KYBERNETIKA — VOLUME 35 (1999), NUMBER 4, PAGES 507–525

ASYMPTOTIC RÉNYI DISTANCES FOR RANDOM FIELDS: PROPERTIES AND APPLICATIONS[1]

MARTIN JANŽURA

The approach introduced in Janžura [6] is further developed and the asymptotic Rényi distances are studied mostly from the point of their monotonicity properties. The results are applied to the problems of statistical inference.

1. INTRODUCTION

Explicit formulas for the asymptotic Rényi distances, as reasonable measures of distance between two random processes or fields, can be obtained within the class of Gibbs random fields (cf. [6]). Their properties in many aspects correspond to the properties of the non-asymptotic Rényi distances which were defined in [9] as the extension of the well-known I-divergence (Kullback–Leibler information – see also [7] and [10]). The crucial problem is given by the existence of phase transitions, which is an inherent property of random fields and which yields the break of continuity or smoothness of some fundamental quantities. Therefore the notions have to be treated carefully and the results cannot be generalized automatically.

In the present paper we deal with the asymptotic Rényi distances of fixed order with varying parameters of the underlying Gibbs random fields. In Section 3 we define several functions of that type and study their properties (as well as the properties of the inverse/pseudoinverse functions), mostly from the point of view of monotonicity.

In Section 4 we first prove the appropriate version of the large deviations theorem with the aid of techniques developed in [6] and the well-known result for the i.i.d. variables (for an alternative proof see also [4]). The result deals essentially with the quantities studied before, and it is applied to the problem of asymptotic behaviour of the error probabilities in testing statistical hypotheses.

Finally, in Section 5 the asymptotic Rényi distances are applied to the problem of parameter estimation. Following the principle of minimum distance estimation we

[1]Supported by the Grant Agency of the Academy of Sciences of the Czech Republic under Grant A 1075601.

introduce the estimates which minimize the asymptotic Rényi distances between the empirical and the theoretical distributions to obtain a class of consistent estimators.

The paper is closely tied to the preceding paper [6], but in order to make it self-contained, the basic definitions and results of [6] are recalled in the following section.

2. PRELIMINARIES

2.1. Asymptotic Rényi distances

For a pair of probability measures P, Q on a measurable space $(\Omega, \mathcal{A})$ the Rényi distance of order $a \in R$ is defined by

$$R_a(P|Q) = (a-1)^{-1} \log \int \left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)^a \mathrm{d}Q \quad \text{for } a \neq 1$$

and

$$R_1(P|Q) = \int \log \frac{\mathrm{d}P}{\mathrm{d}Q}\, \mathrm{d}P,$$

whenever the expression makes sense. Otherwise we set $R_a(P|Q) = \infty$.

Denoting by N the set of positive integers we suppose there exists a system of sub-σ-algebras $\{\mathcal{A}_n\}_{n\in N}$ satisfying $\mathcal{A}_n \nearrow \mathcal{A}$ for $n \to \infty$, and a system of constants $\{K_n\}_{n\in N}$ with $K_n \to \infty$ for $n \to \infty$.

If the limit

$$\mathcal{R}_a(P|Q) = \lim_{n\to\infty} (K_n)^{-1} R_a(P_n|Q_n)$$

exists, where $P_n = P/\mathcal{A}_n$ and $Q_n = Q/\mathcal{A}_n$ are the projections to the σ-algebra $\mathcal{A}_n \subset \mathcal{A}$ for every $n \in N$, we call it the asymptotic Rényi distance of order $a \geq 0$.

For some basic properties of the Rényi distances and the asymptotic Rényi distances cf. [7]. Let us note that we could also consider a generalized sequence (directed set, lattice) instead of N.

The above definition obviously depends on the choice of $\{\mathcal{A}_n\}_{n\in N}$ and $\{K_n\}_{n\in N}$ which are supposed fixed. On the other hand, as we shall see later, only some natural choices give reasonable results.

2.2. Random fields

Let the measurable space $(\Omega, \mathcal{A})$ be given by the infinite-dimensional product

$$(X, \mathcal{B})^T$$

where $(X, \mathcal{B})$ is a fixed standard Borel space and $T = Z^d$ is the d-dimensional integer lattice.

For every $S \subset T$ let us denote by $\mathcal{F}_S = \mathrm{Pr}_S^{-1}(\mathcal{B}^S)$ the sub-σ-algebra generated by the projection function $\mathrm{Pr}_S : X^T \to X^S$, and by $\mathcal{L}_S$ the set of all bounded $\mathcal{F}_S$-measurable functions.

Let

$$\mathcal{L} = \bigcup_{S \in \mathcal{S}} \mathcal{L}_S, \quad \mathcal{S} = \{S \subset T;\, |S| < \infty\},$$

be the set of all local (cylinder) bounded measurable functions.

Let $\mathcal{P}$ denote the set of all probability measures on $(X, \mathcal{B})^T$, which will be called the random fields, and $\mathcal{P}_\Theta \subset \mathcal{P}$ the subset of all shift-invariant (stationary) random fields,

$$P \in \mathcal{P}_\Theta \quad \text{iff} \quad P = P \circ \theta_t^{-1} \quad \text{for every } t \in T,$$

where θ_t is the shift defined by $[\theta_t(x)]_s = x_{t+s}$ for every $t,\, s \in T,\ x \in X^T$. The set $\mathcal{P}$ will be equipped with the topology of "local convergence" which is the coarsest topology on $\mathcal{P}$ making all maps

$$P \mapsto \int f \,\mathrm{d}P, \quad f \in \mathcal{L},$$

continuous. By $\|\cdot\|$ we denote the usual supremum norm.

For the sake of simplicity we consider the system of cubes

$$\{V_n\}_{n \in N},$$

where

$$V_n = \{t \in T;\, |t_i| \le n \text{ for every } i = 1, \dots, d\} \quad \text{for every } n \in N.$$

Thus, $\mathcal{A}_n = \mathcal{F}_{V_n}$ and $P_n = P_{V_n}$ is the restriction of $P \in \mathcal{P}$ to the σ-algebra $\mathcal{F}_{V_n}$. We set $K_{V_n} = |V_n| = (2n+1)^d$ for every n.

Further, let us denote by ω a fixed reference probability measure on $(X, \mathcal{B})$.

Proposition 2.1. For every $P \in \mathcal{P}_\Theta$

$$\mathcal{R}_1(P|\omega^T) = \lim_{n \to \infty} |V_n|^{-1} R_1(P_{V_n}|\omega^T_{V_n})$$

exists and equals

$$\sup_{n \in N} |V_n|^{-1} R_1(P_{V_n}|\omega^T_{V_n}).$$

Moreover,

$$\mathcal{R}_1(\cdot|\omega^T)$$

is affine and lower semicontinuous on $\mathcal{P}_\Theta$, and its level sets

$$\left\{\mathcal{R}_1(\cdot|\omega^T) \le c\right\}, \quad c \ge 0,$$

are compact and sequentially compact.

Proof. Cf. Propositions 15.12, 15.16, 15.14 and 4.15 in [3]. □

We could also understand the measure P_{V_n} on the σ-algebra B^{V_n}. Then we could write ω^{V_n} instead of $\omega^T_{V_n}$. Sometimes we shall not distinguish between these two cases.

2.3. Gibbs random fields

Let $f \in \mathcal{L}$ and $P \in \mathcal{P}_\Theta$. Suppose there exist a constant $c^P(f)$ and a sequence $\delta(V_n, P, f) \to 0$ for $n \to \infty$ such that

$$\left| |V_n|^{-1} \left[\log \frac{\mathrm{d}P_{V_n}}{\mathrm{d}\omega^T_{V_n}} - \sum_{t \in V_n} f \circ \theta_t \right] + c^P(f) \right| \le \delta(V_n, P, f) \quad \text{a.s. } [\omega^T].$$

We write $P \in G(f)$ and call P to be the (stationary) Gibbs random field with respect to the potential $f \in \mathcal{L}$. Let us summarize the main properties.

Proposition 2.2. The constant $c^P(f)$ does not depend on P since

$$c^P(f) = c(f) = \lim_{n \to \infty} |V_n|^{-1} c(V_n, f)$$

where

$$c(V_n, f) = \log \int \exp \left\{ \sum_{t \in V_n} f \circ \theta_t \right\} \mathrm{d}\omega^T \quad \text{for every } n \in N.$$

The limit exists for every $f \in \mathcal{L}$, and the function

$$c : \mathcal{L} \to R,$$

quoted as the pressure, is convex, continuous, and satisfies $|c(f)| \le \|f\|$.

Proof. Cf. Proposition 4.1 and 5.2 in [6]. □

Proposition 2.3. For every $f \in \mathcal{L}$ the set $G(f)$ of (stationary) Gibbs random fields is a non-void compact face (extremal subset) in $\mathcal{P}_\Theta$. In particular, it is equivalently given as

$$G(f) = \left\{ P \in \mathcal{P}_\Theta ; \mathcal{R}_1(P|\omega^T) = \int f \,\mathrm{d}P - c(f) \right\},$$

while for a general $Q \in \mathcal{P}_\Theta$ we have

$$R_1(Q|\omega^T) \ge \int f \,\mathrm{d}Q - c(f).$$

Proof. Cf. Proposition 4.1, Corollary 6.7 and Theorem 7.1 in [6]. □

If $G(f^0) = G(f^1)$ we shall write $f^0 \approx f^1$ and call the potentials equivalent. Some characterization conditions can be found in Theorem 8.4 in [6]. Moreover, potentials f^0, $f^1 \in \mathcal{L}$ are equivalent iff there is a constant c satisfying

$$\operatorname{ess\,sup}_{[\omega^T]} \left| |V_n|^{-1} \sum_{t \in V_n} (f^0 - f^1) \circ \theta_t + c \right| \longrightarrow 0 \quad \text{for } n \to \infty.$$

From the above condition it also follows that $f^0 \approx f^1$ iff $\int (f^0 - f^1) \,\mathrm{d}P + c = 0$ for every $P \in \mathcal{P}_\Theta$.

Let us fix f^0, $f^1 \in \mathcal{L}$. For every real $a \in \mathbb{R}$ we denote $f^a = a f^1 + (1-a) f^0$.

Proposition 2.4. Let $P^0 \in G(f^0)$, $P^1 \in G(f^1)$. Then

$$\mathcal{R}_a(P^1|P^0) = c(f^0) - c(f^1) - \frac{c(f^a) - c(f^1)}{1-a} \quad \text{for } a \neq 1,$$

and

$$R_1(P^1|P^0) = c(f^0) - c(f^1) + \int (f^1 - f^0)\, \mathrm{d}P^1.$$

Proof. Cf. Theorem 8.1 in [6]. □

Let us fix $f \in \mathcal{L}$. For $V \in \mathcal{S}$ and arbitrary A, $B \subset T$ let us denote

$$q^f(V; A|B) = \frac{\int \exp\left\{\sum_{t \in V} f \circ \theta_t\right\} \mathrm{d}\omega^A}{\int \exp\left\{\sum_{t \in V} f \circ \theta_t\right\} \mathrm{d}\omega^B}.$$

For $A \subset B$ we have a (conditional) density, and the corresponding measure will be denoted as $Q^f(V; A|B)$. In the particular case $A = \emptyset$, $B = T$ we have

$$\log q^f(V; \emptyset|T) = \sum_{t \in V} f \circ \theta_t - c(V, f).$$

We denote $A - B = \{a - b;\, a \in A,\, b \in B\}$ for A, $B \subset T$.

For every n, ℓ, $k \in N$ with $n > \ell$ we denote

$$V(n, \ell, k) = \bigcup_{s \in V_k} [V^s_{n-\ell}]$$

where $V^s_{n-\ell} = V_{n-\ell} + (2n+1)\, s$ for every $s \in V_k$.

Note that $V(n, \ell, 0) = V_{n-\ell}$, $V(n, 0, k) = V_{2kn+n+k}$, and $|V(n, \ell, k)| = |V_{n-\ell}| \cdot |V_k|$.

For $S \in \mathcal{S}$ we denote $\ell(S) = 2 \max_{s \in S} \|s\| + 1$ with $\|s\| = \sum_{i=1}^d |s_i|$. Let us also recall that $\operatorname{diam}(S) = \max_{s_1, s_2 \in S} \|s_1 - s_2\| < \ell(S)$.

3. MONOTONICITY PROPERTIES OF THE ASYMPTOTIC RÉNYI DISTANCES

Let f^0, $f^1 \in \mathcal{L}$ be a pair of non-equivalent potentials, let again $P^0 \in G(f^0)$, $P^1 \in G(f^1)$. Moreover, for an arbitrary fixed potential $g \in \mathcal{L}$ we denote $g^b = g + b(f^1 - f^0)$, and suppose some fixed $Q^b_g \in G(g^b)$ for every $b \in \mathbb{R}$.

For every $Q \in \mathcal{P}_\Theta$ (whenever the expressions make sense) introduce the following functions

$$\begin{aligned}
H(Q) &= \int (f^1 - f^0)\, \mathrm{d}Q \\
I^1_a(Q) &= \mathcal{R}_a(Q|P^0) - \mathcal{R}_a(Q|P^1) \\
I^2_a(Q) &= \mathcal{R}_a(P^0|Q) - \mathcal{R}_a(P^1|Q) \\
J^1_a(Q) &= \mathcal{R}_a(Q|Q^0_g) \\
J^2_a(Q) &= \mathcal{R}_a(Q^0_g|Q).
\end{aligned}$$

(In the notation J_a^i, $i = 1, 2$ we omit the dependence on the fixed g. If necessary we shall write $J_a^{i,g}$.) In particular we denote

$$\begin{aligned} H_g(b) &= H(Q_g^b) \\ I_{a,g}^i(b) &= I_a^i(Q_g^b) \quad \text{for } i = 1,2 \\ J_{a,g}^i(b) &= J_a^i(Q_g^b) \quad \text{for } i = 1,2. \end{aligned}$$

We shall see that all the above defined functions obey important monotonicity properties. The proofs are mostly based on the following straightforward but crucial inequality.

Lemma 3.1. For arbitrary f, $g \in \mathcal{L}$ and $Q \in G(g)$ it holds

$$c(f) - c(g) \geq \int (f - g)\,\mathrm{d}Q.$$

Proof. For every $P \in G(f)$ we obviously have

$$0 \leq \mathcal{R}_1(Q|P) = c(f) - c(g) - \int (f - g)\,\mathrm{d}Q. \qquad \square$$

Proposition 3.2.

i) $H_g(b)$ is increasing.

ii) $I_{a,g}^1(b)$ and $I_{a,g}^2(b)$ are increasing if $a > 0$ and decreasing if $a < 0$.

iii) $J_{a,g}^1(b)$ and $J_{a,g}^2(b)$ are decreasing on the negative half-line and increasing on the positive half-line for $a > 0$, and vice-versa for $a < 0$.

Proof. i) For every b^0, $b^1 \in R$, $b^0 \neq b^1$, we have

$$(b^0 - b^1)\,[H_g(b^0) - H_g(b^1)] = \mathcal{R}_1(Q_g^{b^0}|Q_g^{b^1}) + \mathcal{R}_1(Q_g^{b^1}|Q_g^{b^0}) > 0.$$

ii) For $a = 1$ we observe

$$I_{1,g}^1(b^0) - I_{1,g}^1(b^1) = H_g(b^0) - H_g(b^1).$$

With the aid of Lemma 3.1, for $a > 1$ we have

$$\begin{aligned} I_{a,g}^1(b^0) - I_{a,g}^1(b^1) &= \frac{1}{a-1}\Big[c(f^0 + a(g^{b^0} - f^0)) - c(f^1 + a(g^{b^0} - f^1)) \\ &\qquad -c(f^0 + a(g^{b^1} - f^0)) + c(f^1 + a(g^{b^1} - f^1))\Big] \\ &\geq \frac{a(b^0 - b^1)}{a-1}\left[H_{ag+(1-a)f^1}(ab^1 + a - 1) - H_{ag+(1-a)f^1}(ab^0)\right] > 0 \end{aligned}$$

if $0 < b^0 - b^1 < \frac{a-1}{a}$. Similarly, for $a < 1$ we obtain

$$\begin{aligned} I^1_{a,g}(b^0) - I^1_{a,g}(b^1) &\geq \frac{a(b^1 - b^0)}{1-a}\left[H_{ag+(1-a)f^1}(ab^0 + a - 1) - H_{ag+(1-a)f^1}(ab^1)\right] \\ &> 0 \quad \text{if } a < 0 \text{ and } 0 < b^0 - b^1 < \frac{1-a}{-a} \\ &\qquad \text{or if } a \in (0,1) \text{ and } 0 < b^0 - b^1 < \frac{1-a}{a}. \end{aligned}$$

For $I^2_{a,g}$, $a \neq 1$, we can use the relation

$$I^2_{a,g}(b) = \frac{a}{1-a}\, I^1_{1-a,g}(b),$$

while for $a = 1$ we have simply

$$I^2_{1,g}(b^0) - I^2_{1,g}(b^1) = (b^0 - b^1)\left[H_{f_0}(1) - H_{f_0}(0)\right].$$

iii) Again with the aid of Lemma 3.1 we have for $a > 1$

$$J^1_{a,g}(b^0) - J^2_{a,g}(b^1) = \frac{a}{a-1}(b^0 - b^1)\left[H_g(ab^1) - H_g(b^0)\right] > 0$$

if $b^1 < b^0 < ab^1$ or $ab^1 < b^0 < b^1$.
for $a = 1$ we have directly

$$J^1_{1,g}(b^0) - J^1_{1,g}(b^1) \geq b^1(H_g(b^0) - H_g(b^1)) > 0 \quad \text{if } b^1(b^0 - b^1) > 0.$$

Further, for $a \in (0,1)$ we obtain

$$\begin{aligned} J^1_{a,g}(b^0) - J^1_{a,g}(b^1) &\geq \frac{a}{1-a}(b^1 - b^0)\left[H_g(ab^0) - H_g(b^1)\right] > 0 \\ &\qquad \text{if } ab^0 > b^1 > b^0 \text{ or } ab^0 < b^1 < b^0, \end{aligned}$$

and, finally, for $a < 0$ it holds

$$\begin{aligned} J^1_{a,g}(b^0) - J^1_{a,g}(b^1) &\geq \frac{a}{1-a}(b^1 - b^0)\left[H_g(ab^0) - H_g(b^0)\right] > 0 \\ &\qquad \text{if } 0 > b^0 > b^1 \text{ or } 0 < b^0 < b^1. \end{aligned}$$

For $J^2_{a,g}$, $a \neq 1$, we can again use the relation

$$J^2_{a,g}(b) = \frac{a}{1-a}\, J^1_{1-a,g}(b),$$

while for $a = 1$ we have

$$\begin{aligned} J^2_{1,g}(b^0) - J^2_{1,g}(b^1) &\geq (b^0 - b^1)\left[H_g(b^1) - H_g(0)\right] > 0 \\ &\qquad \text{if } (b^0 - b^1) \cdot b^1 > 0. \end{aligned}$$

□

Let us recall that the above studied functions are defined through arbitrary fixed $Q_g^b \in G(g^b)$ for every $b \in \mathbb{R}$. On the other hand only the functions

$$H_g,\ I_{1,g}^1 = \text{const} + H_g, \quad \text{and} \quad J_{1,g}^1(b) = \text{const} + b \cdot H_g(b) - c(g^b)$$

actually depend on the particular choice of Q_g^b. All the other functions are in fact continuous transforms of the potential g^b, and therefore there are continuous functions of b, strictly monotonous on the open real line (half-line). Thus, we may directly define the inverse mappings

$$(I_{a,g}^i)^- = (I_{a,g}^i)^{-1} : R \to R$$

for $i = 1,\ a \neq 0,\ 1$; and $i = 2,\ a \neq 0$

$$(J_{a,g}^i)^- = (J_{a,g}^i)^{-1} : R^+ \to R \text{ (resp. } R^- \to R)$$

for $i = 1,\ a \neq 0,\ 1$; and $i = 2,\ a \neq 0$.

For the remaining functions let us define

$$\begin{aligned} H_g^-(\gamma) &= b \quad \text{if} \quad H(\tilde{Q})_g^b = \gamma \quad \text{for some} \quad \tilde{Q}_g^b \in G(g^b) \\ (I_{1,g}^1)^-(\gamma) &= b \quad \text{if} \quad I_1^1(\tilde{Q}_g^b) = \gamma \quad \text{for some} \quad \tilde{Q}_g^b \in G(g^b) \end{aligned}$$

and

$$(J_{1,g}^1)^-(\gamma) = b \quad \text{if} \quad J_1^1(\tilde{Q}_g^b) = \gamma \quad \text{for some} \quad \tilde{Q}_g^b \in G(g^b)$$

Theorem 3.3. All the functions H_g^-, $(I_{a,g}^i)^-$, $(J_{a,g}^i)^-$ for $i = 1, 2$, $a \neq 0$, are well-defined on open intervals, continuous and monotonous.

Proof. As it is indicated above, with the except of H_g^-, $(I_{1,g})^-$, $(J_{1,g})^-$, the proof is straightforward.

Let us prove the claim for H_g^-.

The proof for $I_{a,g}^1(b) = \text{const} + H_g(b)$ and $J_{a,g}^1(b) = b\, H_g(b) - c(g^b) + \text{const}$ is analogical.

With the aid of Proposition 3.2, for

$$\gamma \in \left(\inf_{b \in R} H_g(b), \sup_{b \in R} H_g(b) \right)$$

there exists b_γ such that

$$H_g(b) \leq \gamma \quad \text{for every } b < b_\gamma$$

and

$$H_g(b) \geq \gamma \quad \text{for every } b > b_\gamma.$$

Therefore, by Proposition 7.2 in [6] there exist

$$b_n^1 \to (b_\gamma)_- \quad \text{with} \quad \lim Q_g^{b_n^1} = Q_\gamma^- \in G(g^{b_\gamma})$$

and

$$b_n^2 \to (b_\gamma)_+ \quad \text{with} \quad \lim Q_g^{b_n^2} = Q_\gamma^+ \in G(g^{b_\gamma}).$$

If $H(Q_\gamma^+) > H(Q_\gamma^-)$, we set

$$Q_\gamma = \nu\, Q_\gamma^- + (1-\nu)\, Q_\gamma^+ \in G(g^{b_\gamma})$$

with $\nu = \frac{H(Q_\gamma^+)-\gamma}{H(Q_\gamma^+)-H(Q_\gamma^-)}$ to obtain $H(Q_\gamma) = \gamma$, and consequently $H_g^-(\gamma) = b_\gamma$. Further, let $\gamma^0 = H(Q_g^{b^0})$, $Q_g^{b^0} \in G(g^{b^0})$, and $\varepsilon > 0$. We set

$$\delta = \min_{Q \in G(g^{b^0-\varepsilon}) \cup G(g^{b^0+\varepsilon})} |H(Q) - H(Q_g^{b^0})| > 0.$$

Thus, for $|\gamma - \gamma^0| < \delta$ we obtain $|H_g^-(\gamma) - H_g^-(\gamma^0)| < \varepsilon$ by monotonicity of H_g. □

From the above results we can conclude that the phase transitions (i. e. $|G(g^b)| > 1$) cause only technical but not essential problems. We must only deal in fact with "multifunctions" like

$$\tilde{H}_g : b \mapsto \{H(Q);\, Q \in G(g^b)\}$$

in Proposition 3.2, and the inverse transforms in Theorem 3.3 may not be strictly monotonous.

The result for $J_{a,g}^2$ can be for $a \geq 1$ even strengthened in the following sense.

Proposition 3.4. Suppose b^0, $b^1 \in R$, $a \in R$, and $\overline{a} = 1 + (a-1)\frac{b^0}{b^1}$. If $\frac{b^0}{b^1} > 1$ then

$$J_{a,g}^2(b^0) > \frac{b^0}{b^1}\, J_{\overline{a},g}^2(b^1).$$

Proof. We may directly verify

$$\begin{aligned} & \mathcal{R}_a(Q_g^0 | Q_g^{b^0}) - \frac{b^0}{b^1}\mathcal{R}_{\overline{a}}(Q_g^0 | Q_g^{b^1}) \\ = \; & \mathcal{R}_1(Q_g^{b^1} | Q_g^{b^0}) + \left(\frac{b^0}{b^1} - 1\right)\mathcal{R}_1(Q_g^{b^1} | Q_g^0) > 0 \quad \text{whenever } \frac{b^0}{b^1} > 1. \end{aligned}$$

□

Corollary 3.5. Let $a \geq 1$ and $b > 1$. Then $\overline{a} = 1 + (a-1)\,b \geq 1$ and $J_{a,g}^2(b) > b\, J_{1,g}^2(1) \to \infty$ for $b \to \infty$.

Proof. The statement follows straightforward from the preceding proposition with the aid of Theorem 8.3 in [6]. □

4. APPLICATION TO TESTING STATISTICAL HYPOTHESES

Suppose f^0, f^1, g, $h \in \mathcal{L}$ with non-equivalent f^0, f^1. Let us fix some $P^0 \in G(f^0)$, $P^1 \in G(f^1)$. For the sake of simplicity we may assume

$$f^0,\, f^1,\, g,\, h \in \mathcal{L}_S \quad \text{with} \quad \ell(S) = \ell.$$

We shall study the asymptotic behaviour of the error probability

$$Q_g^0(C(\gamma_{V_n}))$$

of the test given by the critical region

$$C(\gamma_{V_n}) = \left\{ \frac{\mathrm{d}P^1_{V_n}}{\mathrm{d}P^0_{V_n}} > e^{\gamma_{V_n}} \right\}, \quad \gamma_{V_n} \in R \text{ for every } n \in N,$$

for testing the simple statistical hypothesis

$$H^0 : Q = Q_g^0 \in G(g)$$

against the alternative

$$H^1 : Q = Q_h^0 \in G(h).$$

The main result is contained in the following theorem.

Theorem 4.1. Let $|V_n|^{-1}\gamma_{V_n} \longrightarrow \gamma^0 = I_1^1(Q_g^{a^0})$ with some $Q_g^{a^0} \in G(g^{a^0})$, $a^0 > 0$. Then

$$\lim_{n\to\infty} |V_n|^{-1} \log Q_g^0(C(\gamma_{V_n})) = -J_1^1(Q_g^{a^0}) < 0.$$

Proof. We may write for every $a \geq 0$

$$|V_n|^{-1} \log Q_g^0(C(\gamma_{V_n})) = |V_n|^{-1} \log Q_g^O \left(1 \leq \left(e^{-\gamma_{V_n}} \frac{\mathrm{d}P^1_{V_n}}{\mathrm{d}P^0_{V_n}} \right)^a \right)$$

$$\begin{aligned}
&\leq -|V_n|^{-1}\, \gamma_{V_n} a + |V_n|^{-1} \log \int \left(\frac{\mathrm{d}P^1_{V_n}}{\mathrm{d}P^0_{V_n}} \right)^a \mathrm{d}Q_g^0 \\
&\qquad \longrightarrow -\gamma_0 a + c(g^a) - c(g^0) - a(c(f^1) - c(f^0)) \\
&= \mathcal{R}_1 \left(Q_g^{a^0} | Q_g^a \right) - J_1^1(Q_0^{a^0})
\end{aligned}$$

directly from definitions. Taking minimum over $a \geq 0$, we obtain $-J_1^1(Q_0^{a^0})$ as the upper bound, and $-J_1^1(Q_0^{a_0}) < 0$ since f^0, f^1 are non-equivalent. The lower bound is much more complicated and will be proved with the aid of the following sequence of results. □

Lemma 4.2.

i) Suppose $V(n,0,k_m) \subset V_m \subset V(n,0,k_m+1)$. Then it holds

$$\left\{\frac{\mathrm{d}P^1_{V_m}}{\mathrm{d}P^0_{V_m}} > e^{\gamma_{V_m}}\right\} \supset \left\{\sum_{t\in V(n,\ell,k_m)} (f^1 - f^0)\circ\theta_t > \tilde{\gamma}^n_m\right\}$$

where

$$\begin{aligned}\tilde{\gamma}^n_m &= \gamma_{V_m} + |V_m|\Big\{c(f^1) - c(f^0) + \delta(V_m, P^0, f^0) + \delta(V_m, P^1, f^1) \\ &\quad + \left(1 - \frac{|V_{n-\ell}|\,|V_{k_m}|}{|V_n|\,|V_{k_m+1}|}\right)\|f^1 - f^0\|\Big\}.\end{aligned}$$

ii) For $\tilde{Q}^g_{n,\ell} = \bigotimes_{s\in T}\tilde{Q}^{g,s}_{n,\ell}$ with $\tilde{Q}^{g,s}_{n,\ell} = Q^g(V^s_n;\emptyset|V^s_{n-\ell})$ it holds

$$\frac{\mathrm{d}[Q^0_g]_{V(n,0,k)}}{\mathrm{d}[\tilde{Q}^g_{n,\ell}]_{V(n,0,k)}} \ge \exp\left\{-|V_n|\,|V_k|\left\{4\|g\|\left(1 - \frac{|V_{n-2\ell}|}{|V_n|}\right) + \delta(V(n,0,k),\, Q^0_g,\, g)\right\}\right\}.$$

Proof. i) The assertion follows from the definition of Gibbs random fields and an obvious bound.

ii) Again by definition we have

$$\log[\mathrm{d}Q^0_g]_{V(n,0,k)} \ge \sum_{t\in V(n,0,k)} g\circ\theta_t - |V_n|\,|V_k|\left[c(g) + \delta(V(n,0,k),\, Q^0_g,\, g)\right]$$

and

$$\log[\mathrm{d}\tilde{Q}^g_{n,\ell}]_{V(n,0,k)} = \sum_{t\in V(n,0,k)} g\circ\theta_t - |V_k|\,c(V_n,g) - \sum_{s\in V_k}\log q^g(V^s_n; V^s_{n-\ell}|T),$$

where the latter term can be bounded with the aid of Lemma 6.2 ii) in [6] by

$$|V_k|\,|V_n|\left(2\|g\|\,(1 - |V_n|^{-1}\,|V_{n-2\ell}|)\right).$$

Since by Proposition 5.2 in [6] we have

$$|c(g) - |V_n|^{-1}\,c(V_n,g)| \le 2\|g\|\,\left(1 - |V_n|^{-1}\,|V_{n-\ell}|\right),$$

we obtain the result by combining the bounds. □

Proposition 4.3. For every $n > 2\ell$ it holds

$$\liminf_{m\to\infty} |V_m|^{-1}\log Q^0_g\left(C(\gamma_{V_m})\right) \ge \min_{a\ge 0}(S^n_\ell(a)),$$

whenever the minimum is attained, where

$$\begin{aligned} S_\ell^n(a) &= |V_n|^{-1}\log\int\exp\left\{a\sum_{t\in V_{n-\ell}}(f^1-f^0)\circ\theta_t\right\}\mathrm{d}\tilde{Q}^{g,0}_{n,\ell} \\ &\quad -a\left\{\gamma^0+c(f^1)-c(f^0)+(1-|V_n|^{-1}|V_{n-\ell}|)\,\|f^1-f^0\|\right\} \\ &\quad -4\|g\|\,(1-|V_n|^{-1}|V_{n-2\ell}|). \end{aligned}$$

Proof. Let us denote

$$F_s^{n,\ell}=\sum_{t\in V^s_{n-\ell}}(f^1-f^0)\circ\theta_t\in\mathcal{F}_{V_n^s}.$$

Then with the aid and notation of Lemma 4.2 we observe

$$\begin{aligned} |V_m|^{-1}\log Q_g^0(C(\gamma_{V_m}))&\geq |V_m|^{-1}\log[\tilde{Q}^g_{n,\ell}]_{V(n,0,k_m)}\left(\sum_{s\in V_{k_m}}F_s^{n,\ell}\geq\tilde{\gamma}_m^n\right) \\ &\quad -|V_m|^{-1}\left\{|V_n|\,|V_{k_m}|\left\{4\|g\|\,(1-|V_n|^{-1}|V_{n-2\ell}|)+\delta(V_m,Q_g^0,g)\right\}\right\}. \end{aligned}$$

Since under the product measure

$$[\tilde{Q}^g_{n,\ell}]_{V(n,0,k_m)}=\bigotimes\nolimits_{s\in V_{k_m}}\tilde{Q}^{g,s}_{n,\ell}$$

the variables $\{F_s^{n,\ell}\}_{s\in V_{k_m}}$ are independent, we obtain the result with the aid of the well-known Cramér large deviations theorem for i.i.d. random variables (cf. e.g. Section 2.2 in [2]). □

Lemma 4.4. It holds

$$\begin{aligned} &\left|S_\ell^n(a)-R_1(Q_g^{a^0}|Q_g^a)+R_1(Q_g^{a^0}|Q_g^0)\right| \\ \leq\ &(1-|V_n|^{-1}|V_{n-2\ell}|)\left(10\|g\|+2|a|\,\|f^1-f^0\|\right)=\mathcal{E}_{n,\ell}(a). \end{aligned}$$

Proof. By definition we have

$$\begin{aligned} &S_\ell^n(a)-R_1(Q_g^{a^0}|Q_g^a)+R_1(G_q^{a^0}|Q_g^0) \\ =\ &-4\|g\|\,(1-|V_n|^{-1}|V_{n-2\ell}|)-a(1-|V_n|^{-1}|V_{n-\ell}|)\,\|f^1-f^0\| \\ &+\left[|V_n|^{-1}\log\int\exp\left\{a\sum_{t\in V_{n-\ell}}(f^1-f^0)\circ\theta_t\right\}\mathrm{d}\tilde{Q}^{g,0}_{n,\ell}-c(g^a)+c(g)\right]. \end{aligned}$$

Since by Lemma 6.1 and Lemma 6.2 ii) in [6] we have

$$\left|\log q^g(V_n;\emptyset|V_{n-\ell})-\log q^g(V_{n-\ell;\emptyset|T})\right|\leq 2\|g\|\,(|V_n|-|V_{n-2\ell}|),$$

the latter term above can be bounded by

$$\begin{aligned} & 2\|g\|\,(1-|V_n|^{-1}\,|V_{n-2\ell}|)+\left||V_n|^{-1}[c(V_{n-\ell},g^a)-c(V_{n-\ell},g)]-c(g^a)+c(g)\right| \\ \leq\ & 2\|g\|\,(1-|V_n|^{-1}\,|V_{n-2\ell}|)+|V_n|^{-1}\,|V_{n-\ell}|\,(2\|g^a\|+2\|g\|)\,(1-|V_{n-\ell}|^{-1}\,|V_{n-2\ell}|) \end{aligned}$$

by Proposition 5.2 in [6].

Thus, putting all the terms together and simplifying with the aid of the obvious inequality $|V_{n-2\ell}| < |V_{n-\ell}| < |V_n|$ we obtain the final bound. □

Proposition 4.5. Let $a_n \in \operatorname{argmin}_{a\geq 0} S_\ell^n(a)$. Then

$$a_n \longrightarrow a^0 \quad \text{as} \quad n \longrightarrow \infty.$$

Proof. Let us denote

$$S(a) = R_1(Q_g^{a^0}|Q_g^a) - R_1(Q_g^{a^0}|Q_g^0|),$$

and observe

$$R_1(Q_g^{a^0}|Q_g^a) = J^2_{1,g^{a^0}}(a-a^0).$$

Therefore, by Proposition 3.4 and Corollary 3.5, we have

$$S(a) - S(a^0) = J^2_{1,g^{a^0}}(a-a^0) > \frac{|a-a^0|}{\varepsilon}\,\delta_\varepsilon \quad \text{for} \quad |a-a^0| > \varepsilon > 0$$

where

$$\delta_\varepsilon = \min\left(J^2_{1,g^{a^0}}(\varepsilon),\, J^2_{1,g^{a^0}}(-\varepsilon)\right) > 0.$$

Then we obtain

$$S_\ell^n(a^0) < S(a^0) + \mathcal{E}_{n,\ell}(a^0)$$

while for $|a-a^0| > \varepsilon$ we have

$$S_\ell^n(a) > S(a) - \mathcal{E}_{n,\ell}(a) > S(a^0) + \frac{|a-a^0|\,\delta_\varepsilon}{\varepsilon} - \mathcal{E}_{n,\ell}(a).$$

Let us observe

$$\mathcal{E}_{n,\ell}(a) \leq \mathcal{E}_{n,\ell}(a^0) + \frac{|a-a^0|}{\varepsilon}\,\mathcal{E}_{n,\ell}(\varepsilon).$$

Therefore suppose n to be large enough to satisfy

$$2\mathcal{E}_{n,\ell}(a^0) + \mathcal{E}_{n,\ell}(\varepsilon) < \delta_\varepsilon.$$

Then it holds

$$S_\ell^n(a) > S_\ell^n(a^0) \quad \text{for} \quad |a-a^0| > \varepsilon$$

and, consequently, $|a_n - a^0| \leq \varepsilon$ which proves the claim. □

Remark 4.6. The above result can be also deduced from the convex property of every function S_ℓ^n and the pointwise convergence proved in Lemma 4.4.

Corollary 4.7. It holds

$$\lim_{n\to\infty} \min_{a\geq 0}\{S_\ell^n(a)\} = \min_{a\geq 0} S(a) = -J_1^1(Q_g^{a^0}).$$

Proof. Directly by Lemma 4.4 and Proposition 4.5. □

Remark 4.8. Let us suppose (with the notation of Theorem 4.1)

$$\gamma^0 < I_1^1(Q_g^a) \quad \text{for every} \quad a > 0,\ Q_g^a \in G(g^a).$$

Thus we have by Theorem 4.1

$$0 \geq \liminf_{n\to\infty} |V_n|^{-1} \log Q_g^0(C(\gamma_{V_n})) \geq -J_1^1(Q_g^a)$$

for every $a > 0$, $Q_g^a \in G(g^a)$.

By Proposition 7.2 in [6] we have a sequence

$$a^k \to 0 \quad \text{and} \quad Q^k \in G(g^{a_k}) \quad \text{with} \quad J_1^1(Q^k) \to J_1^1(Q^0) = 0$$

for some $Q^0 \in G(g^0)$.

Thus for such γ^0 we obtain the zero rate. □

Now we can formulate the main result of the asymptotic behaviour of the error probabilities.

Theorem 4.9. For every

$$\gamma^0 \in \left(\max_{Q\in G(g)} I_1^1(Q), \min_{Q\in G(h)} I_1^1(Q)\right)$$

there exist

$$Q_g^{a^0} \in G(g^{a^0}), \quad Q_h^{b^0} \in G(h^{b^0}), \quad a^0 > 0 > b^0,$$

such that

$$\gamma^0 = I_1^1(Q_g^{a^0}) = I_1^1(Q_h^{b^0}),$$

and for every $|V_n|^{-1}\gamma_n \to \gamma^0$ it holds

$$\lim_{n\to\infty} |V_n|^{-1} \log Q_g^0\,(C(\gamma_{V_n})) = -J_1^{1,g}(Q_g^{a^0}) < 0$$

and

$$\lim_{n\to\infty} |V_n|^{-1} \log Q_h^0\,(C(\gamma_{V_n})^c) = -J_1^{1,h}(Q_h^{b^0}) < 0.$$

Proof. It is sufficient to prove the existence of $Q_g^{a^0}$ and $Q_h^{b^0}$ with the above properties. Then the rest of the proof follows from Theorem 4.1.

For every $f \in \mathcal{L}$ let us denote

$$I^+(f) = \max_{Q \in G(f)} I_1^1(Q), \quad I^-(f) = \min_{Q \in G(f)} I_1^1(Q).$$

Thus, we have to prove that the interval $(I^+(g), I^-(h))$ belongs to the definition range of both

$$(I_{1,g}^1)^- \quad \text{and} \quad (I_{1,h}^1)^-.$$

Due to the monotonicity and continuity of the above functions, proved in Theorem 3.3, and since

$$I_1^1(Q_g^a) - I_1^1(Q_h^b) = H(Q_g^a) - H(Q_h^b),$$

it is sufficient to prove

$$\lim_{a \to \infty} \left[H(Q_g^a) - H(Q_h^b)\right] \geq 0 \quad \text{for fixed } b$$

and

$$\lim_{b \to -\infty} \left[H(Q_g^a) - H(Q_h^b)\right] \geq 0 \quad \text{for fixed } a.$$

For $a - b > 0$ we observe

$$0 \leq \frac{R_1(Q_g^a|Q_h^b) + R_1(Q_h^b|Q_g^a)}{a-b} = H(Q_g^a) - H(Q_h^b) + \frac{\int (g-h)\,(\mathrm{d}Q_g^a - \mathrm{d}Q_h^b)}{a-b}$$

where the latter term vanishes as either $a \to \infty$ and $b \to -\infty$.

Remark 4.10. From the above proof it is clear that the definition ranges of $(I_{1,g}^1)^-$ and $(I_{1,h}^1)^-$ coincide for every g, $h \in \mathcal{L}$. This common range $\mathcal{I}$ depends only on the potentials f^0, f^1.

The validity of the above theorem can be extended to every $\gamma^0 \in \mathcal{I}$, but at least one of the rates would be zero.

Remark 4.11. By the famous Neyman–Pearson lemma it is well-known that the optimal tests are based on the likelihood ratios, i.e. in our case we should set $f^0 = g$ and $f^1 = h$. Then it can be proved that at least one of the rates is strictly better to compare with the general "unfitted" testing procedure stated at the beginning of this section (cf. e.g. [5]).

5. APPLICATION TO PARAMETER ESTIMATION

Suppose that $g_1, \ldots, g_K \in \mathcal{L}$ is a fixed collection of mutually non-equivalent potentials, and denote

$$\Gamma = Lin(g_1, \ldots, g_K).$$

Further let us consider some $f \in \mathcal{L}$. We shall interpret Γ as a parameter class and f as the "empirical" potential, based on some given data (we shall discuss the proper meaning later – cf. Remark 5.3).

Let us choose some $P^f \in G(f)$, and $Q^g \in G(g)$ for every $g \in \Gamma$. Then, for fixed $a > 0$, we may define the "estimate" g_a^f simply as the projection

$$g_a^f \in \operatorname{argmin}_{g\in\Gamma} \mathcal{R}_a(P^f|Q^g)$$

whenever it exists.

Theorem 5.1. i) If $a > 1$ then the estimate g_a^f exists for every $f \in \mathcal{L}$. If $||f_n - f|| \longrightarrow 0$ as $n \to \infty$ then

$$||g_a^{f_n} - g_a^f|| \longrightarrow 0.$$

ii) The estimate g_1^f exists for every $f \in \mathcal{L}$ and $P^f \in G(f)$. If $||f_n - f|| \longrightarrow 0$ as $n \to \infty$ then

$$\min_{g_1^f \in M^f} ||g_1^{f_n} - g_1^f|| \longrightarrow 0$$

where

$$M^f = \left\{\operatorname{argmin}_{g\in\Gamma} \mathcal{R}_1(P^f|Q^g);\ P^f \in G(f)\right\}.$$

If $f \in \Gamma$ then $M^f = \{f\}$.

iii) If $a \in (0,1)$ then the estimate g_a^f exists for every f from some open neighborhood $\partial\Gamma$ of Γ in $\mathcal{L}$. If $||f_n - f|| \longrightarrow 0$ as $n \to \infty$ with $f \in \partial\Gamma$ then

$$||g_a^{f_n} - g_a^f|| \longrightarrow 0.$$

P r o o f. i) Let us fix some $g^0 \in \Gamma$. Then

$$\mathcal{R}_a(P^f|Q^{g^0}) \le 2||f - g^0||$$

by Theorem 8.3 in [6].

By Proposition 3.4 and Corollary 3.5 we obtain, providing $||f - g|| > 1$,

$$\mathcal{R}_a(P^f|Q^g) > ||f - g||\, \mathcal{R}_1\left(P^f|Q^{f+\frac{g-f}{||g-f||}}\right) > ||f - g|| \cdot \delta$$

where

$$\delta = \min_{h\in\Gamma_f, ||h||=1} \mathcal{R}_1(P^f|Q^{f+h}) \in (0,2)$$

and $\Gamma_f = Lin(f, g_1, \dots, g_K)$. (We assume $f, g_1, \dots, g_N$ to be mutually non-equivalent since otherwise we would have directly $f \approx g^*$ for some $g^* \in \Gamma$, and consequently $g_a^f = g^*$.)

Thus, for every $g \in \Gamma$ satisfying $||f - g|| > \max\left(1, \frac{2||f-g^0||}{\delta}\right) = \beta$ we have

$$\mathcal{R}_a(P^f|Q^g) > \mathcal{R}_a(P^f|Q^{g^0}),$$

and the minimum is attained in the set

$$\{g \in \Gamma;\ ||g - f|| \le \beta\}.$$

Thus the existence of the estimate is proved, and the consistency

$$g_a^{f_n} \longrightarrow g_a^f$$

follows by the uniform convergence

$$\left|\mathcal{R}_a(P^{f_n}|Q^g) - \mathcal{R}_a(P^f|Q^g)\right| \leq \frac{2a}{a-1}\,||f^n - f||.$$

ii) For $a = 1$ the existence follows from Corollary 3.5 in the same way as above. By Proposition 7.2 in [6] we can find a subsequence

$$\left\{P^{n(j)} \in G(f_{n(j)})\right\}_{j \in N}$$

where

$$P^{n(j)} \longrightarrow P^0 \quad \text{and} \quad \mathcal{R}_1(P^{n(j)}|\omega^T) \longrightarrow \mathcal{R}_1(P^0|\omega^T)$$

for some $P^0 \in G(f)$.

Let $g^0 = \operatorname{argmin}_{g \in \Gamma} \mathcal{R}_1(P^0|Q^g)$. Then for $||g - g^0|| > \varepsilon$ it holds

$$\mathcal{R}_1(P^0|Q^g) - \mathcal{R}_1(P^0|Q^{g^0}) > c_\varepsilon ||g - g^0||$$

with some $c_\varepsilon > 0$.

Thus

$$\mathcal{R}_1(P^{n(j)}|Q^{g^0}) \leq \mathcal{R}_1(P^0|Q^{g^0}) + \delta_j,$$

where

$$\delta_j = \left|\int g^0(\mathrm{d}P^{n(j)} - \mathrm{d}P^0)\right| + \left|\mathcal{R}_1(P^{n(j)}|\omega^T) - \mathcal{R}_1(P^0|\omega^T)\right|$$

and, at the same time, for $||g - g^0|| > \varepsilon$ it holds

$$\mathcal{R}_1(P^{n(j)}|Q^g) \geq \mathcal{R}_1(P^0|Q^{g^0}) + c_\varepsilon ||g - g^0|| - \delta_j - \gamma_j(g - g^0),$$

where $\gamma_j(g - g^0) = \left|\int (g - g^0)\,(\mathrm{d}P^{n(j)} - \mathrm{d}P)\right|$.

Obviously, there exists j_ε large enough to satisfy

$$2\delta_j + \gamma_j(g - g^0) < c_\varepsilon ||g - g^0|| \quad \text{for every } j \geq j_\varepsilon,$$

and, consequently, $||g^{n(j)} - g^0|| \longrightarrow 0$ as $j \to \infty$. Since we can find such a subsequence $\{P^{n(j)}\}_{j \in N}$ with $g^0 \in M^f$ in every subsequence of $\{P^{f_n}\}_{n \in N}$, we conclude the claimed convergence.

If $f \in \Gamma$ we have obviously $f = \operatorname{argmin}_{g \in \Gamma} \mathcal{R}_1(P^f|Q^g)$ for every $P^f \in G(f)$, and therefore $M^f = \{f\}$.

iii) For $a \in (0, 1)$ we cannot apply Proposition 3.4.

Thus, let us fix some $g^0 \in \Gamma$ and denote

$$\delta_L = \min_{g \in \Gamma,\, ||g - g^0|| \geq L} \mathcal{R}_a(Q^{g^0}|Q^g)$$

for some large $L > 0$.

Let $f \in \mathcal{L}$ satisfy $\|f - g^0\| \le \varepsilon$, where $\varepsilon < \delta_L \frac{1-a}{2}$. Then

$$\mathcal{R}_a(P^f | Q^{g^0}) \le 2\|f - g^0\| \le 2\varepsilon$$

and

$$\left|\mathcal{R}_a(P^f | Q^g) - \mathcal{R}_a(Q^{g^0} | Q^g)\right| \le \frac{2a}{1-a} \|f - g^0\| \le \frac{2a\varepsilon}{1-a}.$$

Therefore

$$\mathcal{R}_a(P^f | Q^g) \ge \mathcal{R}_a(Q^{g^0} | Q^g) - \frac{2a\varepsilon}{1-a} \ge \delta_L - \frac{2a\varepsilon}{1-a} > \mathcal{R}_a(P^f | Q^{g^0})$$

if $\|g - g^0\| \ge L$.

Thus the existence is proved and the consistency follows similarly as for $a > 1$.□

Remark 5.2. It is obvious that whenever the estimate in the preceding theorem exists it can be also obtained by solving the system of equations

$$\int g_i \, \mathrm{d}Q^g = \int g_i \, \mathrm{d}P^{af+(1-a)g} \quad \text{for } i = 1, \dots, K$$

for $Q^g \in G(g)$ and some $P^{af+(1-a)g} \in G(af + (1-a)g)$. This equations can be formally used for the numerical solution with the aid of the "stochastic gradient method" (cf. [12] for details).

Remark 5.3. The empirical potential f could be hardly obtained by any non-parametric estimate. Thus we shall calculate it as a parametric estimate. We can realize that for $a = 1$ the above defined estimate depends on the data only through the "empirical integrals"

$$\int g_i \, \mathrm{d}P^f$$

which can be substituted with the empirical mean

$$\int g_i \, \mathrm{d}P^f := \frac{1}{|V_n|} \sum_{t \in V_n} g_i \circ \theta_t$$

where $V_n \subset Z^d$ is an observation region. Then the potential f can be obtained again with the stochastic gradient method (see [12]).

Let us emphasize that the empirical potential can be in such a way estimated from a much large space $\Gamma^0 \supset \Gamma$. Thus the estimate will be usually performed in two steps. First, we find $f \in \Gamma^0$ as the $a = 1$ estimate with the aid of the empirical means, and the final estimate is given as $g_a^f \in \Gamma$ with arbitrary $a > 0$.

(Received March 13, 1998.)

REFERENCES

[1] I. Csiszár: Information-type measures of difference of probability distributions and indirect observations. Stud. Sci. Math. Hungar. *2* (1967), 299–318.

[2] A. Dembo and O. Zeitouni: Large Deviations Techniques and Applications. Jones and Bartlett Publishers, Boston 1993.

[3] H. O. Georgii: Gibbs Measures and Place Transitions. De Gruyter, Berlin 1988.

[4] M. Janžura: Large deviations theorem for Gibbs random fields. In: Proc. 5th Pannonian Symp. on Math. Statist. (W. Grossmann, J. Mogyorodi, I. Vincze and W. Wertz, eds.), Akadémiai Kiadó, Budapest 1987, pp. 97–112.

[5] M. Janžura: Asymptotic behaviour of the error probabilities in the pseudo-likelihood ratio test for Gibbs-Markov distributions. In: Asymptotic Statistics (P. Mandl and M. Hušková, eds.), Physica-Verlag 1994, pp. 285–296.

[6] M. Janžura: On the concept of asymptotic Rényi distances for random fields. Kybernetika *5* (1999), 3, 353–366.

[7] F. Liese and I. Vajda: Convex Statistical Problems. Teubner, Leipzig 1987.

[8] A. Perez: Risk estimates in terms of generalized f-entropies. In: Proc. Coll. on Inform. Theory (A. Rényi, ed.), Budapest 1968.

[9] A. Rényi: On measure of entropy and information. In: Proc. 4th Berkeley Symp. Math. Statist. Prob., Univ of Calif. Press, Berkeley 1961, Vol. 1, pp. 547–561.

[10] I. Vajda: On the f-divergence and singularity of probability measures. Period. Math. Hungar. *2* (1972), 223–234.

[11] I. Vajda: The Theory of Statistical Inference and Information. Kluwer, Dordrecht – Boston – London 1989.

[12] L. Younès: Parametric inference for imperfectly observed Gibbsian fields. Probab. Theory Related Fields *82* (1989), 625–645.

RNDr. Martin Janžura, CSc., Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod vodárenskou věží 4, 182 08 Praha 8. Czech Republic.

e-mail: janzura@utia.cas.cz