Probab. Theory Relat. Fields 96, 241–260 (1993)

Probability Theory and Related Fields

Large deviations for lattice systems I. Parametrized independent fields

Timo Seppäläinen
Department of Mathematics, The Ohio State University, 231 West 18th Avenue, Columbus, OH 43210, USA

Received April 21, 1992; in revised form January 20, 1993

Summary. We prove large deviation theorems for empirical measures of independent random fields whose distributions depend measurably on an auxiliary parameter. This dependence respects the action of the shift group, and a large deviation principle holds whenever a certain ergodicity condition is satisfied. We also investigate the entropy functions for these processes, especially in relation to the usual relative entropy.

Mathematics Subject Classification (1991): 60F10, 60G60

1 Introduction

With a view to applications to nonstationary processes, conditioned random fields, and disordered lattice systems, we study the large deviations of independent but nonstationary random fields. To compensate for the loss of invariance, we make the law $P^{\mathbf{y}}$ of the field measurably dependent on a parameter $\mathbf{y}$ in a way that respects the $\mathbb{Z}^d$-action: $P^{T_{\mathrm{i}}\mathbf{y}} = P^{\mathbf{y}} \circ T_{-\mathrm{i}}$. Suitable assumptions yield large deviation principles of the following type: Suppose we have random variables ξ_n, a sequence of cubes $V_n \uparrow \mathbb{Z}^d$, and a nonnegative rate function I defined on the Polish state space $\mathscr{S}$ of the ξ_n. We say that $\{\xi_n\}$ satisfies a large deviation principle on $\mathscr{S}$ under P with normalization $\{|V_n|\}$ and rate I, if the following inequalities hold for closed subsets F and open subsets G of $\mathscr{S}$:

$$\limsup_{n\to\infty} \frac{1}{|V_n|} \log P^{\mathbf{y}}\{\xi_n \in F\} \leqq -\inf_F I ,$$

and

$$\liminf_{n\to\infty} \frac{1}{|V_n|} \log P^{\mathbf{y}}\{\xi_n \in G\} \geqq -\inf_G I .$$

Let us say that I is a good rate function on $\mathscr{S}$ whenever $I: \mathscr{S} \to [0, \infty]$ is lower semicontinuous and has compact level sets $\{I \leqq c\}$.

After introducing our framework, we state the large deviation principles in Sect. 2 and establish some basic properties of the rate functions. In particular, we find that the rate functions differ from the entropy rates of iid large deviation theory, but can be recovered from entropy via a contraction. Section 3 studies a special case called a skew system where the rate function does coincide with entropy on part of the domain. These results will be applied in a separate paper to disordered lattice systems. Section 4 gives an application of the theorems of Sect. 2 to conditional probabilities. Further applications appear in Seppäläinen (1992).

Large deviation results of this type were first presented by Comets (1989) for conditioned iid lattice systems. A special case of the ergodicity assumption we need was earlier used by Ledrappier (1977) to establish the thermodynamic limit of a random Ising model. Large deviation results for nonstationary processes have recently appeared in Baxter and Jain (1991) and Baxter, Jain, and Seppäläinen.

To describe the setting, fix a Polish space $\mathscr{X}$ and a positive integer d. Put the product topology on the configuration space $\Omega = \mathscr{X}^{\mathbb{Z}^d}$. Equip the spaces $\mathscr{M}_1(\mathscr{X})$ and $\mathscr{M}_1(\Omega)$ of Borel probability measures with their Polish weak topologies generated by the spaces $C_b(\mathscr{X})$ and $C_b(\Omega)$ of bounded continuous functions, respectively. Let V_n denote the cube $\{\mathbf{i} = (\mathbf{i}_1, \ldots, \mathbf{i}_d) \in \mathbb{Z}^d : -n < \mathbf{i}_k < n, k = 1, \ldots, d\}$ with cardinality $|V_n| = (2n-1)^d$. The group of shift transformations $\{T_\mathbf{i} : \mathbf{i} \in \mathbb{Z}^d\}$ acts on Ω by $(T_\mathbf{i}\mathbf{z})_\mathbf{j} = \mathbf{z}_{\mathbf{i}+\mathbf{j}}$.

The variables appearing in our large deviation principles are the empirical distributions $\mathbf{L}_n$ and empirical fields $\mathbf{R}_n$ with values in $\mathscr{M}_1(\mathscr{X})$ and $\mathscr{M}_1(\Omega)$, respectively. They are defined for $\mathbf{z} = (\mathbf{z}_\mathbf{i}) \in \Omega$ by

$$\mathbf{L}_n = \frac{1}{|V_n|} \sum_{\mathbf{i} \in V_n} \delta_{\mathbf{z}_\mathbf{i}}$$

and

$$\mathbf{R}_n = \frac{1}{|V_n|} \sum_{\mathbf{i} \in V_n} \delta_{T_\mathbf{i}\mathbf{z}} .$$

The terms "position level" and "process level" are often used to distinguish between the empirical measures $\{\mathbf{L}_n\}$ and $\{\mathbf{R}_n\}$.

For $W \subset \mathbb{Z}^d$, let $\mathscr{F}_W$ denote the σ-field generated by $(Z_\mathbf{i} : \mathbf{i} \in W)$, where the spin $Z_\mathbf{i} : \Omega \to \mathscr{X}$ at the site $\mathbf{i} \in \mathbb{Z}^d$ is the projection $Z_\mathbf{i}(\mathbf{z}) = \mathbf{z}_\mathbf{i}$, and the Borel field $\mathscr{B}(\mathscr{X})$ on $\mathscr{X}$ is understood. Write simply $\mathscr{F}$ for the Borel field of Ω and $\mathscr{F}_n$ for $\mathscr{F}_{V_n}$.

We study probability measures on Ω that depend on a parameter $\mathbf{y} \in \Sigma$, where Σ is a Polish space equipped with a suitable $\mathbb{Z}^d$-action. More precisely, we have a collection $T_1, \ldots, T_d$ of commuting homeomorphisms on Σ, and writing $T_\mathbf{i} = T_1^{\mathbf{i}_1} \circ \cdots \circ T_d^{\mathbf{i}_d}$ for $\mathbf{i} = (\mathbf{i}_1, \ldots, \mathbf{i}_d) \in \mathbb{Z}^d$ gives a homomorphism from $(\mathbb{Z}^d, +)$ into the group of homeomorphisms on Σ. Y denotes a Σ-valued projection on $\Omega \times \Sigma$.

$\mathbb{Z}^d$ acts on $\Omega \times \Sigma$ by $T_\mathbf{i}(\mathbf{z}, \mathbf{y}) = (T_\mathbf{i}\mathbf{z}, T_\mathbf{i}\mathbf{y})$. The space of invariant probabilities on Ω is denoted by $\mathscr{M}_T(\Omega) = \{Q \in \mathscr{M}_1(\Omega) : Q \circ T_\mathbf{i} = Q \text{ for all } \mathbf{i} \in \mathbb{Z}^d\}$, with an analogous notation for Σ and $\Omega \times \Sigma$. An invariant probability is ergodic if invariant Borel sets have measure either 0 or 1.

We assume given a measurable map $\mathbf{y} \mapsto P^\mathbf{y}$ from Σ into $\mathscr{M}_1(\Omega)$ such that the spins $(Z_\mathbf{i})$ are independent under $P^\mathbf{y}$ and $P^{T_\mathbf{i}\mathbf{y}} = P^\mathbf{y} \circ T_{-\mathbf{i}}$. This is equivalent to having a measurable map $\mathbf{y} \mapsto p^\mathbf{y}$ from Σ into $\mathscr{M}_1(\mathscr{X})$, and then defining

$$P^\mathbf{y} = \bigotimes_{\mathbf{i} \in \mathbb{Z}^d} p^{T_\mathbf{i}\mathbf{y}} . \tag{1.1}$$

With $P^{\mathbf{y}}$ we define $\mathcal{M}_1(\mathcal{M}_1(\Omega))$-valued empirical measures by

$$\mathbf{P}_n(\mathbf{y}) = \frac{1}{|V_n|} \sum_{\mathbf{i} \in V_n} \delta_{\{P^{T_{\mathbf{i}}\mathbf{y}}\}} . \tag{1.2}$$

To clarify its meaning, let f be a bounded Borel function on Ω, so that $F(Q) = Qf$ defines a bounded Borel function on $\mathcal{M}_1(\Omega)$. The integral of F against $\mathbf{P}_n(\mathbf{y})$ is given by

$$\mathbf{P}_n(\mathbf{y}, F) = \frac{1}{|V_n|} \sum_{\mathbf{i} \in V_n} \int f dP^{T_{\mathbf{i}}\mathbf{y}} = \int \mathbf{R}_n(f) dP^{\mathbf{y}} .$$

(Depending on the context, the integral of a Borel function f against a measure Q can be denoted by Qf, $Q(f)$, $Q\{f\}$, or $\int f dQ$.) At position level we define the $\mathcal{M}_1(\mathcal{M}_1(\mathcal{X}))$-valued maps

$$\mathbf{p}_n(\mathbf{y}) = \frac{1}{|V_n|} \sum_{\mathbf{i} \in V_n} \delta_{\{p^{T_{\mathbf{i}}\mathbf{y}}\}} . \tag{1.3}$$

1.4. Definition. A parameter $\mathbf{y}$ is called *P-quasiregular*, if the limit

$$\Upsilon(\mathbf{y}) = \lim_{n \to \infty} \mathbf{P}_n(\mathbf{y})$$

exists in the weak topology of $\mathcal{M}_1(\mathcal{M}_1(\Omega))$, and *p-quasiregular*, if the limit

$$\mathfrak{v}(\mathbf{y}) = \lim_{n \to \infty} \mathbf{p}_n(\mathbf{y})$$

exists in the weak topology of $\mathcal{M}_1(\mathcal{M}_1(\mathcal{X}))$.

If $\mathbf{y}$ is a p-quasiregular parameter and $v \in \mathcal{M}_1(\mathcal{X})$, define

$$K^{\mathbf{y}}(v) = \sup_f \left\{ vf - \int_{\mathcal{M}_1(\mathcal{X})} \log \rho(e^f) \mathfrak{v}(\mathbf{y}, d\rho) \right\} , \tag{1.5}$$

where the supremum is over bounded Borel (equivalently, continuous) functions on $\mathcal{X}$. Next, suppose $\mathbf{y}$ is P-quasiregular and $Q \in \mathcal{M}_T(\Omega)$. For finite rectangles $W \subset \mathbb{Z}^d$, put

$$K^{\mathbf{y}}_{\mathcal{F}(W)}(Q) = \sup_f \left\{ Qf - \int_{\mathcal{M}_1(\Omega)} \log R(e^f) \Upsilon(\mathbf{y}, dR) \right\} , \tag{1.6}$$

where f ranges over bounded, $\mathcal{F}_w$-measurable functions. Then define

$$k^{\mathbf{y}}(Q) = \sup_W \frac{1}{|W|} K^{\mathbf{y}}_{\mathcal{F}(W)}(Q) . \tag{1.7}$$

Extend $k^{\mathbf{y}}$ to all of $\mathcal{M}_1(\Omega)$ by setting $k^{\mathbf{y}}(Q) = \infty$ if Q is not shift-invariant.

2 The large deviation principles and their rates

In this section the measures $P^{\mathbf{y}}$ are defined by (1.1) in terms of a given measurable map $p: \Sigma \to \mathcal{M}_1(\mathcal{X})$.

2.1. Theorem. *Let* $\mathbf{y} \in \Sigma$ *be P-quasiregular. Then* $\{\mathbf{R}_n\}$ *satisfies a large deviation principle on* $\mathscr{M}_1(\Omega)$ *under* $P^{\mathbf{y}}$ *with normalization* $\{|V_n|\}$ *and with rate* $k^{\mathbf{y}}$.

2.2. Theorem. *Let* $\mathbf{y} \in \Sigma$. $\{\mathbf{L}_n\}$ *satisfies a large deviation principle on* $\mathscr{M}_1(\mathscr{X})$ *under* $P^{\mathbf{y}}$ *with normalization* $\{|V_n|\}$ *if and only if* $\mathbf{y}$ *is p-quasiregular. In this case the rate function is* $K^{\mathbf{y}}$.

Proofs of these theorems are in Sect. 5. The equivalence of the large deviation principle and the regularity assumption at position level was also independently observed by S.R.S. Varadhan. We do not know at the moment if this is also the case at process level.

Most of the remainder of the paper studies the rate functions, especially to describe precisely their relationship to relative and specific relative entropy. In this section, Theorems 2.6 and 2.8 show that the rates always dominate the entropy rates of the expected process. In Theorem 2.13 we see that under an additional assumption, the rates are contractions of entropy rates on larger spaces.

Recall that the entropy of a probability ν relative to a probability μ is defined by

$$H(\nu \mid \mu) = \begin{cases} \nu\left(\log \dfrac{d\nu}{d\mu}\right) & \text{if } \nu \ll \mu , \\ \infty & \text{otherwise} . \end{cases}$$

Write $H_{\mathscr{A}}(\nu \mid \mu)$ for the relative entropy of the restrictions of ν and μ to a sub-σ-field $\mathscr{A}$. For Q and R in $\mathscr{M}_T(\Omega)$, define the specific entropy of Q relative to R by

$$h(Q \mid R) = \lim_{n \to \infty} \frac{1}{|V_n|} H_{\mathscr{F}_n}(Q \mid R) , \tag{2.3}$$

whenever this limit exists. To avoid the existence problem also define

$$\bar{h}(Q \mid R) = \sup_{W} \frac{1}{|W|} H_{\mathscr{F}(W)}(Q \mid R) , \tag{2.4}$$

where the supremum is over finite rectangles W in $\mathbb{Z}^d$. We shall make free use of the basic properties of relative entropy. The reader may consult Deuschel and Stroock (1989) and Varadhan (1984) for proofs.

For a p-quasiregular $\mathbf{y}$, define the expectation $\mu \in \mathscr{M}_1(\mathscr{X})$ of $\mathfrak{v}(\mathbf{y})$ by

$$\mu(C) = \int_{\mathscr{M}_1(\mathscr{X})} r(C)\, \mathfrak{v}(\mathbf{y}, dr) , \tag{2.5}$$

for $C \in \mathscr{B}(\mathscr{X})$. Since we always deal with one $\mathbf{y}$ at a time, the dependence of μ on $\mathbf{y}$ is suppressed from the notation.

2.6. Theorem. *Let* $\mathbf{y}$ *be a p-quasiregular parameter. Then* $K^{\mathbf{y}}$ *is a convex, good rate function on* $\mathscr{M}_1(\mathscr{X})$ *with a unique zero at* μ. $K^{\mathbf{y}}(\nu) \geqq H(\nu \mid \mu)$ *for all* $\nu \in \mathscr{M}_1(\mathscr{X})$, *and* $K^{\mathbf{y}}$ *coincides with entropy relative to* μ *if and only if* $\mathfrak{v}(\mathbf{y})$ *is a point mass.*

Proof. Lower semicontinuity and convexity are obvious from the definition (1.5). To get $K^{\mathbf{y}}(\nu) \geq H(\nu \mid \mu)$, apply Jensen's inequality to (1.5) and note that $H(\nu \mid \mu) = \sup\{\nu f - \log \mu(e^f) : f \in C_b(\mathscr{X})\}$.

$K^{\mathbf{y}}(\mu) = 0$ follows from observing that for $\nu = \mu$, the right-hand side of (1.5) is $\leqq 0$ for all f, again by Jensen's inequality. Since $H(\cdot\,|\,\mu)$ is a good rate function with a unique zero at μ, $K^{\mathbf{y}}$ is good and cannot have any other zeroes.

If $\mathfrak{v}(\mathbf{y})$ is a point mass, it is obvious that $K^{\mathbf{y}} = H(\cdot\,|\,\mu)$. Conversely, suppose this is true. For a positive Borel function f which is both bounded and bounded away from zero and satisfies $\mu f = 1$, define $\nu \in \mathcal{M}_1(\mathscr{X})$ by $d\nu = f d\mu$. Then

$$H(\nu\,|\,\mu) = K^{\mathbf{y}}(\nu) \geqq \nu(\log f) - \int \log \rho(f)\,\mathfrak{v}(\mathbf{y}, d\rho)\,,$$

from which $\int \log \rho(f)\,\mathfrak{v}(\mathbf{y}, d\rho) \geqq 0$. By Jensen's inequality $\int \log \rho(f)\,\mathfrak{v}(\mathbf{y}, d\rho) \leqq 0$, so we must have equality. Hence by the strict concavity of log, $\rho f = 1 = \mu f$ for $\mathfrak{v}(\mathbf{y})$-almost all ρ. Let f vary over a countable class of functions that separates measures to conclude that $\mathfrak{v}(\mathbf{y})$-almost every ρ equals μ. $\square$

The expectation P of $\Upsilon(\mathbf{y})$ is defined for $A \in \mathscr{F}$ by

$$P(A) = \int_{\mathcal{M}_1(\Omega)} R(A)\,\Upsilon(\mathbf{y}, dR)\,. \tag{2.7}$$

2.8. Theorem. *Suppose* $\mathbf{y}$ *is a P-quasiregular parameter. Then* $k^{\mathbf{y}}$ *is a convex, good rate function on* $\mathcal{M}_1(\Omega)$ *with a unique zero at P. For* $Q \in \mathcal{M}_T(\Omega)$, $k^{\mathbf{y}}(Q) \geqq \bar{h}(Q\,|\,P)$, *and* $k^{\mathbf{y}}(Q)$ *is given by the infinite volume limit*

$$k^{\mathbf{y}}(Q) = \lim_{n\to\infty} \frac{1}{|V_n|} K^{\mathbf{y}}_{\mathscr{F}(n)}(Q)\,. \tag{2.9}$$

Proof. Put $s(W) = K^{\mathbf{y}}_{\mathscr{F}(W)}(Q)$ for finite rectangles W in $\mathbb{Z}^d$. It follows from Lemma 2.10 below that $s(W) = s(W + \mathbf{i})$ and $s(W) \geqq s(W_1) + \cdots + s(W_m)$ whenever $W_1, \ldots, W_m$ are disjoint and contained in W. A standard superadditivity argument now gives (2.9). The remaining properties follow from applying Theorem 2.6 to the functions $K^{\mathbf{y}}_{\mathscr{F}(W)}$. $\square$

2.10. Lemma. *Suppose* $\mathbf{y}$ *is P-quasiregular. Then* $\Upsilon(\mathbf{y})$*-almost every* $R \in \mathcal{M}_1(\Omega)$ *is of the form* $R = \otimes_{\mathbf{i}\in\mathbb{Z}^d} \rho_{\mathbf{i}}$ *for some collection* $\{\rho_{\mathbf{i}}\} \subset \mathcal{M}_1(\mathscr{X})$. *The map* $\mathbf{y} \mapsto \Upsilon(\mathbf{y})$ *is invariant under the* $\mathbb{Z}^d$*-action on* Σ, *and the measure* $\Upsilon(\mathbf{y})$ *is invariant under the maps* $R \mapsto R \circ T_{\mathbf{j}}$ *on* $\mathcal{M}_1(\Omega)$.

Proof. Let C_m be the closure of the set $\{P^{T_{\mathbf{i}}\mathbf{y}} : \mathbf{i} \in \mathbb{Z}^d \backslash V_m\}$ in $\mathcal{M}_1(\Omega)$, and $C = \bigcap_{m=1}^{\infty} C_m$. Elements of C are of the form $\otimes \rho_{\mathbf{i}}$, so the first statement follows from

$$\Upsilon(\mathbf{y}, C) \geqq \lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{|V_n|} \sum_{\mathbf{i}\in V_n} \mathbf{1}_{C_m}(P^{T_{\mathbf{i}}\mathbf{y}}) = 1\,.$$

For any site $\mathbf{j}$,

$$\mathbf{P}_n(T_{\mathbf{j}}\mathbf{y}) = \frac{1}{|V_n|} \sum_{\mathbf{i}\in V_n + \mathbf{j}} \delta_{\{P^{T_{\mathbf{i}}\mathbf{y}}\}} = \frac{1}{|V_n|} \sum_{\mathbf{i}\in V_n} \delta_{\{P^{T_{\mathbf{i}}\mathbf{y}} \circ T_{-\mathbf{j}}\}}\,.$$

The limiting behavior of the middle expression is clearly independent of $\mathbf{j}$. This implies the invariance statements. $\square$

The rest of this section shows how an additional assumption on the ergodic properties of the parameter allows us to express the rate as the solution to minimizing relative entropy on $\mathcal{M}_1(\Omega \times \Sigma)$ under marginal constraints. Following

Oxtoby (1952), call a parameter $\mathbf{y}$ *quasiregular*, if the limit

$$\mathbf{K}(\mathbf{y}) = \lim_{n\to\infty} \frac{1}{|V_n|} \sum_{\mathbf{i}\in V_n} \delta_{T_\mathbf{i}\mathbf{y}}$$

exists in $\mathscr{M}_1(\Sigma)$. $\mathbf{y} \mapsto \mathbf{K}(\mathbf{y})$ is an $\mathscr{M}_T(\Sigma)$-valued, Borel measurable, invariant map on the Borel set of quasiregular parameters. If $\mathbf{K}(\mathbf{y})$ is ergodic, we call $\mathbf{y}$ *transitive.* Proof of the following proposition can be found in Oxtoby (1952) and in Sect. 2 of Dynkin (1971).

2.11. Proposition. *Let $\nu \in \mathscr{M}_T(\Sigma)$. The set of transitive parameters has ν-measure* 1, *and ν is ergodic if and only if $\nu\{\mathbf{y}: \mathbf{K}(\mathbf{y}) = \nu\} = 1$. In particular, every ergodic probability on Σ arises as a limit $\mathbf{K}(\mathbf{y})$ for some transitive $\mathbf{y}$.*

2.12. Definition. Assuming that the kernels in question exist, we say a parameter $\tilde{\mathbf{y}}$ is *P-regular,* if

$$\Upsilon(\tilde{\mathbf{y}}, A) = \int \mathbf{1}_A(P^{\mathbf{y}})\mathbf{K}(\tilde{\mathbf{y}}, d\mathbf{y})$$

for Borel $A \subset \mathscr{M}_1(\Omega)$, and *p-regular,* if

$$\upsilon(\tilde{\mathbf{y}}, C) = \int \mathbf{1}_C(p^{\mathbf{y}})\mathbf{K}(\tilde{\mathbf{y}}, d\mathbf{y})$$

for Borel $C \subset \mathscr{M}_1(\mathscr{Z})$.

We can now state our result. For a quasiregular $\tilde{\mathbf{y}}$, define $\varphi \in \mathscr{M}_1(\mathscr{Z} \times \Sigma)$ and $\Phi \in \mathscr{M}_T(\Omega \times \Sigma)$ by $\varphi(dz, d\mathbf{y}) = p^{\mathbf{y}}(dz)\mathbf{K}(\tilde{\mathbf{y}}, d\mathbf{y})$ and $\Phi(d\mathbf{z}, d\mathbf{y}) = P^{\mathbf{y}}(d\mathbf{z})\mathbf{K}(\tilde{\mathbf{y}}, d\mathbf{y})$. Let $\mathscr{E}_n = \mathscr{F}_n \vee \mathscr{B}(\Sigma)$.

2.13. Theorem. *Suppose $\tilde{\mathbf{y}}$ is p-regular. Then $K^{\tilde{\mathbf{y}}}(\nu) = \inf_\psi H(\psi \mid \varphi)$, where the infimum is over $\psi \in \mathscr{M}_1(\mathscr{Z} \times \Sigma)$ with marginals ν and $\mathbf{K}(\tilde{\mathbf{y}})$.*

Suppose $\tilde{\mathbf{y}}$ is P-regular. Then

$$h^{\mathscr{E}}(\Psi \mid \Phi) = \lim_{n\to\infty} \frac{1}{|V_n|} H_{\mathscr{E}(n)}(\Psi \mid \Phi) \tag{2.14}$$

exists for $\Psi \in \mathscr{M}_T(\Omega \times \Sigma)$ with Σ-marginal $\mathbf{K}(\tilde{\mathbf{y}})$, and $k^{\tilde{\mathbf{y}}}(Q) = \inf_\Psi h^{\mathscr{E}}(\Psi \mid \Phi)$, where the infimum is over $\Psi \in \mathscr{M}_T(\Omega \times \Sigma)$ with marginals Q and $\mathbf{K}(\tilde{\mathbf{y}})$.

If $\tilde{\mathbf{y}}$ is P-regular and transitive, then the limit in (2.14) *exists for all $\Psi \in \mathscr{M}_T(\Omega \times \Sigma)$, and $k^{\tilde{\mathbf{y}}}$ is affine on $\mathscr{M}_T(\Omega)$.*

Proof is deferred to the end of the section. The next proposition indicates that the P-regular parameters are plentiful.

2.15. Proposition. *The set of P-regular parameters has measure* 1 *under any invariant probability on Σ. If $p^{\mathbf{y}}$ depends continuously on $\mathbf{y}$, quasiregularity implies P-regularity.*

Proof. Suppose $\tilde{\mathbf{y}}$ is quasiregular and $\mathbf{y} \mapsto p^{\mathbf{y}}$ is continuous. It follows that $\mathbf{y} \mapsto P^{\mathbf{y}}$ is continuous, so that if G is a bounded continuous function on $\mathscr{M}_1(\Omega)$, $G(P^{\mathbf{y}})$ is a bounded continuous function of $\mathbf{y}$. By quasiregularity

$$\lim_{n\to\infty} \mathbf{P}_n(\tilde{\mathbf{y}}, G) = \lim_{n\to\infty} \frac{1}{|V_n|} \sum_{\mathbf{i}\in V_n} G(P^{T_\mathbf{i}\tilde{\mathbf{y}}}) = \int_\Sigma G(P^{\mathbf{y}})\mathbf{K}(\tilde{\mathbf{y}}, d\mathbf{y}),$$

which shows that $\tilde{y}$ is P-regular.

If $\mathbf{y} \mapsto P^{\mathbf{y}}$ is only measurable, $G(P^{\mathbf{y}})$ is a bounded Borel function of $\mathbf{y}$, and the convergence above takes place almost surely under any invariant probability on Σ. This gives the first statement. □

2.16. Example. A simple example of a P-quasiregular, but not P-regular, parameter. Take $d = 1$, $\Sigma = [0, 1]^{\mathbb{Z}}$ with the shift map, and pick two distinct probability measures α and β on some Polish space $\mathscr{X}$. For $\mathbf{y} = (\mathbf{y}_n) \in \Sigma$, let

$$p^{\mathbf{y}} = \begin{cases} \alpha & \text{if } 0 \leqq \mathbf{y}_0 < 1\,, \\ \beta & \text{if } \mathbf{y}_0 = 1\,. \end{cases}$$

Let η_n be any sequence in $(0, 1)$ converging to 1 as $n \to \infty$. Put $\tilde{\mathbf{y}}_n = \eta_n$. Then

$$\Upsilon(\tilde{\mathbf{y}}) = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_{\{\alpha^{\mathbb{Z}}\}} = \delta_{\{\alpha^{\mathbb{Z}}\}}\,.$$

But $\mathbf{K}(\tilde{\mathbf{y}}) = \delta_{\mathbf{1}}$, where $\mathbf{1} = (\ldots, 1, 1, 1, \ldots)$, and $P^{\mathbf{1}} = \beta^{\mathbb{Z}}$, so if $\tilde{\mathbf{y}}$ were P-regular, the limit $\Upsilon(\tilde{\mathbf{y}})$ would be $\delta_{\{\beta^{\mathbb{Z}}\}}$.

The next technical lemma will be used in a number of specific situations. Let U and V be Polish spaces, $\kappa \in \mathscr{M}_1(V)$, $v \mapsto \rho^v$ a measurable map from V into $\mathscr{M}_1(U)$, $r^v = \rho^v \otimes \delta_v$ and $\varphi(du, dv) = \rho^v(du)\kappa(dv)$ probabilities on $U \times V$, and Λ a functional on $C_b(U \times V)$ defined by

$$\Lambda(f) = \int_V \log r^v(e^f)\kappa(dv)\,.$$

For $\alpha \in \mathscr{M}_1(U \times V)$ and $\beta \in \mathscr{M}_1(U)$, set

$$J(\alpha) = \sup\{\alpha f - \Lambda(f) : f \in C_b(U \times V)\} \tag{2.17}$$

and

$$L(\beta) = \sup\{\beta g - \Lambda(g) \colon g \in C_b(U)\}\,. \tag{2.18}$$

2.19. Lemma. $J(\alpha) = H(\alpha \,|\, \varphi)$ *if the* V-*marginal of* α *is* κ, *otherwise* $J(\alpha) = \infty$. $L(\beta) = \inf_\alpha J(\alpha)$, *where the infimum is over* α *with* U-*marginal* β.

Proof. Write α^v for the conditional probability of α, given $v \in V$. If $\alpha_V = \kappa$, then the right-hand side of (2.17) is dominated by $\int H(\alpha^v \,|\, r^v)\kappa(dv) = H(\alpha \,|\, \varphi)$. We see that $J(\alpha) < \infty$ only if $\alpha_V = \kappa$ by taking $f \in C_b(V)$ in (2.17). Finally, $J(\alpha) \geqq H(\alpha \,|\, \varphi)$ as in Theorem 2.6.

To prove the second statement, assume that U and V are compact. The general case follows by a compactification argument. Since Λ is strongly continuous and convex on $C(U \times V)$, it is weakly lower semicontinuous and consequently its own double convex dual. Let $g \in C(U)$. Then

$$\begin{aligned} \Lambda(g) &= \Lambda^{**}(g) = J^*(g) \\ &= \sup\{\alpha_U g - J(\alpha) \colon \alpha \in \mathscr{M}_1(U \times V)\} \\ &= \sup\{\beta g - \inf\{J(\alpha) \colon \alpha_U = \beta\} \colon \beta \in \mathscr{M}_1(U)\}\,. \end{aligned}$$

Taking convex duals once more gives $L(\beta) = \inf\{J(\alpha) \colon \alpha_U = \beta\}$. □

2.20. Remark. This lemma contains Donsker and Varadhan's (1976, Theorem 2.1).

Proof of Theorem 2.13 To apply the above lemma, set $V = \Sigma$ and $\kappa = \mathbf{K}(\tilde{\mathbf{y}})$. The first part of the theorem follows by taking $U = \mathscr{X}$ and $\rho^{\mathbf{y}} = p^{\mathbf{y}}$, so that $L = K^{\tilde{\mathbf{y}}}$.

Now suppose $\tilde{\mathbf{y}}$ is P-regular. Let J_n be the function defined by (2.17) with $U = \mathscr{X}^{V(n)}$ and $\rho^{\mathbf{y}}$ equal to the restriction of $P^{\mathbf{y}}$ to $\mathscr{F}_n$. Then $H_{\mathscr{E}(n)}(\Psi \mid \Phi) = J_n(\Psi)$ if Ψ has marginal $\mathbf{K}(\tilde{\mathbf{y}})$. The infinite volume limit exists for the J_n's by the proof of Theorem 2.8, hence so does (2.14).

Again by Lemma 2.19, $H_{\mathscr{E}(n)}(\Psi \mid \Phi) \geqq K_n^{\tilde{\mathbf{y}}}(Q)$ for all n whenever Ψ has marginals Q and $\mathbf{K}(\tilde{\mathbf{y}})$, so $k^{\tilde{\mathbf{y}}}(Q) \leqq \inf_\Psi h^{\mathscr{E}}(\Psi \mid \Phi)$ follows.

To get the converse, use Lemma 2.19 and the properties of relative entropy to pick a probability Γ on $\mathscr{E}_n$ with marginals $Q_{\mathscr{F}(n)}$ and $\mathbf{K}(\tilde{\mathbf{y}})$ such that $H_{\mathscr{E}(n)}(\Gamma \mid \Phi) = K_n^{\tilde{\mathbf{y}}}(Q)$. Let $\{W_{\mathbf{i}} : \mathbf{i} \in H\}$ be a covering of $\mathbb{Z}^d$ with disjoint shifted copies of V_n such that $W_{\mathbf{i}}$ is centered at $\mathbf{i}$. Extend Γ to $\Gamma^n \in \mathscr{M}_1(\Omega \times \Sigma)$ such that under Γ^n the variables $\{(Z_{W(\mathbf{i})}, Y \circ T_{\mathbf{i}}) : \mathbf{i} \in H\}$ have common law Γ, and the variables $\{Z_{W(\mathbf{i})} : \mathbf{i} \in H\}$ are independent conditional on Y. Define $\Psi^n \in \mathscr{M}_T(\Omega \times \Sigma)$ by

$$\Psi^n = \frac{1}{|V_n|} \sum_{\mathbf{i} \in V_n} \Gamma^n \circ T_{\mathbf{i}} .$$

The sequence $\{\Psi^n\}$ is tight because its marginals on Ω converge to Q and those on Σ are constant. Any limit point Ψ is invariant with marginals Q and $\mathbf{K}(\tilde{\mathbf{y}})$. So by the lower semicontinuity and convexity of relative entropy, $k^{\tilde{\mathbf{y}}}(Q) \geqq h^{\mathscr{E}}(\Psi \mid \Phi)$.

Suppose $\tilde{\mathbf{y}}$ is also transitive. Then $H_{\mathscr{E}(n)}(\Psi \mid \Phi) = \infty$ whenever Ψ's Σ-marginal is not $\mathbf{K}(\tilde{\mathbf{y}})$, because $\Psi_\Sigma \ll \mathbf{K}(\tilde{\mathbf{y}})$ forces $\Psi_\Sigma = \mathbf{K}(\tilde{\mathbf{y}})$ by $\mathbf{K}(\tilde{\mathbf{y}})$'s ergodicity.

To prove that $k^{\tilde{\mathbf{y}}}$ is affine on $\mathscr{M}_T(\Omega)$, it suffices to show that $k^{\tilde{\mathbf{y}}}(Q) = \int k^{\tilde{\mathbf{y}}}(R)\alpha(dR)$ whenever $Q = \int R\alpha(dR)$ is the ergodic decomposition of Q. By Deuschel and Stroock's (1989, Lemma 5.4.24), $k^{\tilde{\mathbf{y}}}(Q) \leqq \int k^{\tilde{\mathbf{y}}}(R)\alpha(dR)$. For the opposite inequality, find Ψ such that $k^{\tilde{\mathbf{y}}}(Q) = h^{\mathscr{E}}(\Psi \mid \Phi)$, with ergodic decomposition $\Psi = \int \Gamma\eta(d\Gamma)$. By $\mathbf{K}(\tilde{\mathbf{y}})$'s ergodicity, η-almost every Γ has marginal $\mathbf{K}(\tilde{\mathbf{y}})$. $h^{\mathscr{E}}(\cdot \mid \Phi)$ is affine by the argument on p. 222 of Deuschel and Stroock (1989), so again by their Lemma 5.4.24,

$$k^{\tilde{\mathbf{y}}}(Q) = \int h^{\mathscr{E}}(\Gamma \mid \Phi)\eta(d\Gamma) \geqq \int k^{\tilde{\mathbf{y}}}(\Gamma_\Omega)\eta(d\Gamma) = \int k^{\tilde{\mathbf{y}}}(R)\alpha(dR) .$$

The last equality above follows from the uniqueness of the ergodic decomposition. □

3 Skew system run by the shift group

We study the effect of adjoining the parameter to the process as a deterministic component. We also take the parameter space to be a configuration space with the $\mathbb{Z}^d$ action by shifts. Here we find a very close connection between entropy and our rates.

Assume that $\mathscr{X} = \mathcal{X} \times \mathscr{Y}$ where $\mathcal{X}$ and $\mathscr{Y}$ are Polish spaces, and form the configuration spaces $\Xi = \mathcal{X}^{\mathbb{Z}^d}$, $\Sigma = \mathscr{Y}^{\mathbb{Z}^d}$, and $\Omega = \Xi \times \Sigma$. Write $\mathscr{F}_V^X$ for the σ-field generated by the $\mathcal{X}$-valued spins $(X_{\mathbf{i}} : \mathbf{i} \in V)$, analogously for $\mathscr{Y}$-valued spins, and set $\mathscr{E}_{V,U} = \mathscr{F}_V^X \vee \mathscr{F}_{V+U}^Y$.

The parametrized measures are now $p^{\mathbf{y}} = \pi^{\mathbf{y}} \otimes \delta_{\mathbf{y}_0}$ on $\mathscr{X}$ and $P^{\mathbf{y}} = \Pi^{\mathbf{y}} \otimes \delta_{\mathbf{y}}$ on Ω, where $\mathbf{y} \mapsto \pi^{\mathbf{y}}$ is a measurable map from Σ into $\mathscr{M}_1(\mathcal{X})$, and $\Pi^{\mathbf{y}} \in \mathscr{M}_1(\Xi)$ is defined by $\Pi^{\mathbf{y}} = \otimes \pi^{T_{\mathbf{i}}\mathbf{y}}$. It turns out that the range of the dependence of $\pi^{\mathbf{y}}$ on $\mathbf{y}$ affects the entropy representations of the rate functions. For the remainder of this section, *fix* $U \subset \mathbb{Z}^d$ *so that* $\mathbf{y} \mapsto \pi^{\mathbf{y}}$ *is* $\mathscr{F}_U^Y$*-measurable.*

Weak limits of $\mathbf{R}_n$ under $P^{\mathbf{y}}$ have Σ-marginal $\mathbf{K}(\mathbf{y})$, if $\mathbf{y}$ is quasiregular. So it is of interest to note that the last statement of Proposition 2.11 can be strengthened to cover all shift-invariant measures on Σ.

3.1. Lemma. *Every shift-invariant probability measure on Σ is equal to $\mathbf{K}(\mathbf{y})$ for some quasiregular configuration $\mathbf{y}$.*

Proof. It is well-known that ergodic measures are dense in $\mathcal{M}_T(\Sigma)$, see for example Lemma 3.2 of Föllmer and Orey (1988). Thus it suffices to show that, if $\nu_k \to \nu$ in $\mathcal{M}_1(\Sigma)$ and $\nu_k = \mathbf{K}(\mathbf{y}^k)$ for each k, then $\nu = \mathbf{K}(\mathbf{y})$ for some quasiregular configuration $\mathbf{y}$. The idea is to let $\mathbf{y}$ agree with the $\mathbf{y}^k$'s on successive rectangular shells $V_{n_k} \setminus V_{n_{k-1}}$.

Find $g_1, g_2, g_3, \ldots \in C_b(\Sigma)$ and integers r_k so that $g_1, \ldots, g_k$ are $\mathcal{F}^Y_{r_k}$-measurable and

$$d(\kappa, \lambda) = \sum_{i=1}^{\infty} 2^{-i} \frac{|\kappa g_i - \lambda g_i|}{\|g_i\|} \tag{3.2}$$

metrizes $\mathcal{M}_1(\Sigma)$. Let $\varepsilon_k = 2^{-k}/9$. Without loss assume that, for all k, $|\nu_k(g_i) - \nu(g_i)| < \varepsilon_k \|g_i\|$ for $i = 1, \ldots, k$. Choose $n_k \uparrow$ to satisfy (a)–(b):

(a) For $j \in \{k, k+1\}$, $i = 1, \ldots, k$, and $m \geqq n_k$, $|\mathbf{R}_m(\mathbf{y}^j, g_i) - \nu_j(g_i)| < \varepsilon_k \|g_i\|$.

(b) For $m \geqq n_k$, $|\{\mathbf{i} \in V_m \setminus V_{n_k} : \mathbf{i} + V_{r_k} \not\subset V_m \setminus V_{n_k}\}| < \varepsilon_k |V_m|$.

Set $\mathbf{y}_\mathbf{i} = \mathbf{y}^k_\mathbf{i}$ for $\mathbf{i} \in V_{n_k} \setminus V_{n_{k-1}}$. Then $d(\mathbf{R}_m(\mathbf{y}), \nu) \leqq 2^{-k+2}$ whenever $n_k \leqq m < n_{k+1}$. $\square$

Call a parameter $\mathbf{y}$ *marginally quasiregular*, if the limit

$$\boldsymbol{\kappa}(\mathbf{y}) = \lim_{n \to \infty} \frac{1}{|V_n|} \sum_{\mathbf{i} \in V_n} \delta_{\mathbf{y}_\mathbf{i}}$$

exists in $\mathcal{M}_1(\mathcal{Y})$. In our present setting, it is clear that P-quasiregularity implies quasiregularity, p-quasiregularity implies marginal quasiregularity, the Σ-marginal of the expectation P is $\mathbf{K}(\mathbf{y})$, and the $\mathcal{Y}$-marginal of μ is $\boldsymbol{\kappa}(\mathbf{y})$.

3.3. Theorem. *Suppose $\mathbf{y}$ is p-regular and $U = \{\mathbf{0}\}$. Then*

$$K^{\mathbf{y}}(\nu) = \begin{cases} H(\nu \mid \mu) & \text{if } \nu = \boldsymbol{\kappa}(\mathbf{y}) \text{ on } \mathcal{B}(\mathcal{Y}) , \\ \infty & \text{otherwise} . \end{cases}$$

Proof. Lemma 2.19. Note that if $\mathbf{y}$ is p-regular and $U = \{\mathbf{0}\}$, then μ is the expectation of $y \mapsto p^y$ under $\boldsymbol{\kappa}(\mathbf{y})$. $\square$

3.4. Theorem. *Suppose that $\mathbf{y}$ is a P-regular parameter. Then the specific entropy*

$$h^U(Q \mid P) = \lim_{n \to \infty} \frac{1}{|V_n|} H_{\mathscr{E}(n, U)}(Q \mid P) \tag{3.5}$$

exists for $Q \in \mathcal{M}_T(\Omega)$ with marginal $\mathbf{K}(\mathbf{y})$. For all $Q \in \mathcal{M}_T(\Omega)$,

$$k^{\mathbf{y}}(Q) = \begin{cases} h^U(Q \mid P) & \text{if } Q = \mathbf{K}(\mathbf{y}) \text{ on } \mathcal{F}^Y , \\ \infty & \text{otherwise} . \end{cases} \tag{3.6}$$

Furthermore, $k^{\mathbf{y}}$ is affine on the set $\{Q \in \mathcal{M}_T(\Omega) : Q = \mathbf{K}(\mathbf{y}) \text{ on } \mathcal{F}^Y\}$.

Proof. A special case of the argument given in the proof of Theorem 4.4 below shows that $\mathscr{F}_n$ can be replaced by $\mathscr{E}_{n,U}$ in (2.9). By Lemma 2.19 and P-regularity,

$$K^{\mathbf{y}}_{\mathscr{E}(n,U)}(Q) = \begin{cases} H_{\mathscr{E}(n,U)}(Q\,|\,P) & \text{if } Q = \mathbf{K}(\mathbf{y}) \text{ on } \mathscr{F}^{Y}_{V(n)+U} , \\ \infty & \text{otherwise} . \end{cases} \tag{3.7}$$

This gives everything but the affinity statement, which follows as on p. 222 of Deuschel and Stroock (1989). □

3.8. Remark. As Theorem 2.13, the above proposition holds without the restriction on Σ-marginals provided $\mathbf{y}$ is transitive and we take $U = \mathbb{Z}^d$.

3.9. Example. P-regularity of $\tilde{\mathbf{y}}$ ensures that $P^{\mathbf{y}}$ is a conditional probability of P, given $Y = \mathbf{y}$, which in turn accounts for (3.7) and (3.6). Let us see how things can go wrong: Let $d = 1$, $\mathscr{X} = \{a, b\}$, and $\mathscr{Y} = [0, 1]$. Let $\mathbf{a}, \mathbf{b}, \mathbf{0}$, and $\mathbf{1}$ denote constant sequences of a's, b's, 0's, and 1's, respectively. For $\mathbf{y} = (\mathbf{y}_n) \in \Sigma$, let

$$\pi^{\mathbf{y}} = \begin{cases} \delta_a & \text{if } 0 \leqq \mathbf{y}_0 < 1 , \\ \delta_b & \text{if } \mathbf{y}_0 = 1 . \end{cases}$$

Let η_n be a sequence in (0, 1) converging to 1 as $n \to \infty$. Use Lemma 3.1 to pick a sequence $\mathbf{c} = (c_k)$ of 0's and 1's so that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \delta_{T^k \mathbf{c}} = \frac{1}{2}(\delta_0 + \delta_1) .$$

Put

$$\tilde{\mathbf{y}}_n = \begin{cases} \eta_n & \text{if } c_n = 0 , \\ 1 & \text{if } c_n = 1 . \end{cases}$$

Then $\tilde{\mathbf{y}}$ is P-quasiregular, $\Upsilon(\tilde{\mathbf{y}})$ gives mass 1/2 to $\delta_{(\mathbf{a},\mathbf{1})}$ and $\delta_{(\mathbf{b},\mathbf{1})}$ each, and $P = (\delta_{(\mathbf{a},\mathbf{1})} + \delta_{(\mathbf{b},\mathbf{1})})/2$. Let $Q = \delta_{(\mathbf{a},\mathbf{1})}$. Then $H_{\mathscr{E}(n)}(Q\,|\,P) = \log 2$ for all n but $k^{\tilde{\mathbf{y}}}(Q) = \infty$. This example shows also that, contrary to specific relative entropy, $k^{\tilde{\mathbf{y}}}$ is not necessarily affine.

We conclude this section with a Shannon–McMillan–Perez type result. Write $Q^{\mathbf{y}}$ for the conditional probability of $Q \in \mathscr{M}_1(\Omega)$, given $Y = \mathbf{y}$. It follows from (3.7) that if $k^{\tilde{\mathbf{y}}}(Q) < \infty$, then $Q \ll P$ on $\mathscr{E}_{n,U}$ for all n. Write

$$f_n = \left.\frac{dQ}{dP}\right|_{\mathscr{E}(n,U)}$$

for the Radon–Nikodym derivative.

3.10. Theorem. *Suppose that $\tilde{\mathbf{y}}$ is P-regular and transitive, and $k^{\tilde{\mathbf{y}}}(Q) < \infty$. Then for $\mathbf{K}(\tilde{\mathbf{y}})$-almost all $\mathbf{y}$ and in $L^1(\mathbf{K}(\tilde{\mathbf{y}}))$,*

$$k^{\tilde{\mathbf{y}}}(Q) = \lim_{n\to\infty} \frac{1}{|V_n|} \int \log f_n \, dQ^{\mathbf{y}} .$$

Proof. For finite rectangles $W \subset \mathbb{Z}^d$ define $F_W(\mathbf{y}) = H_{\mathscr{E}(W,U)}(Q^{\mathbf{y}}\,|\,P^{\mathbf{y}})$. $\{F_W\}$ is a nonnegative superadditive process, and the theorem follows from (3.6) and Theorems 6.2.3 and 6.2.9 in Krengel (1985). □

4 Large deviations under conditional probabilities

We return to the general setting of Sect. 1. For a sub-σ-field $\mathscr{D}$ of $\mathscr{F}$ set $\mathscr{A}_n = \bigvee\{T_{\mathbf{j}}\mathscr{D} : \mathbf{j} \in V_n\}$ and $\mathscr{A} = \bigvee_{n=1}^{\infty} \mathscr{A}_n$. Suppose $P \in \mathscr{M}_T(\Omega)$ is ergodic and has the property that, given $\mathscr{D}$, $Z_{\mathbf{0}}$ is independent of $\mathscr{F}_{\{\mathbf{0}\}^c} \vee \mathscr{A}$. Our object is to study the large deviations of $\mathbf{L}_n$ and $\mathbf{R}_n$ under conditional probabilities $P\{\cdot \mid \mathscr{A}\}$.

For probabilities Q defined on $\mathscr{F}_n$, set

$$K_n(Q) = \sup\{Qf - \int \log P\{e^f \mid \mathscr{A}_n\}\, dP\} \tag{4.1}$$

where the supremum is over bounded, $\mathscr{F}_n$-measurable functions. The limit

$$k(Q) = \lim_{n\to\infty} \frac{1}{|V_n|} K_n(Q) \tag{4.2}$$

exists for $Q \in \mathscr{M}_T(\Omega)$ by Theorem 2.8. For nonshift-invariant Q set $k(Q) = \infty$.

4.3. Theorem. *Both $\{\mathbf{R}_n\}$ and $\{\mathbf{L}_n\}$ satisfy a large deviation principle under almost every conditional probability $P\{\cdot \mid \mathscr{A}\}$, the former with rate k and the latter with rate K_1.*

Proof. Set $\Sigma = \Omega$ and let $\mathbf{z} \mapsto p^{\mathbf{z}}$ be a version of the conditional distribution of Z_0, given $\mathscr{D}$. Then $P^{\mathbf{z}}$ defined by (1.1) is a version of $P\{\cdot \mid \mathscr{A}\}$. Almost every $\mathbf{z}$ is P-regular in the sense of Definition 2.12 and satisfies $\mathbf{R}_n(\mathbf{z}) \to P$ as $n \to \infty$. Thus the theorem is an immediate consequence of Theorems 2.1 and 2.2. □

As usual, the rates are convex, good, and have the obvious unique zeroes. Note also that $K_n(Q) = \infty$ unless $Q = P$ on $\mathscr{A}_n \cap \mathscr{F}_n$. As in the skew system case, we can express the rates in terms of entropy:

4.4. Theorem. *Suppose $\mathscr{D} = \bigvee_{n=1}^{\infty} \mathscr{D} \cap \mathscr{F}_n$. Then the specific entropy*

$$h^{\mathscr{A}}(Q \mid P) = \lim_{n\to\infty} \frac{1}{|V_n|} H_{\mathscr{F}_n \vee \mathscr{A}_n}(Q \mid P) \tag{4.5}$$

exists for $Q \in \mathscr{M}_T(\Omega)$ which coincide with P on $\mathscr{A}$. We have

$$k(Q) = \begin{cases} h^{\mathscr{A}}(Q \mid P) & \text{if } Q \text{ is shift-invariant and } Q = P \text{ on } \mathscr{A}\ , \\ \infty & \text{otherwise}\ . \end{cases} \tag{4.6}$$

Assume additionally that $\mathscr{D} \subset \mathscr{F}_1$. Then for $v \in \mathscr{M}_1(\mathscr{X})$

$$K_1(v) = \begin{cases} H(v \mid P \circ Z_0^{-1}) & \text{if } v = P \text{ on } \mathscr{D}\ , \\ \infty & \text{otherwise}\ , \end{cases} \tag{4.7}$$

and $h^{\mathscr{A}}(Q \mid P)$ can be replaced by the usual specific entropy $h(Q \mid P)$ in (4.6).

Proof. Suppose $Q \in \mathscr{M}_T(\Omega)$ and $Q = P$ on $\mathscr{A}$. Let $K_n'(Q)$ be the quantity in (4.1) with the supremum taken over bounded, $\mathscr{F}_n \vee \mathscr{A}_n$-measurable functions. By arguments similar to those of Lemma 2.19, $K_n'(Q) = H_{\mathscr{F}_n \vee \mathscr{A}_n}(Q \mid P)$. Since $K_n'(Q) \geqq K_n(Q)$, to replace $K_n(Q)$ by $K_n'(Q)$ in (4.2) it suffices to show that

$$k(Q) \geqq \frac{1}{|V_m|}\{Qf - \int \log P\{e^f \mid \mathscr{A}_m\}\, dP\} \tag{4.8}$$

for an arbitrary bounded, $\mathscr{F}_m \vee (\mathscr{A}_m \cap \mathscr{F}_\ell)$-measurable function f for any integers ℓ and m. Here we used $\mathscr{A}_m = \bigvee_{n=1}^{\infty} \mathscr{A}_m \cap \mathscr{F}_n$. Let $k > 0$ and pick $n = n(k)$ so that V_n is a disjoint union of k^d shifted copies of V_m centered at sites in a set H_n. Set $g_n = \sum_{\mathbf{i} \in H_n} f \circ T_{\mathbf{i}}$. Then g_n is $\mathscr{F}_{n+\ell}$-measurable, and $P\{e^{g_n} | \mathscr{A}_{n+\ell}\} = \prod_{\mathbf{i} \in H_n} P\{e^f | \mathscr{A}_m\} \circ T_{\mathbf{i}}$. Hence

$$K_{n+\ell}(Q) \geqq Q g_n - \int \log P\{e^{g_n} | \mathscr{A}_{n+\ell}\}\, dP = k^d (Qf - \int \log P\{e^f | \mathscr{A}_m\}\, dP)\,.$$

Divide by $|V_{n+\ell}|$ and let $k \to \infty$ to get (4.8). The rest of the theorem is by now easy. □

Two settings which satisfy the assumptions of this section:

(1) Suppose $\mathscr{Z} = \mathscr{X} \times \mathscr{Y}$ as in Sect. 3, and P is an ergodic probability on Ω such that, for some $U \subset \mathbb{Z}^d$, $X_{\mathbf{0}}$ is independent of $(X_{\{\mathbf{0}\}^c}, Y_{U^c})$, given Y_U. P itself does not necessarily satisfy a large deviation principle if its marginal on Σ is an arbitrary ergodic measure (Orey and Pelikan 1988 give an example due to A. Sokal). Let $\mathscr{D}$ be the σ-field generated by $Y_{\{\mathbf{0}\} \cup U}$. According to Theorem 4.3, large deviation principles hold for almost all conditional probabilities $P\{\cdot | Y = \mathbf{y}\}$. For the case of a product measure $P = P_X \otimes P_Y$ with iid marginals, this result was first proved by Comets (1989) with arguments completely different from ours.

(2) Suppose that P is an iid probability on Ω, and that $\mathscr{D} \subset \mathscr{F}_1$. Then we get conditional versions of the well-known iid large deviation principles, and by Theorem 4.4, the usual entropy rates continue to apply on the set of probabilities with correct restrictions on the conditioning σ-field.

5 Proofs

We now turn to the proofs of Theorems 2.1 and 2.2. The upper bounds are established by a scheme codified in Deuschel and Stroock's (1989, Theorem 2.2.4): We first verify exponential tightness. It is then enough to prove the upper bound for compact sets. This we do by showing the existence of a pressure functional on a class of functions generating the weak topology of $\mathscr{M}_1(\Omega)$. Proof of the lower bound is based on geometric ideas developed by Baxter and Jain (1991).

First we enlarge the class $\{V_n\}$ so that it is closed under a partitioning of each set into two subsets of roughly equal cardinality. The need for this will become evident in the proof of the lower bound. Let $\mathscr{Q}$ denote the set of pairs (q, r) of dyadic rationals satisfying $0 \leqq q < r \leqq 1$ (q is a dyadic rational if $2^k q$ is an integer for some positive integer k). Let $n(q)$ denote the largest integer contained in $nq^{1/d}$. For positive integers n and $(q, r) \in \mathscr{Q}$, set $V_{n,q,r} = V_{n(r)} \setminus V_{n(q)}$. Let $V_0 = \emptyset$, so that $V_{n,0,1} = V_n$. If now $u = (q+r)/2$, then $V_{n,q,r} = V_{n,q,u} \cup V_{n,u,r}$, and

$$\text{(5.1)} \qquad \frac{|V_{n,q,u}|}{|V_{n,q,r}|} \to \frac{1}{2} \quad \text{and} \quad \frac{|V_{n,u,r}|}{|V_{n,q,r}|} \to \frac{1}{2} \quad \text{as } n \to \infty\,.$$

Proof of the next lemma is left to the reader.

5.2. Lemma. *Let $\alpha\colon \mathbb{Z}^d \to \mathbb{R}$ be a bounded function, and define*

$$\alpha(n, p, q) = \frac{1}{|V_{n,p,q}|} \sum_{\mathbf{i} \in V_{n,p,q}} \alpha(\mathbf{i})$$

for positive integers n and $(p, q) \in \mathscr{Q}$. If $\lim_{n\to\infty} \alpha(n, s, t) = \beta$ *for some* $(s, t) \in \mathscr{Q}$, *then* $\lim_{n\to\infty} \alpha(n, p, q) = \beta$ *for all* $(p, q) \in \mathscr{Q}$.

Write $\mathbf{L}_{n,q,r}$ and $\mathbf{R}_{n,q,r}$ for the empirical measures when V_n is replaced by $V_{n,q,r}$.

5.3. Proposition. *Suppose* $\mathbf{y} \in \Sigma$ *is p-quasiregular and* $(q, r) \in \mathscr{Q}$. *Then there are compact sets* $C_\ell \subset \mathscr{M}_1(\mathscr{X})$ *and* $L_\ell \subset \mathscr{M}_1(\Omega)$ *such that for all* $n, \ell \in \mathbb{N}$,

$$P^{\mathbf{y}}\{\mathbf{L}_{n,q,r} \in C_\ell^c\} \leqq e^{-|V_{n,q,r}|\ell} \tag{5.4}$$

and

$$P^{\mathbf{y}}\{\mathbf{R}_{n,q,r} \in L_\ell^c\} \leqq e^{-|V_{n,q,r}|\ell} . \tag{5.5}$$

Proof. Let $\varepsilon_k \downarrow 0$ as $k \to \infty$, and put $b_k = (k + 1 + \log 2)/\varepsilon_k$. By p-quasiregularity and Lemma 5.2, the measures

$$\Gamma_n = \frac{1}{|V_{n,q,r}|} \sum_{\mathbf{i} \in V_{n,q,r}} p^{T_{\mathbf{i}}\mathbf{y}}$$

converge in $\mathscr{M}_1(\mathscr{X})$ as $n \to \infty$, so we may pick compact sets $K_k \subset \mathscr{X}$ such that $\Gamma_n(K_k^c) < e^{-b_k}$ for all $n, k \in \mathbb{N}$. Then

$$\begin{aligned} P^{\mathbf{y}}\{\mathbf{L}_{n,q,r}(K_k^c) > \varepsilon_k\} &= P^{\mathbf{y}}\left\{\frac{1}{|V_{n,q,r}|} \sum_{\mathbf{i} \in V_{n,q,r}} \mathbf{1}_{K_k^c}(Z_{\mathbf{i}}) > \varepsilon_k\right\} \\ &\leqq \exp(-|V_{n,q,r}| b_k \varepsilon_k) \cdot \prod_{\mathbf{i} \in V_{n,q,r}} p^{T_{\mathbf{i}}\mathbf{y}}\{\exp(b_k \mathbf{1}_{K_k^c})\} . \end{aligned}$$

By Jensen's inequality and the choice of the K_k's,

$$\prod_{\mathbf{i} \in V_{n,q,r}} p^{T_{\mathbf{i}}\mathbf{y}}\{\exp(b_k \mathbf{1}_{K_k^c})\} \leqq (\Gamma_n\{\exp(b_k \mathbf{1}_{K_k^c})\})^{|V_{n,q,r}|} \leqq 2^{|V_{n,q,r}|} ,$$

so for all n and k,

$$P^{\mathbf{y}}\{\mathbf{L}_{n,q,r}(K_k^c) > \varepsilon_k\} \leqq e^{-|V_{n,q,r}|(k+1)} .$$

For $\ell \in \mathbb{N}$, let $C_\ell = \{\nu \in \mathscr{M}_1(\mathscr{X}) : \nu(K_k^c) \leqq \varepsilon_k \text{ for } k \geqq \ell\}$. The C_ℓ are convex, compact subsets of $\mathscr{M}_1(\mathscr{X})$, and it is easy to see that they satisfy (5.4). To get (5.5), use (5.4) to choose compact $K_{m,\mathbf{j}} \subset \mathscr{X}$ such that

$$p^{T_{\mathbf{i}}\mathbf{y}}\{\mathbf{L}_{n,q,r}(K_{m,\mathbf{j}}^c) > e^{-(m+|\mathbf{j}|+4d)}\} \leqq e^{-|V_{n,q,r}|(m+|\mathbf{j}|+4d)}$$

for all $m \in \mathbb{N}$ and $\mathbf{j} \in \mathbb{Z}^d$. Define

$$H_m = \{Q \in \mathscr{M}_1(\Omega) : Q(Z_{\mathbf{j}} \in K_{m,\mathbf{j}}^c) \leqq e^{-(m+|\mathbf{j}|+4d)} \text{ for all } \mathbf{j} \in \mathbb{Z}^d\}$$

and $L_\ell = \bigcap_{m \geqq \ell} H_m$. □

Let $\mathscr{C}_0$ be the class of bounded continuous functions on Ω which depend on only finitely many sites. It is a linear subspace of $C_b(\Omega)$ and generates the weak topology of $\mathscr{M}_1(\Omega)$. For $n \in \mathbb{N}$, $(q, r) \in \mathscr{Q}$, $\mathbf{y} \in \Sigma$, $f \in \mathscr{C}_0$, and $R \in \mathscr{M}_1(\Omega)$, let

$$\mathbf{S}_{n,q,r} f = \sum_{\mathbf{i} \in V_{n,q,r}} f \circ T_{\mathbf{i}} ,$$

$$c_{n,q,r}(\mathbf{y}, f) = \frac{1}{|V_{n,q,r}|} \log \int \exp(\mathbf{S}_{n,q,r} f)\, dP^{\mathbf{y}} ,$$

and

$$c_n(R,f) = \frac{1}{|V_n|}\log \int \exp(\mathbf{S}_n f)\, dR\,.$$

5.6. Proposition. *Let* $\mathbf{y}$ *be* P*-quasiregular,* $(q, r) \in \mathscr{Q}$*, and* $f \in \mathscr{C}_0$*. The limit*

$$c(\mathbf{y},f) = \lim_{n\to\infty} c_{n,q,r}(\mathbf{y},f) \tag{5.7}$$

exists and is independent of (q, r)*.* $c(\mathbf{y},f)$ *depends on* $\mathbf{y}$ *only through* $\Upsilon(\mathbf{y})$*, or more precisely*

$$c(\mathbf{y},f) = \lim_{n\to\infty} \int_{\mathscr{M}_1(\Omega)} c_n(R,f)\Upsilon(\mathbf{y}, dR)\,. \tag{5.8}$$

Proof. Take first $(q, r) = (0, 1)$. Usual partitioning arguments, as in the proof of Ledrappier's (1977, Lemma 7), give

$$\int c_m(R,f)\Upsilon(\mathbf{y}, dR) + O\left(\frac{1}{m}\right) \leqq \liminf_{n\to\infty} c_n(\mathbf{y},f) \leqq \limsup_{n\to\infty} c_n(\mathbf{y},f)$$
$$\leqq \int c_m(R,f)\Upsilon(\mathbf{y}, dR) + O\left(\frac{1}{m}\right),$$

which gives (5.7) for $(q, r) = (0, 1)$ and (5.8).

Now let $(q, r) \in \mathscr{Q}$ be arbitrary, suppose f is $\mathscr{F}_k$-measurable, and put $W_n = \{\mathbf{i} \in V_{n(q)} : \mathbf{i} + V_k \not\subset V_{n(q)}\} \cup \{\mathbf{i} \in V_{n,q,r} : \mathbf{i} + V_k \not\subset V_{n,q,r}\}$. $P^{\mathbf{y}}$'s independence and a straightforward approximation give

$$c_{n(r)}(\mathbf{y},f) = \frac{|V_{n(q)}|}{|V_{n(r)}|} c_{n(q)}(\mathbf{y},f) + \frac{|V_{n,q,r}|}{|V_{n(r)}|} c_{n,q,r}(\mathbf{y},f) + R(n)\,,$$

where $|R(n)| \leqq 2\|f\|\,|W_n|\,|V_{n(r)}|^{-1}$. Letting $n \to \infty$ gives (5.7). □

The functional $c(\mathbf{y},\cdot): \mathscr{C}_0 \to \mathbb{R}$ is called the *pressure* (Baxter and Jain 1991). The convex dual of the pressure is defined for P-quasiregular $\mathbf{y}$ and $Q \in \mathscr{M}_1(\Omega)$ by

$$c^*(\mathbf{y}, Q) = \sup\{Qf - c(\mathbf{y},f) : f \in \mathscr{C}_0\}\,.$$

Propositions 5.3 and 5.6 and Deuschel and Stroock's (1989, Theorem 2.2.4) combine to give

5.9. Proposition (Upper bound) *Suppose* $\mathbf{y}$ *is a* P*-quasiregular parameter and* $(q, r) \in \mathscr{Q}$*. Then for any closed set* $F \subset \mathscr{M}_1(\Omega)$*,*

$$\limsup_{n\to\infty} \frac{1}{|V_{n,q,r}|}\log P^{\mathbf{y}}\{\mathbf{R}_{n,q,r} \in F\} \leqq -c^*(\mathbf{y}, F)\,.$$

Before going on to the lower bound, we pause to show that $c^*(\mathbf{y},\cdot)$ agrees with $k^{\mathbf{y}}$. Theorem 2.8 then implies that $c^*(\mathbf{y},\cdot)$ has compact level sets, which we need for the proof of the lower bound.

5.10. Lemma. *Let* $\mathbf{y}$ *be* P*-quasiregular and* $f \in \mathscr{C}_0$*. If* f *is* $\mathscr{F}_k$*-measurable, then*

$$c(\mathbf{y},f) \leqq \frac{1}{|V_k|} \int_{\mathscr{M}_1(\Omega)} \log R\{\exp(|V_k| f)\}\,\Upsilon(\mathbf{y}, dR)\,.$$

Proof. Pick n so that V_n is the union of a collection of shifted copies of V_k. By Hölder's inequality and the independence properties of $P^{\mathbf{y}}$,

$$\int_\Omega \exp(\mathbf{S}_n f)\, dP^{\mathbf{y}} \leqq \prod_{\mathrm{i} \in V_n} \left(\int_\Omega \exp(|V_k| f \circ T_{\mathrm{i}})\, dP^{\mathbf{y}} \right)^{\frac{1}{|V(k)|}} .$$

Now take logarithms, divide by $|V_n|$, and let $n \to \infty$. □

5.11. Proposition. *$c^*(\mathbf{y}, Q) = k^{\mathbf{y}}(Q)$ for all P-quasiregular* $\mathbf{y}$ *and $Q \in \mathscr{M}_1(\Omega)$.*

Proof. If Q is not shift-invariant, then $c^*(\mathbf{y}, Q) = \infty$ because $c(\mathbf{y}, f - f \circ T_{\mathrm{i}}) = 0$.

Suppose Q is shift-invariant. Let f be an arbitrary $\mathscr{F}_k$-measurable bounded continuous function. Then

$$c^*(\mathbf{y}, Q) \geqq |V_k|^{-1} \{ Qf - \int \log R(e^f)\, \Upsilon(\mathbf{y}, dR) \}$$

by Lemma 5.10, so $c^*(\mathbf{y}, Q) \geqq k^{\mathbf{y}}(Q)$. The converse inequality follows from

$$\begin{aligned} K_{m+k}(Q) &\geqq Q(\mathbf{S}_m f) - \int \log R\{\exp(\mathbf{S}_m f)\}\, \Upsilon(\mathbf{y}, dR) \\ &= |V_m| \{ Qf - \int c_m(R, f)\, \Upsilon(\mathbf{y}, dR) \} . \quad \square \end{aligned}$$

5.12. Proposition (Lower bound) *For any P-quasiregular* $\mathbf{y}$, $(q, r) \in \mathscr{Q}$, *and open subset G of $\mathscr{M}_1(\Omega)$,*

$$\liminf_{n \to \infty} \frac{1}{|V_{n,q,r}|} \log P^{\mathbf{y}}\{\mathbf{R}_{n,q,r} \in G\} \geqq -k^{\mathbf{y}}(G) .$$

This proposition is proved after a sequence of lemmas. Fix a P-quasiregular $\mathbf{y}$, and for $Q \in \mathscr{M}_1(\Omega)$, write $k(Q) = k^{\mathbf{y}}(Q)$. For $(q, r) \in \mathscr{Q}$ and open $G \subset \mathscr{M}_1(\Omega)$, define

$$j(q, r, G) = -\liminf_{n \to \infty} \frac{1}{|V_{n,q,r}|} \log P^{\mathbf{y}}\{\mathbf{R}_{n,q,r} \in G\} ,$$

and then for $Q \in \mathscr{M}_1(\Omega)$,

$$J(Q) = \sup\{ j(q, r, G) : (q, r) \in \mathscr{Q},\ Q \in G,\ G \text{ is open} \} .$$

We will prove Proposition 5.12 for the fixed $\mathbf{y}$ by showing that $J(Q) \leqq k(Q)$ for all Q.

5.13. Lemma. *$J : \mathscr{M}_1(\Omega) \to [0, \infty]$ is lower semicontinuous and convex.*

Proof. Lower semicontinuity is immediate from the definition. Convexity is proved as on p. 9 of Lanford (1973), with the help of the partition property of the sets $\{V_{n,q,r}\}$. Suppose $Q = (Q' + Q'')/2$ in $\mathscr{M}_1(\Omega)$. Let G be an open neighborhood of Q. Metrize $\mathscr{M}_1(\Omega)$ with a metric of the type (3.2), and find $\varepsilon > 0$ so that the ball $B(Q, 3\varepsilon) \subset G$. Let $(q, r) \in \mathscr{Q}$, and put $s = (q + r)/2$. For large enough n, we can find measurable sets $D_{n,q,s}$ and $D_{n,s,r}$ depending on disjoint sets of sites and satisfying

$$\{\mathbf{R}_{n,q,s} \in B(Q', \varepsilon)\} \subset D_{n,q,s} \subset \{\mathbf{R}_{n,q,s} \in B(Q', 2\varepsilon)\}$$

and a similar statement with (q, s) and Q' replaced by (s, r) and Q''. Hence

$$P^{\mathbf{y}}\{\mathbf{R}_{n,q,s} \in B(Q', \varepsilon)\} \cdot P^{\mathbf{y}}\{\mathbf{R}_{n,s,r} \in B(Q'', \varepsilon)\} \leqq P^{\mathbf{y}}\{\mathbf{R}_{n,q,r} \in G\} ,$$

from which follow $j(q, r, G) \leqq (J(Q') + J(Q''))/2$ and convexity. □

Next a slightly modified version of Baxter and Jain's (1991, Theorem 4.8). For its proof equip the space $\mathscr{M}(\Omega)$ of real-valued Borel measures on Ω with the locally convex, Hausdorff linear topology generated by $\mathscr{C}_0$.

5.14. Lemma. *Let $g \in \mathscr{C}_0$ be such that $Rg \leqq k(R)$ for all $R \in \mathscr{M}_1(\Omega)$. Suppose Q is an extreme point of the set $A = \{R \in \mathscr{M}_1(\Omega): Rg = k(R)\}$. Then, for every open neighborhood G of Q and every $\varepsilon > 0$, there exists an open neighborhood U of Q contained in G, a function $f \in \mathscr{C}_0$, and a $\delta > 0$ such that $Qf = k(Q)$, $|Rf - Qf| < \varepsilon$ for $R \in U$, and $Rf \leqq k(R) - \delta$ for $R \in \mathscr{M}_1(\Omega) \setminus U$.*

Proof. Suppose first that $A = \{Q\}$. Pick a neighborhood U of Q such that $|Rg - Qg| < \varepsilon$ for $R \in U$. Since k has compact level sets, the number $\eta = \inf\{k(R) - Rg\colon k(R) \leqq \|g\| + 1, R \in U^c\}$ is positive, possibly infinite. Take $f = g$ and $\delta = \eta \wedge 1$ to conclude the lemma.

For the rest of the proof, assume that A contains more than one measure. It is a convex, compact subset of $\mathscr{M}_1(\Omega)$. Put

$$K = \{(R, t) \in \mathscr{M}_1(\Omega) \times \mathbb{R}\colon k(R) \leqq t \leqq \|g\| + 1\} .$$

K is a convex, compact subset of the locally convex linear space $\mathscr{M}(\Omega) \times \mathbb{R}$. Assume $\varepsilon < 1$. Let

$$U = \{R \in \mathscr{M}_1(\Omega)\colon |Rg - Qg| < \varepsilon/2\} ,$$

and

$$V = U \times (k(Q) - \varepsilon/2, k(Q) + \varepsilon/2) .$$

$(Q, k(Q))$ is an extreme point of K, and V is an open neighborhood of $(Q, k(Q))$. By the converse to the Krein–Milman theorem, $(Q, k(Q))$ does not lie in the closed convex hull of $K \setminus V$ (Dunford and Schwartz 1988, p. 440). Let L be the closed convex hull of $(\operatorname{epi} K) \setminus V$, where

$$\operatorname{epi} K = \{(R, t) \in \mathscr{M}_1(\Omega) \times \mathbb{R}\colon k(R) < \infty, t \geqq k(R)\} .$$

It is then easy to see that $(Q, k(Q))$ does not lie in L either. By Dunford and Schwartz's (1988, separating hyperplane theorem V.2.10), there are $h \in \mathscr{C}_0$, $s \in \mathbb{R}$, and $\eta > 0$ such that

$$Qh + sk(Q) = 0 \tag{5.15}$$

and

$$Rh + st < -\eta \quad \text{for all } (R, t) \in L . \tag{5.16}$$

We claim that $s < 0$. Suppose first that $s = 0$. Then $Qh = 0$, and by approaching Q inside A, we can find a measure R such that $k(R) < \infty$ and $-\eta < Rh$. This contradicts (5.16), for $(R, t) \in L$ for all large enough t. Thus $s \neq 0$, and then (5.16) forces $s < 0$. Putting $f' = -h/s$ and $\eta' = -\eta/s$, (5.15–5.16) turn into $Q(f') = k(Q)$ and

$$R(f') < t - \eta' \quad \text{for all } (R, t) \in L . \tag{5.17}$$

Noting that $\{R\} \times [k(R), \infty)$ is a subset of L for any $R \notin U$, (5.17) gives $R(f') \leqq k(R) - \eta'$ for all $R \notin U$. Let $\theta \in (0, 1)$ be such that $2\theta \| f' \| < \varepsilon$. Then $f = (1 - \theta) g + \theta f'$ and $\delta = \theta \eta'$ give the conclusion. □

The key step in the proof of the lower bound is Lemma 5.20. First two simple observations:

5.18. Lemma. *Suppose $r_n \uparrow \infty$, and $\{a_n\}$ and $\{b_n\}$ are nonnegative numbers such that $c = \lim_{n \to \infty} r_n^{-1} \log(a_n + b_n)$ exists. Then*

$$c \leqq \left(\liminf_{n \to \infty} \frac{1}{r_n} \log a_n \right) \vee \left(\limsup_{n \to \infty} \frac{1}{r_n} \log b_n \right).$$

5.19. Lemma. *Suppose $\mathbf{y}$ is P-quasiregular, $(q, r) \in \mathscr{Q}$, and the compact subsets $\{L_\ell\}_{\ell=1}^{\infty}$ of $\mathscr{M}_1(\Omega)$ satisfy (5.5). Let $f \in \mathscr{C}_0$. Then for $\ell > 2\|f\|$,*

$$c(\mathbf{y}, f) = \lim_{n \to \infty} \frac{1}{|V_{n,q,r}|} \log \int_{\mathbf{R}_{n,q,r} \in L_\ell} \exp(\mathbf{S}_{n,q,r} f)\, dP^{\mathbf{y}}.$$

Proof. Obvious from

$$\int_{\mathbf{R}_{n,q,r} \in L_\ell} \exp(\mathbf{S}_{n,q,r} f)\, dP^{\mathbf{y}} \leqq \int \exp(\mathbf{S}_{n,q,r} f)\, dP^{\mathbf{y}}$$

$$\leqq \int_{\mathbf{R}_{n,q,r} \in L_\ell} \exp(\mathbf{S}_{n,q,r} f)\, dP^{\mathbf{y}} + \exp(|V_{n,q,r}|(\|f\| - \ell)). \quad \square$$

5.20. Lemma. *Suppose $g \in \mathscr{C}_0$ is such that $Rg \leqq k(R)$ for all $R \in \mathscr{M}_1(\Omega)$. Then $J(Q) \leqq k(Q)$ for any extreme point Q of the set $\{R \in \mathscr{M}_1(\Omega) \colon Rg = k(R)\}$.*

Proof. Let G be an open neighborhood of Q, $(q, r) \in \mathscr{Q}$, and $\varepsilon > 0$. Let U, f, and δ come from Lemma 5.14. Let $L_\ell \subset \mathscr{M}_1(\Omega)$ be compact and satisfy (5.5). Pick $\ell > 2\|f\|$. For each $R \in L_\ell \setminus U$, find an open neighborhood H_R of R such that $R'f < Rf + \delta/4$ for all $R' \in H_R$, and

$$\inf\{k(R') \colon R' \in \bar{H}_R\} > Rf + \delta/2. \tag{5.21}$$

(5.21) can be satisfied because $k(R) \geqq Rf + \delta$ and k is lower semicontinuous. Pick a cover $H_{R_1}, \ldots, H_{R_m}$ for $L_\ell \setminus U$. Then

$$\limsup_{n \to \infty} \frac{1}{|V_{n,q,r}|} \log \int_{\mathbf{R}_{n,q,r} \in L_\ell \setminus U} \exp(\mathbf{S}_{n,q,r} f)\, dP^{\mathbf{y}}$$

$$\leqq \bigvee_{j=1}^{m} \left(\limsup_{n \to \infty} \frac{1}{|V_{n,q,r}|} \log P^{\mathbf{y}}\{\mathbf{R}_{n,q,r} \in H_{R_j}\} + R_j f + \delta/4 \right)$$

$$\leqq -\delta/4.$$

The last inequality comes by applying Proposition 5.9 on each H_{R_j} and then (5.21) for $R = R_j$. In the next calculation, use first Proposition 5.11, then Lemmas 5.19, 5.18, and 5.14, and the above inequality.

$$\begin{aligned} 0 &= Qf - k(Q) \leqq c(\mathbf{y}, f) \\ &= \lim_{n\to\infty} \frac{1}{|V_{n,q,r}|} \log \int_{\mathbf{R}_{n,q,r}\in L_\ell} \exp(\mathbf{S}_{n,q,r} f)\, dP^{\mathbf{y}} \\ &\leqq \left(\liminf_{n\to\infty} \frac{1}{|V_{n,q,r}|} \log \int_{\mathbf{R}_{n,q,r}\in U} \exp(\mathbf{S}_{n,q,r} f)\, dP^{\mathbf{y}} \right) \\ &\quad \vee \left(\limsup_{n\to\infty} \frac{1}{|V_{n,q,r}|} \log \int_{\mathbf{R}_{n,q,r}\in L_\ell \setminus U} \exp(\mathbf{S}_{n,q,r} f)\, dP^{\mathbf{y}} \right) \\ &\leqq \left(\liminf_{n\to\infty} \frac{1}{|V_{n,q,r}|} \log P^{\mathbf{y}}\{\mathbf{R}_{n,q,r}\in U\} + Qf + \varepsilon \right) \vee (-\delta/4)\,. \end{aligned}$$

From this it follows that

$$-k(Q) \leqq \liminf_{n\to\infty} \frac{1}{|V_{n,q,r}|} \log P^{\mathbf{y}}\{\mathbf{R}_{n,q,r}\in U\} + \varepsilon\,,$$

and consequently $k(Q) \geqq j(q, r, G) - \varepsilon$. Since G, (q, r), and ε were arbitrary, the lemma is proved. $\square$

Proof of Proposition 5.12 It is a consequence of the separating hyperplane theorems that the lower semicontinuous, convex function k can be written as

$$k(Q) = \sup\{Qf : f \in \mathscr{C}_0,\, Rf \leqq k(R) \text{ for all } R \in \mathscr{M}_1(\Omega)\}\,,$$

and a similar formula holds for J. If we can show that

$$Rf \leqq k(R) \quad \text{for all } R \in \mathscr{M}_1(\Omega) \tag{5.22}$$

whenever $f \in \mathscr{C}_0$ is such that $Rf \leqq J(R)$ for all $R \in \mathscr{M}_1(\Omega)$, we have $k \geqq J$ and the proposition. Find a number c such that $R(f + c) \leqq k(R)$ for all $R \in \mathscr{M}_1(\Omega)$ and the set $A = \{R \in \mathscr{M}_1(\Omega) \colon R(f + c) = k(R)\}$ is nonempty. A is convex and compact, and so has an extreme point Q. Then $c = k(Q) - Qf \geqq k(Q) - J(Q)$, so $c \geqq 0$ by Lemma 5.20, which implies (5.22). This completes the proof of Theorem 2.1. $\square$

The argument for the large deviation principle of Theorem 2.2 proceeds along the same lines except that it is simpler. We shall leave the details to the reader and concentrate on proving the converse, namely that existence of a large deviation principle implies p-quasiregularity.

5.23. Lemma. *Let $\mathscr{X}$ be any Polish space and Φ and Λ Borel probabilities on $\mathscr{M}_1(\mathscr{X})$. If*

$$\int_{\mathscr{M}_1(\mathscr{X})} \log \nu(e^g)\Phi(d\nu) = \int_{\mathscr{M}_1(\mathscr{X})} \log \nu(e^g)\Lambda(d\nu)$$

for all $g \in C_b(\mathscr{X})$, then $\Phi = \Lambda$.

Proof. By a compactification argument and the Stone–Weierstrass Theorem, it suffices to show that

$$\int_{\mathscr{M}_1(\mathscr{X})} e^{i(\nu h)}\Phi(d\nu) = \int_{\mathscr{M}_1(\mathscr{X})} e^{i(\nu h)}\Lambda(d\nu)$$

for all $h \in C_b(\mathscr{X})$. Via power series expansions, this in turn will follow from having

$$\int_{\mathscr{M}_1(\mathscr{X})} (vh)^k \Phi(dv) = \int_{\mathscr{M}_1(\mathscr{X})} (vh)^k \Lambda(dv) \tag{5.24}$$

for all $h \in C_b(\mathscr{X})$ and positive integers k. Multiply h by a constant so that $\|h\| < 1$. Let $z \in (-1, 1)$ and $g = \log(1 + zh)$. Then

$$\begin{aligned}\int_{\mathscr{M}_1(\mathscr{X})} \log v(e^g)\Phi(dv) &= \int_{\mathscr{M}_1(\mathscr{X})} \log(1 + zv(h))\Phi(dv) \\ &= -\sum_{k=1}^{\infty} \frac{(-z)^k}{k} \int (vh)^k \Phi(dv)\,.\end{aligned}$$

The last expression is an analytic function of z on $(-1, 1)$, so the coefficients of its power series are determined uniquely. The hypothesis of the lemma now gives (5.24). □

Fix $\mathbf{y}$ and let I be a rate function on $\mathscr{M}_1(\mathscr{X})$ governing the large deviations of $\{\mathbf{L}_n\}$ under $P^{\mathbf{y}}$. Assume I is lower semicontinuous by Proposition 1.1 from Orey (1986). Let $\bar{\mathscr{X}}$ be the compact completion of $\mathscr{X}$ under a totally bounded metric. Think of $p^{\mathbf{y}}$ and $\mathbf{p}_n(\mathbf{y})$ as probabilities on $\bar{\mathscr{X}}$ and $\mathscr{M}_1(\bar{\mathscr{X}})$, respectively. One sees easily that the large deviations of $\{\mathbf{L}_n\}$ on $\mathscr{M}_1(\bar{\mathscr{X}})$ are governed by the lower semicontinuous rate J defined by

$$J(v) = \begin{cases} I(v)\,, & \text{if } v \in \mathscr{M}_1(\mathscr{X})\,, \\ \liminf_{\mathscr{M}_1(\mathscr{X}) \ni \eta \to v} I(\eta) & \text{if } v \in \mathscr{M}_1(\bar{\mathscr{X}}) \setminus \mathscr{M}_1(\mathscr{X})\,. \end{cases}$$

By Varadhan's Theorem, the pressure exists for all $f \in C(\bar{\mathscr{X}})$ and is the convex dual of the rate:

$$\begin{aligned} c(\mathbf{y}, f) &= \lim_{n\to\infty} \frac{1}{|V_n|} \sum_{\mathbf{i} \in V_n} \log \int e^f \, dp^{T_{\mathbf{i}}\mathbf{y}} \\ &= \sup\{vf - J(v) \colon v \in \mathscr{M}_1(\bar{\mathscr{X}})\}\,. \end{aligned} \tag{5.25}$$

Suppose Φ and Λ are limit points of $\{\mathbf{p}_n(\mathbf{y})\}$ in $\mathscr{M}_1(\mathscr{M}_1(\bar{\mathscr{X}}))$. Pass to the limit in (5.25) along suitable subsequences to get

$$\int \log v(e^f)\Phi(dv) = c(\mathbf{y}, f) = \int \log v(e^f)\Lambda(dv) \tag{5.26}$$

for all $f \in C(\bar{\mathscr{X}})$. Thus $\Phi = \Lambda$ by Lemma 5.23, and we conclude that $\mathbf{p}_n(\mathbf{y}) \to \Phi$ as $n \to \infty$.

Let $\mu \in \mathscr{M}_1(\bar{\mathscr{X}})$ be the expectation of Φ. Since $\mathbf{y}$ is p-quasiregular on $\mathscr{M}_1(\bar{\mathscr{X}})$ and a lower semicontinuous rate is unique, $J(v) \geqq H(v \mid \mu)$ by Theorem 2.6. It follows that $\mu(\mathscr{X}) > 0$, for otherwise $I(\eta) = J(\eta) = \infty$ for all $\eta \in \mathscr{M}_1(\mathscr{X})$. Define $\pi \in \mathscr{M}_1(\mathscr{X})$ by $\pi(E) = \mu(E \cap \mathscr{X})/\mu(\mathscr{X})$ and conclude that $I(\eta) \geqq H(\eta \mid \pi) - \log \mu(\mathscr{X})$ for $\eta \in \mathscr{M}_1(\mathscr{X})$. This implies that I has compact level sets in $\mathscr{M}_1(\mathscr{X})$, and consequently $J \equiv \infty$ on $\mathscr{M}_1(\bar{\mathscr{X}}) \setminus \mathscr{M}_1(\mathscr{X})$.

We need to show that Φ is a probability on $\mathscr{M}_1(\mathscr{X})$. Since $\mathscr{X}$ is Polish, there are open $G_j \subset \bar{\mathscr{X}}$ such that $G_j \downarrow \mathscr{X}$. Let $B_j = \{v \in \mathscr{M}_1(\bar{\mathscr{X}}) \colon v(G_j^c) \geqq 1/j\}$. It suffices to show that $\Phi(B_j) = 0$ for all j.

Let $M > 0$ be arbitrary. Let K be a compact subset of $\mathscr{X}$ such that $\eta(K^c) \leqq 1/M$ whenever $I(\eta) \leqq 2M$. Pick $f \in C(\bar{\mathscr{X}})$ so that $0 \leqq f \leqq M$, $f \equiv 0$ on K, and $f \equiv M$ on

G_j^c. Since $c(\mathbf{y},f) \geqq -M$, (5.25) turns into

$$c(\mathbf{y},f) = \sup\{\eta f - I(\eta) \colon \eta \in \mathscr{M}_1(\mathscr{X}), I(\eta) \leqq 2M\} \, ,$$

so by the choice of K and f, $c(\mathbf{y},f) \leqq 1$. On the other hand, $(M - \log j)\Phi(B_j) \leqq c(\mathbf{y},f) \leqq 1$ by (5.26), so letting $M \uparrow \infty$ forces $\Phi(B_j) = 0$. This completes the proof of Theorem 2.2.

Acknowledgement. This paper benefited considerably from the valuable comments by an anonymous referee.

References

Baxter, J.R., Jain, N.C.: Convexity and compactness in large deviation theory. (Preprint 1991)
Baxter, J.R., Jain, N.C., Seppäläinen, T.O.: Large deviations for nonstationary arrays and sequences. Ill. J. Math. (to appear)
Comets, F.: Large deviation estimates for a conditional probability distribution. Applications to random interaction Gibbs measures. Probab. Theory Relat. Fields. **80**, 407–432 (1989)
Deuschel, J.-D., Stroock, D.W.: Large deviations. San Diego: Academic Press 1989
Dunford, N., Schwartz, J.T.: Linear operators, part I: General theory. New York: Wiley 1988
Dynkin, E.B.: The initial and final behaviour of the trajectories of Markov processes. Russ. Math. Surv. **26**, 165–185 (1971).
Föllmer, H., Orey, S.: Large deviations for the empirical field of a Gibbs measure. Ann. Probab. **16**, 961–977 (1988)
Krengel, U.: Ergodic theorems. Berlin: de Gruyter 1985
Lanford, O.E.: Statistical mechanics and mathematical problems. (Lect. Notes Phys., vol. 20, pp. 1–113) Berlin Heidelberg New York: Springer 1973
Ledrappier, F.: Pressure and variational principle for random Ising model. Commun. Math. Phys. **56**, 297–302 (1977)
Orey, S.: Large deviations in ergodic theory. In: Çinlar, E., Chung, K-L., Getoor, R. (eds.) Seminar on stochastic processes, pp. 195–249. Boston: Birkhäuser 1986
Oxtoby, J.C.: Ergodic sets. Bull. **58**, 116–132 (1952)
Seppäläinen, T.: Large deviations for lattice systems. II. Nonstationary independent fields. Probab. Theory Relat. Fields (to appear)
Varadhan, S.R.S.: Large deviations and applications. Philadelphia: SIAM 1984