Bernoulli **4**(2), 1998, 203–272

On large-deviation efficiency in statistical inference

ANATOLII PUHALSKII[1] and VLADIMIR SPOKOINY[2*]

[1]*Institute for Problems in Information Transmission, 19 Bolshoi Karetnii, 101447 Moscow, Russia*
[2]*Institute of Applied Analysis and Stochastics, Mohrenstrasse 39, 10117 Berlin, Germany*
e-mail: spokoiny@wias-berlin.de

We present a general approach to statistical problems with criteria based on probabilities of large deviations. Our main idea, which originates from similarity in the definitions of the large-deviation principle (LDP) and weak convergence, is to develop a large-deviation analogue of asymptotic decision theory. We introduce the concept of the LDP for sequences of statistical experiments, which parallels the concept of weak convergence of experiments, and prove that, in analogy with Le Cam's minimax theorem, the LDP provides an asymptotic lower bound for the sequence of appropriately defined minimax risks. We also show that the bound is tight and give a method of constructing decisions whose asymptotic risk is arbitrarily close to the bound. The construction is further specified for hypothesis testing and estimation problems.

We apply the results to a number of standard statistical models: an independent and identically distributed sample, regression, the change-point model and others. For each model, we check the LDP; then, considering first a hypothesis testing problem and then an estimation problem, we calculate the asymptotic minimax risks and indicate associated decisions.

Keywords: Bahadur efficiency; Chernoff's function; large-deviation efficiency; large-deviation principle; minimax risk; statistical experiments

1. Introduction

The approach to statistical problems that bases its conclusions on the study of probabilities of large deviations has been in use in statistical inference since the papers by Chernoff (1952) and Bahadur (1960).

Chernoff (1952), considering the problem of discriminating between two simple hypotheses, showed that, if the hypotheses are fixed, the error probabilities decrease exponentially fast as the sample size tends to infinity; the corresponding optimal exponent is specified by what is now known as Chernoff's function.

Basu (1956) and Bahadur (1960) proposed a criterion for comparing statistical estimators

*To whom correspondence should be addressed.

1350–7265

based on the view that the quality of an estimator is characterized by the probability that the true value of the parameter is covered by a confidence interval of given width $2c$ with centre at the estimate. If the width $2c$ is held fixed as the sample size grows, then the probabilities that the true value of the parameter is not covered are typically exponentially small. The estimator giving the fastest decay is now called Bahadur efficient. Later Bahadur *et al.* (1980) showed, for the model of independent and identically distributed observations, that in the class of consistent estimators the optimal rate is specified by the Kullback–Leibler information rather than Chernoff's function.

The ideas of Chernoff and Bahadur have been developed in various directions. Ibragimov and Radavicius (1981), Kallenberg (1981), Ibragimov and Khasminskii (1981) and Radavicius (1983; 1991) studied the properties of maximum likelihood estimators from the point of view of Bahadur's criterion. Fu (1982) and Borovkov and Mogulskii (1992a; 1992b) analysed the second- and higher-order terms of asymptotic expansions of Bahadur risks. Kallenberg (1983), Rao (1963), Wieand (1976) and Ermakov (1993) considered intermediate criteria for statistical estimators when the width of the confidence interval goes to zero at a certain rate. Sievers (1978) and Rubin and Rukhin (1983) evaluated Bahadur risks for particular statistical models.

Lately this direction in mathematical statistics has received a new impetus, mostly in papers by Korostelev (1996; 1995) – see also Korostelev and Spokoiny (1996) and Korostelev and Leonov (1995) – where the classical large-deviation (LD) set-up is considered in the minimax nonparametric framework.

Our aim here is to give a unified treatment of statistical problems that use LD considerations. The idea is to capitalize on analogies between LD theory and weak convergence theory (see Lynch and Sethuraman 1987; Vervaat 1988; Puhalskii 1991) and develop an LD analogue of asymptotic decision theory (Strasser 1985). The approach of invoking the methods of weak convergence theory to obtain results about large deviations has proved its worth in various set-ups (Puhalskii 1991; 1993; 1994a; 1994b; 1995; 1996; 1997). We show that it can successfully be applied to statistical problems too.

We begin by defining in Section 2 the concept of the large-deviation principle (LDP) for a sequence of statistical experiments. Analogously to the concept of weak convergence of statistical experiments, it is a short-cut for saying that the distributions of suitably defined likelihood processes satisfy the LDP (Varadhan 1966; 1984). We illustrate the general definition by considering a number of standard statistical models (the Gaussian shift model, the model of independent and identically distributed observations, the 'signal plus white noise' model, the regression model with Gaussian and non-Gaussian errors, with deterministic and random design, and the change-point model). We next study properties of the LDP for statistical experiments and give a sufficient condition for it which is analogous to the local asymptotic normality condition of Le Cam (1960).

The classical minimax theorem of Le Cam states that if statistical experiments weakly converge then the minimax risks are asymptotically bounded from below by the corresponding risk for the limit model (see Le Cam 1972; 1986; Strasser 1985). In Section 3, we show that, similarly, if a sequence of statistical experiments obeys the LDP, then there is an asymptotic lower bound for appropriately defined minimax risks. The problem of evaluating the bound is a minimax optimization problem. Also in Section 3, we study the question of sharpness of the lower bound. We show that it is sharp under a strengthened version of the LDP. This allows us to define LD efficient decisions as those that attain the lower bound. We give a method of obtaining nearly LD efficient decisions, i.e., those whose LD asymptotic risk is arbitrarily close to the lower bound.

Sections 4 and 5 deal with applications. Section 4 adapts the results of Section 3 to the cases of hypothesis testing and estimation problems and presents explicit constructions of nearly LD efficient decisions. In Section 5, we apply the machinery to the models introduced in Section 2: we check the LDP, give conditions when the lower bounds are attained, calculate them for hypothesis testing and estimation problems, and indicate nearly LD efficient decisions. An appendix contains extensions and auxiliary results.

The results of Sections 2–4 are new. The results that we obtain for the models are partly new and partly cover or extend earlier results.

2. The large-deviation principle for statistical experiments

Let $\{\mathscr{E}_n, n \geqslant 1\}$ be a sequence of statistical experiments $\mathscr{E}_n = (\Omega_n, \mathscr{F}_n; P_{n,\theta}, \theta \in \Theta)$ with a parameter set Θ (Strasser 1985). In this section, we give the definition of the LDP for $\{\mathscr{E}_n, n \geqslant 1\}$ and study its properties. We start with the case of dominated experiments.

2.1. The dominated case

Let us assume that each experiment $\mathscr{E}_n = (\Omega_n, \mathscr{F}_n; P_{n,\theta}, \theta \in \Theta)$ is dominated by a probability measure P_n, i.e., $P_{n,\theta} \ll P_n$ for all $\theta \in \Theta$. We abbreviate this by writing $\{\mathscr{E}_n, P_n, n \geqslant 1\}$. Denote

$$Z_{n,\theta} = \left(\frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}\right)^{1/n}, \qquad \theta \in \Theta, \tag{2.1}$$

and let $Z_{n,\Theta} = (Z_{n,\theta}, \theta \in \Theta)$. We endow $\mathbb{R}_+^\Theta$ with the Tihonov (product) topology and the Borel σ-field so that $Z_{n,\Theta}$ is a random element of $\mathbb{R}_+^\Theta$; $\mathscr{L}(Z_{n,\Theta}|P_n)$ denotes the distribution of $Z_{n,\Theta}$ on $\mathbb{R}_+^\Theta$ under P_n. Roughly speaking, the LDP for $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ means that the sequence $\{\mathscr{L}(Z_{n,\Theta}|P_n), n \geqslant 1\}$ of distributions on $\mathbb{R}_+^\Theta$ obeys the LDP, so we recall some basic notions of LD theory.

We use Varadhan's (1966; 1984) original definitions of the rate function and the LDP. Let S be a Hausdorff topological space. We say that a function I: $S \to [0, \infty]$ is a rate function on S if the sets $I^{-1}([0, a])$ are compact in S for all $a \geqslant 0$. A sequence $\{Q_n, n \geqslant 1\}$ of probability measures on the Borel σ-field of S is said to obey the LDP with rate function I if

$$\varliminf_{n\to\infty} \frac{1}{n} \log Q_n(G) \geqslant -\inf_{x\in G} I(x)$$

for all open $G \subset S$ and

$$\overline{\lim_{n\to\infty}}\ \frac{1}{n}\log Q_n(F) \leqslant -\inf_{x\in F} I(x)$$

for all closed $F \subset S$.

We also say that I is a probability rate function if $\inf_{x\in S} I(x) = 0$. Obviously, if I appears in the LDP, it is a probability rate function.

Recall that the contraction principle states that continuous mappings preserve the LDP (Varadhan 1966; 1984).

Next, we say that the sequence $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ satisfies condition (U) if

$$(U)\ \lim_{H\to\infty} \overline{\lim}_{n\to\infty} \mathrm{E}_n^{1/n} Z_{n,\theta}^n 1(Z_{n,\theta} > H) = 0, \qquad \theta \in \Theta.$$

Here and below, E_n denotes an expectation with respect to P_n and, by definition, $\mathrm{E}_n^{1/n}\xi = (\mathrm{E}_n\xi)^{1/n}$, $P_n^{1/n}(A) = (P_n(A))^{1/n}$.

Definition 2.1. *We say that a sequence $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ of dominated statistical experiments obeys the dominated large-deviation principle if:*

1. *the sequence $\{\mathscr{L}(Z_{n,\Theta}|P_n), n \geqslant 1\}$ obeys the LDP with some (probability) rate function I;*
2. *condition (U) holds.*

A critical part of the definition is condition 1. Condition (U) plays a subordinate though essential role. If we disregard condition (U), the definition is analogous to the definition of weak convergence of dominated statistical experiments (Strasser 1985) which states that the likelihood ratios weakly converge. The role of condition (U) will become clear shortly: it ensures the compatibility of this definition with a more general one which does not depend on a choice of dominating measures and incorporates the non-dominated case too. In particular, condition (U) implies that the lower bound that we obtain in Section 3 for the sequence of so-called LD risks does not depend on dominating measures either (see Remark 3.2 below). Note that an analogue of condition (U) in the theory of weak convergence of statistical experiments is a consequence of weak convergence of the likelihood ratios and does not have to be singled out.

In applications, rather than considering $Z_{n,\theta}$, it is more convenient to deal with log-likelihood ratios $\Xi_{n,\theta}$ defined as

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}.$$

Let us introduce $\Xi_{n,\Theta} = (\Xi_{n,\theta}, \theta \in \Theta)$ and denote by $\mathscr{L}(\Xi_{n,\Theta}|P_n)$ the distribution of $\Xi_{n,\Theta}$ on $\mathbb{R}^\Theta$ under P_n, where $\mathbb{R}^\Theta$ is supplied with the Tihonov topology and the Borel σ-field. If the $\Xi_{n,\theta}$ are well defined then, by the contraction principle, the LDP for the sequence $\{\mathscr{L}(\Xi_{n,\Theta}|P_n), n \geqslant 1\}$ implies the LDP for the sequence $\{\mathscr{L}(Z_{n,\Theta}|P_n), n \geqslant 1\}$.

Now we consider a number of statistical models which, on the one hand, show that the LDP for the log-likelihood ratios arises quite naturally and, on the other hand, motivate and illustrate theoretical developments below. We stop short of giving rigorous proofs of the LDP for the models, deferring this until Section 5.

Example 2.1 *Gaussian observations*. Let us observe a sample of n independent real-valued random variables $\mathbf{X}_n = (X_{1,n}, \ldots, X_{n,n})$ normally distributed with $\mathcal{N}(\theta, 1)$, $\theta \in \Theta \subset \mathbb{R}$. For this model, $\Omega_n = \mathbb{R}^n$ and $P_{n,\theta} = (\mathcal{N}(\theta, 1))^n$, $\theta \in \Theta$. We take $P_{n,0}$ as a dominating measure P_n. Then the corresponding log-likelihood ratios are of the form

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(\mathbf{X}_n) = \frac{1}{n}\sum_{k=1}^{n}\left(\theta X_{k,n} - \frac{1}{2}\theta^2\right) = \theta Y_n - \frac{1}{2}\theta^2,$$

where

$$Y_n = \frac{1}{n}\sum_{k=1}^{n} X_{k,n}, \qquad n \geqslant 1.$$

The sequence $\{\mathscr{L}(Y_n|P_n),\ n \geqslant 1\}$ obeys the LDP in $\mathbb{R}$ with rate function $I^N(y) = y^2/2$, $y \in \mathbb{R}$ (see, e.g., Freidlin and Wentzell 1979). This yields by the contraction principle the LDP for the log-likelihood ratios $\Xi_{n,\theta}$.

Example 2.2 *An independent and identically distributed sample*. Let $\mathbf{X}_n = (X_{1,n}, \ldots, X_{n,n})$ be an independent and identically distributed sample from a distribution P_θ, $\theta \in \Theta$, on the real line. We do not specify the nature of the parameter set Θ. For example, it can be a subset of a finite-dimensional space, a set of distributions on $\mathbb{R}$ (or their probability density functions), etc. We assume that the family $\mathscr{P}$ is dominated by a probability measure P, i.e., $P_\theta \ll P$, $\theta \in \Theta$. This model is described by dominated experiments $\mathscr{E}_n = (\Omega_n, \mathscr{F}_n; P_{n,\theta}, \theta \in \Theta)$ with $\Omega_n = \mathbb{R}^n$, $\mathscr{F}_n = \mathscr{B}(\mathbb{R}^n)$, $P_{n,\theta} = P_\theta^n$, $\theta \in \Theta$ and $P_n = P^n$.

We have

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(\mathbf{X}_n) = \sum_{k=1}^{n} \frac{1}{n} \log \frac{\mathrm{d}P_\theta}{\mathrm{d}P}(X_{k,n}) = \int_{\mathbb{R}} \log \frac{\mathrm{d}P_\theta}{\mathrm{d}P}(x) F_n(\mathrm{d}x),$$

where

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} 1(X_{k,n} \leqslant x), \qquad x \in \mathbb{R},$$

is an empirical distribution function.

Let $\mathscr{Y}$ be the space of cumulative distribution functions on $\mathbb{R}$ with the topology of weak convergence of associated probability measures. By Sanov's theorem (Sanov 1957; Deuschel and Stroock 1989, Section 3.2.17), the sequence $\{\mathscr{L}(F_n|P_n),\ n \geqslant 1\}$ obeys the LDP in $\mathscr{Y}$ with rate function $I^S(F) = K(F, P)$, $F \in \mathscr{Y}$, where $K(F, P)$ denotes the Kullback–Leibler information:

$$K(F, P) = \begin{cases} \displaystyle\int_{\mathbb{R}} \frac{\mathrm{d}F}{\mathrm{d}P}(x) \log \frac{\mathrm{d}F}{\mathrm{d}P}(x) P(\mathrm{d}x), & \text{if } F \ll P, \\ \infty, & \text{otherwise.} \end{cases}$$

Let us also denote, for $\theta \in \Theta$ and $F \in \mathscr{Y}$,

$$\zeta_\theta(F) = \int_{\mathbb{R}} \log \frac{\mathrm{d}P_\theta}{\mathrm{d}P}(x) F(\mathrm{d}x).$$

If the density functions $(\mathrm{d}P_\theta/\mathrm{d}P)(x)$ are bounded from above, bounded away from zero and continuous in x for all $\theta \in \Theta$, then the $\zeta_\theta(F)$ are continuous functions on $\mathscr{Y}$ and, since $\Xi_{n,\theta} = \zeta_\theta(F_n)$, the contraction principle yields the LDP for the sequence $\{\Xi_{n,\Theta},\ n \geqslant 1\}$.

Example 2.3 '*Signal plus white noise*'. We observe a real-valued stochastic process $X_n = (X_n(t),\ t \in [0, 1])$ obeying the stochastic differential equation

$$\mathrm{d}X_n(t) = \theta(t)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\,\mathrm{d}W(t), \qquad 0 \leqslant t \leqslant 1,$$

where $W = (W(t),\ t \in [0, 1])$ is a standard Wiener process and $\theta(\cdot)$ is an unknown function assumed to belong to some set Θ of real-valued continuous functions on [0, 1].

This model is described by statistical experiments $\mathscr{E}_n = (\Omega_n, \mathscr{F}_n;\ P_{n,\theta},\ \theta \in \Theta)$, where Ω_n is $C[0, 1]$, the space of continuous functions on [0, 1] with the uniform metric and Borel σ-field, and $P_{n,\theta}$ is the distribution of X_n on $C[0, 1]$ for θ. We take $P_n = P_{n,0}$, where $P_{n,0}$ corresponds to the zero function $\theta(\cdot) \equiv 0$. Then $P_{n,\theta} \ll P_n$ and, moreover, by Girsanov's formula, P_n-almost surely,

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n) = \int_0^1 \theta(t)\,\mathrm{d}X_n(t) - \frac{1}{2}\int_0^1 \theta^2(t)\,\mathrm{d}t. \tag{2.2}$$

Let $C_0[0, 1]$ be the subset of $C[0, 1]$ of the functions $x(\cdot)$ that are absolutely continuous with respect to Lebesgue measure and equal to 0 at 0. Then the sequence $\{\mathscr{L}(X_n|P_n),\ n \geqslant 1\}$ obeys the LDP in $C[0, 1]$ with rate function

$$I^W(x(\cdot)) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^1 (\dot{x}(t))^2\,\mathrm{d}t, & \text{if } x(\cdot) \in C_0[0, 1], \\ \infty, & \text{otherwise,} \end{cases}$$

$\dot{x}(t)$ denoting the derivative of $x(\cdot)$ at t (see, e.g., Freidlin and Wentzell, 1979).

Let us denote, for functions $\theta(\cdot) \in \Theta$ and $x(\cdot) \in C_0[0, 1]$,

$$\zeta_\theta(x) = \int_0^1 \theta(t)\,\mathrm{d}x(t) - \frac{1}{2}\int_0^1 \theta^2(t)\,\mathrm{d}t,$$

where the integral is understood as a Lebesgue–Stieltjes integral.

Again the log-likelihood ratio $\Xi_{n,\theta}$ can formally be represented as $\Xi_{n,\theta} = \zeta_\theta(X_n)$. Note, however, that the first integral in (2.2) is an Itô integral, so the latter equality as well as the continuity property for ζ_θ actually holds for functions $\theta(\cdot)$ of a special sort (e.g., piecewise constant or differentiable). For these functions, the contraction principle again implies the LDP for $\{\Xi_{n,\Theta},\ n \geqslant 1\}$. A general case is studied in Section 5.

Example 2.4 *Gaussian regression.* We consider the regression model

$$X_{k,n} = \theta(t_{k,n}) + \xi_{k,n}, \qquad t_{k,n} = \frac{k}{n},\ k = 1, \ldots, n, \tag{2.3}$$

where errors $\xi_{k,n}$ are independent standard normal and $\theta(\cdot)$ is an unknown real-valued continuous function.

In this model, $\Omega_n = \mathbb{R}^n$, $\Theta \subset C[0, 1]$ and $P_{n,\theta}$ is the distribution of $\mathbf{X}_n = (X_{1,n}, \ldots, X_{n,n})$ for $\theta(\cdot)$. As above, we take $P_n = P_{n,0}$. Then

$$\begin{aligned}\Xi_{n,\theta} &= \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n) \\ &= \frac{1}{n}\sum_{k=1}^{n} \theta(t_{k,n})X_{k,n} - \frac{1}{2n}\sum_{k=1}^{n} \theta^2(t_{k,n}) \\ &= \int_0^1 \theta(t)\,\mathrm{d}X_n(t) - \frac{1}{2n}\sum_{k=1}^{n} \theta^2(t_{k,n}),\end{aligned}$$

where

$$X_n(t) = \frac{1}{n}\sum_{k=1}^{[nt]} X_{k,n}, \qquad 0 \leqslant t \leqslant 1.$$

Let $\mathscr{Y}$ be the space of right-continuous functions on [0, 1] with left-hand limits and with the uniform metric (for measurability of X_n, see Billingsley 1968, Section 8).

Since the $X_{k,n}$ are distributed as $\mathscr{N}(0, 1)$ under P_n, the sequence $\{\mathscr{L}(X_n|P_n), n \geqslant 1\}$ obeys the LDP in $\mathscr{Y}$ with rate function I^W (Mogulskii 1976, Theorem 2).

Since the function $\theta(\cdot)$ is continuous, we have, for large n, the approximate equality

$$\frac{1}{n}\sum_{k=1}^{n} \theta^2(t_{k,n}) \approx \int_0^1 \theta^2(t)\,\mathrm{d}t$$

and hence $\Xi_{n,\theta} \approx \zeta_\theta(X_n)$, with the same function ζ_θ as in the preceding example. If the θ are differentiable, integration by parts shows that the $\Xi_{n,\theta}$ are continuous functions of the X_n, and the LDP for $\{\Xi_{n,\Theta}, n \geqslant 1\}$ follows by the contraction principle. Again, a general case is deferred until Section 5.

Example 2.5 *Non-Gaussian regression.* We consider the same regression model (2.3) but now assume that independent and identically distributed errors $\xi_{k,n}$ have a distribution P with a positive probability density function $p(x)$ with respect to Lebesgue measure on the real line. An unknown regression function $\theta(\cdot)$ is assumed to be continuous, so $\Theta \subset C[0, 1]$.

As above, for a regression function $\theta(\cdot)$, we denote by $P_{n,\theta}$ the distribution of $X_n = (X_{1,n}, \ldots, X_{n,n})$. We have, with $P_n = P_{n,0}$,

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n) = \frac{1}{n}\sum_{k=1}^{n} \log \frac{p(X_{k,n} - \theta(t_{k,n}))}{p(X_{k,n})}.$$

Introducing the empirical process $F_n = F_n(x, t)$, $x \in \mathbb{R}$, $t \in [0, 1]$, by

$$F_n(x,\ t) = \frac{1}{n}\sum_{k=1}^{[nt]} 1(X_{k,n} \leqslant x),$$

we have that

$$\Xi_{n,\theta} = \int_0^1\int_{\mathbb{R}} \log \frac{p(x-\theta(t))}{p(x)} F_n(\mathrm{d}x,\ \mathrm{d}t). \tag{2.4}$$

Let us define $\mathscr{Y}$ as the space of cumulative distribution functions $F = F(x,\ t)$, $x \in \mathbb{R}$, $t \in [0,\ 1]$, on $\mathbb{R} \times [0,\ 1]$ with the weak topology. Let $\mathscr{Y}_0$ be the subset of $\mathscr{Y}$ of functions $F(x,\ t)$ absolutely continuous with respect to Lebesgue measure on $\mathbb{R} \times [0,\ 1]$ and with densities $p_t(x)$ such that $\int_{\mathbb{R}} p_t(x)\,\mathrm{d}x = 1$ for $t \in [0,\ 1]$.

It is shown in Dembo and Zajic (1995) – see also Theorem 1 in Puhalskii (1996) – that the sequence $\{\mathscr{L}(F_n|P_n),\ n \geqslant 1\}$ obeys the LDP in $\mathscr{Y}$ with rate function $I^{SK}(F)$ given by

$$I^{SK}(F) = \begin{cases} \displaystyle\int_0^1\int_{\mathbb{R}} \log \frac{p_t(x)}{p(x)} p_t(x)\,\mathrm{d}x\,\mathrm{d}t, & \text{if } F \in \mathscr{Y}_0,\\ \infty, & \text{otherwise.} \end{cases}$$

Denote, for $F \in \mathscr{Y}_0$ and $\theta \in \Theta$,

$$\zeta_\theta(F) = \int_0^1\int_{\mathbb{R}} \log \frac{p(x-\theta(t))}{p(x)} F(\mathrm{d}x,\ \mathrm{d}t).$$

Then by (2.4), $\Xi_{n,\theta} = \zeta_\theta(F_n)$ and if the logs in the integrals in the definition of the ζ_θ are bounded and continuous, we have the LDP for $\{\Xi_{n,\Theta},\ n \geqslant 1\}$.

Example 2.6 *The change-point model.* Let us observe a sample $X_n = (X_{1,n},\ \ldots,\ X_{n,n})$ of real-valued random variables, where, for some $k_n \geqslant 1$, the observations $X_{1,n},\ \ldots,\ X_{k_n,n}$ are independent and identically distributed with a distribution P_0 and the observations $X_{k_n+1,n},\ \ldots,\ X_{n,n}$ are independent and identically distributed with a distribution P_1. We assume that P_0 and P_1 are known and k_n is unknown. Let us also assume that $k_n = [n\theta]$, where $\theta \in \Theta = [0,\ 1]$. For this model, $\Omega_n = \mathbb{R}^n$ and $P_{n,\theta}$ stands for the distribution of X_n for θ.

Let a probability measure P dominate P_0 and P_1 and

$$f_0(x) = \frac{\mathrm{d}P_0}{\mathrm{d}P}(x),\ f_1(x) = \frac{\mathrm{d}P_1}{\mathrm{d}P}(x), \qquad x \in \mathbb{R},$$

be respective densities. Assume that $f_0(x)$ and $f_1(x)$ are positive and continuous. Denoting $P_n = P^n$, we have

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n) = \frac{1}{n}\sum_{i=1}^{[n\theta]} \log f_0(X_{i,n}) + \frac{1}{n}\sum_{i=[n\theta]+1}^{n} \log f_1(X_{i,n}),$$

so that, defining an empirical process again by

$$F_n(x,\ t) = \frac{1}{n}\sum_{i=1}^{[nt]} 1(X_{i,n} \leqslant x), \qquad x \in \mathbb{R},\ t \in [0,\ 1],$$

we obtain the representation

$$\Xi_{n,\theta} = \int_0^\theta \int_{\mathbb{R}} \log f_0(x) F_n(\mathrm{d}x,\ \mathrm{d}t) + \int_\theta^1 \int_{\mathbb{R}} \log f_1(x) F_n(\mathrm{d}x,\ \mathrm{d}t).$$

Let a space $\mathscr{Y}$ be defined as for the preceding model and $\mathscr{Y}_P$ be the set of those $F \in \mathscr{Y}$ that are absolutely continuous with respect to the measure $P(\mathrm{d}x) \times \mathrm{d}t$ and admit densities $p_t(x)$ such that $\int_{\mathbb{R}} p_t(x) P(\mathrm{d}x) = 1,\ t \in [0,\ 1]$. As above, the F_n obey the LDP with rate function I_P^{SK} of the form

$$I_P^{SK}(F) = \begin{cases} \displaystyle\int_0^1 \int_{\mathbb{R}} p_t(x) \log p_t(x) P(\mathrm{d}x)\, \mathrm{d}t, & \text{if } F \in \mathscr{Y}_P, \\ \infty, & \text{otherwise.} \end{cases}$$

Define next for $F \in \mathscr{Y}_P$

$$\zeta_\theta(F) = \int_0^\theta \int_{\mathbb{R}} \log f_0(x) F(\mathrm{d}x,\ \mathrm{d}t) + \int_\theta^1 \int_{\mathbb{R}} \log f_1(x) F(\mathrm{d}x,\ \mathrm{d}t).$$

Then again $\Xi_{n,\theta} = \zeta_\theta(F_n)$, and the LDP for $\{\Xi_{n,\Theta},\ n \geqslant 1\}$ holds, e.g., when $\log f_0(x)$ and $\log f_1(x)$ are bounded and continuous.

Example 2.7 *Regression with random design.* We consider the model

$$X_{k,n} = \theta(t_{k,n}) + \xi_{k,n}, \qquad k = 1,\ \ldots,\ n,$$

where real-valued errors $\xi_{k,n}$ and design points $t_{k,n}$ are independent with respective distributions P and Π dominated by Lebesgue measure. We denote the respective densities by $p(x)$ and $\pi(t)$. We also assume that the prior measure Π has a compact support D, $\pi(t)$ is continuous and positive on the support, $p(x)$ is continuous and positive on $\mathbb{R}$, and an unknown regression function $\theta(\cdot)$ is continuous.

In this model, $P_{n,\theta}$ is the joint distribution of $X_n = (X_{1,n},\ \ldots,\ X_{n,n})$, and $t_n = (t_{1,n},\ \ldots,\ t_{n,n})$ for θ. Let F_n be the joint empirical distribution function of X_n and t_n:

$$F_n(A,\ B) = \frac{1}{n}\sum_{k=1}^{n} 1(X_{k,n} \in A,\ t_{k,n} \in B)$$

for Borel sets $A \subset \mathbb{R}$, $B \subset D$, and let $\mathscr{Y}$ be the space of distributions on $\mathbb{R} \times D$ with the weak topology. Set also $P_n = P_{n,0} = (P \times \Pi)^n$.

With these definitions,

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n,\ t_n)$$
$$= \frac{1}{n}\sum_{k=1}^{n}\log\frac{p(X_{k,n} - \theta(t_{k,n}))}{p(X_{k,n})}$$
$$= \int_D\int_R \log\frac{p(x-\theta(t))}{p(x)}F_n(\mathrm{d}x,\ \mathrm{d}t).$$

Let $\mathscr{Y}_1$ be the set of the cumulative distribution functions on $\mathbb{R}^2$ that are absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^2$ and have support in $\mathbb{R}\times D$. Under P_n, the random pairs $(X_{k,n},\ t_{k,n})$ are independent and identically distributed with the distribution $P\times\Pi$, and hence, by Sanov's theorem, the LDP holds for the F_n with rate function $I^{SS}(F)$ given by

$$I^{SS}(F) = \begin{cases} \displaystyle\int_D\int_{\mathbb{R}}\log\frac{p(x,\ t)}{p(x)\pi(t)}p(x,\ t)\,\mathrm{d}x\,\mathrm{d}t, & \text{if } F\in\mathscr{Y}_1,\\ \infty, & \text{otherwise.}\end{cases}$$

Here $F(\mathrm{d}x,\ \mathrm{d}t) = p(x,\ t)\,\mathrm{d}x\,\mathrm{d}t$. The LDP for this model follows now in a manner similar to the case of an independent and identically distributed sample.

We end this subsection with a simple but useful remark. It is noticeable that the definition of the LDP given above uses the same letter n both to subscript probability measures and associated random elements, and to denote a scaling parameter. One might wonder whether this is not a loss of generality and how n should be chosen when considering particular models. The answer to the first question is in the negative and making n play the two roles economizes on notation. Indeed, if we have a sequence of probability measures $\{Q_n,\ n\geqslant 1\}$ with $\log Q_n$ having the right rate $b_n\to\infty$ as $n\to\infty$, we can always reduce this case to the above 'standard' set-up by 'relabelling' the measures, i.e., by introducing measures Q'_n such that $Q'_{b_n} = Q_n$; taking b_n as a new n then gives $\log Q'_n$ the rate n as required. This argument, originating from Varadhan (1984), also answers the second question: n in our formalism has the meaning of the right scale rather than 'the natural parameter of the model'. Of course, the two can coincide, as in most of the examples we considered where n is a sample size, but not always, as Example 2.3 shows. On the other hand, it is clear from the above that if we want n to be 'the natural parameter', we can do this by introducing some $b_n\to\infty$ as a scale.

2.2. Sufficient conditions for the dominated LDP

We now study properties of the LDP for statistical experiments and begin with a sufficient condition for the LDP. The condition serves two purposes: first, in particular statistical models it is easier to check than the definition of the LDP; and second, this condition is useful when constructing asymptotically optimal decisions (see Section 4). The idea behind the condition is similar to that used in the condition of local asymptotic normality by Le Cam

(1960) for studying weak convergence of experiments, or, more generally, in the condition of λ-convergence by Shiryaev and Spokoiny (1997).

Given a sequence of dominated statistical experiments $\{\mathscr{E}_n, P_n, n \geq 1\}$, assume that there exist statistics Y_n on $(\Omega_n, \mathscr{F}_n)$ with values in a Hausdorff space $\mathscr{Y}$ such that the sequence $\{\mathscr{L}(Y_n|P_n), n \geq 1\}$ obeys the LDP and the Y_n are asymptotically sufficient in the sense that $Z_{n,\theta} \approx \mathfrak{z}_\theta(Y_n)$ for some non-random functions $\mathfrak{z}_\theta$ on $\mathscr{Y}$. In the above examples the statistic Y_n is easily identified: it is the empirical mean $(X_{1,n} + \ldots + X_{n,n})/n$ in the case of a sample from the normal distribution in Example 2.1, the empirical distribution function F_n in the case of an independent and identically distributed sample in Example 2.2, the observation process X_n for the 'signal plus white noise' model, the empirical process F_n for the regression model with non-Gaussian errors and the change-point model, etc.

If the functions $\mathfrak{z}_\theta$ are continuous then, by the contraction principle, the LDP for the sequence $\{\mathscr{L}(Y_n|P_n), n \geq 1\}$ implies the LDP for the sequence $\{\mathscr{L}(\mathfrak{z}_\theta(Y_n)|P_n), n \geq 1\}$ and hence for $\{\mathscr{L}(Z_{n,\theta}|P_n), n \geq 1\}$. Unfortunately, by contrast with the theory of weak convergence of experiments, in applications the functions $\mathfrak{z}_\theta$ typically are not continuous. For instance, the functions $\zeta_\theta(y) = \log \mathfrak{z}_\theta(y)$ generally are not continuous in the above examples for an independent and identically distributed sample, the 'signal plus white noise' model, the regression models and the change-point model. To overcome this difficulty, we need to introduce 'regularizations' $\mathfrak{z}_{\theta,\delta}(y)$ of $\mathfrak{z}_\theta(y)$ that, on the one hand, are continuous functions and, on the other hand, converge to $\mathfrak{z}_\theta(y)$ as $\delta \to 0$.

Before stating the condition, let us review some more facts about large deviations used below. Recall (Varadhan 1966; 1984; Deuschel and Stroock 1989; Bryc 1990) that if a sequence of probability measures $\{Q_n, n \geq 1\}$ on the Borel σ-field of a Hausdorff space S obeys the LDP with rate function I then, for all non-negative bounded continuous functions f on S,

$$\lim_{n\to\infty} \left[\int_S (f(x))^n Q_n(\mathrm{d}x)\right]^{1/n} = \sup_{x\in S} f(x)V(x), \tag{2.5}$$

where $V(x) = \exp(-I(x))$. If S is a metric, or, more generally, a Tihonov (i.e., completely regular) space (Engelking 1977; Kelley 1957) then (2.5) also is sufficient for the LDP (Puhalskii 1993).

Moreover, the LDP implies (2.5) also for unbounded continuous non-negative functions f under 'the uniform exponential integrability condition' (Varadhan 1984; Deuschel and Stroock 1989)

$$\lim_{H\to\infty} \overline{\lim_{n\to\infty}} \left[\int_S (f(x))^n 1(f(x) > H) Q_n(\mathrm{d}x)\right]^{1/n} = 0. \tag{2.6}$$

Also, if f is a lower semi-continuous non-negative function then

$$\underline{\lim_{n\to\infty}} \left[\int_S (f(x))^n Q_n(\mathrm{d}x)\right]^{1/n} \geq \sup_{x\in S} f(x)V(x). \tag{2.7}$$

The function $V(x)$ is henceforth referred to as a deviability. Equivalently, a deviability is defined as a function $V\colon S \to [0, 1]$ such that $\sup_{x\in S} V(x) = 1$ and the inverse images

$V^{-1}([a, 1])$ are compact sets for all $a>0$. Obviously, there is one-to-one correspondence between probability rate functions and deviabilities. We say that $\{Q_n, n \geq 1\}$ LD converges to V and write $Q_n \overset{l.d.}{\to} V$ $(n \to \infty)$ if (2.5) holds for all bounded continuous non-negative functions f (Puhalskii 1994a). Below we use the fact that, if S is metric, then one can only require that the functions f be uniformly continuous (analogously to weak convergence theory; Billingsley 1968, Theorem 2.1). By the above, if S is a Tihonov space then $Q_n \overset{l.d.}{\to} V$ $(n \to \infty)$ if and only if $\{Q_n\}$ obeys the LDP with $I = -\log V$. All the spaces we consider below are Tihonov and we mostly use the formulation of the LDP as LD convergence as more convenient in theoretical considerations.

Next, let S and S' be Hausdorff spaces and V a deviability on S. Denote

$$\Phi_V(a) = \{x \in S\colon V(x) \geq a\}, \qquad a>0. \tag{2.8}$$

As in Puhalskii (1997) – cf. Schwartz (1973) – we say that a map $\varphi\colon S \to S'$ is V-Luzin measurable if it is continuous in restriction to each set $\Phi_V(a)$, $a>0$. The term V-Luzin is motivated by the following analogy with Luzin's theorem in measure theory. Let us extend V to a set function on S by defining $V(\Gamma) = \sup_{x\in\Gamma} V(x)$, $\Gamma \subset S$. Then V as a set function is an analogue of probability (for a discussion see Puhalskii, 1991; 1994; 1995), and, equivalently, a function φ is V-Luzin measurable if, for every $\varepsilon>0$, there exists a set $A \subset S$ with $V(S\backslash A) < \varepsilon$ such that φ is continuous in restriction to A. It is also interesting to note that one can prove an analogue of Egorov's theorem for sequences of Luzin measurable functions Puhalskii (1991, 1997). Deviabilities are preserved under Luzin measurable maps: for any V-Luzin measurable map φ, the function $V \circ \varphi^{-1}$ on S', defined by $V \circ \varphi^{-1}(x') = \sup_{x\in\varphi^{-1}(x')} V(x)$, $x' \in S'$, is a deviability on S' – see Deuschel and Stroock (1989, Section 2.1.4); the argument of Puhalskii (1991, Lemma 2.1) also applies.

Also, we say that a function $\varphi\colon S \to S'$ is V-almost everywhere (V-a.e.) continuous if it is continuous at every $x \in S$ with $V(x)>0$. Obviously, each V-a.e. continuous function is V-Luzin measurable.

Some more notational conventions are in order. We denote by $\mathscr{A}(\Theta)$ the family of all finite subsets of Θ. Elements of $\mathbb{R}_+^\Theta$ are denoted by $z_\Theta = (z_\theta, \theta \in \Theta)$, and elements of $\mathbb{R}_+^\Lambda$, where $\Lambda \in \mathscr{A}(\Theta)$, by $z_\Lambda = (z_\theta, \theta \in \Lambda)$. Maps π_Λ and $\pi_{\Lambda'\Lambda}$, where $\Lambda \in \mathscr{A}(\Theta)$, $\Lambda' \in \mathscr{A}(\Theta)$ and $\Lambda \subset \Lambda'$, are the natural projections of $\mathbb{R}_+^\Theta$ onto $\mathbb{R}_+^\Lambda$ and of $\mathbb{R}_+^{\Lambda'}$ onto $\mathbb{R}_+^\Lambda$, respectively: $\pi_\Lambda(z_\theta, \theta \in \Theta) = (z_\theta, \theta \in \Lambda)$ and $\pi_{\Lambda'\Lambda}(z_\theta, \theta \in \Lambda') = (z_\theta, \theta \in \Lambda)$. Since $\mathbb{R}_+^\Theta$ and $\mathbb{R}_+^\Lambda$, $\Lambda \in \mathscr{A}(\Theta)$, are supplied with the Tihonov topology, the projections are continuous.

We now state and prove the sufficient condition for the LDP. We thereby assume that the statistics Y_n take values in a metric space which is enough for applications, though this restriction can be relaxed.

Lemma 2.1. *Let $\{\mathscr{E}_n, P_n, n \geq 1\}$ be a sequence of dominated experiments and $Z_{n,\theta}$, $\theta \in \Theta$, be defined by (2.1). Assume that the following condition holds:*

(Y) *there exist statistics $Y_n\colon \Omega_n \to \mathscr{Y}$ with values in a metric space $\mathscr{Y}$ with the Borel σ-field, functions $\mathfrak{z}_\theta\colon \mathscr{Y} \to \mathbb{R}_+$, $\theta \in \Theta$, and $\mathfrak{z}_{\theta,\delta}\colon \mathscr{Y} \to \mathbb{R}_+$, $\theta \in \Theta$, $\delta>0$, such that*

(Y.1) *the sequence $\{\mathscr{L}(Y_n|P_n), n \geq 1\}$ of distributions on $\mathscr{Y}$ LD converges to a deviability $V(y)$, $y \in \mathscr{Y}$;*

(Y.2) *for all* $\delta > 0$, *the functions* $\mathfrak{z}_{\theta,\delta}\colon \mathscr{Y} \to \mathbb{R}_+$, $\theta \in \Theta$, *are Borel measurable and* V*-a.e. continuous;*
(Y.3) $\lim_{\delta\to 0} \overline{\lim}_{n\to\infty} P_n^{1/n}(|Z_{n,\theta} - \mathfrak{z}_{\theta,\delta}(Y_n)| > \varepsilon) = 0$ *for all* $\varepsilon > 0$ *and* $\theta \in \Theta$;
(Y.4) $\lim_{\delta\to 0} \sup_{y\in\Phi_V(a)} |\mathfrak{z}_{\theta,\delta}(y) - \mathfrak{z}_\theta(y)| = 0$ *for all* $a > 0$ *and* $\theta \in \Theta$.

Then $\mathscr{L}(Z_{n,\Theta}|P_n) \overset{l.d.}{\to} V_\Theta$ $(n \to \infty)$, *where* $V_\Theta = V \circ \mathfrak{z}_\Theta^{-1}$, $\mathfrak{z}_\Theta = (\mathfrak{z}_\theta, \theta \in \Theta)$.

Proof. Conditions (Y.2) and (Y.4) obviously imply that $\mathfrak{z}_\Theta\colon \mathscr{Y} \to \mathbb{R}_+^\Theta$ is V-Luzin measurable, hence V_Θ is a deviability on $\mathbb{R}_+^\Theta$.

Let $\Lambda \in \mathscr{A}(\Theta)$. We first prove that

$$\mathscr{L}(Z_{n,\Lambda}|P_n) \overset{l.d.}{\to} V_\Lambda, \qquad n \to \infty, \tag{2.9}$$

where $Z_{n,\Lambda} = (Z_{n,\theta}, \theta \in \Lambda)$, $V_\Lambda = V \circ \mathfrak{z}_\Lambda^{-1}$ and $\mathfrak{z}_\Lambda = (\mathfrak{z}_\theta, \theta \in \Lambda)$. Let $f\colon \mathbb{R}_+^\Lambda \to \mathbb{R}_+$ be bounded and uniformly continuous. Since, by the definition of V_Λ, $\sup_{z_\Lambda\in\mathbb{R}_+^\Lambda} f(z_\Lambda)V_\Lambda(z_\Lambda) = \sup_{y\in\mathscr{Y}} f(\mathfrak{z}_\Lambda(y))V(y)$, we need to prove that

$$\lim_{n\to\infty} \mathrm{E}_n^{1/n} f^n(Z_{n,\Lambda}) = \sup_{y\in\mathscr{Y}} f(\mathfrak{z}_\Lambda(y))V(y). \tag{2.10}$$

Let $\mathfrak{z}_{\Lambda,\delta} = (\mathfrak{z}_{\theta,\delta}, \theta \in \Lambda)$. Condition ($Y$.3) implies, in view of the boundedness and uniform continuity of f, that

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}} |\mathrm{E}_n^{1/n} f^n(Z_{n,\Lambda}) - \mathrm{E}_n^{1/n} f^n(\mathfrak{z}_{\Lambda,\delta}(Y_n))| = 0. \tag{2.11}$$

Since the sequence $\{\mathscr{L}(Y_n|P_n), n \geqslant 1\}$ LD converges to V and the map $\mathfrak{z}_{\Lambda,\delta}\colon \mathscr{Y} \to \mathbb{R}_+^\Lambda$ is V-a.e. continuous, the sequence $\{\mathscr{L}(\mathfrak{z}_{\Lambda,\delta}(Y_n)|P_n), n \geqslant 1\}$ LD converges to $V \circ (\mathfrak{z}_{\Lambda,\delta})^{-1}$ (Puhalskii 1991, Theorem 2.2). Thus, since f is non-negative, bounded and continuous,

$$\lim_{n\to\infty} \mathrm{E}_n^{1/n} f^n(\mathfrak{z}_{\Lambda,\delta}(Y_n)) = \sup_{y\in\mathscr{Y}} f(\mathfrak{z}_{\Lambda,\delta}(y))V(y). \tag{2.12}$$

By (2.11) and (2.12), for (2.10) it remains to show that

$$\lim_{\delta\to 0} \sup_{y\in\mathscr{Y}} f(\mathfrak{z}_{\Lambda,\delta}(y))V(y) = \sup_{y\in\mathscr{Y}} f(\mathfrak{z}_\Lambda(y))V(y), \tag{2.13}$$

which is an easy consequence of condition (Y.4). Convergence (2.9) is proved. The assertion of the lemma now follows by the Dawson–Gärtner theorem on the projective limits of LD systems (Dawson and Gärtner 1987, Theorem 3.3) if we note that $\mathscr{L}(Z_{n,\Theta}|P_n)$ is the projective limit of $\{\mathscr{L}(Z_{n,\Lambda}|P_n), \Lambda \in \mathscr{A}(\Theta)\}$ and $V_\Lambda = V_\Theta \circ \pi_\Lambda^{-1}$, $\Lambda \in \mathscr{A}(\Theta)$. □

Remark 2.1. Since $\mathbb{R}_+^\Theta$ is a Tihonov space, the lemma implies that, under conditions (Y) and (U), the sequence $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ obeys the dominated LDP.

Remark 2.2. As we have seen, in applications it is more convenient to manipulate rate functions and log-likelihood ratios given by

$$\Xi_{n,\theta} = \log Z_{n,\theta} = \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}, \qquad \theta \in \Theta.$$

Accordingly, it is useful to state condition (Y) in these terms. Assume that the $\Xi_{n,\theta}$ are well defined. It is easy to see that condition (Y) is implied by the following condition:

(Y') *there exist statistics* $Y_n\colon \Omega_n \to \mathscr{Y}$ *with values in a metric space* $\mathscr{Y}$ *with the Borel* σ-*field, functions* $\zeta_\theta\colon \mathscr{Y} \to \mathbb{R}$, $\theta \in \Theta$, *and* $\zeta_{\theta,\delta}\colon \mathscr{Y} \to \mathbb{R}$, $\theta \in \Theta$, $\delta > 0$, *such that*

- $(Y'.1)$ *the sequence* $\{\mathscr{L}(Y_n|P_n),\ n \geqslant 1\}$ *of distributions on* $\mathscr{Y}$ *obeys the LDP with rate functions* $I(y)$, $y \in \mathscr{Y}$;
- $(Y'.2)$ *for all* $\delta > 0$, *the functions* $\zeta_{\theta,\delta}\colon \mathscr{Y} \to \mathbb{R}$, $\theta \in \Theta$, *are Borel measurable and continuous at each point* y *such that* $I(y) < \infty$;
- $(Y'.3)$ $\lim_{\delta\to 0} \overline{\lim}_{n\to\infty} P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(Y_n)| > \varepsilon) = 0$ *for all* $\varepsilon > 0$ *and* $\theta \in \Theta$;
- $(Y'.4)$ $\lim_{\delta\to 0} \sup_{y\in\Phi'_I(a)} |\zeta_{\theta,\delta}(y) - \zeta_\theta(y)| = 0$ *for all* $a \geqslant 0$ *and* $\theta \in \Theta$, *where* $\Phi'_I(a) = \{y \in \mathscr{Y}\colon I(y) \leqslant a\}$.

Condition (U) takes the form

(U') $\lim_{H\to\infty} \overline{\lim}_{n\to\infty} \mathrm{E}_n^{1/n} \exp(n\Xi_{n,\theta}) 1(\Xi_{n,\theta} > H) = 0$, $\theta \in \Theta$.

By Lemma 2.1, conditions (Y') and (U') imply the dominated LDP.

2.3. The general case

The above definition of the LDP for statistical experiments covers only the dominated case and depends on a choice of dominating measures. We now present another definition which is free of these defects. It is motivated by Le Cam's definition of weak convergence of experiments (see, e.g., Strasser 1985).

Let $|\Lambda|$ denote the number of elements in $\Lambda \in \mathscr{A}(\Theta)$. For $z_\Lambda = (z_\theta,\ \theta \in \Lambda) \in \mathbb{R}_+^\Lambda$ and $z_\Theta = (z_\theta,\ \theta \in \Theta) \in \mathbb{R}_+^\Theta$, we set $\|z_\Lambda\|_\Lambda = \max_{\theta\in\Lambda} z_\theta$ and $\|z_\Theta\|_\Theta = \max_{\theta\in\Theta} z_\theta$, respectively, and define $S_\Lambda = \{z_\Lambda \in \mathbb{R}_+^\Lambda\colon \|z_\Lambda\|_\Lambda = 1\}$ and $S_\Theta = \{z_\Theta \in \mathbb{R}_+^\Theta\colon \|z_\Theta\|_\Theta = 1\}$. In order not to overburden notation, we sometimes omit the subscript Λ in $\|\cdot\|_\Lambda$ if there is no risk of confusion.

Next, given a sequence of statistical experiments $\{\mathscr{E}_n,\ n \geqslant 1\}$, where $\mathscr{E}_n = (\Omega_n, \mathscr{F}_n;\ P_{n,\theta},\ \theta \in \Theta)$, set, for $\Lambda \in \mathscr{A}(\Theta)$,

$$P_{n,\Lambda} = \frac{1}{|\Lambda|} \sum_{\theta\in\Lambda} P_{n,\theta},$$

$$\mathbf{Z}_{n,\theta;\Lambda} = \left(\frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_{n,\Lambda}}\right)^{1/n}, \qquad \theta \in \Lambda, \tag{2.14}$$

$$\mathbf{Z}_{n,\Lambda} = (\mathbf{Z}_{n,\theta;\Lambda},\ \theta \in \Lambda).$$

The definitions immediately imply that, $P_{n,\Lambda}$-almost surely,

$$\sum_{\theta\in\Lambda} \mathbf{Z}^n_{n,\theta;\Lambda} = |\Lambda| \tag{2.15}$$

and

$$1 \leqslant \|\mathbf{Z}_{n,\Lambda}\| \leqslant |\Lambda|^{1/n}. \tag{2.16}$$

Definition 2.2. *A sequence $\{\mathscr{E}_n,\ n \geqslant 1\}$ of statistical experiments obeys the LDP if, for each $\Lambda \in \mathscr{A}(\Theta)$, the sequence $\{\mathscr{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}),\ n \geqslant 1\}$ of distributions on $\mathbb{R}_+^\Lambda$ obeys the LDP with some rate function.*

Remark 2.3. Equivalently, $\{\mathscr{E}_n,\ n \geqslant 1\}$ obeys the LDP if $\mathscr{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \overset{l.d.}{\to} \mathbf{V}_\Lambda$, $\Lambda \in \mathscr{A}(\Theta)$, where $\mathbf{V}_\Lambda$ is a deviability on $\mathbb{R}_+^\Lambda$.

We next study consequences of the definition and, particularly, prove that the definitions of the LDP for the dominated and general cases are consistent. We start by giving another characterization of the LDP. Let $\mathscr{H}_\Lambda$ denote the set of all non-negative, continuous and positively homogeneous functions on $\mathbb{R}_+^\Lambda$: $h \in \mathscr{H}_\Lambda$ if $h(z_\Lambda) \geqslant 0$, h is continuous and $h(\lambda z_\Lambda) = \lambda h(z_\Lambda)$ for all $z_\Lambda \in \mathbb{R}_+^\Lambda$ and $\lambda \geqslant 0$. We say that a deviability $\mathbf{V}_\Lambda$ has support in S_Λ if $\mathbf{V}_\Lambda(z_\Lambda) = 0$ for $z_\Lambda \notin S_\Lambda$.

Lemma 2.2. *Let $\Lambda \in \mathscr{A}(\Theta)$. Then $\mathscr{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \overset{l.d.}{\to} \mathbf{V}_\Lambda$ if and only if $\mathbf{V}_\Lambda$ has support in S_Λ and*

$$\lim_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) = \sup_{z_\Lambda\in\mathbb{R}_+^\Lambda} h(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda) \qquad \textit{for every } h \in \mathscr{H}_\Lambda.$$

In particular, if $\mathscr{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \overset{l.d.}{\to} \mathbf{V}_\Lambda$ then, for all $\theta \in \Lambda$,

(R) $\sup_{z_\Lambda\in\mathbb{R}_+^\Lambda} \pi_\theta z_\Lambda \mathbf{V}_\Lambda(z_\Lambda) = 1.$

Proof. Let $\mathscr{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \overset{l.d.}{\to} \mathbf{V}_\Lambda$. We have, using the equivalence of LD convergence and the LDP on $\mathbb{R}_+^\Lambda$, that, for $\varepsilon > 0$,

$$\varliminf_{n\to\infty} P_{n,\Lambda}^{1/n}(|\,\|\mathbf{Z}_{n,\Lambda}\| - 1| > \varepsilon) \geqslant \sup_{z_\Lambda:|\|z_\Lambda\|-1|>\varepsilon} \mathbf{V}_\Lambda(z_\Lambda).$$

Inequalities (2.16) imply that the left-hand side equals zero. Since ε is arbitrary, $\mathbf{V}_\Lambda$ has support in S_λ. The claimed limit follows by the definition of LD convergence since, by (2.16), $h(\mathbf{Z}_{n,\Lambda}) = \hat{h}(\mathbf{Z}_{n,\Lambda})$ $P_{n,\Lambda}$-almost surely, where $\hat{h}(z_\Lambda) = h(z_\Lambda)[(2 - \|z_\Lambda\|/\Lambda) \wedge 1 \vee 0]$, and the latter function is non-negative, bounded and continuous.

For the converse, pick a non-negative continuous bounded function f on $\mathbb{R}_+^\Lambda$. We need to prove that

$$\lim_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} f^n(\mathbf{Z}_{n,\Lambda}) = \sup_{z_\Lambda\in\mathbb{R}_+^\Lambda} f(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda). \tag{2.17}$$

We define a function $\tilde{f}$ by

$$\tilde{f}(z_\Lambda) = \begin{cases} \|z_\Lambda\| f\left(\dfrac{z_\Lambda}{\|z_\Lambda\|}\right), & \text{if } \|z_\Lambda\| > 0, \\ 0, & \text{if } \|z_\Lambda\| = 0. \end{cases}$$

Note that f and $\tilde{f}$ coincide on S_Λ and, since $\mathbf{V}_\Lambda$ is supported by S_Λ, we can change f to $\tilde{f}$ on the right-hand side of (2.17). The continuity of f and the inequalities (2.16) easily imply that the random variables $f(\mathbf{Z}_{n,\Lambda})$ and $\tilde{f}(\mathbf{Z}_{n,\Lambda})$ are uniformly bounded and

$$\lim_{n\to\infty} |\mathrm{E}_{n,\Lambda}^{1/n} f^n(\mathbf{Z}_{n,\Lambda}) - \mathrm{E}_{n,\Lambda}^{1/n} \tilde{f}^n(\mathbf{Z}_{n,\Lambda})| = 0.$$

Since $\tilde{f} \in \mathscr{H}_\Lambda$, taking $h = \tilde{f}$ in the conditions of the lemma, we obtain

$$\lim_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} \tilde{f}^n(\mathbf{Z}_{n,\Lambda}) = \sup_{z_\Lambda \in \mathbb{R}_+^\Lambda} \tilde{f}(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda),$$

concluding the proof of (2.17).

Property (R) follows by taking $h(z_\Lambda) = \pi_\theta z_\Lambda$. □

We now show that if $\Lambda \subset \Lambda' \in \mathscr{A}(\Theta)$ then the deviability $\mathbf{V}_\Lambda$ is a sort of projection of the deviability $\mathbf{V}_{\Lambda'}$, the property being inherited from associated probabilities. Recall the notation $\pi_{\Lambda'\Lambda}$ and π_Λ for the projections from $\mathbb{R}_+^{\Lambda'}$ onto $\mathbb{R}_+^\Lambda$ and $\mathbb{R}_+^\Theta$ onto $\mathbb{R}_+^\Lambda$, respectively, and let $\Pi_{\Lambda'\Lambda}$ and Π_Λ stand for normalized projections:

$$\Pi_{\Lambda'\Lambda} z_{\Lambda'} = \pi_{\Lambda'\Lambda} z_{\Lambda'} / \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda, \qquad z_{\Lambda'} \in \mathbb{R}_+^{\Lambda'},\ \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda > 0,$$

$$\Pi_\Lambda z_\Theta = \pi_\Lambda z_\Theta / \|\pi_\Lambda z_\Theta\|_\Lambda, \qquad z_\Theta \in \mathbb{R}_+^\Theta,\ \|\pi_\Lambda z_\Theta\|_\Lambda > 0.$$

Also we adhere to the convention that $\sup_\varnothing = 0$.

Lemma 2.3. *Let* $\Lambda \subset \Lambda' \in \mathscr{A}(\Theta)$. *If* $\mathscr{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \xrightarrow{l.d.} \mathbf{V}_\Lambda$ *and* $\mathscr{L}(\mathbf{Z}_{n,\Lambda'}|P_{n,\Lambda'}) \xrightarrow{l.d.} \mathbf{V}_{\Lambda'}$ *then the following conditions hold:*

(C) $\sup_{z_\Lambda \in \mathbb{R}_+^\Lambda} h(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_{\Lambda'} \in \mathbb{R}_+^{\Lambda'}} h(\pi_{\Lambda'\Lambda} z_{\Lambda'})\mathbf{V}_{\Lambda'}(z_{\Lambda'}), \qquad h \in \mathscr{H}_\Lambda;$
(S) $\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_{\Lambda'} \in \Pi_{\Lambda'\Lambda}^{-1} z_\Lambda} \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}), \qquad z_\Lambda \in \mathbb{R}_+^\Lambda,$

where $\Pi_{\Lambda'\Lambda}^{-1} z_\Lambda = \{z_{\Lambda'} \in \mathbb{R}_+^{\Lambda'}\colon \Pi_{\Lambda'\Lambda} z_{\Lambda'} = z_\Lambda\}$.

Proof. Define

$$\mathbf{Z}_{n,\Lambda;\Lambda'} = \left(\frac{\mathrm{d}P_{n,\Lambda}}{\mathrm{d}P_{n,\Lambda'}}\right)^{1/n}.$$

By (2.14),

$$\pi_{\Lambda'\Lambda}\mathbf{Z}_{n,\Lambda'} = \mathbf{Z}_{n,\Lambda}\mathbf{Z}_{n,\Lambda;\Lambda'} \qquad P_{n,\Lambda'}\text{-almost surely},$$

and, since $h \in \mathscr{H}_\Lambda$, we have that

$$\mathrm{E}_{n,\Lambda}^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) = \mathrm{E}_{n,\Lambda'}^{1/n}[h(\mathbf{Z}_{n,\Lambda'})\mathbf{Z}_{n,\Lambda;\Lambda'}]^n = \mathrm{E}_{n,\Lambda'}^{1/n} h^n(\pi_{\Lambda'\Lambda}\mathbf{Z}_{n,\Lambda'}).$$

Applying Lemma 2.2 to the leftmost and rightmost sides, we obtain condition (C).

Now, condition (S), for a given $\hat{z}_\Lambda \in S_\Lambda$, can formally be obtained by substituting $\hat{h}(z_\Lambda) = 1(z_\Lambda = \|z_\Lambda\|\hat{z}_\Lambda)\|z_\Lambda\|$ into condition (C) and using the fact that $\mathbf{V}_\Lambda$ has support in S_Λ. However, the function $\hat{h}$ is not continuous, so we approximate it with a sequence of continuous functions $h_k \in \mathscr{H}_\Lambda$, $k \geqslant 1$, as follows. Let

$$h_k(z_\Lambda) = (\|z_\Lambda\| - k\|z_\Lambda - \hat{z}\|z_\Lambda\|\,\|)^+.$$

Since the h_k are from $\mathscr{H}_\Lambda$, they satisfy condition (C). Also $h_k(z_\Lambda) \downarrow \hat{h}(z_\Lambda)$ as $k \to \infty$. From the fact that the $h_k(z_\Lambda)$ are continuous and $\mathbf{V}_\Lambda$, and $\mathbf{V}_{\Lambda'}$ are deviabilities, it is not difficult to check by using Dini's theorem (for a proof see, e.g., Lemmas A.1 and A.4 in Puhalskii 1997) that one can take the limit as $k \to \infty$ in condition (C) for the h_k, as required. □

Remark 2.4. Property (S) implies that condition (C) holds for non-continuous positively homogeneous non-negative functions, too.

In analogy with statistical decision theory (Strasser 1985), we call a family of deviabilities $\{V_\Lambda, \Lambda \in \mathscr{A}(\Theta)\}$, where V_Λ is defined on $\mathbb{R}_+^\Lambda$, *conical* if it satisfies (C). If, in addition, $V_\Lambda(z_\Lambda) = 0$ for all $z_\Lambda \notin S_\Lambda$, the family is called *standard*. The proof of Lemma 2.3 shows that a family is standard if it meets condition (S).

The next result is of particular important for the minimax theorem below. It states that every standard family of deviabilities admits an extension to a function on $\mathbb{R}_+^\Theta$ which preserves the conical property.

Lemma 2.4. *For every standard family of deviabilities* $\{\mathbf{V}_\Lambda, \Lambda \in \mathscr{A}(\Theta)\}$, *there exists a function* $\mathbf{V}_\Theta$ *on* $\mathbb{R}_+^\Theta$ *such that the following conditions hold:*

(i) the function $\mathbf{V}_\Theta$ *is upper semi-continuous, assumes values in* [0, 1],

$$\sup_{z_\Theta \in \mathbb{R}_+^\Theta} \mathbf{V}_\Theta(z_\Theta) = 1 \quad \textit{and}\ \mathbf{V}_\Theta(z_\Theta) = 0 \ \textit{if}\ z_\Theta \notin S_\Theta;$$

(ii) for all $\Lambda \in \mathscr{A}(\Theta)$ *and* $h \in \mathscr{H}_\Lambda$,

$$\sup_{z_\Lambda \in \mathbb{R}_+^\Lambda} h(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta \in \mathbb{R}_+^\Theta} h(\pi_\Lambda z_\Theta)\mathbf{V}_\Theta(z_\Theta);$$

(iii) for all $z_\Lambda \in \mathbb{R}_+^\Lambda$,

$$\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta \in \Pi_\Lambda^{-1} z_\Lambda} \|\pi_\Lambda z_\Theta\|_\Lambda \mathbf{V}_\Theta(z_\Theta),$$

where $\Pi_\Lambda^{-1} z_\Lambda = \{z_\Theta \in \mathbb{R}_+^\Theta\colon \Pi_\Lambda z_\Theta = z_\Lambda\}$.

We relegate the proof to the Appendix.

We conclude this section by showing consistency of the above definitions of the LDP for statistical experiments.

Lemma 2.5. *Let $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ be a sequence of dominated statistical experiments. If it obeys the dominated LDP, then it obeys the LDP. More specifically, denoting by V_Θ the deviability on $\mathbb{R}_+^\Theta$ that is the LD limit of $\mathscr{L}(Z_{n,\Theta}|P_n)$ as $n \to \infty$, we have that $\mathscr{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}) \xrightarrow{l.d.} \mathbf{V}_\Lambda$, $\Lambda \in \mathscr{A}(\Theta)$, where*

$$\mathbf{V}_\Lambda(z_\Lambda) = \begin{cases} \sup_{z_\Theta \in \Pi_\Lambda^{-1} z_\Lambda} \|\pi_\Lambda z_\Theta\| V_\Theta(z_\Theta), & \text{if } z_\Lambda \in S_\Lambda, \\ 0, & \text{otherwise}. \end{cases}$$

Also, denoting by $\mathbf{V}_\Theta$ the extension of the standard family $\{\mathbf{V}_\Lambda, \Lambda \in \mathscr{A}(\Theta)\}$ defined in Lemma 2.4, we have that, for every $\Lambda \in \mathscr{A}(\Theta)$ and $h \in \mathscr{H}(\Lambda)$,

$$\sup_{z_\Theta \in \mathbb{R}_+^\Theta} h(\pi_\Lambda z_\Theta) V_\Theta(z_\Theta) = \sup_{z_\Theta \in \mathbb{R}_+^\Theta} h(\pi_\Lambda z_\Theta) \mathbf{V}_\Theta(z_\Theta).$$

Proof. We first prove that, for all $\Lambda \in \mathscr{A}(\Theta)$ and $h \in \mathscr{H}_\Lambda$,

$$\lim_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) = \sup_{z_\Theta \in \mathbb{R}_+^\Theta} h(\pi_\Lambda z_\Theta) V_\Theta(z_\Theta). \tag{2.18}$$

Since by (2.1) and (2.14),

$$Z_{n,\theta} = \mathbf{Z}_{n,\theta;\Lambda} \left(\frac{\mathrm{d}P_{n,\Lambda}}{\mathrm{d}P_n} \right)^{1/n} \qquad P_n\text{-almost surely, } \theta \in \Lambda,$$

and h is positively homogeneous, we have that

$$\mathrm{E}_{n,\Lambda}^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) = \mathrm{E}_n^{1/n} h^n(\mathbf{Z}_{n,\Lambda}) \frac{\mathrm{d}P_{n,\Lambda}}{\mathrm{d}P_n} = \mathrm{E}_n^{1/n} h^n(\pi_\Lambda Z_{n,\Theta}). \tag{2.19}$$

Now using the assumed LD convergence $\mathscr{L}(Z_{n,\Theta}|P_n) \xrightarrow{l.d.} V_\Theta$, we want to prove that

$$\lim_{n\to\infty} \mathrm{E}_n^{1/n} h^n(\pi_\Lambda Z_{n,\Theta}) = \sup_{z_\Theta \in \mathbb{R}_+^\Theta} h(\pi_\Lambda z_\Theta) V_\Theta(z_\Theta), \tag{2.20}$$

which by (2.19) would yield (2.18). The function h being non-negative and continuous but not bounded, (2.20) would follow if the uniform exponential integrability condition introduced in (2.6) holds:

$$\lim_{H\to\infty} \overline{\lim_{n\to\infty}} \mathrm{E}_n^{1/n} h^n(\pi_\Lambda Z_{n,\Theta}) 1(h(\pi_\Lambda Z_{n,\Theta}) > H) = 0. \tag{2.21}$$

It is here that we need condition (U). Let $h^* = \sup_{z_\Lambda \in S_\Lambda} h(z_\Lambda)$, which is finite by the continuity of h. Since $h \in \mathscr{H}_\Lambda$, it follows that $h(Z_{n,\Lambda}) \leqslant h^* \|Z_{n,\Lambda}\|$, so, in view of condition (U),

$$\overline{\lim_{n\to\infty}} \mathrm{E}_n^{1/n} h^n(\pi_\Lambda Z_{n,\Theta}) 1(h(\pi_\Lambda Z_{n,\Theta}) > H) \leqslant \overline{\lim_{n\to\infty}} \sum_{\theta\in\Lambda} \mathrm{E}_n^{1/n} h^{*n} Z_{n,\theta}^n 1(h^* Z_{n,\theta} > H)$$
$$\to 0 \quad \text{as } H \to \infty.$$

So, (2.20) and hence (2.18) have been proved.

Since by the definition of $\mathbf{V}_\Lambda$,

$$\sup_{z_\Lambda\in\mathbb{R}_+^\Lambda} h(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta\in\mathbb{R}_+^\Theta} h(\pi_\Lambda z_\Theta)V_\Theta(z_\Theta), \tag{2.22}$$

Lemma 2.2 implies that the proof of the first claim of the lemma is completed by checking that $\mathbf{V}_\Lambda$ is a deviability on $\mathbb{R}_+^\Lambda$.

Limit (2.21), in view of the LD convergence of $\mathscr{L}(Z_{n,\Theta}|P_n)$ to V_Θ, implies (using property (2.7)) that

$$\lim_{H\to\infty} \sup_{z_\Theta\in\mathbb{R}_+^\Theta} \|\pi_\Lambda z_\Theta\|_\Lambda 1(\|\pi_\Lambda z_\Theta\|_\Lambda > H)V_\Theta(z_\Theta) = 0.$$

Therefore, for every $\varepsilon > 0$ there exists H_ε such that

$$\{z_\Theta \in \mathbb{R}_+^\Theta\colon \|\pi_\Lambda z_\Theta\|_\Lambda V_\Theta(z_\Theta) \geqslant \varepsilon\} \subset \left\{z_\Theta \in \mathbb{R}_+^\Theta\colon V_\Theta(z_\Theta) \geqslant \frac{\varepsilon}{H_\varepsilon}\right\}$$

so that the set on the left is compact. Since also $\|\pi_\Lambda z_\Theta\|_\Lambda \geqslant \varepsilon$ when $\|\pi_\Lambda z_\Theta\|_\Lambda V_\Theta(z_\Theta) \geqslant \varepsilon$, and Π_Λ is continuous on $\{z_\Theta \in \mathbb{R}_+^\Theta\colon \|\pi_\Lambda z_\Theta\|_\Lambda \geqslant \varepsilon\}$, it follows that the set $\Pi_\Lambda\{z_\Theta \in \mathbb{R}_+^\Theta\colon \|\pi_\Lambda z_\Theta\|_\Lambda V_\Theta(z_\Theta) \geqslant \varepsilon\}$ is compact. Since, for $a > 0$,

$$\{z_\Lambda \in \mathbb{R}_+^\Lambda\colon \mathbf{V}_\Lambda(z_\Lambda) \geqslant a\} = \bigcap_{n=1}^{\infty} \Pi_\Lambda\left\{z_\Theta \in \mathbb{R}_+^\Theta\colon \|\pi_\Lambda z_\Theta\|_\Lambda V_\Theta(z_\Theta) \geqslant a\left(1 - \frac{1}{n+1}\right)\right\},$$

we conclude that the sets $\{z_\Lambda \in \mathbb{R}_+^\Lambda\colon \mathbf{V}_\Lambda(z_\Lambda) \geqslant a\}$ are compact. Thus, we are left to check that

$$\sup_{z_\Lambda\in\mathbb{R}_+^\Lambda} \mathbf{V}_\Lambda(z_\Lambda) = 1. \tag{2.23}$$

By (2.18) with $h(z_\Lambda) = \pi_\theta z_\Lambda$, $\theta \in \Lambda$,

$$\sup_{z_\Theta\in\mathbb{R}_+^\Theta} \pi_\theta z_\Theta V_\Theta(z_\Theta) = 1,$$

hence,

$$\sup_{z_\Theta\in\mathbb{R}_+^\Theta} \|\pi_\Lambda z_\Theta\|_\Lambda V_\Theta(z_\Theta) = \sup_{\theta\in\Lambda} \sup_{z_\Theta\in\mathbb{R}_+^\Theta} \pi_\theta z_\Theta V_\Theta(z_\Theta) = 1,$$

and (2.23) follows by the definition of $\mathbf{V}_\Lambda$.

The second claim of the lemma follows by (2.22) and Lemma 2.4. The lemma is proved. □

Remark 2.5. Equality (2.22) and Lemmas 2.2 and 2.3 imply that projections V_Λ, $\Lambda \in \mathscr{A}(\Theta)$, of V_Θ defined by

$$V_\Lambda(z_\Lambda) = \sup_{z_\Theta\in\pi_\Lambda^{-1}z_\Lambda} V_\Theta(z_\Theta)$$

constitute a family of deviabilities with properties (C) and (R).

3. A minimax theorem

We start this section by showing that, in analogy with the classical asymptotic theory of statistical experiments (Strasser 1985), the LDP for statistical experiments allows us to obtain an asymptotic lower bound for appropriately defined risks, which, in fact, has been the purpose of introducing the concept of the LDP for sequences of statistical experiments. We next prove that under additional conditions the bound is tight, and study the problem of constructing decisions attaining it.

We consider a sequence of statistical experiments $\{\mathscr{E}_n, n \geqslant 1\}$, where $\mathscr{E}_n = (\Omega, \mathscr{F}_n; P_{n,\theta}, \theta \in \Theta)$, and assume that it obeys the LDP. The associated deviabilities are denoted by $\mathbf{V}_\Lambda$, $\Lambda \in \mathscr{A}(\Theta)$, and $\mathbf{V}_\Theta$ denotes the extension defined in Lemma 2.4.

We introduce some more notation common in statistical decision theory (see, e.g., Strasser 1985). We denote by $\mathscr{D}$ a Hausdorff topological space with the Borel σ-field which we take as a decision space; $W_\theta = (W_\theta(r), r \in \mathscr{D})$, $\theta \in \Theta$, are, for each θ, non-negative and lower semi-continuous functions on $\mathscr{D}$ which play the role of loss functions; $\mathscr{R}_n$ denotes the set of all measurable mappings ρ_n: $\Omega_n \to \mathscr{D}$, i.e., $\mathscr{R}_n$ is the set of all decision functions with values in $\mathscr{D}$. We define the LD risk of a decision $\rho_n \in \mathscr{R}_n$ in the experiment $\mathscr{E}_n$ by

$$R_n(\rho_n) = \sup_{\theta\in\Theta} \mathrm{E}_{n,\theta}^{1/n} W_\theta^n(\rho_n). \tag{3.1}$$

Obviously, this is an analogue of the risk in minimax decision theory (cf. Strasser 1985).

Recall that a function f: $U \to \mathbb{R}$ on a topological space U is level-compact if it is bounded from below and the sets $\{u \in U$: $f(u) \leqslant \alpha\}$ are compact for all $\alpha < \sup_{u\in U} f(u)$ (Strasser 1985, Definition 6.3). Obviously, if U is Hausdorff, a level-compact function is lower semi-continuous and the supremum of a family of level-compact functions is level-compact. For what follows, it is also worth mentioning that level-compact functions attain infima on closed sets.

Theorem 3.1. *Let the sequence $\{\mathscr{E}_n, n \geqslant 1\}$ obey the LDP. Assume that the functions W_θ, $\theta \in \Theta$, are level-compact. Then*

$$\varliminf_{n\to\infty} \inf_{\rho_n\in\mathscr{R}_n} R_n(\rho_n) \geqslant R^*,$$

where

$$R^* = \sup_{z_\Theta\in\mathbb{R}_+^\Theta} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta} W_\theta(r) z_\theta \mathbf{V}_\Theta(z_\Theta).$$

In particular, if $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ obeys the dominated LDP and V_Θ is the associated deviability then the lower bound can be rewritten as

$$R^* = \sup_{z_\Theta\in\mathbb{R}_+^\Theta} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta} W_\theta(r) z_\theta V_\Theta(z_\Theta). \tag{3.2}$$

If, moreover, conditions (Y) and (U) hold then

$$R^* = \sup_{y\in\mathscr{Y}} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta} W_\theta(r) \mathfrak{z}_\theta(y) V(y).$$

Proof. Let $\Lambda \in \mathscr{A}(\Theta)$. We first prove that

$$\varliminf_{n\to\infty} \inf_{\rho_n} \sup_{\theta\in\Lambda} \mathrm{E}_{n,\theta}^{1/n} W_\theta^n(\rho_n) \geqslant \sup_{z_\Lambda\in\mathbb{R}_+^\Lambda} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta \mathbf{V}_\Lambda(z_\Lambda). \tag{3.3}$$

Let $\{\rho_n,\ n \geqslant 1\}$ be an arbitrary sequence of decisions. We have, by the definition of $\mathbf{Z}_{n,\Lambda}$ (see (2.14)),

$$\begin{aligned}\varliminf_{n\to\infty} \sup_{\theta\in\Lambda} \mathrm{E}_{n,\theta}^{1/n} W_\theta^n(\rho_n) &= \varliminf_{n\to\infty} \sup_{\theta\in\Lambda} \mathrm{E}_{n,\Lambda}^{1/n} W_\theta^n(\rho_n)\mathbf{Z}_{n,\theta;\Lambda}^n \\ &\geqslant \varliminf_{n\to\infty} \left[\frac{1}{|\Lambda|}\mathrm{E}_{n,\Lambda}\sum_{\theta\in\Lambda} W_\theta^n(\rho_n)\mathbf{Z}_{n,\theta;\Lambda}^n\right]^{1/n} \\ &\geqslant \varliminf_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} \sup_{\theta\in\Lambda} W_\theta^n(\rho_n)\mathbf{Z}_{n,\theta;\Lambda}^n \\ &\geqslant \varliminf_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} w^n(\mathbf{Z}_{n,\Lambda}),\end{aligned}$$

where

$$w(z_\Lambda) = \inf_{r\in\mathscr{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta, \qquad z_\Lambda = (z_\theta,\ \theta\in\Lambda) \in \mathbb{R}_+^\Lambda.$$

Since the set Λ is finite and the functions W_θ are level-compact, it is not difficult to see that the function $w(\cdot)$ is lower semi-continuous (cf. Aubin 1984, Proposition 1.7). So by the LD convergence of $\mathscr{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda})$ to $\mathbf{V}_\Lambda$,

$$\varliminf_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} w^n(\mathbf{Z}_{n,\Lambda}) \geqslant \sup_{z_\Lambda\in\mathbb{R}_+^\Lambda} w(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda),$$

implying (3.3).

Since the function $w(\cdot)$ belongs to $\mathscr{H}_\Lambda$, an application of Lemma 2.4(ii) yields

$$\sup_{z_\Lambda\in\mathbb{R}_+^\Lambda} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta \mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta\in\mathbb{R}_+^\Theta} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta \mathbf{V}_\Theta(z_\Theta),$$

so by (3.3)

$$\varliminf_{n\to\infty} \inf_{\rho_n} \sup_{\theta\in\Lambda} \mathrm{E}_{n,\theta}^{1/n} W_\theta^n(\rho_n) \geqslant \sup_{z_\Theta\in\mathbb{R}_+^\Theta} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta \mathbf{V}_\Theta(z_\Theta).$$

Now the proof of the lower bound is completed by observing that, for every $z_\Theta = (z_\theta,\ \theta\in\Theta) \in \mathbb{R}_+^\Theta$,

$$\sup_{\Lambda\in\mathscr{A}(\Theta)} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Lambda} W_\theta(r) z_\theta = \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta} W_\theta(r) z_\theta \tag{3.4}$$

(for a proof see Lemma A.3 in the Appendix; or Aubin and Ekeland 1984, Theorem 6, Section 2, Chapter 6).

If $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ obeys the dominated LDP, then by Lemma 2.5

$$\sup_{z_\Theta \in \mathbb{R}_+^\Theta} \inf_{r \in \mathscr{D}} \sup_{\theta \in \Lambda} W_\theta(r) z_\theta V_\Theta(z_\Theta) = \sup_{z_\Theta \in \mathbb{R}_+^\Theta} \inf_{r \in \mathscr{D}} \sup_{\theta \in \Lambda} W_\theta(r) z_\theta \mathbf{V}_\Theta(z_\Theta),$$

and representation (3.2) follows by (3.4). The last representation for R^* in the statement of the theorem follows since, by Lemma 2.1, $V_\Theta = V \circ \mathfrak{z}_\Theta^{-1}$. □

Remark 3.1. Note that the proof only uses what is known as a lower bound in the LDP.

Remark 3.2. We are now in a position to explain why we consider condition (U) in the definition of the dominated LDP to be important. Assume that $\{\mathscr{E}_n, n \geqslant 1\}$ is a dominated family with dominating measures P_n such that, for a deviability V_Θ on $\mathbb{R}_+^\Theta$, we have the LD convergence $\mathscr{L}(Z_{n,\Theta}|P_n) \xrightarrow{l.d.} V_\Theta$. The proof of Theorem 3.1 with $\mathbf{V}_\Theta$ replaced by V_Θ and $\mathbf{V}_\Lambda$ replaced by $V_\Theta \circ \pi_\Lambda^{-1}$ (which would not use condition (U)) would still give the right-hand side of (3.2) as a lower bound. However, these lower bounds can generally be different for different sequences of dominating measures. The role of condition (U) is to eliminate this possibility by making sure that equality (3.2) holds so that the lower bounds do not depend on a choice of dominating measures.

In applications, as we will see, the assumption that the loss functions are level-compact is normally met. However, in the appendix we give a variant of Theorem 3.1 for more general loss functions. As in the classical theory, tackling this case requires considering generalized decisions (cf. Strasser 1985).

We now turn to the question of tightness of the above lower bound and start with defining the concept of LD efficiency. We say that a sequence of decisions $\{\rho_n^*, n \geqslant 1\}$ is LD efficient if, for any other sequence of decisions $\{\rho_n\}$,

$$\overline{\lim_{n\to\infty}} (R_n(\rho_n^*) - R_n(\rho_n)) \leqslant 0.$$

Theorem 3.1 implies that to construct LD efficient decisions one can apply an approach similar to that used in the classical asymptotic decision theory. Indeed, by Theorem 3.1, if the W_θ, $\theta \in \Theta$, are level-compact, then, for any sequence of decisions $\{\rho_n, n \geqslant 1\}$,

$$\underline{\lim}_{n\to\infty} R_n(\rho_n) \geqslant R^*.$$

Now if a sequence $\{\rho_n^*, n \geqslant 1\}$ is such that $R_n(\rho_n^*) \to R^*$ as $n \to \infty$, it is obviously LD efficient.

Furthermore, motivated by applications, we assume that the sequence $\{\mathscr{E}_n, n \geqslant 1\}$ is dominated and conditions (Y) and (U) hold. Then, by Theorem 3.1, the asymptotic minimax risk can be written as

$$R^* = \sup_{y \in \mathscr{Y}} \inf_{r \in \mathscr{D}} \sup_{\theta \in \Theta} W_\theta(r) \mathfrak{z}_\theta(y) V(y). \tag{3.5}$$

Representation (3.5) prompts considering for each $y \in \mathscr{Y}$ the subproblem

(Q) $Q^*(y) = \inf_{r \in \mathscr{D}} \sup_{\theta \in \Theta} W_\theta(r) \mathfrak{z}_\theta(y).$

Since the functions W_θ are level-compact for each $\theta \in \Theta$, it follows that, given $y \in \mathscr{Y}$, we can find $r^*(y) \in \mathscr{D}$ that delivers the infimum in (Q). The value $r^*(y)$ can be viewed as 'the best decision if the value of Y_n is y'. Hence, provided the function $r^*(y)$: $\mathscr{Y} \to \mathscr{D}$ is Borel measurable, the decisions $r^*(Y_n)$ are natural candidates for the LD efficient decisions. Unfortunately, we cannot prove this without requiring that $Q^*(y)$ be continuous (or upper semi-continuous) which usually is not fulfilled in applications. The reason for the latter, as in condition (Y) above, is that the $\mathfrak{z}_\theta(y)$ typically are not continuous as maps from $\mathscr{Y}$ into $\mathbb{R}_+$. Therefore, as in condition (Y), we invoke the idea of regularization. We require that there exist functions $\mathfrak{z}_{\theta,\delta}(y)$ such that functions $Q_\delta(y)$ defined by

$$(Q_\delta)\ \ Q_\delta(y) = \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta} W_\theta(r)\mathfrak{z}_{\theta,\delta}(y), \qquad y \in \mathscr{Y},$$

are continuous in y, on the one hand, and approximate $Q^*(y)$ for small δ, on the other hand. A rigorous formulation is given by condition (sup Y), which strengthens condition (Y) to the effect that the requirements of (Y) hold uniformly in $\theta \in \Theta$. This way of handling the technical difficulties does not allow us, however, to get LD efficient decisions: as the next theorem shows, in general we are only able to obtain decisions whose asymptotic risk is arbitrarily close to the lower bound. Still, we succeed in proving that the lower bound of Theorem 3.1 is tight and LD efficient decisions exist. We next state the condition. Recall that $Z_{n,\theta} = (\mathrm{d}P_{n,\theta}/\mathrm{d}P_n)^{1/n}$.

(sup Y) *There exist statistics* Y_n: $\Omega_n \to \mathscr{Y}$ *with values in a metric space* $\mathscr{Y}$ *with the Borel* σ*-field, functions* $\mathfrak{z}_\theta$: $\mathscr{Y} \to \mathbb{R}_+$, $\theta \in \Theta$, *and* $\mathfrak{z}_{\theta,\delta}$: $\mathscr{Y} \to \mathbb{R}_+$, $\theta \in \Theta$, $\delta > 0$, *such that:*

- (Y.1) *the sequence* $\{\mathscr{L}(Y_n|P_n),\ n \geqslant 1\}$ *LD converges to a deviability* $V(y)$, $y \in \mathscr{Y}$;
- (sup Y.2) *for the uniform topology on* $\mathbb{R}_+^\Theta$, *the functions* $\mathfrak{z}_{\Theta,\delta} = (\mathfrak{z}_{\theta,\delta},\ \theta \in \Theta)$: $\mathscr{Y} \to \mathbb{R}_+^\Theta$, $\delta > 0$, *are Borel measurable and continuous* V *a.e.;*
- (sup Y.3) $\lim_{\delta\to 0} \overline{\lim}_{n\to\infty} \sup_{\theta\in\Theta} P_n^{1/n}(|Z_{n,\theta} - \mathfrak{z}_{\theta,\delta}(Y_n)| > \varepsilon) = 0$ *for all* $\varepsilon > 0$;
- (sup Y.4) $\lim_{\delta\to 0} \sup_{\theta\in\Theta} \sup_{y\in\Phi_V(a)} |\mathfrak{z}_{\theta,\delta}(y) - \mathfrak{z}_\theta(y)| = 0$ *for all* $a > 0$.

In the next theorem, condition (sup Y) is used together with condition (sup U) which strengthens (U):

$$(\sup U)\ \ \lim_{H\to\infty} \overline{\lim}_{n\to\infty} \sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} Z_{n,\theta}^n 1(Z_{n,\theta} > H) = 0$$

Theorem 3.2. *Let a sequence of dominated experiments* $\{\mathscr{E}_n,\ P_n,\ n \geqslant 1\}$ *satisfy conditions* (sup Y) *and* (sup U). *Let the function* $W_\theta(r)$ *be bounded in* (θ, r) *and level-compact in* r *for each* $\theta \in \Theta$. *Assume that there exist Borel functions* $r_\delta(y)$: $\mathscr{Y} \to \mathscr{D}$ *such that the infimum in* (Q_δ) *is attained at* $r_\delta(y)$, *and denote* $\rho_{n,\delta} = r_\delta(Y_n)$. *Then*

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}} R_n(\rho_{n,\delta}) = \lim_{\delta\to 0} \underline{\lim}_{n\to\infty} R_n(\rho_{n,\delta}) = R^*$$

so that

$$\lim_{n\to\infty} \inf_{\rho_n\in\mathscr{R}_n} R_n(\rho_n) = R^*.$$

In particular, for some sequence ρ_n^*,

$$\lim_{n\to\infty} R_n(\rho_n^*) = R^*.$$

Proof. Since (sup Y) implies (Y), by Lemma 2.1, $\mathscr{L}(Z_{n,\Theta}|P_n) \xrightarrow{l.d.} V_\Theta = V \circ \mathfrak{z}_\Theta^{-1}$, so by Theorem 3.1, for each δ,

$$\varliminf_{n\to\infty} R_n(\rho_{n,\delta}) \geqslant R^*.$$

The proof of the first set of equalities would be finished if

$$\varlimsup_{\delta\to 0} \varlimsup_{n\to\infty} R_n(\rho_{n,\delta}) \leqslant R^*. \tag{3.6}$$

Let C be an upper bound for W: $W_\theta(r) \leqslant C$. Since

$$R_n(\rho_{n,\delta}) = \sup_{\theta\in\Theta} \mathrm{E}_{n,\theta}^{1/n} W_\theta^n(\rho_{n,\delta}) = \sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} W_\theta^n(\rho_{n,\delta}) Z_{n,\theta}^n,$$

we have that, for any $H > 0$,

$$R_n(\rho_{n,\delta}) \leqslant \sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} W_\theta^n(\rho_{n,\delta})(Z_{n,\theta} \wedge H)^n + C \sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} Z_{n,\theta}^n 1(Z_{n,\theta} > H).$$

The second term on the right tends to 0 as $n \to \infty$ and $H \to \infty$ by condition (sup U), so the required limit would follow by

$$\varlimsup_{\delta\to 0} \varlimsup_{n\to\infty} \sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} W_\theta^n(\rho_{n,\delta})(Z_{n,\theta} \wedge H)^n \leqslant R^*. \tag{3.7}$$

Since

$$\Big|\sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} W_\theta^n(\rho_{n,\delta})(Z_{n,\theta} \wedge H)^n - \sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} W_\theta^n(\rho_{n,\delta})(\mathfrak{z}_{n,\delta}(Y_n) \wedge H)^n\Big| \leqslant C \sup_{\theta\in\Theta} \mathrm{E}_n^{1/n}(|Z_{n,\theta} - \mathfrak{z}_{\theta,\delta}(Y_n)| \wedge H)^n,$$

condition (sup Y.3) implies that

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} \Big|\sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} W_\theta^n(\rho_{n,\delta})(Z_{n,\theta} \wedge H)^n - \sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} W_\theta^n(\rho_{n,\delta})(\mathfrak{z}_{\theta,\delta}(Y_n) \wedge H)^n\Big| = 0. \tag{3.8}$$

Next, using the definitions of Q_δ and $\rho_{n,\delta}$ and the inequality $W_\theta(r) \leqslant C$, we obtain

$$\begin{aligned}\sup_{\theta\in\Theta} \mathrm{E}_n^{1/n} W_\theta^n(\rho_{n,\delta})(\mathfrak{z}_{\theta,\delta}(Y_n) \wedge H)^n &\leqslant \mathrm{E}_n^{1/n}(\sup_{\theta\in\Theta} (W_\theta^n(r_\delta(Y_n))\mathfrak{z}_{\theta,\delta}(Y_n)) \wedge CH)^n \\ &= \mathrm{E}_n^{1/n}(Q_\delta(Y_n) \wedge CH)^n.\end{aligned} \tag{3.9}$$

The last two expectations in (3.9) are well defined since the assumptions of the theorem imply that $Q_\delta(y) = \sup_{\theta\in\Theta} W_\theta(r_\delta(y))\mathfrak{z}_{\theta,\delta}(y)$ is a Borel function.

By the boundedness of $W_\theta(r)$ and (sup Y.2), the function $Q_\delta(y)$ is V-a.e. continuous. Since $\mathscr{L}(Y_n|P_n) \xrightarrow{l.d.} V$, we obtain

$$\lim_{n\to\infty} \mathrm{E}_n^{1/n}(Q_\delta(Y_n) \wedge CH)^n = \sup_{y\in\mathscr{Y}} (Q_\delta(y) \wedge CH)V(y). \tag{3.10}$$

By (Q), (Q_δ) and the inequality $W_\theta(r) \leqslant C$, we have that

$$|\sup_{y\in\mathscr{Y}}(Q_\delta(y)\wedge CH)V(y)-\sup_{y\in\mathscr{Y}}(Q^*(y)\wedge CH)V(y)|\leqslant C\sup_{y\in\mathscr{Y}}\sup_{\theta\in\Theta}(|\mathfrak{z}_{\theta,\delta}(y)-\mathfrak{z}_\theta(y)|\wedge H)V(y),$$

and (sup Y.4) easily implies that the right-hand side tends to 0 as $\delta\to 0$. Thus,

$$\begin{aligned}\lim_{\delta\to 0}\sup_{y\in\mathscr{Y}}(Q_\delta(y)\wedge CH)V(y) &= \sup_{y\in\mathscr{Y}}(Q^*(y)\wedge CH)V(y)\\ &\leqslant \sup_{y\in\mathscr{Y}}Q^*(y)V(y)=R^*,\end{aligned}\tag{3.11}$$

where the last equality follows by (3.5) and (Q). Putting together (3.8)–(3.11) proves (3.7) and hence (3.6).

The second claim of the theorem follows by (3.6) and the string of inequalities the first of which is Theorem 3.1:

$$R^*\leqslant \underline{\lim}_{n\to\infty}\inf_{\rho_n}R_n(\rho_n)\leqslant \overline{\lim}_{n\to\infty}\inf_{\rho_n}R_n(\rho_n)\leqslant \overline{\lim}_{n\to\infty}R_n(\rho_{n,\delta}).$$

□

Remark 3.3. Obviously, $r_\delta(y)$ chosen so that

$$\sup_{\theta\in\Theta}W_\theta(r_\delta(y))\mathfrak{z}_{\theta,\delta}(y)\geqslant Q_\delta(y)-\epsilon_\delta,$$

where $\epsilon_\delta\to 0$ as $\delta\to 0$, would work too.

Remark 3.4. If condition (sup Y) holds with $\mathfrak{z}_{\theta,\delta}(y)=\mathfrak{z}_\theta(y)$, then the $r_\delta(y)$ in the theorem do not depend on δ and the decisions $\rho_n^*:=\rho_{n,\delta}$ are LD efficient.

Remark 3.5. As with condition (Y), in applications it is more convenient to deal with a logarithmic form of condition (sup Y). Specifically, defining $\Xi_{n,\Theta}$ and $\Phi_I'(a)$ as in Remark 2.2, let us introduce condition (sup Y'):

(sup Y') *there exist statistics Y_n: $\Omega_n\to\mathscr{Y}$ with values in a metric space $\mathscr{Y}$ with the Borel σ-field, functions ζ_θ: $\mathscr{Y}\to\mathbb{R}$, $\theta\in\Theta$, and $\zeta_{\theta,\delta}$: $\mathscr{Y}\to\mathbb{R}$, $\theta\in\Theta$, $\delta>0$, such that*

- (Y'.1) *the sequence $\{\mathscr{L}(Y_n|P_n),\ n\geqslant 1\}$ obeys the LDP with rate function $I(y)$, $y\in\mathscr{Y}$;*
- (sup Y'.2) *for the uniform topology on $\mathbb{R}^\Theta$, the functions $\zeta_{\Theta,\delta}=(\zeta_{\theta,\delta},\ \theta\in\Theta)$: $\mathscr{Y}\to\mathbb{R}^\Theta$, $\delta>0$, are Borel measurable and continuous at each point y such that $I(y)<\infty$;*
- (sup Y'.3) $\lim_{\delta\to 0}\overline{\lim}_{n\to\infty}\sup_{\theta\in\Theta}P_n^{1/n}(|\Xi_{n,\theta}-\zeta_{\theta,\delta}(Y_n)|>\varepsilon)=0$ *for all* $\varepsilon>0$*;*
- (sup Y'.4) $\lim_{\delta\to 0}\sup_{\theta\in\Theta}\sup_{y\in\Phi_I'(a)}|\zeta_{\theta,\delta}(y)-\zeta_\theta(y)|=0$ *for all* $a\geqslant 0$.

Then condition (sup Y) is implied by condition (sup Y'). Similarly, condition (sup U) follows from the condition

$$(\sup U')\quad \lim_{H\to\infty}\overline{\lim}_{n\to\infty}\sup_{\theta\in\Theta}\mathrm{E}_n^{1/n}\exp(n\Xi_{n,\theta})1(\Xi_{n,\theta}>H)=0.$$

We henceforth refer to the decisions $\rho_{n,\delta}$ as nearly LD efficient.

4. Asymptotic LD risks and efficient decisions for hypothesis testing and estimation problems

This section specifies the asymptotic minimax bound of Theorem 3.1 and (nearly) LD efficient decisions for some typical statistical set-ups by considering hypothesis testing and estimation with Bahadur-type criteria. We consider indicator loss functions, i.e.,

$$W_\theta(r) = 1(r \notin A_\theta), \qquad r \in \mathscr{D},\ \theta \in \Theta,$$

where A_θ are closed subsets of $\mathscr{D}$. Then the LD risk of a decision ρ_n in the nth experiment is

$$R_n(\rho_n) = \sup_{\theta\in\Theta} P_{n,\theta}^{1/n}(\rho_n \notin A_\theta).$$

For applications, it is convenient to introduce the logarithmic risk

$$R'_n(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(\rho_n \notin A_\theta). \tag{4.1}$$

Accordingly, we consider the logarithm of the lower bound R^*:

$$R'^* = \sup_{\zeta_\Theta\in\mathbb{R}^\Theta} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta: A_\theta\not\ni r} (\zeta_\theta - \mathbf{I}_\Theta(\zeta_\Theta)),$$

where $\mathbf{I}_\Theta(\zeta_\Theta) = -\log \mathbf{V}_\Theta(z_\Theta)$ for $z_\Theta = (\exp(\zeta_\theta),\ \theta \in \Theta)$, $\zeta_\Theta = (\zeta_\theta,\ \theta \in \Theta)$. Theorem 3.1 then yields the following result.

Theorem 4.1. *Assume that the A_θ, $\theta \in \Theta$, are compact. If the sequence $\{\mathscr{E}_n,\ n \geqslant 1\}$ obeys the LDP then*

$$\varliminf_{n\to\infty} \inf_{\rho_n\in\mathscr{R}_n} R'_n(\rho_n) \geqslant R'^*.$$

Let us assume now that the sequence $\{\mathscr{E}_n,\ n \geqslant 1\}$ is dominated and conditions (Y') and (U') hold. According to Remark 2.2 and Theorem 3.1, we then have that

$$R'^* = \sup_{y\in\mathscr{Y}} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta: A_\Theta\not\ni r} (\zeta_\theta(y) - I(y)). \tag{4.2}$$

Similarly, subproblems (Q) and (Q_δ) of Section 3 take the form

(Q') $\displaystyle Q'^*(y) = \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta: A_\theta\not\ni r} \zeta_\theta(y), \qquad y \in \mathscr{Y},$

and

(Q'_δ) $\displaystyle Q'_\delta(y) = \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta: A_\theta\not\ni r} \zeta_{\theta,\delta}(y), \qquad y \in \mathscr{Y}.$

Obviously,

$$R'^* = \sup_{y\in\mathscr{Y}} (Q'^*(y) - I(y)).$$

Let the infimum in (Q'_δ) be attained at some point $r'_\delta(y)$ which is the case, e.g., if the A_θ, $\theta \in \Theta$, are compact. We denote $\rho'_{n,\delta} = r'_\delta(Y_n)$.

Combining Theorem 4.1 and Theorem 3.2, and taking into account Remarks 2.2 and 3.5, we obtain the following theorem.

Theorem 4.2. *Assume that $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ is a dominated sequence of statistical experiments and the A_θ, $\theta \in \Theta$, are compact.*

1. *If conditions (Y') and (U') hold then*

$$\varliminf_{n\to\infty} \inf_{\rho_n \in \mathscr{R}_n} R'_n(\rho_n) \geqslant R'^*.$$

2. *Let the functions r'_δ, $\delta > 0$, which map $\mathscr{Y}$ into $\mathscr{D}$, be Borel measurable. If conditions* $(\sup Y')$ *and* $(\sup U')$ *hold then*

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R'_n(\rho'_{n,\delta}) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R'_n(\rho'_{n,\delta}) = R'^*.$$

so that

$$\lim_{n\to\infty} \inf_{\rho_n \in \mathscr{R}_n} R'_n(\rho_n) = R'^*.$$

4.1. Hypothesis testing

Let Θ_0 and Θ_1 be non-intersecting subsets of the parameter set Θ: $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$, $\Theta_0 \cap \Theta_1 = \varnothing$. We want to test the hypothesis H_0: $\theta \in \Theta_0$ versus the alternative H_1: $\theta \in \Theta_1$.

The decision space $\mathscr{D}$ consists of two points: $\mathscr{D} = \{0, 1\}$. We endow it with the discrete topology and, for any decision (test) ρ, we treat the event $\{\rho = 0\}$ (or $\{\rho = 1\}$) as accepting (or rejecting) the null hypothesis.

An associated loss function $W_\theta(r)$ is the indicator of the wrong choice:

$$W_\theta(r) = 1(\theta \notin \Theta_r), \qquad r = 0, 1, \tag{4.3}$$

and the logarithmic risk $R'(\rho_n)$ of a decision ρ_n in (4.1) takes the form

$$R^T_n(\rho_n) = \max\left\{ \sup_{\theta\in\Theta_0} \frac{1}{n} \log P_{n,\theta}(\rho_n = 1),\ \sup_{\theta\in\Theta_1} \frac{1}{n} \log P_{n,\theta}(\rho_n = 0) \right\}. \tag{4.4}$$

Denoting the corresponding asymptotic minimax risk R'^* by T^*, we have by (4.2) that

$$T^* = \sup_{y\in\mathscr{Y}} \min\big(\sup_{\theta\in\Theta_0} (\zeta_\theta(y) - I(y)),\ \sup_{\theta\in\Theta_1} (\zeta_\theta(y) - I(y))\big\}. \tag{4.5}$$

For what follows, it is more convenient to use another representation for T^*,

$$T^* = \sup_{\theta\in\Theta_0, \theta'\in\Theta_1} S(\theta, \theta'), \tag{4.6}$$

where

$$S(\theta, \theta') = \sup_{y\in\mathscr{Y}} \min\{\zeta_\theta(y) - I(y),\ \zeta_{\theta'}(y) - I(y)\}. \tag{4.7}$$

Next, subproblem (Q'_δ) for this case is

$$T'_\delta(y) = \min_{r=0,1} \sup_{\theta\in\Theta_{1-r}} \zeta_{\theta,\delta}(y), \qquad y \in \mathscr{Y}.$$

It has the solution

$$r^T_\delta(y) = 1(\sup_{\theta\in\Theta_0} \zeta_{\theta,\delta}(y) < \sup_{\theta\in\Theta_1} \zeta_{\theta,\delta}(y)),$$

which leads us to tests of the form

$$\rho^T_{n,\delta} = 1(\sup_{\theta\in\Theta_0} \zeta_{\theta,\delta}(Y_n) < \sup_{\theta\in\Theta_1} \zeta_{\theta,\delta}(Y_n)). \tag{4.8}$$

In the case of two simple hypotheses θ_0 and θ_1, the tests reduce to a regularization of the Neyman–Pearson test:

$$\rho^T_{n,\delta} = 1(\zeta_{\theta_0,\delta}(Y_n) < \zeta_{\theta_1,\delta}(Y_n)).$$

Applying Theorem 4.2, we obtain the following theorem.

Theorem 4.3. *Let Θ_0 and Θ_1 be non-intersecting subsets of Θ. If a sequence of dominated experiments $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ satisfies conditions (Y') and $U')$ then*

$$\varliminf_{n\to\infty} \inf_{\rho_n\in\mathscr{R}_n} R^T_n(\rho_n) \geqslant T^*.$$

If conditions $(\sup Y')$ *and* $(\sup U')$ *hold then*

$$\lim_{n\to\infty} \inf_{\rho_n\in\mathscr{R}_n} R^T_n(\rho_n) = T^*,$$

and the tests $\rho^T_{n,\delta}$ are nearly LD efficient:

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R^T_n(\rho^T_{n,\delta}) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R^T_n(\rho^T_{n,\delta}) = T^*.$$

4.2. Parameter estimation

Let Θ be a subset of a normed space $\mathscr{B}$ with norm $\|\cdot\|$. We are interested in estimating a parameter θ under the Bahadur-type loss function

$$W_\theta(r) = 1(\|r-\theta\| > c) \tag{4.9}$$

for a given positive c. The logarithmic risk of an estimator ρ_n is

$$R^E_n(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(\|\rho_n - \theta\| > c). \tag{4.10}$$

We assume that the decision space $\mathscr{D}$ is either a compact subset of $\mathscr{B}$ with the induced topology or a closed convex subset of $\mathscr{B}$ with the weak topology; in the latter case, $\mathscr{B}$ is assumed to be a reflexive Banach space. For both cases, the functions W_θ, $\theta \in \Theta$, are level-compact on $\mathscr{D}$.

In this set-up, we denote the asymptotic minimax risk R'^* from (4.2) by E^*:

$$E^* = \sup_{y\in\mathscr{Y}} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta:\|r-\theta\|>c} (\zeta_\theta(y) - I(y)), \tag{4.11}$$

and the corresponding subproblem (Q'_δ) is

(E_δ) $E_\delta(y) = \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta:\|r-\theta\|>c} \zeta_{\theta,\delta}(y), \qquad y\in\mathscr{Y}.$

We next describe solutions to (E_δ). Consider a real-valued function $f(\theta)$, $\theta\in\Theta$, and let

$$A(h) = \{\theta\in\Theta\colon f(\theta) > h\}, \qquad h\in\mathbb{R}, \tag{4.12}$$

$$r(h) = \inf_{r\in\mathscr{D}} \sup_{\theta\in A(h)} \|r-\theta\|, \qquad h\in\mathbb{R}, \tag{4.13}$$

$$h_c = \inf\,(h\colon r(h)\leqslant c).$$

We assume that $h_c < \infty$ (e.g., $f(\theta)$ is bounded). Note that, for both definitions of $\mathscr{D}$, the infimum in (4.13) is attained since the functions $r \to \|r-\theta\|$ from $\mathscr{D}$ to $\mathbb{R}_+$ are level-compact for all $\theta\in\Theta$.

Lemma 4.1. *The set* $D_c = \{r\in\mathscr{D}\colon \sup_{\theta\in A(h_c)}\|r-\theta\|\leqslant c\}$ *is non-empty and consists of all* $r_c\in\mathscr{D}$ *at which* $\inf_{r\in\mathscr{D}}\sup_{\theta\in\Theta:\|r-\theta\|>c} f(\theta)$ *is attained. Also the latter infimum equals* h_c.

Proof. Since the function $(r, h)\to\sup_{\theta\in A(h)}\|r-\theta\|$ is decreasing in h and level-compact in $r\in\mathscr{D}$, the function $r(h)$ is decreasing and right-continuous. Hence, $r(h_c)\leqslant c$ and, since $\inf_{r\in\mathscr{D}}\sup_{\theta\in A(h_c)}\|r-\theta\| = r(h_c)$ and the infimum is attained, the set D_c is non-empty.

Now let $r_c\in D_c$. By definition, $\|r_c-\theta\|\leqslant c$ for all $\theta\in\Theta$ such that $f(\theta)>h_c$. Hence,

$$\sup_{\theta\in\Theta:\|r_c-\theta\|>c} f(\theta)\leqslant h_c. \tag{4.14}$$

On the other hand, if $h<h_c$, then $r(h)>c$, which implies that, for every $r\in\mathscr{D}$, $\sup_{\theta\in A(h)}\|r-\theta\|>c$ or, equivalently, there exists θ such that $f(\theta)>h$ and $\|r-\theta\|>c$ so that $\inf_{r\in\mathscr{D}}\sup_{\theta\in\Theta:\|r-\theta\|>c} f(\theta)\geqslant h$. Since h is arbitrarily close to h_c, we conclude that

$$\inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta:\|r-\theta\|>c} f(\theta)\geqslant h_c,$$

which by (4.14) proves that $\inf_{r\in\mathscr{D}}\sup_{\theta\in\Theta:\|r-\theta\|>c} f(\theta) = h_c$ and r_c delivers the infimum.

Finally, if $r\notin D_c$ then $\sup_{\theta\in A(h_c)}\|r-\theta\|>c$, i.e., there exists θ such that $\|r-\theta\|>c$ and $f(\theta)>h_c$, which yields the inequality $\sup_{\theta\in\Theta:\|r-\theta\|>c} f(\theta)>h_c$. □

Remark 4.1. Informally, $r(h)$ is the smallest radius of the balls that contain all the θ with $f(\theta)>h$, and h_c is the lowest level h for which there exists a ball of radius c with this property. The lemma states, in particular, that h_c is the infimum over all the balls of radius c of the largest values of $f(\theta)$ outside the balls. For a one-dimensional parameter θ, the construction in the lemma chooses the largest level set of the function f contained in an interval of length $2c$, and the r_c are the centres of the intervals.

Let $r_c(f)$ denote an element of the set D_c in the lemma and, taking $f(\theta) = \zeta_{\theta,\delta}(y)$, let $r^E_{\delta,c}(y) = r_c(\zeta_{\Theta,\delta}(y))$. We assume that the functions $r^E_{\delta,c}(y)$: $\mathscr{Y} \to \mathscr{D}$ are Borel measurable. We can then define the estimators

$$\rho^E_{n,\delta} = r^E_{\delta,c}(Y_n). \tag{4.15}$$

Motivated by Remark 4.1, we call these estimators *interval-median*.

A version of Theorem 4.2 for this case is the following.

Theorem 4.4. *Assume that either $\mathscr{B}$ is a normed space and $\mathscr{D}$ is its compact subset with the induced topology, or $\mathscr{B}$ is a reflexive Banach space and $\mathscr{D}$ is its closed convex subset with the weak topology. Let $\Theta \subset \mathscr{B}$. If a sequence of dominated experiments $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ satisfies conditions (Y') and (U') then*

$$\varliminf_{n\to\infty} \inf_{\rho_n \in \mathscr{R}_n} R^E_n(\rho_n) \geqslant E^*.$$

If conditions $(\sup Y')$ *and* $(\sup U')$ *hold then*

$$\lim_{n\to\infty} \inf_{\rho_n \in \mathscr{R}_n} R^E_n(\rho_n) = E^*,$$

and the interval-median estimators $\rho^E_{n,\delta} = r^E_{\delta,c}(Y_n)$ are nearly LD efficient:

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R^E_n(\rho^E_{n,\delta}) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R^E_n(\rho^E_{n,\delta}) = E^*.$$

Remark 4.2. If $\mathscr{B}$ is a separable reflexive Banach space then the Borel σ-fields for the strong and weak topologies coincide, hence the condition of measurability of $r^E_{\delta,c}$ does not depend on which topology on $\mathscr{B}$ has been chosen.

4.3. Estimation of linear functionals

Let Θ be a subset of a vector space and $L(\cdot)$ a linear functional on the vector space. Consider the problem of estimating $L(\theta)$. We take $\mathscr{D} = \mathbb{R}$, the real line. As above, we consider Bahadur-type criteria: the loss function is

$$W_\theta(r) = 1(|r - L(\theta)| > c), \qquad \theta \in \Theta,\ r \in \mathbb{R},$$

where $c > 0$ is fixed, and the risk of an estimator ρ_n is given by

$$R^F_n(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(|\rho_n - L(\theta)| > c). \tag{4.16}$$

The asymptotic minimax lower bound R'^* assumes the form

$$F^* = \sup_{y\in\mathscr{Y}} \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta:|r-L(\theta)|>c} (\zeta_\theta(y) - I(y)), \tag{4.17}$$

and subproblem (Q'_δ) becomes

(F_δ) $F_\delta(y) = \inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta:|r-L(\theta)|>c} \zeta_{\theta,\delta}(y), \qquad y \in \mathscr{Y}.$

Associated solutions $r'_\delta(y)$ can be constructed along the same lines as for the parameter estimation problem. Specifically, fixing y and δ, let us denote $f(\theta) = \zeta_{\theta,\delta}(y)$ and, for $h \in \mathbb{R}$ and $A(h)$ from (4.12), denote by $L \circ A(h)$ the image of $A(h)$ on the real line for the mapping L:

$$L \circ A(h) = \{L(\theta) \colon \theta \in A(h)\}.$$

Let $B(h)$ be the smallest closed interval in $\mathbb{R}$ containing $L \circ A(h)$. Furthermore, denoting by $d(B(h))$ the length of $B(h)$, set

$$h_{c,L} = \inf\{h \colon d(B(h)) \leqslant 2c\}.$$

Finally, consider the intervals $B_{c,L}$ of length $2c$ that contain $B(h_{c,L})$ (note that $d(B(h_{c,L})) \leqslant 2c$), and let $D_{c,L}$ be the set of the centres of all such intervals. The argument of the proof of Lemma 4.1 yields the following lemma.

Lemma 4.2. *The set $D_{c,L}$ is non-empty and consists of all $r_{c,L} \in \mathscr{D}$ at which $\inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta:|r-L(\theta)|>c} f(\theta)$ is attained. Also, the latter infimum equals $h_{c,l}$.*

To emphasize dependence on f, let us denote the elements of $D_{c,L}$ by $r_{c,L}(f)$. By the lemma, $r^F_{\delta,c}(y) = r_{c,L}(\zeta_{\Theta,\delta}(y))$ solves (F_δ). Assuming that the $r^F_{\delta,c}(y)$ are Borel functions from $\mathscr{Y}$ into $\mathbb{R}$, we introduce the estimators $\rho^F_{n,\delta}$ of $L(\theta)$ by

$$\rho^F_{n,\delta} = r_{c,L}(\zeta_{\Theta,\delta}(Y_n)), \tag{4.18}$$

and call them also interval-median. Applying Theorem 4.2, we obtain the following result.

Theorem 4.5. *If a sequence of dominated experiments $\{\mathscr{E}_n, P_n, n \geqslant 1\}$ satisfies conditions (Y') and (U') then*

$$\varliminf_{n\to\infty} \inf_{\rho_n\in\mathscr{R}_n} R^F_n(\rho_n) \geqslant F^*.$$

If conditions $(\sup Y')$ and $(\sup U')$ hold then

$$\lim_{n\to\infty} \inf_{\rho_n\in\mathscr{R}_n} R^F_n(\rho_n) = F^*,$$

and the interval-median estimators $\rho^F_{n,\delta} = r_{c,L}(\zeta_{\Theta,\delta}(Y_n))$ are nearly LD efficient:

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R^F_n(\rho^F_{n,\delta}) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R^F_n(\rho^F_{n,\delta}) = F^*.$$

We conclude the section by giving a more explicit representation for F^*.

Lemma 4.3. *Under the above notation and conditions,*

$$F^* = \sup_{\theta,\theta':|L(\theta-\theta')|>2c} S(\theta, \theta'),$$

where $S(\theta, \theta')$ is defined by (4.7).

Proof. We fix $y \in \mathscr{Y}$ with $I(y) < \infty$, set $f(\theta) = \zeta_\theta(y)$ and define $h_{c,L}$ as above. We show that

$$h_{c,L} = \sup_{\theta,\theta' : |L(\theta-\theta')| > 2c} \min\{f(\theta), f(\theta')\}.$$

By (4.17) and Lemma 4.2, this implies the claim.

Since $d(B(h)) \leqslant 2c$ for $h > h_{c,L}$, we have that if $\theta, \theta' \in \Theta$ are such that $|L(\theta - \theta')| > 2c$ then $\min(f(\theta), f(\theta')) \leqslant h_{c,L}$. Conversely, if $h < h_{c,L}$ then $d(B(h)) > 2c$, hence there exist $\theta, \theta' \in \Theta$ such that $L(\theta - \theta') > 2c$ and $f(\theta) > h$, $f(\theta') > h$, which, by the arbitrariness of $h < h_{c,L}$ completes the proof. □

Remark 4.3. The latter case of functional estimation includes the case of the estimation of a one-dimensional parameter θ if we take $L(\theta) = \theta$, so the result of Lemma 4.3 can be used for evaluating E^* from (4.11) too.

5. Statistical applications

In this section, we go back to the statistical models introduced in Section 2 and apply to them the general results of Sections 3 and 4. We first verify the LDP for the models by checking conditions (Y') and (U'). This is done under weaker assumptions than in Section 2. After that we give conditions that imply $(\sup Y')$ and $(\sup U')$. Next, considering certain hypothesis testing and estimation problems for the models, we calculate the asymptotic minimax risks and indicate (nearly) LD efficient decisions.

Each of the subsections below uses its own notation. We mention it if different subsections reuse certain symbols for the same objects. For the reader's convenience, we repeat the main points of the analysis of the models in Section 2 and recall the models themselves. Also, we implicitly assume that the functions we choose as estimators are properly measurable.

5.1. Gaussian observations

We observe a sample of n independent real-valued random variables $\mathbf{X}_n = (X_{1,n}, \ldots, X_{n,n})$ normally distributed as $\mathscr{N}(\theta, 1)$, $\theta \in \Theta \subset \mathbb{R}$. For this model, $\Omega_n = \mathbb{R}^n$ and $P_{n,\theta} = (\mathscr{N}(\theta, 1))^n$, $\theta \in \Theta$. We take $P_{n,0}$ as a dominating measure P_n. Then

$$\frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(\mathbf{X}) = \frac{1}{n} \sum_{k=1}^{n} \left(\theta X_k - \frac{1}{2}\theta^2 \right), \qquad \mathbf{X} = (X_1, \ldots, X_n) \in \mathbb{R}^n.$$

Thus, it is natural to take

$$Y_n = \frac{1}{n} \sum_{k=1}^{n} X_{k,n}, \qquad n \geqslant 1,$$

so that

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(\mathbf{X}_n) = \theta Y_n - \frac{1}{2}\theta^2.$$

Then $\{\mathscr{L}(Y_n|P_n),\ n \geqslant 1\}$ obeys the LDP in $\mathbb{R}$ with rate function $I^N(y) = y^2/2$, $y \in \mathbb{R}$ (see, e.g., Freidlin and Wentzell, 1979). This verifies condition $(Y'.1)$.

We next take

$$\zeta_\theta(y) = \zeta_{\theta,\delta}(y) = \theta y - \frac{1}{2}\theta^2. \tag{5.1}$$

Conditions $(Y'.2)$–$(Y'.4)$ are then obvious. Condition (U') follows by Chebyshev's inequality since

$$\mathrm{E}_n^{1/n} \exp(n\Xi_{n,\theta}) 1(\Xi_{n,\theta} > H) \leqslant \mathrm{e}^{-H} \mathrm{E}_n^{1/n} \exp(2n\Xi_{n,\theta}) \to \mathrm{e}^{-H}\mathrm{e}^{\theta^2}.$$

By Remark 2.2, the sequence $\{\mathscr{E}_n,\ n \geqslant 1\}$ obeys the LDP. Moreover, condition $(\sup Y)$ trivially holds. If, in addition, Θ is bounded, it readily follows that condition $(\sup U')$ is met as well.

We now turn to hypothesis testing and estimation problems and begin with calculating, for $\theta,\ \theta' \in \Theta$, the value of the function $S(\theta, \theta')$ from (4.7).

Lemma 5.1. *For all $\theta,\ \theta' \in \Theta$,*

$$S(\theta, \theta') := \sup_{y\in\mathbb{R}} \min\{\zeta_\theta(y) - I^N(y),\ \zeta_{\theta'}(y) - I^N(y)\} = -\frac{(\theta-\theta')^2}{8}.$$

Proof. By (5.1) and the definition of I^N, $\zeta_\theta(y) - I(y) = -(y-\theta)^2/2$, so

$$S(\theta, \theta') = \sup_{y\in\mathbb{R}} \min\left\{-\frac{(y-\theta)^2}{2},\ -\frac{(y-\theta')^2}{2}\right\} = -\frac{(\theta-\theta')^2}{8}.$$

□

5.1.1. Testing $\theta = 0$ versus $|\theta| \geqslant 2c$

Assume that Θ contains 0 as an internal point. We test the simple hypothesis H_0: $\theta = 0$ versus the two-sided alternative H_1: $|\theta| \geqslant 2c$ with some $c > 0$ such that the interval $[-2c, 2c]$ is contained in Θ. The logarithmic risk of a test ρ_n is given by (see (4.4))

$$R_n^T(\rho_n) = \max\left\{\frac{1}{n}\log P_{n,0}(\rho_n = 1),\ \frac{1}{n}\sup_{|\theta|\geqslant 2c}\log P_{n,\theta}(\rho_n = 0)\right\}.$$

Now, using (4.6) with $\Theta_0 = \{0\}$ and $\Theta_1 = \{\theta \in \Theta\colon |\theta| \geqslant 2c\}$ and Lemma 5.1, we readily obtain

$$T^* = \sup_{|\theta'|\geqslant 2c} S(0, \theta') = -\frac{c^2}{2}.$$

Next, by Theorem 4.3 and Remark 3.4, LD efficient tests ρ_n^T can be taken in the form

$$\rho_n^T = 1(\sup_{|\theta|\geqslant 2c} \zeta_\theta(Y_n) > \zeta_0(Y_n)) = 1\left(\sup_{|\theta|\geqslant 2c}\left(\theta Y_n - \frac{\theta^2}{2}\right) > 0\right) = 1(|Y_n| > c).$$

Applying Theorem 4.3 and Remark 3.4, we arrive at the following result.

Proposition 5.1. *Let* $[-2c, 2c] \subset \Theta$. *Then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) \geqslant -\frac{c^2}{2}.$$

If Θ *is bounded then*

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) = -\frac{c^2}{2},$$

and the tests ρ_n^T *are LD efficient:*

$$\lim_{n\to\infty} R_n^T(\rho_n^T) = -\frac{c^2}{2}.$$

5.1.2. Parameter estimation

We now consider the problem of estimating the parameter θ. We take the real line as a decision space $\mathscr{D}$. Recall (see (4.10)) that, for a given $c>0$, the risk of an estimator ρ_n is defined by

$$R_n^E(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta| > c).$$

In view of Remark 4.3, the asymptotic minimax risk E^* is given by Lemma 4.3:

$$E^* = \sup_{\theta,\theta'\in\Theta:|\theta-\theta'|>2c} S(\theta, \theta').$$

Lemma 5.1 implies that if Θ contains an interval of length greater than $2c$, then $E^* = -c^2/2$. An application of Theorem 4.4 and Remark 3.4 yields the following result.

Proposition 5.2. *Let* Θ *contain an interval of length greater than* $2c$. *Then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^E(\rho_n) \geqslant -\frac{c^2}{2}.$$

If Θ *is bounded then*

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^E(\rho_n) = -\frac{c^2}{2},$$

and the interval-median estimators $\rho_n^E = r_c(\zeta_\Theta(Y_n))$ *(see Section 4.2) are LD efficient:*

$$\lim_{n\to\infty} R_n^E(\rho_n^E) = -\frac{c^2}{2}.$$

Remark 5.1. It is easy to see that the estimator $\rho_n^E = r_c(\zeta_n(Y_n))$ coincides with Y_n if $Y_n - c \in \Theta$ and $Y_n + c \in \Theta$. Direct calculations show that the estimators $\hat{\rho}_n = Y_n$ are also LD efficient, i.e., $\lim_n R_n^E(\hat{\rho}_n) = -c^2/2$. The latter estimator is of simpler structure and does not depend on either c or Θ. However, the ρ_n^E seem to perform better at points outside or close to the boundary of Θ. In particular, if $Y_n \notin \Theta$ then $\hat{\rho}_n \notin \Theta$, whereas, for Θ convex, ρ_n^E always belongs to Θ.

5.2. An independent and identically distributed sample

We observe an independent and identically distributed sample $\mathbf{X}_n = (X_{1,n}, \ldots, X_{n,n})$ from a distribution P_θ, $\theta \in \Theta$, on the real line. We assume that the family $\mathscr{P} = \{P_\theta, \theta \in \Theta\}$ is dominated by a probability measure P, i.e., $P_\theta \ll P$, $\theta \in \Theta$. This model is described by dominated experiments $\mathscr{E}_n = (\Omega_n, \mathscr{F}_n; P_{n,\theta}, \theta \in \Theta)$ with $\Omega_n = \mathbb{R}^n$, $\mathscr{F}_n = \mathscr{B}(\mathbb{R}^n)$, $P_{n,\theta} = P_\theta^n$, $\theta \in \Theta$, and $P_n = P^n$.

Assume that the family $\mathscr{P}$ satisfies the following regularity conditions:

(R.1) *the densities* $\mathrm{d}P_\theta/\mathrm{d}P(x)$, $\theta \in \Theta$, *are continuous and positive functions of* $x \in \mathbb{R}$;
(R.2) $\int_{\mathbb{R}}((\mathrm{d}P_\theta/\mathrm{d}P)(x))^\gamma P(\mathrm{d}x) < \infty$, $\theta \in \Theta$, *for all* $\gamma \in \mathbb{R}$.

We have that

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(\mathbf{X}_n) = \sum_{k=1}^{n} \frac{1}{n} \log \frac{\mathrm{d}P_\theta}{\mathrm{d}P}(X_{k,n}) = \int_{\mathbb{R}} \log \frac{\mathrm{d}P_\theta}{\mathrm{d}P}(x) F_n(\mathrm{d}x),$$

where

$$F_n(x) = \frac{1}{n} \sum_{k=1}^{n} 1(X_{k,n} \leqslant x), \quad x \in \mathbb{R}, \tag{5.2}$$

are empirical distribution functions.

We take the latter as statistics Y_n in condition (Y). The underlying space $\mathscr{Y}$ is the space of cumulative distribution functions on $\mathbb{R}$ which we denote by $\mathscr{F}$ and endow with the topology of weak convergence of associated probability measures. By Sanov's theorem (Sanov 1957; Deuschel and Stroock 1989, Section 3.2.17), the sequence $\{\mathscr{L}(Y_n|P_n), n \geqslant 1\}$ obeys the LDP with rate function $I^S(F) = K(F, P)$, $F \in \mathscr{F}$, where $K(F, P)$ is the Kullback–Leibler information:

$$K(F, P) = \begin{cases} \int_{\mathbb{R}} \frac{\mathrm{d}F}{\mathrm{d}P}(x) \log \frac{\mathrm{d}F}{\mathrm{d}P}(x) P(\mathrm{d}x), & \text{if } F \ll P, \\ \infty, & \text{otherwise.} \end{cases} \tag{5.3}$$

This checks condition (Y'.1). The verification of the rest of condition (Y') is more intricate than in the previous example.

Denote for $\theta \in \Theta$, $x \in \mathbb{R}$ and $\delta > 0$,

$$L_\theta(x) = \log \frac{\mathrm{d}P_\theta}{\mathrm{d}P}(x),$$
$$L_{\theta,\delta}(x) = L_\theta(x) \wedge \delta^{-1} \vee (-\delta^{-1})$$

and let

$$\zeta_{\theta,\delta}(F) = \int_{\mathbb{R}} L_{\theta,\delta}(x) F(\mathrm{d}x), \qquad F \in \mathscr{F}.$$

By $(R.1)$, the functions $\zeta_{\theta,\delta}$ are continuous on $\mathscr{F}$, so $(Y'.2)$ holds.

We now check $(Y'.3)$. Condition $(R.2)$ implies that, for all $\gamma > 0$,

$$\lim_{\delta\to 0} \int_{\mathbb{R}} [\exp(\gamma |L_\theta(x) - L_{\theta,\delta}(x)|) - 1] P(\mathrm{d}x) = 0. \tag{5.4}$$

Then, for $\gamma > 0$, $\varepsilon > 0$, with the use of Chebyshev's inequality,

$$\begin{aligned} P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon) &\leqslant P_n^{1/n}\left(\int_{\mathbb{R}} |L_\theta(x) - L_{\theta,\delta}(x)| F_n(\mathrm{d}x) > \varepsilon\right) \\ &\leqslant \exp(-\gamma\varepsilon) \mathrm{E}_n^{1/n} \exp\left(n\gamma \int_{\mathbb{R}} |L_\theta(x) - L_{\theta,\delta}(x)| F_n(\mathrm{d}x)\right) \\ &= \exp(-\gamma\varepsilon) \int_{\mathbb{R}} \exp(\gamma |L_\theta(x) - L_{\theta,\delta}(x)|) P(\mathrm{d}x). \end{aligned}$$

By (5.4), it then follows that

$$\overline{\lim_{\delta\to 0}}\ \overline{\lim_{n\to\infty}}\ P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon) \leqslant \exp(-\gamma\varepsilon).$$

Since γ is arbitrary, $(Y'.3)$ follows.

We next check $(Y'.4)$ with

$$\zeta_\theta(F) = \begin{cases} \int_{\mathbb{R}} L_\theta(x) F(\mathrm{d}x), & \text{if } I^S(F) < \infty, \\ 0, & \text{otherwise.} \end{cases} \tag{5.5}$$

To begin, we show that the ζ_θ are well defined. Since the functions $x \log x - x + 1$ and $\exp x - 1$ are convex conjugates (Rockafellar 1970), by the Young–Fenchel inequality (Rockafellar 1970; Krasnoselskii and Rutickii 1961), for $F \ll P$,

$$\begin{aligned} \int_{\mathbb{R}} \left| L_\theta(x) \frac{\mathrm{d}F}{\mathrm{d}P}(x) \right| P(\mathrm{d}x) &\leqslant \int_{\mathbb{R}} [\exp(|L_\theta(x)|) - 1] P(\mathrm{d}x) \\ &\quad + \int_{\mathbb{R}} \left(\frac{\mathrm{d}F}{\mathrm{d}P}(x) \log \frac{\mathrm{d}F}{\mathrm{d}P}(x) - \frac{\mathrm{d}F}{\mathrm{d}P}(x) + 1 \right) P(\mathrm{d}x) \\ &\leqslant 1 + \int_{\mathbb{R}} \left(\frac{\mathrm{d}P_\theta}{\mathrm{d}P}(x) \right)^{-1} P(\mathrm{d}x) + I^S(F). \end{aligned}$$

In view of $(R.2)$, this proves that the ζ_θ are well defined.

Now, for F with $I^S(F) < \infty$, we have, for $\gamma > 0$, using the Young–Fenchel inequality again,

$$\begin{aligned}\gamma|\zeta_{\theta,\delta} - \zeta_\theta(F)| &\leqslant \int_{\mathbb{R}} \gamma|L_{\theta,\delta}(x) - L_\theta(x)|F(\mathrm{d}x)\\ &\leqslant \int_{\mathbb{R}} [\exp(\gamma|L_{\theta,\delta}(x) - L_\theta(x)|) - 1]P(\mathrm{d}x)\\ &\quad + \int_{\mathbb{R}} \left(\frac{\mathrm{d}F}{\mathrm{d}P}(x)\log\frac{\mathrm{d}F}{\mathrm{d}P}(x) - \frac{\mathrm{d}F}{\mathrm{d}P}(x) + 1\right)P(\mathrm{d}x)\\ &= \int_{\mathbb{R}} [\exp(\gamma|L_{\theta,\delta}(x) - L_\theta(x)|) - 1]P(\mathrm{d}x) + I^S(F).\end{aligned}$$

Hence, by (5.4)

$$\overline{\lim_{\delta\to 0}} \sup_{F\in\Phi'_{I^S}(a)} |\zeta_{\theta,\delta}(F) - \zeta_\theta(F)| \leqslant \frac{a}{\gamma},$$

and letting $\gamma \to \infty$, we arrive at $(Y'.4)$. Remark 2.2 then implies that the the LDP holds for $\{\mathscr{L}(\Xi_{n,\Theta}|P_n),\ n \geqslant 1\}$.

It remains to check (U'). Using Chebyshev's inequality once again, we obtain, for $H > 0$,

$$\begin{aligned}\mathrm{E}_n^{1/n}\exp(n\Xi_{n,\theta})1(\Xi_{n,\theta} > H) &\leqslant \exp(-H)\mathrm{E}_n^{1/n}\exp(2n\Xi_{n,\theta})\\ &= \exp(-H)\int_{\mathbb{R}}\left(\frac{\mathrm{d}P_\theta}{\mathrm{d}P}(x)\right)^2 P(\mathrm{d}x),\end{aligned}$$

and the result follows by condition $(R.2)$.

Conditions (Y') and (U') have been verified, and thus the LDP holds.

Remark 5.2. It is possible to do without condition $(R.1)$. Then the functions $L_{\theta,\delta} = (L_{\theta,\delta}(x),\ x \in \mathbb{R})$, $\delta > 0$, $\theta \in \Theta$, should be chosen bounded, continuous and so that (5.4) holds. The existence of such functions follows from $(R.2)$.

To check $(\sup Y')$ and $(\sup U')$, we assume that stronger versions of conditions $(R.1)$ and $(R.2)$ hold:

- $(\sup R.1)$ *the functions* $(\mathrm{d}P_\theta/\mathrm{d}P)(x)$, $\theta \in \Theta$, *are positive and equicontinuous at each* $x \in \mathbb{R}$;
- $(\sup R.2)$ $\sup_{\theta\in\Theta} \int_{\mathbb{R}}((\mathrm{d}P_\theta/\mathrm{d}P)(x))^\gamma P(\mathrm{d}x) < \infty$, *for all* $\gamma \in \mathbb{R}$.

Defining ζ_θ, $\zeta_{\theta,\delta}$, L_θ and $L_{\theta,\delta}$ as above, we have, by $(\sup R.2)$, that for all $\gamma > 0$

$$\lim_{\delta\to 0}\sup_{\theta\in\Theta}\int_{\mathbb{R}}[\exp(\gamma|L_\theta(x) - L_{\theta,\delta})|) - 1]P(\mathrm{d}x) = 0.$$

The latter equality enables us to check conditions $(\sup Y'.3)$ and $(\sup Y'.4)$ in the same way as conditions $(Y'.3)$ and $(Y'.4)$. Condition $(\sup U')$ is also checked analogously to condition (U'), with the use of $(\sup R.2)$. Condition $(Y'.1)$ has already been checked.

It remains to check $(\sup Y'.2)$. We show that the functions $(\zeta_{\theta,\delta}(F),\ \theta \in \Theta)$ are continuous in F for the uniform topology on $\mathbb{R}_+^\Theta$ which obviously implies $(\sup Y'.2)$. Since the weak topology on $\mathscr{F}$ is metrizable, it is enough to check sequential continuity. Let $F^{(n)}$ weakly converge to F as $n \to \infty$. Then the definition of the $L_{\theta,\delta}$, and (sup R.1) imply that the $L_{\theta,\delta}(x)$, $\theta \in \Theta$, for δ fixed, are uniformly bounded and equicontinuous at each $x \in \mathbb{R}$ so that (see, e.g., Billingsley 1968, Problem 8, Section 2)

$$\sup_{\theta\in\Theta}\left|\int_{\mathbb{R}} L_{\theta,\delta}(x)F^{(n)}(\mathrm{d}x) - \int_{\mathbb{R}} L_{\theta,\delta}(x)F(\mathrm{d}x)\right| \to 0$$

verifying (sup Y'.2). Conditions (sup Y') and (sup U') have been checked.

We now proceed to considering concrete statistical problems for the model. For this we need the following result by Chernoff (1952); see also Kullback (1959).

Lemma 5.2. *Let $\mathscr{P}$ be the space of probability measures on a Polish space $\mathbb{E}$ with the Borel σ-field, and let measures P, $Q \in \mathscr{P}$ be dominated by a measure μ and have respective densities $p(x)$ and $q(x)$. Then*

$$\inf_{F\in\mathscr{P}} \max\{K(F, P), K(F, Q)\} = C(P, Q),$$

where $K(F, P)$ is the Kullback–Leibler information (5.3) and $C(P, Q)$ is Chernoff's function:

$$C(P, Q) = -\inf_{\gamma\in[0,1]} \log \int_{\mathbb{E}} p^{\gamma}(x)q^{1-\gamma}(x)\mu(\mathrm{d}x).$$

We next apply Lemma 5.2 to calculating the function $S(\theta, \theta')$ from (4.7).

Lemma 5.3. *For θ, $\theta' \in \Theta$,*

$$S(\theta, \theta') := \sup_{F\in\mathscr{F}} \min\{\zeta_{\theta}(F) - I^S(F), \zeta_{\theta'}(F) - I^S(F)\} = -C(P_{\theta}, P_{\theta'}).$$

Proof. Let $I^S(F) < \infty$. Then $F \ll P$ and, since the densities $\mathrm{d}P_{\theta}/\mathrm{d}P(x)$, $\theta \in \Theta$, are positive, we also have that $F \ll P_{\theta}$ and P-almost surely

$$\frac{\mathrm{d}F}{\mathrm{d}P} = \frac{\mathrm{d}F}{\mathrm{d}P_{\theta}}\frac{\mathrm{d}P_{\theta}}{\mathrm{d}P}.$$

Therefore, by the definitions of ζ_{θ}, and I^S,

$$\begin{aligned}\zeta_{\theta}(F) - I^S(F) &= \int_{\mathbb{R}} \log\frac{\mathrm{d}P_{\theta}}{\mathrm{d}P}(x)F(\mathrm{d}x) - \int_{\mathbb{R}} \log\frac{\mathrm{d}F}{\mathrm{d}P}F(\mathrm{d}x)\\ &= -\int_{\mathbb{R}} \log\frac{\mathrm{d}F}{\mathrm{d}P_{\theta}}F(\mathrm{d}x) = -K(F, P_{\theta}),\end{aligned}$$

and the result follows by Lemma 5.2. □

We now give an application to hypothesis testing problems. Consider the tests from (4.8):

$$\rho^T_{n,\delta} = 1\big(\sup_{\theta\in\Theta_0} \zeta_{\theta,\delta}(F_n) < \sup_{\theta\in\Theta_1} \zeta_{\theta,\delta}(F_n)\big).$$

As above, the risk $R_n^T(\rho_n)$ of a test ρ_n is defined by (4.4). By (4.6) and Lemma 5.3,

$$T^* = -\inf_{\theta\in\Theta_0,\theta'\in\Theta_1} C(P_\theta, P_{\theta'}),$$

so Theorem 4.3 yields the following.

Proposition 5.3. *Let* Θ_1 *and* Θ_2 *be non-intersecting subsets of* Θ. *If conditions* $(R.1)$ *and* $(R.2)$ *hold then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) \geqslant -\inf_{\theta\in\Theta_0,\theta'\in\Theta_1} C(P_\theta, P_{\theta'}).$$

If conditions $(\sup R.1)$ *and* $(\sup R.2)$ *hold then*

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) = -\inf_{\theta\in\Theta_0,\theta'\in\Theta_1} C(P_\theta, P_{\theta'}),$$

and the tests $\rho_{n,\delta}^T$ *are nearly LD efficient, i.e.,*

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R_n^T(\rho_{n,\delta}^T)$$

$$= -\inf_{\theta\in\Theta_0,\theta'\in\Theta_1} C(P_\theta, P_{\theta'}).$$

In a similar manner one can tackle estimation problems for θ or linear functionals of θ.

5.3. 'Signal plus white noise'

We observe a real-valued stochastic process $X_n = (X_n(t),\ t\in[0, 1])$ obeying the stochastic differential equation

$$\mathrm{d}X_n(t) = \theta(t)\,\mathrm{d}t + \frac{1}{\sqrt{n}}\,\mathrm{d}W(t), \qquad 0 \leqslant t \leqslant 1, \tag{5.6}$$

where $W = (W(t),\ t\in[0, 1])$ is a standard Wiener process and $\theta(\cdot)$ is an unknown continuous function.

This model is described by statistical experiments $\mathscr{E}_n = (\Omega_n, \mathscr{F}_n;\ P_{n,\theta},\ \theta\in\Theta)$, where $\Omega_n = C[0, 1]$, the space of continuous functions on $[0, 1]$ with the uniform metric, $\Theta \subset C[0, 1]$ and $P_{n,\theta}$ is the distribution of X_n on $C[0, 1]$ for θ. We take $P_n = P_{n,0}$, where $P_{n,0}$ corresponds to the zero function $\theta(\cdot)\equiv 0$. Then $P_{n,\theta} \ll P_n$ and, moreover, by Girsanov's formula, P_n-almost surely,

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n) = \int_0^1 \theta(t)\,\mathrm{d}X_n(t) - \frac{1}{2}\int_0^1 \theta^2(t)\,\mathrm{d}t. \tag{5.7}$$

So, to check condition (Y'), we take $Y_n = X_n$ and $\mathscr{Y} = C[0, 1]$.

Let $C_0[0, 1]$ be the subset of $C[0, 1]$ of the functions $x(\cdot)$ that are absolutely continuous with respect to Lebesgue measure and equal to 0 at 0. Since the sequence $\{\mathscr{L}(X_n|P_n),\ n\geqslant 1\}$ obeys the LDP in $C[0, 1]$ with rate function

$$I^W(x(\cdot)) = \begin{cases} \dfrac{1}{2}\displaystyle\int_0^1 (\dot{x}(t))^2\, dt, & \text{if } x(\cdot) \in C_0[0, 1], \\ \infty, & \text{otherwise,} \end{cases} \tag{5.8}$$

where $\dot{x}(t)$ denotes the derivative of $x(\cdot) \in C[0, 1]$ at t (see, e.g., Freidlin and Wentzell, 1979), condition $(Y'.1)$ holds.

We next take

$$\zeta_{\theta,\delta}(x(\cdot)) = \int_0^1 \theta_\delta(t)\, dx(t) - \frac{1}{2}\int_0^1 \theta^2(t)\, dt, \qquad x(\cdot) \in C[0, 1], \tag{5.9}$$

where

$$\theta_\delta(t) = \sum_{k=0}^{[1/\delta]} \theta(k\delta) 1(t \in [k\delta, (k+1)\delta)), \qquad t \in [0, 1], \tag{5.10}$$

the first integral on the right of (5.9) being understood as a finite sum.

By the continuity of $\theta(\cdot)$,

$$\lim_{\delta\to 0} \int_0^1 (\theta(t) - \theta_\delta(t))^2\, dt = 0. \tag{5.11}$$

The $\zeta_{\theta,\delta}$ are obviously continuous in $x(\cdot) \in C[0, 1]$, so $(Y'.2)$ holds. Next, by (5.7) and (5.9), we have, for $\varepsilon > 0$ and $\gamma > 0$, in view of Chebyshev's inequality,

$$\begin{aligned} P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(X_n)| > \varepsilon) &\leqslant P_n^{1/n}\left(\left|\int_0^1 (\theta(t) - \theta_\delta(t)) \frac{1}{\sqrt{n}}\, dW(t)\right| > \varepsilon\right) \\ &\leqslant 2e^{-\gamma\varepsilon} \exp\left(\frac{\gamma^2}{2}\int_0^1 (\theta(t) - \theta_\delta(t))^2\, dt\right), \end{aligned}$$

and by (5.11)

$$\overline{\lim_{\delta\to 0}}\ \overline{\lim_{n\to\infty}}\ P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(X_n)| > \varepsilon) \leqslant 2\exp(-\gamma\varepsilon),$$

which proves $(Y'.3)$ by the arbitrariness of γ.

For condition $(Y'.4)$, we take

$$\zeta_\theta(x(\cdot)) = \begin{cases} \displaystyle\int_0^1 \theta(t)\dot{x}(t)\, dt - \frac{1}{2}\int_0^1 \theta^2(t)\, dt, & \text{if } I^W(x(\cdot)) < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

The ζ_θ are well defined, since, by the Cauchy–Schwarz inequality and (5.8), if $x(\cdot)$ is absolutely continuous then

$$\int_0^1 |\theta(t)\dot{x}(t)|\, dt \leqslant \left(\int_0^1 \theta^2(t)\, dt\right)^{1/2} (2I^W(x(\cdot)))^{1/2}.$$

Moreover, if $I^W(x(\cdot)) < \infty$ then

$$|\zeta_{\theta,\delta}(x(\cdot)) - \zeta_\theta(x(\cdot))| \leqslant \int_0^1 |\theta_\delta(t) - \theta(t)|\,|\dot{x}(t)|\,\mathrm{d}t$$
$$\leqslant \left(\int_0^1 (\theta_\delta(t) - \theta(t))^2\,\mathrm{d}t\right)^{1/2} \left(\int_0^1 (\dot{x}(t))^2\,\mathrm{d}t\right)^{1/2},$$

so

$$\sup_{x(\cdot)\in\Phi'_{I^W}(a)} |\zeta_{\theta,\delta}(x(\cdot)) - \zeta_\theta(x(\cdot))| \leqslant (2a)^{1/2}\left(\int_0^1 (\theta_\delta(t) - \theta(t))^2\,\mathrm{d}t\right)^{1/2},$$

and the latter goes to 0 as $\delta \to 0$ by (5.11). Condition (Y') has been verified.

It remains to check (U'). Using the model equation (5.6), (5.7) and Chebyshev's inequality once again, we have that

$$\mathrm{E}_n^{1/n} \exp(n\Xi_{n,\theta}) 1(\Xi_{n,\theta} > H) \leqslant \exp(-H)\mathrm{E}_n^{1/n} \exp(2n\Xi_{n,\theta})$$
$$= \exp(-H)\exp\left(\int_0^1 \theta^2(t)\,\mathrm{d}t\right) \to 0 \qquad \text{as } H \to \infty,$$

verifying condition (U').

Remark 5.3. The condition of continuity of the functions $\theta(\cdot)$ can be weakened to the condition

$$\int_0^1 \theta^2(t)\,\mathrm{d}t < \infty.$$

The functions θ_δ should then be chosen as step functions for which (5.11) holds.

For conditions $(\sup Y')$ and $(\sup U')$, we require that the functions $\theta(\cdot)$ belong to a compact set in $C[0, 1]$. More specifically, for fixed $\beta \in (0, 1]$, $M > 0$ and $K > 0$, we introduce the Hölder class

$$\Sigma(\beta, M) = \{\theta(\cdot)\colon |\theta(t) - \theta(s)| \leqslant M|t - s|^\beta, \text{ for all } s, t \in [0, 1]\}, \tag{5.12}$$

define $\Sigma_K(\beta, M)$ to be the subset of $\Sigma(\beta, M)$ of functions θ such that $\sup_{t\in[0,1]} |\theta(t)| \leqslant K$ and assume that $\Theta \subset \Sigma_K(\beta, M)$. By the Arzelà–Ascoli theorem, the set $\Sigma_K(\beta, M)$ is compact in $C[0, 1]$. Also

$$\sup_{\theta(\cdot)\in\Sigma_K(\beta,M)} \int_0^1 \theta^2(t)\,\mathrm{d}t < \infty \tag{5.13}$$

and

$$\lim_{\delta\to 0} \sup_{\theta(\cdot)\in\Sigma_K(\beta,M)} \int_0^1 (\theta(t) - \theta_\delta(t))^2\,\mathrm{d}t = 0. \tag{5.14}$$

Now conditions (sup $Y'.3$) and (sup $Y'.4$) are checked as conditions ($Y'.3$) and ($Y'.4$), respectively, with the use of (5.14) in place of (5.11). Condition (sup $Y'.2$) follows by the uniform boundedness of functions from $\Sigma_K(\beta, M)$, which implies that $x(\cdot) \to (\zeta_{\theta,\delta}(x(\cdot)),\ \theta \in \Sigma_K(\beta, M))$ is a continuous map from $C[0, 1]$ into $\mathbb{R}_+^\Theta$ with the uniform topology.

Finally, condition (sup U') follows in analogy with condition (U') with the use of (5.13). This completes verification of conditions (sup Y') and (sup U').

We now calculate the function $S(\theta, \theta')$ for the model.

Lemma 5.4. *For all $\theta, \theta' \in C[0, 1]$,*

$$S(\theta, \theta') := \sup_{x(\cdot)\in C[0,1]} \min\{\zeta_\theta(x(\cdot)) - I^W(x(\cdot)), \zeta_{\theta'}(x(\cdot)) - I^W(x(\cdot))\}$$

$$= -\frac{1}{8}\int_0^1 [\theta(t) - \theta'(t)]^2 \,\mathrm{d}t.$$

Proof. Since by the definitions of I^W and ζ_θ, for $x(\cdot)$ with $I^W(x(\cdot)) < \infty$,

$$\zeta_\theta(x(\cdot)) - I^W(x(\cdot)) = -\frac{1}{2}\int_0^1 (\dot{x}(t) - \theta(t))^2 \,\mathrm{d}t,$$

we obtain, by the inequality $\max(a^2, b^2) \geqslant (a-b)^2/4$,

$$S(\theta, \theta') = -\inf_{x(\cdot)\in C[0,1]} \max\left\{\frac{1}{2}\int_0^1 [\dot{x}(t) - \theta(t)]^2 \,\mathrm{d}t, \frac{1}{2}\int_0^1 [\dot{x}(t) - \theta'(t)]^2 \,\mathrm{d}t\right\}$$

$$\leqslant -\frac{1}{8}\int_0^1 [\theta(t) - \theta'(t)]^2 \,\mathrm{d}t.$$

On the other hand, for $x(\cdot)$ with $\dot{x}(t) = [\theta(t) + \theta'(t)]/2$, we have that

$$\frac{1}{2}\int_0^1 [\dot{x}(t) - \theta(t)]^2 \,\mathrm{d}t = \frac{1}{2}\int_0^1 [\dot{x}(t) - \theta'(t)]^2 \,\mathrm{d}t = \frac{1}{8}\int_0^1 [\theta(t) - \theta'(t)]^2 \,\mathrm{d}t,$$

and the result follows. □

Now we apply these formulae and the general results from Section 4 to two statistical problems concerning the value of the function $\theta(\cdot)$ at an internal point t_0 of [0, 1].

5.3.1. *Testing $\theta(t_0) = 0$ versus $|\theta(t_0)| \geqslant 2c$*

Given $c > 0$, denote $\Theta_0 = \{\theta \in \Theta\colon \theta(t_0) = 0\}$, $\Theta_1 = \{\theta \in \Theta\colon |\theta(t_0)| \geqslant 2c\}$ and define the risk $R_n^T(\rho_n)$ of a test ρ_n by (4.4). Introduce

$$t^* = (c/M)^{1/\beta}. \tag{5.15}$$

Proposition 5.4. *Let c, β, M, K and t_0 be such that $[t_0 - t^*, t_0 + t^*] \subset [0, 1]$ and $K \geqslant 2c$. If $\Theta = \Sigma(\beta, M)$ then*

$$\underline{\lim}_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) \geqslant -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)}\left(\frac{c}{M}\right)^{1/\beta}.$$

If $\Theta = \Sigma_K(\beta, M)$ then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)}\left(\frac{c}{M}\right)^{1/\beta},$$

and the tests $\rho_{n,\delta}^T$ from (4.8) are nearly LD efficient, i.e.,

$$\lim_{\delta\to 0} \overline{\lim}_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = \lim_{\delta\to 0} \underline{\lim}_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)}\left(\frac{c}{M}\right)^{1/\beta}.$$

Proof. By Theorem 4.3, we need only to calculate T^* from (4.6). Denote

$$\theta^*(t) = [c - M|t - t_0|^\beta]^+, \tag{5.16}$$

where $a^+ = \max(a, 0)$. If $\theta \in \Theta_0$ and $\theta' \in \Theta_1$ then the inequality $|\theta(t_0) - \theta'(t_0)| \geqslant 2c$ and the Hölder constraints (5.12) imply that $|\theta(t) - \theta'(t)| \geqslant 2[c - M|t - t_0|^\beta]^+ = 2\theta^*(t)$, and hence

$$\int_0^1 (\theta(t) - \theta'(t))^2 \, dt \geqslant \int_0^1 4(\theta^*(t))^2 \, dt.$$

This yields, by Lemma 5.4,

$$S(\theta, \theta') \leqslant -\frac{1}{8} 4\int_0^1 (\theta^*(t))^2 \, dt = -\int_0^{t^*} (c - Mt^\beta)^2 \, dt$$

$$= -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)}\left(\frac{c}{M}\right)^{1/\beta}.$$

On the other hand, evidently, $c - \theta^* \in \Theta_0$, $c + \theta^* \in \Theta_1$ and $S(c - \theta^*, c + \theta^*) = -\frac{1}{2}\int_0^1 (\theta^*(t))^2 \, dt$ so that

$$T^* = \sup_{\theta\in\Theta_0, \theta'\in\Theta_1} S(\theta, \theta') = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta = 1)}\left(\frac{c}{M}\right)^{1/\beta},$$

and the proof is complete. □

5.3.2. *Estimating $\theta(t_0)$*

Treating $\theta(t_0)$ as a linear functional of $\theta(\cdot)$, we define the risk of an estimator ρ_n of $\theta(t_0)$ by

$$R_n^F(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta(t_0)| > c).$$

Proposition 5.5. *Let c, β, M, K and t_0 be such that $[t_0 - t^*,\, t_0 + t^*] \subset [0, 1]$ and $K > c$. If $\Theta = \Sigma(\beta, M)$ then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) \geqslant -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)}\left(\frac{c}{M}\right)^{1/\beta}.$$

If $\Theta = \Sigma_K(\beta, M)$ then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)}\left(\frac{c}{M}\right)^{1/\beta},$$

and the interval-median estimators $\rho_{n,\delta}^F$ from (4.18) are nearly LD efficient, i.e.,

$$\lim_{\delta\to 0} \varlimsup_{n\to\infty} R_n^T(\rho_{n,\delta}^F) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R_n^T(\rho_{n,\delta}^F) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)}\left(\frac{c}{M}\right)^{1/\beta}.$$

Proof. By Theorem 4.5 and Lemma 4.3,

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) \geqslant F^* = \sup_{\theta,\theta' : |\theta(t_0)-\theta'(t_0)| > 2c} S(\theta, \theta').$$

Repeating the above calculation for the testing problem, we obtain with $\theta^*(t)$ from (5.16)

$$F^* = S(\theta^*, -\theta^*) = -\frac{2\beta^2 c^2}{(\beta+1)(2\beta+1)}\left(\frac{c}{M}\right)^{1/\beta},$$

and we are done. □

Remark 5.4. The latter problem has been studied by Korostelev (1996), who suggests different upper estimators, namely, the kernel estimators

$$\hat{\rho}_n = \int K(t_0 - t)\,\mathrm{d}X_n(t)$$

with the kernel $K(t) = [(\beta+1)/(2c\beta)](M/c)^{1/\beta}[c - M|t - t_0|]^+$. These estimators have proved to be asymptotically efficient in the sense that $R_n^T(\hat{\rho}_n) \to F^*$ as $n \to \infty$.

5.4. Gaussian regression

We consider the regression model

$$X_{k,n} = \theta(t_{k,n}) + \xi_{k,n}, \qquad t_{k,n} = \frac{k}{n},\ k = 1, \ldots, n, \tag{5.17}$$

where errors $\xi_{k,n}$ are independent standard normal and $\theta(\cdot)$ is an unknown continuous function.

In this model, $\Omega_n = \mathbb{R}^n$, $\Theta \subset C[0, 1]$ and $P_{n,\theta}$ is the distribution of $\mathbf{X}_n = (X_{1,n}, \ldots, X_{n,n})$ for $\theta(\cdot)$. As above, we take $P_n = P_{n,0}$. Then

$$\begin{aligned}\Xi_{n,\theta} &= \frac{1}{n}\log\frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n)\\ &= \frac{1}{n}\sum_{k=1}^{n}\theta(t_{k,n})X_{k,n} - \frac{1}{2n}\sum_{k=1}^{n}\theta^2(t_{k,n})\\ &= \int_0^1 \theta(t)\,\mathrm{d}X_n(t) = \frac{1}{2n}\sum_{k=1}^{n}\theta^2(t_{k,n}),\end{aligned} \tag{5.18}$$

where

$$X_n(t) = \frac{1}{n}\sum_{k=1}^{[nt]} X_{k,n}, \qquad 0 \leqslant t \leqslant 1.$$

This prompts taking the process $X_n = (X_n(t),\ t \in [0, 1])$ as a statistic Y_n in condition (Y'). We define $\mathscr{Y}$ to be the space of right-continuous functions on [0, 1] with left-hand limits and with the uniform metric.

Since the $X_{k,n}$ are distributed as $\mathscr{N}(0, 1)$ under P_n, the sequence $\{\mathscr{L}(X_n|P_n),\ n \geqslant 1\}$ obeys the LDP with I^W from (5.8) (Mogulskii 1976). This verifies condition $(Y'.1)$.

Next, we define $\zeta_{\theta,\delta}(x(\cdot))$ as in Section 5.3, i.e.,

$$\zeta_{\theta,\delta}(x(\cdot)) = \int_0^1 \theta_\delta(t)\,\mathrm{d}x(t) - \frac{1}{2}\int_0^1 \theta^2(t)\,\mathrm{d}t,\ x(\cdot) \in \mathscr{Y}, \tag{5.19}$$

with $\theta_\delta(t)$ as in (5.10). Note that the $\zeta_{\theta,\delta}$ are measurable with respect to the Borel σ-field on $\mathscr{Y}$ and continuous at $x(\cdot)$ with $I^W(x(\cdot)) = \infty$ since they are continuous at continuous functions and $I^W(x(\cdot)) = \infty$ when $x(\cdot)$ is not absolutely continuous. This verifies condition $(Y'.2)$.

Now, by (5.18) and (5.19),

$$\begin{aligned}P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(X_n)| > \varepsilon) \leqslant 1\left(\left|\int_0^1 \theta^2(t)\,\mathrm{d}t - \frac{1}{n}\sum_{k=1}^{n}\theta^2(k/n)\right| > \varepsilon/4\right)\\ + P_n^{1/n}\left(\left|\int_0^1 (\theta(t) - \theta_\delta(t))\,\mathrm{d}X_n(t)\right| > \varepsilon/2\right).\end{aligned}$$

The first term on the right is zero for all n large enough by the continuity of $\theta(\cdot)$. The second, for $\gamma > 0$, is not greater than

$$\mathrm{e}^{-\gamma\varepsilon/2}\mathrm{E}_n^{1/n}\exp\left(n\gamma\left|\int_0^1 (\theta(t) - \theta_\delta(t))\,\mathrm{d}X_n(t)\right|\right) \leqslant 2\mathrm{e}^{-\gamma\varepsilon/2}\exp\left(\frac{\gamma^2}{2n}\sum_{k=1}^{n}(\theta(k/n) - \theta_\delta(k/n))^2\right).$$

Since the $\theta(\cdot)$ are continuous and the $\theta_\delta(\cdot)$ are step functions,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}(\theta(k/n) - \theta_\delta(k/n))^2 = \int_0^1 (\theta(t) - \theta_\delta(t))^2\,\mathrm{d}t,$$

and the latter goes to 0 as $\delta \to 0$. Since γ is arbitrary, condition $(Y'.3)$ follows.

Conditions $(Y'.4)$ and (U') are checked as for the 'signal plus white noise' model (with the same choice of ζ_θ).

Remark 5.5. As in the 'signal plus white noise' model, instead of continuity of $\theta(\cdot)$, we could require that it be square-integrable on [0, 1].

To obtain nearly LD efficient decisions, we assume that the $\theta(\cdot)$ belong to the class $\Sigma_K(\beta, M)$ defined above. Conditions $(\sup Y'.2)$, $(\sup Y'.3)$, $(\sup Y'.4)$ and $(\sup U')$ are checked as for the 'signal plus white noise' model if, in addition, we take into account that

$$\lim_{n\to\infty} \sup_{\theta(\cdot)\in\Sigma_K(\beta,M)} \int_0^1 (\theta([nt]+1/n) - \theta(t))^2 \, \mathrm{d}t = 0.$$

Condition $(\sup Y'.2)$ is obvious.

Since here we have the same functions $I^W(x)$ and $\zeta_\theta(x)$ as for the 'signal plus white noise' model, the statistical problems of Section 5.3 are solved in the same way.

5.5. Non-Gaussian regression

We consider the regression model (5.17) but now assume that independent and identically distributed errors $\xi_{k,n}$ have a distribution P on the real line with a probability density function $p(x)$ with respect to Lebesgue measure. An unknown regression function $\theta(\cdot)$ is again assumed to be continuous, so $\Theta \subset C[0, 1]$.

Next, we assume that the density $p(x)$ obeys the following condition, cf. conditions $(R.1)$ and $(R.2)$ for the model of an independent and identically distributed sample:

(P) *the density $p(x)$ is positive and continuous, and the function*

$$H_\gamma(s) = \int_{\mathbb{R}} p^\gamma(x) p^{1-\gamma}(x-s)\, \mathrm{d}x$$

is bounded over s from bounded domains for all $\gamma \in \mathbb{R}$.

As above, for a regression function $\theta(\cdot)$, we denote by $P_{n,\theta}$ the distribution of $X_n = (X_{1,n}, \ldots, X_{n,n})$. We have, with $P_n = P_{n,0}$,

$$\Xi_{n,\theta} = \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n) = \frac{1}{n}\sum_{k=1}^{n} \log \frac{p(X_{k,n} - \theta(k/n))}{p(X_{k,n})}.$$

As in the case of an independent and identically distributed sample, this representation suggests taking for Y_n the empirical process $F_n = F_n(x, t)$, $x \in \mathbb{R}$, $t \in [0, 1]$, defined by $F_n(x, 0) = 0$ and

$$F_n(x, t) = \frac{1}{n}\sum_{k=1}^{[nt]} 1(X_{k,n} \leqslant x), \qquad 0 < t \leqslant 1. \tag{5.20}$$

Then

$$\Xi_{n,\theta} = \int_0^1 \int_{\mathbb{R}} \log \frac{p(x-\theta(t))}{p(x)} F_n(\mathrm{d}x,\, \mathrm{d}t). \tag{5.21}$$

We define $\mathscr{Y}$ as the space of cumulative distribution functions $F = F(x,\, t)$, $x \in \mathbb{R}$, $t \in [0,\, 1]$, on $\mathbb{R} \times [0,\, 1]$ with the weak topology. Let $\mathscr{Y}_0$ be the subset of $\mathscr{Y}$ of absolutely continuous functions $F(x,\, t)$ with respect to Lebesgue measure on $\mathbb{R} \times [0,\, 1]$ and with densities $p_t(x)$ satisfying the condition $\int_{\mathbb{R}} p_t(x)\,\mathrm{d}x = 1$, $t \in [0,\, 1]$. As follows from Dembo and Zajic (1995) or Theorem 1 of Puhalskii (1996), the sequence $\{\mathscr{L}(F_n|P_n),\, n \geqslant 1\}$ obeys the LDP in $\mathscr{Y}$ with rate function $I^{SK}(F)$ given by

$$I^{SK}(F) = \begin{cases} \displaystyle\int_0^1 \int_{\mathbb{R}} \log \frac{p_t(x)}{p(x)}\, p_t(x)\,\mathrm{d}x\,\mathrm{d}t, & \text{if } F \in \mathscr{Y}_0, \\ \infty, & \text{otherwise.} \end{cases}$$

This verifies $(Y'.1)$.

To define $\zeta_{\theta,\delta}(F)$, introduce the functions

$$L_\theta(x,\, t) = \log \frac{p(x-\theta(t))}{p(x)},$$

$$L_{\theta,\delta}(x,\, t) = L_\theta(x,\, t) \vee (-\delta^{-1}) \wedge \delta^{-1}, \qquad x \in \mathbb{R},\, t \in [0,\, 1].$$

The functions $L_{\theta,\delta}$ are bounded, continuous and, in view of (P), satisfy the relations

$$\lim_{\delta\to 0} \int_0^1 \int_{\mathbb{R}} [\exp(\gamma|L_\theta(x,\, t) - L_{\theta,\delta}(x,\, t)|) - 1]\, p(x)\,\mathrm{d}x\,\mathrm{d}t = 0, \qquad \gamma > 0, \tag{5.22}$$

and, for every $\gamma > 0$,

$$\overline{\lim_{n\to\infty}} \int_0^1 \int_{\mathbb{R}} \left[\exp\left(\gamma \left| L_\theta\left(x,\, \frac{[nt]}{n}\right) - L_{\theta,\delta}\left(x,\, \frac{[nt]}{n}\right) \right| \right) - 1 \right] p(x)\,\mathrm{d}x\,\mathrm{d}t \to 0 \tag{5.23}$$

as $\delta \to 0$. We set

$$\zeta_{\theta,\delta}(F) = \int_0^1 \int_{\mathbb{R}} L_{\theta,\delta}(x,\, t) F(\mathrm{d}x,\, \mathrm{d}t). \tag{5.24}$$

Then condition $(Y'.2)$ holds by the definition of the topology on $\mathscr{Y}$ and choice of the $L_{\theta,\delta}$.

For condition $(Y'.3)$, write, for $\gamma > 0$, using Chebyshev's inequality, and (5.20), (5.21) and (5.24),

$$\begin{aligned} &\frac{1}{n} \log P_n(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon) \\ &\qquad \leqslant \frac{1}{n} \log P_n\left(\int_0^1 \int_{\mathbb{R}} |L_\theta(x,\, t) - L_{\theta,\delta}(x,\, t)| F_n(\mathrm{d}x,\, \mathrm{d}t) > \varepsilon \right) \\ &\qquad \leqslant -\gamma\varepsilon + \frac{1}{n} \sum_{k=1}^n \log \int_{\mathbb{R}} \exp(\gamma|L_\theta(x,\, k/n) - L_{\theta,\delta}(x,\, k/n)|)\, p(x)\,\mathrm{d}x. \end{aligned}$$

Limit (5.23) yields

$$\overline{\lim_{\delta\to 0}}\ \overline{\lim_{n\to\infty}}\ \frac{1}{n}\log P_n(|\varXi_{n,\theta}-\zeta_{\theta,\delta}(F_n)|>\varepsilon)\leqslant -\gamma\varepsilon,$$

which proves $(Y'.3)$ since γ is arbitrary.

For condition $(Y'.4)$, we take

$$\zeta_\theta(F)=\begin{cases}\int_0^1\int_{\mathbb{R}} L_\theta(x,\,t)F(\mathrm{d}x,\,\mathrm{d}t), & \text{if } I^{SK}(F)<\infty,\\ 0, & \text{otherwise.}\end{cases}$$

The ζ_θ are well defined since, by the Young–Fenchel inequality, if $F(x,\,t)=\int_0^t\int_{-\infty}^x p_s(y)\,\mathrm{d}y\,\mathrm{d}s$ then

$$\begin{aligned}\int_0^1\int_{\mathbb{R}}|L_\theta(x,\,t)|\frac{p_t(x)}{p(x)}p(x)\,\mathrm{d}x\,\mathrm{d}t &\leqslant \int_0^1\int_{\mathbb{R}}[\exp(|L_\theta(x,\,t)|)-1]p(x)\,\mathrm{d}x\,\mathrm{d}t\\ &\quad+\int_0^1\int_{\mathbb{R}}\left(\frac{p_t(x)}{p(x)}\log\frac{p_t(x)}{p(x)}-\frac{p_t(x)}{p(x)}+1\right)p(x)\,\mathrm{d}x\,\mathrm{d}t\\ &\leqslant 1+\int_0^1\int_{\mathbb{R}}p^2(x)(p(x-\theta(t)))^{-1}\,\mathrm{d}x\,\mathrm{d}t+I^{SK}(F),\end{aligned}$$

which is finite when $I^{SK}(F)<\infty$ by condition (P).

Next, once again by the Young–Fenchel inequality, we have, for $\gamma>0$,

$$\begin{aligned}\gamma|\zeta_{\theta,\delta}(F)-\zeta_\theta(F)| &\leqslant \int_0^1\int_{\mathbb{R}}\gamma|L_{\theta,\delta}(x,\,t)-L_\theta(x,\,t)|F(\mathrm{d}x,\,\mathrm{d}t)\\ &\leqslant \int_0^1\int_{\mathbb{R}}[\exp(\gamma|L_{\theta,\delta}(x,\,t)-L_\theta(x,\,t)|)-1]p(x)\,\mathrm{d}x\,\mathrm{d}t+I^{SK}(F),\end{aligned}$$

so by (5.22)

$$\overline{\lim_{\delta\to 0}}\sup_{F\in\Phi'_{I^{SK}}(a)}|\zeta_{\theta,\delta}(F)-\zeta_\theta(F)|\leqslant\frac{a}{\gamma},$$

which proves $(Y'.4)$ since γ is arbitrary.

Condition (U') is checked as in the case of an independent and identically distributed sample with the use of condition (P).

We now check conditions $(\sup Y')$ and $(\sup U')$. For this purpose, we assume that the $\theta(\cdot)$ are again from the set $\Sigma_K(\beta,\,M)$ defined in Section 5.3. Then limits (5.22) and (5.23) hold uniformly over $\theta\in\Sigma_K(\beta,\,M)$, which allows us to check $(\sup Y'.3)$, $(\sup Y'.4)$ and $(\sup U')$ analogously to $(Y'.3)$, $(Y'.4)$ and (U'), respectively. Condition $(\sup Y'.2)$ follows from the fact that the $L_{\theta,\delta}(x,\,t)$, $\theta\in\Sigma_K(\beta,\,M)$, are equicontinuous at each $(x,\,t)$ and uniformly bounded, so the $(\zeta_{\theta,\delta},\,\theta\in\Theta)$: $\mathscr{Y}\to\mathbb{R}_+^\Theta$ are continuous for the uniform topology on $\mathbb{R}_+^\Theta$.

We now calculate the function $S(\theta, \theta')$, $\theta, \theta' \in \Theta$. This is carried out with the use of a generalization of Chernoff's result in Lemma 5.2 which we state and prove next. Let $\mathbb{E}$ be a Polish space with the Borel σ-field $\mathscr{E}$ and $\mathscr{P}(\mathbb{E})$, the space of probability measures on $(\mathbb{E}, \mathscr{E})$. As above, for $F, P \in \mathscr{P}(\mathbb{E})$, we denote by $K(F, P)$ the Kullback–Leibler information:

$$K(F, P) = \begin{cases} \int_{\mathbb{E}} \log \frac{\mathrm{d}F}{\mathrm{d}P}(x) F(\mathrm{d}x), & \text{if } F \ll P, \\ \infty, & \text{otherwise.} \end{cases}$$

Recall that $K(F, P)$, for P fixed, is convex and is a rate function in F for the weak topology on $\mathscr{P}(\mathbb{E})$ (Deuschel and Stroock 1989, Section 3.2.17).

If the role of $\mathbb{E}$ is taken over by $\mathbb{E} \times [0, 1]$ with the product topology, then given a probability Borel measure ν on $[0, 1]$, we denote by $\mathscr{P}_\nu(\mathbb{E} \times [0, 1])$ the subset of $\mathscr{P}(\mathbb{E} \times [0, 1])$ of measures F such that $F(\mathbb{E} \times [0, t]) = \nu([0, t])$, $t \in [0, 1]$.

Our version of Chernoff's result is the following lemma.

Lemma 5.5. *Let $\mathbb{E}$ be a Polish space. Let probability measures $P, Q \in \mathscr{P}(\mathbb{E} \times [0, 1])$ be dominated by the product measure $\mu \times \nu$, where μ and ν are Borel measures on $\mathbb{E}$ and $[0, 1]$, respectively, with $\nu([0, 1]) = 1$. Then*

$$\inf_{F \in \mathscr{P}_\nu(\mathbb{E} \times [0,1])} \max\{K(F, P), K(F, Q)\} = -\inf_{\gamma \in [0,1]} \int_0^1 \log\left[\int_{\mathbb{E}} p_t^\gamma(x) q_t^{1-\gamma}(x) \mu(\mathrm{d}x)\right] \nu(\mathrm{d}t),$$

where $p_t(x)$ and $q_t(x)$ are the respective densities of P and Q relative to $\mu \times \nu$.

Proof. Obviously,

$$\max\{K(F, P), K(F, Q)\} = \sup_{\gamma \in [0,1]} (\gamma K(F, P) + (1 - \gamma) K(F, Q)). \tag{5.25}$$

Let $\mathscr{P}(\mathbb{E} \times [0, 1])$ be endowed with the weak topology. Since $K(F, P)$ is convex and is a rate function in F, we deduce that the function $\gamma K(F, P) + (1 - \gamma) K(F, Q)$, $\gamma \in [0, 1]$, $F \in \mathscr{P}_\nu(\mathbb{E} \times [0, 1])$, meets the conditions of a minimax theorem (see, e.g., Aubin and Ekeland 1984, Theorem 7, Section 2, Chapter 6). Hence,

$$\inf_{F \in \mathscr{P}_\nu(\mathbb{E} \times [0,1])} \sup_{\gamma \in [0,1]} (\gamma K(F, P) + (1 - \gamma) K(F, Q)) = \sup_{\gamma \in [0,1]} \inf_{F \in \mathscr{P}_\nu(\mathbb{E} \times [0,1])} (\gamma K(F, P) + (1 - \gamma) K(F, Q)). \tag{5.26}$$

The latter infimum can equivalently be taken over F dominated by P and Q, and hence by $\mu \times \nu$. Denote by $f_t(x)$ the density of F with respect to $\mu \times \nu$. Since, by the definition of $\mathscr{P}_\nu(\mathbb{E} \times [0, 1])$,

$$F(\mathbb{E} \times [0, t]) = \int_0^t \int_{\mathbb{E}} f_t(x) \mu(\mathrm{d}x) \nu(\mathrm{d}t) = \nu([0, t]), \qquad t \in [0, 1],$$

we have that

$$\int_{\mathbb{E}} f_t(x)\mu(\mathrm{d}x) = 1 \qquad \nu\text{-almost everywhere.} \tag{5.27}$$

Next, by the definition of the Kullback–Leibler information,

$$\gamma K(F, P) + (1-\gamma)K(F, Q) = \int_0^1 \int_{\mathbb{E}} \log \frac{f_t(x)}{p_t^{\gamma}(x) q_t^{1-\gamma}(x)} f_t(x)\mu(\mathrm{d}x)\nu(\mathrm{d}t), \tag{5.28}$$

where $0/0 = 0$, $0 \log 0 = 0$. Since the function $x \log x$, $x \geqslant 0$, is convex, an application of Jensen's inequality and (5.27) gives that ν-almost everywhere in $t \in [0, 1]$

$$\int_{\mathbb{E}} \log \frac{f_t(x)}{p_t^{\gamma}(x) q_t^{1-\gamma}(x)} f_t(x)\mu(\mathrm{d}x) \geqslant -\log \int_{\mathbb{E}} p_t^{\gamma}(x) q_t^{1-\gamma}(x)\mu(\mathrm{d}x).$$

On the other hand, taking

$$f_t(x) = p_t^{\gamma}(x) q_t^{1-\gamma}(x) \left(\int_{\mathbb{E}} p_t^{\gamma}(x) q_t^{1-\gamma}(x)\mu(\mathrm{d}x) \right)^{-1}, \tag{5.29}$$

we get equality above. Since the measure F with the density defined by (5.29) belongs to $\mathscr{P}_\nu(\mathbb{E} \times [0, 1])$, we obtain by (5.28) that

$$\inf_{F \in \mathscr{P}_\nu(\mathbb{E}\times[0,1])} [\gamma k(F, P) + (1-\gamma)K(F, Q)] = -\int_0^1 \log \left[\int_{\mathbb{E}} p_t^{\gamma}(x) q_t^{1-\gamma}(x)\mu(\mathrm{d}x) \right] \nu(\mathrm{d}t)$$

which, by (5.25) and (5.26), concludes the proof. □

Remark 5.6. Chernoff's result follows when ν is a Dirac measure.

We now apply Lemma 5.5 to evaluating the function $S(\theta, \theta')$.

Lemma 5.6. *For all θ, $\theta' \in \Theta$,*

$$S(\theta, \theta') = \inf_{\gamma \in [0,1]} \int_0^1 \log H_\gamma(\theta'(t) - \theta(t))\,\mathrm{d}t.$$

Proof. We have, for $F \in \mathscr{Y}_0$ with $I^{SK}(F) < \infty$,

$$\zeta_\theta(F) - I^{SK}(F) = -K(F, \bar{P}_\theta),$$

where $\bar{P}_\theta(\mathrm{d}x, \mathrm{d}t) = p(x - \theta(t))\,\mathrm{d}x\,\mathrm{d}t$, and the claim follows by (4.7) and Lemma 5.5 with $\mathbb{E} = \mathbb{R}$, $\mu(\mathrm{d}x) = \mathrm{d}x$, $\nu(\mathrm{d}t) = \mathrm{d}t$, $P = \bar{P}_\theta$ and $Q = \bar{P}_{\theta'}$. □

The latter result enables us to calculate asymptotic minimax risks for various statistical problems. To compare with the Gaussian case, let us consider the same statistical problems as in Sections 5.3 and 5.4 dealing with the value of $\theta(t_0)$ for a given t_0. Sets $\Sigma(\beta, M)$ and $\Sigma_K(\beta, M)$ are defined above.

5.5.1. Testing $\theta(t_0) = 0$ versus $|\theta(t_0)| \geqslant 2c$

Given $c > 0$, let $\Theta_0 = \{\theta \in \Theta\colon \theta(t_0) = 0\}$, $\Theta_1 = \{\theta \in \Theta\colon |\theta(t_0)| \geqslant 2c\}$ and define the risk $R_n^T(\rho_n)$ of a test ρ_n by (4.4). Recall that t^* was defined in (5.15).

Proposition 5.6. *Let c, β, M, K and t_0 be such that $[t_0 - t^*, t_0 + t^*] \subset [0, 1]$ and $K \geqslant 2c$. Let the measure P satisfy condition (P) and the function $H_\gamma(s)$ monotonically decrease in $s \geqslant 0$ for each $\gamma \in [0, 1]$. If $\Theta = \Sigma(\beta, M)$ then*

$$\underline{\lim}_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) \geqslant \inf_{\gamma\in[0,1]} 2\int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\,\mathrm{d}t.$$

If $\Theta = \Sigma_K(\beta, M)$ then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) = \inf_{\gamma\in[0,1]} 2\int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\,\mathrm{d}t,$$

and the tests $\rho_{n,\delta}^T$ from (4.8) are nearly LD efficient, i.e.,

$$\lim_{\delta\to 0} \overline{\lim}_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = \lim_{\delta\to 0} \underline{\lim}_{n\to\infty} R_n^T(\rho_{n,\delta}^T)$$

$$= \inf_{\gamma\in[0,1]} 2\int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\,\mathrm{d}t.$$

Proof. By Theorem 4.3 we need only to evaluate T^* from (4.6). A straightforward calculation using Lemma 5.6 and the monotonicity of $H_\gamma(s)$ shows that

$$T^* := \sup_{\theta\in\Theta_0, \theta'\in\Theta_1} S(\theta, \theta') = \inf_{\gamma\in[0,1]} 2\int_0^1 \log H_\gamma(2\theta^*(t))\,\mathrm{d}t,$$

where $\theta^*(t) = [c - M|t - t_0|^\beta]^+$. The claim follows. □

5.5.2. Estimating $\theta(t_0)$

For the problem of estimating $\theta(t_0)$, the risk of an estimator ρ_n is defined by

$$R_n^F(\rho_n) = \sup_{\theta\in\Theta} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta(t_0)| > c).$$

Proposition 5.7. *Let the conditions of Proposition 5.6 hold. If $\Theta = \Sigma(\beta, M)$ then*

$$\underline{\lim}_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) \geqslant \inf_{\gamma\in[0,1]} 2\int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\,\mathrm{d}t.$$

If $\Theta = \Sigma_K(\beta, M)$ then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) = \inf_{\gamma\in[0,1]} 2\int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\,\mathrm{d}t,$$

and the interval-median estimators $\rho_{n,\delta}^F$ from (4.18) are nearly LD efficient, i.e.,

$$\lim_{\delta\to 0}\overline{\lim_{n\to\infty}}\ R_n^F(\rho_{n,\delta}^F) = \lim_{\delta\to 0}\underline{\lim_{n\to\infty}}\ R_n^F(\rho_{n,\delta}^F)$$
$$= \inf_{\gamma\in[0,1]} 2\int_0^{t^*} \log H_\gamma(2(c - Mt^\beta))\,\mathrm{d}t.$$

Proof. By Theorem 4.5 and Remark 4.3 it suffices to calculate the asymptotic minimax risk given by Lemma 4.3:

$$F^* = \sup_{\theta,\theta'\in\Theta:|\theta(t_0)-\theta'(t_0)|>2c} S(\theta, \theta')$$

which is done as for the 'signal plus white noise' model. □

Remark 5.7. The latter problem of estimating $\theta(t_0)$ has been considered by Korostelev and Spokoiny (1996) under the assumption that $\log p(x)$ is concave upwards, and by Korostelev and Leonov (1995), who study the double asymptotics as $n \to \infty$ and then $c \to 0$.

5.6. The change-point model

Let us observe a sample $X_n = (X_{1,n}, \ldots, X_{n,n})$ of real-valued random variables, where, for some $k_n \geqslant 1$, the observations $X_{1,n}, \ldots, X_{k_n,n}$ are independent and identically distributed with a distribution P_0 and the observations $X_{k_n+1,n}, \ldots, X_{n,n}$ are independent and identically distributed with a distribution P_1. We assume that P_0 and P_1 are known and k_n is unknown. Let us also assume that $k_n = [n\theta]$, where $\theta \in \Theta = [0, 1]$. For this model, $\Omega_n = \mathbb{R}^n$ and $P_{n,\theta}$ denotes the distribution of X_n for θ.

Let a probability measure P dominate P_0 and P_1, and let

$$f_0(x) = \frac{\mathrm{d}P_0}{\mathrm{d}P}(x),\ f_1(x) = \frac{\mathrm{d}P_1}{\mathrm{d}P}(x), \qquad x \in \mathbb{R},$$

be the respective densities. We assume that $f_0(x)$ and $f_1(x)$ are positive and continuous and

$$\int_{\mathbb{R}} f_0^\gamma(x)P(\mathrm{d}x) < \infty,\ \int_{\mathbb{R}} f_1^\gamma(x)P(\mathrm{d}x) < \infty \qquad \text{for all } \gamma \in \mathbb{R}. \tag{5.30}$$

Introducing $P_n = P^n$, we have

$$\Xi_{n,\theta} = \frac{1}{n}\log\frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n) = \frac{1}{n}\sum_{i=1}^{[n\theta]}\log f_0(X_{i,n}) + \frac{1}{n}\sum_{i=[n\theta]+1}^{n}\log f_1(X_{i,n}),$$

so that, defining an empirical process by

$$F_n(x, t) = \frac{1}{n}\sum_{i=1}^{[nt]} 1(X_{i,n} \leqslant x), \qquad x \in \mathbb{R},\ t \in [0, 1],$$

we obtain the representation

$$\Xi_{n,\theta} = \int_0^\theta\int_{\mathbb{R}} \log f_0(x)F_n(\mathrm{d}x, \mathrm{d}t) + \int_\theta^1\int_{\mathbb{R}} \log f_1(x)F_n(\mathrm{d}x, \mathrm{d}t).$$

We define statistics Y_n and a space $\mathscr{Y}$ as for the non-Gaussian regression model. Let $\mathscr{Y}_P$ consist of the functions $F \in \mathscr{Y}$ that are absolutely continuous with respect to the measure $P(\mathrm{d}x) \times \mathrm{d}t$ with densities $p_t(x)$ such that $\int_{\mathbb{R}} p_t(x)P(\mathrm{d}x) = 1$, $t \geqslant 0$. As for the non-Gaussian regression model, condition $(Y'.1)$ holds with

$$I_P^{SK}(F) = \begin{cases} \int_0^1 \int_{\mathbb{R}} p_t(x) \log p_t(x) P(\mathrm{d}x)\, \mathrm{d}t, & \text{if } F \in \mathscr{Y}_P, \\ \infty, & \text{otherwise.} \end{cases}$$

We next take, for $F(\cdot, \cdot) \in \mathscr{Y}$,

$$\zeta_{\theta,\delta}(F) = \int_0^1 \int_{\mathbb{R}} L_{0,\delta}(x) g_\delta(\theta - t) F(\mathrm{d}x, \mathrm{d}t) + \int_0^1 \int_{\mathbb{R}} L_{1,\delta}(x) g_\delta(t - \theta) F(\mathrm{d}x, \mathrm{d}t),$$

where

$$L_{i,\delta}(x) = \log f_i(x) \wedge \delta^{-1} \vee (-\delta^{-1}), \qquad i = 0, 1,$$

$$g_\delta(t) = 0 \vee (\tfrac{1}{2} + \delta^{-2} t) \wedge 1.$$

The functions $L_{i,\delta}$ and g_δ are bounded, continuous and

$$\lim_{\delta \to 0} \int_{\mathbb{R}} [\exp(\gamma |\log f_i(x) - L_{i,\delta}(x)|) - 1] P(\mathrm{d}x) = 0, \qquad i = 0, 1, \gamma > 0. \tag{5.31}$$

The $\zeta_{\theta,\delta}$ are easily seen to be continuous, so $(Y'.2)$ holds.

For $(Y'.3)$, write, by Chebyshev's inequality, for $\gamma > 0$, $\varepsilon > 0$,

$$\begin{aligned} P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon) &\leqslant P_n^{1/n}\left(\int_0^1 \int_{\mathbb{R}} |\log f_0(x) - L_{0,\delta}(x)| F_n(\mathrm{d}x, \mathrm{d}t) + 2\delta > \frac{\varepsilon}{2} \right) \\ &\quad + P_n^{1/n}\left(\int_0^1 \int_{\mathbb{R}} |\log f_1(x) - L_{1,\delta}(x)| F_n(\mathrm{d}x, \mathrm{d}t) + 2\delta > \frac{\varepsilon}{2} \right) \\ &\leqslant \exp(-\gamma\varepsilon/2) \exp(2\gamma\delta)[\mathrm{E} \exp(\gamma |\log f_0(X_{1,n}) - L_{0,\delta}(X_{1,n})|) \\ &\quad + \mathrm{E} \exp(\gamma |\log f_1(X_{1,n}) - L_{1,\delta}(X_{1,n})|)], \end{aligned}$$

so

$$\begin{aligned} \overline{\lim_{n\to\infty}} P_n^{1/n}(|\Xi_{n,\theta} - \zeta_{\theta,\delta}(F_n)| > \varepsilon) &\leqslant \exp(-\gamma\varepsilon/2) \exp(2\gamma\delta) \left(\int_{\mathbb{R}} \exp(\gamma |\log f_0(x) - L_{0,\delta}(x)|) P(\mathrm{d}x) \right. \\ &\quad \left. + \int_{\mathbb{R}} \exp(\gamma |\log f_1(x) - L_{1,\delta}(x)|) P(\mathrm{d}x) \right), \end{aligned}$$

and, by (5.31), this goes to $2\exp(-\gamma\varepsilon/2)$ as $\delta \to 0$. Since γ is arbitrary, condition $(Y'.3)$ is verified.

To check $(Y'.4)$, we take

$$\zeta_\theta(F) = \begin{cases} \int_0^\theta \int_{\mathbb{R}} \log f_0(x) F(\mathrm{d}x,\, \mathrm{d}t) + \int_\theta^1 \int_{\mathbb{R}} \log f_1(x) F(\mathrm{d}x,\, \mathrm{d}t), & \text{if } I_P^{SK}(F) < \infty, \\ 0, & \text{otherwise.} \end{cases}$$

The facts that the ζ_θ are well defined and $(Y'.4)$ holds are proved as for the non-Gaussian regression model with the use of (5.30). Condition (U') also is easily checked.

Remark 5.8. The continuity condition on $f_0(x)$ and $f_1(x)$ can be omitted. One should then choose the $L_{i,\delta}$ bounded, continuous and satisfying (5.31).

Next, the argument used for (Y') and (U') checks also conditions $(\sup Y')$ and $(\sup U')$ (the verification of $(\sup Y'.2)$ uses the fact that the function $g_\delta(t-\theta)$ is equicontinuous for $\theta \in [0, 1]$ at each $t \in [0, 1]$).

The next step is evaluating $S(\theta, \theta')$ for $\theta, \theta' \in [0, 1]$.

Lemma 5.7. *For all $\theta, \theta' \in [0, 1]$,*

$$S(\theta, \theta') = -|\theta - \theta'| C(P_0, P_1).$$

Proof. In a manner similar to the case of non-Gaussian regression, we have, for any $F \in \mathscr{Y}_P$, $I_P^{SK}(F) < \infty$, with $F(\mathrm{d}x, \mathrm{d}t) = p_t(x) P(\mathrm{d}x)\, \mathrm{d}t$,

$$\zeta_\theta(F) - I_P^{SK}(F) = -\int_0^\theta \int_{\mathbb{R}} \log \frac{p_t(x)}{p_0(x)} p_t(x) P(\mathrm{d}x)\, \mathrm{d}t$$
$$- \int_\theta^1 \int_{\mathbb{R}} \log \frac{p_t(x)}{p_1(x)} p_t(x) P(\mathrm{d}x)\, \mathrm{d}t = -K(F, \bar{P}_\theta),$$

where $\bar{P}_\theta(\mathrm{d}x, \mathrm{d}t) = (f_0(x) 1(t \leqslant \theta) + f_1(x) 1(t > \theta)) P(\mathrm{d}x)\, \mathrm{d}t$. The claim follows by (4.7), Lemma 5.5 with $\mathbb{E} = \mathbb{R}$, $\mu(\mathrm{d}x) = P(\mathrm{d}x)$, $\nu(\mathrm{d}t) = \mathrm{d}t$, $P = \bar{P}_\theta$ and $Q = \bar{P}_{\theta'}$ and the definition of Chernoff's function in Lemma 5.2. □

We apply this result and the general theorems from Section 4 to the problem of estimating the parameter θ. The risk of an estimator ρ_n is defined in a standard way:

$$R_n^F(\rho_n) = \sup_{\theta \in [0,1]} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta| > c). \tag{5.32}$$

Proposition 5.8. *For each $c < 1/2$,*

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) = -2cC(P_0, P_1).$$

If $\rho_{n,\delta}^F$ are the interval-median estimators from (4.18) then

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}} R_n^F(\rho_{n,\delta}^F) = \lim_{\delta\to 0} \underline{\lim_{n\to\infty}} R_n^F(\rho_{n,\delta}^F) = -2cC(P_0, P_1).$$

Proof. We apply Theorem 4.5. One needs only to calculate the minimax risk F^*. Using Lemmas 4.3 and 5.6, we obtain

$$F^* = \sup_{\theta,\theta':|\theta-\theta'|>2c} S(\theta, \theta') = -2cC(P_0, P_1). \qquad \square$$

Remark 5.9. The same result has been obtained by Korostelev (1995), who uses another kind of an upper estimator. The construction is based on considering the concave hull of a sample path of the likelihood process. By Lemma 4.2 this estimator is a particular case of the interval-median estimators $\rho^F_{n,\delta}$.

5.7. Regression with random design

We consider the model

$$X_{k,n} = \theta(t_{k,n}) + \xi_{k,n}, \qquad k = 1, \ldots, n, \tag{5.33}$$

where real-valued errors $\xi_{k,n}$ are independent with a common distribution P having a density $p(x)$ that obeys condition (P) of Section 5.5, and design points $t_{k,n}$ are real-valued independent random variables with a common distribution Π and are independent of the $\xi_{k,n}$. We impose a standard condition on the design measure Π.

(Π) *The measure* Π *is compactly supported and has a positive density with respect to Lebesgue measure on the support.*

We denote the support by D. An unknown regression function $\theta(\cdot)$ is assumed to be continuous. In this model, $P_{n,\theta}$ is the joint distribution of $X_n = (X_{1,n}, \ldots, X_{n,n})$ and $t_n = (t_{1,n}, \ldots, t_{n,n})$ for θ.

Let us take for Y_n the joint empirical distribution function F_n of X_n and t_n:

$$F_n(A, B) = \frac{1}{n}\sum_{k=1}^{n} 1(X_{k,n} \in A,\ t_{k,n} \in B) \tag{5.34}$$

for Borel sets $A \subset \mathbb{R}$, $B \subset D$. We take $\mathscr{Y}$ to be the space of distributions on $\mathbb{R} \times D$ with the weak topology. Let also $P_n = P_{n,0} = (P \times \Pi)^n$.

With these definitions,

$$\begin{aligned} \Xi_{n,\theta} &= \frac{1}{n} \log \frac{\mathrm{d}P_{n,\theta}}{\mathrm{d}P_n}(X_n, t_n) \\ &= \frac{1}{n}\sum_{k=1}^{n} \log \frac{p(X_{k,n} - \theta(t_{k,n}))}{p(X_{k,n})} \\ &= \int_D \int_{\mathbb{R}} \log \frac{p(x-\theta(t))}{p(x)} F_n(\mathrm{d}x, \mathrm{d}t). \end{aligned}$$

Let $\mathscr{Y}_1$ be the subset of the set $\mathscr{Y}$ of the cumulative distribution functions on $\mathbb{R}^2$ that are absolutely continuous with respect to Lebesgue measure on $\mathbb{R}^2$ and have support in $\mathbb{R} \times D$.

Under P_n, the random pairs $(X_{k,n}, t_{k,n})$ are independent and identically distributed with the distribution $P \times \Pi$, and hence, by Sanov's theorem, the LDP holds for the F_n with rate function $I^{SS}(F)$ defined by

$$I^{SS}(F) = \begin{cases} \int_D \int_{\mathbb{R}} \log \dfrac{p(x, t)}{p(x)\pi(t)} p(x, t)\, \mathrm{d}x\, \mathrm{d}t, & \text{if } F \in \mathscr{Y}_1, \\ \infty, & \text{otherwise.} \end{cases}$$

Here $F(\mathrm{d}x, \mathrm{d}t) = p(x, t)\, \mathrm{d}x\, \mathrm{d}t$. This verifies $(Y'.1)$.

Next set, for $F \in \mathscr{Y}$,

$$\zeta_\theta(F) = \begin{cases} \int_D \int_{\mathbb{R}} \log \dfrac{p(x - \theta(t))}{p(x)} F(\mathrm{d}x, \mathrm{d}t), & \text{if } I^{SS}(F) < \infty, \\ 0, & \text{otherwise,} \end{cases}$$

$$\zeta_{\theta,\delta}(F) = \int_D \int_{\mathbb{R}} \left[\log \frac{p(x - \theta(t))}{p(x)} \right] \wedge \delta^{-1} \vee (-\delta^{-1}) F(\mathrm{d}x, \mathrm{d}t).$$

With this notation, the rest of condition (Y') and condition (U') are verified in analogy with the case of non-Gaussian regression. This proves the LDP for the model.

For conditions $(\sup Y')$ and $(\sup U')$, we again assume that $\theta \in \Sigma_K(\beta, M)$, where the set $\Sigma_K(\beta, M)$ was defined above. The conditions are then checked as for the non-Gaussian regression model.

We now calculate the function $S(\theta, \theta')$ from (4.7). Recall that the function $H_\gamma(s)$ was defined in condition (P).

Lemma 5.8. *Under conditions* (P) *and* (Π),

$$S(\theta, \theta') = \inf_{\gamma \in [0,1]} \log \int_D H_\gamma(\theta'(t) - \theta(t)) \pi(t)\, \mathrm{d}t.$$

Proof. Given $F \in \mathscr{Y}_1$ with $I^{SS}(F) < \infty$, we easily obtain

$$\zeta_\theta(F) - I^{SS}(F) = -K(F, \bar{P}_\theta),$$

where $\bar{P}_\theta(\mathrm{d}x, \mathrm{d}t) = p(x - \theta(t))\pi(t)\, \mathrm{d}x\, \mathrm{d}t$, and the claim follows by (4.7) and Lemma 5.2 with $\mathbb{E} = \mathbb{R} \times D$, $\mu(\mathrm{d}x, \mathrm{d}t) = \mathrm{d}x\, \mathrm{d}t$, $P = \bar{P}_\theta$ and $Q = \bar{P}_{\theta'}$. □

We now consider the same two statistical problems as in Section 5.5 and compare the results for the cases of random and deterministic design.

5.7.1. Testing $\theta(t_0) = 0$ versus $|\theta(t_0)| \geqslant 2c$

Given $t_0 \in D$ and $c > 0$, consider the hypothesis testing problem: $\theta(t_0) = 0$ versus $|\theta(t_0)| \geqslant 2c$. The risk $R_n^T(\rho_n)$ of a test ρ_n, as well as the sets $\Sigma(\beta, M)$ and $\Sigma_0(\beta, M)$, and t^* are defined as above.

Proposition 5.9. *Let $D = [0, 1]$. Let c, β, M, K and t_0 be such that $[t_0 - t^*, t_0 + t^*] \subset [0, 1]$ and $K \geqslant 2c$. Let conditions (P) and (Π) hold and the function $H_\gamma(s)$ monotonically decrease in $s \geqslant 0$ for each $\gamma \in [0, 1]$. If $\Theta = \Sigma(\beta, M)$ then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) \geqslant T^*,$$

where

$$T^* = \inf_{\gamma\in[0,1]} \log\left(1 + \int_{t_0-t^*}^{t_0+t^*} [H_\gamma(2(c - M|t - t_0|^\beta)) - 1]\pi(t)\,\mathrm{d}t\right).$$

If $\Theta = \Sigma_K(\beta, M)$ then

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^T(\rho_n) = T^*,$$

and the tests $\rho_{n,\delta}^T$ from (4.8) are nearly LD efficient, i.e.,

$$\lim_{\delta\to\infty} \varlimsup_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = \lim_{\delta\to 0} \varliminf_{n\to\infty} R_n^T(\rho_{n,\delta}^T) = T^*.$$

Proof. Theorem 4.3 reduces the proof to calculating T^* from (4.6) Using the result of Lemma 5.8 and proceeding in analogy with the case of deterministic design, we conclude that

$$\begin{aligned} T^* &= S(c - \theta^*, c + \theta^*) \\ &= \inf_{\gamma\in[0,1]} \log\left(\int_0^{t_0-t^*} \pi(t)\,\mathrm{d}t + \int_{t_0-t^*}^{t_0+t^*} H_\gamma(2(c - M|t - t_0|^\beta))\pi(t)\,\mathrm{d}t + \int_{t_0+t^*}^1 \pi(t)\,\mathrm{d}t\right). \end{aligned}$$

Now the claim follows by the equality $\int_D \pi(t)\,\mathrm{d}t = 1$. □

5.7.2. Estimating $\theta(t_0)$

As above, when estimating $\theta(t_0)$, we define the risk of an estimator ρ_n by

$$R_n^F(\rho_n) = \sup_{\theta\in\Sigma_K(\beta,M)} \frac{1}{n} \log P_{n,\theta}(|\rho_n - \theta(t_0)| > c).$$

Proposition 5.10. *Let the conditions of Proposition 5.9 hold. If $\Theta = \Sigma(\beta, M)$ then*

$$\varliminf_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) \geqslant F^*,$$

where

$$F^* = \inf_{\gamma\in[0,1]} \log\left(1 + \int_{t_0-t^*}^{t_0+t^*} [H_\gamma(2(c - M|t - t_0|^\beta)) - 1]\pi(t)\,\mathrm{d}t\right).$$

If $\Theta = \Sigma_K(\beta, M)$ *then*

$$\lim_{n\to\infty} \inf_{\rho_n} R_n^F(\rho_n) = F^*,$$

and the interval-median estimators $\rho_{n,\delta}^F$ *from (4.18) are nearly LD efficient, i.e.,*

$$\lim_{\delta\to 0} \overline{\lim_{n\to\infty}} R_n^F(\rho_{n,\delta}^F) = \lim_{\delta\to 0} \underline{\lim_{n\to\infty}} R_n^F(\rho_{n,\delta}^F) = F^*.$$

Proof. By Theorem 4.5 it suffices to calculate the asymptotic minimax risk F^* from Lemma 4.3, which is done in analogy with the proof of Proposition 5.9. □

Remark 5.10. If we consider the uniform random design on [0, 1], i.e., take $\pi(t) = 1$, Jensen's inequality easily implies that its asymptotic minimax risks are not greater than the corresponding risks for regression with deterministic design (see Section 5.5). This fact also follows from comparing Lemma 5.2 and Lemma 5.5.

Remark 5.11. The problem of estimating $\theta(t_0)$ for the uniform random design has been considered by Korostelev (1995), who studies the double asymptotics as $n \to \infty$ and then $c \to 0$.

Appendix

Proof of Lemma 2.4

Let $\{\mathbf{V}_\Lambda, \Lambda \in \mathscr{A}(\Theta)\}$ be a standard family of deviabilities so that for all $\Lambda \subset \Lambda' \in \mathscr{A}(\Theta)$ and $z_\Lambda \in S_\Lambda$,

$$\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_{\Lambda'}\in\Pi_{\Lambda'\Lambda}^{-1} z_\Lambda} \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}). \tag{A.1}$$

We define

$$\mathbf{V}_\Theta(z_\Theta) = \begin{cases} \inf_{\Lambda\in\mathscr{A}(\Theta)} \|\pi_\Lambda z_\Theta\|_\Lambda^{-1} \mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta), & z_\Theta \in S_\Theta, \\ 0, & \text{otherwise}, \end{cases} \tag{A.2}$$

where we set $\mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta) = 1$ and $\|\pi_\Lambda z_\Theta\|_\Lambda^{-1}\mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta) = \infty$ when $\|\pi_\Lambda z_\Theta\|_\Lambda = 0$.

The functions $\|\pi_\Lambda z_\Theta\|_\Lambda^{-1}\mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta)$, $\Lambda \in \mathscr{A}(\Theta)$, are easily seen to be upper semi-continuous on S_Θ, so $(\mathbf{V}_\Theta(z_\Theta), z_\Theta \in \mathbb{R}_+^\Theta)$ is upper semi-continuous as the infimum of a family of upper semi-continuous functions. Moreover, since, for every $z_\Theta \in S_\Theta$ and $\varepsilon > 0$, there exists $\Lambda \in \mathscr{A}(\Theta)$ such that $\|\pi_\Lambda z_\Theta\|_\Lambda > 1 - \varepsilon$, and since $\mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta) \leqslant 1$, we conclude that $\mathbf{V}_\Theta(z_\Theta) \leqslant 1$. Since (ii) obviously follows by (iii), we are left to prove (iii) and the equality

$$\sup_{z_\Theta \in S_\Theta} \mathbf{V}_\Theta(z_\Theta) = 1. \tag{A.3}$$

We begin with (iii). Let us fix Λ and z_Λ, assuming that $z_\Lambda \in S_\Lambda$. Definition (A.2) implies that

$$\mathbf{V}_\Lambda(z_\Lambda) \geqslant \sup_{z_\Theta \in \Pi_\Lambda^{-1} z_\Lambda} \|\pi_\Lambda z_\Theta\|_\Lambda \mathbf{V}_\Theta(z_\Theta),$$

so we need to prove that

$$\mathbf{V}_\Lambda(z_\Lambda) \leqslant \sup_{z_\Theta \in \Pi_\Lambda^{-1} z_\Lambda} \|\pi_\Lambda z_\Theta\|_\Lambda \mathbf{V}_\Theta(z_\Theta). \tag{A.4}$$

First, we note that (A.2) and (A.1) imply that

$$\mathbf{V}_\Theta(z_\Theta) = \inf_{\substack{\Lambda' \in \mathscr{A}(\Theta) \\ \Lambda' \supset \Lambda}} \|\pi_{\Lambda'} z_\Theta\|_{\Lambda'}^{-1} \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} z_\Theta), \qquad z_\Theta \in S_\Theta. \tag{A.5}$$

Indeed, by (A.1), if $\Lambda \subset \Lambda' \in \mathscr{A}(\Theta)$ and $z_\Theta \in S_\Theta$ is such that $\|\pi_\Lambda z_\Theta\|_\Lambda > 0$ then

$$\mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta) \geqslant \|\pi_{\Lambda'\Lambda} \Pi_{\Lambda'} z_\Theta\|_\Lambda \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} z_\Theta),$$

and hence, since $\pi_{\Lambda'\Lambda} \Pi_{\Lambda'} z_\Theta = \pi_\Lambda z_\Theta / \|\pi_{\Lambda'} z_\Theta\|_{\Lambda'}$,

$$\|\pi_{\Lambda'} z_\Theta\|_{\Lambda'}^{-1} \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} z_\Theta) \leqslant \|\pi_\Lambda z_\Theta\|_\Lambda^{-1} \mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta),$$

which, in view of (A.2), proves (A.5).

Next, we obviously can assume that $a := \mathbf{V}_\Lambda(z_\Lambda) > 0$. For $\Lambda' \supset \Lambda$, $\Lambda' \in \mathscr{A}(\Theta)$, introduce the sets

$$A_{\Lambda'} = \{z_{\Lambda'} \in S_{\Lambda'}\colon \Pi_{\Lambda'\Lambda} z_{\Lambda'} = z_\Lambda \text{ and } \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}) = a\}. \tag{A.6}$$

We show that $A_{\Lambda'}$ is non-empty. Since $\mathbf{V}_{\Lambda'}(z_{\Lambda'}) \leqslant 1$, the supremum on the right of (A.1) can equivalently be taken over the set $\Pi_{\Lambda'\Lambda}^{-1} z_\Lambda \cap \{\|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \geqslant a/2\}$. This set is closed since the projection $\Pi_{\Lambda'\Lambda}$ is continuous in restriction to the set $\{z_{\Lambda'}\colon \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \geqslant a/2\}$. Since $\mathbf{V}_{\Lambda'}$ is a deviability, it attains suprema on closed sets, so the supremum on the right of (A.1) is attained, which is equivalent to $A_{\Lambda'}$ being non-empty. Next, $A_{\Lambda'}$ is closed and hence compact since $\mathbf{V}_{\Lambda'}$ is upper semi-continuous and, by (A.1) and the definition of a, $\|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}) = a$ if and only if $\|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}) \geqslant a$.

Now we introduce for each $\Lambda' \in \mathscr{A}(\Theta)$, $\Lambda' \supset \Lambda$,

$$\mathbf{A}_{\Lambda'} = \{z_\Theta \in [0, 1]^\Theta\colon \Pi_{\Lambda'} z_\Theta \in A_{\Lambda'} \text{ and } \|\pi_{\Lambda'} z_\Theta\|_{\Lambda'} \geqslant a\}.$$

These sets are easily seen to be non-empty (e.g., if $z_{\Lambda'} \in A_{\Lambda'}$ then $z_\Theta = (z_\theta, \theta \in \Theta)$, defined by $(z_\theta, \theta \in \Lambda') = z_{\Lambda'}$ and $z_\theta = 0$, $\theta \notin \Lambda'$, belongs to $\mathbf{A}_{\Lambda'}$) and compact for the Tihonov topology on $[0, 1]^\Theta$ (the latter holds because $\Pi_{\Lambda'}$ is continuous in restriction to the set $\{z_\Theta\colon \|\pi_{\Lambda'} z_\Theta\|_{\Lambda'} \geqslant a\}$ and $A_{\Lambda'}$ is closed).

We next show that, for all elements Λ' and Λ'' of $\mathscr{A}(\Theta)$ containing Λ, the sets $\mathbf{A}_{\Lambda'}$ and

$\mathbf{A}_{\Lambda''}$ have a non-empty intersection. Indeed, let $\Lambda''' = \Lambda' \cup \Lambda''$ and $z_\Theta \in [0, 1]^\Theta$ be such that $z_\Theta \in \mathbf{A}_{\Lambda'''}$ and $\|\pi_{\Lambda'''} z_\Theta\| = 1$ (such a z_Θ obviously exists). We prove that $z_\Theta \in \mathbf{A}_{\Lambda'}$ and $z_\Theta \in \mathbf{A}_{\Lambda''}$.

Denote $z_{\Lambda'''} = \Pi_{\Lambda'''} z_\Theta$, $z_{\Lambda'} = \Pi_{\Lambda'} z_\Theta$, the latter being well defined since the definitions of $\mathbf{A}_{\Lambda'''}$ and $A_{\Lambda'''}$ imply that $\|\pi_\Lambda z_\Theta\|_\Lambda \geqslant a$. First, note that

$$\Pi_{\Lambda'\Lambda} z_{\Lambda'} = \Pi_\Lambda z_\Theta = \Pi_{\Lambda'''\Lambda} z_{\Lambda'''} = z_\Lambda, \tag{A.7}$$

where the last equality follows by the fact that $z_{\Lambda'''} \in A_{\Lambda'''}$. This and (A.1) yield, in view of the equality $\Pi_{\Lambda'''\Lambda'} z_{\Lambda'''} = z_{\Lambda'}$,

$$\mathbf{V}_\Lambda(z_\Lambda) \geqslant \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \mathbf{V}_{\Lambda'}(z_{\Lambda'}), \tag{A.8}$$

$$\mathbf{V}_{\Lambda'}(z_{\Lambda'}) \geqslant \|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'} \mathbf{V}_{\Lambda'''}(z_{\Lambda'''}). \tag{A.9}$$

Next, by the definitions of $z_{\Lambda'''}$ and $z_{\Lambda'}$,

$$\|\pi_{\Lambda'''\Lambda} z_{\Lambda'''}\|_\Lambda = \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \cdot \|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'},$$

so that, by (A.8) and (A.9),

$$\mathbf{V}_\Lambda(z_\Lambda) \geqslant \|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \cdot \|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'} \mathbf{V}_{\Lambda'''}(z_{\Lambda'''}) = \|\pi_{\Lambda'''\Lambda} z_{\Lambda'''}\|_\Lambda \mathbf{V}_{\Lambda'''}(z_{\Lambda'''}).$$

Since $z_{\Lambda'''} \in A_{\Lambda'''}$, we actually have equality here and hence in (A.8) and (A.9). (A.8) and (A.7) prove that $z_{\Lambda'} \in A_{\Lambda'}$. Equalities in (A.8) and (A.9) together imply, since $\mathbf{V}_{\Lambda'''}(z_{\Lambda'''}) \leqslant 1$ and $\|\pi_{\Lambda'\Lambda} z_{\Lambda'}\|_\Lambda \leqslant 1$, that $\|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'} \geqslant \mathbf{V}_{\Lambda'}(z_{\Lambda'}) \geqslant \mathbf{V}_\Lambda(z_\Lambda) = a$; since also $\|\pi_{\Lambda'''} z_\Theta\|_{\Lambda'''} = 1$, we obtain

$$\|\pi_{\Lambda'} z_\Theta\|_{\Lambda'} = \|\pi_{\Lambda'''} z_\Theta\|_{\Lambda'''} \cdot \|\pi_{\Lambda'''\Lambda'} z_{\Lambda'''}\|_{\Lambda'} \geqslant a.$$

This concludes the proof of the inclusion $z_\Theta \in \mathbf{A}_{\Lambda'}$. The inclusion $z_\Theta \in \mathbf{A}_{\Lambda''}$ is proved by the same argument.

Thus, finite intersections of the compact sets $\mathbf{A}_{\Lambda'}$, $\Lambda' \supset \Lambda$, are non-empty, hence $\cap_{\Lambda' \supset \Lambda} \mathbf{A}_{\Lambda'} \neq \varnothing$. Pick z_Θ from this intersection and let $\hat{z}_\Theta = z_\Theta / \|z_\Theta\|_\Theta$. We prove that

$$\Pi_\Lambda \hat{z}_\Theta = z_\Lambda \tag{A.10}$$

and

$$\mathbf{V}_\Theta(\hat{z}_\Theta) = \|\pi_\Lambda \hat{z}_\Theta\|_\Lambda^{-1} \mathbf{V}_\Lambda(z_\Lambda), \tag{A.11}$$

which yields (A.4) since $\hat{z}_\Theta \in S_\Theta$. Let $\Lambda' \in \mathscr{A}(\Theta)$ with $\Lambda \subset \Lambda'$. Since $\Pi_{\Lambda'} \hat{z}_\Theta = \Pi_{\Lambda'} z_\Theta \in A_{\Lambda'}$, it follows by the definition of $A_{\Lambda'}$ that $\Pi_\Lambda \hat{z}_\Theta = \Pi_{\Lambda'\Lambda} \Pi_{\Lambda'} \hat{z}_\Theta = z_\Lambda$ verifying (A.10); also

$$\mathbf{V}_\Lambda(z_\Lambda) = a = \|\pi_{\Lambda'\Lambda} \Pi_{\Lambda'} \hat{z}_\Theta\|_\Lambda \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} \hat{z}_\Theta) = \frac{\|\pi_\Lambda \hat{z}_\Theta\|_\Lambda}{\|\pi_{\Lambda'} \hat{z}_\Theta\|_{\Lambda'}} \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} \hat{z}_\Theta),$$

so

$$\|\pi_\Lambda \hat{z}_\Theta\|_\Lambda^{-1} \mathbf{V}_\Lambda(z_\Lambda) = \|\pi_{\Lambda'} \hat{z}_\Theta\|_{\Lambda'}^{-1} \mathbf{V}_{\Lambda'}(\Pi_{\Lambda'} \hat{z}_\Theta).$$

In view of (A.5), this implies (A.11), and (A.4) follows. Assertion (iii) has been proved.

Finally, according to (iii),

$$1 = \sup_{z_\Lambda \in S_\Lambda} \mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta \in S_\Theta} \|\pi_\Lambda z_\Theta\|_\Lambda \mathbf{V}_\Theta(z_\Theta) \leqslant \sup_{z_\Theta \in S_\Theta} \mathbf{V}_\Theta(z_\Theta),$$

proving (A.3). □

Remark A.1. Equality (A.5) shows that $\mathbf{V}_\Theta$ can equivalently be defined as

$$\mathbf{V}_\Theta(z_\Theta) = \lim_{\Lambda \in \mathscr{A}(\Theta)} \mathbf{V}_\Lambda(\Pi_\Lambda z_\Theta), \qquad z_\Theta \in S_\Theta,$$

where the limit is with respect to the partial ordering by inclusion: $\Lambda \leqslant \Lambda'$ if $\Lambda \subset \Lambda'$.

A minimax theorem for non-level-compact loss functions

This subsection contains a minimax theorem for generalized risks and non-level-compact loss functions. We assume the setting described at the beginning of Section 3 and start by introducing an extension of the decision space (cf. Strasser 1985).

Denote by $\mathscr{C}_+(\mathscr{D})$ the set of all non-negative bounded continuous functions on $\mathscr{D}$, and let $\mathbf{B}(\mathscr{D})$ be the set of all functionals b: $\mathscr{C}_+(\mathscr{D}) \to \mathbb{R}_+$ with the following properties:

(1) $b(\mathbf{0}) = 0$, $b(\mathbf{1}) = 1$, where $\mathbf{0}$ ($\mathbf{1}$) denotes the element of $\mathscr{C}_+(\mathscr{D})$ identically equal to 0 (1);
(2) $b(f) \leqslant b(g)$ if $f \leqslant g$, $f, g \in \mathscr{C}_+(\mathscr{D})$;
(3) $b(\lambda f) = \lambda b(f)$, $f \in \mathscr{C}_+(\mathscr{D})$, $\lambda \in \mathbb{R}_+$;
(4) $b(f + g) \leqslant b(f) + b(g)$, $f, g \in \mathscr{C}_+(\mathscr{D})$.

Also let $\mathbf{B}_1(\mathscr{D})$ be the subset of those $b \in \mathbf{B}(\mathscr{D})$ for which, in addition,

(5) $b(f \vee g) = b(f) \vee b(g)$, $f, g \in \mathscr{C}_+(\mathscr{D})$,

where $f \vee g$ denotes the maximum of f and g.

We endow $\mathbf{B}(\mathscr{D})$ with the weak topology which is the topology induced by the Tihonov (product) topology on $\mathbb{R}_+^{\mathscr{C}_+(\mathscr{D})}$, i.e., a net $\{b_\sigma, \sigma \in \Sigma\}$ of elements of $\mathbf{B}(\mathscr{D})$, where Σ is a directed set, converges to $b \in \mathbf{B}(\mathscr{D})$ if $\lim_{\sigma \in \Sigma} b_\sigma(f) = b(f)$ for all $f \in \mathscr{C}_+(\mathscr{D})$. Obviously, $\mathbf{B}(\mathscr{D})$ is closed in $\mathbb{R}_+^{\mathscr{C}_+(\mathscr{D})}$.

We extend the domain of the functionals b to the set $\underline{\mathscr{C}}_+(\mathscr{D})$ of lower semi-continuous non-negative functions on $\mathscr{D}$ by letting

$$b(g) = \sup\{b(f): f \leqslant g, f \in \mathscr{C}_+(\mathscr{D})\}, \qquad g \in \underline{\mathscr{C}}_+(\mathscr{D}). \tag{A.12}$$

It is easily seen that the map $b \to b(g)$ is lower semi-continuous on $\mathbf{B}(\mathscr{D})$ for each $g \in \underline{\mathscr{C}}_+(\mathscr{D})$.

Next, let us denote by $\mathscr{B}_n$ the set of all random elements on $(\Omega_n, \mathscr{F}_n)$ with values in $\mathbf{B}(\mathscr{D})$. We call the elements of $\mathscr{B}_n$ generalized decision functions (or generalized decisions). Given loss functions W_θ, $\theta \in \Theta$, which are lower semi-continuous by definition, and a generalized decision $\beta_n \in \mathscr{B}_n$, we define $\beta_n(W_\theta^n)$ according to (A.12), and define the LD risk $B_n(\beta_n)$ of a generalized decision $\beta_n \in \mathscr{B}_n$ in the experiment $\mathscr{E}_n = (\Omega_n, \mathscr{F}_n; P_{n,\theta}, \theta \in \Theta)$ by

$$B_n(\beta_n) = \sup_{\theta \in \Theta} \mathrm{E}_{n,\theta}^{1/n} \beta_n(W_\theta^n). \tag{A.13}$$

Theorem A.1. *Let $\{\mathscr{E}_n, n \geqslant 1\}$ satisfy the LDP. Then*

$$\varliminf_{n \to \infty} \inf_{\beta_n \in \mathscr{B}_n} B_n(\beta_n) \geqslant B^*,$$

where

$$B^* = \sup_{z_\Theta \in \mathbb{R}_+^\Theta} \inf_{b \in \mathbf{B}_1(\mathscr{D})} \sup_{\theta \in \Theta} b(W_\theta) z_\theta \mathbf{V}_\Theta(z_\Theta).$$

For a proof, we need to study properties of $\mathbf{B}(\mathscr{D})$ and $\mathbf{B}_1(\mathscr{D})$.

Lemma A.1. *Let $f_1, f_2, \ldots, f_k \in \mathscr{C}_+(\mathscr{D})$ and $\{b_n, n \geqslant 1\}$ be a sequence of elements of $\mathbf{B}(\mathscr{D})$. Then there exists $b \in \mathbf{B}_1(\mathscr{D})$ such that $b(f_i)$ is an accumulation point of the sequence $\{b_n^{1/n}(f_i^n), n \geqslant 1\}$ for $i = 1, \ldots, k$.*

Proof. Let $\|\cdot\|$ denote the uniform norm on $\mathscr{C}_+(\mathscr{D})$. Define $\mathscr{C}_{1,+}(\mathscr{D})$ as the subset of $\mathscr{C}_+(\mathscr{D})$ of functions f with $\|f\| \leqslant 1$. Introduce the functionals $\bar{b}_n(f) = b_n^{1/n}(f^n)$, $f \in \mathscr{C}_{1,+}(\mathscr{D})$. Then the set $B = \{\bar{b}_n, n \geqslant 1\}$ is contained in the set $[0, 1]^{\mathscr{C}_{1,+}(\mathscr{D})}$ By Tihonov's theorem, $[0, 1]^{\mathscr{C}_{1,+}(\mathscr{D})}$ with the product topology is compact, and hence B is relatively compact. Let $\tilde{b}$ denote any accumulation point. We extend $\tilde{b}$ to a functional on $\mathscr{C}_+(\mathscr{D})$ by letting $\tilde{b}(\lambda f) = \lambda \tilde{b}(f)$, $\lambda > 0$, $f \in \mathscr{C}_{1,+}(\mathscr{D})$. Since $b_n \in \mathbf{B}(\mathscr{D})$, it is easy to see that $\tilde{b} \in \mathbf{B}(\mathscr{D})$. Also, since the topology on $\mathbf{B}(\mathscr{D})$ is the restriction of the product topology on $\mathbb{R}_+^{\mathscr{C}_+(\mathscr{D})}$, it follows that $\tilde{b}$ is an accumulation point of $\{\bar{b}_n, n \geqslant 1\}$, where the $\bar{b}_n$ are extended to functionals on $\mathscr{C}_+(\mathscr{D})$ by letting $\bar{b}_n(\lambda f) = \lambda \bar{b}_n(f)$, $\lambda > 0$, $f \in \mathscr{C}_{1,+}(\mathscr{D})$. This implies, by the definition of the $\bar{b}_n$, that $\tilde{b}(f_i)$ is an accumulation point of $\{b_n^{1/n}(f_i^n), n \geqslant 1\}$ for $i = 1, \ldots, k$.

We complete the proof by showing that $\tilde{b} \in \mathbf{B}_1(\mathscr{D})$. Let $f, g \in \mathscr{C}_+(\mathscr{D})$. Then, since $\tilde{b}$ is an accumulation point of $\{\bar{b}_n, n \geqslant 1\}$, it follows that $\tilde{b}(f)$, $\tilde{b}(g)$ and $\tilde{b}(f \vee g)$ are respective accumulation points of $\{\bar{b}_n(f), n \geqslant 1\}$, $\{\bar{b}_n(g), n \geqslant 1\}$ and $\{\bar{b}_n(f \vee g), n \geqslant 1\}$. Hence, by the definition of the $\bar{b}_n$, for a subsequence (n'), we have that $b_{n'}^{1/n'}(f^{n'}) \to \tilde{b}(f)$, $b_{n'}^{1/n'}(g^{n'}) \to \tilde{b}(g)$ and $b_{n'}^{1/n'}((f \vee g)^{n'}) \to \tilde{b}(f \vee g)$. By properties (2) and (4) of $\mathbf{B}(\mathscr{D})$,

$$b_n^{1/n}(f^n) \vee b_n^{1/n}(g^n) \leqslant b_n^{1/n}((f \vee g)^n) \leqslant 2^{1/n}[b_n^{1/n}(f^n) \vee b_n^{1/n}(g^n)],$$

and we conclude that $\tilde{b}(f \vee g) = \tilde{b}(f) \vee \tilde{b}(g)$. □

Lemma A.2. *The set* $\mathbf{B}_1(\mathscr{D})$ *is compact.*

Proof. An argument similar to that used in the proof of Lemma A.1 shows that the set of functionals $\{(b(f), f \in \mathscr{C}_{1,+}(\mathscr{D})), b \in \mathbf{B}_1(\mathscr{D})\}$ is closed in $[0, 1]^{\mathscr{C}_{1,+}(\mathscr{D})}$ and hence compact, which obviously is equivalent to the compactness of $\mathbf{B}_1(\mathscr{D})$. □

The next lemma is motivated by and extends Proposition 8.2 of Aubin (1984).

Lemma A.3. *Let* T *be an arbitrary set and* U *a topological space. Assume that a real-valued function* $g(t, u)$, $t \in T$, $u \in U$, *has the following properties:*

(a) $g(t, u)$ *is level-compact in* $u \in U$ *for every* $t \in T$,

(b) for every $t_1, t_2 \in T$, *there exists* $t_3 \in T$ *such that* $g(t_3, u) \geqslant g(t_1, u) \vee g(t_2, u)$ *for all* $u \in U$.

Then

$$\sup_{t\in T} \inf_{u\in U} g(t, u) = \inf_{u\in U} \sup_{t\in T} g(t, u).$$

Remark A.2. Condition (a) holds when $g(t, u)$ is lower semi-continuous in u and U is a compact topological space.

Remark A.3. If T is a directed set, condition (b) holds when $g(t, u)$ is increasing in t for all u, i.e., $g(t_1, u) \leqslant g(t_2, u)$, $u \in U$, for $t_1 \leqslant t_2$ (the latter inequality is with respect to the order on T).

Proof. We proceed analogously to Aubin (1984). Pick $\alpha > \sup_{t\in T} \inf_{u\in U} g(t, u)$. We need to prove that

$$\alpha \geqslant \inf_{u\in U} \sup_{t\in T} g(t, u). \tag{A.14}$$

Let $T_0 = \{t \in T\colon \sup_{u\in U} g(t, u) > \alpha\}$. If T_0 is empty, the proof is complete. So we assume that $T_0 \neq \varnothing$. By the choice of α, the sets $A_t = \{u \in U\colon g(t, u) \leqslant \alpha\}$ are non-empty for all $t \in T$, and they are, moreover, compact for all $t \in T_0$, since the $g(t, u)$, $u \in U$, are level-compact. Condition (b) implies that, whatever $t_1, t_2 \in T$, there exists $t_3 \in T$ such that $A_{t_1} \cap A_{t_2} \supset A_{t_3} \neq \varnothing$, which shows that finite intersections of the compact sets A_t, $t \in T_0$, are non-empty, and hence $\bigcap_{t\in T_0} A_t \neq \varnothing$. The latter is equivalent to

$$\alpha \geqslant \inf_{u\in U} \sup_{t\in T_0} g(t, u).$$

Since by the definition of T_0, $\alpha \geqslant \sup_{t\in T\setminus T_0} g(t, u)$, $u \in U$, (A.14) is proved. □

Proof of Theorem A.1. We need to prove that, for an arbitrary sequence β_n, $n \geqslant 1$, of generalized decisions,

$$\varliminf_{n\to\infty} B_n(\beta_n) \geqslant B^*. \tag{A.15}$$

The argument is similar to that in the proof of Theorem 3.1. Let $f_\theta(r)$, $\theta \in \Theta$, be some non-negative bounded functions continuous in $r \in \mathscr{D}$. Fix a non-empty $\Lambda \in \mathscr{A}(\Theta)$. We have, by the definition of $\mathbf{Z}_{n,\Lambda}$ (see (2.14)),

$$\begin{aligned}
\varliminf_{n\to\infty} \sup_{\theta\in\Lambda} \mathrm{E}_{n,\theta}^{1/n} \beta_n(f_\theta^n) &= \varliminf_{n\to\infty} \sup_{\theta\in\Lambda} \mathrm{E}_{n,\Lambda}^{1/n} \beta_n(f_\theta^n) \mathbf{Z}_{n,\Theta;\Lambda}^n \\
&\geqslant \varliminf_{n\to\infty} \left[\frac{1}{|\Lambda|} \mathrm{E}_{n,\Lambda} \sum_{\theta\in\Lambda} \beta_n(f_\theta^n) \mathbf{Z}_{n,\theta;\Lambda}^n \right]^{1/n} \\
&\geqslant \varliminf_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} \sup_{\theta\in\Lambda} \beta_n(f_\theta^n) \mathbf{Z}_{n,\theta;\Lambda}^n \\
&\geqslant \varliminf_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} u_n^n(\mathbf{Z}_{n,\Lambda}), \qquad \text{(A.16)}
\end{aligned}$$

where

$$u_n(z_\Lambda) = \inf_{b\in\mathbf{B}(\mathscr{D})} \sup_{\theta\in\Lambda} b^{1/n}(f_\theta^n) z_\theta, \qquad z_\Lambda = (z_\theta,\, \theta \in \Lambda) \in \mathbb{R}_+^\Lambda. \qquad \text{(A.17)}$$

Note that the $u_n(z_\Lambda)$, $n = 1, 2, \ldots,$ are upper semi-continuous (recall that Λ is finite) and hence measurable so that the expectations on the rightmost side of (A.16) are well defined.

Let us introduce

$$u(z_\Lambda) = \inf_{b\in\mathbf{B}_1(\mathscr{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta, \qquad z_\Lambda \in \mathbb{R}_+^\Lambda, \qquad \text{(A.18)}$$

and prove that

$$\varliminf_{n\to\infty} u_n(z_\Lambda(n)) \geqslant u(z_\Lambda), \qquad z_\Lambda \in \mathbb{R}_+^\Lambda, \qquad \text{(A.19)}$$

for each sequence $z_\Lambda(n) \to z_\Lambda$.

Let $b_n \in \mathbf{B}(\mathscr{D})$ be such that

$$\varliminf_{n\to\infty} u_n(z_\Lambda(n)) = \varliminf_{n\to\infty} \sup_{\theta\in\Lambda} b_n^{1/n}(f_\theta^n) z_\theta(n). \qquad \text{(A.20)}$$

By Lemma A.1 and since Λ is finite, there exists $\tilde{b} \in \mathbf{B}_1(\mathscr{D})$ such that $\tilde{b}(f_\theta)$ is an accumulation point of $\{b_n^{1/n}(f_\theta^n),\, n \geqslant 1\}$ for all $\theta \in \Lambda$. Therefore, we have, for a subsequence (n'),

$$\lim_{n'} b_{n'}^{1/n'}(f_\theta^{n'}) = \tilde{b}(f_\theta), \qquad \theta \in \Lambda,$$

$$\lim_{n'} \sup_{\theta\in\Lambda} b_{n'}^{1/n'}(f_\theta^{n'}) z_\theta(n') = \varliminf_{n\to\infty} \sup_{\theta\in\Lambda} b_n^{1/n}(f_\theta^n) z_\theta(n).$$

Since Λ is finite and $z_\Lambda(n') \to z_\Lambda$, we conclude that

$$\varliminf_{n\to\infty} \sup_{\theta\in\Lambda} b_n^{1/n}(f_\theta^n) z_\theta(n) = \sup_{\theta\in\Lambda} \tilde{b}(f_\theta) z_\theta$$

which, in view of (A.20) and (A.18), proves (A.19).

By (A.19) and the LD convergence of $\{\mathscr{L}(\mathbf{Z}_{n,\Lambda}|P_{n,\Lambda}),\ n \geqslant 1\}$ to $\mathbf{V}_\Lambda$, we have (see Varadhan 1984; Chaganty 1993; Puhalskii (1995)

$$\underline{\lim}_{n\to\infty} \mathrm{E}_{n,\Lambda}^{1/n} u_n^n(\mathbf{Z}_{n,\Lambda}) \geqslant \sup_{z_\Lambda \in \mathbb{R}_+^\Lambda} u(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda). \tag{A.21}$$

Since by (A.18) $u \in \mathscr{H}_\Lambda$, property (ii) of $\mathbf{V}_\Theta$ in Lemma 2.4 yields

$$\sup_{z_\Lambda \in \mathbb{R}_+^\Lambda} u(z_\Lambda)\mathbf{V}_\Lambda(z_\Lambda) = \sup_{z_\Theta \in \mathbb{R}_+^\Theta} u(\pi_\Lambda z_\Theta)\mathbf{V}_\Theta(z_\Theta).$$

Relations (A.16) and (A.21) imply then that

$$\underline{\lim}_{n\to\infty} \sup_{\theta\in\Lambda} \mathrm{E}_{n,\theta}^{1/n} \beta_n(f_\theta^n) \geqslant \sup_{z_\Theta \in \mathbb{R}_+^\Theta} u(\pi_\Lambda z_\theta)\mathbf{V}_\Theta(z_\Theta),$$

so, by the definition of the function u in (A.18),

$$\underline{\lim}_{n\to\infty} \sup_{\theta\in\Lambda} \mathrm{E}_{n,\theta}^{1/n} \beta_n(f_\theta^n) \geqslant \sup_{z_\Theta \in \mathbb{R}_+^\Theta} \inf_{b\in\mathbf{B}_1(\mathscr{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta \mathbf{V}_\Theta(z_\Theta).$$

Hence, since $\Lambda \in \mathscr{A}(\Theta)$ and $\beta_n(f)$ are increasing in f from $\mathscr{C}_+(\mathscr{D})$, it follows that

$$\underline{\lim}_{n\to\infty} \sup_{\theta\in\Theta} \mathrm{E}_{n,\theta}^{1/n} \beta_n(W_\theta^n) \geqslant \sup_{z_\Theta \in \mathbb{R}_+^\Theta} \sup_{\substack{\Lambda\in\mathscr{A}(\Theta)\\ f_\Theta\in C_W}} \inf_{b\in\mathbf{B}_1(\mathscr{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta \mathbf{V}_\Theta(z_\Theta),$$

where $C_W = \{f_\Theta = (f_\theta,\ \theta\in\Theta) \in \mathscr{C}_+(\mathscr{D})^\Theta\colon f_\theta \leqslant W_\theta,\ \theta\in\Theta\}$. Thus, (A.15) and the theorem would follow if, for every $z_\Theta = (z_\theta,\ \theta\in\Theta) \in \mathbb{R}_+^\Theta$,

$$\sup_{\substack{\Lambda\in\mathscr{A}(\Theta)\\ f_\Theta\in C_W}} \inf_{b\in\mathbf{B}_1(\mathscr{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta = \inf_{b\in\mathbf{B}_1(\mathscr{D})} \sup_{\theta\in\Theta} b(W_\theta) z_\theta. \tag{A.22}$$

Fixing z_Θ, introduce, for $\Lambda \in \mathscr{A}(\Theta)$, $f_\Theta \in \mathscr{C}_+(\mathscr{D})^\Theta$, $b \in \mathbf{B}_1(\mathscr{D})$,

$$g((\Lambda, f_\Theta), b) = \sup_{\theta\in\Lambda} b(f_\theta) z_\theta.$$

We check that $g((\Lambda, f_\Theta), b)$ satisfies the conditions of Lemma A.3. Endow the set $\mathscr{A}(\Theta) \times C_W$ with the natural order: $(\Lambda, f_\Theta) \leqslant (\Lambda', f'_\Theta)$ if $\Lambda \subset \Lambda'$ and $f_\theta \leqslant f'_\theta$, $\theta\in\Theta$. It is easily seen that $\mathscr{A}(\Theta) \times C_W$ is a directed set and $g((\Lambda, f_\Theta), b)$ is increasing in (Λ, f_Θ) for each b. Also, since Λ is finite, the definition of the topology on $\mathbf{B}(\mathscr{D})$ implies that $g((\Lambda, f_\Theta), b)$ is continuous in b for each (Λ, f_Θ). Therefore, since $\mathbf{B}_1(\mathscr{D})$ is compact by Lemma A.2, $g((\Lambda, f_\Theta), b)$ is level-compact in b. Thus, by Lemma A.3,

$$\sup_{(\Lambda, f_\Theta)\in\mathscr{A}(\Theta)\times C_W} \inf_{b\in\mathbf{B}_1(\mathscr{D})} g((\Lambda, f_\Theta), b) = \inf_{b\in\mathbf{B}_1(\mathscr{D})} \sup_{(\Lambda, f_\Theta)\in\mathscr{A}(\Theta)\times C_W} g((\Lambda, f_\Theta), b).$$

Recalling the definition of g and using the fact that by (A.12),

$$b(W_\theta) = \sup\{b(f_\theta)\colon f_\theta \leqslant W_\theta,\ f_\theta \in \mathscr{C}_+(\mathscr{D})\}, \qquad \theta\in\Theta,$$

we obtain (A.22). □

It is interesting to relate Theorem A.1 with Theorem 3.1. Let us associate with each $r \in \mathscr{D}$ an element b_r of $\mathbf{B}_1(\mathscr{D})$ defined by

$$b_r(f) = f(r), \qquad f \in \mathscr{C}_+(\mathscr{D}). \tag{A.23}$$

Then $b_{\rho_n} \in \mathscr{B}_n$ when $\rho_n \in \mathscr{B}_n$. Therefore, in view of extension (A.12) and definitions (3.1) and (A.13), $B_n(b_{\rho_n}) \leqslant R_n(\rho_n)$, so

$$\underline{\lim}_{n\to\infty} \inf_{\rho_n \in \mathscr{R}_n} R_n(\rho_n) \geqslant \underline{\lim}_{n\to\infty} \inf_{\rho_n \in \mathscr{R}_n} B_n(b_{\rho_n}) \geqslant \underline{\lim}_{n\to\infty} \inf_{\beta_n \in \mathscr{B}_n} B_n(\beta_n).$$

Similarly, $R^* \geqslant B^*$ so that Theorem 3.1 follows from Theorem A.1 if $B^* = R^*$. The next lemma establishes conditions for the latter.

Lemma A.4. *If the loss functions W_θ are such that*

$$W_\theta = \sup\{f_\theta\colon f_\theta \leqslant W_\theta,\ f_\theta \in \mathscr{C}_+(\mathscr{D}),\ f_\theta \text{ are level-compact}\}, \qquad \theta \in \Theta,$$

then $R^* = B^*$.

Remark A.4. The conditions of the lemma hold when the W_θ are level-compact and $\mathscr{D}$ is locally compact (cf. Strasser 1985, Theorem 6.4). So, if $\mathscr{D}$ is locally compact, Theorem A.1 implies Theorem 3.1.

The proof is preceded by two lemmas. We first derive an analogue of the partition of the unity (cf. Strasser 1985, Lemma 6.6).

Lemma A.5. *Let $f_1, \ldots, f_k \in \mathscr{C}_+(\mathscr{D})$. For every $\varepsilon > 0$, there exist $h_1, \ldots, h_m \in \mathscr{C}_+(\mathscr{D})$ with the following properties:*

(i) $\max_{1\leqslant j\leqslant m} h_j(r) = 1$, *$r \in \mathscr{D}$;*
(ii) $\max_{1\leqslant i\leqslant k} |f_i(r_1) - f_i(r_2)| \leqslant \varepsilon$ *for all r_1 and r_2 such that $h_j(r_1) > 0$ and $h_j(r_2) > 0$ for some $j = 1, \ldots, m$.*

Proof. The argument is similar to that in Strasser (1985). Assume, first, that $k = 1$ and $\sup_{r\in\mathscr{D}} f_1(r) = 1$. Choose m such that $3/m \leqslant \varepsilon$ and define, for $x \geqslant 0$,

$$g_j(x) = (x - (j-2))^+ \wedge (j + 1 - x)^+ \wedge 1, \qquad 1 \leqslant j \leqslant m.$$

Let

$$h_j(r) = g_j(mf_1(r)), \qquad 1 \leqslant j \leqslant m,\ r \in \mathscr{D}.$$

It is readily seen, since $g_j(x) = 1$ when $j - 1 \leqslant x \leqslant j$ and $0 \leqslant f_1(r) \leqslant 1$, that $\max_{1\leqslant j\leqslant m} h_j(r) = 1$, $r \in \mathscr{D}$.

Next, since, given $j = 1, \ldots, m$, we have $g_j(x) = 0$ when $x \notin [(j-2)^+, j+1]$, it follows that if $h_j(r_1) > 0$ and $h_j(r_2) > 0$, then $|mf_1(r_1) - mf_1(r_2)| \leqslant 3$, i.e., $|f_1(r_1) - f_1(r_2)| \leqslant 3/m \leqslant \varepsilon$ as required.

Now, if $\sup_{r\in\mathscr{D}} f_1(r) = a > 0$, then the h_j chosen as above for f_1/a and ε/a satisfy (i) and (ii).

Finally, if $k > 1$, choose, for each $i = 1, \dots, k$, functions $h_{i,j}$, $1 \leqslant j \leqslant m_i$, that satisfy (i) and (ii). Then the functions

$$h_{j_1,\dots,j_k}(r) = \prod_{i=1}^{k} h_{i,j_i}(r), \qquad 1 \leqslant j_i \leqslant m_i,\ r \in \mathscr{D},$$

meet the requirement with $m = m_1 \dots m_k$. □

Denote by T_1 the set of non-negative (upper semi-continuous) functions of finite support $(t(r),\ r \in \mathscr{D})$ such that $\sup_{r\in\mathscr{D}} t(r) = 1$. Define $\mathbf{B}_2(\mathscr{D})$ as the set of those $b \in \mathbf{B}_1(\mathscr{D})$ that can be represented as $b(f) = \sup_{r\in\mathscr{D}} f(r)t(r)$, $f \in \mathscr{C}_+(\mathscr{D})$, for some $(t(r),\ r \in \mathscr{D}) \in T_1$. The next lemma parallels Strasser (1985, Theorem 42.5).

Lemma A.6. *The set $\mathbf{B}_2(\mathscr{D})$ is dense in $\mathbf{B}_1(\mathscr{D})$ for the weak topology.*

Proof. We proceed as in the proof of Strasser (1985, Theorem 42.5). Fix $b \in \mathbf{B}_1(\mathscr{D})$ and $f_1, \dots, f_k \in \mathscr{C}_+(\mathscr{D})$. We have to check that for any $\varepsilon > 0$ there exists $\tilde{b} \in \mathbf{B}_2(\mathscr{D})$ such that $|b(f_i) - \tilde{b}(f_i)| \leqslant \varepsilon$, $1 \leqslant i \leqslant k$.

Let functions h_j, $1 \leqslant j \leqslant m$, be as in Lemma A.5. Obviously we can assume that they are not identically equal to 0. For each $j = 1, \dots, m$, choose r_j such that $h_j(r_j) > 0$. By the definition of the h_j, we have that, on the one hand,

$$|f_i(r)h_j(r) - f_i(r_j)h_j(r)| \leqslant \varepsilon, \qquad 1 \leqslant i \leqslant k,\ r \in \mathscr{D},$$

and, on the other hand,

$$f_i(r) = \max_{1\leqslant j\leqslant m} f_i(r)h_j(r), \qquad 1 \leqslant i \leqslant k,\ r \in \mathscr{D}.$$

Hence,

$$|f_i(r) - \max_{1\leqslant j\leqslant m} f_i(r_j)h_j(r)| \leqslant \max_{1\leqslant j\leqslant m} |f_i(r)h_j(r) - f_i(r_j)h_j(r)| \leqslant \varepsilon, \qquad 1 \leqslant i \leqslant k,\ r \in \mathscr{D}.$$

Properties (1), (3) and (4) of $\mathbf{B}(\mathscr{D})$ then yield

$$|b(f_i) - b(\max_{1\leqslant j\leqslant m} f_i(r_j)h_j)| \leqslant \varepsilon, \qquad 1 \leqslant i \leqslant k. \tag{A.24}$$

Now, since $b \in \mathbf{B}_1(\mathscr{D})$ and by property (3) again,

$$b(\max_{1\leqslant j\leqslant m} f_i(r_j)h_j) = \max_{1\leqslant j\leqslant m} f_i(r_j)b(h_j), \qquad 1 \leqslant i \leqslant k. \tag{A.25}$$

Define

$$t(r) = \begin{cases} \max_{l:r_l=r_j} b(h_l), & \text{if } r = r_j \text{ for some } j = 1, \dots, m, \\ 0, & \text{otherwise,} \end{cases}$$

and let

$$\tilde{b}(f) = \sup_{r\in\mathscr{D}} f(r)t(r), \qquad f \in \mathscr{C}_+(\mathscr{D}).$$

By properties (1) and (5) of $\mathbf{B}_1(\mathscr{D})$ and the choice of h_j,

$$\sup_{r\in\mathscr{D}} t(r) = \max_{1\leqslant j\leqslant m} b(h_j) = b(\max_{1\leqslant j\leqslant m} h_j) = b(\mathbf{1}) = 1,$$

so $(t(r)) \in T_1$.

Also, by the definitions of $t(r)$ and $\tilde{b}$, the right-hand side of (A.25) equals $\tilde{b}(f_i)$, and (A.25) and (A.24) yield the result. □

Proof of Lemma A.4. Since $R^* \geqslant B^*$, we prove the opposite inequality. Let f_θ, $\theta \in \Theta$, belong to $\mathscr{C}_+(\mathscr{D})$, be level-compact and be less than or equal to W_θ, $\theta \in \Theta$. By the definition of B^*,

$$B^* \geqslant \sup_{z_\Theta \in \mathbb{R}_+^\Theta} \inf_{b\in\mathbf{B}_1(\mathscr{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta \mathbf{V}_\Theta(z_\Theta), \qquad \Lambda \in \mathscr{A}(\Theta). \tag{A.26}$$

By Lemma A.6 and the definition of $\mathbf{B}_2(\mathscr{D})$, for $z_\Theta \in \mathbb{R}_+^\Theta$, $\Lambda \in \mathscr{A}(\Theta)$,

$$\begin{aligned}\inf_{b\in\mathbf{B}_1(\mathscr{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta &= \inf_{b\in\mathbf{B}_2(\mathscr{D})} \sup_{\theta\in\Lambda} b(f_\theta) z_\theta \\ &= \inf_{(t(r))\in T_1} \sup_{r\in\mathscr{D}} \sup_{\theta\in\Lambda} t(r) f_\theta(r) z_\theta \\ &= \inf_{r\in\mathscr{D}} \sup_{\theta\in\Lambda} f_\theta(r) z_\theta.\end{aligned}$$

Since the f_θ are level-compact, an application of Lemma A.3 shows, in analogy with the end of the proof of Theorem A.1, that the supremum of the latter quantity over the f_θ and $\Lambda \in \mathscr{A}(\Theta)$ equals $\inf_{r\in\mathscr{D}} \sup_{\theta\in\Theta} W_\theta(r) z_\theta$, which by (A.26) proves that $B^* \geqslant R^*$. □

Acknowledgement

Puhalskii's research was partially supported by Russian Foundation for Basic Research grant no. 95-01-01074a.

References

Aubin, J.-P. (1984) *L'Analyse non Linéaire et ses Motivations Économiques*. Paris: Masson.

Aubin, J.-P. and Ekeland, I. (1984) *Applied Nonlinear Analysis*. New York: Wiley.

Bahadur, R. (1960) On the asymptotic efficiency of tests and estimators. *Sankhyā*, **22**, 229–252.

Bahadur, R., Zabell, S., and Gupta, J. (1980) Large deviations, tests, and estimates. In I.M. Chaterabarli (ed.), *Asymptotic Theory of Statistical Tests and Estimation*, pp. 33–64. New York: Academic Press.

Basu, D. (1956) On the concept of asymptotic efficiency. *Sankhyā*, **17**, 193–196.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

Borovkov, A. and Mogulskii, A. (1992a) Large deviations and statistical invariance principle. *Theory Probab. Appl.*, **37**, 11–18.

Borovkov, A. and Mogulskii, A. (1992b) Large deviations and testing of statistical hypotheses. *Proc. Inst. Math. Russian Acad. Sci., Siberian Division*, **19**. English transl.: *Siberian Adv. Math.*, **2**(3, 4), 1992; **3**(1, 2), 1993.

Bryc, W. (1990) Large deviations by the asymptotic value method. In M.A. Pinsky (ed.), *Diffusion Processes and Related Problems in Analysis*, Vol. 1, pp. 447–472. Birkhäuser.

Chaganty, N. (1993) Large deviations for joint distributions and statistical applications. Technical Report TR93-2, Department of Mathematics and Statistics, Old Dominion University, Norfolk, VA.

Chernoff, H. (1952) A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.*, **23**, 497–507.

Dawson, D. and Gärtner, J. (1987) Large deviations from the McKean–Vlasov limit for weakly interacting diffusions. *Stochastics Stochastics Rep.*, **20**, 247–308.

Dembo, A. and Zajic, T. (1995) Large deviations: from empirical mean and measure to partial sums processes. *Stochastic Process Appl.*, **57**, 191–224.

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

Engelking, R. (1977) *General Topology.* Warszawa: PWN.

Ermakov, M. (1993) Large deviations of empirical measures and hypothesis testing. *Zap. Nauchn. Sem. LOMI RAN*, **207**, 37–60.

Freidlin, M. and Wentzell, D. (1979) *Random Perturbations of Dynamical Systems*. Moscow: Nauka [in Russian]. English translation: Springer (1984).

Fu, J. (1982) Large sample point estimation: a large deviation theory approach. *Ann. Statist.*, **10**, 762–771.

Ibragimov, I. and Khasminskii, R. (1981) *Statistical Estimation: Asymptotic Theory.* New York: Springer.

Ibragimov, I. and Radavicius, M. (1981) Probability of large deviations for the maximum likelihood estimator. *Soviet Math. Dokl.*, **23**(2), 403–406.

Kallenberg, W. (1981) Bahadur deficiency of likelihood ratio tests in exponential families. *J. Multivariate Anal.*, **11**, 506–531.

Kallenberg, W. (1983) Intermediate efficiency, theory and examples. *Ann. Statist.*, **11**, 170–182.

Kelley, J. (1955) *General Topology.* New York: Van Nostrand.

Korostelev, A. (1991) A minimaxity criterion in nonparametric regression based on large deviations probabilities. *Ann. Statist.*, **24**, 1075–1083.

Korostelev, A. (1995) Minimax large deviations risk in change-point problems. *Stochastic Process. Appl.* (To appear).

Korostelev, A. and Leonov, S. (1995) Minimax Bahadur efficiency for small confidence intervals. Discussion paper 37, Sonderforschungsbereich 373, Humboldt University, Berlin.

Korostelev, A. and Spokoiny, V. (1996) Exact asymptotics of minimax Bahadur risk in Lipschitz regression. *Statistics.*, **28**, 13–24.

Krasnoselskii, M. and Rutickii, Y. (1961) *Convex Functions and Orlicz Spaces*. Groningen: Noordhoff.

Kullback, S. (1959) *Information Theory and Statistics*. New York: Wiley.

Le Cam, L. (1960) Locally asymptotically normal families of distributions. *Univ. California Publ. Statist.*, **3**, 27–98.

Le Cam, L. (1972) Limits of experiments. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 245–261. Berkeley: University of California Press.

Le Cam, L. (1986) *Asymptotic Methods in Statistical Decision Theory.* New York: Springer-Verlag.

Lynch, J. and Sethuraman, J. (1987) Large deviations for processes with independent increments. *Ann. Probab.*, **15**(2), 610–627.

Mogulskii, A.A. (1976) Large deviations for trajectories of multidimensional random walks. *Theory Probab. Appl.*, **21**(2), 300–315.

Puhalskii, A. (1991) On functional principle of large deviations. In V. Sazonov and T. Shervashidze (eds), *New Trends in Probability and Statistics*, Vol. 1, pp. 198–218. Utrecht: VSP/Moks'las.

Puhalskii, a. (1993) On the theory of large deviations. *Theory Probab. Appl.*, **38**(3), 490–497.

Puhalskii, A. (1994a) Large deviations of semimartingales via convergence of the predictable characteristics. *Stochastics Stochastics Rep.*, **49**, 27–85.

Puhalskii, A. (1994b) The method of stochastic exponentials for large deviations. *Stochastic Process Appl.*, **54**, 45–70.

Puhalskii, A. (1995) Large deviation analysis of the single server queue. *Queueing Systems Theory Appl.*, **21**, 5–66.

Puhalskii, A. (1997) Large deviations of semimartingales: a maxingale problem approach. I. Limits as solutions to a maxingale problem. (To appear in *Stoch. Stoch. Rep.*)

Puhalskii, A. (1996) Large deviations of the statistical empirical process. In A.N. Shiryaev *et al.* (eds), *Frontiers in Pure and Applied Probability II*, pp. 163–170. Moscow: TVP.

Radavicius, M. (1983) On the probability of large deviations of maximum likelihood estimators. *Soviet Math. Dokl.*, **27**(1), 127–131.

Radavicius, M. (1991) From asymptotic efficiency in minimax sense to Bahadur efficiency. In V. Sazonov and T. Shervashidze (eds), *New Trends in Probability and Statistics*, Vol. 1, pp. 629–635. Utrecht: VSP/Moks'las.

Rao, C. (1963) Criteria of estimation in large samples. *Sankhyā A*, **25**, 189–206.

Rockafellar, R. (1970) *Convex Analysis*. Princeton, NJ: Princeton University Press.

Rubin, H. and Rukhin, A. (1983) Convergence rates of large deviations probabilities for point estimators. *Statist. Probab. Lett.*, **1**, 197–202.

Sanov, I. (1957) On the probability of large deviations of random variables. *Mat. Sb.*, **42** [in Russian]. English transl.: *Sel. Transl. Math. Statist. Probab.*, **1**, 213–244 (1961).

Schwartz, L. (1973) *Radon Measures on Arbitrary Topological Spaces and Cylindrical Measures*. Oxford: Oxford University Press.

Shiryaev, A. and Spokoiny, V. (1997) *Statistical Experiments and Decisions: Asymptotic Theory.* Springer. Forthcoming.

Sievers, G. (1978) Estimates of location: a large deviations comparison. *Ann. Statist.*, **6**, 610–618.

Strasser, H. (1985) *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory.* Berlin: de Gruyter.

Varadhan, S. (1966) Asymptotic probabilities and differential equations. *Comm. Pure Appl. Math.*, **19**(3), 261–286.

Varadhan, S. (1984) *Large Deviations and Applications.* Philadelphia: SIAM.

Vervaat, W. (1988) Narrow and vague convergence of set functions. *Statist. Probab. Lett.*, **6**(5), 295–298.

Wieand, H. (1976) A condition under which the Pitman and Bahadur approaches to efficiency coincide. *Ann. Statist.*, **4**, 1003–1011.

Received January 1996 and revised December 1996.