))\,dx = y \right\},$$

which equals the rate function of $\{-\frac{\sqrt{n}}{a(n)}\int_{\mathbb{R}}(F_n(x)-F(x))J(F(x))\,dx, n\geq 1\}$, that is, $I^L(y)=\frac{y^2}{2\sigma^2(J,F)}$. $\square$

Now, let us remove the trimming restrictions on J. Set

$$\tilde{\mathcal{D}}_\Phi=\biggl\{\tilde{G}(x)=G(x)I_{(-\infty,0)}(x)+\bigl(G(x)-1\bigr)I_{[0,\infty)}(x);\\ G\in\mathcal{D}_\Phi,\int|x|\,dG(x)<\infty\biggr\}$$

and $\tilde{\mathcal{D}}_0=\{a(\tilde{G}-\tilde{F})\equiv a(G-F); a\in\mathbb{R}, \tilde{G}\in\tilde{\mathcal{D}}_\Phi\}$. Then $\tilde{\mathcal{D}}_\Phi,\tilde{\mathcal{D}}_0\subset L^1(\mathbb{R})$. Define $\tilde{\Phi}:\tilde{\mathcal{D}}_\Phi\mapsto\mathbb{R}$ by $\tilde{\Phi}(\tilde{G})=\Phi(G)$ for all $\tilde{G}\in\tilde{\mathcal{D}}_\Phi$.

LEMMA 4.2. *If J is Lipschitz continuous on* $[0,1]$, *then* $\tilde{\Phi}:\tilde{\mathcal{D}}_\Phi\mapsto\mathbb{R}$ *is Hadamard-differentiable at $\tilde{F}$ tangentially to $\tilde{\mathcal{D}}_0$ with respect to L^1-convergence, and*

$$\tilde{\Phi}'_{\tilde{F}}(\alpha)=-\int_{\mathbb{R}}\alpha(x)J(F(x))\,dx,\qquad \alpha\in\tilde{\mathcal{D}}_0.$$

PROOF. By integration by parts, we can write [cf. Boos (1979), Shao (1989)]

$$\tilde{\Phi}(\tilde{G})-\tilde{\Phi}(\tilde{F})+\int_{\mathbb{R}}\bigl(G(x)-F(x)\bigr)J(F(x))\,dx=R(G,F)\qquad\text{for any }\tilde{G}\in\tilde{\mathcal{D}}_\Phi,$$

where $R(G,F)=\int_{\mathbb{R}}W_{G,F}(x)(G(x)-F(x))\,dx$, and

$$W_{G,F}(x)=\begin{cases}\dfrac{\int_{F(x)}^{G(x)}(J(t)-J(F(x)))\,dt}{G(x)-F(x)}, & \text{if } G(x)\neq F(x),\\ 0, & \text{if } G(x)=F(x).\end{cases}$$

By the Lipschitz continuity of J, there exists a constant $C>0$ such that

$$|R(G,F)|\leq C\int_{\mathbb{R}}\bigl(G(x)-F(x)\bigr)^2\,dx=C\int_{\mathbb{R}}\bigl(\tilde{G}(x)-\tilde{F}(x)\bigr)^2\,dx.$$

For any $t_n\to 0+$ and $H_n\to\alpha\in\tilde{\mathcal{D}}_0$ in $(L^1(\mathbb{R}),\|\cdot\|_{L^1})$ with $\tilde{F}+t_nH_n\in\tilde{\mathcal{D}}_\Phi$, then $|H_n|\leq 2/t_n$ and

$$\int_{\mathbb{R}}|H_n(x)-\alpha(x)|^2\,dx\leq(\|\alpha\|+2/t_n)\int_{\mathbb{R}}|H_n(x)-\alpha(x)|\,dx,$$

where $\|\alpha\|=\sup_{x\in\mathbb{R}}|\alpha(x)|$. Therefore,

$$\frac{1}{t_n}\int_{\mathbb{R}}\bigl(\tilde{F}(x)+t_nH_n(x)-\tilde{F}(x)\bigr)^2\,dx\\ \leq 2t_n\int_{\mathbb{R}}|H_n(x)-\alpha(x)|^2\,dx+2t_n\int_{\mathbb{R}}|\alpha(x)|^2\,dx\to 0,$$

and so

$$\lim_{n\to\infty}\left|\frac{\tilde{\Phi}(\tilde{F}+t_nH_n)-\tilde{\Phi}(\tilde{F})}{t_n}+\int H_n(x)J(F(x))\,dx\right|=0,$$

which yields that $\tilde{\Phi}$ is Hadamard-differentiable at $\tilde{F}$ tangentially to $\tilde{\mathcal{D}}_0$ with respect to L^1-convergence, and $\tilde{\Phi}'_{\tilde{F}}(\alpha)=-\int_{\mathbb{R}}\alpha(x)J(F(x))\,dx$. $\square$

LEMMA 4.3. *Let X be a random variable with values in a separable Banach space $\mathbf{B}$ and $E(\|X\|^2)<\infty$. Then $(\mathbf{B}_1^*,d)$ is totally bounded, where $\mathbf{B}_1^*$ is the unit ball of the dual space $\mathbf{B}^*$ of $\mathbf{B}$, and*

$$d(g,h)=\bigl(E\bigl(\bigl(g\bigl(X-E(X)\bigr)-h\bigl(X-E(X)\bigr)\bigr)^2\bigr)\bigr)^{1/2},\qquad g,h\in\mathbf{B}_1^*.$$

PROOF. Noting $|g(X-E(X))-h(X-E(X))|\le 2\|X-E(X)\|$ for all $g,h\in\mathbf{B}_1^*$ and $E(\|X-E(X)\|^2)<\infty$, by the dominated convergence theorem, we know that the function $(g,h)\mapsto d(g,h)$ is continuous from $\mathbf{B}_1^*\times\mathbf{B}_1^*$ to $\mathbb{R}$ with respect to w^*-topology. Let d^* denote a compatible metric on (B_1^*,w^*). Since $\mathbf{B}_1^*$ is w^*-compact and $d(g,g)=0$, then, for any $\varepsilon>0$, there exists some $\delta>0$ such that $d(g,h)<\varepsilon$, if $d^*(g,h)<\delta$. Choose finite points $h_1,\dots,h_m\in\mathbf{B}_1^*$ such that $B_1^*\subset\bigcup_{i=1}^m\{g;d^*(g,h_i)<\delta\}$, then $B_1^*\subset\bigcup_{i=1}^m\{g;d(g,h_i)<\varepsilon\}$. Therefore, $(\mathbf{B}_1^*,d)$ is totally bounded. $\square$

Define

$$\Lambda_{2,1}(X)=\int_0^\infty\sqrt{P(|X|>t)}\,dt.$$

Then [cf. del Barrio, Giné and Matrán (1999), page 1014], $\Lambda_{2,1}(X)<\infty$ if and only if $\int_{-\infty}^{\infty}\sqrt{F(x)(1-F(x))}\,dx<\infty$.

LEMMA 4.4. *Assume that $\Lambda_{2,1}(X)<\infty$. If (4.15) holds and*

$$\limsup_{n\to\infty}\frac{1}{a^2(n)}\log\bigl(nP\bigl(|X|\ge\sqrt{n}a(n)\bigr)\bigr)=-\infty,\tag{4.26}$$

then $\{\frac{\sqrt{n}}{a(n)}(F_n-F)=\frac{\sqrt{n}}{a(n)}(\tilde{F}_n-\tilde{F}),n\ge1\}\subset\tilde{\mathcal{D}}_0$ satisfies the LDP in $(L^1(\mathbb{R}),\|\cdot\|_{L^1})$ with speed $a^2(n)$ and rate function I_F.

PROOF. Set $\xi_i=I_{\{X_i\le x\}}-F(x),x\in\mathbb{R}$, then

$$\|\xi_i\|_{L^1}=2\biggl(X_iF(X_i)-\int_{(-\infty,X_i)}x\,dF(x)\biggr).$$

Therefore, the condition of the lemma implies

$$\limsup_{n\to\infty}\frac{1}{a^2(n)}\log\bigl(nP\bigl(\|\xi_1\|_{L^1}\ge\sqrt{n}a(n)\bigr)\bigr)=-\infty,$$

and by Theorem 2.1(b) of del Barrio, Giné and Matrán (1999), we also have $\frac{1}{a(n)}\|\sum_{i=1}^n \xi_i\|_{L^1} \xrightarrow{P} 0$. By Lemma 4.3, $(\mathbf{B}_1^*, d)$ is totally bounded, where

$$\mathbf{B}_1^* := \Big\{g \in L^\infty; \|g\|_\infty := \operatorname*{esssup}_{x\in\mathbb{R}} |g(x)| \le 1\Big\}$$

and

$$d(g,h) = \left(E\left(\left(\int_{\mathbb{R}} (g(x)-h(x))\xi_1(x)\,dx\right)^2\right)\right)^{1/2}.$$

Therefore, by Theorem 2.8 in Arcones (2003a), the conclusion of the lemma holds. □

By Lemmas 4.4 and 4.2 and Theorem 3.1, we obtain the following result.

THEOREM 4.9. *Assume that* $\Lambda_{2,1}(X) < \infty$, (4.15) *and* (4.26) *hold. If J is Lipschitz continuous on* $[0, 1]$, *then* $\{\frac{\sqrt{n}}{a(n)}(L_n - m(J, F)), n \ge 1\}$ *satisfies the LDP in* $\mathbb{R}$ *with speed* $a^2(n)$ *and rate function* $I^L(x) = \frac{x^2}{2\sigma^2(J,F)}$.

REMARK 4.3. From Remark 4.1, the moment condition in Theorem 4.9 is weaker than the conditions given in Vandemaele and Veraverbeke (1982), Bentkus and Zitikis (1990) and Aleskeviciene (1991). In particular, if $E(|X|^{2+\delta}) < \infty$ and $a(n) = \sqrt{\log\log n}$, then the condition of Lemma 4.4 is valid, and so, for any $r > 0$,

$$\lim_{n\to\infty} \frac{1}{\log\log n} \log P\left(\sqrt{\frac{n}{\log\log n}}|L_n - m(J, F)| \ge r\right) = -\frac{r^2}{2\sigma^2(J, F)}.$$

5. Application: Statistical hypothesis testing. In this section, we applied the moderate deviations to hypothesis testing problems. We only consider the right-censored data model. The method can be applied to other models.

Let F be the unknown distribution function in the right-censored data model considered in Section 4.2 and let $\hat{F}_n$ be the Kaplan–Meier estimator of F. Consider the following hypothesis testing:

$$H_0 : F = F_0 \quad \text{and} \quad H_1 : F = F_1,$$

where F_0 and F_1 are two distribution functions such that $F_0(x_0) \ne F_1(x_0)$ for some $x_0 \in [0, \tau]$. Similar to the Kolmogorov–Smirnov test, we take the Kaplan–Meier statistic $T_n := \sup_{x\in[0,\tau]} |\hat{F}_n(x) - F_0(x)|$ as test statistic. Suppose that the rejection region for testing the null hypothesis H_0 against H_1 is $\{\frac{\sqrt{n}}{a(n)} T_n \ge c\}$, where c is a positive constant. Then the probability α_n of Type I error and the probability β_n of Type II error are

$$\alpha_n = P\left(\frac{\sqrt{n}}{a(n)} T_n \ge c \,\Big|\, F = F_0\right) \quad \text{and} \quad \beta_n = P\left(\frac{\sqrt{n}}{a(n)} T_n < c \,\Big|\, F = F_1\right),$$

respectively. Then

$$\beta_n \leq P\left(\frac{\sqrt{n}}{a(n)} \sup_{x\in[0,\tau]} |\hat{F}_n(x) - F_1(x)| \right.$$
$$\left. \geq \frac{\sqrt{n}}{a(n)} \sup_{x\in[0,\tau]} |F_0(x) - F_1(x)| - c \middle| F = F_1 \right).$$

Therefore, Theorem 4.4 implies that

$$\lim_{n\to\infty} \frac{1}{a^2(n)} \log \alpha_n = -\frac{c^2}{2\sigma_{KM}^2}, \qquad \lim_{n\to\infty} \frac{1}{a^2(n)} \log \beta_n = -\infty,$$

where

$$\sigma_{KM}^2 = \sup_{t\in[0,\tau]} (1 - F_0(t))^2 \int_{(0,t]} \frac{1}{(1 - \Delta\Lambda(u))\overline{H}_0(u)} \, d\Lambda_0(u),$$

$$\Lambda_0(t) = \int_{[0,t]} \frac{1}{1 - F_0(s-)} \, dF_0(s), \qquad \overline{H}_0(t) = P(Z \geq t | F = F_0),$$

and Z is as defined in Section 4.2.

The above result tells us that if the rejection region for the test is $\{\frac{\sqrt{n}}{a(n)} T_n \geq c\}$, then the probability of Type I error tends to 0 with decay speed

$$\exp\{-c^2 a^2(n)/(2\sigma_{KM}^2)\},$$

and the probability of Type II error tends to 0 with decay speed $\exp\{-ra^2(n)\}$ for all $r > 0$.

6. Concluding remarks. This article discussed the large deviations of transformed statistics. For the problem, an extended contraction principle was developed and a general delta method in large deviation theory was proposed. The new method was used to establish the moderate deviation principles for the Wilcoxon statistic, the Kaplan–Meier estimator, the empirical quantile estimator and the empirical copula estimator, which have not been addressed in the literature. The proposed method was also used to improve the existing moderate deviation results for M-estimators and L-statistics, where our proofs are different from others but simpler by the new method. Moreover, our moderate deviation results are very useful for statistical hypothesis testing. As shown in Section 5, a moderate deviation result can be used to construct a test of a statistical hypothesis such that the probabilities of both Type I and Type II errors tend to 0 with an exponentially decay speed as $n \to \infty$.

Note that the asymptotics for multivariate trimming and general Z-estimators have been studied by using the delta method in a weak convergence; see Nolan (1992) and van der Vaart and Wellner (1996). Similar to those presented in Section 4, the moderate deviations for these estimators can be established by using the proposed delta method in large deviations.

These applications show that the proposed method is very powerful for deriving moderate deviation principles on estimators. The method will play an important role in large sample theory of statistics like the functional delta method in weak convergence. Theoretically speaking, we can apply the proposed delta method to obtain moderate deviations for estimators where the classical delta method can be applied.

APPENDIX: PROOF OF THE EXTENDED CONTRACTION PRINCIPLE

Step 1. First of all, let us prove $\{I < \infty\} \subset \mathcal{D}_\infty$, where $\mathcal{D}_\infty$ denotes the set of all x for which there exists a sequence x_n with $x_n \in \mathcal{D}_n$ and $x_n \to x$.

In fact, by the definition of $\mathcal{D}_\infty$, $x \in \mathcal{D}_\infty$ if and only if for any $k \geq 1$, there exists a positive integer n_k such that $B_d(x, 1/k) \cap \mathcal{D}_n \neq \varnothing$ for all $n \geq n_k$, where $B_d(x, 1/k) = \{y \in \mathcal{X}; d(y, x) < 1/k\}$. Therefore, for any $x \notin \mathcal{D}_\infty$, there exist an open neighborhood U of x and a subsequence $\{\mathcal{D}_{n_k}, k \geq 1\}$ such that $\mathcal{D}_{n_k} \cap U = \varnothing$ for all $k \geq 1$. Then by the lower bound of the large deviations for $\{X_n, n \geq 1\}$, we have

$$-\infty = \liminf_{k\to\infty} \frac{1}{\lambda(n_k)} \log P_{n_k *}(X_{n_k} \in U) \geq -I(x),$$

which implies $\{I < \infty\} \subset \mathcal{D}_\infty$, where $P_{n_k *}$ is the inner measure corresponding to P_{n_k} as defined in Section 2.

Step 2. Let us prove that if some subsequence $x_{n_k} \to x \in \{I < \infty\}$ with $x_{n_k} \in \mathcal{D}_{n_k}$, then $f_{n_k}(x_{n_k}) \to f(x)$ and the restriction of the function f to $\{I < \infty\}$ is continuous.

The proof is similar to that of the extended mapping theorem [see Theorem 1.11.1 in van der Vaart and Wellner (1996)], which is given below. Let a subsequence $x_{n_k} \to x \in \{I < \infty\}$ be given. Since $x \in \mathcal{D}_\infty$, there exists a sequence $y_n \to x$ with $y_n \in \mathcal{D}_n$ for each $n \geq 1$. Define $x_n = x_n I_{\{n_k, k\geq 1\}}(n) + y_n I_{\{n_k, k\geq 1\}^c}(n)$. Then $x_n \in \mathcal{D}_n$ for each $n \geq 1$ and $x_n \to x$. Therefore, by condition (ii), $f_n(x_n) \to f(x)$, and so $f_{n_k}(x_{n_k}) \to f(x)$. To prove the continuity of f on $\{I < \infty\}$, let $x_m \to x$ in $\{I < \infty\}$. For every m, there is a sequence $x_{m,n} \in \mathcal{D}_n$ with $x_{m,n} \to x_m$ as $n \to \infty$. Since $x_m \in \{I < \infty\}$, then $f_n(x_{m,n}) \to f(x_m)$ as $n \to \infty$. For every m, take n_m such that n_m is increasing with m satisfying $d(x_{m,n_m}, x_m) < 1/m$ and $\rho(f_{n_m}(x_{m,n_m}), f(x_m)) < 1/m$. Then $x_{m,n_m} \to x$, and by the first conclusion in Step 2, $f_{n_m}(x_{m,n_m}) \to f(x)$ as $m \to \infty$. This yields $f(x_m) \to f(x)$.

Step 3. Let us prove that $[I_f \leq L] = f([I \leq L])$ for any $L \geq 0$ and I_f is inf-compact, that is, for any $L \in [0, +\infty)$, $[I_f \leq L]$ is compact. This can be shown by the continuity of $f|_{\{I<\infty\}}$ obtained in Step 2.

Step 4. Next, we show the upper bound of large deviations.

Let F be a closed subset in $\mathcal{Y}$. Then, using the arguments similar to the proof of the extended continuous mapping theorem [see Theorem 1.11.1 in van der Vaart and Wellner (1996)],

$$\bigcap_{n=1}^{\infty} \overline{\bigcup_{m=n}^{\infty} f_m^{-1}(F)} \subset f^{-1}(F) \cup (\{I < \infty\})^c. \tag{A.1}$$

Now for every fixed k, by the large deviation principle of $\{X_n, n \geq 1\}$, for each $L > 0$, there exists a compact subset K_L such that for any $\delta > 0$,

$$\limsup_{n\to\infty} \frac{1}{\lambda(n)} \log P_n^*\big(X_n \in (K_L^\delta)^c\big) \leq -L,$$

and so

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{\lambda(n)} \log P_n^*\big(f_n(X_n) \in F\big) \\
&\quad \leq \limsup_{n\to\infty} \frac{1}{\lambda(n)} \log P_n^*\left(X_n \in \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F)}\right) \\
&\quad \leq \max\left\{- \inf_{x \in \overline{K_L^\delta} \cap \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F)}} I(x), -L\right\},
\end{aligned}$$

where $K_L^\delta = \{y; d(y,x) < \delta \text{ for some } x \in K_L\}$ and P_n^* is the outer measure corresponding to P_n as defined in Section 2. Since K_L is compact and I is lower semi-continuous, then, when $\delta \downarrow 0$,

$$\inf_{x \in \overline{K_L^\delta} \cap \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F)}} I(x) \uparrow \inf_{x \in K_L \cap \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F)}} I(x).$$

Hence it follows that

$$\limsup_{n\to\infty} \frac{1}{\lambda(n)} \log P_n^*\big(f_n(X_n) \in F\big) \leq \max\left\{- \inf_{x \in K_L \cap \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F)}} I(x), -L\right\}.$$

Choose a sequence $x_k \in K_L \cap \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F)}$, $k \geq 1$ such that $I(x_k) = \inf_{x \in K_L \cap \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F)}} I(x)$, and then choose a subsequence $\{x_{k_m}, m \geq 1\}$ and $x_0 \in K_L$ such that $x_{k_m} \to x_0$. Then we have

$$x_0 \in K_L \cap \left(\bigcap_{k=1}^{\infty} \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F)}\right) \subset K_L \cap \big(f^{-1}(F) \cup (\{I < \infty\})\big).$$

Letting $k \to \infty$, we have

$$\liminf_{k\to\infty} I(x_k) \geq I(x_0) \geq \inf_{x \in K_L \cap (f^{-1}(F) \cup (\{I<\infty\})^c)} I(x) \geq \inf_{x \in f^{-1}(F)} I(x).$$

Now letting $L \to \infty$, we conclude that

$$\limsup_{n\to\infty} \frac{1}{\lambda(n)} \log P_n^*\big(f_n(X_n) \in F\big) \leq - \inf_{x \in f^{-1}(F)} I(x) = - \inf_{x \in F} I_f(x).$$

Step 5. Finally, we show the lower bound of large deviations: for any $y_0 \in \mathcal{Y}$ with $I_f(y_0) < \infty$,

$$\liminf_{n\to\infty} \frac{1}{\lambda(n)} \log P_{n*}\big(f_n(X_n) \in B(y_0, \delta)\big) \geq -I_f(y_0).$$

For any $a > I_f(y_0)$, there is some $x_0 \in \mathcal{X}$ with $f(x_0) = y_0$ and $I(x_0) < a$. For any $\delta > 0$, set $B(\delta) = B_\rho(y_0, \delta) = \{y \in \mathcal{Y}; \rho(y_0, y) < \delta\}$ and $F(\delta) = B(\delta)^c$. Then, by (A.1), we have

$$\bigcup_{n=1}^{\infty} \overline{\bigcup_{m=n}^{\infty} f_m^{-1}(F(\delta))}^{c} \supset f^{-1}(B(\delta)) \cap (\{I < \infty\}) \ni x_0. \tag{A.2}$$

Now for every fixed k, by the large deviation principle of $\{X_n\}$, we have

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{\lambda(n)} \log P_{n*}\big(f_n(X_n) \in B(\delta)\big) \\
&\qquad \geq \liminf_{n\to\infty} \frac{1}{\lambda(n)} \log P_{n*}\left(X_n \in \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F(\delta))}^{c}\right) \geq - \inf_{x \in \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F(\delta))}^{c}} I(x).
\end{aligned}$$

Since $x_0 \in f^{-1}(B(\delta)) \subset \bigcup_{n=1}^{\infty} \overline{\bigcup_{m=n}^{\infty} f_m^{-1}(F(\delta))}^{c}$, there is some $k \geq 1$ such that $x_0 \in \overline{\bigcup_{m=k}^{\infty} f_m^{-1}(F(\delta))}^{c}$. Therefore,

$$\liminf_{n\to\infty} \frac{1}{\lambda(n)} \log P_{n*}\big(f_n(X_n) \in B(\delta)\big) \geq -I(x_0) > -a.$$

Letting $a \downarrow I_f(y_0)$, we obtain the lower bound of large deviations.

REMARK A.1. When $\mathcal{D}_n = \mathcal{X}$ for all $n \geq 1$, the continuity of f can be proved directly by the following property [see Theorem 2.1 in Arcones (2003a)]: Given $\varepsilon > 0$, for any $x_0 \in \{I < \infty\}$, there are $\delta > 0$ and a positive integer n_0 such that for all $n \geq n_0$, $f_n(B(x_0, \delta)) \subset B(f(x_0), \varepsilon)$. However, when $\mathcal{D}_n \neq \mathcal{X}$, $f_n(B(x_0, \delta))$ is not well defined since $B(x_0, \delta) \not\subset \mathcal{D}_n$. Thus, the above property cannot be used for proving the continuity of f in this case.

Acknowledgments. The authors are very grateful to the Editors, Professors Tony Cai and Bernard W. Silverman, the Associate Editor and the three referees for their many valuable comments and suggestions that greatly improved the paper. The authors also wish to thank Professor Qiman Shao and Professor Liming Wu for their useful discussions and suggestions.

REFERENCES

ALESKEVICIENE, A. K. (1991). Large and moderate deviations for L-statistics. *Lithuanian Math. J.* **31** 145–156.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York. MR1198884

ARCONES, M. A. (2002). Moderate deviations for M-estimators. *Test* **11** 465–500. MR1947607

ARCONES, M. A. (2003a). Moderate deviations of empirical processes. In *Stochastic Inequalities and Applications* (E. Giné, C. Houdré and D. Nualart, eds.). *Progress in Probability* **56** 189–212. Birkhäuser, Basel. MR2073434

ARCONES, M. A. (2003b). The large deviation principle for stochastic processes. I. *Theory Probab. Appl.* **47** 567–583. MR2001788

ARCONES, M. A. (2004). The large deviation principle for stochastic processes. II. *Theory Probab. Appl.* **48** 19–44. MR2013408

ARCONES, M. A. (2006). Large deviations for M-estimators. *Ann. Inst. Statist. Math.* **58** 21–52. MR2281205

BAHADUR, R. R. (1967). Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* **38** 303–324. MR0207085

BAHADUR, R. R. and ZABELL, S. L. (1979). Large deviations of the sample mean in general vector spaces. *Ann. Probab.* **7** 587–621. MR0537209

BENTKUS, V. and ZITIKIS, R. (1990). Probabilities of large deviations for L-statistics. *Lithuanian Math. J.* **30** 215–222.

BERCU, B. (2001). On large deviations in the Gaussian autoregressive process: Stable, unstable and explosive cases. *Bernoulli* **7** 299–316. MR1828507

BITOUZÉ, D., LAURENT, B. and MASSART, P. (1999). A Dvoretzky–Kiefer–Wolfowitz type inequality for the Kaplan–Meier estimator. *Ann. Inst. H. Poincaré Probab. Statist.* **35** 735–763. MR1725709

BOOS, D. D. (1979). A differential for L-statistics. *Ann. Statist.* **7** 955–959. MR0536500

BOROVKOV, A. A. and MOGUL′SKII, A. A. (1992). Large deviations and testing statistical hypotheses. II. Large deviations of maximum points of random fields. *Siberian Adv. Math.* **2** 43–72. MR1209905

CHEN, X. (1991). Probabilities of moderate deviations of independent random vectors in a Banach space. *Chinese J. Appl. Probab. Statist.* **7** 24–32. MR1204537

DEL BARRIO, E., GINÉ, E. and MATRÁN, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *Ann. Probab.* **27** 1009–1071. MR1698999

DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. *Applications of Mathematics* (*New York*) **38**. Springer, New York. MR1619036

DEUSCHEL, J.-D. and STROOCK, D. W. (1989). *Large Deviations*. *Pure and Applied Mathematics* **137**. Academic Press, Boston, MA. MR0997938

DINWOODIE, I. H. (1993). Large deviations for censored data. *Ann. Statist.* **21** 1608–1620. MR1241281

DJELLOUT, H., GUILLIN, A. and WU, L. (2006). Moderate deviations of empirical periodogram and non-linear functionals of moving average processes. *Ann. Inst. H. Poincaré Probab. Statist.* **42** 393–416. MR2242954

ERMAKOV, M. S. (2008). On semiparametric statistical inferences in the moderate deviation zone. *J. Math. Sci.* **152** 869–874.

FU, J. C. (1982). Large sample point estimation: A large deviation theory approach. *Ann. Statist.* **10** 762–771. MR0663430

GAO, F. (2001). Moderate deviations for the maximum likelihood estimator. *Statist. Probab. Lett.* **55** 345–352. MR1877638

GAO, F. (2003). Moderate deviations and large deviations for kernel density estimators. *J. Theoret. Probab.* **16** 401–418. MR1982035

GILL, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method (Part 1). *Scand. J. Statist.* **16** 97–128. MR1028971

GROENEBOOM, P., OOSTERHOFF, J. and RUYMGAART, F. H. (1979). Large deviation theorems for empirical probability measures. *Ann. Probab.* **7** 553–586. MR0537208

HE, X. and SHAO, Q.-M. (1996). Bahadur efficiency and robustness of Studentized score tests. *Ann. Inst. Statist. Math.* **48** 295–314. MR1405934

HEESTERMAN, C. C. and GILL, R. D. (1992). A central limit theorem for M-estimators by the von Mises method. *Statist. Neerlandica* **46** 165–177. MR1178477

HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73–101. MR0161415

INGLOT, T. and KALLENBERG, W. C. M. (2003). Moderate deviations of minimum contrast estimators under contamination. *Ann. Statist.* **31** 852–879. MR1994733

INGLOT, T. and LEDWINA, T. (1990). On probabilities of excessive deviations for Kolmogorov–Smirnov, Cramér–von Mises and chi-square statistics. *Ann. Statist.* **18** 1491–1495. MR1062723

JOUTARD, C. (2004). Large deviations for M-estimators. *Math. Methods Statist.* **13** 179–200. MR2090471

JUREČKOVÁ, J., KALLENBERG, W. C. M. and VERAVERBEKE, N. (1988). Moderate and Cramér-type large deviation theorems for M-estimators. *Statist. Probab. Lett.* **6** 191–199. MR0919374

KALLENBERG, W. C. M. (1983). On moderate deviation theory in estimation. *Ann. Statist.* **11** 498–504. MR0696062

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481. MR0093867

KESTER, A. D. M. and KALLENBERG, W. C. M. (1986). Large deviations of estimators. *Ann. Statist.* **14** 648–664. MR0840520

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. Springer, New York. MR2724368

LEDOUX, M. (1992). Sur les déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi. *Ann. Inst. H. Poincaré Probab. Statist.* **28** 267–280. MR1162575

LEI, L. and WU, L. (2005). Large deviations of kernel density estimator in $L^1(\mathbb{R}^d)$ for uniformly ergodic Markov processes. *Stochastic Process. Appl.* **115** 275–298. MR2111195

LOUANI, D. (1998). Large deviations limit theorems for the kernel density estimator. *Scand. J. Statist.* **25** 243–253. MR1614292

MASSART, P. (1990). The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. *Ann. Probab.* **18** 1269–1283. MR1062069

NIKITIN, Y. (1995). *Asymptotic Efficiency of Nonparametric Tests*. Cambridge Univ. Press, Cambridge. MR1335235

NOLAN, D. (1992). Asymptotics for multivariate trimming. *Stochastic Process. Appl.* **42** 157–169. MR1172513

PUHALSKII, A. and SPOKOINY, V. (1998). On large-deviation efficiency in statistical inference. *Bernoulli* **4** 203–272. MR1632979

REEDS, J. A. (1976). On the definition of von Mises functionals. Ph.D. dissertation, Dept. Statistics, Harvard Univ., Cambridge, MA.

RÖMISCH, W. (2005). Delta method, infinite dimensional. In *Encyclopedia of Statistical Sciences*, 2nd ed. (S. Kotz, C. B. Read, N. Balakrishnan and B. Vidakovic, eds.). Wiley, New York.

SANOV, I. N. (1957). On the probability of large deviations of random magnitudes. *Mat. Sb. N. S.* **42 (84)** 11–44. MR0088087

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. MR0595165

SHAO, J. (1989). Functional calculus and asymptotic theory for statistical analysis. *Statist. Probab. Lett.* **8** 397–405. MR1040799

SIEDERS, A. and DZHAPARIDZE, K. (1987). A large deviation result for parameter estimators and its application to nonlinear regression analysis. *Ann. Statist.* **15** 1031–1049. MR0902244 van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes with Applications to Statistics*. Springer, New York. MR1385671

Vandemaele, M. and Veraverbeke, N. (1982). Cramér type large deviations for linear combinations of order statistics. *Ann*. *Probab*. **10** 423–434. MR0647514

Wellner, J. A. (2007). On an exponential bound for the Kaplan–Meier estimator. *Lifetime Data Anal*. **13** 481–496. MR2394284

Worms, J. (2001). Moderate deviations of some dependent variables. II. Some kernel estimators. *Math*. *Methods Statist*. **10** 161–193. MR1851746

Wu, L. M. (1994). Large deviations, moderate deviations and LIL for empirical processes. *Ann*. *Probab*. **22** 17–27. MR1258864

School of Mathematics and Statistics
Wuhan University
Wuhan 430072
China
E-mail: fqgao@whu.edu.cn

Department of Applied Mathematics
Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
China
and
School of Statistics and Mathematics
Zhongnan University of Economics and Law
Wuhan 430073
China
E-mail: xingqiu.zhao@polyu.edu.hk