Large deviations for cascades and cascades of large deviations

Alain Rouault *

February 4, 2004

Abstract. In a Mandelbrot's multiplicative cascade on $[0,1]$, let r be the number of cells, and Z_r^n the mass of the measure at the height n in the r-ary tree. Most of the known results deal with the limit $n \to \infty$ with r fixed. In some previous papers, we began the study of $r \to \infty$ when n is fixed, showing a law of large numbers and describing a new large deviations phenomenon when the cascade generator is bounded. Actually, there are self-similar rate functions at geometrical scales, reflecting the multiplicative structure of the process. Here we describe the behavior of a typical cascade, knowing that Z_r^n is given different from its mean 1. The main result is a Sanov type theorem, at different rates, leading to a Gibbs conditioning principle.

Key words. Multiplicative cascades, large deviations, Sanov heorem, Gibbs conditional principle.

AMS 2000 subject classification. Primary 60F10.

1 Introduction

Cascades are a very popular mathematical model for rainfall, internet packet traffic, market prices, etc. The literature is now very broad since the foundation (Kolmogorov [12], Mandelbrot [17], until more recent improvements. Extensive bibliographies are in Barral [1] or Liu [13], and Ossiander and Waymire [18] for statistical aspects.

Let $W \geq 0$ a random variable of mean 1. The *cascade generators* are given by a family $\{W_{\mathbf{i}}\}$ of independent copies of W, indexed by the set of all finite sequences $\mathbf{i} = i_1 \cdots i_n, n \geq 1$ of positive integers. We are interested in the random variable

$$Z_r^n = r^{-n} \sum_{1 \leq i_1, \ldots, i_n \leq r} W_{i_1} W_{i_1 i_2} \cdots W_{i_1 \cdots i_n}$$

which is the mass of the cascade at height n. Let $\mathcal{F}_n$ be the σ-field generated by all the $W_{\mathbf{i}}$ for those $\mathbf{i}$ of height less or equal to n. Then for r fixed, $\{Z_r^n\}_n$ is a $\{\mathcal{F}_n\}_n$ nonnegative martingale and we denote by Z_r^∞ its a.s limit when $n \to \infty$. It is a.s. stricly positive as soon as $\log r > EW \log W$ (Kahane-Peyrière [11]).

Another asymptotics consists in letting $r \to \infty$ with n fixed (finite or infinite). For $n = 1$ it is the classical sum of i.i.d. random variables. For $n \geq 2$, the summands are dependent: the terms corresponding to $\mathbf{i}$ and $\mathbf{j}$ have in their products a number of common random variables equal to the height of the last common ancestor $\mathbf{i} \wedge \mathbf{j}$ of $\mathbf{i}$ and $\mathbf{j}$. Notice that it is also posiible to write

$$Z_r^n = r^{-1} \sum_{k=1}^r W_k \left(Z_r^{n-1} \circ \Theta_k\right) \quad \text{a.s.},$$

*LAMA CNRS UMR8100, Bâtiment Fermat, Université de Versailles, 78035 Versailles, France. e-mail: rouault@math.uvsq.fr

where Θ_k is the tree shifted at the node of label k, and Z_r^n appears as a sum of r i.i.d. random variables, with a distribution depending on r. In [15], [14], Liu, Rio and Rouault extended results known for $n = 1$, in particular the law of large numbers. For $n \leq \infty$ fixed, they proved that

$$\lim_{r\to\infty} Z_r^n = 1 \quad \text{a.s.}$$

(remember $EW = 1$). It means roughly that only the first level contributes to the limit. It is also true for other types of results (CLT, LIL).

In the study of large deviations, a new phenomenon occurs. In [15] it is proved that the family of distributions of Z_r^∞ satisfies the following tail estimates, at different rates, under the assumption

$$\overline{w} := \text{ess sup } W < \infty\,. \tag{1.1}$$

Roughly, every level contributes to large deviations, at an appropriate scale and appropriate rate, with a kind of self-similarity, justifying the title of the present paper.

Theorem 1.1 *([15]) If assumption (1.1) holds, then for any $k \geq 0$ and any $x \in (\overline{w}^k, \overline{w}^{k+1}]$,*

$$\lim_{r\to\infty} r^{-(k+1)} \log \mathbb{P}(Z_r^\infty \geq x) = -\Lambda^*(x\overline{w}^{-k})\,,$$

where

$$\Lambda^*(y) = \sup_\theta \theta y - \Lambda(\theta) \quad \textit{and} \quad \Lambda(\theta) = \log \mathbb{E}e^{\theta W}\,.$$

Actually, at intermediate levels (n fixed finite) we can consider the weaker assumption

$$\forall \tau > 0 \qquad \mathbb{E}\exp \tau W < \infty \tag{1.2}$$

and state the following result [1].

Proposition 1.2 *i) If assumption (1.2) holds, then for any $x > 1$*

$$\lim_{r\to\infty} r^{-1} \log \mathbb{P}(Z_r^n \geq x) = -\Lambda^*(x)\,.$$

ii) If assumption (1.1) holds, then for any $0 \leq k < n$ and $x \in (\overline{w}^k, \overline{w}^{k+1}]$,

$$\lim_{r\to\infty} r^{-(k+1)} \log \mathbb{P}(Z_r^n \geq x) = -\Lambda^*(x\overline{w}^{-k}).$$

It should be clear that the same results hold when one considers left tails, i.e. $\mathbb{P}(Z_r^n \leq x)$ for $x < 1$. For the analog of i) we put no assumption, and for ii) we assume $\underline{w} := \text{ess inf } W > 0$.

A quite natural question is then the description of a typical cascade of branching number r (large), of mass approximately $a \neq 1$. We use large deviations techniques for which we recall some definitions.

Let $a_r, r \in \mathbb{N}$ an increasing sequence of positive real numbers with $\lim_{r\to\infty} a_r = \infty$. We say that a sequence of probability measures $(\mathcal{P}_r)$ on a regular Hausdorff space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ satisfies a Large Deviation Principle (LDP) with rate function I at scale a_r if I is a lower semicontinuous function $I : \mathcal{X} \to [0, \infty]$ such that

a) For any closed subset F of $\mathcal{X}$

$$\limsup_{r\to\infty} \frac{1}{a_r} \log \mathcal{P}_r(F) \leq -\inf_{x\in F} I(x)$$

[1] Part i) is a consequence of [14]. The proof of ii) is similar to the proof of [15] Theorem 1.4. We could also use [16].

b) For any open subset G of $\mathcal{X}$

$$\limsup_{r\to\infty} \frac{1}{a_r} \log \mathcal{P}_r(G) \leq - \inf_{x\in G} I(x)\,.$$

The rate function is good if for every c the level set $\{x : I(x) \leq c\}$ is compact. The sequence $(\mathcal{P}_r)$ is said to be exponentially tight at scale a_r if for every $L > 0$ there is a compact set K_L such that $\limsup_r a_r \log \mathcal{P}_r(K_L^c) < -L$.

As a first answer to the question of typical cascade, our main result (Theorem 4.2) is a Gibbs conditioning principle. If the cascade generator is bounded (with maximum $\bar{w}$), then for $a \in (\bar{w}^{k-1}, \bar{w}^k)$, and $i_1, ...i_n$ fixed, the distribution of $(W_{i_1}, \cdots W_{i_1\cdots i_n})$ conditioned on $Z_r^n \in [a-\delta, a+\delta]$ converges weakly to $\delta_{\bar{w}}^{k-1} \otimes Q^\theta \otimes Q^{n-k}$ as $r \to \infty$ followed by $\delta \to 0$. Here δ_c denotes the Dirac mass at c, Q^θ is obtained by tilting Q to get mean $a\bar{w}^{-(k-1)}$ instead of mean 1. It is a consequence of a Sanov type theorem (Theorem 2.1) giving the LDP for empirical distributions:

$$\mathcal{L}_r^n = \frac{1}{r^n} \sum_{1\leq i_1,\ldots,i_n\leq r} \delta_{W_{i_1}\,\ldots,W_{i_1\ldots i_n}}\,. \tag{1.3}$$

This empirical distribution is a random measure on $[0,\infty)^n$. In Section 3, we use the contraction principle to study again the masses Z_r^n, and in Section 4, we deduce the typical behavior along a fixed branch. It would be interesting to have a more general Gibbs principle, taking into account the limiting distribution along a fixed number (> 1) of branches, like in Corollary 7.3.5 of [4].

The structure of our problem[2] leads us to conditioning at stage n upon the σ-field $\mathcal{F}_{n-1}$, and use induction. In particular, we have to argue about conditional LDP, or in other words, we have to manage mixture of distributions which satisfy LDP. Various authors have studied this topic, Dinwoodie-Zabell [6] for exchangeable vectors (and more recently Trashorras [20]), Chaganthy [3] for statistical purposes, Grunwald for statistical mechanics ([10] section 2), Finnoff for evolutionary games [9], and Biggins for random graphs in [2]. Here, we have tried an (almost) self-contained treatment of this question adapted to our model.

Let us give some notations used in the sequel. First, Σ is a Polish space in Section 2, and is $\mathbb{R}^+$ in the end of Section 3 (Corollary 3.2) and in Section 4. We denote by $bm(\Sigma)$ the set of bounded Borel functions from Σ to $\mathbb{R}$. Let $\mathcal{M}_1(\Sigma)$ be the set of probabilities on Σ equipped with the weak convergence denoted by $\Rightarrow$. If $f \in bm(\Sigma)$ and $\nu \in \mathcal{M}_1(\Sigma)$ we denote as usual,

$$< f, \nu > = \int_\Sigma f d\nu\,.$$

The relative entropy or Kullback information is defined by:

$$H(\alpha\|\beta) = \int_\Sigma \left(\log \frac{d\alpha}{d\beta}\right) d\alpha$$

whenever $\alpha \in \mathcal{M}_1(\Sigma)$ is absolutely continuous with respect to β. Otherwise $H(\alpha\|\beta) = +\infty$ (see [4] Appendix D3). We will use the following variational formula ([7] Lemma 1.4.3 p.36):

$$H(\alpha\|\beta) = \sup\{< \varphi, \alpha > - \log < e^\varphi, \beta > :\ \varphi \in \mathcal{C}_b(\Sigma)\} \tag{1.4}$$

For $\nu \in \mathcal{M}_1(\Sigma^n)$ and $1 \leq j \leq n$, let ν_j or $\pi_j\nu$ be the projection of ν on the first j coordinates.

The random variable W is Σ valued, and its distribution is denoted by Q. In Section 3 and 4, we assume that $\mathbb{E}W = 1$. For $j \geq 1$ let $Q^j = Q^{\otimes j}$ (and $\nu \otimes Q^0 = \nu$ for every measure ν).

For $f : \Sigma^n \to \mathbb{R}$ let $\pi f = \mathbb{E} f(\cdot, W)$ or in other words $\pi f(\mathbf{x}) = \int_\Sigma f(\mathbf{x}, y) Q(dy)$.

[2]This model involves a dependence between strings of variables different from that of U-empirical measures (see [8]).

2 A Sanov type theorem

In the whole section we fix $n < \infty$. We consider $\mathcal{L}_r^n$ as (random) element of $\mathcal{M}_1(\Sigma^n)$ equipped with the weak topology. Let S be the support of Q.

Theorem 2.1 *a) The family $\{\mathbb{P}(\mathcal{L}_r^n \in .)\}_r$ satisfies the LDP at scale r with good rate function*

$$\begin{aligned} \mathcal{I}_1^n(\nu) &= H(\nu_1 \parallel Q) \text{ if } \nu = \nu_1 \otimes Q^{n-1} \\ &= +\infty \quad \text{otherwise}. \end{aligned} \tag{2.1}$$

b) If S is compact, then for every $2 \le k \le n$, the family $\{\mathbb{P}(\mathcal{L}_r^n \in .)\}_r$ satisfies the LDP in $\mathcal{M}_1(S^n)$ at scale r^k with rate function

$$\begin{aligned} \mathcal{I}_k^n(\nu) &= H(\nu_k \parallel \nu_{k-1} \otimes Q) \text{ if } \nu = \nu_k \otimes Q^{n-k} \\ &= +\infty \qquad \text{otherwise}. \end{aligned} \tag{2.2}$$

As usual, we need exponential tightness.

Lemma 2.2 *a) For $n \ge 1$, the family $\{\mathbb{P}(\mathcal{L}_r^n \in .)\}_r$ is exponentially tight at scale r.*

b) For $n \ge 2$, and for every sequence $\{\xi_r\}_r \in \mathcal{M}_1(\Sigma^n)$ such that $\xi_r \Rightarrow \xi$, the family $\{\mathbb{P}\left(\mathcal{L}_r^n \in \cdot \,|\, \mathcal{L}_r^{n-1} = \xi_r\right)\}_r$ is exponentially tight at scale r^n.

To prove this lemma, we use a slightly modified version of the criteria of exponential tightness of Deuschel-Stroock Lemma 3.2.7 p.67 of [5].

Lemma 2.3 *If $(\gamma_r)_r$ is a family of random probabilities on a Polish space $\mathcal{Y}$, then the family $\{\mathbb{P}(\gamma_r \in \cdot)\}_r$ is exponentially tight at scale a_r as soon as there is a tight family of probabilities $(\rho_r)_r$ such that for every $\varphi \in bm(\mathcal{Y})$*

$$\mathbb{E} \exp < \varphi, a_r \gamma_r > \le < \exp \varphi, \rho_r >^{a_r} . \tag{2.3}$$

Proof of Lemma 2.3 We give a construction similar to that of Deuschel-Stroock's (see also Dembo-Zeitouni's Lemma 6.2.6 [4]). For every $L \ge 1$, we have to find a compact set $C_L \in \mathcal{M}_1(S^n)$ such that for every $r \ge 1$

$$\mathbb{P}(\gamma_r \notin C_L) \le 2 \exp -L a_r . \tag{2.4}$$

For $\ell \ge 1$ let $\epsilon_\ell = e^{-\ell^2} 2^{-\ell}$. Since $\{\rho_r\}_r$ is tight, one can find a compact K_ℓ such that $\rho_r(K_\ell) \ge 1 - \epsilon_\ell$ for every r. We claim that

$$C_L = \cap_{\ell \ge L} \left\{ \nu : \nu(K_\ell) \ge 1 - \ell^{-1} \right\} \tag{2.5}$$

suits. We have for $f_\ell = \ell^2 + \ell \log 2$ on K_ℓ^c and 0 on K_ℓ

$$\mathbb{P}(\gamma_r(K_\ell) < 1 - \ell^{-1}) \le \mathbb{E}(\exp < f_\ell, a_r \gamma_r >) \left(2 e^\ell\right)^{-a_r} . \tag{2.6}$$

Now, from (2.3)

$$\mathbb{E}(\exp < f_\ell, a_r \gamma_r >) \ \le < \exp f_\ell, \rho_r >^{a_r}$$

and from the definition of K_ℓ,

$$< \exp f_\ell, \rho_r > = \rho_r(K_\ell) + e^{\ell^2} 2^\ell (1 - \rho_r(K_\ell)) \ \le \ 2 .$$

With the three last displays and the union of events bound we get $\mathbb{P}(\gamma_r \notin C_L) \le \sum_{\ell \ge L} e^{-\ell r^n}$, which yields (2.4). ■

<u>Proof of Lemma 2.2</u> In a) and b) we will assume that S is unbounded (otherwise it is trivial). Let us begin with a display often used in the sequel. If $h \in bm(\Sigma^n)$, we have

$$\mathbb{E}\big(\exp < r^n h, \mathcal{L}_r^n > \mid \mathcal{F}_{n-1}\big) = \exp < r^n \tilde{h}, \mathcal{L}_r^{n-1} > , \tag{2.7}$$

where $\tilde{h}(x_1, x_2, ..., x_{n-1}) = \log \mathbb{E} \exp h(x_1, x_2, ..., x_{n-1}, W)$, and in particular

$$r^{-n} \log \mathbb{E}\left(\exp < r^n h, \mathcal{L}_r^n > \mid \mathcal{L}_r^{n-1} = \mu\right) = < \log \mathbb{E} \exp h(\cdot, W), \mu > \ . \tag{2.8}$$

a) If $n = 1$ it is classical. For $n \ge 2$ and $\varphi \in bm(\Sigma^n)$ we have from (2.7)

$$\mathbb{E} \exp < \varphi, r\mathcal{L}_r^n >= \mathbb{E} \exp < \tilde{\varphi}, r\mathcal{L}_r^{n-1} >$$

with $\tilde{\varphi}(x) = r^{n-1} \log \mathbb{E} \exp \left(r^{1-n}\ \varphi(x, W)\right)$, hence by induction and Jensen inequality

$$\mathbb{E} \exp < \varphi, r\mathcal{L}_r^n > \le \ \{\mathbb{E} \exp \varphi(W_1, .., W_n)\}^r \ ,$$

which is (2.3) with $\rho_r = Q^{\otimes n}$, $a_r = r$.

b) From (2.8) and Jensen's inequality,

$$\mathbb{E}\left(\exp < \varphi, r^n \mathcal{L}_r^n > \mid \mathcal{L}_r^{n-1} = \xi_r\right) \le < \exp \varphi,\ \xi_r \otimes Q >^{r^n} \ .$$

If $\xi_r \Rightarrow \xi$, the family $(\xi_r \otimes Q)_r$ is tight, so we have (2.3) with $\rho = \xi_r \otimes Q$, $a_r = r^n$. ■

<u>Proof of Theorem 2.1</u>: Both following claims will be used in the sequel. For $g \in bm(\Sigma)$

$$\lim_{a \to 0} a^{-1} \log \mathbb{E} \exp ag(W) \ = \ \mathbb{E} g(W) \tag{2.9}$$

$$\lim_{a \to \infty} a^{-1} \log \mathbb{E} \exp ag(W) \ = \ \operatorname{essup} g(W) \, . \tag{2.10}$$

For convenience, we begin by the case $k = n$, consider conditional LDP and make an induction.

Let $\xi_r \Rightarrow \xi$. We first claim that the family $\{\mathbb{P}\left(\mathcal{L}_r^n \in \cdot \mid \mathcal{L}_r^{n-1} = \xi_r\right)\}_r$ satisfies the LDP at scale r^n with good rate function

$$\mathcal{J}_\xi(\nu) := \sup \{< f, \nu > - < \log \mathbb{E} \exp f(\cdot, W)\ ,\ \xi > \ ; \ \ f \in bm(\Sigma^n)\} \ . \tag{2.11}$$

It is actually a consequence of Corollary 4.5.27 of Baldi's theorem ([4] p.160) since from the one hand Lemma 2.2 b) gives exponential tightness and from the other hand (2.8) gives

$$\lim_r\ r^{-n} \log \mathbb{E}\left(\exp < r^n f, \mathcal{L}_r^n > \mid \mathcal{L}_r^{n-1} = \xi_r\right) =< \log \mathbb{E} \exp f(\cdot, W)\ ,\ \xi > \ .$$

This implies a uniform conditional LDP, i.e. that for every closed set F (resp. open G), every $\epsilon > 0$ and every μ one can find an open neighborhood V_μ of μ such that

$$\limsup_r \sup_{\xi \in V_\mu} r^{-n} \log \mathbb{P}\left(\mathcal{L}_r^n \in F \mid \mathcal{L}_r^{n-1} = \xi\right) \ \le \ -\inf_{\nu \in F} \mathcal{J}_\mu(\nu) + \epsilon \tag{2.12}$$

$$\liminf_r \inf_{\xi \in V_\mu} r^{-n} \log \mathbb{P}\left(\mathcal{L}_r^n \in G \mid \mathcal{L}_r^{n-1} = \xi\right) \ \ge \ -\inf_{\nu \in G} \mathcal{J}_\mu(\nu) - \epsilon \tag{2.13}$$

Let us give another (non variational) expression of $\mathcal{J}_\xi$. If f involves only the first $n-1$ coordinates, i.e. $f(\mathbf{x},y)=\psi(\mathbf{x})$ then

$$< f,\nu > - < \log \mathbb{E} \exp f(\cdot, W)\ ,\ \xi > = < \psi, \nu_{n-1} > - < \psi, , \xi > ,$$

and the supremum taken on ψ is infinite if $\nu_{n-1} \neq \xi$. So

$$\mathcal{J}_\xi(\nu) = +\infty \ \text{ if } \nu_{n-1} \neq \xi\,. \tag{2.14}$$

Now if $\nu_{n-1}=\xi$ we may write $\nu = \xi \otimes \bar\nu$ where $\bar\nu$ is a regular probability kernel so that

$$\mathcal{J}_\xi(\nu) = \sup_f \int_{\Sigma^{n-1}} \left(\int_\Sigma f(\mathbf{x},y)\bar\nu(dy|\mathbf{x}) - \log \mathbb{E}\exp f(\mathbf{x}, W) \right) \xi(d\mathbf{x}).$$

Applying (1.4) with $\alpha = \bar\nu(\cdot|\mathbf{x})$ and $\beta = Q$, we get

$$\mathcal{J}_\xi(\nu) \leq \int_{\Sigma^{n-1}} H\left(\bar\nu(.|\mathbf{x}) \parallel Q\right)\xi(d\mathbf{x}) = H\left(\xi\otimes\bar\nu \parallel \xi \otimes Q\right)\,. \tag{2.15}$$

Applying again (1.4) but with $\alpha = \xi\otimes\bar\nu = \nu$ and $\beta = \xi\otimes Q$ we obtain

$$H\left(\xi\otimes\bar\nu \parallel \xi\otimes Q\right) = \sup_g \left\{ < g,\nu > - \log < e^g, \xi\otimes Q > \right\}.$$

Jensen's inequality yields

$$\log < e^g, \xi \otimes Q > \geq \int_{\Sigma^{n-1}} \log \left(\int_\Sigma \exp g(\mathbf{x},y) Q(dy)\right) \xi(d\mathbf{x})$$

so that

$$H\left(\xi\otimes\bar\nu \parallel \xi\otimes Q\right) \leq \sup_g \left\{ < g,\nu > - \int_{\Sigma^{n-1}} \log \left(\int_\Sigma \exp g(\mathbf{x},y) Q(dy)\right) \xi(d\mathbf{x}) \right. \tag{2.16}$$

which is $\mathcal{J}$ by (2.11). Gathering (2.15), (2.16) and (2.14) we conclude

$$\begin{aligned} \mathcal{J}_\xi(\nu) &= H(\nu \parallel \xi\otimes Q) \ \text{ if } \nu_{n-1} = \xi \\ &= +\infty \ \text{ otherwise.} \end{aligned} \tag{2.17}$$

Let us prove now the lower bound of the (unconditioned) LDP. Let G be an open neighbourhood of ν such that $\mathcal{I}_n^n(\nu) < \infty$ and let $\epsilon > 0$. We have

$$\mathbb{P}\left(\mathcal{L}_r^n \in G\right) \geq \inf_{\xi \in \pi G \cap V_{\pi\nu}} \mathbb{P}\left(\mathcal{L}_r^n \in G \mid \mathcal{L}_r^{n-1} = \xi\right) \mathbb{P}\left(\mathcal{L}_r^{n-1} \in \pi G \cap V_{\pi\nu}\right)$$

From (2.13)

$$\liminf_r r^{-n} \log \inf_{\xi \in \pi G \cap V_{\pi\nu}} \mathbb{P}\left(\mathcal{L}_r^n \in G \mid \mathcal{L}_r^{n-1} = \xi\right) \geq - \inf_{\rho\in G} \mathcal{I}_{\pi\nu}(\rho) - \epsilon \geq -\mathcal{I}_{\pi\nu}(\nu) - \epsilon\,.$$

It remains to find a lowerbound for $r^{-n}\log \mathbb{P}(\mathcal{L}_r^{n-1} \in \pi G \cap V_{\pi\nu})$. We want to apply the LDP at order $n-1$ and use induction.

Let us begin with a very simple lemma which is surely known but for which we don't know any reference.

Lemma 2.4 *If S is a compact metric space and P a probability on S with full support. Then each non empty open subset $\mathcal{O}$ of $\mathcal{M}_1(S)$ contains a probability absolutely continuous with respect to P.*

Proof of Lemma 2.4 Every probability on S may be approximated in the Lévy metric by a weighted finite sum of atoms. By linearity, it is then sufficient to work with an open neighbourhood of a Dirac mass δ_{x_0} for $x_0 \in S$. Now, if $B_j := \{x \in S : d(x, x_0) < j^{-1}\}$ we have $P(B_j) > 0$ for every j, by definition of the support of P. The probability $\xi_j = P(\cdot|B_j)$ is absolutely continuous with respect to P and $\xi_j \Rightarrow \delta_{x_0}$ as $j \to \infty$ (apply the Portmanteau theorem [4] p.356). ■

End of the proof of Theorem 2.1 In $G' := \pi G \cap V_{\pi\nu}$ pick some qQ^{n-1}, then for every $M > 0$ truncate q into $\tilde{q}_M = (M^{-1} \vee q) \wedge M$ and set $q_M =< \tilde{q}_M, Q >^{-1} \tilde{q}_M$. The dominated convergence theorem implies $q_M Q^{n-1} \Rightarrow qQ^{n-1}$ as $M \to \infty$. Choosing M such that $\bar{\nu} := q_M Q^{n-1} \in G'$, we deduce

$$\mathcal{I}_{n-1}^{n-1}(\bar{\nu}) = H(\bar{\nu}_{n-1} \parallel \bar{\nu}_{n-2} \otimes Q) \leq M \log M^2 < \infty\,.$$

The LDP at order $n-1$ yields

$$\liminf_r r^{1-n} \log \mathbb{P}\left(\mathcal{L}_r^{n-1} \in \pi G \cap V_{\pi\nu}\right) \geq -\inf\left\{\mathcal{I}_{n-1}^n(\xi)\ ;\ \xi \in \pi G \cap V_{\pi\nu}\right\} \geq -M \log M^2$$

for every $M > 0$, so that

$$\liminf_r r^{-n} \log \mathbb{P}\left(\mathcal{L}_r^{n-1} \in \pi G \cap V_{\pi\nu}\right) = 0\,,$$

and we get the lower bound.

For the upper bound, we adapt the proof of [4] p.150. Let $\delta > 0$ and $f_\nu \in \mathcal{C}_b$ such that

$$< f_\nu, \nu > -\Lambda_{\pi\nu}(f_\nu) \geq I^\delta(\nu) := \min\{I(\nu) - \delta,\ \delta^{-1}\}\,. \tag{2.18}$$

By continuity, we can find a neighbourhood A_ν of ν such that

$$\inf_{\xi \in A_\nu} \{< f_\nu, \xi > - < f_\nu, \nu >\} \geq -\delta\ ,\quad \sup_{\mu \in \pi A_\nu} |\Lambda_{\pi\nu}(f_\nu) - \Lambda_\mu(f_\nu)| \leq \delta\,. \tag{2.19}$$

Now

$$\begin{aligned} \mathbb{P}\left(\mathcal{L}_r^n \in A_\nu\right) &= \int_{\pi A_\nu} \mathbb{P}\left(\mathcal{L}_r^n \in A_\nu \mid \mathcal{L}_r^{n-1} = \mu\right) P\left(\mathcal{L}_r^{n-1} \in d\mu\right) \\ &\leq \sup_{\mu \in \pi A_\nu} \mathbb{P}\left(\mathcal{L}_r^n \in A_\nu \mid \mathcal{L}_r^{n-1} = \mu\right)\,. \end{aligned} \tag{2.20}$$

The Markov exponential inequality gives

$$\begin{aligned} \mathbb{P}\left(\mathcal{L}_r^n \in A_\nu \mid \mathcal{L}_r^{n-1} = \mu\right) \leq \mathbb{E}\left[\exp r^n\{< f_\nu, \mathcal{L}_r^n > - < f_\nu, \nu >\} \mid \mathcal{L}_r^{n-1} = \mu\right] \times \\ \times \exp -r^n \inf_{\xi \in A_\nu} (< f_\nu, \xi > - < f_\nu, \nu >)\,, \end{aligned}$$

which yields, owing to (2.19) and (2.20):

$$\mathbb{P}\left(\mathcal{L}_r^n \in A_\nu\right) \leq \exp\{2\delta r^n - < f_\nu, \nu > r^n + r^n \Lambda_{\pi\nu}(f_\nu)\}$$

and we end the proof of the LDP for the case $k = n$, using (2.18). Let us remark that the rate function is

$$\mathcal{I}_n^n(\nu) = \mathcal{J}_{\pi\nu}(\nu) = \int_{\Sigma^{n-1}} H(\nu(.|\mathbf{x}) \parallel Q)\nu_{n-1}(d\mathbf{x}) = H(\nu \parallel \nu_{n-1} \otimes Q)\,. \tag{2.21}$$

For intermediate levels, we make an induction on n. Let us assume that b) is true at order $n-1$. From (2.7), we have for $k < n$

$$r^{-k} \log \mathbb{E}\big(\exp < r^k f, \mathcal{L}_r^n > \mid \mathcal{L}_r^{n-1} = \mu \big) = < r^{n-k} \log \mathbb{E} \exp \left[r^{k-n} f(\cdot, W) \right] \, , \, \mu > ,$$

and its limit, as $r \to \infty$ is $< f, \mu \otimes Q > =: < \pi f, \mu >$ (see (2.9). Moreover it is possible to replace in the above display μ by a sequence $\xi_r \Rightarrow \mu$ without change in the conclusion. With the same argument as above, we get a conditional LDP with rate function $\chi(.|\mu \otimes Q)$, where $\chi(a|b) = 0$ if $a = b$ and $\chi(a|b) = \infty$ if $a \neq b$. Then we apply the induction hypothesis gives a (unconditional) LDP of rate function

$$\mathcal{I}_k^n(\nu) = \inf_{\mu \in \mathcal{M}_{n-1}} \chi(\nu|\mu \otimes Q) + \mathcal{I}_k^{n-1}(\mu) \tag{2.22}$$

Now $\mathcal{I}_k^n(\nu) < \infty$ forces $\nu = \mu \otimes Q$ with $\mathcal{I}_k^{n-1}(\mu) < \infty$ i.e. (induction assumption) $\mu = \mu_k \otimes Q^{n-1-k}$, hence $\nu = \mu_k \otimes Q^{n-k}$and in particular $\nu_k = \mu_k$. Moreover, in this case, (2.22) yields $\mathcal{I}_k^n(\nu) = \mathcal{I}_k^{n-1}(\mu) = H(\nu_k \parallel \nu_{k-1} \otimes Q)$, which ends the proof of Theorem 2.1. ■

Remark 1) The rate function $\mathcal{I}_k^n$ is very similar to the rate function of the empirical measure of k-tuples in a sample of size r of i.i.d. copies of W (as it appears in [4] Theorem 6.5.12).

2) Formula (2.22) is of the same type as Lemma 2.3 in [9], (2) in [2], or (2.4) in [3].

3 Contraction and LDP for masses

We first study linear functionals of the empirical measure, and then take the particular case of masses. We assume in this section that Σ is compact.

Proposition 3.1 *Fix n, $k \in \{1, \cdots, n\}$ and $f \in \mathcal{C}(\Sigma^n, \mathbb{R})$. The family of distributions of $\{< f, \mathcal{L}_r^n >\}_r$ satisfies the LDP at scale r^k and good rate function*

$$I_k(c) = \inf\{\mathcal{I}_n^k(\nu) \ : \nu \in \mathcal{M}_1(\Sigma^n) \ , \ < f, \nu > = c\} . \tag{3.1}$$

Moreover

$$I_1(c) = \sup_{\theta \in \mathbb{R}} \left\{ \theta c - \log \int_\Sigma \exp\left(\theta \ \pi_{n-1} f(y)\right) Q(dy) \right\} , \tag{3.2}$$

$$I_k(c) = \sup_{\theta \in \mathbb{R}} \left\{ \theta c - \log \sup_{\mathbf{x} \in \Sigma^{k-1}} \int_\Sigma \exp\left(\theta \ \pi_{n-k} f(\mathbf{x}, y)\right) Q(dy) \right\} \quad \text{if } k \geq 2 . \tag{3.3}$$

Proof: The first claim is a consequence of the contraction principle, since $\nu \mapsto < f, \nu >$ is continuous. To prove (3.2), notice that $\mathcal{I}_1^n(\nu) = H(\nu_1 \parallel Q)$ if $\nu = \nu_1 \otimes Q$, and then $< f, \nu > = < \pi_{n-1} f, \nu_1 >$.

Fix $k \geq 2$. Let $\ell^*(c)$ the right hand side of (3.3) so that we have to prove

$$I_k = \ell^* . \tag{3.4}$$

Set, for $\mathbf{x} \in \Sigma^{k-1}, \lambda \in \mathbb{R}, \ \xi \in \mathcal{M}_1(\Sigma^{k-1})$

$$\Lambda_{\mathbf{x}}(\lambda) := \log \int_\Sigma \exp\left(\lambda \ \pi_{n-k} f(\mathbf{x}, y)\right) Q(dy) \tag{3.5}$$

$$\Lambda_\xi(\lambda) := < \Lambda_\cdot(\lambda), \xi > \ \text{ and } \ \Lambda_\xi^*(c) := \sup_\lambda (\lambda c - \Lambda_\xi(\lambda)) , \tag{3.6}$$

From the obvious equality

$$\sup\left\{\Lambda_\xi(\lambda)\ ;\ \xi\in\mathcal{M}_1(\Sigma^{k-1})\right\}=\sup\left\{\Lambda_{\mathbf{x}}(\lambda)\ ;\ \mathbf{x}\in\Sigma^{k-1}\right\},$$

and the minimax theorem (the ξ set is compact - see [4] p. 151) we deduce

$$\ell^*(c)=\sup_\lambda\left(\lambda c-\sup_\xi\Lambda_\xi(\lambda)\right)=\sup_\lambda\inf_\xi\left(\lambda c-\Lambda_\xi(\lambda)\right)=\inf_\xi\Lambda_\xi^*(c)$$

Let $\nu\in\mathcal{M}_1(\Sigma^n)$ such that $<f,\nu>=c$ and $\mathcal{I}_k^n(\nu)=c$. Then $\nu=\nu_k\otimes Q^{n-k}$ and $<g,\nu_k>=c$ where $g:=\pi^{n-k}f$. Applying once more (1.4), we have for $\mathbf{x}\in\Sigma^{k-1}$,

$$H\left(\nu_k(.|\mathbf{x})\parallel Q\right)\geq\lambda\int_\Sigma g(\mathbf{x},y)\nu_k(dy|\mathbf{x})-\Lambda_{\mathbf{x}}(\lambda)\,. \tag{3.7}$$

If ν_{k-1} satisfies

$$\int_{\Sigma^{k-1}\times\Sigma}g(\mathbf{x},y)\nu_{k-1}(d\mathbf{x})\nu_k(dy|\mathbf{x})=c\,,$$

an integration of (3.7) and a maximization in λ yield

$$\int_{\Sigma^{k-1}}H\left(\nu_k(.|\mathbf{x})\parallel Q\right)\nu_{k-1}(d\mathbf{x})\geq\Lambda^*_{\nu_{k-1}}(c)$$

and taking infimum on ν_{k-1} we get

$$I_k(c)\geq\ell^*(c)\,.$$

Let us prove the reverse inequality. Fix $\xi\in\mathcal{M}_1(\Sigma^{k-1})$ and assume $\Lambda_\xi^*(c)<\infty$. The function Λ_ξ is convex, infinitely differentiable and strictly increasing, and

$$\Lambda'_\xi(\infty)=\int_{\Sigma^{k-1}}\operatorname{essup}g(\mathbf{x},W)\xi(d\mathbf{x})\ ,\quad\Lambda'_\xi(-\infty)=\int_{\Sigma^{k-1}}\operatorname{essinf}g(\mathbf{x},W)\xi(d\mathbf{x})\,.$$

If $c\in\left(\Lambda'_\xi(-\infty),\Lambda'_\xi(\infty)\right)$, let λ the unique solution of $\Lambda'_\xi(\lambda)=c$. This λ reaches the supremum in (3.6). For $\mathbf{x}\in\Sigma^{k-1}$ let $\eta(dy|\mathbf{x})=\exp\left(\lambda g(\mathbf{x},y)-\Lambda_{\mathbf{x}}(\lambda)\right)\ Q(dy)$ which yields easily

$$\int_{\Sigma^{k-1}}H\left(\eta(.|\mathbf{x})\parallel Q\right)\xi(d\mathbf{x})=\Lambda_\xi^*(c)\,. \tag{3.8}$$

For fixed ξ, we get the inequality

$$\Lambda_\xi^*(c)\geq\inf\left\{\int_{\Sigma^{k-1}}H\left(\eta(.|\mathbf{x})\parallel Q\right)\xi(d\mathbf{x})\ ;\int_{\Sigma^{k-1}\times\Sigma}g(\mathbf{x},y)\xi(d\mathbf{x})\eta(dy|\mathbf{x})=c\right\}\,. \tag{3.9}$$

If $c\notin[\Lambda'_\xi(-\infty),\Lambda'_\xi(\infty)]$, we have $\Lambda_\xi^*(c)=\infty$ and (3.9) holds. If $c=\Lambda'_\xi(\infty)<\infty$, we have

$$\Lambda_\xi^*(c)=-\int_{\Sigma^{k-1}}\log\mathbb{P}(g(\mathbf{x},W)=c)\ \xi(d\mathbf{x})<\infty\,,$$

so that $\mathbb{P}(g(\mathbf{x},W)=c)\neq 0$ for ν a.e. $\mathbf{x}$ and choosing

$$\nu_k(dy|\mathbf{x})=\mathbb{I}_{f(\mathbf{x},y)=c}\{\mathbb{P}(f(\mathbf{x},W)=c)\}^{-1}\ Q(dy)$$

we see that (3.8) holds in this case (and also if $c=\Lambda'_\xi(-\infty)$). Henceforth (3.9) holds in all cases. Taking infimum on ξ we get

$$\ell^*(c)=\inf_\xi\Lambda_\xi^*(c)\geq I_k(c) \tag{3.10}$$

and finally equation (3.4) holds, which ends the proof of Proposition 3.1. ■

In the particular case of masses, we assume (1.1) and take $\Sigma=[0,\bar{w}]$. We denote $p_n(\mathbf{x}):=x_1x_2\cdots x_n$ for $\mathbf{x}\in\Sigma^n$ and take $f=p_n$ in the above proposition. This gives the following result, which recovers Theorem 1.2 (see [15] for the case with $n=\infty$).

Corollary 3.2 *1) The family of distributions of Z_r^n satisfies the LDP at scale r with good rate function Λ^*.*

2) For $2 \leq k \leq n$ the family of distributions of Z_r^n satisfies the LDP at scale r^k with good rate function I_k^n given by

$$\begin{array}{rcll} I_k^n(x) & = & 0 & \text{if } x \in [\underline{w}^{k-1}, \bar{w}^{k-1}] \\ I_k^n(x) & = & \Lambda^*(x\bar{w}^{-(k-1)}) & \text{if } x \in [\bar{w}^{k-1}, \bar{w}^k) \\ I_k^n(x) & = & \Lambda^*(x\underline{w}^{-(k-1)}) & \text{if } x \in [\underline{w}^k, \underline{w}^{k-1}) \\ I_k^n(x) & = & +\infty & \text{if } x \notin [\underline{w}^k, \bar{w}^k]\,. \end{array} \tag{3.11}$$

Proof: Start from $\pi_{n-k}f(\mathbf{x}, y) = x_1 \cdots x_{k-1}y$. With the notation of (3.5) we have $\Lambda_{\mathbf{x}}(\lambda) = \Lambda(x_1x_2\cdots x_{k-1}\lambda)$. This gives $\sup_{\mathbf{x}\in S^{k-1}} \Lambda_{\mathbf{x}}(\lambda) = \Lambda(\bar{w}^{k-1}\lambda)$ if $\lambda > 0$ and $\Lambda(\underline{w}^{k-1}\lambda)$ si $\lambda < 0$. ■

4 Gibbs conditioning principle

We want to characterize typical behaviors along a given branch, by means of the Gibbs conditioning principle, ([19] or Chapter 7.3 in [4]). Mass plays the role of energy. For $\delta > 0$, let $A_\delta = \{\nu \in \mathcal{M}_1(\Sigma^n) : | < p_n, \nu > - a | \leq \delta\}$ be the "energy constraint". The following lemma identifies the Gibbs states. Assume $\bar{w} < \infty$.

Let

$$F_a = \{\nu \in \mathcal{M}_1(\mathbb{R}^n) \ : < p_n, \nu >= a\}\ .$$

For $\theta \in \mathbb{R}$ let

$$Q_\theta(dy) := e^{\theta y - \Lambda(\theta}Q(dy)\,,$$

whose mean is $\Lambda'(\theta)$. For every $x \in (\underline{w}, \bar{w})$ let $\theta(x)$ be the unique solution of $\Lambda'(\theta) = x$.

Lemma 4.1 *1) When $\underline{w} < a < \bar{w}$ and $n \geq 1$, the probability $\nu_n^{(a)} = Q_{\theta(a)} \otimes Q^{n-1}$ is the unique solution of*

$$\inf\{\mathcal{I}_1^n(\nu); \nu \in F_a\} = \mathcal{I}_1^n(\nu_n^{(a)})\,. \tag{4.1}$$

2) When $a \in [\bar{w}^{k-1}, \bar{w}^k)$ for $k \geq 2$, the probability $\nu_n^{(a)} = \delta_{\bar{w}}^{k-1} \otimes Q_{\theta(a\bar{w}^{-(k-1)})} \otimes Q^{n-k}$ is the unique solution of

$$\inf\{\mathcal{I}_k^n(\nu); \nu \in F_a\} = \mathcal{I}_k^n(\nu_n^{(a)})\,. \tag{4.2}$$

3) When $a \in [\underline{w}^k, \underline{w}^{k-1})$, the probability $\nu_n^{(a)} = \delta_{\underline{w}}^{k-1} \otimes Q_{\theta(a\underline{w}^{-(k-1)})} \otimes Q^{n-k}$ is the unique solution of (4.2).

Here is the main result of this section.

Theorem 4.2 *Fix $n > 0$ and $k \leq n$. Let Γ be an open neighbourhood of $\nu_n^{(a)}$*

$$\limsup_{\delta\to 0}\limsup_{r}\ r^{-k}\log \mathbb{P}\left(\mathcal{L}_r^n \notin \Gamma \mid \mathcal{L}_r^n \in A_\delta\right) < 0\,.$$

Consequently, for any $i_1, i_2, \cdots, i_n$, the distribution of $(W_{i_1}, \cdots, W_{i_1\cdots i_n})$ conditioned upon $Z_r^n \in [a-\delta, a+\delta]$ converges to $\nu_n^{(a)}$, as $r \to \infty$, followed by $\delta \to 0$.

Proof of Lemma 4.1 To simplify, we treat only the second case and set $\tilde{a} = a\bar{w}^{-(k-1)}$, $\tilde{\theta} = \theta(\tilde{a})$, $\tilde{\nu} = \nu_n^{(a)}$. From the previous section, we know that

$$\mathcal{I}_k^n(\tilde{\nu})) = H(Q_{\tilde{\theta}} \parallel Q) = \tilde{\theta}\tilde{a} - \Lambda(\tilde{\theta})\,.$$

So $\tilde{\nu}$ is a solution. Let $\mu \in F_a$ another one. Since $\mathcal{I}_k^n(\mu) < \infty$ we have $\mu = \mu_k \otimes Q^{n-k}$ and

$$\mathcal{I}_k^n(\mu) = H\big(\mu_k \parallel \mu_{k-1} \otimes Q\big) = H\big(\mu_k \parallel \mu_{k-1} \otimes Q_{\tilde{\theta}}\big) + \tilde{\theta} \int_{\Sigma^k} x_k \mu_k(d\mathbf{x}) - \Lambda(\tilde{\theta})\,.$$

Now since μ_k is absolutely continuous with respect to $\mu_{k-1} \otimes Q$, the support of $\mu_k(dy|\mathbf{x})$ is a subset of S. Since $\mathcal{I}_k^n(\mu) < \infty$ we know from Theorem 2.1 b) that $\mu = \mu_k \otimes Q^{n-k}$, which gives

$$a = < p_n, \mu > = < p_k, \mu_k > \leq \ \bar{w}^{k-1} \int_{\Sigma^k} x_k \mu_k(d\mathbf{x})\,. \tag{4.3}$$

Since $\tilde{\theta} > 0$, we get $\Lambda^*(\tilde{a}) - H\left(\mu_k \parallel \mu_{k-1} \otimes Q_{\tilde{\theta}}\right) \geq \tilde{\theta} a \bar{w}^{-(k-1)} - \Lambda(\tilde{\theta}) = \Lambda^*(\tilde{a})$ which needs $\mu_k = \mu_{k-1} \otimes Q_{\tilde{\theta}}$. Carrying into (4.3) gives $\bar{w}^{k-1} = < p_{k-1}, \mu_{k-1} >$ which forces $\mu_{k-1} = \delta_{\bar{w}^{k-1}}$. ■

Proof of Theorem 4.2: Clearly,

$$\begin{aligned}
&\limsup_{\delta\to 0} \limsup_{r} \ r^{-k} \log \mathbb{P}\left(\mathcal{L}_r^n \notin \Gamma \mid \mathcal{L}_r^n \in A_\delta\right) \quad \leq \\
&\quad \leq \lim_{\delta\to 0} \limsup_{r} \ r^{-k} \log \mathbb{P}\left(\mathcal{L}_r^n \in \Gamma^c \cap \ A_\delta\right) \quad - \quad \lim_{\delta\to 0} \liminf_{r} \ r^{-k} \log \mathbb{P}\left(\mathcal{L}_r^n \in A_\delta\right)
\end{aligned}$$

Using the upper bound of the LDP (Theorem 2.1 b)), we see that the first term of the right hand side is less than $-\lim_{\delta\to 0} \ \inf\left\{H(\mu|Q^{\otimes n}) \ ; \ \mu \in \Gamma^c \cap \ A_\delta\right\}$. Since the sets $\Gamma^c \cap \ A_\delta$ are closed and nested (see [4] Lemma 4.1.6), the bound is equal to $-\inf\left\{H(\mu|Q^{\otimes n}) \ ; \ \mu \in \Gamma^c \cap \ F_a\right\}$, which is strictly smaller than $-I_k(a)$. For the second term, we apply the lower bound of the LDP to A_δ so that

$$\liminf_r \ r^{-k} \log \mathbb{P}\left(\mathcal{L}_r^n \in A_\delta\right) \geq -\inf\{H(\mu|Q^{\otimes n}) \ ; \ \mu \in int A_\delta\} \geq -H(\nu_n^{(a)} \| Q^{\otimes n}) = -I_k(a) \ ■$$

Acknowledgement. I thank Wendelin Werner for asking me this question.

References

[1] J. Barral. Generalized vector multiplicative cascades. *Adv. in Appl. Probab.*, 33(4):874–895, 2001.

[2] J.D. Biggins. Large deviations for mixtures. Available at http://www.shef.ac.uk/ st1jdb/.

[3] N.R. Chaganthy. Large deviations for joint distributions and statistical applications. *Sankhya*, 59:147–166, 1997.

[4] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, 2nd edition, 1998.

[5] J.D. Deuschel and D.W. Stroock. *Large deviations*. Academic Press, 1989.

[6] I.H. Dinwoodie and S.L. Zabell. Large deviations for exchangeable random vectors. *Ann. Probab.*, pages 1147–1166, 1992.

[7] P. Dupuis and R. Ellis. *A weak convergence approach to the theory of large deviations*. Wiley, 1997.

[8] P. Eichelsbacher and U. Schmock. Large deviations of U-empirical measures in strong topologies and applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5):779–797, 2002.

[9] W. Finnoff. Integration of large-deviation kernels and applications to large deviations for evolutionary games. *Probab. Theory Relat Fields*, 122:141–162, 2002.

[10] M. Grunwald. Sanov results for Glauber spin-glass dynamics. *Probab. Theory Relat Fields*, 106:187–232, 1996.

[11] J.-P. Kahane and J. Peyrière. Sur certaines martingales de Benoit Mandelbrot. *Advances in Math.*, 22(2):131–145, 1976.

[12] A. N. Kolmogorov. The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers. *Proc. Roy. Soc. London Ser. A*, 434(1890):9–13, 1991. Translated from the Russian by V. Levin, Turbulence and stochastic processes: Kolmogorov's ideas 50 years on.

[13] Q. Liu. On generalized multiplicative cascades. *Stochastic Process. Appl.*, 86(2):263–286, 2000.

[14] Q. Liu, E. Rio, and A. Rouault. Limit theorems for Multiplicative Processes. to appear in *J. of Theoret. Probab.*, 2004.

[15] Q. Liu and A. Rouault. Limit theorems for Mandelbrot's multiplicative cascades. *Annals of Appl. Probab.*, 10(1):218–239, 2000.

[16] M. Löwe. Iterated large deviations. *Stat. and Probab. Letters*, 26(3):219–223, February 1996.

[17] B. Mandelbrot. Multiplications aléatoires itérées et distributions invariantes par moyenne pondérée aléatoire: quelques extensions. *C. R. Acad. Sci. Paris Sér. A*, 278:355–358, 1974.

[18] M. Ossiander and E.C. Waymire. Statistical estimation for multiplicative cascades. *Ann. Statist.*, 28(6):1533–1560, 2000.

[19] D.W. Stroock and O. Zeitouni. Microcanonical distributions, Gibbs states, and the equivalence of ensembles. In Durrett R. and H. Kesten, editors, *Festschrift in Honour of F.Spitzer*, pages 399–424. Birkhauser, 1991.

[20] J. Trashorras. Large deviations for a triangular array of exchangeable random variables. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5):649–680, 2002.