KYBERNETIKA — VOLUME *35* (1999), NUMBER 3, PAGES 353 – 366

ON THE CONCEPT OF THE ASYMPTOTIC RÉNYI DISTANCES FOR RANDOM FIELDS[1]

Martin Janžura

The asymptotic Rényi distances are explicitly defined and rigorously studied for a convenient class of Gibbs random fields, which are introduced as a natural infinite-dimensional generalization of exponential distributions.

1. INTRODUCTION

The Rényi distance of a general order $a \geq 0$ was introduced in [5] as a "continuous extension" of the well-known I-divergence (Kullback–Leibler information) with which it coincides for $a = 1$. Any order distance exhibits properties of a reasonable measure of divergence, namely it assumes zero for a pair of identical probability measures and infinity for a pair of singular ones.

The notion has been thoroughly studied by many authors (cf., e. g., [1, 4, 6]), mostly in the frame of general f-divergences of probability measures (cf. [3]). Many applications for statistical procedures and decision making were proved (see [7] for a survey). Since the distributions of stochastic processes and fields are often mutually singular, in order to obtain meaningful results it seems necessary to replace the distances by the asymptotic rates. The particular rate indicates the speed of divergence between the finite-dimensional projections of the infinite-dimensional distributions. The rates will be called the asymptotic Rényi distances and their properties imitate in many aspects the properties of the "non-asymptotical" distances.

Unfortunately, the transition from the distances to the asymptotic rates is not only mechanical, there are arising many new problems that concern the properties of measures on infinite-dimensional product spaces. Moreover, the problem of evaluating the asymptotic distances is in general extremely difficult (for some particular cases cf. [3]).

Therefore, when dealing with the asymptotic Rényi distances for random process, we have first to choose a reasonable class of distributions for which the distances can be explicitly expressed, and their properties can be rigorously studied. In order to seek for a class of "easily treatable" distributions, let us recall that in the case

[1]Supported by the Grant Agency of AS CR through grant No. A 1075601.

of exponential distributions the formulas for the (non-asymptotical) Rényi distances assume a rather simple form, namely they can be expressed with the aid of the moment generating function. Following this basic observation, we shall consider the Gibbs random fields which can be understood as an infinite-dimensional counterpart of the exponential distributions. In order to emphasize the "exponential-like" form of Gibbs random fields, a new definition of the notion is introduced in Section 4, and an original technique is developed in Sections 5 and 6 to prove the equivalence with the standard definition (cf. [2]) as well as to show the existence and some basic properties of Gibbs random fields in Section 7. Finally, the main results concerning the asymptotic Rényi distances for Gibbs random fields are obtained in Section 8.

2. ASYMPTOTIC RÉNYI DISTANCES

For a pair of probability measures P, Q on a measurable space $(\Omega, \mathcal{A})$, the Rényi distance of order $a \geq 0$ 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. [3]. Let us note that we could also consider a generalized sequence (directed set, lattice) instead of N.

Let us note that sometimes the normalizing term $(a(a-1))^{-1}$ is used in the definition of R_a (cf. [3]). Such modification yields slightly different properties with more symmetric role of P and Q. Nevertheless, for our purposes we shall keep the above definition.

3. 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 (i. e. equivalent to a complete separable metric space with the σ-algebra of Borel sets) and $T = Z^{\mathrm{d}}$ 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 smallest topology on $\mathcal{P}$ making all maps

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

continuous. By $\|f\|$ 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 $K_{V_n} = |V_n| = (2n+1)^{\mathrm{d}}$ for every n.

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

Let us emphasize that the quantities below strongly depend on the choice of ω. E. g., we have $\mathcal{R}_1(P|\omega)^T = +\infty$ if P_{V_n} is not absolutely continuous with respect to $\omega_{V_n}^T$ for some $n \in N$. In what follows, we shall consider ω as a fixed hidden parameter which will be mostly suppressed in the notation.

Proposition 3.1. For every $P \in \mathcal{P}_\Theta$

$$\mathcal{R}_1(P|\omega^T) = \lim_{n \to \infty} |V_n|^{-1} R_1(P_{V_n}|\omega_{V_n}^T) \ge 0$$

exists and equals

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

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.

P r o o f . Cf. Propositions 15.12, 15.16, 15.14 and 4.15 in [2]. □

We could also understand the measure P_{V_n} on the σ-algebra $\mathcal{B}^{V_n}$. Then we could write ω^{V_n} instead of $\omega^T_{V_n}$. Sometimes we shall not distinguish between these two cases. But, in principle, we prefer to deal with measures P_{V_n} defined on sub-σ-algebras $\mathcal{F}_{V_n} \subset \mathcal{B}^T$, and functions f (potentially measurable with respect to some $\mathcal{F}_{V_n}$) defined on the whole space X^T.

4. GIBBS RANDOM FIELDS

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

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

We write $P \in G(f)$ and call P to be the (stationary) Gibbs random field with respect to the potential $f \in \mathcal{L}$. We can easily observe the following assertion.

Proposition 4.1. If there exists some $P \in G(f)$ then it holds

$$R_1(P|\omega^T) = \int f \,\mathrm{d}P - c^P(f),$$

and $c^P(f)$ does not depend on P since

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

where

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

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

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

P r o o f . Since $P \in G(f)$ and P_{V_n} is a probability measure, we have

$$1 = e^{|V_n| \left[-c^P(f) \pm \delta_n \right] + c(V_n, f)},$$

and therefore

$$c^P(f) = |V_n|^{-1} \, c(V_n, f) \pm \delta_n.$$

Since $\delta_n = \delta(V_n, P, f) \to 0$ for $n \to \infty$, the limit exists and does not depend on P.

Similarly, for $Q \in P_\Theta$ we have

$$\int f \,\mathrm{d}Q - c(f) \pm \delta_n = |V_n|^{-1} \int \log \frac{\mathrm{d}P_{V_n}}{\mathrm{d}\omega^T_{V_n}} \,\mathrm{d}Q$$

$$= \quad |V_n|^{-1}\, R_1(Q_{V_n}|\omega^T_{V_n}) - |V_n|^{-1}\, R_1(Q_{V_n}|P_{V_n}) \le |V_n|^{-1} R_1(Q_{V_n}|\omega^T_{V_n})$$

and the remaining statements follow. □

The opposite statement is more complicated. Before proving it we need some deeper results.

5. PRESSURE

The function

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

will be quoted as the pressure. First, we have to prove its existence for every $f \in \mathcal{L}$.

Lemma 5.1.

i) $|V|^{-1}\, |c(V, f_1) - c(V, f_2)| \le \|f_1 - f_2\|$ holds for every $V \in \mathcal{S};\ f_1,\ f_2 \in \mathcal{L}$;

ii) $|V|^{-1}\, |c(V, f) - c(W, f)| \le \big(1 - |V|^{-1}\, |W|\big)\, \|f\|$ holds for every $W \subset V \in \mathcal{S};\ f \in \mathcal{L}$.

The proofs follow directly from definitions with the aid of elementary bounds. □

For every $n,\ \ell,\ k \in N$ 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 $\ell(S) = 2 \max\limits_{s \in S} \|s\| + 1$. Let us also recall that $\mathrm{diam}(S) = \max\limits_{s_1, s_2 \in S} \|s_1 - s_2\| < \ell(S)$.

Proposition 5.2. For every $f \in \mathcal{L}$ there exists

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

In particular, if $f \in \mathcal{L}_S$ with $\ell(S) = \ell < \infty$ then

$$\big|c(f) - |V_n|^{-1}\, c(V_n, f)\big| \le 2\|f\|\, \big(1 - |V_n|^{-1}\, |V_{n-\ell}|\big)$$

holds for every $n > \ell$. Moreover, it holds $|c(f)| \leq \|f\|$ and $\mathcal{R}_1(P|\omega^T) \leq 2\|f\|$ for every $P \in G(f)$.

Proof. Under the assumptions it holds

$$c(V(n,\ell,k),\ f) = |V_k|\, c(V_{n-\ell}, f),$$

and therefore we obtain

$$\begin{aligned} & \left||V(n,0,k)|^{-1}\, c(V(n,0,k),\ f) - |V_n|^{-1}\, c(V_n, f)\right| \\ \leq\ & |V(n,0,k)|^{-1}\, |c(V(n,0,k),\ f) - c(V(n,\ell,k),\ f)| + |V_n|^{-1}\, |c(V_n, f) - c(V_{n-\ell}, f)| \\ \leq\ & 2\|f\| \cdot \left(1 - |V_n|^{-1}\, |V_{n-\ell}|\right) \end{aligned}$$

by Lemma 5.1 ii).

For general $m > 2n$ there exists some $k(m) \geq 1$ with

$$V(n,0,k(m)) \subset V_m \subset V(n,0,k(m)+1)$$

and again by Lemma 5.2 ii) with the aid of i) we obtain

$$\begin{aligned} & \left||V_m|^{-1}\, c(V_m, f) - |V(n,0,k(m))|^{-1}\, c(V(n,0,k(m)),\ f)\right| \\ \leq\ & 2\|f\| \cdot \left(1 - |V_m|^{-1}\, |V_n|\, |V_{k(m)}|\right) \leq 2\|f\| \left(1 - |V_{k(m)+1}|^{-1}\, |V_{k(m)}|\right). \end{aligned}$$

By combining the estimates we prove the existence of the limit. The rest of the proof is obvious. □

6. EQUIVALENCE

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

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

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

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

The following auxiliary results will be useful. We denote $A - B = \{a - b;\, a \in A,\, b \in B\}$ for $A,\ B \subset T$.

Lemma 6.1. For $W \subset V \in \mathcal{S}$ it holds

$$|\log q(V; A|B) - \log q(W; A|B)| \leq 2\|f\| \cdot |V \setminus W|.$$

Proof. The assertion is a straightforward extension of Lemma 5.1 ii). □

Lemma 6.2. i) It holds

$$q(V;A|B) = q(V \cap (A \cup B - S); A|B).$$

Therefore

$$q(V;A|B) \in \mathcal{L}_M \quad \text{with} \quad M = [V \cap (A \cup B - S)] + S$$

and

$$|\log q(V;A|B)| \leq 2\|f\| \, |V \cap (A \cup B - S)|.$$

ii) It holds

$$|\log q(V;A|B)| \leq 2\|f\| \cdot |V \cap [(A \cap B)^c - S]|.$$

P r o o f. Obviously, in order to prove i) we may write

$$\begin{aligned} & \int \exp\left\{\sum_{t\in V} f\circ\theta_t\right\} \mathrm{d}\omega^A \\ = \; & \int \exp\left\{\sum_{t\in V\cap[(A\cup B)-S]} f\circ\theta_t\right\} \mathrm{d}\omega^A \cdot \exp\left\{\sum_{t\in V\setminus[(A\cup B)-S]} f\circ\theta_t\right\} \end{aligned}$$

and similarly for B. On the other hand, for ii) we may observe

$$\int \exp \sum_{t\in V\cap[(A\cap B)^c-S]^c} f\circ\theta_t \,\mathrm{d}\omega^A = \int \exp\left\{\sum_{t\in V\cap[(A\cap B)^c-S]^c} f\circ\theta_t\right\} \mathrm{d}\omega^T$$

and the same holds for B. $\square$

Lemma 6.3. If $(V^c - S) \cap (A \cup B - S) = \emptyset$ then $M \subset V$.

P r o o f. Under the assumption we have $M = [V \cap (V^c - S)^c \cap (A \cup B - S)] + S$. For $v \in V \cap (V^c - S)^c$ we have $v + S \subset V$ for every $S \in \mathcal{S}$ which proves the claim. $\square$

Lemma 6.4. For V, $W \in \mathcal{S}$, $V \cap W = \emptyset$, and an arbitrary probability measure λ^W on $\mathcal{F}_W$ it holds

$$\begin{aligned} & \left|\log \int q(V\cup W;\emptyset|V)\,\mathrm{d}\lambda^W - \log q(V;\emptyset|T)\right| \\ \leq \; & 4\|f\| \, |V \cap (V^c - S)| + 2\|f\| \cdot |W|. \end{aligned}$$

P r o o f. Since $[V \cap (V^c - S)^c] + S \subset V$, and $W \subset V^c$, we observe

$$\begin{aligned} & \int q(V \cap (V^c - S)^c; \emptyset|V)\,\mathrm{d}\lambda^W \\ = \; & q(V \cap (V^c - S)^c; \emptyset|V) = q(V \cap (V^c - S)^c; \emptyset|T). \end{aligned}$$

By using twice Lemma 6.1 we obtain the upper bound

$$\begin{aligned} & 2\|f\|\,|(V \cup W) \setminus [V \cap (V^c - S)^c] + 2\|f\|\|\,|V \setminus [V \cap (V^c - S)^c]| \\ = \; & 4\|f\|\,|V \cap (V^c - S)| + 2\|f\|\,|W|. \end{aligned}$$ □

Lemma 6.5. If $(A - S) \cap (B - S) = \emptyset$ then

$$q(V; A \cup B|D) = \frac{q(V; A|D)}{q(V; \emptyset|B)}.$$

P r o o f. Under the assumption we can find a decomposition

$$V = V_1 \cup V_2, \quad V_1 \cap V_2 = \emptyset \quad \text{with} \quad (V_1 - S) \cap B = \emptyset \quad \text{and} \quad (V_2 - S) \cap A = \emptyset$$

to prove $q(V; A \cup B|A) = [q(V; \emptyset|B)]^{-1}$. Since obviously $q(V; A \cup B|D) = q(V; A|D) \cdot q(V; A \cup B|A)$ we have the claim. □

Now, we can prove the main result of this section.

Theorem 6.6. If $f \in \mathcal{L}$ and $P \in \mathcal{P}_\Theta$ with $\mathcal{R}_1(P|\omega^T) = \int f \,\mathrm{d}P - c(f)$ then

$$P \in G(f).$$

P r o o f. Let us suppose $f \in \mathcal{L}_S,\ \ell = \ell(S)$, and $n > 2\ell$.

Due to the assumption, Proposition 3.1, and Proposition 5.2 we have

$$0 = \lim_{k \to \infty} |V(n,0,k)|^{-1} \mathcal{R}_1 \left(P_{V(n,0,k)} | Q(V(n,0,k); \emptyset | T) \right).$$

Since

$$|V(n,0,k) \cap (V(n,0,k)^c - S)| \le |V(n)| \left(|V(k)| - |V(k-1)| \right),$$

the same holds by Lemma 6.2 i) also for $Q(V(n,0,k); V(n,0,k)^c|T)$.

Further, since the system $V(n,0,k)^c - S,\ \{V^s_{n-\ell} - S\}_{s \in V(k)}$ is given by pairwise disjoint sets, with the aid of Lemma 6.5 we obtain

$$q(V(n,0,k); V(n,0,k)^c|T) = \hat{q}^{n,\ell,k} \cdot \prod_{s \in V_k} \tilde{q}^{n,\ell}_s$$

where

$$\hat{q}^{n,\ell,k} = q\left(V(n,0,k); V(n,0,k)^c \cup V(n,\ell,k)|T\right)$$

and

$$\tilde{q}^{n,\ell}_s = q\left(V^s_n; \emptyset | V^s_{n-\ell}\right) \quad \text{for every } s \in V(k).$$

By Lemma 6.3 we have $\tilde{q}_s^{n,\ell} \in \mathcal{L}_{V_n^s}$ for every $s \in V(k)$ and obviously $\hat{q}^{n,\ell,k} \in \mathcal{L}_{V(n,0,k)\setminus V(n,\ell,k)}$. Therefore, we may write

$$\begin{aligned} & |V(n,0,k)|^{-1}\,\mathcal{R}_1\left(P_{V(n,0,k)}|Q(V(n,0,k);V(n,0,k)^c|T)\right) \\ = \;& |V(n,0,k)|^{-1}\,\mathcal{R}_1\left(P_{V(n,0,k)}|(\widehat{PQ})^{n,\ell,k}\right) \\ & + |V_n|^{-1}\int \mathcal{R}_1\left(P_{V_{n-\ell}|(V_n\setminus V_{n-\ell})}|\tilde{Q}_0^{n,\ell}\right)\mathrm{d}P \end{aligned}$$

where

$$(\widehat{PQ})^{n,\ell,k} = \bigotimes\nolimits_{s\in V_k} P_{V_{n-\ell}^s|(V_n^s\setminus V_{n-\ell}^s)} \otimes \hat{Q}^{n,\ell,k}.$$

Note that the basic regularity conditions are satisfied, and all the (translation invariant) conditional probabilities are well defined.

Thus, since both the above terms are nonnegative and tending to zero as $k \to \infty$, we must have directly

$$\mathrm{d}P_{V_{n-\ell}|(V_n\setminus V_{n-\ell})} = \tilde{q}_0^{n,\ell} \quad \text{a.s.}\ \left[\omega^{V_{n-\ell}} \otimes P_{V_n\setminus V_{n-\ell}}\right]$$

and consequently

$$\mathrm{d}P_{V_{n-\ell}} = \int \tilde{q}_0^{n,\ell}\,\mathrm{d}P_{(V_n\setminus V_{n-\ell})} \quad \text{a.s.}\ \left[\omega^{V_{n-\ell}}\right].$$

Finally, by Lemma 6.4 and Proposition 5.2 we obtain a. s.

$$\begin{aligned} & \left| |V_{n-\ell}|^{-1}\left[\log \mathrm{d}P_{V_{n-\ell}} - \sum_{t\in V_{n-\ell}} f\circ\theta_t\right] + c(f)\right| \\ \le \;& |V_{n-\ell}|^{-1}\left|\log\int \tilde{q}_0^{n,\ell}\,\mathrm{d}P_{V_n\setminus V_{n-\ell}} - \log q(V_{n-\ell};\emptyset/T)\right| \\ & + \left||V_{n-\ell}|^{-1}\,c(V_{n-\ell},f) - c(f)\right| \\ \le \;& 4\|f\|\,\frac{|V_{n-\ell}\setminus V_{n-2\ell}|}{|V_{n-\ell}|} + 2\|f\|\,\frac{|V_n\setminus V_{n-\ell}|}{|V_{n-\ell}|} + 2\|f\|\left(1 - \frac{|V_{n-2\ell}|}{|V_{n-\ell}|}\right) \\ = \;& 4\|f\|\left(1 - \frac{|V_{n-2\ell}|}{|V_{n-\ell}|}\right) + 2\|f\|\left(\frac{|V_n| - |V_{n-2\ell}|}{|V_{n-\ell}|}\right) \\ = \;& \delta_{n-\ell}(V,P,f). \end{aligned}$$ □

Thus, we have the fundamental characterization property.

Corollary 6.7. For every $f \in \mathcal{L}$ it holds

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

Proof. Directly from Theorem 6.6 and Proposition 4.1. □

7. EXISTENCE

In the preceding section we have proved among others that our definition of Gibbs random fields is equivalent to the standard one (cf. e.g. Chapter 15 in [2]). Moreover, our techniques provide also a direct proof of the basic existence properties.

Theorem 7.1. For every $f \in \mathcal{L}$ the set $G(f)$ of (stationary) Gibbs random fields is a non-void compact face in $\mathcal{P}_\Theta$.

Proof. Let us denote

$$Q_n = \bigotimes_{s\in T} \tilde{Q}_s^{n,\ell}$$

as the product measure on $\mathcal{B}^T$, where again $\tilde{q}_s^{n,\ell} = q(V_n^s;\emptyset|V_{n-\ell}^s) \in \mathcal{L}_{V_n^s}$.

In order to make the field stationary we set

$$\overline{Q}_n = |V_n|^{-1} \sum_{t\in V_n} Q_n \circ \theta_t.$$

Obviously we have

$$\begin{aligned} \mathcal{R}_1(\overline{Q}_n|\omega^T) &= |V_n|^{-1}\, R_1\left(\tilde{Q}_0^{n,\ell}|\omega^{V_n}\right) \\ = \quad & |V_n|^{-1} \int \left[\log \tilde{q}_\theta^{n,\ell}\right] \mathrm{d}\tilde{Q}_0^{n,\ell} \;\le\; 2\|f\| \end{aligned}$$

by Lemma 6.2, and therefore there exists a cluster point

$$P^* = \lim_{k\to\infty} \overline{Q}_{n(k)} \in \mathcal{P}_\Theta.$$

By Proposition 3.1 with the aid of Proposition 5.2 and Lemma 6.2 ii) we finally have

$$\begin{aligned} & \int f\,\mathrm{d}P^* - c(f) \;\le\; \mathcal{R}_1(P^*|\omega^T) \\ \le \quad & \lim_{k\to\infty} \mathcal{R}_1(\overline{Q}_{n(k)}|\omega^T) \\ = \quad & \lim_{k\to\infty} |V_{n(k)}|^{-1} \int \left\{ \sum_{t\in V_{n(k)}} f\circ\theta_t - \log \int e^{\sum_{t\in V_{n(k)}} f\circ\theta_t} \mathrm{d}\omega^{V_{n(k)-\ell}} \right\} \mathrm{d}\tilde{Q}_0^{n(k),\ell} \\ = \quad & \lim_{k\to\infty} \left(\int f\,\mathrm{d}\overline{Q}_{n(k)} - |V_{n(k)}|^{-1} c(V_{n(k)}, f) \right. \\ & \qquad \left. + |V_{n(k)}|^{-1} \int \log q(V_{n(k)}; V_{n(k)-\ell}|T)\,\mathrm{d}\tilde{Q}_0^{n,\ell} \right) \\ = \quad & \int f\,\mathrm{d}P^* - c(f), \end{aligned}$$

which proves $P^* \in G(f)$. By similar arguments we can show $G(f)$ to be a closed subset of a compact set. Since $\mathcal{R}_1(\cdot|\omega^T)$ is affine by Proposition 3.1 we obtain that $G(f)$ is a face. □

Proposition 7.2. Suppose $f_n, f \in \mathcal{L}$ with $\|f_n - f\| \to 0$ as $n \to \infty$. Then there exist a subsequence $\{P^{n(k)}\}_{k=1}^{\infty}, P^{n(k)} \in G(f_{n(k)})$, and $P \in G(f)$ such that

$$P = \lim_{k\to\infty} P^{n(k)} \quad \text{and} \quad \mathcal{R}_1(P|\omega^T) = \lim_{k\to\infty} \mathcal{R}_1(P^{n(k)}|\omega^T).$$

If $G(f) = \{P^0\}$ then

$$P^0 = \lim_{n\to\infty} P^n \quad \text{and} \quad \mathcal{R}_1(P^0|\omega^T) = \lim_{n\to\infty} \mathcal{R}_1(P^n|\omega^T)$$

for every sequence $\{P^n\}_{n=1}^{\infty},\ P^n \in G(f_n)$.

P r o o f. Since

$$\mathcal{R}_1(P^n|\omega^T) \le 2\|f_n\| \le 2\,(\|f\| + \|f - f_n\|)$$

for every $P^n \in G(f_n),\ n \in N$, again by Proposition 3.1 we can choose a convergent subsequence $\{P^{n(k)}\}_{k=1}^{\infty}$ with a limit $P \in \mathcal{P}_\Theta$ to obtain

$$\int f\,\mathrm{d}P - c(f) \le \mathcal{R}_1(P|\omega^T) \le \lim_{k\to\infty} \mathcal{R}_1(P^{n(k)}|\omega^T)$$
$$\le \quad \lim_{k\to\infty} \left[\|f_{n(k)} - f\| + \int f\,\mathrm{d}P^{n(k)} - c(f_{n(k)})\right] = \int f\,\mathrm{d}P - c(f).$$

Therefore $P \in G(f)$ and the proof is completed.

In the case of uniqueness the same result holds for every subsequence and consequently for the whole original sequence. □

8. ASYMPTOTIC RÉNYI DISTANCES FOR GIBBS RANDOM FIELDS

The definition of Gibbs random fields has been chosen in order to facilitate easy evaluation of the asymptotic Rényi distances.

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

Theorem 8.1. Let $P^0 \in G(f^0),\ P^1 \in G(f^1)$. Then

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

and

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

P r o o f. The formulas follow straightforward from the definitions and Proposition 5.2. □

From the above theorem we conclude that in this case the asymptotic Rényi distance can be directly defined for every real order $a \in \mathbb{R}$, and we shall in general treat it as a real function.

There are deep connections between the various distances with the crucial role of the basic I-divergence. Some of the relations are introduced in the following proposition.

Proposition 8.2. Let $P^0 \in G(f^0)$, $P^1 \in G(f^1)$, $P^a \in G(f^a)$, $P^b \in G(f^b)$, $a \neq 1$, $b \neq 1$. Then it holds

i) $\mathcal{R}_a(P^1|P^0) - \mathcal{R}_b(P^1|P^0) = \dfrac{(a-b)\,\mathcal{R}_1(P^b|P^1) + (b-1)\,R_1(P^b|P^a)}{(a-1)\,(b-1)}$;

ii) $\mathcal{R}_a(P^1|P^0) - \mathcal{R}_b(P^1|P^0) = \dfrac{(a-b)\,\mathcal{R}_1(P^a|P^1) - (a-1)\,R_1(P^a|P^b)}{(a-1)\,(b-1)}$;

iii) $\mathcal{R}_a(P^1|P^0) - \mathcal{R}_1(P^1|P^0) = \dfrac{\mathcal{R}_1(P^1|P^a)}{a-1}$;

iv) $\mathcal{R}_a(P^1|P^0) = \mathcal{R}_1(P^a|P^0) + \dfrac{a}{1-a}\,\mathcal{R}_1(P^a|P^1)$.

P r o o f . All the expressions can be verified by direct calculations. □

The expression iii) can be understood as the "limiting version" of i) for $b \to 1$, or of ii) for $a \to 1$. The nature of the expression iv) is a bit different from the preceding three expressions, but it is also very useful, especially for $a \in (0,1)$. The proof of the following theorem is based on these relations.

Theorem 8.3. The function

$$F(a) = \mathcal{R}_a(P^1|P^0)$$

is bounded and non-decreasing with $F(0) = 0$ and $|F(a)| \leq 2\|f^0 - f^1\|$. For $a \neq 1$ it is continuous and equal to $\frac{a}{1-a}\,\mathcal{R}_{1-a}(P^0|P^1)$.

At $a = 1$ it holds $\lim_{a\to 1_-} F(a) \leq F(1) \leq \lim_{a\to 1_+} F(a)$ with both equalities if $G(f^1) = \{P^1\}$.

P r o o f . By Lemma 5.1 i) we have $|c(f) - c(g)| \leq \|f - g\|$ for every f, $g \in \mathcal{L}$, and in particular

$$|c(f^a) - c(f^b)| \leq |a-b|\,\|f^1 - f^0\|.$$

Therefore $|F(a)| \leq 2\|f^1 - f^0\|$ and the boundedness is proved.

From Proposition 8.2 we obtain

$$\begin{aligned} F(a) \geq F(b) \quad & \text{for} \quad a > b > 1 \quad \text{by i)}, \\ & \text{for} \quad a > b = 1 \quad \text{by iii)}, \\ & \text{for} \quad 1 > a > b \quad \text{by ii)}, \\ \text{and} \quad & \text{for} \quad 1 = a > b \quad \text{by iii)}. \end{aligned}$$

Further, with the aid of Proposition 8.2 i) we obtain

$$|F(a) - F(b)| \leq \frac{2|a-b| \cdot \|f^0 - f^1\|}{|a-1|}$$

which proves the continuity for $a \neq 1$, while the inequalities at $a = 1$ follow from the monotonicity.

The equality $\mathcal{R}_a(P^1|P^0) = \frac{a}{1-a}\,\mathcal{R}_{1-a}(P^0|P^1)$ can be easily verified since $\mathcal{R}_{1-a}(P^0|P^1) = c(f^1) - c(f^0) - \frac{c(f^a)-c(f^0)}{a}$.

Let $G(f^1) = \{P^1\}$ and $a_n \to 1$. Then, according to Proposition 7.2, for a sequence $\{P^n\}$ with $P^n \in G(f^{a_n})$ we have $P^1 = \lim_{k\to\infty} P^n$ and $\lim_{k\to\infty} \mathcal{R}_1(P^n|P^0) = \mathcal{R}_1(P^1|P^0)$. Moreover, by Proposition 8.2 iii) it holds

$$\begin{aligned} |F(a_n) - F(1)| &\leq \frac{1}{|1-a_n|}\,\mathcal{R}_1(P^1|P^n) \;\leq\; \frac{1}{|1-a_n|}\left[\mathcal{R}_1(P^n|P^1) + \mathcal{R}_1(P^1|P^n\right] \\ &= \left|\int (f^1 - f^0)\,\mathrm{d}P^1 - \int (f^1 - f^0)\,\mathrm{d}P^n\right| \longrightarrow 0 \quad \text{for } n \to \infty. \qquad \square \end{aligned}$$

An important characterization property is given in the following proposition.

Proposition 8.4. Let $P^1 \in G(f^1)$, $P^0 \in G(f^0)$. Then the following statements are equivalent:

a) $G(f^0) \cap G(f^1) \neq \emptyset$,

b) $G(f^0) = G(f^1)$,

c) $\mathcal{R}_a(P^1|P^0) = 0$ for some $a \neq 0$,

d) $\mathcal{R}_a(P^1|P^0) = 0$ for every $a \in \mathbb{R}$.

P r o o f. Let $P^* \in G(f^0) \cap G(f^1)$. Then by definition we conclude

$$\left| |V_n|^{-1} \sum_{t\in V_n} (f^0 - f^1)\circ\theta_t + c(f^1) - c(f^0)\right| \leq \delta(V_n, P^*, f^0) + \delta(V_n, P^*, f^1) \quad \text{a.s. } [\omega^T],$$

Therefore

$$\int f^0\,\mathrm{d}Q - c(f^0) = \int f^1\,\mathrm{d}Q - c(f^1)$$

for every $Q \in \mathcal{P}_\Theta$. This proves a) $\Rightarrow$ b), and also

$$c(f^a) = a\,c(f^1) + (1-a)\,c(f^0)$$

for every $a \in \mathbb{R}$, which proves a) $\Rightarrow$ d).

Let $\mathcal{R}_a(P^1|P^0) = 0$ for some $a > 0$. By monotonicity we may assume $a < 1$ and by Proposition 8.2 iv) we obtain $P^a \in G(f^a)$ with $\mathcal{R}_1(P^a|P^0) = \mathcal{R}_1(P^a|P^1) = 0$. Therefore $P^a \in G(f^0) \cap G(f^1)$, and c) $\Rightarrow$ a) is proved.

For $a < 0$ we may consider $\mathcal{R}_{1-a}(P^0|P^1) = 0$ thanks to Theorem 8.3, and by symmetry and again the monotonicity we obtain the same result.

Since b) $\Rightarrow$ a) and d) $\Rightarrow$ c) are straightforward, all desired implications are proved. $\square$

If $G(f^0) = G(f^1)$ we shall write $f^0 \approx f^1$ and call the potentials equivalent. Following the proof of the preceding proposition, we may consider the condition

$$\left| |V_n|^{-1} \sum_{t \in V_n} (f^0 - f^1) \circ \theta_t + c(f^1) - c(f^0) \right| \le \Delta(V_n, f^0, f^1) \quad \text{a.s. } [\omega^T]$$

where $\Delta(V_n, f^0, f^1) \to 0$ for $n \to \infty$, as a characterization of equivalent potentials. Thus, 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 proposition it also follows that $f^0 \approx f^1$ iff $\int (f^0 - f^1)\,\mathrm{d}P + c = 0$ for every $P \in \mathcal{P}_\Theta$.

Corollary 8.5. Let $f^0 \not\approx f^1$. Then $F(a) = \mathcal{R}_a(P^1|P^0)$ is a strictly increasing function.

P r o o f . All the appropriate terms in Proposition 8.2 which are used to prove the monotonicity in Proposition 8.3 are now positive for $a \neq b$ due to Proposition 8.4. □

(Received December 15, 1997.)

REFERENCES

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

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

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

[4] A. Perez: Risk estimates in terms of generalized f–entropies. In: Proc. Colloq. Inform. Theory (A. Rényi, ed.), Budapest 1968.

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

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

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

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