Estimating Potential Functions of One–Dimensional Gibbs States Under Constraints *

by

Chuanshu Ji
Purdue University and Columbia University

Technical Report #88-13

Department of Statistics
Purdue University

April 1988

* Research partially supported by a David Ross Fellowship

*

§1. Introduction

For time series in which data are categorical rather than numerical, linear models and normality assumptions are not appropriate. In such cases, the so-called "Gibbs states" may be appropriate models.

Gibbs states were originally conceived as models in statistical mechanics (cf. Ruelle [5]), and they are also important in topological dynamics (cf. Bowen [1]). Multi-dimensional Gibbs states have been proposed as models for certain types of spatial data (cf. Ripley [4]). However, using one-dimensional Gibbs states to model categorical time series seems to be a new idea.

A one-dimensional Gibbs state μ_f is a probability measure on the space $\Sigma^+ = \prod_{i=0}^{\infty} \{1, \ldots, r\}$. Each element of Σ^+ is a sequence $x = (x_0, x_1, \ldots)$ whose coordinates x_i have possible states $1, \ldots, r$, $i = 0, 1, \ldots$. Define the forward shift operator $\sigma : \Sigma^+ \to \Sigma^+$ by $(\sigma x)_n = x_{n+1}$, $n = 0, 1, \ldots$, for $x \in \Sigma^+$. The Gibbs measure μ_f is the unique σ-invariant probability measure on Σ^+ satisfying

$$c_1 \leq \frac{\mu_f(y : y_i = x_i,\ 0 \leq i \leq m-1)}{\exp\{-mp + \sum_{j=0}^{m-1} f(\sigma^j x)\}} \leq c_2 \tag{1.1}$$

for some constants $c_1, c_2 \in (0, \infty)$ and for all $x \in \Sigma^+$, $m \in \mathbb{N}$, where p is called the pressure for f, and f is a real-valued function defined on Σ^+, called the potential (or energy) function. It is clear from (1.1) that f determines the dependence in the stationary sequence $\{X_n\}$.

Traditionally, categorical time series $X = (X_0, X_1, \ldots)$ are modeled by finite state stationary Markov chains, or more generally, k-step Markov dependent chains with k being an arbitrary positive integer. When f depends only on a finite number k of coordinates, X under μ_f is just a k-step Markov dependent sequence. Therefore the family of Gibbs states includes all k-step Markov models, $k = 1, 2, \ldots$.

The inequalities (1.1) reveal that the family of Gibbs states looks like an infinite-dimensional exponential family, where the potential function f plays the role of the natural parameter. There is a formal similarity between (1.1) and the likelihood function for a stationary Gaussian sequence;

however, for Gaussian measures the potential function is quadratic.

Assuming the potential function f is unknown and the observations $X_0, \ldots, X_{n-1}$ are given. One may want to estimate f based on those n observations. However, since two different functions f and g may induce the same Gibbs measure $\mu_f (= \mu_g)$, f is not identifiable; only μ_f is. Two approaches are adopted to resolve the identifiability problem: reparametrization and normalization constraints. In [2], instead of estimating f we estimate the linear functional $\theta \triangleq \int \psi d\mu_f$, where ψ is a known function. Estimators of maximum likelihood type are constructed and shown to be strongly consistent, asymptotically normal and asymptotically efficient. In this paper, we show that under appropriate normalization constraints f is identifiable. Strongly consistent estimators (in sup-norm) T_n for the unknown function e^f are constructed.

We first introduce Ruelle-Perron-Frobenius theory and define Gibbs states rigorously.

(1) **Forward shift:** Let A be an irreducible, aperiodic, $r \times r$ matrix of zeros and ones ($r > 1$), and let

$$\Sigma_A^+ = \left\{ x \in \prod_{i=0}^{\infty} \{1, \ldots, r\} : A_{x_i x_{i+1}} = 1. \quad \forall i \in \mathbb{N} \right\},$$

where A_{jk}, $j, k = 1, \ldots, r$ are entries of A. The space Σ_A^+ is compact and metrizable in the product topology.

Define the forward shift operator $\sigma : \Sigma_A^+ \to \Sigma_A^+$ by $(\sigma x)_n = x_{n+1}$, $n \in \mathbb{N}$, $x \in \Sigma_A^+$. Observe that σ, although continuous and surjective, is not generally $1-1$.

Remark: Σ^+ is a special case of Σ_A^+ with $A_{jk} = 1$ for all $j, k = 1, \ldots, r$. The reason for introducing Σ_A^+ is to cover those cases in which certain transitions $j \to k$ are not allowed.

(2) **Hölder continuity:** Let $C(\Sigma_A^+)$ denote the space of continuous, complex-valued functions on Σ_A^+. For $f \in C(\Sigma_A^+)$ define

$$\text{var}_n f = \sup\{|f(x) - f(y)| : x_i = y_i,\ 0 \le i < n\};$$

for $0 < \rho < 1$ let

$$|f|_\rho = \sup_{n \in \mathbb{N}} \frac{\mathrm{var}_n f}{\rho^n}$$

and

$$\mathcal{F}_\rho^+ = \{f \in C(\Sigma_A^+) : |f|_\rho < \infty\}.$$

Elements of $\mathcal{F}_\rho^+$ are referred to as Hölder continuous functions. The space $\mathcal{F}_\rho^+$ is a Banach algebra when endowed with the norm $\|\cdot\|_\rho = |\cdot|_\rho + \|\cdot\|_\infty$.

(3) **Ruelle-Perron-Frobenius (RPF) operators:** For $f, g \in C(\Sigma_A^+)$, define $\mathcal{L}_f : C(\Sigma_A^+) \to C(\Sigma_A^+)$ by

$$\mathcal{L}_f g(x) = \sum_{y:\sigma y = x} e^{f(y)} g(y), \; x \in \Sigma_A^+.$$

Theorem 1.1. *For each real-valued $f \in \mathcal{F}_\rho^+$, there exists $\lambda_f \in (0, \infty)$, a simple eigenvalue of $\mathcal{L}_f : \mathcal{F}_\rho^+ \to \mathcal{F}_\rho^+$, with strictly positive eigenfunction h_f and a Borel measure ν_f on Σ_A^+ such that $\mathcal{L}_f^* \nu_f = \lambda_f \nu_f$. Moreover, spectrum $(\mathcal{L}_f) \backslash \{\lambda_f\}$ is contained in a disc of radius strictly less than λ_f. Finally,*

$$\lim_{n \to \infty} \|\mathcal{L}_f^n g / \lambda_f^n - (\int g d\nu_f) h_f\|_\infty = 0, \; \forall g \in C(\Sigma_A^+).$$

The proof may be found in [1], [5].

(4) **Gibbs states:** Assume that $\int h_f d\nu_f = 1$. For each $f \in \mathcal{F}_\rho^+$, the Gibbs measure μ_f is defined by

$$\frac{d\mu_f}{d\nu_f} = h_f.$$

It is easy to verify that μ_f is an invariant probability measure under σ.

Let $M_\sigma(\Sigma_A^+)$ denote the set of all σ-invariant probability measures on Σ_A^+.

Theorem 1.2. *For each $f \in \mathcal{F}_\rho^+$, there exist constants $c_1, c_2 \in (0, \infty)$ such that*

$$(1.2) \qquad c_1 \le \frac{\mu_f(x_0, \ldots, x_{m-1})}{\exp\{-mp + \sum_{j=0}^{m-1} f(\sigma^j x)\}} \le c_2, \quad \forall x \in \Sigma_A^+, \; m \in \mathbb{N}^+ = \mathbb{N} \backslash \{0\};$$

and μ_f is the unique element in $M_\sigma(\Sigma_A^+)$ such that (1.2) holds, where $\mu_f(x_0,\ldots,x_{m-1})$ $= \mu_f(y \in \sum_A^+ : y_i = x_i,\ 0 \le i \le m-1)$. Here $p = p(f) = \log \lambda_f$ is called the pressure for f.

The proof is given in [1].

Remark 1.3. (1.2) is an extension of (1.1) for the case Σ_A^+.

Remark 1.4. Two functions $f, g \in C(\Sigma_A^+)$ are said to be homologous, written $f \sim g$, if there exists $\phi \in C(\Sigma_A^+)$ such that

$$f - g = \phi \circ \sigma - \phi.$$

Homology is clearly an equivalence relation. It can be shown (cf. [1]) that $\mu_f = \mu_g$ iff $f - g \sim$ constant; otherwise $\mu_f \perp \mu_g$, because μ_f and μ_g are ergodic measures.

Remark 1.5. Gibbs states have the following special cases: Let $X = (X_0, X_1, \ldots)$ be a stationary sequence with underlying distribution μ_f, then

(i) In the case of Σ^+, if $f(x) \equiv c$, for all $x \in \Sigma^+$, then X is a sequence of iid random variables with discrete uniform distribution.

(ii) In the case of Σ^+, if $f(x) = f(x_0)$, for all $x \in \Sigma^+$, i.e., f only depends on the first coordinate, then X is a sequence of iid random variables with $P(X_0 = l) = ce^{f(l)}$, $l = 1, \ldots, r$, where $c = 1/\sum_{l=1}^r e^{f(l)}$.

(iii) In the case of Σ_A^+, if $f(x) = f(x_0)$, then X forms a stationary Markov chain with state space $\{1, \ldots, r\}$ and suitable transition probabilities.

(iv) In the case of Σ_A^+, if $f(x) = f(x_0, \ldots, x_{k-1})$, $k \in \mathbb{N}^+$, i.e., f only depends on the first k coordinates, then X is a k-step Markov dependent chain.

In fact the family of Gibbs states includes all finite state stationary k-step Markov chains, $k \in \mathbb{N}^+$.

§2. Construction of Consistent Estimators for e^f under certain constraints on f

The reason that the identifiability problem arises when estimating the potential function f is because all potential functions equivalent to f in the sense of homology induce the same Gibbs state

μ_f. The next lemma indicates that in each equivalence class there is a unique distinguished element which satisfies certain normalization conditions. We will construct estimators of this distinguished element later on.

Lemma 2.1. *For every $f \in \mathcal{F}_\rho^+$, there uniquely exists $\tilde{f} \in \mathcal{F}_\rho^+$ such that*

(i) $\lambda_{\tilde{f}} = 1$;

(ii) $h_{\tilde{f}} \equiv 1$;

(iii) $\tilde{f} \sim f +$ *constant.*

Proof. Let

$$\tilde{f} = f + \log h_f - \log h_f \cdot \sigma - \log \lambda_f, \tag{2.1}$$

then (i), (ii), (iii) are straightforward.

Furthermore, by [3] Proposition 1 we have

$$\mu_f(x_0 | x_1, x_2, \ldots) = \frac{e^{f(x)} h_f(x)}{\lambda_f h_f(\sigma x)}, \quad \forall x \in \Sigma_A^+, \tag{2.2}$$

where the LHS is the conditional probability of x_0 appearing in the slot 0 given that $x_1, x_2, \ldots$ appear in the slots 1, 2, ... Since the martingale convergence theorem implies that the limit

$$\lim_{m\to\infty} \mu_f(x_0 | x_1, \ldots, x_{m-1}) = \lim_{m\to\infty} \frac{\mu_f(x_0, \ldots, x_{m-1})}{\mu_f(x_1, \ldots, x_{m-1})} \tag{2.3}$$

exists for almost every $x \in \Sigma_A^+$ under μ_f, the LHS in (2.2) is well–defined as the limit in (2.3). Therefore, uniqueness follows from (2.2). □

Let $\mathcal{H} \subset \mathcal{F}_\rho^+$ be the set of all functions that satisfy (i) and (ii) in Lemma 2.1. In the sequel we just use the notation f to denote the generic element in $\mathcal{H}$ when there is no confusion.

Assume that $X = (X_0, X_1, \ldots)$ is a stationary sequence with probability distribution μ_f, $f \in \mathcal{H}$ and let $x = (x_0, x_1, \ldots)$ denote a specific value of X. We want to estimate the unknown function

e^f based on observations $X_0, \ldots, X_{n-1}$. f and e^f are in 1-1 correspondence. Hence Lemma 2.1 guarantees that e^f is identifiable for $f \in \mathcal{H}$.

For simplicity we only consider the case with the configuration space Σ^+. Similar results can be derived for the case Σ_A^+. Our goal is to construct a random function T_n on Σ^+ based on $X_0, \ldots, X_{n-1}$ such that for every $f \in \mathcal{H}$

$$(2.4) \qquad \sup_{y \in \Sigma^+} |T_n(y) - e^{f(y)}| \to 0, \text{ a.s. under } \mu_f \quad \text{as } n \to \infty.$$

The random function T_n satisfying (2.4) is called a *strongly consistent estimator* of e^f.

Notice that Lemma 2.1 (i) and (ii) are equivalent to the normalization constraint

$$\sum_{x_0} e^{f(x_0, x_1, \ldots)} = 1, \; \forall \; x \in \Sigma^+.$$

Moreover, for $f \in \mathcal{H}$, by (2.2)

$$(2.5) \qquad \mu_f(x_0 | x_1, x_2, \ldots) = e^{f(x)}, \quad \forall x \in \Sigma^+.$$

So e^f may be regarded as an infinite–step backward transition function, which suggests the following plan for constructing T_n.

First of all, we may use a sequence of finite–step (backward) transition functions $\{\mu_f(x_0|x_1, \ldots, x_{m-1}),\ m \in \mathbb{N},\ x \in \Sigma^+\}$ to approximate e^f. Then at each stage m we estimate $\mu_f(x_0|x_1, \ldots, x_{m-1})$ by the "sample transition function". Given n observations, the correct order for the "step–length" m should be $c \log n$, where $c \in (0,1)$ also depends on f, hence is unknown. Certain adaptive procedures are proposed to guarantee the strong consistency of the estimator T_n.

Construction of Consistent Estimators

Given observations $X_0, \ldots, X_{n-1}$ we first construct n periodic sequences $\sigma^j X(n),\ j = 0, 1, \ldots, n-1$ with

$$X(n) = (X_0, \ldots, X_{n-1}; X_0, \ldots, X_{n-1}; \ldots).$$

Then for every $y \in \Sigma^+$ and $m < n$ define

$$N_m^{(n)}(y) = \sum_{j=0}^{n-1} I_{\{(\sigma^j X(n))_k = y_k,\ k=0,1,\ldots,m-1\}},$$

$$N_{m-1}^{(n)}(y) = \sum_{j=0}^{n-1} I_{\{(\sigma^j X(n))_k = y_k,\ k=1,\ldots,m-1\}},$$

where $(\sigma^j X(n))_k$ represents the k–th coordinate of the sequence $\sigma^j X(n)$. And define

$$R_m^{(n)}(y) = \begin{cases} \frac{N_m^{(n)}(y)}{N_{m-1}^{(n)}(y)}, & \text{if } N_{m-1}^{(n)}(y) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

$R_m^{(n)}(y)$, also written as $\frac{N_m^{(n)}(y)}{n} / \frac{N_{m-1}^{(n)}(y)}{n}$, is the "sample conditional probability" of y_0 appearing in the slot 0 given that $y_1, \ldots, y_{m-1}$ appear in the slots $1, \ldots, m-1$. The next two theorems show that under certain conditions $R_m^{(n)}$ is a strongly consistent estimator of e^f.

Theorem 2.2. *Suppose f is an unknown potential function satisfying*

(A1) $f \in \mathcal{H}$;

(A2) $\|f\|_\rho \leq K$ *for a known constant* $K > 0$.

Let

(2.6) $$\overline{a} = \frac{2K}{1-\rho} \quad \text{and}$$

(2.7) $$m = [c \log n],$$

where $c \in (0,1)$ satisfies

(2.8) $$1 - \overline{a}c > 0;$$

the notation $[z]$ represents the integer part of z.

Define

$$T_n(y) = R_m^{(n)}(y),\ y \in \Sigma^+.$$

then (2.4) holds for T_n .

Theorem 2.3. *Under the assumptions in Theorem 2.2 without (A2), T_n defined by the following adaptive procedure also satisfies (2.4).*

Procedure 2.4. Choose a sequence of positive constants $\{c_n,\ n \in \mathbb{N}\}$, such that $c_n \downarrow 0$ as $n \to \infty$ with arbitrarily slow rate (e.g. $c_n \log n \to \infty$ as $n \to \infty$). Set

$$m = [c_n \log n],$$

then define

$$T_n(y) = R_m^{(n)}(y),\ y \in \Sigma^+.$$

The proofs of Theorem 2.2 and Theorem 2.3 will be given in Section 3.

§3. Exponential Decay of Certain Large Deviation Probabilities

In this section the deviation of the estimator T_n (or $R_m^{(n)}$) from the estimated function e^f is investigated in detail. The main result is that the related large deviation probabilities drop to zero exponentially as n tends to infinity. As a corollary, the strong consistency of T_n is established.

The next lemma provides uniform bounds for certain conditional probabilities, which will be used very often.

Lemma 3.1. *For every $f \in \mathcal{H}$, there exists a positive constant a which depends on f, such that*

$$e^{-a} \le \mu_f(y_{m-1}|y_0, \ldots, y_{m-2}) \le 1 - e^{-a}, \tag{3.1}$$

$$e^{-a} \le \mu_f(y_0|y_1, \ldots, y_{m-1}) \le 1 - e^{-a}, \tag{3.2}$$

uniformly for all $y \in \Sigma^+$ and all $m \in \mathbb{N}$.

Proof. For $f \in \mathcal{H}$, (1.1) implies that

$$\mu_f(y_m|y_0,\ldots,y_{m-2}) \geq \frac{c_1}{c_2}e^{f(\sigma^{m-1}y)} \qquad \text{and}$$

$$\mu_f(y_0|y_1,\ldots,y_{m-1}) \geq \frac{c_1}{c_2}e^{f(y)},\ \forall\, y \in \Sigma^+,\ m \in \mathbb{N}.$$

Bowen [1] gives $\begin{cases} c_1 = e^{-\|f\|_\infty - \eta} \\ c_2 = e^{\eta} \end{cases}$ with

$$\eta = \sum_{k=0}^{\infty} \mathrm{var}_k f \leq \frac{|f|_\rho}{1-\rho}.$$

Therefore, (3.1) and (3.2) follow by setting

$$a = \frac{2\|f\|_\rho}{1-\rho}. \qquad \square \tag{3.3}$$

For $y \in \Sigma^+$ and $m < n$, let

$$P_m^{(n)}(y) = \mu_f(x \in \Sigma^+ :\ x_i = y_i,\ i = 0,\ldots,m-1);$$

and

$$P_{m-1}^{(n)}(y) = \mu_f(x \in \Sigma^+ :\ x_i = y_i,\ i = 1,\ldots,m-1).$$

Then

$$\begin{cases} E_f N_m^{(n)}(y) = nP_m^{(n)}(y), \\ E_f N_{m-1}^{(n)}(y) = nP_{m-1}^{(n)}(y), \end{cases}$$

and

$$\mu_f(y_0|y_1,\ldots,y_{m-1}) = \frac{P_m^{(n)}(y)}{P_{m-1}^{(n)}(y)}.$$

By (2.5), $\frac{P_m^{(n)}(y)}{P_{m-1}^{(n)}(y)}$ is close to $e^{f(y)}$ for every y when m is large.

Notice that

$$|T_n(y) - e^{f(y)}| \tag{3.4}$$

$$\leq \left|\frac{P_m^{(n)}(y)}{P_{m-1}^{(n)}(y)} - e^{f(y)}\right| + I_{(N_{m-1}^{(n)}(y)=0)} \cdot \frac{P_m^{(n)}(y)}{P_{m-1}^{(n)}(y)} + I_{(N_{m-1}^{(n)}(y)>0)}\left|\frac{N_m^{(n)}(y)}{N_{m-1}^{(n)}(y)} - \frac{P_m^{(n)}(y)}{P_{m-1}^{(n)}(y)}\right|$$

$$\triangleq D_n^{(1)}(y) + D_n^{(2)}(y) + D_n^{(3)}(y).$$

The first term has a uniform upper bound. For m sufficiently large,

$$\sup_{y\in\Sigma+} D_n^{(1)}(y) \le e^{\|f\|_\infty}\left(e^{\mathrm{var}_m f} - 1\right) \le 2e^{\|f\|_\infty}\mathrm{var}_m f. \tag{3.5}$$

In what follows we simply denote the probability of event A under μ_f by $P(A)$, and the corresponding expectation operator by $E(\cdot)$.

For every $\varepsilon \in (0, \frac{1}{2})$,

$$P(D_n^{(2)}(y) > \varepsilon) = P(N_{m-1}^{(n)}(y) = 0) \le P\left(|\frac{N_{m-1}^{(n)}(y)}{nP_{m-1}^{(n)}(y)} - 1| > \varepsilon\right). \tag{3.6}$$

Lemma 3.2. *For every $\varepsilon > 0$,*

$$P(D_n^{(3)}(y) > 2\varepsilon) \le P\left(|\frac{N_m^{(n)}(y)}{nP_m^{(n)}(y)} - 1| > \delta_1\right) + P\left(\frac{N_{m-1}^{(n)}(y)}{nP_{m-1}^{(n)}(y)} - 1| > \delta_2\right), \tag{3.7}$$

where $\delta_1 = \frac{\varepsilon}{1+\varepsilon}$, $\delta_2 = (\frac{\varepsilon}{1-e^{-a}})/(1 + \frac{\varepsilon}{1-e^{-a}})$.

Proof. Since

$$D_n^{(3)}(y) \le I_{(N_{m-1}^{(n)}(y)>0)} \cdot \frac{|N_m^{(n)}(y) - nP_m^{(n)}(y)|}{N_{m-1}^{(n)}(y)} + I_{(N_{m-1}^{(n)}(y)>0)} \cdot \frac{P_m^{(n)}(y)}{P_{m-1}^{(n)}(y)} \cdot \frac{|N_{m-1}^{(n)}(y) - nP_{m-1}^{(n)}(y)|}{N_{m-1}^{(n)}(y)},$$

and $N_{m-1}^{(n)}(y) \ge N_m^{(n)}(y)$, we obtain that

$$\begin{aligned}
&P(D_n^{(3)}(y) > 2\varepsilon)\\
&\le P\left(|N_m^{(n)}(y) - nP_m^{(n)}(y)| > \varepsilon N_{m-1}^{(n)}(y)\right) + P\left(|N_{m-1}^{(n)}(y) - nP_{m-1}^{(n)}(y)| > \frac{\varepsilon}{1-e^{-a}} \cdot N_{m-1}^{(n)}(y)\right)\\
&\le P\left((1+\varepsilon)|N_m^{(n)}(y) - nP_m^{(n)}(y)| > \varepsilon nP_m^{(n)}(y)\right)\\
&\quad + P\left((1 + \frac{\varepsilon}{1-e^{-a}})|N_{m-1}^{(n)}(y) - nP_{m-1}^{(n)}(y)| > \frac{\varepsilon}{1-e^{-a}} \cdot nP_{m-1}^{(n)}(y)\right)\\
&= P\left(|\frac{N_m^{(n)}(y)}{nP_m^{(n)}(y)} - 1| > \delta_1\right) + P\left(|\frac{N_{m-1}^{(n)}(y)}{nP_{m-1}^{(n)}(y)} - 1| > \delta_2\right). \quad \square
\end{aligned}$$

(3.6) and (3.7) indicate that it suffices to evaluate $P(|\frac{N_m^{(n)}(y)}{nP_m^{(n)}(y)} - 1| > \varepsilon)$ for large n.

Remark 3.3. Here is the motivation of choosing the step length $m = [c \log n]$ (cf. (2.7)). To have consistent estimators for e^f, the ratio $\frac{N_m^{(n)}(y)}{nP_m^{(n)}(y)}$ has to be close to one for every y. Hence both $N_m^{(n)}(y)$ and its expectation $nP_m^{(n)}(y)$ should be large. For $m = [c \log n]$, by (3.1)

$$n^{1-ac} \le nP_m^{(n)}(y) \le n^{1-bc}, \ \forall \quad y \in \Sigma^+,$$

where $b = -\log(1 - e^{-a}) > 0$. Hence $\log n$ is the proper order and we may choose the constant $c \in (0, 1)$ such that

$$1 - ac > 0. \tag{3.8}$$

However, since a depends on the unknown function f, we should adopt either (2.8) to choose c or Procedure 2.4 to choose the sequence $\{c_n\}$.

Now let

$$Z_j = I_{\{(\sigma^j X(n))_k = y_k, \ k=0,1,\ldots,m-1\}} - P_m^{(n)}(y), \qquad j = 0, 1, \ldots, n-1;$$

Then

$$N_m^{(n)}(y) - nP_m^{(n)}(y) = \sum_{j=0}^{n-1} Z_j,$$

and

$$P\left(|\frac{N_m^{(n)}(y)}{nP_m^{(n)}(y)} - 1| > \varepsilon \right) = P\left(|\sum_{j=0}^{n-1} Z_j| > \varepsilon n P_m^{(n)}(y) \right).$$

This is the large deviation probability for partial sum of a double–array, mean zero, mixing sequence. The following "splitting" procedure turns out to be useful.

For a small number $\lambda \in (0, \frac{1}{4})$.

Set

$$p = [n^{\frac{1}{2}+\lambda}],$$

$$q = [n^{\frac{1}{2}-\lambda}],$$

and

$$k = [\frac{n - m + 1 + q}{p + q}], \qquad \text{i.e.}$$

k satisfies

$$kp + (k-1)q \leq n - m + 1 < (k+1)p + kq.$$

Let

$$U_1 = Z_0 + \ldots + Z_{p-1},$$

$$U_2 = Z_{p+q} + \ldots + Z_{2p+q-1},$$

$$\ldots$$

$$U_k = Z_{(k-1)(p+q)} + \ldots + Z_{kp+(k-1)q-1};$$

And

$$V_1 = Z_p + \ldots + Z_{p+q-1},$$

$$V_2 = Z_{2p+q} + \ldots + Z_{2p+2q-1},$$

$$\ldots$$

$$V_k = \begin{cases} Z_{n-m+1} + \ldots + Z_{n-1}, & \text{if } kp + (k-1)q = n - m + 1, \\ Z_{kp+(k-1)q} + \ldots + Z_{n-m} + Z_{n-m+1} + \ldots + Z_{n-1}, & \text{if } kp + (k-1)q < n - m + 1. \end{cases}$$

Each U_i, $i = 1, \ldots, k$ contains p Z–terms; Each V_j, $j = 1, \ldots, k-1$ contains q Z–terms. In particular, V_k contains s Z–terms with

$$m - 1 \leq s \leq (p + q - 1) + (m - 1).$$

The idea is that for large n both $\{U_i,\ i = 1, \ldots, k\}$ and $\{V_j,\ j = 1, \ldots, k-1\}$ behave approximately like iid sequences. And V_k does not affect the magnitude of $\sum_{j=0}^{n-1} Z_j$ very much.

Denote $nP_m^{(n)}(y)$ by b_n^2 and note that

$$\sum_{j=0}^{n-1} Z_j = \sum_{i=1}^{k} U_i + \sum_{j=1}^{k-1} V_j + V_k.$$

Therefore,

$$P\left(|\sum_{j=0}^{n-1} Z_j| > \varepsilon b_n^2\right)$$

$$\leq P\left(|\sum_{i=1}^{k} U_i| > \delta b_n^2\right) + P\left(|\sum_{j=1}^{k-1} V_j| > \delta b_n^2\right) + P\left(|V_k| > \delta b_n^2\right),$$

with $\delta = \frac{\varepsilon}{3}$.

Recall the following weak Bernoulli property of μ_f (cf. [1] Theorem 1.25).

Let $\mathcal{A}_{m-1}$ be the σ–field generated by $(X_0, \ldots, X_{m-1})$; $\mathcal{A}_{m+n,\infty}$ be the σ–field generated by $(X_i,\ i \geq m+n)$. Then there exist constants $C > 0$ and $\beta \in (0,1)$, which only depend on f, such that

$$|\frac{P(A \cap B)}{P(A) \cdot P(B)} - 1| \leq C\beta^n \tag{3.9}$$

uniformly for all $A \in \mathcal{A}_{m-1},\ B \in \mathcal{A}_{m+n,\infty}$ and all $m, n \in \mathbb{N}$.

Lemma 3.4.

$$|\frac{E(Z_0 Z_\ell)}{EZ_0^2}| = O(\beta^{\ell-m}),\ \forall \quad \ell \geq m. \tag{3.10}$$

Proof. (3.9) implies that

$$|E(Z_0 Z_\ell) - EZ_0 \cdot EZ_\ell| \leq C \cdot E|Z_0| \cdot E|Z_\ell| \cdot \beta^{\ell-m},\ \forall \quad \ell \geq m.$$

(3.10) follows since $EZ_j = 0,\ \forall \quad j \in \mathbb{N}$. $\square$

Lemma 3.5. *Let $\nu \in \mathbb{N}$ satisfy $\nu \sim n^b$ as $n \to \infty$ with $b \in (0,1]$. Then*

$$\frac{E(Z_0 + \ldots + Z_{\nu-1})^2}{\nu \cdot EZ_0^2} = O(1), \quad \text{as} \qquad n \to \infty. \tag{3.11}$$

Proof.

$$\begin{aligned} LHS &= 1 + 2\sum_{\ell=1}^{\nu-1}(1 - \frac{\ell}{\nu}) \cdot \frac{E(Z_0 Z_\ell)}{EZ_0^2} \\ &= 1 + 2\sum_{\ell=1}^{m-1} \frac{E(Z_0 Z_\ell)}{EZ_0^2} + 2\sum_{\ell=m+1}^{\nu-1} \frac{E(Z_0 Z_\ell)}{EZ_0^2} - \frac{2}{\nu}\sum_{\ell=1}^{\nu-1} \ell \cdot \frac{E(Z_0 Z_\ell)}{EZ_0^2}. \end{aligned}$$

By (3.10),

$$2\sum_{\ell=m+1}^{\nu-1}\frac{E(Z_0Z_\ell)}{EZ_0^2}=O(1), \text{ as } \quad n\to\infty.$$

Moreover, for $1\le \ell\le m$.

$$E(Z_0Z_\ell)=P\left((X_0,\dots,X_{m-1})=(X_\ell,\dots,X_{\ell+m-1})=(y_0,\dots,y_{m-1})\right)-\left(P_m^{(n)}(y)\right)^2,$$

$$EZ_0^2=P_m^{(n)}(y)\cdot\left(1-P_m^{(n)}(y)\right);$$

And

$$P\left((X_0,\dots,X_{m-1})=(X_\ell,\dots,X_{\ell+m-1})=(y_0,\dots,y_{m-1})\right)/P_m^{(n)}(y)$$

$$=P\left((X_\ell,\dots,X_{\ell+m-1})=(y_0,\dots,y_{m-1})|(X_0,\dots,X_{m-1})=(y_0,\dots,y_{m-1})\right)$$

$$=P\left(X_m=y_{m-\ell},\dots,X_{\ell+m-1}=y_{m-1}|X_0=y_0,\dots,X_{m-1}=y_{m-1}\right)$$

$$=P\left(X_m=y_{m-\ell}|X_0=y_0,\dots,X_{m-1}=y_{m-1}\right)$$

$$\cdot P\left(X_{m+1}=y_{m-\ell+1}|X_0=y_0,\dots,X_{m-1}=y_{m-1},X_m=y_{m-\ell}\right)$$

$$\dots$$

$$\cdot P\left(X_{m+\ell-1}=y_{m-1}|X_0=y_0,\dots,X_{m-1}=y_{m-1},X_m=y_{m-\ell},\dots,X_{m+\ell-2}=y_{m-2}\right)$$

$$\le e^{-b\ell} \qquad \text{by (3.1)}\cdot \qquad \left(b=-\log(1-e^{-a})\right)$$

Therefore,

$$\left|\frac{2}{\nu}\sum_{\ell=1}^{m}\ell\cdot\frac{E(Z_0Z_\ell)}{EZ_0^2}\right|\le\frac{2}{\nu(1-P_m^{(n)}(y))}\sum_{\ell=1}^{m}\ell e^{-b\ell}+\frac{2}{\nu(1-P_m^{(n)}(y))}\sum_{\ell=1}^{m}\ell P_m^{(n)}(y)$$
$$\to 0, \qquad \text{as } n\to\infty;$$

And by the Kronecker lemma,

$$\left|\frac{2}{\nu}\sum_{\ell=m+1}^{\nu-1}\ell\cdot\frac{E(Z_0Z_\ell)}{EZ_0^2}\right|\le\frac{2C}{\nu}\sum_{\ell=m+1}^{\nu-1}\ell\beta^{\ell-m}\to 0, \quad \text{as} \quad n\to\infty.$$

Finally,

$$\left|2\sum_{\ell=1}^{m}\frac{E(Z_0Z_\ell)}{EZ_0^2}\right|\le 2\sum_{\ell=1}^{m}\frac{e^{-b\ell}}{1-P_m^{(n)}(y)}+2\sum_{\ell=1}^{m}\frac{P_m^{(n)}(y)}{1-P_m^{(n)}(y)}$$
$$\to\frac{2\alpha}{1-\alpha}, \qquad \text{as} \quad n\to\infty.$$

Thus (3.11) follows. □

The next lemma indicates that $\{U_i,\ i=1,\ \ldots,\ k\}$ is similar to an iid sequence.

Lemma 3.6. *For every* $t>0$,

$$E\left[\exp\left(\frac{t}{b_n}\sum_{i=1}^{k}U_i\right)\right]=\left\{E\left[\exp\left(\frac{t}{b_n}U_1\right)\right]\right\}^k(1+o(1)),\quad \text{as}\quad n\to\infty. \tag{3.12}$$

Proof. Applying (3.9) to the sequence $\{U_i,\ i=1,\ \ldots,\ k\}$ iteratively gives that

$$\left(1-C\beta^{q-m}\right)^{k-1}\le\frac{E\left[\exp\left(\frac{t}{b_n}\sum_{i=1}^{k}U_i\right)\right]}{\left\{E\left[\exp\left(\frac{t}{b_n}U_1\right)\right]\right\}^k}\le\left(1+C\beta^{q-m}\right)^{k-1}.$$

Since

$$|\left(1\pm C\beta^{q-m}\right)^{k-1}-1|\le Ck\beta^{q-m}\to 0,\qquad \text{as } n\to\infty,$$

(3.12) follows. □

Lemma 3.7. *For every* $t>0$,

$$\left\{E\left[\exp\left(\frac{t}{b_n}U_1\right)\right]\right\}^k=O(1),\quad \text{as}\quad n\to\infty. \tag{3.13}$$

Proof. By Taylor expansion,

$$E\left[\exp\left(\frac{t}{b_n}U_1\right)\right]=1+\frac{t^2}{2}\cdot\frac{EU_1^2}{b_n^2}+\frac{\theta t^3}{3!}\cdot\frac{EU_1^3}{b_n^3},$$

where $|\theta|\le 1$ may be different on each appearance.

By (3.11),

$$\frac{EU_1^2}{b_n^2}=O\left(\frac{p}{n}\right)=O\left(\frac{1}{n^{\frac{1}{2}-\lambda}}\right),\qquad \text{as } n\to\infty;$$

And the same argument as in [6] Lemma 5.4.8 implies that

$$E|U_1|^3 = O\left((EU_1^2)^{\frac{3}{2}}\right) \text{ as } n \to \infty.$$

Hence as $n \to \infty$

$$k \cdot \frac{EU_1^2}{b_n^2} = O(1),$$

and

$$k \cdot \frac{EU_1^3}{b_n^3} = o(1).$$

Therefore,

$$\left\{E\left[\exp\left(\frac{t}{b_n}U_1\right)\right]\right\}^k = \left(1 + \frac{t^2}{2} \cdot \frac{EU_1^2}{b_n^2} + \frac{\theta t^3}{3!}\frac{EU_1^3}{b_n^3}\right)^k = O(1), \quad \text{as } n \to \infty. \qquad \square$$

The main result is

Theorem 3.8. *For every $\delta > 0$, there exist $\gamma > 0$ and $n_0 \in \mathbb{N}$ such that*

(3.14) $$p\left(|\sum_{i=1}^k U_i| > \delta b_n^2\right) \le e^{-\delta n^\gamma},$$

uniformly for all $y \in \Sigma^+$ and all $n > n_0$.

Proof. It suffices to verify the inequality

(3.15) $$p\left(\sum_{i=1}^k U_i > \delta b_n^2\right) \le e^{-\delta n^\gamma}.$$

For every $t > 0$ and n sufficiently large,

$$\begin{aligned} P\left(\sum_{i=1}^k U_i > \delta b_n^2\right) &= P\left(\exp\left(\frac{t}{b_n}\sum_{i=1}^k U_i\right) > e^{t\delta b_n}\right) \\ &\le e^{-t\delta b_n} E\left[\exp\left(\frac{t}{b_n}\sum_{i=1}^k U_i\right)\right] \\ &= e^{-t\delta b_n} \cdot \left\{E\left[\exp\left(\frac{t}{b_n}U_1\right)\right]\right\}^k (1 + o(1)) \qquad \text{by (3.12)} \\ &= e^{-t\delta b_n} \cdot O(1) \qquad \text{by (3.13)}. \end{aligned}$$

(3.15) follows by setting $0 < \gamma < \frac{1-ac}{2}$. $\square$

Since the same argument shows that

$$P\left(|\sum_{j=1}^{k-1} V_j| > \delta b_n^2\right) \le e^{-\delta n^\gamma}, \tag{3.16}$$

and

$$P\left(|V_k| > \delta b_n^2\right) \le e^{-\delta n^\gamma}, \tag{3.17}$$

uniformly for all $y \in \Sigma^+$ and $n > n_0$, by combining (3.14), (3.16) and (3.17) we obtain

Corollary 3.9. *For every $\varepsilon > 0$,*

$$P\left(|\frac{N_m^{(n)}(y)}{b_n^2} - 1| > \varepsilon\right) \le e^{-\varepsilon n^\gamma} \tag{3.18}$$

uniformly for all $y \in \Sigma^+$ and $n > n_0$.

Proof of Theorem 2.2 and Theorem 2.3.

First by (3.4)

$$\sup_{y\in\Sigma^+} |T_n(y) - e^{f(y)}| \le \sup_{y\in\Sigma^+} D_n^{(1)}(y) + \sup_{y\in\Sigma^+} D_n^{(2)}(y) + \sup_{y\in\Sigma^+} D_n^{(3)}(y).$$

Then recall that each coordinate of $y \in \Sigma^+$ may take r different values. Thus

$$P\left(\sup_{y\in\Sigma^+} D_n^{(i)}(y) > \varepsilon\right) \le r^m P\left(D_n^{(i)}(y) > \varepsilon\right), \; i = 2, \; 3.$$

Hence Theorem 2.2 follows from (3.5), (3.6), (3.7), (3.18) and the Borel–Cantelli lemma.

Furthermore, for every $f \in \mathcal{H}$, the quantity $a = \frac{2\|f\|_\rho}{1-\rho}$ satisfies

$$1 - ac_n > 0$$

for n sufficiently large. Theorem 2.3 is proved just like Theorem 2.2.

Acknowledgement

This work constitutes a part of the author's doctoral dissertation, which was written under the supervision of Professor Steven Lalley. The author gratefully acknowledges Professor Lalley's guidance and support.

References

[1] Bowen,, R. (1975). Equilibrium states and the ergodic theory of Anosov diffeomorphisms. Lecture Notes in Math. 470. Springer-Verlag, New York.

[2] Ji, C. (1987). Estimating functionals of one-dimensional Gibbs states. Technical Report #87-33, Department of Statistics, Purdue University.

[3] Lalley, S. P. (1985). Ruelle's Perron-Frobenius theorem and the central limit theorem for additive functionals of one-dimensional Gibbs states. Proc. Conf. in honor of H. Robbins.

[4] Ripley, B. D. (1981). Spatial statistics. John Wiley and Sons, New York.

[5] Ruelle, D. (1978). Theomodynamic formalizm. Addison-Wesley, Reading, Massachusetts.

[6] Stout, W.F. (1974). Almost sure convergence. Academic Press, New York.