ESTIMATING FUNCTIONALS OF ONE-DIMENSIONAL GIBBS STATES

by

Chuanshu Ji

Department of Statistics
University of North Carolina
Chapel Hill, NC 27599-3260
USA

SUMMARY

Some estimators of maximum likelihood type are constructed for estimating functionals of one-dimensional Gibbs states. We also show that those estimators are strongly consistent, asymptotically normal and asymptotically efficient.

Key words and phrases: Gibbs state, maximum likelihood estimator, asymptotic efficiency.

Running Title: Estimating Functionals of Gibbs States

*Research partially supported by NSF grant DMS-8702620

0. Introduction

Gibbs states were originally conceived as models in statistical mechanics (cf. Ruelle [10]), 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 [9]). However, using one-dimensional Gibbs states to model categorical time series seems to be a new idea.

Important examples of categorical time series arise in communications engineering. These series typically consist of long strings of symbols from a finite alphabet. The dependence in such series may be quite complicated: Gibbs states might be useful models in these contexts.

Further discussion of modeling binary time series may be found in Kedem [5].

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

$$(0.1) \quad c_1 \le \frac{\mu_f(y : y_i = x_i,\ 0 \le i \le m-1)}{\exp\{-mp + \Sigma_{j=0}^{m-1} f(\sigma^j x)\}} \le c_2$$

Define the forward shift operator σ: $\Sigma^+ \to \Sigma^+$ 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

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

for $0 < \rho < 1$ let

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

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], [10].

(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 measure 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*

$$(1.1) \quad c_1 \leq \frac{\mu_f(x_0,\ldots,x_{m-1})}{\exp\{-mp + \Sigma_{j=0}^{m-1} f(\sigma^j x)\}} \leq c_2, \quad \forall x \in \Sigma^+,\ m \in \mathbb{N}^+ = \mathbb{N}\backslash\{0\};$$

and μ_f *is the unique element in* $M_\sigma(\Sigma^+)$ *such that* (1.1) *holds, where* $\mu_f(x_0,\ldots,x_{m-1}) = \mu_f(y \in \Sigma^+;\ y_i = x_i,\ 0 \leq i \leq m-1)$.

The proof is given in [1]. Note that $p = p(f) = \log \lambda_f$ is called the pressure for f.

Remark 1.3. Two functions f, $g \in C(\Sigma^+)$ are said to be homologous, written $f \sim g$, if there exists $\phi \in C(\Sigma^+)$ 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 1.4. 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

(i) 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 $P(X_0=l) =$

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

(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 MLE for Estimating Certain Functionals of One-dimensional Gibbs States

In what follows, we assume that $X = (X_0, X_1, \ldots)$ is a stationary sequence with probability distribution μ_f and $x = (x_0, x_1, \ldots)$ is a specific value of X.

For real-valued f, $\psi \in \mathcal{F}_\rho^+$ define $\theta = \theta(\mu_f) = \int \psi \, d\mu_f = E_{\mu_f}\psi$. Suppose ψ is given but f is unknown. We consider the problem of estimating θ based on a finite number of observations $X_0,\ldots,X_{n-1}$. The parameter θ is just the expectation of the random variable ψ defined on Σ^+ under μ_f. For instance, when $\psi = I(x_0,\ldots,x_{k-1})$, θ equals the probability that the first k coordinates of X are exactly $x_0,\ldots,x_{k-1}$.

This estimation problem is connected with the weak topology on the space $M_\sigma(\Sigma^+)$. If μ_n is a sequence in $M_\sigma(\Sigma^+)$, then $\mu_n \to \mu$ in the weak (Lévy) topology ($\mu_n \xrightarrow{W} \mu$) iff $\int \psi d\mu_n \to \int \psi d\mu$ for all $\psi \in \mathcal{F}_\rho^+$.

Recall that $f \sim g$ implies $\mu_f = \mu_g$. As we mentioned in Section 0, considering the identifiability problem we estimate θ instead of f itself.

For every $\nu \in M_\sigma(\Sigma^+)$, recall that $\nu(x_0,\dots,x_{n-1}) = \nu(y \in \Sigma^+ : y_i = x_i, 0 \le i \le n-1)$, i.e. the probability of the cylinder set with the first n coordinates $x_0,\dots,x_{n-1}$.

Definition 2.1. Given $x_0,\dots,x_{n-1}$, if there exists $\hat{\mu}_n \in M_\sigma(\Sigma^+)$ such that $\hat{\mu}_n(x_0,\dots,x_{n-1}) \ge \nu(x_0,\dots,x_{n-1})$, for all $\nu \in M_\sigma(\Sigma^+)$, then $\hat{\mu}_n$ is called *a maximum likelihood summary (MLS)*, and $\theta(\hat{\mu}_n) = \int \psi d\hat{\mu}_n$ is called *an MLE* of θ.

The key to constructing MLE is to define periodic sequences.

Definition 2.2. $x \in \Sigma^+$ is said to be *a periodic sequence* with *period* $m \in \mathbb{N}^+$ if $\sigma^m x = x$. We call m *the smallest period* of x, denoted by $l(x)$, when $\sigma^m x = x$ but $\sigma^j x \ne x$ for all $j < m$. The set of all periodic sequences in Σ^+ is denoted by $\mathscr{C}$.

Observe that for every $x_0,\dots,x_{n-1}$ there exists a unique $x^{(l_n)} \in \mathscr{C}$ with smallest period l_n such that $x_j^{(l_n)} = x_j$, $j = 0,1,\dots,n-1$; and $l_n \le n$ for all $x_0,\dots,x_{n-1}$.

Define $\hat{\mu}_n \in M_\sigma(\Sigma^+)$ by

$$\hat{\mu}_n(\sigma^j x^{(l_n)}) = \frac{1}{l_n}, \quad j = 1,\dots,l_n.$$

Note that $\hat{\mu}_n$ is a σ-invariant probability measure which puts all its mass on the orbit of the sequence $x^{(l_n)}$.

Example 2.3. If $(x_0,x_1,x_2) = (1,2,1)$, $n = 3$, then $l_3 = 2$, $x^{(l_3)} = (1,2,1,2,\dots)$ and $\hat{\mu}_3(1,2,1,2,\dots) = \hat{\mu}_3(2,1,2,1,\dots) = \frac{1}{2}$.

Example 2.4. If $(x_0,x_1,x_2) = (1,2,3)$, $n = 3$, then $l_3 = 3$, $x^{(l_3)} = (1,2,3,1,2,3,\dots)$ and $\hat{\mu}_3(1,2,3,1,2,3,\dots) = \hat{\mu}_3(2,3,1,2,3,1,\dots) =$

$\hat{\mu}_3(3,1,2,3,1,2,\dots) = \frac{1}{3}$.

The next lemma shows that every element in $M_\sigma(\Sigma^+)$ can be approximated by a sequence of σ-invariant measures concentrated on $\mathcal{C}$.

Lemma 2.5. *For every $\nu \in M_\sigma(\Sigma^+)$ there exist $\nu_n \in M_\sigma(\Sigma^+)$, $n \in \mathbb{N}^+$ such that supp $\nu_n \subset \mathcal{C}$ for all $n \in \mathbb{N}^+$ and $\nu_n \overset{W}{\to} \nu$ as $n \to \infty$.*

Proof. For every $x_0,\dots,x_{n-1}$, define $x(n) \in \mathcal{C}$ by $x(n) = (x_0,\dots,x_{n-1}; x_0,\dots,x_{n-1};\dots)$, then define ν_n by

(a) $\nu_n(x(n)) = \frac{1}{n}\{\nu(x_0,\dots,x_{n-1}) + \nu(x_1,\dots,x_{n-1}, x_0) + \dots + \nu(x_{n-1},x_0,\dots,x_{n-2})\}$

and

(b) $\nu_n(\sigma^j\, x(n)) = \nu_n(x(n))$, $j = 1,\dots,n$.

Observe that $\nu_n \in M_\sigma(\Sigma^+)$ only assigns positive mass to the periodic sequences with period n, hence *supp* $\nu_n \subset \mathcal{C}$. To show $\nu_n \overset{W}{\to} \nu$, only cylinder sets need to be considered. For any $m \in \mathbb{N}^+$ and $x_0,\dots,x_{m-1}$, when $n > m$

$$\nu_n(x_0,\dots,x_{m-1}) = \sum_{x_m,\dots,x_{n-1}} \nu_n(x_0,\dots,x_{m-1},x_m,\dots,x_{n-1})$$

$$= \frac{1}{n}\{ \sum_{x_m,\dots,x_{n-1}} \nu(x_0,\dots,x_{m-1},\underline{x_m,\dots,x_{n-1}})$$

$$+ \sum_{x_m,\dots,x_{n-1}} \nu(x_1,\dots,x_{m-1}, \underline{x_m,\dots,x_{n-1}},x_0)$$

$$+ \dots + \sum_{x_m,\dots,x_{n-1}} \nu(\underline{x_{n-1}},x_0,\dots,x_{m-1},\underline{x_m,\dots,x_{n-2}})\},$$

Since $\nu \in M_\sigma(\Sigma^+)$, at most m-1 sums in $\{\dots\}$ differ from $\nu(x_0,\dots,x_{m-1})$. Therefore,

$$|\nu_n(x_0,\dots,x_{m-1}) - \nu(x_0,\dots,x_{m-1})| \le \frac{m-1}{n} \to 0$$

as $n \to \infty$. ■

Theorem 2.6. *$\hat{\mu}_n$ is an MLS.*

Proof. For an arbitrary $\nu \in M_\sigma(\Sigma^+)$ let $\{\nu_k\}$ be a corresponding approximating sequence given in Lemma 2.5. It suffices to show that for every $x_0,\dots,x_{n-1}$,

$$\hat{\mu}_n(x_0,\dots,x_{n-1}) \ge \nu_k(x_0,\dots,x_{n-1}),\ \forall\, k \in \mathbb{N}^+.$$

Introduce an equivalence relation $\sim$ in $\mathcal{C}$ as follows: for $x, y \in \mathcal{C}$, $x \sim y$ if $\sigma^j x = y$ for some $j \in \mathbb{N}^+$. Then we have

(i) $x \sim y$ implies $l(x) = l(y)$;

(ii) $\mathcal{C}$ is a countable union of disjoint equivalence classes;

(iii) for each $k \in \mathbb{N}^+$, there exist a finite number of equivalence classes $E_1,\dots,E_K$ such that $supp\ \nu_k = \cup_{i=1}^K E_i$.

Let $y^{(i)}$ be a representative element of the class E_i, then $E_i = \{y^{(i)}, \sigma y^{(i)},\dots,\sigma^{l(y^{(i)})-1} y^{(i)}\}$. Let $B_n = \{y \in \mathcal{C} : y_j = x_j,\ j = 0,1,\dots,n-1\}$, then

$$\nu_k(x_0,\dots,x_{n-1}) = \sum_{i=1}^{K} \nu_k(E_i \cap B_n)$$

$$= \sum_{i=1}^{K} \nu_k(y^{(i)}) \cdot \#\{j : 1 \le j \le l(y^{(i)}),\ \sigma^j(y^{(i)}) \in B_n\}$$

$$= \sum_{i=1}^{K} \{\nu_k(E_i)/l(y^{(i)})\} \cdot \#\{j : 1 \le j \le l(y^{(i)}),\ \sigma^j(y^{(i)}) \in B_n\}$$

$$\le \max_{1 \le i \le K} [\#\{j : 1 \le j \le l(y^{(i)}),\ \sigma^j(y^{(i)}) \in B_n\}/l(y^{(i)})]$$

$$\le \#\{j : 1 \le j \le l_n,\ \sigma^j x^{(l_n)} \in B_n\}/l_n$$

$$= \hat{\mu}_n(x_0,\ldots,x_{n-1})$$

The last inequality follows from the fact that those $\sigma^j x^{(l_n)}$, $j = 1,\ldots,l_n$ in *supp* ν_k must belong to some equivalence class E_m, hence $l_n = l(y^{(m)})$ and

$$\#\{j : 1 \le j \le l_n,\ \sigma^j x^{(l_n)} \in B_n \cap \mathit{supp}\ \nu_k\}/l_n$$

$$= \#\{j : 1 \le j \le l(y^{(m)}),\ \sigma^j y^{(m)} \in B_n\}/l(y^{(m)})$$

$$= \max_{1\le i\le K} [\#\{j : 1 \le j \le l(y^{(i)}),\ \sigma^j y^{(i)} \in B_n\}/l(y^{(i)})]. \qquad \blacksquare$$

3. Consistency and Asymptotic Normality of MLE

The main result of this section is

Theorem 3.1.

(i) *Strong consistency of MLE:* $\theta(\hat{\mu}_n) \to \theta$ *a.s. as* $n \to \infty$ *under* μ_f.

(ii) *Asymptotic normality of MLE: Let*

$$\beta_f(z) = \log \frac{\lambda_{f+z\psi}}{\lambda_f},\quad z \in \mathbb{R},$$

and $\Phi(t)$ *be the standard normal cdf. Then for every* $t \in \mathbb{R}$

$$\mu_f(x \in \Sigma^+ : \sqrt{n/\beta_f''(0)} \cdot (\theta(\hat{\mu}_n)(x) - \theta) \le t) \to \Phi(t) \quad \text{as } n \to \infty,$$

where $\beta_f''(0) = \dfrac{d^2\beta_f(z)}{dz^2} \Big|\ z=0.$

To prove Theorem 3.1 we need several lemmas.

Lemma 3.2. *(Weak Bernoulli property of Gibbs states)*

Let $\mathcal{A}_{m-1}$ *and* $\mathcal{A}_{m+n,\infty}$ *be the σ-fields generated by* $(X_0,\ldots,X_{m-1})$ *and* $(X_i,$ $i \ge m+n)$ *respectively. Then there exist* $B > 0$, $\alpha \in (0,1)$ *such that*

$$\left| \frac{\mu_f(A_1 \cap A_2)}{\mu_f(A_1) \cdot \mu_f(A_2)} - 1 \right| \leq B\, \alpha^n$$

uniformly for all $A_1 \in \mathcal{A}_{m-1}$, $A_2 \in \mathcal{A}_{m+n,\infty}$ *and all* m,n, $\in \mathbb{N}$.

See Bowen [1], Theorem 1.25 for the proof.

Lemma 3.3. *For each Gibbs state* μ_f, *there exists* $\beta \in (0,1)$ *such that*

$$\mu_f(x_0,\dots,x_{n-1}) \leq \beta^n,\ \forall\ x_0,\dots,x_{n-1}.$$

Proof. (1.1) implies that

$$\epsilon \leq \mu_f(y \in \Sigma^+ : y_{m-1} = x_{m-1} | y_j = x_j,\ j = 0,\dots,m-2) \leq 1 - \epsilon$$

for some $\epsilon \in (0,1)$ and for all $x_0,\dots,x_{m-1}$ and $m \in \mathbb{N}^+$.

The lemma follows by setting $\beta = 1-\epsilon$. ■

Lemma 3.4. *Under* μ_f, *as* $n \to \infty$

(i) $\dfrac{n-l_n}{\sqrt{n}} \to 0$ *a.s.*

and

(ii) $\dfrac{l_n}{n} \to 1$ *a.s.*

Proof. It suffices to prove (i). Let P(E) be the probability of an event E under μ_f. For any $\delta \in (0,1)$

$$P\left[\frac{n-l_n}{\sqrt{n}} > \delta\right] \leq \sum_{k=1}^{[n-\delta\sqrt{n}]} P(l_n = k),$$

where [c] denotes the integer part of $c \in \mathbb{R}$. If we can show that for every k = $1,\dots,[n-\delta\sqrt{n}]$, $P(l_n{=}k)$ goes to zero exponentially as $n \to \infty$, then the Borel-Cantelli lemma implies (i).

<u>Case 1</u>: $[\sqrt{n}] + 1 \leq k \leq [n-\delta\sqrt{n}]$

Since $\ell_n = k$ implies $(X_0,\dots,X_{[n^{1/4}]-1}) = (X_k,\dots,X_{k+[n^{1/4}]-1})$, for sufficiently large n we have

$$P(l_n=k) \le P((X_0,\dots,X_{[n^{1/4}]-1}) = (X_k,\dots,X_{k+[n^{1/4}]-1}))$$

$$\le \sum_{x_0,\dots,x_{[n^{1/4}]-1}} \mu_f(y \in \Sigma^+ : y_j = y_{j+k} = x_j,\ j = 0,\dots,[n^{1/4}] - 1)$$

$$\le \sum_{x_0,\dots,x_{[n^{1/4}]-1}} \left\{\mu_f^2(x_0,\dots,x_{[n^{1/4}]-1}) + B\alpha^{k-[n^{1/4}]}\right\} \quad \text{(by Lemma 3.2)}$$

$$\le \beta^{[n^{1/4}]} + r^{[n^{1/4}]} \cdot B\alpha^{k-[n^{1/4}]} \quad \text{(by Lemma 3.3)}$$

$$\le C\,\gamma^{[n^{1/4}]}, \quad \text{for some } C > 0,\ \gamma \in (0,1).$$

Case 2: $1 \le k \le [\sqrt{n}]$

Since $\ell_n=k$ implies $(X_0,\dots,X_{k-1}) = (X_k,\dots,X_{2k-1}) = \dots = (X_{(m-1)k},\dots,X_{mk-1})$, where $m = [\frac{n}{k}] \ge [\sqrt{n}]$, we let $\tau = [\frac{m}{3}]$ and derive for sufficiently large n that

$$P(l_n=k) \le P((X_0,\dots,X_{k-1}) = (X_k,\dots,X_{2k-1}) = \dots = (X_{(3\tau-1)k},\dots,X_{3\tau k-1}))$$

$$\le \sum_{x_0,\dots,_{k-1}} \{\mu_f^2(x_0,\dots,x_{\tau k-1}) + B\alpha^{\tau k}\} \quad \text{(by Lemma 3.2)}$$

$$\le \beta^{\tau k} + (r\alpha^{\sqrt{\tau}})^k \cdot B\alpha^{(\tau-\sqrt{\tau})k} \quad \text{(by Lemma 3.3)}$$

$$\le C\gamma^{[n^{1/4}]}, \quad \text{for some } C > 0,\ \gamma \in (0,1). \qquad \blacksquare$$

Lemma 3.5. *(CLT for additive functionals of one-dimensional Gibbs states)* *For every* $t \in \mathbb{R}$,

$$\mu_f\left[x \in \Sigma^+ : \frac{1}{\sqrt{n\beta_f''(0)}} \left[\sum_{j=0}^{n-1} \psi(\sigma^j x) - n \int \psi d\mu_f\right] \le t\right] \to \Phi(t) \quad \text{as } n \to \infty.$$

A proof is given in Lalley [7].

Proof of Theorem 3.1.

To prove (i), notice that

(a) $\frac{1}{n} \Sigma_{j=0}^{n-1} \psi(\sigma^j X) \to \theta$ a.s. as $n \to \infty$ by Birkhoff's Ergodic Theorem;

(b) $\frac{1}{l_n} \Sigma_{j=0}^{l_n-1} \psi(\sigma^j X) \to \theta$ a.s. as $n \to \infty$, by Lemma 3.4 (ii);

and

(c) $\frac{1}{l_n} \Sigma_{j=0}^{l_n-1} |\psi(\sigma^j X^{(l_n)}) - \psi(\sigma^j X)| \le \frac{1}{l_n} \Sigma_{j=0}^{l_n-1} \mathrm{var}_{l_n-j}\, \psi \to 0$ a.s.

as $n \to \infty$, since $\psi \in \mathcal{F}_\rho^+$.

(a), (b), (c) imply (i).

To prove (ii), consider the following decomposition:

$$\sqrt{n}[\theta(\hat{\mu}_n) - \theta] = H_1 + H_2 + H_3 + H_4 + H_5,$$

where
$$H_1 = \sqrt{n}\left[\frac{1}{n} \sum_{j=0}^{n-1} \psi(\sigma^j X) - \theta\right];$$

$$H_2 = \frac{1}{\sqrt{l_n}} \sum_{j=0}^{l_n-1} [\psi(\sigma^j X^{(l_n)}) - \psi(\sigma^j X)];$$

$$H_3 = \left(\frac{\sqrt{n}}{l_n} - \frac{1}{\sqrt{l_n}}\right) \sum_{j=0}^{l_n-1} \psi(\sigma^j X^{(l_n)});$$

$$H_4 = \left(\frac{1}{\sqrt{l_n}} - \frac{1}{\sqrt{n}}\right) \sum_{j=0}^{l_n-1} \psi(\sigma^j X);$$

$$H_5 = \frac{-1}{\sqrt{n}} \sum_{j=l_n}^{n-1} \psi(\sigma^j (X).$$

Let $\xrightarrow{d}$ and $\xrightarrow{P}$ denote the convergence in distribution and in probability respectively under μ_f. Then as $n \to \infty$

(a') $H_1 \xrightarrow{d} N(0,\beta''_f(0))$ by Lemma 3.5;

(b') $H_i \xrightarrow{P} 0$ a.s., $i = 2,3,4,5$, by Lemma 3.4.

(a'), (b') plus the Slutsky Theorem imply (ii). ■

4. Asymptotic Efficiency of the MLE

Theorem 3.1 (ii) suggests that $\lim_{n\to\infty} E_f[\sqrt{n}(\theta(\hat{\mu}_n) - \theta)]^2 = \beta''_f(0)$ for every real-valued $f \in \mathcal{F}^+_\rho$, where $E_f(\cdot) = E_{\mu_f}(\cdot)$. (See Lemma 4.10 for a stronger result.) If every statistic T_n based on the observations $X_0,\ldots,X_{n-1}$ satisfied

$$\liminf_{n\to\infty} E_f[\sqrt{n}(T_n-\theta)]^2 \geq \beta''_f(0)$$

for every real-valued $f \in \mathcal{F}^+_\rho$, then $\theta(\hat{\mu}_n)$ would be an asymptotically efficient estimator of θ. Unfortunately, this is not true in general because there exist some superefficient estimators (cf. [2]). Nevertheless in the vicinity of the point of superefficiency, there are some points in $\mathcal{F}^+_\rho$ where the superefficient estimators behave badly. From the minimax point of view the MLE is superior to other estimators.

The following definition is suggested by [2].

Definition 4.1. The estimator $\hat{T}_n$ is said to be *asymptotically efficient* at real-valued $f_0 \in \mathcal{F}^+_\rho$ if

(*) $\liminf_{\delta\to 0} \liminf_{n\to\infty} \sup_{f\in\overline{U}_{f_0}(\delta)} E_f[\sqrt{n}(\hat{T}_n-\theta)]^2 = \beta''_{f_0}(0)$;

and for any other statistic T_n based on $X_0,\ldots,X_{n-1}$

$$(**)\quad \liminf_{\delta\to 0}\ \liminf_{n\to\infty}\ \sup_{f\in\overline{U}_{f_0}(\delta)} E_f[\sqrt{n}(T_n-\theta)]^2 \geq \beta''_{f_0}(0),$$

where $\overline{U}_{f_0}(\delta) = \{f \in \mathcal{F}_\rho^+ : \|f-f_0\|_\rho \leq \delta\}$.

Notice that $\overline{U}_{f_0}(\delta)$ contains complex-valued f, for which λ_f, h_f, ν_f (hence μ_f) are well-defined when δ is sufficiently small. The justification follows from Lemma 4.5.

The main result of this section is

Theorem 4.2. *The MLE* $\theta(\hat{\mu}_n)$ *is asymptotically efficient at every real-valued* $f_0 \in \mathcal{F}_\rho^+$.

Since f is an unknown function, we have an infinite-dimensional estimation problem. Stein [11] points out that for estimating a single real-valued functional of the unknown state of nature it frequently happens that through each state of nature there is a one-dimensional problem which is, for large samples, at least as difficult as any other problems at that point. We call this one-dimensional problem the "least favorable" one.

We first consider the one-dimensional problem of estimating $\theta_z(g) = \int \psi d\mu_{f+zg}$.

Assume that f, g, $\psi \in \mathcal{F}_\rho^+$ are real-valued known functions and $z \in \mathbb{R}$ is an unknown parameter. Here $\theta_z(g)$ should be thought of as the quantity θ perturbed by the magnitude z along the direction specified by g.

Asymptotic efficiency in this parametric problem is closely related to the concept of "information".

Definition 4.3. We call

$$I_z^{(n)}(g) = E_{f+zg}[\frac{d}{dz} \log \mu_{f+zg}(X_0,\ldots,X_{n-1})]^2$$

the *Fisher information* contained in the sample $X_0,\dots,X_{n-1}$ associated with the one-parameter family $\{\mu_{f+zg} : z \in \mathbb{R}\}$, and

(4.1) $$I_z(g) = \lim_{n\to\infty} \frac{1}{n} I_z^{(n)}(g)$$

the *average information* associated with $\{\mu_{f+zg} : z \in \mathbb{R}\}$ when the limit exists.

Proposition 4.4. *For every real-valued* $f, g \in \mathcal{F}_\rho^+$, *the limit* (4.1) *exists and equals* $\frac{d}{dz} \int g d\mu_{f+zg}$.

To show Proposition 4.4 we need two lemmas.

Lemma 4.5. *(Perturbation theory for RPF operators)*

Let $f_1,\dots,f_k \in \mathcal{F}_\rho^+$ *be real-valued functions,*

$$f = (f_1,\dots,f_k)',\ z = (z_1,\dots,z_k)' \in \mathbb{C}^k;$$

let

$$\mathcal{L}_z = \mathcal{L}_{\langle z|f\rangle},\ \lambda_z = \lambda_{\langle z|f\rangle},\ h_z = h_{\langle z|f\rangle},$$

$$\nu_z = \nu_{\langle z|f\rangle},\ \mu_z = \mu_{\langle z|f\rangle}$$

for all z *at which these quantities exist, where* $\langle z|f\rangle = \Sigma_{\ell=1}^k z_\ell f_\ell \in \mathcal{F}_\rho^+$. *Observe that* $\mathcal{L}_z$ *is defined for all* $z \in \mathbb{C}^k$, *and* $\lambda_z, h_z, \nu_z, \mu_z$ *are well-defined for all* $z \in \mathbb{R}^k$.

(i) *The maps* $z \to \lambda_z$, $z \to h_z$ *have analytic extensions to a neighborhood* $\Omega = \Omega(f)$ *of* $\mathbb{R}^k$ *in* $\mathbb{C}^k$, *such that*

$$\mathcal{L}_z h_z = \lambda_z h_z,\quad z \in \Omega$$

and

$$\int h_z d\nu_0 = 1,\ z \in \Omega.$$

(ii) *The map* $z \to \nu_z$ *extends to a weak-* analytic* $M_\sigma(\Sigma^+)$*-valued function on* Ω

such that

$$\mathcal{L}^{*}_{z}\nu_{z} = \lambda_{z}\nu_{z}, \quad z \in \Omega$$

and

$$\int h_{z} d\nu_{z} = 1, \quad z \in \Omega.$$

Note. Weak-$*$ analytic means that for each $\phi \in \mathcal{F}^{+}_{\rho}$ the map $z \to \int \phi d\nu_{z}$ is analytic.

The lemma is stated as Proposition 4 in Lalley [8], Appendix 1.

Lemma 4.6. *Let* $g_{n}(z)$, $n \in \mathbb{N}^{+}$ *be analytic on* $\overline{U}_{2\delta} = \{z \in \mathbb{C} : |z| \leq 2\delta\}$, $\delta > 0$. *If* $|g_{n}(z)| \leq C$ *for some* $C > 0$ *and for all* $n \in \mathbb{N}^{+}$, $z \in \overline{U}_{2\delta}$, *then for some* $K > 0$

$$\left|\frac{dg_{n}(z)}{dz}\right| \leq K, \ \forall\, n \in \mathbb{N}^{+}, \ z \in \overline{U}_{\delta}.$$

The proof follows from the Cauchy integral representation of $g_{n}(z)$.

Proof of Proposition 4.4. For every $x \in \Sigma^{+}$

$$\begin{aligned}\mu_{f+zg}(x_{0},\ldots,x_{n-1}) &= \nu_{f+zg}(h_{f+zg}I_{(x_{0},\ldots,x_{n-1})}) \\ &= \lambda^{-n}_{f+zg} \cdot \nu_{f+zg}(\mathcal{L}^{n}_{f+zg}(h_{f+zg}\, I_{(x_{0},\ldots,x_{n-1})})),\end{aligned}$$

where the notation $\nu_{f+zg}(\psi)$ means $\int \psi d\nu_{f+zg}$ for $\psi \in C(\Sigma^{+})$.

Define $S_{0}f = 0$, $S_{n}f = \Sigma^{n-1}_{j=0} f \circ \sigma^{j}$, $n \in \mathbb{N}^{+}$. Note that

$$\mathcal{L}^{n}_{f+zg}(h_{f+zg}I_{(x_{0},\ldots,x_{n-1})})(y)$$

$$= \sum_{u:\sigma^{n}u=y} \exp(S_{n}(f+zg)(u)) \cdot h_{f+zg}(u) \cdot I_{(x_{0},\ldots,x_{n-1})}(u)$$

$$= \exp(S_{n}(f+zg)(x)) \cdot \exp\{S_{n}(f+zg)(\zeta) - S_{n}(f+zg)(x)\} \cdot h_{f+zg}(\zeta),$$

where $\zeta = (x_{0},\ldots,x_{n-1};y_{0},y_{1},\ldots) \in \Sigma^{+}$. Therefore,

$$\log \mu_{f+zg}(x_0,\dots,x_{n-1}) = S_n f(x) + zS_n g(x) - n \log \lambda_{f+zg} + \log Q_n(z),$$

where $Q_n(z) = \nu_{f+zg}(\exp\{S_n(f+zg)(\zeta) - S_n(f+zg)(x)\} \cdot h_{g+zg}(\zeta))$.

Consequently,

(4.2) $$\frac{d}{dz} \log \mu_{f+zg}(x_0,\dots,x_{n-1}) = S_n g(x) - n\, \beta'_{f,g}(z) + \frac{1}{Q_n(z)} \cdot \frac{d}{dz} Q_n(z),$$

where $\beta_{f,g}(z) = \log \dfrac{\lambda_{f+zg}}{\lambda_f}$. Recall the notation in Theorem 3.1: $\beta_f(z) = \beta_{f,\psi}(z)$.

Since $\beta'_{f,g}(z) = \int g d\mu_{f+zg}$ (cf. [8] p 161 (e)),

(4.3) $$\frac{1}{\sqrt{n}} (S_n g(x) - n\beta'_{f,g}(z)) \xrightarrow{d} N(0,\beta''_{f,g}(z)) \quad \text{as} \quad n \to \infty$$

under μ_{f+zg} by Lemma 3.5.

Since $Q_n(z) = \mu_{f+zg}(x_0,\dots,x_{n-1})/\exp\{-np(f+zg) + S_n(f+zg)(x)\}$, it follows from (1.1) and Bowen [1] that for every f and g there exist constants c_1 and c_2 such that

(4.4) $$0 < c_1 \le Q_n(z) \le c_2 < \infty, \quad \forall n \in \mathbb{N}^+, \; z \in \overline{U}_{2\delta}.$$

Lemma 4.6 implies that there exists $C > 0$ such that

(4.5) $$\left|\frac{d}{dz} Q_n(z)\right| \le C, \quad \forall n \in \mathbb{N}^+, \; z \in \overline{U}_{\delta}.$$

Therefore, (4.2)-(4.5) imply that under μ_{f+zg}

(4.6) $$\frac{1}{\sqrt{n}} \left[\frac{d}{dz} \log \mu_{f+zg}(X_0,\dots,X_{n-1})\right] \xrightarrow{d} N(0,\beta''_{f,g}(z)) \quad \text{as } n \to \infty.$$

Moreover, the moment convergence (4.1) follows from (4.2)-(4.5) and Theorem 1 of Lalley [7]. ■

The next thing is to determine the least favorable direction. When we estimate $\theta_z(g) = \int \psi d\mu_{f+zg}$ as a function of an unknown parameter

$z \in \mathbb{R}$, the asymptotic variance of unbiased estimators has Cramér-Rao lower bound

$$L_z(g) = [\frac{d}{dz}\, \theta_z(g)]^2 / I_z(g).$$

The following proposition shows that among all directions ψ itself represents the least favorable one.

Proposition 4.7. *For all real-valued* $g \in \mathcal{F}_\rho^+$

$$\lim_{z\to 0} L_z(g) \le \lim_{z\to 0} L_z(\psi).$$

Proof. By Proposition 4.4 $I_z(\psi) = \frac{d}{dz} \int \psi d\mu_{f+z\psi}$, so

$$L_z(\psi) = \frac{d}{dz} \int \psi d\mu_{f+z\psi}.$$

Let $B(z) = p(\langle z|G\rangle)$ with $z = (z_1, z_2, z_3)' \in \mathbb{C}^3$, $G = (f, \psi, g)'$, f, ψ, $g \in \mathcal{F}_\rho^+$; then

$$B_{z_2 z_2} \cdot B_{z_3 z_3} > B_{z_3 z_2} \cdot B_{z_2 z_3} \quad \text{(cf. [8] p 161 (c))},$$

where $B_{z_i z_j} = \frac{\partial^2 B(z)}{\partial z_i \partial z_j}$, $i, j = 2, 3$. By [8], p 161 (e), $B_{z_2} = \int \psi d\mu_{\langle z|G\rangle}$, $B_{z_3} = \int g d\mu_{\langle z|G\rangle}$. So

$$\frac{d}{dz_2} \int \psi d\mu_{\langle z|G\rangle} \cdot \frac{d}{dz_3} \int g d\mu_{\langle z|G\rangle} > \left[\frac{d}{dz_3} \int \psi d\mu_{\langle z|G\rangle}\right]^2,$$

where $\frac{d}{dz_3} \int g d\mu_{\langle z|G\rangle} > 0$ since $B(z)$ is strictly convex. Hence

$$\frac{d}{dz_2} \int \psi d\mu_{\langle z|G\rangle} > \left[\frac{d}{dz_3} \int \psi d\mu_{\langle z|G\rangle}\right]^2 \Big/ \frac{d}{dz_3} \int g d\mu_{\langle z|G\rangle}.$$

Let $z_1 \to 1$, $z_2 \to 0$, $z_3 \to 0$, then the analyticity of $B(z)$ implies that

(iii) $B_n^{(2)} \xrightarrow{P} 0$.

In fact, (i) follows from (4.1) and (4.6). To verify (ii), notice that

$$\frac{d^2}{dz^2} \log \mu_{f+zg}(X_0,\ldots,X_{n-1})$$

$$= -n\beta''_{f,g}(z) + \frac{1}{Q_n(z)} \cdot \frac{d^2}{dz^2} Q_n(z) - \left[\frac{1}{Q_n(z)} \cdot \frac{d}{dz} Q_n(z)\right]^2 .$$

By (4.4) and (4.5),

$$\left|\frac{1}{Q_n(z)} \frac{d}{dz} Q_n(z)\right| \le C_1, \quad \forall\, n \in \mathbb{N}^+,\ z \in \overline{U}_\delta \text{ and some } C_1 > 0;$$

and by (4.5) and Lemma 4.6,

$$\left|\frac{d^2}{dz^2} Q_n(z)\right| \le C_2, \quad \forall\, n \in \mathbb{N}^+,\ z \in \overline{U}_{\delta/2} \text{ and some } C_2 > 0.$$

Hence (ii) follows from (4.1).

Furthermore,

$$\frac{d^2}{dz^2} \log \mu_{f+zg} (X_0,\ldots,X_{n-1}) = -n\beta''_{f,g}(z) + R_n(z),$$

where

$$R_n(z) = \frac{1}{Q_n(z)} \frac{d^2}{dz^2} Q_n(z) - \left[\frac{1}{Q_n(z)} \cdot \frac{d}{dz} Q_n(z)\right]^2 .$$

So

$$\frac{d^3}{dz^3} \log \mu_{f+zg}(X_0,\ldots,X_{n-1}) = -n\, \beta'''_{f,g}(z) = \frac{d}{dz} R_n(z) .$$

For every $u \in \mathbb{R}$, $z \in \overline{U}_\delta$.

$$|\xi_{n,z}(u) - z| \to 0 \quad \text{as} \quad n \to \infty,$$

so

$$\beta'''_{f,g}(\xi_{n,z}(u)) \to \beta'''_{f,g}(z) \quad \text{as} \quad n \to \infty,$$

hence

$$\text{(4.7)} \qquad \frac{-n}{[I_z^{(n)}(g)]^{3/2}} \cdot \beta'''_{f,g}(\xi_{n,z}(u)) \to 0 \quad \text{as} \quad n \to \infty.$$

Since $|R_n(z)| \leq C \quad \forall\, n \in \mathbb{N}^+,\ z \in \overline{U}_{\delta/2}$ and some $C > 0$,

$$\left|\frac{d}{dz} R_n(z)\right| \leq C_1 \quad \forall\, n \in \mathbb{N}^+,\ z \in \overline{U}_{\delta/4} \quad \text{and} \quad \text{some } C_1 > 0.$$

To verify (iii), we still need to show that for every $u \in \mathbb{R}$, there exists $N \in \mathbb{N}^+$ such that

$$\text{(4.8)} \quad \xi_{n,z}(u) \in \overline{U}_{\delta/4} \quad \forall\, n > N,\ z \in \overline{U}_{\delta/8}.$$

Note that

$$|\xi_{n,z}(u)| \leq |z| + \frac{|u|}{\sqrt{I_z^{(n)}(g)}} \leq |z| + \frac{|u|}{\sqrt{n I_z(g)/2}}$$

$$\leq \frac{\delta}{8} + \frac{2|u|}{\sqrt{n \cdot m(g)}} \quad \forall\, n > N_1 \quad \text{and some } N_1 \in \mathbb{N}^+,$$

where $m(g) = \min_{m \in \overline{U}_{\delta/8}} I_z(g) > 0$ since $I_z(g) > 0$ for every $z \in \overline{U}_{\delta/8}$ and every $g \in \mathcal{F}_\rho^+$ which is not homologous to constant, and $I_z(g)$ is continuous for $z \in \overline{U}_{\delta/8}$.

Now (4.8) holds for all $n > N$ provided

$$\frac{2|u|}{\sqrt{N \cdot m(g)}} < \frac{\delta}{8}. \qquad \blacksquare$$

Lemma 4.12. *For every real-valued $f \in \mathcal{F}_\rho^+$ and any statistic T_n based on $X_0, \ldots, X_{n-1}$,*

Proposition A1. *For every real-valued* $f_0 \in \mathcal{F}_\rho^+$, *there exists* $\delta > 0$ *such that*

$$\text{(A1)}\quad \lim_{n\to\infty} \|\mathcal{L}_f^n h/\lambda_f^n - \nu_f(h)\cdot h_f\|_\infty = 0$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$ *and* $h \in \mathcal{F}_\rho^+$ *with* $\|h\|_\rho \le 1$.

Corollary A2. *Let* $Y_n(f) = \dfrac{S_n\psi - n\beta_f'(0)}{\sqrt{n}}$. *Then there exists* $\delta > 0$ *such that*

$$\text{(A2)}\quad \lim_{n\to\infty} E_f e^{zY_n(f)} = e^{z^2\beta_f''(0)/2}$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$ *and* $z \in \overline{U}_\delta$.

Proof. Taylor expansion implies that

$$E_f e^{zY_n(f)} = \exp\{z^2\,\beta_f''(0)/2 + z^3\beta_f^{(3)}(\xi_{n,z})/3!\sqrt{n}\}\cdot \int \frac{\mathcal{L}^n_{f+zn^{-1/2}\psi}\, h_f}{\lambda^n_{f+zn^{-1/2}\psi}}\cdot d\nu_f,$$

where $|\xi_{n,z}| \le |z|$. Since $f \to \lambda_f$ is real-analytic, and $\beta_f(z) = \log \dfrac{\lambda_{f+z\psi}}{\lambda_f}$,

$$z^3\,\beta_f^{(3)}(\xi_{n,z})/3!\sqrt{n} \to 0, \quad \text{as}\quad n\to\infty$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$, $|z| \le \delta$, when $\delta > 0$ is sufficiently small. Moreover, by (A1) and [3], Lemma A3 (ii), (iii),

$$\lim_{n\to\infty} \|\mathcal{L}^n_{f+zn^{-1/2}\psi}\, h_f/\lambda^n_{f+zn^{-1/2}\psi} - \left(\int h_f d\nu_{f+zn^{-1/2}\psi}\right) h_{f+zn^{-1/2}\psi}\|_\infty = 0$$

and

$$\lim_{n\to\infty} \|\left(\int h_f d\nu_{f+zn^{-1/2}\psi}\right)\cdot h_{f+zn^{-1/2}\psi} - h_f\|_\infty = 0$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$, $|z| \le \delta$. Therefore (A2) holds. ■

Corollary A3. *There exists $\delta > 0$ such that*

$$\lim_{n\to\infty} E_f Y_n^2(f) = \beta_f''(0)$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$.

Proof. Assume that (A2) holds uniformly for all $f \in \overline{U}_{f_0}(2\delta)$ and let

$$C = \{z \in \mathbb{C} : |z| = \delta\}.$$

Since $E_f e^{zY_n(f)}$ is analytic in $\overline{U}_\delta$,

$$\frac{d^2}{dz^2} E_f e^{zY_n(f)} = \frac{1}{\pi i} \oint_c \frac{E_f e^{\zeta Y_n(f)}}{(\zeta - z)^3} d\zeta .$$

Therefore,

$$E_f Y_n^2(f) = \frac{1}{\pi i} \oint_c \frac{E_f e^{\zeta Y_n(f)}}{\zeta^3} d\zeta .$$

By the Dominated Convergence Theorem,

$$\lim_{n\to\infty} E_f(Y_n^2(f)) = \beta_f''(0)$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$. ■

Now let $Z_n(f) = \sqrt{n}[\theta(\hat{\mu}_n) - \theta]$. If we can prove that

(A3) $$\lim_{n\to\infty} E_f[Z_n(f) - Y_n(f)]^2 = 0$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$, then Corollary A3 plus the Cauchy-Schwarz inequality implies that

(A4) $$\lim_{n\to\infty} E_f Z_n^2(f) = \beta_f''(0)$$

References

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

[2] Ibragimov, I.A. and Has'minskii, R.Z. (1981). *Statistical estimation-asymptotic theory*. Springer-Verlag, New York.

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

[4] Kato, T. (1976). *Perturbation theory for linear operators*, 2nd ed. Springer-Verlag, New York.

[5] Kedem, B. (1980). *Binary time series*. Lecture Notes in Pure and Applied Math. 52. Marcel Dekker, New York.

[6] Koshevnik, Y.A. and Levit, B.Y. (1976). On a non-parametric analogue of the information matrix. Theory of Probability and Its Applications 4, 738-753.

[7] 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.

[8] Lalley, S.P. (1987). Distribution of periodic orbits of symbolic and axiom A flows. Adv. in Applied Math. 8, 154-193.

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

[10] Ruelle, D. (1978). *Thermodynamic formalism*. Addison-Wesley, Reading, Massachusetts.

[11] Stein, C. (1956). Efficient non-parametric testing and estimation. Proc. Third Berkeley Sympos., Math. Stat. Prob. 1, 187-195.