ESTIMATING POTENTIAL FUNCTIONS OF ONE-DIMENSIONAL GIBBS STATES UNDER CONSTRANTS*

by

Chuanshu Ji

Department of Statistics
University of North Carolina
Chapel Hill, NC 27514
USA

SUMMARY

Some consistent estimators are constructed for estimating potential functions of one-dimensional Gibbs states. Certain normalization constraints are imposed to resolve the identifiability problem. The step-length selection is also discussed in terms of the convergence rates of those estimators.

Key Words and Phrases: potential function, Gibbs state, consistent estimator

Running Title: Estimation of Potential Functions for Gibbs States

*Research partially supported by a David Ross Fellowship at Purdue University

1. Introduction and Background

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 $1,\ldots,r$. 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

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

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 observed that f determines the dependence in the stationary sequence $X = (X_0, X_1,\ldots)$ which has the probability distribution μ_f.

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. The motivation for considering such a problem is mentioned in [2]. 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: reparameterization 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 (in sup-norm) estimators T_n for the unknown function e^f are constructed.

After renormalization e^f becomes an infinite-step backward transition function (See (2.5)). This suggests us to use a sequence of finite-step (backward) transition functions $\{g_m, m \in \mathbb{N}\}$ to approximate e^f, and at each step m to estimate g_m by a "sample transition function" which is a ratio of two empirical measures. A key question is what is the appropriate order for the step-length m as the sample size n tends to infinity. Some heuristic arguments indicate that m should be of the order log n so that T_n can achieve the "nearly best" convergence rate among all consistent estimators of e^f.

For simplicity, we only consider the case of the sample space Σ^+ in this paper. However, all results here can be extended to the case of a more general sample space Σ_A^+ in which transitions between certain states are not allowed. The definition and description of Σ_A^+ are given in [2].

Now we define Gibbs states rigorously by Ruelle-Perron-Frobenius theory.

(1) **Forward shift:** Recall that our sample space is $\Sigma^+ = \prod_{i=0}^{\infty} \{1,\dots,r\}$, which is compact and metrizable in the product topology.

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

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

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

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 $f, g \in C(\Sigma^+)$, define $\mathcal{L}_f : C(\Sigma^+) \to C(\Sigma^+)$ by

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

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^+ 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,*

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

The proof may be found in [1], [4].

(4) **Gibbs states**: Assume that $\int h_f d\nu_f = 1$. For each real-valued $f \in \mathcal{F}_\rho^+$, 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 $M_\sigma(\Sigma^+)$ denote the set of all σ-invariant probability measures on Σ^+.

Theorem 1.2. *For each real-valued $f \in \mathcal{F}_\rho^+$, there exist constants $c_1, c_2 \in (0,\infty)$ such that (1.1) holds for all $x \in \Sigma^+$ and all $m \in \mathbb{N}$; and μ_f is the* *unique element in* $M_\sigma(\Sigma^+)$ *satisfying* (1.1). *In* (1.1), $p = p(f) = \log \lambda_f$ *is the pressure for* f.

The proof is given in [1].

Remark 1.3. Two functions f, $g \in C(\Sigma^+)$ are said to be homologous, written $f \sim g$, if there exists $\varphi \in C(\Sigma^+)$ such that

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

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 1.4. The Gibbs state model includes the following special cases: Let $X = (X_0, X_1, \ldots)$ be a stationary sequence with underlying distribution μ_f, then

(1) If $f(x) \equiv c$, for all $x \in \Sigma^+$, then X is a sequence of iid random variables with discrete uniform distribution.

(ii) 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 $P(X_0=\ell) = ce^{f(\ell)}$, $\ell=1,\ldots,r$, where $c = 1/\Sigma_{\ell=1}^r e^{f(\ell)}$.

(iii) If $f(x) = f(x_0,x_1)$, for all $x \in \Sigma^+$, i.e., f only depends on the first two coordinates, then X forms a stationary Markov chain with state space $\{1,\ldots,r\}$ and suitable transition probabilities.

(iv) If $f(x) = f(x_0,\ldots,x_k)$, for all $x \in \Sigma^+$ and some $k \in \mathbb{N}$, i.e., f only depends on the first k+1 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 (See Remark 1.3). 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) $\tilde{f} \sim f + \text{constant}$.

Proof. Let

(2.1) $\tilde{f} = f + \log h_f - \log h_f \circ \sigma - \log \lambda_f$,

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

Furthermore, by [3] Proposition 1 we have

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

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

$$(2.3) \quad \lim_{m\to\infty} \mu_f(x_0|x_1,\dots,x_{m-1}) = \lim_{m\to\infty} \frac{\mu_f(y : y_i=x_i,\ 0 \le i \le m-1)}{\mu_f(y : y_i=x_i,\ 1 \le i \le m-1)}$$

exists for almost every $x \in \Sigma^+$ under μ_f, the LHS in (2.2) is well-defined as the limit in (2.3). Therefore, the 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 $f \in \mathcal{H}$.

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) $\sup_{y \in \Sigma^+} |T_n(y) - e^{f(y)}| \to 0$, a.s. under μ_f 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 constraints

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

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

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

So e^f may be regarded as an infinite-step backward transition function, which sheds light on the construction of 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 in that situation. Further discussion on the choice of the step-length m will be given in Section 4.

Construction of Consistent Estimator T_n

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

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

$R_m^{(n)}(y)$, also written as $\dfrac{N_m^{(n)}(y)}{n} \Big/ \dfrac{N_{m-1}^{(n)}(y)}{n}$, is the "sample conditional frequency" 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 just a strongly consistent estimator of e^f.

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

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

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

Let

(2.6) $\bar{a} = \frac{2K}{1-\rho}$ *and*

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

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

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

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 procedure also satisfies* (2.4).

Procedure 2.4. Choose a sequence of positive constants $\{c_n,\ n \in N\}$, 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 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* $f \in \mathcal{F}$, *there exists a positive constant* a *which depends on* f, *such that*

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

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

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

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

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

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

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

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

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

$$(3.3)\quad a = \frac{2\|f\|_\rho}{1-\rho}.$$

■

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

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

and

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

Then

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

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

$$(3.4)\quad |T_n(y) - e^{f(y)}| \le \left|\frac{P_m^{(n)}(y)}{P_{m-1}^{(n)}(y)} - e^{f(y)}\right| + I_{(N_{m-1}^{(n)}(y)=0)} \cdot \frac{P_m^{(n)}(y)}{P_{m-1}^{(n)}(y)}$$

$$+ I_{(N_{m-1}^{(n)}(y)>0)} \left|\frac{N_m^{(n)}(y)}{N_{m-1}^{(n)}(y)} - \frac{P_m^{(n)}(y)}{P_{m-1}^{(n)}(y)}\right| \triangleq D_n^{(1)}(y) + D_n^{(2)}(y) + D_n^{(3)}(y).$$

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

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

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

For every $\epsilon \in (0,½)$,

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

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

$$(3.7)\quad P(D_n^{(3)}(y) > 2\epsilon) \le P\left[\left|\frac{N_{m-1}^{(n)}(y)}{nP_{m-1}^{(n)}(y)} - 1\right| > \delta_1\right] + P\left[\left|\frac{N_{m-1}^{(n)}(y)}{nP_{m-1}^{(n)}(y)} - 1\right| > \delta_2\right] .$$

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

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

$$P(D_n^{(3)}(y) > 2\epsilon)$$

$$\le P(|N_m^{(n)}(y) - nP_m^{(n)}(y)| > \epsilon N_m^{(n)}(y))$$

$$+ P(|N_{m-1}^{(n)}(y)-nP_{m-1}^{(n)}(y)| > \frac{\epsilon}{1-e^{-a}} \cdot N_{m-1}^{(n)}(y))$$

$$\le P((1+\epsilon)|N_m^{(n)}(y) - nP_m^{(n)}(y)| > \epsilon nP_m^{(n)}(y))$$

$$+ P((1 + \frac{\epsilon}{1-e^{-a}})|N_{m-1}^{(n)}(y) - nP_{m-1}^{(n)}(y)| > \frac{\epsilon}{1-e^{-a}} \cdot nP_{m-1}^{(n)}(y))$$

$$= P(|\frac{N_m^{(n)}(y)}{nP_m^{(n)}(y)} - 1| > \delta_1) + P(|\frac{N_{m-1}^{(n)}(y)}{nP_{m-1}^{(n)}(y)} - 1| > \delta_2). \qquad \blacksquare$$

(3.6) and (3.7) indicate that it suffices to evaluate

$P(|\frac{N_m^{(n)}(y)}{nP_m^{(n)}(y)} - 1| > \epsilon)$ for large n.

Now let

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

Then

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

and

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

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

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

k satisfies

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

Let

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

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

$$\dots$$

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

And

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

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

$$\dots$$

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

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

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

The idea is that for large n both $\{U_i,\ i=1,\dots,k\}$ and $\{V_j,\ j=1,\dots,k-1\}$ 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,

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

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

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

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

$$(3.8) \quad \left|\frac{P(A\cap B)}{P(A)\cdot P(B)} - 1\right| \le C\beta^n$$

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

Lemma 3.3.

$$(3.9)\qquad \left|\frac{E(Z_0Z_\ell)}{E\,Z_0^2}\right| = O(\beta^{\ell-m}),\ \forall\ \ell \geq m.$$

Proof. (3.8) implies that

$$|E(Z_0Z_\ell) - E\,Z_0 \cdot E\,Z_\ell| \leq C\cdot E|Z_0| \cdot E(Z_\ell| \cdot \beta^{\ell-m},\ \forall\ \ell \geq m.$$

(3.9) follows since $E\,Z_j = 0,\ \forall\ j \in \mathbb{N}$. ■

Lemma 3.4. *Let $\nu \in \mathbb{N}$ satisfy $\nu \sim n^b$ as $n \to \infty$ with $b \in (0,1]$. Then*

$$(3.10)\qquad \frac{E(Z_0+\dots+Z_{\nu-1})^2}{\nu\cdot E\,Z_0^2} = O(1),\ as\quad n \to \infty.$$

Proof. $\text{LHS} = 1 + 2 \sum_{\ell=1}^{\nu-1} (1 - \frac{\ell}{\nu}) \cdot \frac{E(Z_0Z_\ell)}{E\,Z_0^2}$

$$= 1 + 2 \sum_{\ell=1}^{m-1} \frac{E(Z_0Z_t)}{E\,Z_0^2} + 2 \sum_{\ell=m+1}^{\nu-1} \frac{E(Z_0Z_\ell)}{E\,Z_0^2} - \frac{2}{\nu} \sum_{\ell=1}^{\nu-1} \ell \cdot \frac{E(Z_0Z_\ell)}{E\,Z_0^2}.$$

By (3.9),

$$2 \sum_{\ell=m+1}^{\nu-1} \frac{E(Z_0Z_\ell)}{E\,Z_0^2} = O(1),\quad \text{as } n \to \infty.$$

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

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

$$E\,Z_0^2 = P_m^{(n)}(y) \cdot (1 - P_m^{(n)}(y));$$

And

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

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

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

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

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

$$\dots$$

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

$$\le e^{-b\ell} \text{ by (3.1).} \quad (b = -\log(1-e^{-a}))$$

Therefore,

$$\left|\frac{2}{\upsilon}\sum_{\ell=1}^{m} \ell \cdot \frac{E(Z_0 Z_\ell)}{E\, Z_0^2}\right| \le \frac{2}{\upsilon(1-P_m^{(n)}(y))}\sum_{\ell=1}^{m} \ell e^{-b\ell} + \frac{2}{\upsilon(1-P_m^{(n)}(y))}\sum_{\ell=1}^{m} \ell P_m^{(n)}(y)$$

$$\to 0, \text{ as } n \to \infty;$$

And by the Kronecker lemma,

$$\left|\frac{2}{\upsilon}\sum_{\ell=m+1}^{\upsilon-1} \ell \cdot \frac{E(Z_0 Z_\ell)}{E\, Z_0^2}\right| \le \frac{2C}{\upsilon}\sum_{\ell=m+1}^{\upsilon-1} \ell\beta^{\ell-m} \to 0, \text{ as } n \to \infty.$$

Finally,

$$\left|2\sum_{\ell=1}^{m} \frac{E(Z_0 Z_\ell)}{E\, Z_0^2}\right| \le 2\sum_{\ell=1}^{m} \frac{e^{-b\ell}}{1-P_m^{(n)}(y)} + 2\sum_{\ell=1}^{m} \frac{P_m^{(n)}(y)}{1-P_m^{(n)}(y)}$$

$$\to \frac{2\alpha}{1-\alpha}, \text{ as } n \to \infty.$$

Thus (3.10) follows. ∎

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

Lemma 3.5. *For every $t > 0$,*

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

Proof. Applying (3.8) to the sequence $\{U_i, i=1,\ldots,k\}$ iteratively gives that

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

Since

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

(3.11) follows. ■

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

(3.12) $$\{E[\exp(\frac{t}{b_n}U_1)]\}^k = O(1), \quad \textit{as } n \to \infty.$$

Proof. By Taylor expansion,

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

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

By (3.10),

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

And the same argument as in [5] Lemma 5.4.8 implies that

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

Hence $n \to \infty$

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

and

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

Therefore,

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

The main result is

Theorem 3.7. *For every $\delta > 0$, there exist $\gamma > 0$ and $n_0 \in \mathbb{N}$ such that*

$$(3.13) \quad P(|\sum_{i=1}^{k} U_i| > \delta b_n^2) \le e^{-\delta n^\gamma},$$

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

Proof. It suffices to verify the inequality

$$(3.14) \quad P(\sum_{i=1}^{k} U_i > \delta b_n^2) \le e^{-\delta n^\gamma}.$$

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

$$\begin{aligned} P(\sum_{i=1}^{k} U_i > \delta b_n^2) &= P(\exp(\frac{t}{b_n} \sum_{i=1}^{k} U_i) > e^{t\delta b_n}) \\ &\le e^{-t\delta b_n} E[\exp(\frac{t}{b_n} \sum_{i=1}^{k} U_i)] \\ &= e^{-t\delta b_n} \cdot \{E[\exp(\frac{t}{b_n} U_1)]\}^k (1 + o(1)) \qquad \text{by (3.11)} \\ &= e^{-t\delta b_n} \cdot O(1) \qquad \text{by (3.12)}. \end{aligned}$$

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

Since the same argument shows that

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

and

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

uniformly for all $y \in \Sigma^+$ and $n > n_0$, by combining (3.13), (3.15) and (3.16) we obtain

Corollary 3.8. *For every* $\epsilon > 0$,

(3.17) $$P(|\frac{N_m^{(n)}(y)}{b_n^2} - 1| > \epsilon) \le e^{-\epsilon n^\gamma},$$

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

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

Hence Theorem 2.2 follows from (3.5), (3.6), (3.7), (3.17) and the Borel–Cantelli lemma.

Furthermore, for every $f \in \mathcal{F}$, 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.

4. Remark on the step-length selection

Many consistent estimators T_n could be constructed in the same way as in Section 2 provided the step-length m tends to infinity "not too fast". Therefore their convergence rates need to be taken into consideration. In this section we explain why m should be of the order log n and what is the corresponding convergence rate.

First of all, we have a stronger theorem than Theorem 2.2.

Theorem 4.1. *Suppose f is an unknown potential function satisfying (A1) and (A2) in Theorem 2.2. Let* $\bar{a} = \frac{2K}{1-\rho}$ *and* $m = [c \log n]$ *(same as (2.6), (2.7)), where the constant c satisfies*

$$(4.1) \quad \frac{\lambda}{-\log \rho} < c < \frac{1-2\lambda}{\bar{a}} ;$$

and λ *is a constant satifying*

$$(4.2) \quad 0 < \lambda < \frac{1}{2 + \frac{\bar{a}}{-\log \rho}} .$$

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

$$(4.3) \quad n^{\lambda} \sup_{y \in \Sigma^+} |T_n(y) - e^{f(y)}| \to 0, \quad \textit{a.s. under } \mu_f \textit{ as } n \to \infty.$$

Proof. The first inequality in (4.1) implies that

$$(4.4) \quad n^{\lambda} \rho^{m} \to 0 \quad \text{as} \quad n \to \infty.$$

Hence by (3.5),

$$(4.5) \quad n^{\lambda} \sup_{y\in\Sigma^+} D_n^{(1)}(y) \to 0, \quad \text{a.s. under } \mu_f \text{ as } n \to \infty.$$

Moreover, the second inequality in (4.1) allows us to obtain a stronger result than (3.17): For every $\epsilon > 0$, there exist $\gamma > 0$ and $n_0 \in \mathbb{N}$ such that

$$(4.6) \quad P\left(n^{\lambda} \left|\frac{N_m^{(n)}(y)}{b_n^2} - 1\right| > \epsilon\right) \leq e^{-\epsilon n^{\gamma}}$$

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

It follows from the same arguments in Section 3 that

$$(4.7) \quad n^{\lambda} \sup_{y\in\Sigma^+} D_n^{(i)}(y) \to 0, \quad i=2,3, \quad \text{a.s. under } \mu_f \text{ as } n \to \infty.$$

Therefore, (4.3) holds. ∎

Theorem 4.1 shows the sufficiency of the order log n for step-length m. Is it also necessary? Notice that the empirical measure $\frac{N_m^{(n)}(\cdot)}{n}$ plays a role of sufficient statistics in this nonparametric estimation problem. To derive consistent estimator T_n in (2.4), the ratio $\frac{N_m^{(n)}(y)}{n} / P_m^{(n)}(y)$ has to be close to one for every y. Hence $n\ P_m^{(n)}(y)$ should be large for every y. By (3.1) we have

$$n\ e^{-ma} \leq n\ P_m^{(n)}(y) \leq n\ e^{-mb}$$

uniformly for $y \in \Sigma^+$ and all $m \in \mathbb{N}$, where $b = -\log(1-e^{-a}) > 0$. So m should grow no faster than c log n for some $c > 0$. On the other hand, (3.4) suggests that there is a trade-off between the good approximation (evaluated by $|\mu_f(y_0|y_1,\ldots,y_{m-1}) - e^{f(y)}|$) and the accurate estimation at each step (evaluated by $|T_n(y)-\mu_f(y_0|y_1,\ldots,y_{m-1})|$). The convergence rate of the former part will be damaged if m grows too slowly. Therefore, log n is the right order for m and the constant c is determined by (4.1).

Let $\Lambda = \dfrac{1}{2 + \dfrac{\bar{a}}{-\log \rho}}$. Then for $\lambda \geq \Lambda$ no constant c will satisfy (4.1). Therefore (4.3) can not be established under our construction of T_n. We conjecture that in that situation no other methods can produce the result (4.3), either. i.e. if $\lambda \geq \Lambda$, let T_n be an arbitrary consistent estimator of e^f in the sense of (2.4). Then (4.3) fails for some f satisfying A(1) and (A2). For the time being, the rigorous proof is still in the process of development.

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 in H. Robbins.

[4] Ruelle, D. (1978). *Theomodynamic Formalism*. Addison-Wesley, Reading, Massachusetts.

[5] Stout, W.F. (1974). *Almost Sure Convergence*. Academic Press, New York.