entropy

MDPI

Article

Rare Event Analysis for Minimum Hellinger Distance Estimators via Large Deviation Theory

Anand N. Vidyashankar [1,*] and Jeffrey F. Collamore [2]

[1] Department of Statistics, George Mason University, Fairfax, VA 22030, USA
[2] Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark; collamore@math.ku.dk
* Correspondence: avidyash@gmu.edu

Abstract: Hellinger distance has been widely used to derive objective functions that are alternatives to maximum likelihood methods. While the asymptotic distributions of these estimators have been well investigated, the probabilities of rare events induced by them are largely unknown. In this article, we analyze these rare event probabilities using large deviation theory under a potential model misspecification, in both one and higher dimensions. We show that these probabilities decay exponentially, characterizing their decay via a "rate function" which is expressed as a convex conjugate of a limiting cumulant generating function. In the analysis of the lower bound, in particular, certain geometric considerations arise that facilitate an explicit representation, also in the case when the limiting generating function is nondifferentiable. Our analysis involves the modulus of continuity properties of the affinity, which may be of independent interest.

Keywords: Hellinger distance; large deviations; divergence measures; rare event probabilities

Citation: Vidyashankar, A.N.; Collamore, J.F. Rare Event Analysis for Minimum Hellinger Distance Estimators via Large Deviation Theory. *Entropy* **2021**, *23*, 386. https://doi.org/10.3390/e23040386

Academic Editor: Leandro Pardo

Received: 22 February 2021
Accepted: 15 March 2021
Published: 24 March 2021

1. Introduction

In a variety of applications, the use of divergence-based inferential methods is gaining momentum, as these methods provide robust alternatives to traditional maximum likelihood-based procedures. Since the work of [1,2], divergence-based methods have been developed for various classes of statistical models. A comprehensive treatment of these ideas is available, for instance, in [3,4]. The objective of this paper is to study the large deviation tail behavior of the minimum divergence estimators and, more specifically, the minimum Hellinger distance estimators (MHDE).

To describe the general problem, suppose $\Theta \subset \mathbb{R}^d$, and let $\mathfrak{F} = \{f_{\boldsymbol{\theta}}(\cdot) : \boldsymbol{\theta} \in \Theta\}$ denote a family of densities indexed by $\boldsymbol{\theta}$. Let $\{X_n : n \geq 1\}$ denote a class of i.i.d. random variables, postulated to have a continuous density with respect to Lebesgue measure and belonging to the family $\mathfrak{F}$, and let X be a generic element of this class. We denote by $g(\cdot)$ the true density of X.

Before providing an informal description of our results, we begin by recalling that the square of the Hellinger distance (SHD) between two densities $h_1(\cdot)$ and $h_2(\cdot)$ on $\mathbb{R}$ is given by

$$\mathrm{HD}^2(h_1, h_2) = \left\| h_1^{\frac{1}{2}} - h_2^{\frac{1}{2}} \right\|_2^2 = 2 - 2\int_{\mathbb{R}} (h_1(x)h_2(x))^{\frac{1}{2}} dx.$$

The quantity $\int_{\mathbb{R}} (h_1(x)h_2(x))^{\frac{1}{2}} dx$ is referred to as the *affinity* between $h_1(\cdot)$ and $h_2(\cdot)$ and denoted by $\mathscr{A}(h_1, h_2)$. Hence, the SHD between the postulated density and the true density is given by $\mathrm{SHD}(\boldsymbol{\theta}) = \mathrm{HD}^2(f_{\boldsymbol{\theta}}, g)$. When Θ is compact, it is known that there exists a unique $\boldsymbol{\theta}_g \in \Theta$ minimizing the $\mathrm{SHD}(\boldsymbol{\theta})$. Furthermore, when $g(\cdot) = f_{\boldsymbol{\theta}_0}(\cdot)$ and $\mathfrak{F}$ satisfies an identifiability condition, it is well known that $\boldsymbol{\theta}_g$ coincides with $\boldsymbol{\theta}_0$; cf. [1]. Turning to the

sample version, we replace $g(\cdot)$ by $g_n(\cdot)$ in the definition of SHD, obtaining the objective function $\mathrm{SHD}_n(\boldsymbol{\theta}) = \mathrm{HD}^2(f_{\boldsymbol{\theta}}, g_n)$ and

$$g_n(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b_n}\right), \tag{1}$$

where the kernel $K(\cdot)$ is a probability density function and $b_n \searrow 0$ and $nb_n \nearrow \infty$ as $n \to \infty$.

It is known that when the parameter space Θ is compact, there exists a unique $\hat{\boldsymbol{\theta}}_n \in \Theta$ minimizing $\mathrm{SHD}_n(\boldsymbol{\theta})$, and that $\hat{\boldsymbol{\theta}}_n$ converges almost surely to $\boldsymbol{\theta}_g$ as $n \to \infty$; cf. [1]. Furthermore, under some natural assumptions,

$$n^{\frac{1}{2}}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_g) \xrightarrow{d} G, \tag{2}$$

where, under the probability measure associated with $g(\cdot)$, G is a Gaussian random vector with mean vector $\mathbf{0}$ and covariance matrix Σ_g. If $g(\cdot) = f_{\boldsymbol{\theta}_0}(\cdot)$, then the variance of G coincides with the inverse of the Fisher information matrix $\mathfrak{I}(\boldsymbol{\theta}_0)$, yielding statistical efficiency. When the true distribution $g(\cdot)$ does not belong to $\mathfrak{F}$, we will call this the "model misspecifed case," while when $g \in \mathfrak{F}$, we will say that the "postulated model" holds.

In this paper, we focus on the large deviation behavior of $\{\hat{\boldsymbol{\theta}}_n : n \geq 1\}$; namely, the asymptotic probability that the estimate $\hat{\boldsymbol{\theta}}_n$ will achieve values within a set *away* from the central tendency described in (2). We establish results of the form

$$\log \boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n \in B) \approx -n \inf_{\boldsymbol{\theta} \in B} I(\boldsymbol{\theta}), \tag{3}$$

for some "rate function" I and given Borel subset $B \subset \Theta$. Similar large deviation estimates for maximum likelihood estimators (MLE) have been investigated in [5–7], and for general M-estimators in [8,9]. These results allow for a precise description of the probabilities of Type I and Type II error in both the Neymann–Pearson and likelihood ratio test frameworks. Furthermore, large deviation bounds allow one to identify the best exponential rate of decrease of Type II error amongst all tests that satisfy a bound on the Type I error, as in Stein's lemma (cf. [10]). Additional evidence of the importance of large deviation results for statistical inference has been described in [11] and in the book [12].

One of our initial goals was to derive sharp probability bounds for Type I and Type II error in the context of robust hypothesis testing using Hellinger deviance tests. This article is a first step towards this endeavor. A key issue that distinguishes our work from earlier works is that, in our case, the objective function is a nonlinear function of the smoothed empirical measure, and the analysis of this case requires more involved methods compared with those currently existing in the statistical literature on large deviations. Consistent with large deviation analysis more generally, we identify the rate function I as the convex conjugate of a certain limiting cumulant generating function, although in our problem, we uncover a subtle asymmetry between the upper and lower bounds when our limiting generating function is nondifferentiable. In the classical large deviation literature, similar asymmetries have been studied in other one-dimensional contexts (e.g. [13]), although the statistical problem is still quite different, as the dependence on the parameter $\boldsymbol{\theta}$ arises explicitly—inhibiting the use of convexity methods typically exploited in the large deviation literature—and hence requiring novel techniques.

1.1. Large Deviations

In this subsection we provide relevant definitions and properties from large deviation theory required in the sequel. In the following, $\mathbb{R}_+$ will denote the set of non-negative real numbers.

Definition 1. *A collection of probability distributions* $\{P_n : n \geq 1\}$ *on a topological space* $(\mathscr{X}, \mathscr{B})$ *is said to satisfy the weak large deviation principle if*

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(F) \leq -\inf_{x\in F} I(x), \quad \text{for all closed } F \in \mathscr{B},$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \geq -\inf_{x\in G} I(x) \quad \text{for all open sets } F \in \mathscr{B}$$

for some lower semicontinuous function $I : \mathscr{X} \to [0, \infty]$. *The function* I *is called the* rate function. *If the level sets of* I *are compact, we call* I *a* good rate function *and we say that* $\{P_n\}$ *satisfies the* large deviation principle *(LDP).*

We begin with a brief review of large deviation results for i.i.d. random variables and empirical measures. Let $\{X_n\} \subset \mathbb{R}$ be an i.i.d. sequence of real-valued random variables, and let P_n denote the distribution of the sample mean $\bar{X}_n$. If the moment generating function of X_1 is finite in a neighborhood of the origin, then Cramér's theorem states that $\{P_n\}$ satisfies the LDP with good rate function Λ^*, where Λ^* is the convex conjugate (or Legendre–Fenchel transform) of Λ, and where $\Lambda(\alpha) = \log \boldsymbol{E}[e^{\alpha X_1}]$ is the cumulant generating function of X_1 (cf. [10], Section 2.2).

Next, consider the empirical measures $\{\mu_n\}$, defined by

$$\mu_n(B) = \frac{1}{n} \sum_{i=1}^{n} I_{\{X_i \in B\}}, \quad B \in \mathscr{B}, \tag{4}$$

where $\mathscr{B}$ denotes the collection of Borel subsets of $\mathbb{R}$. It is well known (cf. [14]) that $\{\mu_n\}$ converges weakly to P, namely to the distribution of X_1. Then Sanov's theorem asserts that $\{\mu_n\}$ satisfies a large deviation principle with rate function I_P given by

$$I_P(\nu) = \begin{cases} \mathrm{KL}(\nu, P) & \text{if } \nu \ll P, \\ \infty & \text{otherwise,} \end{cases} \tag{5}$$

where $\mathrm{KL}(\nu, P)$ is the *Kullback–Leibler information* between the probability measures ν and P. When ν and P each possesses a density with respect to Lebesgue measure (say p and g, respectively), the above expression becomes

$$\mathrm{KL}(p, g) := \begin{cases} \int_S p(x) \log\left(\frac{p(x)}{g(x)}\right) d\mu(x) & \text{if } p \ll g, \\ \infty & \text{otherwise.} \end{cases} \tag{6}$$

In Sanov's theorem, the rate function I_P is defined on the space of probability measures, which is a metric space with the open sets induced by weak convergence. Extensions of Sanov's theorem to strong topologies have been investigated in the literature; cf., e.g., [15].

We now turn to a general result, which will play a central role in this paper, namely Varadhan's integral lemma (cf. [10], Theorem 4.3.1). This result will allow us to infer the scaled limit of a sequence of generating functions from the existence of the large deviation principle.

Lemma 1 (Varadhan)**.** *Let* $\{Y_n\}$ *be a sequence of random variables taking values in a regular topological space* $(\mathscr{X}, \mathscr{B})$, *and assume that the probability law of* $\{Y_n\}$ *satisfies the LDP with good rate function* I. *Then for any bounded continuous function* $F : \mathscr{X} \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \log \boldsymbol{E}[\exp(nF(Y_n))] = \sup_{x\in\mathscr{X}} \{F(x) - I(x)\}. \tag{7}$$

1.2. Minimum Hellinger Distance Estimator and Large Deviations

We first observe that the MHDE is obtained by maximizing

$$\mathscr{A}_n(\boldsymbol{\theta}) \equiv \mathscr{A}_n(\boldsymbol{\theta}, g_n) := \int_{\mathbb{R}} f_{\boldsymbol{\theta}}^{\frac{1}{2}}(x) g_n^{\frac{1}{2}}(x) d\mu(x), \tag{8}$$

which involves solving the equation $\nabla \mathscr{A}_n(\boldsymbol{\theta}) = 0$. The idea behind the large deviation analysis is to observe that the large deviation behavior of the maximizer can be extracted from that of the objective function $\nabla \mathscr{A}_n(\boldsymbol{\theta})$ near $\mathbf{0}$. By the Gärtner–Ellis theorem (cf. [10], Section 2.3), this amounts to investigating the asymptotic behavior as $n \to \infty$ of

$$\frac{1}{a_n} \log \boldsymbol{E}_g[\exp\{a_n \langle \boldsymbol{\alpha}, \nabla \mathscr{A}_n(\boldsymbol{\theta}) \rangle\}], \quad \boldsymbol{\alpha} \in \mathbb{R}^d, \tag{9}$$

where $a_n \nearrow \infty$ as $n \to \infty$. In the case of maximum likelihood estimation (MLE) or minimum contrast estimation (MCE), the objective function can be expressed as

$$\sum_{i=1}^{n} h_{\boldsymbol{\theta}}(X_i) = n \int_{\mathbb{R}} h_{\boldsymbol{\theta}}(x) d\mu_n(x), \tag{10}$$

where $\{\mu_n : n \geq 1\}$ is the empirical measure associated with $\{X_k : 1 \leq k \leq n\}$. Thus, while the objective functions associated with the MLE and MCE are linear functions of the empirical measure, the affinity is a nonlinear function of the empirical measure. This creates certain complications in identifying the rate function $I(\cdot)$ alluded to in (3). Of course, in the case of likelihood and minimum contrast estimator analysis, an explicit formula for $I(\cdot)$ ensues as the Legendre–Fenchel transform of the cumulant generating function of $h_{\boldsymbol{\theta}}(X_1)$, viz. $\log \boldsymbol{E}_{\boldsymbol{\theta}_0}[\exp(\alpha h_{\boldsymbol{\theta}}(X_1))]$. One approach to evaluating the limiting generating function is to apply Varadhan's lemma as given above in (7). In the context of our problem, that requires an investigation into the large deviation principle for the density estimators $g_n(\cdot)$ viewed as elements of $L_1(S)$, viz. the space of integrable functions on S. Equivalently, we require a version of Sanov's theorem in L_1-space, which leads to certain topological considerations. The main issue here is that, when L_1 is equipped with a norm topology, the sequence of kernel density estimates $\{g_n(\cdot)\}$ possesses large deviation bounds, but the associated rate function may not have compact level sets, as is required for a typical application of Varadhan's lemma. Nonetheless, one obtains a full LDP when $L_1(S)$ is equipped with the weak topology.

The asymptotic properties of MHDE, such as consistency and asymptotic normality, are established using the norm convergence of $g_n(\cdot)$ to $g(\cdot)$. For this reason, we focus on a subclass of densities $\mathscr{G}$ (see Proposition 1 below) possessing certain equicontinuity properties where norm convergence prevails. These issues are handled in Section 2, where the precise statements of our main results can also be found. Section 3 is devoted to the proofs of the main results. Section 4 contains some concluding remarks.

2. Notation, Assumptions, and Main Results

Let $f_{\boldsymbol{\theta}}(\cdot)$ denote the postulated density of $\{X_n\}$, defined on a measure space $(\Omega, \mathscr{F})$. Let $S \subset \mathbb{R}$ denote the support of X and $s_{\boldsymbol{\theta}}(\cdot) = f_{\boldsymbol{\theta}}^{\frac{1}{2}}(\cdot)$. Let the true density of $\{X_n\}$ be given by $g(\cdot)$. Throughout the paper, we assume that the following regularity conditions hold.

Hypothesis 1. *Θ is a compact and convex subset of $\mathbb{R}^d$.*

Hypothesis 2. *The family $\mathfrak{F}$ is identifiable; namely, if $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$, $f_{\boldsymbol{\theta}_1}(\cdot) \neq f_{\boldsymbol{\theta}_2}(\cdot)$ on a set of positive Lebesgue measure.*

Hypothesis 3. *For every $\boldsymbol{\theta} \in \Theta$, $s_{\boldsymbol{\theta}}$ is three times continuously differentiable with respect to all components of $\boldsymbol{\theta}$. Denote by $\nabla s_{\boldsymbol{\theta}}$ the gradient of $s_{\boldsymbol{\theta}}$ and its components by $\dot{s}^{i}_{\boldsymbol{\theta}}(\cdot)$. Let $\mathscr{H}_{\boldsymbol{\theta}}$ denote the matrix of second partial derivatives of $s_{\boldsymbol{\theta}}(\cdot)$ with respect to $\boldsymbol{\theta}$ and $\ddot{s}^{ij}_{\boldsymbol{\theta}}$ the $(i,j)^{th}$ element of $\mathscr{H}_{\boldsymbol{\theta}}$.*

Hypothesis 4. *Let the matrix of second partial derivatives of $\mathscr{A}_n(\boldsymbol{\theta})$ and $\mathscr{A}(\boldsymbol{\theta})$ be denoted by $H_{\mathscr{A}_n}(\boldsymbol{\theta})$ and $H_{\mathscr{A}}(\boldsymbol{\theta})$, respectively. Assume that $H_{\mathscr{A}_n}(\boldsymbol{\theta})$ and $H_{\mathscr{A}}(\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$ and that $H_{\mathscr{A}}(\boldsymbol{\theta})$ is positive definite for every $\boldsymbol{\theta} \in \Theta$. For $p \in \mathscr{G}$ and $\boldsymbol{\theta} \in \Theta$, let $\lambda_{\boldsymbol{\theta}}(p)$ denote the smallest eigenvalue of the matrix $\int_S \mathscr{H}_{\boldsymbol{\theta}}(x) p^{\frac{1}{2}}(x) dx$. Assume that $\inf\{\lambda_{\boldsymbol{\theta}}(p) : p \in \mathscr{G}\} \geq c > 0$, where c is independent of $\boldsymbol{\theta}$.*

These hypotheses on the family $\mathfrak{F}$ are generally standard and are used to establish the asymptotic properties of the MHDE. Sufficient conditions on $\mathfrak{F}$ for the validity of these hypotheses are described in [3,16], and [17]. A remark on Hypothesis 4 is warranted here. When $p = g$, this assumption is related to the positive definiteness of the Fisher information matrix. If one assumes $\mathscr{G} = \mathfrak{F}$, then this hypothesis reduces to the condition that $\inf\{\lambda_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\} \geq c > 0$, which is standard. Finally, we remark that we have not attempted to provide the weakest regularity conditions, and we do believe some of these conditions can possibly be relaxed.

Recall that the MHDE of $\boldsymbol{\theta}$ can be obtained by solving the equation

$$\nabla \mathscr{A}_n(\boldsymbol{\theta}) := \nabla_{\boldsymbol{\theta}} \mathscr{A}(f_{\boldsymbol{\theta}}, g_n) = \frac{1}{2} \int_{\mathbb{R}} u_{\boldsymbol{\theta}}(x) s_{\boldsymbol{\theta}}(x) g_n^{\frac{1}{2}}(x) dx = 0, \tag{11}$$

where $u_{\boldsymbol{\theta}}(x) = \nabla_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}(x)(f_{\boldsymbol{\theta}}(x))^{-1}$ is the *score function*, which is obtained using $\nabla_{\boldsymbol{\theta}} s(x;\boldsymbol{\theta}) = \frac{1}{2} u(x;\boldsymbol{\theta}) s(x;\boldsymbol{\theta})$.

We begin by providing some heuristics for the case $d = 1$. Let $\dot{\mathscr{A}}_n(\theta)$ denote the derivative of $\mathscr{A}_n(\theta)$ when $d = 1$. Let $\hat{\theta}_n$ denote the argzero of the function $\dot{\mathscr{A}}_n(\theta)$ obtained from (11) above. Let $\hat{\theta}_{n,l} = \inf\{\theta \in \Theta : \dot{\mathscr{A}}_n(\theta) \leq 0\}$ and $\hat{\theta}_{n,u} = \sup\{\theta \in \Theta : \dot{\mathscr{A}}_n(\theta) \geq 0\}$. Since $\hat{\theta}_{n,l} \leq \hat{\theta}_n \leq \hat{\theta}_{n,u}$, we obtain using Markov's inequality that for any $\epsilon > 0$,

$$\boldsymbol{P}_g(\hat{\theta}_{n,l} \geq \theta_g + \epsilon) \leq \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta_g + \epsilon) \geq 0) \leq \boldsymbol{E}_g[\exp(n\alpha \dot{\mathscr{A}}_n(\theta_g + \epsilon)], \tag{12}$$

where $\alpha > 0$. Similarly, for $\alpha < 0$, it can be seen that

$$\boldsymbol{P}_g(\hat{\theta}_{n,u} \leq \theta_g - \epsilon) \leq \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta_g - \epsilon) \leq 0) \leq \boldsymbol{E}_g[\exp(n\alpha \dot{\mathscr{A}}_n(\theta_g - \epsilon)]. \tag{13}$$

Thus, an evaluation of (9) will allow us to obtain the logarithmic upper bound for $\hat{\theta}_{n,l}$ and $\hat{\theta}_{n,u}$. Next, using the inequalities

$$\boldsymbol{P}_g(\hat{\theta}_{n,l} \geq \theta_g + \epsilon) \leq \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta_g + \epsilon) \geq 0) \leq \boldsymbol{P}_g(\hat{\theta}_{n,u} \geq \theta_g + \epsilon), \tag{14}$$

$$\boldsymbol{P}_g(\hat{\theta}_{n,u} \leq \theta_g - \epsilon) \leq \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta_g - \epsilon) \leq 0) \leq \boldsymbol{P}_g(\hat{\theta}_{n,l} \leq \theta_g - \epsilon), \tag{15}$$

under additional hypotheses, one can derive large deviation lower bounds for $\hat{\theta}_n$. Deriving these bounds for MLE and MCE is rather standard, since the objective functions and their derivatives are *linear* functionals of the empirical distribution, as stated in (10), but this is not the case for the Hellinger distance.

Observe that the probabilities in (12) and (13) represent rare-event probabilities since, under the hypotheses described previously, $\hat{\theta}_n$ converges to θ_g almost surely as $n \to \infty$. The distributional results concerning $\hat{\theta}_n$ rely on the continuity and differentiability properties of $\nabla \mathscr{A}_n(\theta)$, which depend nonlinearly on g_n, and the norm convergence of g_n to g.

Let $\mathscr{G}$ denote the collection of all probability densities with support S. By Scheffe's theorem, the pointwise convergence of g_n to g implies $g_n \xrightarrow{L_1} g$ as $n \to \infty$. Additionally, when $g_n(\cdot)$ is the kernel density estimator, then Glick's Theorem guarantees that $g_n \xrightarrow{L_1} g$ almost surely as $n \to \infty$ when $b_n \searrow 0$ and $n \nearrow \infty$; cf. [18]. Since the MHDE are

functionals of density estimators, it is natural to expect that the large deviations of density estimators will play a significant role in our analysis. For this reason, one is forced to consider the topological issues that arise in the large deviation analysis of density estimators. Interestingly, it turns out that the weak topology on $L_1(S)$ plays a prominent role. This, in turn, leads to the question of whether certain continuity properties, which were part of the traditional theory of MHD analysis, continue to hold if $\mathscr{G}$ were viewed as a subset of $L_1(S)$ equipped with weak topology. Expectedly, while the answer in general is no (cf. [19]), Proposition 1 provides sufficient conditions on the family $\mathscr{G}$ under which one additionally obtains norm convergence.

Before proceeding, we now introduce some further regularity conditions, as follows.

Hypothesis 5. *$u_{\boldsymbol{\theta}} s_{\boldsymbol{\theta}} \in L_2(S)$ and is an $L_2(S)$-continuous function of $\boldsymbol{\theta}$.*

Hypothesis 6. *The family $\mathfrak{F}$ consists of bounded equicontinuous densities.*

Hypothesis 7. *The family $\mathscr{G}$ consists of bounded and equicontinuous densities.*

Hypothesis 8. *$u_{\boldsymbol{\theta}} g \in L_2(S)$ and is an $L_2(S)$-continuous function of $\boldsymbol{\theta}$.*

Here, we note that Hypotheses 6 and 7 are related. Furthermore, if one is willing to assume that $\mathscr{G} = \mathfrak{F}$, then one does not need Hypothesis 7. On the other hand, if one believes that parametric distributions are approximations to $\mathscr{G}$, then one needs to work with Hypothesis 7. For this reason, we have maintained both of these hypotheses in our main results. Hypotheses 5 and 8 are related to finiteness of the Fisher information and are standard in the statistical literature.

Before we state the first proposition, we recall the definition of weak topology on L_1 (cf. [19]). A sequence $\{h_n : n \geq 1\}$ is said to converge weakly in L_1 if $\int_S h_n(x)b(x)dx \to \int_S h(x)b(x)dx$ as $n \to \infty$ for every $b \in L_\infty(S)$, where $L_\infty(S)$ is a class of essentially bounded functions. We assume throughout the paper that the topology on Θ is the standard topology generated by the Euclidean metric.

Proposition 1. *Let $\mathscr{G}$ denote the class of densities, equipped with the weak topology. Further assume that Hypotheses 1–7 hold. Let $\Theta \otimes \mathscr{G}$ be equipped with the product topology. Then the mapping $\nabla\mathscr{A} : \Theta \otimes \mathscr{G} \to \mathbb{R}^d$ defined by*

$$\nabla\mathscr{A}(\boldsymbol{\theta}, g) := \int_{\mathbb{R}} u_{\boldsymbol{\theta}}(x) s_{\boldsymbol{\theta}}(x) g^{\frac{1}{2}}(x) dx \tag{16}$$

is jointly continuous in $(\boldsymbol{\theta}, g)$. Furthermore, if $g_n \xrightarrow{w} g$, then

$$\lim_{n\to\infty} \sup_{\boldsymbol{\theta}\in\Theta} ||\nabla\mathscr{A}(\boldsymbol{\theta}, g_n) - \nabla\mathscr{A}(\boldsymbol{\theta}, g)|| = 0. \tag{17}$$

Finally, under Hypothesis 7, the family $\mathscr{G}$ is a weakly sequentially closed subset of $L_1(S)$.

Our next result is concerned with the limit behavior of the generating function of $\nabla\mathscr{A}_n(\boldsymbol{\theta})$. In the following we use the notation $p \ll g$ to mean the probability measures associated with $p(\cdot)$ and $g(\cdot)$ are absolutely continuous.

Theorem 1. *Assume that Hypotheses 1–7 hold, and set*

$$\Lambda_{n,\boldsymbol{\theta}}(\boldsymbol{\alpha}) := \frac{1}{n} \log \boldsymbol{E}_g[\exp(n\langle \boldsymbol{\alpha}, \nabla\mathscr{A}_n(\boldsymbol{\theta})\rangle)], \quad \boldsymbol{\alpha} \in \mathbb{R}^d. \tag{18}$$

Then $\Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) := \lim_{n\to\infty} \Lambda_{n,\boldsymbol{\theta}}(\boldsymbol{\alpha})$ *exists and is a convex function given by*

$$\Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) = \sup_{p\in\mathscr{G}}\left\{\int_S \langle \boldsymbol{\alpha}, u_{\boldsymbol{\theta}}(x)\rangle s_{\boldsymbol{\theta}}(x) p^{\frac{1}{2}}(x)dx - \mathrm{KL}(p,g)\right\}, \tag{19}$$

where

$$\mathrm{KL}(p,g) = \begin{cases} \int_S p(x)\log\left(\frac{p(x)}{g(x)}\right)dx & \text{if } p \ll g, \\ \infty & \text{otherwise.} \end{cases} \tag{20}$$

Remark 1. Since $\Lambda_{\boldsymbol{\theta}}$ is defined via a limiting operation, it is hard to extract its qualitative properties. However, we can obtain a simple lower bound by observing that $\mathrm{KL}(p,g) = 0$ if and only if $p = g$, and an upper bound using that the Kullback–Leibler information is nonnegative. This results in the following bounds:

$$\int_S \langle \boldsymbol{\alpha}, u_{\boldsymbol{\theta}}(x)\rangle s_{\boldsymbol{\theta}}(x) g^{\frac{1}{2}}(x)dx \le \Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) \le \sup_{p\in\mathscr{G}}\left[\int_S \langle \boldsymbol{\alpha}, u_{\boldsymbol{\theta}}(x)\rangle s_{\boldsymbol{\theta}}(x) p^{\frac{1}{2}}(x)dx\right]. \tag{21}$$

Furthermore, if all densities in $\mathscr{G}$ are bounded by one, then $p^{\frac{1}{2}}(\cdot) \ge p(\cdot)$ implies

$$\Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) \quad \ge \quad \sup_{p\in\mathscr{G}}\left\{\int_S \langle \boldsymbol{\alpha}, u_{\boldsymbol{\theta}}(x)\rangle s_{\boldsymbol{\theta}}(x) p(x)dx - \mathrm{KL}(p,g)\right\}. \tag{22}$$

Using a variational argument, it can be shown that the supremum on the right-hand side is attained at p^* given by

$$p^*(x) := \frac{\exp(\langle \boldsymbol{\alpha}, u_{\boldsymbol{\theta}}(x)\rangle) s_{\boldsymbol{\theta}}(x)}{\int_S \langle \boldsymbol{\alpha}, u_{\boldsymbol{\theta}}(x)\rangle s_{\boldsymbol{\theta}}(x) g(x)dx}; \tag{23}$$

cf. [20]. Furthermore, the maximum that results from this choice of $p^*(\cdot)$ is

$$\log \int_S \exp(\langle \boldsymbol{\alpha}, u_{\boldsymbol{\theta}}(x)\rangle) s_{\boldsymbol{\theta}}(x) g(x)dx,$$

yielding yet another lower bound for $\Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha})$, although the comparison of these two lower bounds is not immediate.

Returning to our main discussion, recall from [21] that the convex conjugate of the function $\Lambda_{\boldsymbol{\theta}}$ is defined by

$$\Lambda^*_{\boldsymbol{\theta}}(\boldsymbol{x}) = \sup_{\boldsymbol{\alpha}\in\mathbb{R}^d}\{\langle \boldsymbol{\alpha}, \boldsymbol{x}\rangle - \Lambda(\boldsymbol{\alpha})\}, \quad \mathbf{x} \in \mathbb{R}^d. \tag{24}$$

Let $\mathfrak{D}_{\boldsymbol{\theta}}$ denote the domain of $\Lambda_{\boldsymbol{\theta}}$; namely,

$$\mathfrak{D}_{\boldsymbol{\theta}} = \{\boldsymbol{\alpha} \in \mathbb{R}^d : \Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) < \infty\}; \tag{25}$$

and let $\mathfrak{R}_{\boldsymbol{\theta}}$ denote the range of the gradient map $\nabla\Lambda_{\boldsymbol{\theta}}$; that is,

$$\mathfrak{R}_{\boldsymbol{\theta}} = \left\{\boldsymbol{x} \in \mathbb{R}^d : \nabla\Lambda_{\boldsymbol{\theta}}(\alpha) = \boldsymbol{x}, \quad \text{some } \boldsymbol{\alpha} \in \mathbb{R}^d\right\}.$$

We begin with the discussion of the case $d = 1$. In this case, the generating function Λ_θ reduces to

$$\Lambda_\theta(\alpha) = \sup_{p\in\mathscr{G}}\left\{\alpha \int_S \exp(n\alpha\dot{\mathscr{A}}_n(\theta)s(x;\theta)p^{\frac{1}{2}}(x)dx - \mathrm{KL}(p,g)\right\}. \tag{26}$$

By the convexity of $\Lambda_\theta(\cdot)$, this function is differentiable almost everywhere (cf. [21]), and in the proof, we would like to exploit the differentiability of this function at the point α_θ^* where it attains its minimum value. If Λ_θ is not differentiable at this point, it is helpful to consider the directional derivatives of Λ_θ. Specifically, let $\Lambda'_{\theta,+}(\cdot)$ and $\Lambda'_{\theta,-}(\cdot)$ denote the right and left derivatives of $\Lambda_\theta(\cdot)$, respectively. When $x \in \Big(\Lambda'_{\theta,-}(\alpha), \Lambda'_{\theta,+}(\alpha)\Big)$, then it is well known that $\Lambda_\theta^*(x) = \alpha x - \Lambda_\theta(\alpha)$, but this observation will not be sufficient to obtain a proper lower bound. For that to hold, we need a stronger condition, namely that $0 \in \mathfrak{R}_\theta$, which will only be true if Λ_θ is differentiable at its point of minimum, α_θ^*. Otherwise, the expected lower bound turns out to be $\Lambda_\theta^*(x)$, where $x = \Lambda'_{\theta,+}(\alpha_\theta^*)$; cf. [13].

We now turn to our large deviation theorem in $\mathbb{R}^1$, where we study the rare-event probabilities $\boldsymbol{P}_g(\hat{\theta}_n \in C)$ for sets C that are away from the true value θ_g. Specifically, we establish an analogue of the LDP, but where a subtle difference arises in the lower bound in the absence of differentiability of Λ_θ.

We recall that $\hat{\theta}_n$ is defined using the kernel density estimator $g_n(\cdot)$ defined in (1), whose behavior is dictated by the bandwidth sequence $\{b_n\}$.

Theorem 2. *Assume $d = 1$, Hypotheses 1–8 are satisfied, and $\hat{\theta}_n$ is the unique zero of $\dot{\mathscr{A}}_n(\theta) = 0$. Further assume that $b_n \searrow 0$ and $nb_n \nearrow \infty$ as $n \to \infty$. Then for any closed set F not containing θ_g,*

$$\limsup_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \in F) \leq -\inf_{\theta\in F} \Lambda_\theta^*(0). \tag{27}$$

Moreover, for any open set G not including θ_g,

$$\liminf_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \in G) \geq -\inf_{\theta\in G} I(\theta), \tag{28}$$

where

$$I(\theta) = \inf\{\Lambda_\theta^*(x) : x \in \mathfrak{R}_\theta \cap [0,\infty)\}, \tag{29}$$

and the infimum is taken to be infinity if the set $\mathfrak{R}_\theta \cap [0,\infty)$ is empty.

Remark 2. If $F = [\theta,\infty)$ where $\theta > \theta_g$, then in both the upper and lower bounds, it is sufficient to evaluate the infimum at the boundary point θ. That is,

$$\limsup_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \in [\theta,\infty)) \leq -\Lambda_\theta^*(0).$$

Similarly, if $G = (\theta,\infty)$ where $\theta > \theta_g$, then

$$\liminf_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \in (\theta,\infty)) \geq -I(\theta).$$

Furthermore, if $\inf_\alpha \Lambda_\theta(\alpha)$ is achieved at a unique point α_θ^* and Λ_θ is differentiable at α_θ^*, then the right-hand side of (28) reduces to $\Lambda_\theta^*(0)$, i.e., the upper and lower bounds coincide and the limits exist. Since the rate function appearing in the upper and lower bounds coincide in this case, we obtain a proper LDP *if* the resulting rate function has the required regularity properties, in particular, $I(\theta) = \Lambda_\theta^*(0)$ is lower semicontinuous and has compact level sets.

The proof of the above theorem relies on (14) and (15) combined with Theorem 1, together with a change of measure argument characteristic of large deviation analysis. The comparison inequalities in (14) and (15) are critical to obtaining the characterizations in the above theorem, but these are essentially one-dimensional results and their analogues in higher dimensions ($d \geq 2$) are not immediate. Consequently, when $\Lambda_{\boldsymbol{\theta}}$ is not differentiable,

new complications arise, which lead to a slightly different, and less explicit, representation of the lower bound.

Next we establish a large deviation theorem for $\mathbb{R}^d$, generalizing the previous theorem to higher dimensions. In the following, let $\text{dist}(x, G) = \inf_{y \in G} ||x - y||$ denote the distance between a point $x \in \mathbb{R}^d$ and a set $G \subset \mathbb{R}^d$.

Theorem 3. *Assume Hypotheses 1–8 are satisfied, and assume that* $b_n \searrow 0$ *and* $nb_n \nearrow \infty$ *as* $n \to \infty$. *Then for any closed set* F *not containing* $\boldsymbol{\theta}_g$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n \in F) \leq - \inf_{\boldsymbol{\theta} \in F} \Lambda^\star_{\boldsymbol{\theta}}(\mathbf{0}). \tag{30}$$

Moreover, for any open set G *not including* $\boldsymbol{\theta}_g$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n \in G) \geq - \inf_{\boldsymbol{\theta} \in G} I(\boldsymbol{\theta}), \tag{31}$$

where $I(\boldsymbol{\theta}) = \inf\{\Lambda^*_{\boldsymbol{\theta}}(\boldsymbol{x}) : \boldsymbol{x} \in \mathfrak{R}_{\boldsymbol{\theta}} \cap B(\mathbf{0}; c_{\boldsymbol{\theta}})\}$ *and* $c_{\boldsymbol{\theta}} = b\ \text{dist}(\boldsymbol{\theta}, \Theta - G)$ *for some universal constant* $b \in (0, \infty)$, *and the infimum is taken to be infinity if the set* $\mathfrak{R}_{\boldsymbol{\theta}} \cap B(\mathbf{0}; c_{\boldsymbol{\theta}})$ *is empty.*

Remark 3. As we noted for the one-dimensional case in Remark 2, under a differentiability assumption on Λ_θ, the function $I(\boldsymbol{\theta})$ can be identified as $\Lambda^*_\theta(\mathbf{0})$, but in full generality, it is not immediately known that $I(\boldsymbol{\theta})$ is even nontrivial. Moreover, without differentiability, the infimum in the definition of $I(\boldsymbol{\theta})$ is *more* restrictive than what we encountered in the one-dimensional problem. However, if one assumes additional geometry on G, such as a translated cone structure, then one obtains improved estimates in the sense that one can take unbounded regions in the definition of $I(\boldsymbol{\theta})$, just as we saw in Theorem 2.2. For further remarks in this direction, see the discussion given after the proof of the theorem.

3. Proofs

We turn first to Proposition 1.

Proof of Proposition 1. Since $\Theta \otimes \mathscr{G}$ is equipped with product topology, it is sufficient to show that if $\boldsymbol{\theta}_n \to \boldsymbol{\theta}$ and $g_n \xrightarrow{w} g$, then $\nabla \mathscr{A}_n(\boldsymbol{\theta})$ converges to $\nabla \mathscr{A}(\boldsymbol{\theta})$, where

$$\nabla \mathscr{A}(\boldsymbol{\theta}) = \int_S u_{\boldsymbol{\theta}}(x) s_{\boldsymbol{\theta}}(x) g^{\frac{1}{2}}(x) dx. \tag{32}$$

Let $r_{\boldsymbol{\theta}}(x) = u_{\boldsymbol{\theta}}(x) s_{\boldsymbol{\theta}}(x)$, and observe that

$$\begin{aligned} |\nabla \mathscr{A}(\boldsymbol{\theta}_n, g_n) - \nabla \mathscr{A}(\boldsymbol{\theta}, g)| &\leq \int_S |r_{\boldsymbol{\theta}_n}(x)| |g_n^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x)| dx + \int_S |r_{\boldsymbol{\theta}_n}(x) - r_{\boldsymbol{\theta}}(x)| g^{\frac{1}{2}}(x) dx \\ &\leq ||r_{\boldsymbol{\theta}}||_2 \text{HD}(g_n, g) + \int_S |r_{\boldsymbol{\theta}_n}(x) - r_{\boldsymbol{\theta}}(x)| g^{\frac{1}{2}}(x) dx \\ &= T_{n,1} + T_{n,2}, \end{aligned} \tag{33}$$

where the penultimate equation follows by applying the Cauchy–Schwarz inequality. Then by the Cauchy–Schwarz inequality and Hypothesis 5, $T_{n,2} \to 0$. Since Hellinger distance is dominated by the L_1-distance, in order to complete the proof, it is sufficient to show that $||g_n - g||_1 \to 0$. Now since $g_n \xrightarrow{w} g$, it follows that as $n \to \infty$,

$$G_n(x) := \int_S g_n(y) I_{\{y \leq x\}} dy \to \int_S g(y) I_{\{y \leq x\}} dy := G(x). \tag{34}$$

Evidently, $G_n(\cdot)$ and $G(\cdot)$ are nondecreasing and right continuous. Furthermore, if $x_* = \inf\{x : x \in S\}$ and $x^* = \sup\{x : x \in S\}$, then $G_n(x_*) \to G(x_*)$ and $G_n(x^*) \to G(x^*)$, where $G_n(x_*) = \lim_{x \to x_*} G_n(x)$, $G_n(x^*) = \lim_{x \to x*} G_n(x)$, $G(x_*) = \lim_{x \to x_*} G(x)$,

$G(x^*) = \lim_{x \to x*} G(x)$. Thus G_n converges to G, which is a proper distribution function. Then by Lemma 1 of Boos [22], $g_n(\cdot)$ converges to $g(\cdot)$ uniformly on compact sets. This, in turn, implies the L_1 convergence of $g_n(\cdot)$ to $g(\cdot)$ (by Scheffe's lemma), which establishes the convergence of $T_{n,1}$ to 0, thus completing the proof of the joint continuity of $\nabla \mathscr{A}(\boldsymbol{\theta}, g)$.

Next, the uniform convergence (17) follows by Hypothesis 5, since

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \Theta} |\nabla \mathscr{A}(\boldsymbol{\theta}, g_n) - \nabla \mathscr{A}(\boldsymbol{\theta}, g)| &\leq \int_S |r_{\boldsymbol{\theta}}(x)| |g_n^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x)| dx \\ &\leq \sup_{\boldsymbol{\theta} \in \Theta} ||r_{\boldsymbol{\theta}}||_2 \mathrm{HD}(g_n, g) \to 0. \end{aligned}$$

Finally, to prove that $\mathscr{G}$ is weakly sequentially closed, note that convergence in weak topology implies pointwise convergence, yielding $g(\cdot) \geq 0$. Noting that

$$\int_S g(x) d\mu(x) = 1 + \int_S (g(x) - g_n(x)) dx, \tag{35}$$

it follows that $g(\cdot)$ integrates to one, using L_1 convergence, thus completing the proof of the proposition. □

We now turn to the proof of Theorem 1. The proof relies on the large deviation theorem for the kernel density estimator $g_n(\cdot)$ in the weak topology of $\mathscr{G}$. The next proposition is concerned with the LDP for $\{g_n\}$ in $\mathscr{G}$, equipped with the inherited weak topology from $L_1(S)$. This issue has received considerable attention recently (cf. [23,24]), where it is established that the full LDP may *not* hold for $\{g_n\}$ in norm topology, but does hold under the weak topology.

Proposition 2. *Assume Hypotheses 1–8 and that* $b_n \searrow 0$ *and* $nb_n \nearrow \infty$ *as* $n \to \infty$. *Then* $\{g_n\}$ *satisfies the LDP in the weak topology of* $L_1(S)$ *with good rate function* I *given by*

$$I(p) = \begin{cases} \int_S p(x) \log\left(\frac{p(x)}{g(x)}\right) dx & \text{if } g \ll p, \\ \infty & \text{otherwise.} \end{cases} \tag{36}$$

Proof of Theorem 1. As before, let $\mathscr{G}$ be equipped with the weak topology. Set $r_{\boldsymbol{\theta}}(x) = u_{\boldsymbol{\theta}}(x) s_{\boldsymbol{\theta}}(x)$, and define $F : \mathscr{G} \to \mathbb{R}$ as follows:

$$F(h) = \int_S \langle \boldsymbol{\alpha}, r_{\boldsymbol{\theta}}(x) \rangle h^{\frac{1}{2}}(x) dx. \tag{37}$$

By Hypothesis 5, $r_{\boldsymbol{\theta}} \in L_2(S)$. To show that $F(\cdot)$ is continuous, let $h_n \xrightarrow{w} h$ as $n \to \infty$. Then

$$\begin{aligned} |F(h_n) - F(h)| &\leq \int_S r_{\boldsymbol{\theta}}(x) |h_n^{\frac{1}{2}}(x) - h^{\frac{1}{2}}(x)| d\mu(x) \\ &\leq ||r_{\boldsymbol{\theta}}||_2 \mathrm{HD}(h_n, h) \leq ||r_{\boldsymbol{\theta}}||_2 ||h_n - h||_1 \to 0 \quad \text{as} \quad n \to \infty, \end{aligned} \tag{38}$$

where we have used the Cauchy–Schwarz inequality that the L_1 distance dominates the Hellinger distance in (38). Now by Hypothesis 7, as in the proof of Proposition 1, we have that $||h_n - h||_1 \to 0$ as $n \to \infty$, establishing the continuity of $F(\cdot)$. Next, to show that $F(\cdot)$ is

bounded, note that $\sup\{F(p) : p \in \mathcal{G}\} \leq ||r_{\boldsymbol{\theta}}||_2$ by the Cauchy–Schwarz inequality. Then by Proposition 2, it follows by Varadhan's integral lemma (see [10], Theorem 4.3.1) that

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \log \boldsymbol{E}[\exp(nF(g_n(x))] &= \lim_{n\to\infty} \frac{1}{n} \log \boldsymbol{E}\left[\exp\left(n \int_S \langle \boldsymbol{\alpha}, u_{\boldsymbol{\theta}}(x) s_{\boldsymbol{\theta}}(x)\rangle g_n^{\frac{1}{2}}(x) dx\right)\right] \\ &= \sup_{p\in\mathcal{G}} \left\{ \int_S \langle \boldsymbol{\alpha}, u_{\boldsymbol{\theta}}(x)\rangle s_{\boldsymbol{\theta}}(x) p^{\frac{1}{2}}(x) dx - \mathrm{KL}(p, g) \right\} \\ &:= \Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}). \end{aligned} \tag{39}$$

This completes the proof of the theorem. □

The proofs of our main results will involve probability bounds on the modulus of continuity of $\mathscr{A}_n(\boldsymbol{\theta})$ and $\nabla\mathscr{A}_n(\boldsymbol{\theta})$, respectively. Recall that the modulus of continuity $\omega(h; r)$ of a function $h : \mathbb{R}^d \to \mathbb{R}$ is given by

$$\omega(h; r) := \sup_{||x_1 - x_2|| \leq r} |h(x_1) - h(x_2)|, \quad r > 0. \tag{40}$$

Observe that when $h(\cdot)$ is replaced by $\mathscr{A}_n(\boldsymbol{\theta})$ or $\nabla\mathscr{A}_n(\boldsymbol{\theta})$, the modulus of continuity becomes a random quantity. Our next proposition summarizes the continuity properties of $\mathscr{A}_n(\boldsymbol{\theta})$ and $\nabla\mathscr{A}_n(\boldsymbol{\theta})$ via their modulus of continuity as real-valued functionals from $\mathcal{G}$ equipped with the weak topology.

Proposition 3. *Assume that Hypotheses 1–8 hold and that $b_n \searrow 0$ and $nb_n \nearrow \infty$ as $n \to \infty$. Then, with respect to $\{\mathscr{A}_n\}$ and $\mathscr{A}$, the modulus of continuity satisfies the following relations, each with probability one:*

$$\text{(i)} \lim_{n\to\infty} \omega(\mathscr{A}_n; r) = \omega(\mathscr{A}, r); \quad \text{(ii)} \lim_{r\to 0} \omega(\mathscr{A}_n; r) = 0; \quad \text{and} \quad \text{(iii)} \lim_{r\to 0} \omega(\mathscr{A}; r) = 0.$$

Similarly, the sequence $\{\nabla\mathscr{A}_n\}$ and $\nabla\mathscr{A}$ satisfy the analogous relations with probability one; namely,

$$\text{(iv)} \lim_{n\to\infty} \omega(\nabla\mathscr{A}_n; r) = \omega(\nabla\mathscr{A}; r); \ \text{(v)} \lim_{r\to 0} \omega(\nabla\mathscr{A}_n; r) = 0; \text{ and (vi)} \lim_{r\to 0} \omega(\nabla\mathscr{A}; r) = 0.$$

Proof. First observe that $\mathscr{A}_n(\boldsymbol{\theta})$ converges uniformly to $\mathscr{A}(\boldsymbol{\theta})$. To see this, note that if $g_n \xrightarrow{w} g$, then by Proposition 1, it converges in L_1. Hence

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta} |\mathscr{A}_n(\boldsymbol{\theta}) - \mathscr{A}(\boldsymbol{\theta})| &\leq \sup_{\boldsymbol{\theta}\in\Theta} \int_{\mathbb{R}} s_{\boldsymbol{\theta}}(x) |g_n^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x)| dx \\ &\leq ||g_n^{\frac{1}{2}}(x) - g^{\frac{1}{2}}(x)||_2 \leq ||g_n - g||_1 \to 0, \end{aligned} \tag{41}$$

where the last inequality follows using that the Hellinger distance is dominated by the L_1-distance. We now prove (i). For this we invoke the properties of the modulus of continuity. Observe that

$$\omega(\mathscr{A}_n; r) = \omega(\mathscr{A}_n - \mathscr{A} + \mathscr{A}; r) \leq \omega(\mathscr{A}_n - \mathscr{A}; r) + \omega(\mathscr{A}; r), \tag{42}$$

which yields

$$|\omega(\mathscr{A}_n; r) - \omega(\mathscr{A}; r)| \leq \omega(\mathscr{A}_n - \mathscr{A}; r). \tag{43}$$

Next observe that

$$\begin{aligned}\omega(\mathscr{A}_n - \mathscr{A}; r) &= \sup_{||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|| \le r} |(\mathscr{A}_n - \mathscr{A})(\boldsymbol{\theta}_1) - (\mathscr{A}_n - \mathscr{A})(\boldsymbol{\theta}_2)| \\ &\le 2 \sup_{\boldsymbol{\theta} \in \Theta} |\mathscr{A}_n(\boldsymbol{\theta}) - \mathscr{A}(\boldsymbol{\theta})| \to 0, \end{aligned} \tag{44}$$

where the last convergence follows from the uniform convergence of $(\mathscr{A}_n - \mathscr{A})(\boldsymbol{\theta})$ to 0 as shown in (42). The proof of (iv) is similar, and specifically is obtained by using that

$$\omega(\nabla(\mathscr{A}_n - \mathscr{A}); r) \le 2 \sup_{\boldsymbol{\theta} \in \Theta} ||\nabla \mathscr{A}_n(\boldsymbol{\theta}) - \nabla \mathscr{A}(\boldsymbol{\theta})|| \to 0, \tag{45}$$

where the above convergence follows from (17).

We now turn to the proof of (ii). Using the Cauchy–Schwarz inequality and the definition of Hellinger distance,

$$\begin{aligned}\omega(\mathscr{A}_n; r) &= \sup_{||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|| \le r} |\mathscr{A}_n(\boldsymbol{\theta}_1) - \mathscr{A}_n(\boldsymbol{\theta}_2)| \\ &= \sup_{||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|| \le r} \left| \int_{\mathbb{R}} (s_{\boldsymbol{\theta}_1}(x) - s_{\boldsymbol{\theta}_2}(x)) g_n^{\frac{1}{2}}(x) dx \right| \le \mathrm{HD}(f_{\boldsymbol{\theta}_1}, f_{\boldsymbol{\theta}_2}) \\ &\le \sup_{||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|| \le r} ||f_{\boldsymbol{\theta}_1} - f_{\boldsymbol{\theta}_2}||_1 := \omega(H; r), \end{aligned} \tag{46}$$

where $H : (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \to ||f_{\boldsymbol{\theta}_1} - f_{\boldsymbol{\theta}_2}||_1$ is continuous since $\mathfrak{F}$ is continuous in $\boldsymbol{\theta}$. Also, since $\Theta \times \Theta$ is compact, $H(\cdot, \cdot)$ is uniformly continuous. Since the modulus of continuity converges to 0 if and only if $H(\cdot, \cdot)$ is uniformly continuous, (ii) follows. Turning to (v), notice that, as before,

$$\omega(\nabla \mathscr{A}_n; r) \le \sup_{||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2|| \le r} ||u_{\boldsymbol{\theta}_1} s_{\boldsymbol{\theta}_1} - u_{\boldsymbol{\theta}_2} s_{\boldsymbol{\theta}_2}||_2. \tag{47}$$

Now, since $u_{\boldsymbol{\theta}} s_{\boldsymbol{\theta}}$ is L_2 continuous, by Hypothesis 5, the proof follows as in (ii) due to to the compactness of Θ. The proofs of (iii) and (vi) are similar to (ii) and (v), respectively, and are therefore omitted. □

Proposition 4. *For any $0 < M < \infty$ and $\delta > 0$, there exists a positive number $r(M, \delta)$ such that*

$$\boldsymbol{P}_g(\omega(\mathscr{A}_n; r) \ge \delta) \le e^{-Mn} \quad \textit{and} \quad \boldsymbol{P}_g(\omega(\nabla \mathscr{A}_n; r) \ge \delta) \le e^{-Mn}. \tag{48}$$

Proof. By Markov's inequality and (46), it follows that for any $\beta > 0$,

$$\boldsymbol{P}_g(\omega(\mathscr{A}_n; r) \ge \delta) \le \boldsymbol{E}_g[e^{n\beta\omega(\mathscr{A}_n; r)}] e^{-n\beta\delta} \le e^{-n\beta(\delta - \omega(H; r))}. \tag{49}$$

Since $\omega(H; r) \to 0$ as $r \searrow 0$, there exists an r_0 such that for all $r \le r_0$, $(\delta - \omega(H; r)) > 0$. Since $\beta > 0$ is arbitrary, the proposition follows by taking $\beta = M(\delta - \omega(H; r))^{-1}$, for some $r \le r_0$. The proof of the second inequality is similar, using (47). □

Proof of Theorem 2. We begin with the proof of the upper bound. Since we assume that the equation $\dot{\mathscr{A}}_n(\theta) = 0$ has a unique solution, it follows from the inequality in (12) that for any $\alpha > 0$ and $\theta > \theta_g$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \ge \theta) \le \limsup_{n \to \infty} \frac{1}{n} \log \boldsymbol{E}_g[\exp(n\alpha \dot{\mathscr{A}}_n(\theta))] = \Lambda_\theta(\alpha), \tag{50}$$

where the last equality follows by applying Theorem 1 with $d = 1$. Since the inequality holds for every $\alpha > 0$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \geq \theta) \quad \leq \quad \sup_{\alpha>0} \Lambda_\theta(\alpha) \leq \sup_{\alpha\in\mathbb{R}} \Lambda_\theta(\alpha). \tag{51}$$

Now, noticing that $\sup_{\alpha\in\mathbb{R}} \Lambda_\theta(\alpha) = -\inf_{\alpha\in\mathbb{R}} -\Lambda_\theta(\alpha) = -\Lambda_\theta^*(0)$, we then obtain

$$\limsup_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \geq \theta) \leq -\Lambda_\theta^*(0). \tag{52}$$

Similarly, for $\theta < \theta_g$, using (13), one can show by an analogous calculation that

$$\limsup_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \leq \theta) \quad \leq \quad -\Lambda_\theta^*(0). \tag{53}$$

Now let $\theta_1 = \inf\{\theta > \theta_g : \theta \in F\}$ and $\theta_2 = \sup\{\theta < \theta_g : \theta \in F\}$. Then

$$\boldsymbol{P}_g(\hat{\theta}_n \in F) \leq \boldsymbol{P}_g(\hat{\theta}_n \geq \theta_1) + \boldsymbol{P}_g(\hat{\theta}_n \leq \theta_2), \tag{54}$$

and so by (52) and (53), it follows that

$$\limsup_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \in F) \leq -\min_{\theta\in\{\theta_1,\theta_2\}} \Lambda_\theta^*(0) \leq -\inf_{\theta\in F} \Lambda_\theta^*(0), \tag{55}$$

where the last step follows since F closed implies $\{\theta_1, \theta_2\} \subset F$.

Next we turn now to the proof of the lower bound. Let G be an open set, and let $\theta \in G$. Then there exists an $\epsilon > 0$ (to be chosen) such that $I_\epsilon := (\theta - \epsilon, \theta + \epsilon) \subsetneq G$. Note that

$$\begin{aligned} \{\hat{\theta}_n \in I_\epsilon\} \quad &= \quad \{\dot{\mathscr{A}}_n(\hat{\theta}_n) = 0,\ \hat{\theta}_n \in I_\epsilon\} \\ &\supset \quad \{\dot{\mathscr{A}}_n(\theta) - \dot{\mathscr{A}}_n(\hat{\theta}_n) \geq \delta\} \cup \{\ \hat{\theta}_n \in I_\epsilon,\ \sup_{\theta_1,\theta_2\in I_\epsilon} |\dot{\mathscr{A}}_n(\theta_1) - \dot{\mathscr{A}}_n(\theta_2)| \leq \delta\}. \end{aligned}$$

Thus,

$$\begin{aligned} \boldsymbol{P}_g(\hat{\theta}_n \in I_\epsilon) \quad &\geq \quad \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta) - \dot{\mathscr{A}}_n(\hat{\theta}_n) \geq \delta) - \boldsymbol{P}_g(\hat{\theta}_n \notin I_\epsilon,\ \sup_{\theta_1,\theta_2\in I_\epsilon} |\dot{\mathscr{A}}_n(\theta_1) - \dot{\mathscr{A}}_n(\theta_2| > \delta)) \\ &\geq \quad \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta) - \dot{\mathscr{A}}_n(\hat{\theta}_n) \geq \delta) - \boldsymbol{P}_g(\sup_{\theta_1,\theta_2\in I_\epsilon} |\dot{\mathscr{A}}_n(\theta_1) - \dot{\mathscr{A}}_n(\theta_2)| > \delta) \\ &= \quad \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta) \geq \delta) - \boldsymbol{P}_g(\sup_{\theta_1,\theta_2\in I_\epsilon} |\dot{\mathscr{A}}_n(\theta_1) - \dot{\mathscr{A}}_n(\theta_2)| > \delta) \\ &= \quad \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta) \geq \delta) - \boldsymbol{P}_g(\omega(\dot{\mathscr{A}}_n; \epsilon) > \delta). \end{aligned} \tag{56}$$

We now investigate $\boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta) \geq \delta)$. Let Q_n denote the distribution of $\dot{\mathscr{A}}_n(\theta)$, and define $Q_{n,\alpha}$ as follows:

$$Q_{n,\alpha}(B) = \frac{1}{\Lambda_{n,\theta}(\alpha)} \int_B e^{-n\alpha y} dQ_n(y), \quad B \in \mathscr{B}. \tag{57}$$

Let $B = (x - \eta, x + \eta)$, for some $\eta > 0$, where $B \subset (\delta, \infty)$ and $x \in \mathfrak{R}_\theta$. Then

$$Q_n(B) \geq \exp\{-n\alpha x - n\eta|\alpha| + n\Lambda_{n,\theta}(\alpha)\} Q_{n,\alpha}(B). \tag{58}$$

Taking the logarithm, dividing by n, and then taking the limit as $n \to \infty$, we obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(B) \geq -\alpha x - \eta|\alpha| - \Lambda_\theta(\alpha) + \liminf_{n\to\infty} \frac{1}{n} \log Q_{n,\alpha}(B). \tag{59}$$

Now since $x \in \mathfrak{R}_\theta$, we can apply Theorem IV.1 of [25] to obtain that the last term on the right-hand side of the previous equation converges to zero. Upon letting $\eta \to 0$, it follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(B) \geq -\Lambda_\theta^*(x). \tag{60}$$

Since the above inequality holds for all $x \in \mathfrak{R}_\theta \cap (\delta, \infty)$, we conclude that

$$\lim_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta) \geq \delta) \geq -I_\delta(\theta), \tag{61}$$

where $I_\delta(\theta) = \inf_{x\in\mathfrak{R}_\theta\cap(\delta,\infty)} \Lambda_\theta^*(x)$.

By Proposition 4, choosing $M > I_\delta(\theta)$, one can find $\epsilon > 0$ such that

$$\boldsymbol{P}_g(\omega(\dot{\mathscr{A}}_n; \epsilon) > \delta) \leq e^{-Mn}. \tag{62}$$

Since

$$\boldsymbol{P}_g(\hat{\theta}_n \in G) \geq \boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta) \geq \delta) \left(1 - \frac{\boldsymbol{P}_g(\omega(\dot{\mathscr{A}}_n; \epsilon))}{\boldsymbol{P}_g(\dot{\mathscr{A}}_n(\theta) \geq \delta)}\right), \tag{63}$$

by the choice of M, it follows from (61) that

$$\liminf_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\theta}_n \in G) \geq -I_\delta(\theta). \tag{64}$$

Taking the supremum on left- and right-hand side over all $\delta > 0$ yields the required lower bound. □

Turning to the higher dimensional case, we first need the following result, which provides a uniform bound on the Hessian of the objective function $\mathscr{A}_n(\boldsymbol{\theta})$.

Lemma 2. *Under Hypotheses 1–8, there exists a finite constant* $0 < C < \infty$ *such that with probability one,*

$$\sup_{n\geq 1} \sup_{\boldsymbol{\theta}\in\Theta} ||H_{\mathscr{A}_n}(\boldsymbol{\theta})||_2 \leq C. \tag{65}$$

Proof. This is standard. Specifically, note that the $(i,j)^{\text{th}}$ element of the matrix $H_{\mathscr{A}_n}(\boldsymbol{\theta})$ is given by

$$h_{n,ij} = \int_S \ddot{s}_{\boldsymbol{\theta}}^{ij}(x) g_n^{\frac{1}{2}}(x) dx. \tag{66}$$

Next, writing down the expression for $\ddot{s}_{\boldsymbol{\theta}}^{ij}$ in terms of the derivatives of the score function $u_{\boldsymbol{\theta}}$, using the Cauchy–Schwarz inequality along with Hypotheses 3, 4, 6, and 8, and the definition of the matrix norm, the lemma follows. □

In the proof of the lower bound, we will take a somewhat different approach, involving the analysis of k constraints, and our strategy will be to reduce this to a problem involving a single constraint. Specifically, in (67) below, we establish that, instead of studying k constraints on a quantity $\mathscr{D}_n$ (which we are about to define), we can cast the problem in terms of a d-dimensional vector $\boldsymbol{Y}_n$ (defined in (70) below) belonging to a ball centered at $\mathbf{0}$ and of appropriate radius.

To be more precise, let $G \subset \mathbb{R}^d$ be open, and consider the probability that we obtain an estimated value $\boldsymbol{\theta} \in G$. Let $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k\} \subset \Theta - G$, and for any $\delta > 0$, set

$$d_n(j) = \mathscr{A}_n(\boldsymbol{\theta}) - \mathscr{A}_n(\boldsymbol{\theta}_j) - \delta, \quad j = 1, \ldots, k$$

and $\mathscr{D}_n(\boldsymbol{\theta}) = (d_n(1), \cdots d_n(k))$. If $\boldsymbol{\theta}$ is chosen as the estimate, then we must have $\mathscr{A}_n(\boldsymbol{\theta}) - \mathscr{A}_n(\boldsymbol{\theta}_j) \geq 0$ for all j, so, in particular,

$$\boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n \in G) \geq \boldsymbol{P}_g(\mathscr{D}_n(\boldsymbol{\theta}) \geq \mathbf{0}) \tag{67}$$

(by which we mean that $d_n(j) \geq 0$ for all j in this last probability).

To evaluate the latter probability, observe that by a second-order Taylor expansion,

$$s_{\boldsymbol{\theta}}(x) - s_{\boldsymbol{\theta}_j}(x) = \langle \boldsymbol{\theta} - \boldsymbol{\theta}_j, \nabla s_{\boldsymbol{\theta}}(x) \rangle + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_j)\mathscr{H}(x; \boldsymbol{\theta}_j^*)(\boldsymbol{\theta} - \boldsymbol{\theta}_j)'. \tag{68}$$

Using the positive definiteness and uniform boundedness of the matrix $\int_{\mathbb{R}} \mathscr{H}(x;\boldsymbol{\theta}) p^{\frac{1}{2}}(x)dx$, by Hypothesis 4, we have that for any unit vector $\boldsymbol{v} \in \mathbb{R}^d$,

$$\sup_{p \in \mathscr{G}} \inf_{\boldsymbol{\eta} \in \Theta} \left\{ \boldsymbol{v} \left(\int_{\mathbb{R}} \mathscr{H}(x;\boldsymbol{\eta}) p^{\frac{1}{2}}(x)dx \right) \boldsymbol{v}' \right\} \geq c,$$

where c is a positive constant independent of $\boldsymbol{v}$. Thus, for each j,

$$\sup_{p \in \mathscr{G}} \inf_{\boldsymbol{\eta} \in \Theta} \left\{ (\boldsymbol{\theta} - \boldsymbol{\theta}_j) \left(\int_{\mathbb{R}} \mathscr{H}(x;\boldsymbol{\eta}) p^{\frac{1}{2}}(x)dx \right) (\boldsymbol{\theta} - \boldsymbol{\theta}_j)' \right\} \geq c \left\| \boldsymbol{\theta} - \boldsymbol{\theta}_j \right\|^2. \tag{69}$$

Integrating with respect to $g_n^{\frac{1}{2}}(\cdot)$ and using the definition of $\mathscr{A}_n(\cdot)$, we then obtain that

$$d_n(j) = \int_{\mathbb{R}} \left[\langle \boldsymbol{\theta} - \boldsymbol{\theta}_j, \nabla s(x, \boldsymbol{\theta}) \rangle \right] g_n^{\frac{1}{2}}(x)dx + \mathscr{R}(\boldsymbol{\theta}, \boldsymbol{\theta}_j), \tag{70}$$

where

$$\mathscr{R}(\boldsymbol{\theta}, \boldsymbol{\theta}_j) \geq c \left\| \boldsymbol{\theta} - \boldsymbol{\theta}_j \right\|^2 - \delta.$$

Let $\boldsymbol{Y}_n(\boldsymbol{\theta}) = (Y_{n,1}, \ldots, Y_{n,d})$, where for $s(x;\boldsymbol{\theta}) := s_{\boldsymbol{\theta}}(x)$:

$$Y_{n,j} = \int_S \frac{\partial}{\partial \theta_j} s(x;\boldsymbol{\theta}) g_n^{\frac{1}{2}}(x)dx, \quad 1 \leq j \leq k. \tag{71}$$

(We have suppressed $\boldsymbol{\theta}$ in the notation for $Y_{n,j}$.) Then the inequality $d_n(j) \geq 0$ corresponds to an event $\mathscr{E}_{n,j}$ described by the occurrence of the inequality

$$\left\langle \frac{\boldsymbol{\theta} - \boldsymbol{\theta}_j}{||\boldsymbol{\theta} - \boldsymbol{\theta}_j||}, \boldsymbol{Y}_n \right\rangle \geq -c||\boldsymbol{\theta} - \boldsymbol{\theta}_j|| + \delta(||\boldsymbol{\theta} - \boldsymbol{\theta}_j||)^{-1}, \tag{72}$$

where the right-hand side is always negative for small δ (since $\text{dist}(\boldsymbol{\theta}, \Theta - G) > 0$) and behaves like a constant multiple of $\text{dist}(\boldsymbol{\theta}, \Theta - G)$ as this distance tends to infinity. Thus, we can choose a positive constant a_δ such that

$$a_\delta \, \text{dist}(\boldsymbol{\theta}, \Theta - G) \leq c||\boldsymbol{\theta} - \boldsymbol{\theta}_j|| - \delta(||\boldsymbol{\theta} - \boldsymbol{\theta}_j||)^{-1}, \quad j = 1, \ldots, k,$$

and set $c_{\boldsymbol{\theta}}(\delta) := a_\delta \, \text{dist}(\boldsymbol{\theta}, \Theta - G)$. Finally, let $\tilde{\mathscr{E}}_n$ denote the event that

$$\left\langle \frac{\boldsymbol{\theta} - \boldsymbol{\theta}_j}{||\boldsymbol{\theta} - \boldsymbol{\theta}_j||}, \boldsymbol{Y}_n \right\rangle \geq -c_{\boldsymbol{\theta}}(\delta). \tag{73}$$

Then for all j, $\mathscr{E}_{n,j} \supset \tilde{\mathscr{E}}_n$, where we recall that $\mathscr{E}_{n,j}$ was defined via (72). Now, since the definition of the event $\tilde{\mathscr{E}}_n$ does not depend on any specific vector $\boldsymbol{\theta}_j$, one can replace the vector $(\boldsymbol{\theta} - \boldsymbol{\theta}_j)(||\boldsymbol{\theta} - \boldsymbol{\theta}_j||)^{-1}$ by any unit vector $\boldsymbol{v}$ in $\mathbb{R}^d$. Hence

$$\boldsymbol{P}_g(\mathscr{D}_n \geq \mathbf{0}) \geq \boldsymbol{P}_g(\langle \boldsymbol{v}, \boldsymbol{Y}_n \rangle \geq -c_{\boldsymbol{\theta}}(\delta), \text{ for all unit vectors } \boldsymbol{v}) = \boldsymbol{P}_g(\boldsymbol{Y}_n \in \overline{B}(\mathbf{0}; c_{\boldsymbol{\theta}}(\delta))), \quad (74)$$

and we now derive a large deviation lower bound for the probability on the right-hand side.

Proposition 5. *Assume that Hypotheses 1–8 hold, and suppose that G is an open subset of $\mathbb{R}^d$. Assume that $b_n \searrow 0$ and $nb_n \nearrow \infty$ as $n \to \infty$. Then for any $\boldsymbol{\theta} \in G$ and $r > 0$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\boldsymbol{Y}_n \in B(\mathbf{0}; r)) \geq -I_r(\boldsymbol{\theta}), \quad (75)$$

where $I_r(\boldsymbol{\theta}) = \inf\{\Lambda^*_{\boldsymbol{\theta}}(\boldsymbol{x}) : \boldsymbol{x} \in \mathfrak{R}_{\boldsymbol{\theta}} \cap B(\mathbf{0}; r)\}$ and the infimum is taken to be infinity if the set $\mathfrak{R}_{\boldsymbol{\theta}} \cap B(\mathbf{0}; r)$ is empty.

Proof. We begin by studying the limiting generating function of $\boldsymbol{Y}_n$. By Varadhan's integral lemma, it follows that

$$\lim_{n\to\infty} \Lambda_{n,\boldsymbol{\theta}}(\boldsymbol{\alpha}) := \lim_{n\to\infty} \frac{1}{n} \log \boldsymbol{E}_g[\exp(n\langle \boldsymbol{\alpha}, \boldsymbol{Y}_n \rangle] = \Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}), \quad (76)$$

where

$$\Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) = \sup_{p \in \mathscr{G}} \left[\int_S \langle \boldsymbol{\alpha}, \nabla s_{\boldsymbol{\theta}}(x) \rangle p^{\frac{1}{2}}(x) dx - \mathrm{KL}(p, g) \right]. \quad (77)$$

Define the $\boldsymbol{\alpha}$-shifted distribution by

$$Q_{n,\boldsymbol{\alpha}}(B) = \frac{1}{\Lambda_{n,\boldsymbol{\theta}}(\boldsymbol{\alpha})} \int_B e^{n\langle \boldsymbol{\alpha}, \boldsymbol{y} \rangle} dQ_n(\boldsymbol{y}), \quad (78)$$

where Q_n denotes the distribution of $\boldsymbol{Y}_n$. Note by the convexity of $\Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha})$ that it is almost everywhere differentiable. Fix $x \in \mathfrak{R}_{\boldsymbol{\theta}} \cap B(\mathbf{0}; r)$ and choose $\boldsymbol{\alpha}$ such that $\nabla \Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) = \boldsymbol{x}$. Let $\delta > 0$ be such that $B(\boldsymbol{x}; \delta) \subsetneq B(\mathbf{0}; r)$. Then

$$\begin{aligned} Q_n(B(\boldsymbol{x}; \delta)) &= \exp(n\Lambda_{n,\boldsymbol{\theta}}(\boldsymbol{\alpha})) \int_{B(\boldsymbol{x};\delta)} \exp\left(-n\langle \boldsymbol{\alpha}, \boldsymbol{y} \rangle\right) dQ_{n,\boldsymbol{\alpha}}(\boldsymbol{y}) \\ &\geq \exp(n(-\langle \boldsymbol{\alpha}, \boldsymbol{x} \rangle + \Lambda_{n,\boldsymbol{\theta}}(\boldsymbol{\alpha}) + ||\boldsymbol{\alpha}||\delta)) Q_{n,\boldsymbol{\alpha}}(B(\boldsymbol{x}; \delta)), \end{aligned} \quad (79)$$

implying

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_n(B(\boldsymbol{x}; \delta)) \geq -\langle \boldsymbol{\alpha}, \boldsymbol{x} \rangle + \Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha}) - ||\boldsymbol{\alpha}||\delta + \liminf_{n\to\infty} \frac{1}{n} \log Q_{n,\boldsymbol{\alpha}}(B(\boldsymbol{x}; \delta)). \quad (80)$$

Now, notice that the limiting cumulant generating function of $\boldsymbol{Y}_n$ under the measure $Q_{n,\boldsymbol{\alpha}}$ is given by

$$\tilde{\Lambda}_{\boldsymbol{\theta}}(\boldsymbol{\beta}) = \Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha} + \boldsymbol{\beta}) - \Lambda_{\boldsymbol{\theta}}(\boldsymbol{\beta}). \quad (81)$$

Since $\tilde{\Lambda}_{\boldsymbol{\theta}}$ is a proper convex function, it is continuous since $\Lambda_{\boldsymbol{\theta}}(\boldsymbol{\alpha})$ is finite in the $\mathbb{R}^d$, and moreover, by the choice of $\boldsymbol{x}$, it is differentiable at $\mathbf{0}$. Hence Condition II.1 of [25] is satisfied. Now, using Theorem IV.1 of [25], it follows that

$$\liminf_{n\to\infty} \frac{1}{n} \log Q_{n,\boldsymbol{\alpha}}(B(\boldsymbol{x}; \delta)) = 0. \quad (82)$$

Substituting the above into (80), we obtain

$$\liminf_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\boldsymbol{Y}_n \in B(\mathbf{0};r)) \geq -\Lambda^*_{\boldsymbol{\theta}}(\boldsymbol{x}). \tag{83}$$

Taking the supremum in $\boldsymbol{x} \in \mathfrak{R}_{\boldsymbol{\theta}} \cap B(\mathbf{0};r)$, the proposition follows. □

Proof of Theorem 3: Upper Bound. Let F be a closed subset of Θ. Note Θ compact implies that F is compact. Let $\{B(\boldsymbol{\theta};r) : \boldsymbol{\theta} \in \Theta\}$ denote an open cover of F, and let $\{B(\boldsymbol{\theta}_1;r),\ldots,B(\boldsymbol{\theta}_k;r)\}$ denote the finite subcover. Using that $\nabla \mathscr{A}_n(\hat{\boldsymbol{\theta}}_n) = \mathbf{0}$, we then obtain that for any $\boldsymbol{\alpha} \in \mathbb{R}^d$,

$$\begin{aligned} \boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n \in F) &\leq \sum_{j=1}^{k} \boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n \in B(\boldsymbol{\theta}_k;r)) \\ &= \sum_{j=1}^{k} \boldsymbol{E}_g[\exp(n\langle \boldsymbol{\alpha}, \mathscr{A}'_n(\hat{\boldsymbol{\theta}}_n)\rangle) I_{\left\{\hat{\boldsymbol{\theta}}_n \in B(\boldsymbol{\theta}_j;r)\right\}}] := \sum_{j=1}^{k} T_n(j). \end{aligned} \tag{84}$$

Adding and subtracting $\nabla \mathscr{A}_n(\boldsymbol{\theta}_j)$ to $\nabla \mathscr{A}_n(\boldsymbol{\theta})$ and then applying Hölder's inequality yields $T_n(j) \leq T_n(1,j,p) T_n(2,j,q)$, where

$$\begin{aligned} \log T_n(1,j,p) &= \frac{1}{p} \log \boldsymbol{E}_g[\exp(np\langle \boldsymbol{\alpha}, \nabla \mathscr{A}_n(\boldsymbol{\theta}_j)\rangle) I_{\left\{\boldsymbol{\theta} \in B(\boldsymbol{\theta}_j;r)\right\}}], \\ \log T_n(2,j,q) &= \frac{1}{q} \log \boldsymbol{E}_g[\exp(nq\langle \boldsymbol{\alpha}, \nabla(\mathscr{A}_n(\hat{\boldsymbol{\theta}}_n) - \mathscr{A}_n(\boldsymbol{\theta}_j))\rangle) I_{\{\hat{\boldsymbol{\theta}}_n \in B(\boldsymbol{\theta}_j;r)\}}]. \end{aligned}$$

First we study $T_n(2,j,q)$. For $\hat{\boldsymbol{\theta}}_n \in B(\boldsymbol{\theta}_j, r_j)$ and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$, the Cauchy–Schwarz inequality gives

$$\begin{aligned} |\langle \boldsymbol{\alpha}, \nabla \mathscr{A}_n(\hat{\boldsymbol{\theta}}_n) - \nabla \mathscr{A}_n(\boldsymbol{\theta}_j)\rangle| &\leq ||\boldsymbol{\alpha}||_2 \sup_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2 \in B(\boldsymbol{\theta}_j,r)} ||\nabla \mathscr{A}_n(\boldsymbol{\theta}_1) - \nabla \mathscr{A}_n(\boldsymbol{\theta}_2))||_2 \\ &\leq ||\boldsymbol{\alpha}||_2 |r| \sup_{\boldsymbol{\theta} \in B(\boldsymbol{\theta}_j,r)} ||H_{\mathscr{A}_n}(\boldsymbol{\theta})||_2 \\ &\leq ||\boldsymbol{\alpha}||_2 |r| \max_{1\leq j\leq k} \left[\sup_{\boldsymbol{\theta} \in B(\boldsymbol{\theta}_j,r_j)} ||H_{\mathscr{A}_n}(\boldsymbol{\theta})||_2 \right], \end{aligned}$$

where $H_{\mathscr{A}_n}(\boldsymbol{\theta})$ is the Hessian matrix consisting of the second partial derivatives of $\mathscr{A}_n(\boldsymbol{\theta})$. Hence we obtain for any $1 \leq j \leq k$ that

$$\begin{aligned} \frac{1}{n} \log T_n(2,j,q) &\leq r \frac{1}{nq} (nq||\boldsymbol{\alpha}||_2) \max_{1\leq j\leq k} \left\{ \sup_{\boldsymbol{\theta} \in B(\boldsymbol{\theta}_j,r)} ||H_{\mathscr{A}_n}(\boldsymbol{\theta})||_2 \right\} \\ &= r||\alpha||_2 \max_{1\leq j\leq k} \left\{ \sup_{\boldsymbol{\theta} \in B(\boldsymbol{\theta}_j,r)} ||H_{\mathscr{A}_n}(\boldsymbol{\theta})||_2 \right\}. \end{aligned} \tag{85}$$

Now by Lemma 2,

$$\limsup_{n\to\infty} \frac{1}{n} \log T_n(2,j,q) \leq Cr. \tag{86}$$

Also, for each $1 \leq j \leq k$, Theorem 1 provides that

$$\limsup_{n\to\infty} \frac{1}{n} \log T_n(1,j,p) \leq \frac{1}{p} \Lambda_{\boldsymbol{\theta}_j}(p\boldsymbol{\alpha}). \tag{87}$$

Thus

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n\in F) &\leq \max_{1\leq j\leq k}\limsup_{n\to\infty}\frac{1}{n}\log T_n(1,j,p)+\max_{1\leq j\leq k}\limsup_{n\to\infty}\frac{1}{n}\log T_n(2,j,p)\\ &\leq \max_{1\leq j\leq k}\frac{1}{p}\Lambda_{\boldsymbol{\theta}_j}(p\alpha)+Cr. \end{aligned}\tag{88}$$

Since the last inequality holds for all $p>1$,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n\in F) &\leq \max_{1\leq j\leq k}\frac{1}{p}\Lambda_{\boldsymbol{\theta}_j}(p\boldsymbol{\alpha})+Cr\\ &\to \max_{1\leq j\leq k}\Lambda_{\boldsymbol{\theta}_j}(\boldsymbol{\alpha})+Cr \quad \text{as } p\searrow 0. \end{aligned}\tag{89}$$

Moreover, for each j,

$$\Lambda_{\boldsymbol{\theta}_j}(\boldsymbol{\alpha})\leq \sup_{\boldsymbol{\alpha}\in\mathbb{R}^d}\Lambda_{\boldsymbol{\theta}_j}(\boldsymbol{\alpha}):=-\Lambda_{\boldsymbol{\theta}_j}(\mathbf{0}).$$

Hence

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n\in F) &\leq \max_{1\leq j\leq k}-\Lambda^*_{\boldsymbol{\theta}_j}(\mathbf{0})+Cr\\ &\leq -\inf_{\boldsymbol{\theta}\in F}\Lambda^*_{\boldsymbol{\theta}}(\mathbf{0})+Cr. \end{aligned}\tag{90}$$

The upper bound follows by letting $r\searrow 0$. □

Proof of Theorem 3: Lower Bound. Let G be an open subset of Θ, and let $\boldsymbol{\theta}\in G$. Then $G^c=\Theta-G$ is compact, and there exists a collection $\mathbb{T}=\{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_k\}\subset G^c$ such that $B(\boldsymbol{\theta}_1;\epsilon),\ldots,B(\boldsymbol{\theta}_k;\epsilon)$ forms a finite subcover of $\Theta-G$, where $\epsilon>0$. Since

$$\begin{aligned}\left\{\mathscr{A}_n(\boldsymbol{\theta})\geq\sup_{\boldsymbol{t}\in\mathbb{T}}\mathscr{A}_n(\boldsymbol{t})\right\} &\supset \left\{\mathscr{A}_n(\boldsymbol{\theta})\geq\max_{1\leq j\leq k}\mathscr{A}_n(\boldsymbol{\theta}_j)+\max_{1\leq j\leq k}\sup_{\boldsymbol{t}\in B(\boldsymbol{\theta}_j;\epsilon)}[\mathscr{A}_n(\boldsymbol{t})-\mathscr{A}_n(\boldsymbol{\theta}_j)]\right\}\\ &\supset \left\{\mathscr{A}_n(\boldsymbol{\theta})\geq\max_{1\leq j\leq k}\mathscr{A}_n(\boldsymbol{\theta}_j)+\sup_{||\boldsymbol{\theta}_1-\boldsymbol{\theta}_2||<\epsilon}|\mathscr{A}_n(\boldsymbol{\theta}_1)-\mathscr{A}_n(\boldsymbol{\theta}_2)|\right\}\end{aligned}\tag{91}$$

it follows that

$$\begin{aligned}\boldsymbol{P}_g\hat{\boldsymbol{\theta}}_n\in G &\geq \boldsymbol{P}_g\left(\mathscr{A}_n(\boldsymbol{\theta})>\max_{1\leq j\leq k}\mathscr{A}_n(\boldsymbol{\theta}_j)+\delta,\ \sup_{||\boldsymbol{\theta}_1-\boldsymbol{\theta}_2||<\epsilon}[\mathscr{A}_n(\boldsymbol{\theta}_1)-\mathscr{A}_n(\boldsymbol{\theta}_2)]\leq\delta\right)\\ &\geq J_{n,1}-J_{n,2}, \end{aligned}\tag{92}$$

where

$$\begin{aligned}J_{n,1} &:= \boldsymbol{P}_g\left(\mathscr{A}_n(\boldsymbol{\theta})>\max_{1\leq j\leq k}\mathscr{A}_n(\boldsymbol{\theta}_j)+\delta\right),\\ J_{n,2} &:= \boldsymbol{P}_g\left(\sup_{||\boldsymbol{\theta}_1-\boldsymbol{\theta}_2||\leq r}[\mathscr{A}_n(\boldsymbol{\theta}_1)-\mathscr{A}_n(\boldsymbol{\theta}_2)]\geq\delta\right):=\boldsymbol{P}_g(\omega(\mathscr{A}_n;\epsilon)\geq\delta).\end{aligned}$$

We now investigate the behavior of $J_{n,1}$ and $J_{n,2}$. Starting with $J_{n,1}$, note that

$$J_{n,1} \geq \boldsymbol{P}_g\left(\min_{1\leq j\leq k}(\mathscr{A}_n(\boldsymbol{\theta}) - \mathscr{A}_n(\boldsymbol{\theta}_j) - \delta) \geq 0\right) = \boldsymbol{P}_g(\mathscr{D}_n \geq \mathbf{0}). \tag{93}$$

Now by (74), it follows that

$$J_{n,1} \geq \boldsymbol{P}_g(\boldsymbol{Y}_n \in B(\mathbf{0}; r)), \tag{94}$$

where Y_n is as in (71) and $r = c_{\boldsymbol{\theta}}(\delta)$. Applying Proposition 3.4, we obtain

$$\lim_{n\to\infty} \frac{1}{n} \log J_{n,1} \geq -I_r(\boldsymbol{\theta}), \tag{95}$$

where $I_r(\boldsymbol{\theta}) = \inf\{\Lambda^*_{\boldsymbol{\theta}}(\boldsymbol{x}) : \boldsymbol{x} \in \mathfrak{R}_{\boldsymbol{\theta}} \cap B(\mathbf{0}; r)\}$, and we now observe that r may be chosen to be $c_{\boldsymbol{\theta}} := \lim_{\delta\downarrow 0} c_{\boldsymbol{\theta}}(\delta) > 0$, where $c_{\boldsymbol{\theta}}(\delta)$ is given as in (73). Hence we may replace $I_r(\cdot)$ with $I(\cdot)$ on the right-hand side of the previous equation. Next, using Proposition 4 yields that

$$\liminf_{n\to\infty} \frac{1}{n} \log \boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n \in G) \geq \liminf_{n\to\infty} \frac{1}{n} \log J_{n,1} + \lim_{n\to\infty} \log\left(1 - \frac{J_{n,2}}{J_{n,1}}\right) \geq -I(\boldsymbol{\theta}). \tag{96}$$

Finally, the required lower bound is obtained by maximizing the right-hand side over all $\boldsymbol{\theta} \in G$. □

In the proof of the lower bound, it is clear that the choice of $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k\}$ plays a central role, and the rate function $I(\boldsymbol{\theta})$ will be minimized when k is small. As a simple example, suppose that our goal is to obtain a lower bound for $\boldsymbol{P}_g(\hat{\boldsymbol{\theta}}_n \in G)$, where

$$G = \{(\theta_1, \theta_2) : \theta_1 > a_1 \text{ or } \theta_2 > a_2\} \subset \mathbb{R}^2, \quad \boldsymbol{\theta}_g \notin G,$$

which is a union of two halfspaces, This can be expressed as $\boldsymbol{a} + \mathscr{C}$, where $\boldsymbol{a} = (a_1, a_2)$ and $\mathscr{C} = \{(\theta_1, \theta_2) : \theta_1 > 0 \text{ or } \theta_2 > 0\}$, which is an example of a *translated* cone. Now if $\boldsymbol{\theta} \in G$, then we can find two elements which generate the entire set $\Theta - G$, in the sense that all other normalized differences lie between these two unit vectors. These two representative points are the unit vectors $\mathbf{e}_1 = (-1, 0)$ and $\mathbf{e}_2 = (0, -1)$, and all other normalized differences $(\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}/\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\|$ lie between these vectors for all $\tilde{\boldsymbol{\theta}} \in \Theta - G$. Now going back to (73), we see that this equation again holds. Furthermore, (74) holds with $B(\mathbf{0}; c_\delta(\boldsymbol{\theta}))$ now replaced by an intersection of *two* halfspaces rather than of all halfspaces, yielding an unbounded region in the definition of $I(\boldsymbol{\theta})$. This potentially improves the quality of the lower bound compared with what is presented in the statement of Theorem 3. This idea can be potentially generalized to other sets, such as other unions of halfspaces, and so from a practical perspective, could apply somewhat generally.

4. Concluding Remarks

In this article, we have derived large deviation results for the minimum Hellinger distance estimators of a family of continuous distributions satisfying an equicontinuity condition. These results extend large deviation asymptotics for M-estimators given, e.g., in [6,9]. In contrast to the case for M-estimators, our setting is complicated due to its inherent nonlinearity, leading to complications in the proofs of both the upper and lower bounds, and an unexpected subtlety in the form of the rate function for the lower bound. Our results suggest that one can, under additional hypotheses, establish saddlepoint approximations to the density of MHDE, which would enable one to sharpen inference for small samples.

Similar results are expected to hold for discrete distributions. However, the equicontinuity condition is not required in that case, since ℓ_1, unlike $L_1(S)$, possesses the *Schur property*. Hence the LDP in the weak topology of ℓ_1 can be derived (more easily) using a standard Gärtner–Ellis argument, and utilizing this, one can, in principle, repeat all of the

arguments above to derive results analogous to Theorems 2 and 3. Large deviations for other divergences under weak family regularity (such as noncompactness of the parameter space Θ)—and their connections to estimation and test efficiency—are interesting open problems requiring new techniques beyond those described in this article.

Author Contributions: Conceptualization, A.N.V. and J.F.C.; Methodology, A.N.V. and J.F.C.; Validation, A.N.V. and J.F.C.; Writing—original draft, A.N.V. and J.F.C.; Writing—review & editing, A.N.V. and J.F.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Beran, R. Minimum Hellinger distance estimates for parametric models. *Ann. Stat.* **1977**, *5*, 445–463.
2. Lindsay, B.G. Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Ann. Stat.* **1994**, *22*, 1081–1114.
3. Basu, A.; Shioya, H.; Park, C. *Statistical Inference: The Minimum Distance Approach*; CRC Press: Boca Raton, FL, USA, 2011.
4. Pardo, L. *Statistical Inference Based on Divergence Measures*; CRC Press: Boca Raton, FL, USA, 2006.
5. Bahadur, R.R. Rates of convergence of estimates and test statistics. *Ann. Math. Stat.* **1967**, *38*, 303–324.
6. Borovkov, A.A.; Mogulskii, A.A. Large Deviations and Testing Statistical Hypotheses. *Sib. Adv. Math. 2*, 43–72.
7. Fu, J.C. On a theorem of Bahadur on the rate of convergence of point estimators. *Ann. Stat.* **1973**, *1*, 745–749.
8. Arcones, M.A. Large deviations for M-estimators. *Ann. Inst. Stat. Math.* **2006**, *58*, 21–52.
9. Joutard, C. Large deviations for M-estimators. *Math. Methods Stat.* 2004, *13*, 179–200.
10. Dembo, A.; Zeitouni, O. *Large Deviations Techniques and Applications*; Springer: Berlin, Germany, 1998.
11. Puhalskii, A.; Spokoiny,V. On large-deviation efficiency in statistical inference. *Bernoulli* **1998**, *4*, 203–272.
12. Nikitin, Y. *Asymptotic Efficiency of Nonparametric Tests*; Cambridge University Press: Cambridge, UK, 1995.
13. Biggins, J.; Bingham, N. Large deviations in the supercritical branching process. *Adv. Appl. Probab.* **1993**, *25*, 757–772.
14. Billingsley, P. *Convergence of Probability Measures*, 2nd ed.; John Wiley & Sons, Inc.: New York, NY, USA, 1999.
15. de Acosta, A. On large deviations of empirical measures in the τ-topology. *J. Appl. Probab.* **1993**, *31*, 41–47.
16. Basu, A.; Sarkar, S.; Vidyashankar, A.N. Minimum negative exponential disparity estimation in parametric models. *J. Statist. Plann. Inference* **1997**, *58*, 349–370.
17. Cheng, A.-L.; Vidyashankar, A.N. Minimum Hellinger distance estimation for randomized play the winner design. *J. Statist. Plann. Inference* **2006**, *136*, 1875–1910.
18. Devroye, L.; Györfi, L. *Nonparametric Density Estimation: The L_1 View*; Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics; John Wiley & Sons, Inc.: New York, NY, USA, 1985.
19. Conway, J.B. *A Course in Functional Analysis*; Springer: New York, NY, USA, 1990.
20. Dupuis, P.; Ellis, R.S. *A Weak Convergence Approach to the Theory of Large Deviations*; John Wiley & Sons: New York, NY, USA, 1997.
21. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.
22. Boos, D.D. A converse to Scheffé's theorem. *Ann. Stat.* **1985**, *13*, 423–427.
23. Lei, L. Large Deviations for Kernel Density Estimators and Study for Random Decrement Estimator. Ph. D. Thesis, Université Blaise Pascal-Clermont-Ferrand II, Clermont-Ferrand, France, 2005.
24. Louani, D.; Maouloud, S.M.O. Some functional large deviations principles in nonparametric function estimation. *J. Theor. Probab.* **2012**, *25*, 280–309.
25. Ellis, R.S. Large deviations for a general class of random vectors. *Ann. Probab.* **1984**, *12*, 1–12.