ESTIMATING FUNCTIONALS OF
ONE-DIMENSIONAL GIBBS STATES*

by

Chuanshu Ji
Purdue University
and
Columbia University

Technical Report 87–33

Department of Statistics
Purdue University

* Research partially supported by NSF grant DMS–8702620

ESTIMATING FUNCTIONALS OF ONE-DIMENSIONAL GIBBS STATES*

by

Chuanshu Ji

Purdue University
and
Columbia University

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.

* Research partially supported by NSF grant DMS–8702620

§0. 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 [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.

Categorical times series also arise in social sciences, e.g., in studies of social mobility. In these situations one would not usually have long series of observations, but rather "panel" type data. Thus, inference about complicated dependence would be problematic. The usefulness of Gibbs states as models in these contexts may be somewhat more limited.

Further discussion of modeling binary time series may be found in Kedem [4].

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

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

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 from (0.1) that f determines the dependence in the stationary sequence $\{X_n\}$ which has the distribution μ_f.

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 (0.1) 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 (0.1) and the likelihood function for a stationary Gaussian sequence; however, for Gaussian measures the potential function is quadratic.

The purpose of this paper is to estimate some functionals of Gibbs states μ_f when f is unknown. It is possible that two different functions f and g induce the same Gibbs measure $\mu_f (= \mu_g)$ which gives rise to the stationary sequence $X = (X_0, X_1, \ldots)$. Therefore based on the observations $X_0, \ldots, X_{n-1}$, f is not identifiable; only μ_f is. That is why we do not estimate the function f itself.

Unfortunately, the exact form of the likelihood function $\mu_f(y : y_i = x_i, 0 \leq i \leq n-1)$, denoted by $\mu_f(x_0, \ldots, x_{n-1})$ in what follows, is not available. So there is no direct method of finding a maximum likelihood estimator (MLE). Nevertheless, we shall construct an MLE, and use (0.1) to study its asymptotic properties.

Section 1 summarizes some background information about Gibbs states.

In Section 2 the estimation problem is formulated. Using the mixing properties of the shift operator on the sample space we construct estimators of maximum likelihood type. This is done for the case of the sample space Σ^+ first, then extended to the situation with a more general sample space Σ_A^+.

In Section 3 we prove the strong consistency and the asymptotic normality of MLEs. The former follows Birkhoff's ergodic theorem and the latter is derived from the central limit theorem (CLT) for sums of stationary sequences.

Section 4 establishes the asymptotic efficiency of MLEs. The nature of our estimation problem is infinite-dimensional since the function f is unknown. We follow Stein's idea of simplifying the problem of estimating a functional of the unknown f by considering the "least favorable" one-dimensional parametric subfamily (cf. Stein [11]). The main tools are provided by the perturbation theory for Ruelle-Perron-Frobenius operators and the theory of locally asymptotic normality (LAN) due to Hájek and LeCam.

§1. Background: Gibbs States and Ruelle-Perron-Frobenius Theory

(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

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

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

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

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

and μ_f is the unique element in $M_\sigma(\Sigma_A^+)$ such that (1.1) 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. (1.1) is an extension of (0.1) 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. 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) 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 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_A^+ 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 $M_\sigma(\Sigma_A^+)$. If μ_n is a sequence in $M_\sigma(\Sigma_A^+)$, 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 $\nu \in M_\sigma(\Sigma_A^+)$, recall that $\nu(x_0, \ldots, x_{n-1}) = \nu(y \in \Sigma_A^+ : y_i = x_i,\ 0 \le i \le n-1)$, i.e. the probability of the cylinder set with the first n coordinates $x_0, \ldots, x_{n-1}$.

Definition 2.1. Given $x_0, \ldots, x_{n-1}$, if there exists $\hat{\mu}_n \in M_\sigma(\Sigma_A^+)$ such that $\hat{\mu}_n(x_0, \ldots, x_{n-1}) \ge \nu(x_0, \ldots, x_{n-1})$, for all $\nu \in M_\sigma(\Sigma_A^+)$, 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.

We construct an MLE of θ in the cases with sample space Σ^+ and Σ_A^+ separately.

§2.1. The case Σ^+

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 $\sigma^j x \neq x$ for all $j < m$. The set of all periodic sequences in Σ^+ is denoted by $\mathcal{C}$.

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

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

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

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 $l_3 = 2$, $x^{(l_3)} = (1,2,1,2,\ldots)$ and $\hat{\mu}_3(1,2,1,2,\ldots) = \hat{\mu}_3(2,1,2,1,\ldots) = \frac{1}{2}$.

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

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 $\nu_n \xrightarrow{w} \nu$ as $n \to \infty$.*

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

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

and

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

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

$$\begin{aligned} &\nu_n(x_o, \ldots, x_{m-1}) \\ &= \sum_{x_m, \ldots, x_{n-1}} \nu_n(x_0, \ldots, x_{m-1}, x_m, \ldots, x_{n-1}) \\ &= \frac{1}{n}\{\nu(x_0, \ldots, x_{m-1}) + \nu(x_1, \ldots, x_{m-1}, \underline{n-m}, x_0) + \ldots + \nu(\underline{1}, x_0, \ldots, x_{m-1}, \underline{n-m-1})\}, \end{aligned}$$

where the pattern $\underline{k}$ represents a string of length k with arbitrary components. Since $\nu \in M_\sigma(\Sigma^+)$, at most $m-1$ terms in { } differ from $\nu(x_0, \ldots, x_{m-1})$. Therefore,

$$|\nu_n(x_o, \ldots x_{m-1}) - \nu(x_0, \ldots, x_{m-1})| \leq \frac{m-1}{n} \to 0$$

as $n \to \infty$. □

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

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

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

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

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

(ii) C is a countable union of disjoint equivalence classes;

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

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

$$\begin{aligned}
&\nu_k(x_0, \ldots, x_{n-1}) \\
&= \sum_{i=1}^K \nu_k(E_i \cap B_n) \\
&= \sum_{i=1}^K \nu_k(y^{(i)}) \cdot \#\{j : 1 \leq j \leq l(y^{(i)}),\ \sigma^j(y^{(i)}) \in B_n\} \\
&= \sum_{i=1}^K \{\nu_k(E_i)/l(y^{(i)})\} \cdot \#\{j : 1 \leq j \leq l(y^{(i)}),\ \sigma^j(y^{(i)}) \in B_n\} \\
&\leq \max_{1 \leq i \leq K} [\#\{j : 1 \leq j \leq l(y^{(i)}), \sigma^j(y^{(i)}) \in B_n\}/l(y^{(i)})] \\
&\leq \#\{j : 1 \leq j \leq l_n,\ \sigma^j x^{(l_n)} \in B_n\}/l_n \\
&= \hat{\mu}_n(x_0, \ldots, x_{n-1})
\end{aligned}$$

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 one equivalence class E_m, hence $l_n = l(y^{(m)})$ and

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

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

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

§2.2. The case Σ_A^+

Example 2.7. (cf. Example 2.4) Let $(x_0, x_1, x_2) = (1, 2, 3)$, $n = 3$, $A_{31} = 0$. Recall that $l_3 = 3$, $x^{(l_3)} = (1, 2, 3, 1, 2, 3, \ldots)$ in Example 2.4. Now $x^{(l_3)} \notin \Sigma_A^+$ since the transition $3 \to 1$ is disallowed. Therefore, the definition of l_n needs to be revised for those $x_0, \ldots, x_{n-1}$ satisfying

(a) $x \in \mathcal{C} \cap B_n$, $l(x) > n - 1$

and

(b) $A_{x_{n-1}, x_0} = 0$, i.e., $(x_0, \ldots, x_{n-1}; x_0, \ldots, x_{n-1}; \ldots) \notin \Sigma_A^+$.

For any $x_0, \ldots, x_{n-1}$ define

$$l^* = \inf\{l \in \mathbb{N} : \text{ there exist } a_0, \ldots, a_{l+1} \text{ with } a_j \in \{1, \ldots, r\} \ \forall 0 \leq j \leq l+1,$$

$$\text{such that } A_{a_j a_{j+1}} = 1 \ \forall 0 \leq j \leq l \text{ and } a_0 = x_{n-1}, \ a_{l+1} = x_0\}.$$

Since A is irreducible, l^* is well-defined (cf. [1], Lemma 1.3.). Notice that the choice of $a_1, \ldots, a_{l^*}$ need not be unique and l^* depends on x_0, x_{n-1} in general.

Definition 2.8. For given $x_0, \ldots, x_{n-1}$ satisfying (a) and (b) define

$$x^{(l_n)} = (x_0, \ldots, x_{n-1}, \ a_1, \ldots, a_{l^*}; \ x_0, \ldots, x_{n-1}, \ a_1, \ldots, a_{l^*}; \ \ldots) \in \mathcal{C}$$

for some choice of $a_1, \ldots, a_{l^*}$ with the smallest period $l(x^{(l_n)}) = n + l^* \triangleq l_n$; and define $\hat{\mu}_n \in M_\sigma(\Sigma_A^+)$ by

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

The next lemma is the analogue of Lemma 2.5.

Lemma 2.9. *For every $\nu \in M_\sigma(\Sigma_A^+)$ there exist $\nu_n \in M_\sigma(\Sigma_A^+)$, $n \in \mathbb{N}^+$ such that supp $\nu_n \subset \mathcal{C}$ for all $n \in \mathbb{N}^+$ and $\nu_n \xrightarrow{w} \nu$ as $n \to \infty$.*

Proof. Let $L^* = \max_{x_0, x_{n-1}} l^*$ and $L_n = n + L^*$, then for every $x_0, \ldots, x_{n-1}$ and for a possible choice of $a_1, \ldots, a_{L^*}$ define

$$x(n) = (x_0, \ldots, x_{n-1},\ a_1, \ldots, a_{L^*};\ x_0, \ldots, x_{n-1},\ a_1, \ldots, a_{L^*}; \ldots) \in C$$

and ν_n by

$$\text{(a)} \qquad \nu_n(x(n)) = \frac{1}{L_n}\{\nu(x_0, \ldots, x_{n-1}, a_1, \ldots, a_{L^*}) + \nu(x_1, \ldots, x_{n-1}, a_1, \ldots, a_{L^*}, x_0) + \cdots + \nu(a_{L^*}, x_0, \ldots, x_{n-1}, a_1, \ldots, a_{L^*-1})\}$$

and

$$\text{(b)} \qquad \nu_n(\sigma^j x(n)) = \nu_n(x(n)), \quad j = 1, \ldots, L_n,$$

then $\nu_n \in M_\sigma(\Sigma_A^+)$, *supp* $\nu_n \subset C$, $n \in \mathbb{N}^+$. Furthermore, for any $m \in \mathbb{N}^+$ and $x_0, \ldots, x_{m-1}$, when $n > m$ the same argument as in the proof of Lemma 2.5 implies that

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

as $n \to \infty$. Hence $\nu_n \xrightarrow{w} \nu$. □

Theorem 2.10. *$\hat{\mu}_n$ is an MLS.*

The proof is the same as in Theorem 2.6.

§3. Consistency and Asymptotic Normality of MLE

The main result of this section is

Theorem 3.1.

(i) *Strong consistency of MLE: for the MLE $\theta(\hat{\mu}_n)$, we have $\theta(\hat{\mu}_n) \to \theta$* 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 $t \in \mathbb{R}$

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

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

To prove Theorem 3.1 we need several lemmas.

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

Let

$$\mu_f\big((x_0, \dots, x_{m-1}) \cap (x_{m+n}, \dots, x_{m+n+k-1})\big)$$

$$= \mu_f(y \in \Sigma_A^+ : y_j = x_j,\ j = 0, \dots, m-1, m+n, \dots, m+n+k-1).$$

Then there exist $B > 0$, $\alpha \in (0,1)$ and $N \in \mathbb{N}^+$ such that

$$\Big|\frac{\mu_f\big((x_0, \dots, x_{m-1}) \cap (x_{m+n}, \dots, x_{m+n+k-1})\big)}{\mu_f(x_0, \dots, x_{m-1}) \cdot \mu_f(x_{m+n}, \dots, x_{m+n+k-1})} - 1\Big| \le B\alpha^n$$

uniformly for all x_j, $j = 0, \dots, m-1, m+n, \dots, m+n+k-1$ and for all $k \in \mathbb{N}^+$, $n > N$.

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

Lemma 3.3. *For each Gibbs state μ_f, there exist $\beta \in (0,1)$ and $N \in \mathbb{N}^+$ such that*

$$\mu_f(x_0, \dots, x_{n-1}) \le \beta^n, \quad \forall x_0, \dots, x_{n-1} \text{ and } n > N.$$

Proof. (1.1) implies that

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

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

The lemma follows by setting $\beta = 1 - \varepsilon$. □

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

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

and

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

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

$$P(|\frac{n-l_n}{\sqrt{n}}| > \delta) = P(\frac{n-l_n}{\sqrt{n}} > \delta) + P(\frac{l_n - n}{\sqrt{n}} > \delta).$$

The second term identically equals zero for all sufficiently large n because

$$l_n \le L_n = n + L^* \quad \forall x_0, \ldots x_{n-1}.$$

On the other hand,

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

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

Case 1: $[\sqrt{n}] + 1 \le k \le [n - \delta\sqrt{n}]$

Since $(X_0, \ldots, X_{[n^{\frac{1}{4}}]-1}) = (X_k, \ldots, X_{k+[n^{\frac{1}{4}}]-1})$ for sufficiently large n,

$$\begin{aligned}
P(l_n = k) &\le P\left((X_0, \ldots, X_{[n^{\frac{1}{4}}]-1}) = (X_k, \ldots, X_{k+[n^{\frac{1}{4}}]-1})\right) \\
&\le \sum_{x_0, \ldots, x_{[n^{\frac{1}{4}}]-1}} \mu_f(y \in \Sigma_A^+ : y_j = y_{j+k} = x_j,\ j = 0, \ldots, [n^{\frac{1}{4}}] - 1) \\
&\le \sum_{x_0, \ldots, x_{[n^{\frac{1}{4}}]-1}} \left\{\mu_f^2(x_0, \ldots, x_{[n^{\frac{1}{4}}]-1}) + B\alpha^{k-[n^{\frac{1}{4}}]}\right\} \quad \text{(by Lemma 3.2)} \\
&\le \beta^{[n^{\frac{1}{4}}]} + r^{[n^{\frac{1}{4}}]} \cdot B\alpha^{k-[n^{\frac{1}{4}}]} \quad \text{(by Lemma 3.3)} \\
&\le C\gamma^{[n^{\frac{1}{4}}]}, \quad \text{for some } C > 0, \gamma \in (0,1).
\end{aligned}$$

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

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

$$P(l_n = k) \le P\left((X_0, \ldots, X_{k-1}) = (X_k, \ldots, X_{2k-1}) = \cdots = (X_{(3r-1)k}, \ldots, X_{3rk-1})\right)$$

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

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

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

Lemma 3.5. *(CLT for additive functionals of one-dimensional Gibbs states)*

For every $t \in \mathbb{R}$,

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

A proof is given in Lalley [6].

Proof of Theorem 3.1.

To prove (i), notice that

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

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

and

(c) $\frac{1}{l_n}\sum_{j=0}^{l_n-1}|\psi(\sigma^j X^{(l_n)}) - \psi(\sigma^j X)| \leq \frac{1}{l_n}\sum_{j=0}^{l_n-1} \mathrm{var}_{l_n-j}\psi \to 0$ 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′) $H_1 \xrightarrow{d} N\big(0, \beta_f''(0)\big)$ by Lemma 3.5;

(b′) $H_2 \xrightarrow{P} 0$ since $|H_2| \leq \frac{1}{\sqrt{l_n}} \sum_{j=0}^{\infty} \text{var}_j\, \psi \to 0$ a.s.;

(c′) $H_3 \xrightarrow{P} 0$ since $|H_3| \leq \|\psi\|_\infty \cdot |\frac{n-l_n}{2\sqrt{l_n}}| \to 0$ a.s., by Lemma 3.4;

(d′) $H_4 \xrightarrow{P} 0$ since $|H_4| \leq \|\psi\|_\infty \cdot |\frac{1}{\sqrt{l_n}} - \frac{1}{\sqrt{n}}| \cdot l_n \to 0$ a.s., by Lemma 3.4;

(e′) $H_5 \xrightarrow{P} 0$ since $|H_5| \leq \|\psi\|_\infty \cdot |\frac{n-l_n}{\sqrt{n}}| \to 0$ a.s., by Lemma 3.4.

(a′)–(e′) plus the Slutsky Theorem imply (ii). □

§4. Asymptotic Efficiency of the MLE

Theorem 3.1 (ii) suggests that $\lim_{n\to\infty} E_f\left[\sqrt{n}\big(\theta(\hat{\mu}_n) - \theta\big)\right]^2 = \beta_f''(0)$ for every real-valued $f \in \mathcal{F}_\rho^+$, where $E_f(\cdot) = E_{\mu_f}(\cdot)$. (See Lemma 4.12 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.

Example 4.1. (Superefficient estimator)

Let T_n be an arbitrary asymptotically unbiased and asymptotically normal estimator of $\theta = \theta(\mu_f)$, i.e.,

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N\big(0, \tau^2(f)\big) \text{ as } n \to \infty \text{ under } \mu_f.$$

Suppose $\tau^2(f_0) > 0$. Define

$$\tilde{T}_n = \begin{cases} T_n, & \text{if } |T_n - \theta(\mu_{f_0})| > n^{-\frac{1}{4}} \\ \theta(\mu_{f_0}), & \text{if } |T_n - \theta(\mu_{f_0})| \leq n^{-\frac{1}{4}}, \end{cases} \quad n \in \mathbb{N}^+.$$

It can be verified that

$$\sqrt{n}(\tilde{T}_n - \theta) \xrightarrow{d} N\big(0, \tilde{\tau}^2(f)\big) \text{ as } n \to \infty \text{ under } \mu_f,$$

where

$$\tilde{\tau}^2(f) = \begin{cases} \tau^2(f), & \text{at } f \neq f_0, \\ 0, & \text{at } f = f_0. \end{cases}$$

Therefore $\tilde{T}_n$ strictly improves T_n at $f = f_0$ in terms of asymptotic variance $\tau^2(f)$. Nevertheless it is also observed that usually in the vicinity of the point of superefficiency, there are some points in $\mathcal{F}_\rho^+$ where the superefficient estimator behaves badly. From the minimax point of view the MLE is superior to other estimators.

Definition 4.2. The estimator $\hat{T}_n$ is said to be *asymptotically efficient* at rea-valued $f_0 \in \mathcal{F}_\rho^+$ if

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

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

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

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 4.3. *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.

§4.1. 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 the parametric problem is closely related to the concept of "information".

Definition 4.4. We call

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

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

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

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

We will show that this limit does exist and equals $\frac{d}{dz}\int g d\mu_{f+zg}$.

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

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

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

let

$$\mathcal{L}_\mathbf{z} = \mathcal{L}_{<\mathbf{z}|f>}, \quad \lambda_\mathbf{z} = \lambda_{<\mathbf{z}|f>}, \quad h_\mathbf{z} = h_{<\mathbf{z}|f>},$$

$$\nu_\mathbf{z} = \nu_{<\mathbf{z}|f>}, \quad \mu_\mathbf{z} = \mu_{<\mathbf{z}|f>}$$

for all $\mathbf{z}$ at which these quantities exist, where $< \mathbf{z}|f >= \sum_{l=1}^k z_l f_l \in \mathcal{F}_\rho^+$. Observe that $\mathcal{L}_\mathbf{z}$ is defined for all $\mathbf{z} \in \mathbb{C}^k$, and $\lambda_\mathbf{z}, h_\mathbf{z}, \nu_\mathbf{z}, \mu_\mathbf{z}$ are well-defined for all $\mathbf{z} \in \mathbb{R}^k$.

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

$$\mathcal{L}_\mathbf{z} h_\mathbf{z} = \lambda_\mathbf{z} h_\mathbf{z}, \quad \mathbf{z} \in \Omega$$

and

$$\int h_\mathbf{z} d\nu_0 = 1, \quad \mathbf{z} \in \Omega.$$

(ii) *The map* $\mathbf{z} \to \nu_{\mathbf{z}}$ *extends to a weak-∗ analytic* $M_\sigma(\Sigma_A^+)$*-valued function on* Ω *such that*

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

and

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

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

This lemma is stated as Proposition 4 in Lalley [7], Appendix 1.

Lemma 4.6. *Let* $g_n(z)$, $n \in \mathbb{N}^+$ *be analytic on* $\overline{U}_{2\delta} = \{z \in \mathbb{C} : |z| \leq 2\delta\}$, $\delta > 0$; *let* $L = \{z \in \mathbb{C} : |z| = 2\delta\}$ *be the boundary of* $\overline{U}_{2\delta}$. *Assume* $|g_n(z)| \leq C$ *for some* $C > 0$ *and for all* $n \in \mathbb{N}^+$, $z \in \overline{U}_{2\delta}$. *Then for some* $K > 0$

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

Proof. $g_n(z)$ has the Cauchy integral representation

$$g_n(z) = \frac{1}{2\pi i} \oint_L \frac{g_n(\varsigma)}{\varsigma - z} d\varsigma,$$

so

$$\frac{dg_n(z)}{dz} = \frac{1}{2\pi i} \oint_L \frac{g_n(\varsigma)}{(\varsigma - z)^2} d\varsigma, \quad z \in U_{2\delta} = \{z \in \mathbb{C} : |z| < 2\delta\}.$$

Therefore,

$$\left| \frac{dg_n(z)}{dz} \right| \leq \frac{1}{2\pi} \cdot \frac{C}{\delta^2} \cdot 2\pi \cdot 2\delta = \frac{2C}{\delta} \triangleq K, \quad \forall\, n \in \mathbb{N}^+,\ z \in \overline{U}_\delta. \quad \square$$

Proposition 4.7. *For every real-valued* $f, g \in \mathcal{F}_\rho^+$, *the limit (4.1) exists and equals* $\frac{d}{dz} \int g d\mu_{f+zg}$.

Proof. For every $x \in \Sigma_A^+$

$$\begin{aligned}\mu_{f+zg}(x_0, \ldots, x_{n-1}) &= \nu_{f+zg}(h_{f+zg} I_{(x_0,\ldots,x_{n-1})}) \\ &= \lambda_{f+zg}^{-n} \cdot \nu_{f+zg}\left(\mathcal{L}_{f+zg}^n (h_{f+zg} I_{(x_0,\ldots,x_{n-1})})\right),\end{aligned}$$

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

Define $S_0 f = 0$, $S_n f = \sum_{j=0}^{n-1} f \circ \sigma^j$, $n \in \mathbb{N}^+$. Note that

$$\begin{aligned}
&\mathcal{L}^n_{f+zg}\left(h_{f+zg} I_{(x_0,\ldots,x_{n-1})}\right)(y) \\
&= \sum_{u:\sigma^n u = y} \exp\left(S_n(f+zg)(u)\right) \cdot h_{f+zg}(u) \cdot I_{(x_0,\ldots,x_{n-1})}(u) \\
&= \exp\left(S_n(f+zg)(x)\right) \cdot \exp\left\{S_n(f+zg)(\varsigma) - S_n(f+zg)(x)\right\} \cdot h_{f+zg}(\varsigma),
\end{aligned}$$

where $\varsigma = (x_0, \ldots, x_{n-1}; y_0, y_1, \ldots) \in \Sigma_A^+$. Therefore,

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

where $Q_n(z) = \nu_{f+zg}\big(\exp\left\{S_n(f+zg)(\varsigma) - S_n(f+zg)(x)\right\} \cdot h_{f+zg}(\varsigma)\big)$.

Consequently,

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

where $\beta_{f,g}(z) = \log \frac{\lambda_{f+zg}}{\lambda_f}$. 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. [7] p161 (e)),

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

under μ_{f+zg} by Lemma 3.5, and

$$0 < c_1 \le Q_n(z) \le c_2 < \infty \quad \forall\, n \in \mathbb{N}^+ \tag{4.4}$$

because $Q_n(z) = \mu_{f+zg}(x_0, \ldots, x_{n-1}) \;/\; \exp\{-np(f+zg) + S_n(f+zg)(x)\}$ (cf. (1.1)).

To prove Proposition 4.7 it suffices to show that there exists $C > 0$ such that

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

In fact, $\frac{d}{dz} Q_n(z) = U_n^{(1)}(z) + U_n^{(2)}(z) + U_n^{(3)}(z)$. where

$$U_n^{(1)}(z) = \int [S_n g(\varsigma) - S_n g(x)] \cdot \exp\{S_n(f+zg)(\varsigma) - S_n(f+zg)(x)\} \cdot h_{f+zg}(\varsigma) d\nu_{f+zg};$$

$$U_n^{(2)}(z) = \int \exp\{S_n(f+zg)(\varsigma) - S_n(f+zg)(x)\} \cdot \frac{d}{dz} h_{f+zg}(\varsigma) \cdot d\nu_{f+zg};$$

and

$$U_n^{(3)}(z) = [\frac{d}{d\tau} \int \exp\{S_n(f+zg)(\varsigma) - S_n(f+zg)(x)\} \cdot h_{f+zg}(\varsigma) d\nu_{f+\tau g}]_{\tau=z}.$$

Observe that for $z \in \overline{U}_{2\delta}$

(i) $|S_n g(\varsigma) - S_n g(x)| \le \sum_{j=0}^{n-1} \mathrm{var}_{n-j}\, g \le \sum_{j=1}^{\infty} \mathrm{var}_j\, g < \infty$;

(ii) $|S_n(f + zg)(\varsigma) - S_n(f + zg)(x)| \le \sum_{j=1}^{\infty} \mathrm{var}_j(f + zg) \le \|f + zg\|_\rho \cdot \sum_{j=1}^{\infty} \rho^j$
$\le (\|f\|_\rho + 2\delta\|g\|_\rho) \frac{1}{1-\rho} < \infty$;

(iii) $\|h_{f+zg}\|_\rho \le K_1$ for some $K_1 > 0$;

(iv) $\|\frac{d}{dz} h_{f+zg}\|_\rho \le K_2$ for some $K_2 > 0$, by Lemma (4.5) (i).

Therefore, (i), (ii), (iii) imply that

$$|U_n^{(1)}(z)| \le C_1 \quad \forall\, n \in \mathbb{N}^+ \text{ and some } C_1 > 0, \text{ and}$$

(ii), (iv) imply that

$$|U_n^{(2)}(z)| \le C_2 \quad \forall\, n \in \mathbb{N}^+ \text{ and some } C_2 > 0.$$

Let $g_n(\tau) = \int \exp\{S_n(f + zg)(\varsigma) - S_n(f + zg)(x)\} \cdot h_{f+zg}(\varsigma) \cdot d\nu_{f+\tau g}$. By Lemma (4.5) (ii), $g_n(\tau)$ is analytic for $\tau \in \overline{U}_{2\delta}$ given z. Since (ii), (iii) imply that

$$|g_n(\tau)| \le K_3 \quad \forall\, n \in \mathbb{N}^+ \text{ and some } K_3 > 0,$$

by Lemma 4.6

$$|U_n^{(3)}(z)| = \left| [\frac{d}{d\tau} g_n(\tau)]_{\tau=z} \right| \le C_3 \quad \forall\, n \in \mathbb{N}^+,\ z \in \overline{U}_\delta \text{ and some } C_3 > 0.$$

Hence $|\frac{d}{dz} Q_n(z)| \le C \ \ \forall\, n \in \mathbb{N}^+$, $z \in \overline{U}_\delta$ and some $C > 0$.

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

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

Moreover, the moment convergence (4.1) follows from (4.2)–(4.5) and Theorem 1 of Lalley [6]. □

§4.2. 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 4.8. *For all* $g \in \mathcal{F}_\rho^+$

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

Proof. By Proposition 4.7 $I_z(\psi) = \frac{d}{dz}\int \psi d\mu_{f+z\psi}$, so

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

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

$$B_{z_2 z_2} \cdot B_{z_3 z_3} > B_{z_3 z_2} \cdot B_{z_2 z_3} \quad \big(\text{ cf. [7] p161 (c)}\big),$$

where $B_{z_i z_j} = \frac{\partial^2 B(z)}{\partial z_i \partial z_j}$, $i, j = 2, 3$. By [7], p161 (e), $B_{z_2} = \int \psi d\mu_{<\mathbf{z}|G>}$, $B_{z_3} = \int g d\mu_{<\mathbf{z}|G>}$. So

$$\frac{d}{dz_2}\int \psi d\mu_{<\mathbf{z}|G>} \cdot \frac{d}{d_{z_3}}\int g d\mu_{<\mathbf{z}|G>} > \left(\frac{d}{dz_3}\int \psi d\mu_{<\mathbf{z}|G>}\right)^2,$$

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

$$\frac{d}{dz_2}\int \psi d\mu_{<\mathbf{z}|G>} > \left(\frac{d}{dz_3}\int \psi d\mu_{<\mathbf{z}|G>}\right)^2 / \frac{d}{dz_3}\int g d\mu_{<\mathbf{z}|G>}.$$

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

$$\lim_{z_2\to 0}\frac{d}{dz_2}\int \psi d\mu_{f+z_2\psi} \geq \lim_{z_3\to 0}\left\{\left(\frac{d}{dz_3}\int \psi d\mu_{f+z_3 g}\right)^2 / \frac{d}{dz_3}\int g d\mu_{f+z_3 g}\right\},$$

i.e.,

$$\lim_{z\to 0} L_z(\psi) \geq \lim_{z\to 0} L_z(g). \quad \square$$

§4.3. Local asymptotic normality (LAN) of the one-parameter family $\{\mu_{f+zg} : z \in \mathbb{R}\}$

Hajek, LeCam and others proved that many important properties of point estimators follow from the asymptotic normality of the logarithm of the likelihood ratio for parameters close to each other, regardless of the relation between the observations which produce the given likelihood function (cf. [2]). The family $\{\mu_{f+zg} : z \in \mathbb{R}\}$ enjoys this property.

Proposition 4.9. *For every $u \in \mathbb{R}$, let*

$$Z_{n,z}(u) = \log \frac{\mu_{f+(z+u/\sqrt{I_z^{(n)}(g)})g}(X_0, \dots, X_{n-1})}{\mu_{f+zg}(X_0, \dots, X_{n-1})}.$$

Then

$$Z_{n,z}(u) = u \cdot \triangle_{n,z} - \frac{u^2}{2} + B_{n,z}(u),$$

where

$$\triangle_{n,z} \xrightarrow{d} N(0,1)$$

and

$$B_{n,z}(u) \xrightarrow{P} 0$$

under μ_{f+zg} as $n \to \infty$.

Proof. By Taylor expansion

$$Z_{n,z}(u) = u \cdot \triangle_{n,z} + \frac{u^2}{2} \cdot B_n^{(1)} + \frac{u^3}{3!} B_n^{(2)}.$$

where

$$\triangle_{n,z} = \frac{1}{\sqrt{I_z^{(n)}(g)}} \cdot \frac{d}{dz} \log \mu_{f+zg}(X_0, \dots, X_{n-1}),$$

$$B_n^{(1)} = \frac{1}{I_z^{(n)}(g)} \cdot \frac{d^2}{dz^2} \log \mu_{f+zg}(X_0, \dots, X_{n-1});$$

$$B_n^{(2)} = \frac{1}{[I_z^{(n)}(g)]^{3/2}} \cdot \left[\frac{d^3}{dz^3} \log \mu_{f+zg}(X_0, \dots, X_{n-1})\right]_{z=\xi_{n,z}(u)}$$

with $|\xi_{n,z}(u) - z| \le \frac{|u|}{\sqrt{I_z^{(n)}(g)}}$.

We claim that under μ_{f+zg} as $n \to \infty$

(i) $\triangle_{n,z} \xrightarrow{d} N(0,1)$;

(ii) $B_n^{(1)} \xrightarrow{P} -1$;

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

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

$$\begin{aligned}&\frac{d^2}{dz^2}\log \mu_{f+zg}(X_0,\dots,X_{n-1})\\ &\quad= -n\beta''_{f,g}(z) + \frac{1}{Q_n(z)}\cdot\frac{d^2}{dz^2}\,Q_n(z) - \left[\frac{1}{Q_n(z)}\cdot\frac{d}{dz}\,Q_n(z)\right]^2.\end{aligned}$$

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.

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

Hence (ii) follows from (4.1).

Furthermore,

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

where

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

So

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

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

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

so

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

hence

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

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

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

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

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

Note that

$$\begin{aligned} |\xi_{n,z}(u)| &\le |z| + \frac{|u|}{\sqrt{I_z^{(n)}(g)}} \le |z| + \frac{|u|}{\sqrt{nI_z(g)/2}} \\ &\le \frac{\delta}{8} + \frac{2|u|}{\sqrt{n \cdot m(g)}} \quad \forall\, n > N_1 \text{ and some } N_1 \in \mathbb{N}^+, \end{aligned}$$

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

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

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

§4.4. The minimax lower bound on asymptotic variance of estimators

Lemma 4.10. *$\beta''_f(0) \to \beta''_{f_0}(0)$ as $\|f - f_0\|_\rho \to 0$.*

Proof. Since $\beta_f(z) = p(f + z\psi) - p(f)$ and the map $f \mapsto p(f)$ is real analytic on $\mathcal{F}_\rho^+$ (cf. [10], Corollary 5.27), Lemma 4.10 follows. $\square$

Corollary 4.11. *Let $\overline{U}_{f_0}(\delta) = \{f \in \mathcal{F}_\rho^+ : \|f - f_0\|_\rho \le \delta\}$, then*

$$\liminf_{\delta \to 0} \sup_{f \in \overline{U}_{f_0}(\delta)} \beta''_f(0) = \beta''_{f_0}(0).$$

The next lemma is the key to proving the asymptotic efficiency of the MLE $\theta(\hat{\mu}_n)$. The proof will be given in the Appendix.

Lemma 4.12. *For every $f_0 \in \mathcal{F}_\rho^+$ and $\delta > 0$,*

$$\lim_{n\to\infty} E_f[\sqrt{n}(\theta(\hat{\mu}_n) - \theta)]^2 = \beta_f''(0)$$

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

Lemma 4.13. *For every $f \in \mathcal{F}_\rho^+$ and any statistic T_n based on $X_0, \ldots, X_{n-1}$,*

$$\lim_{b\to\infty} \liminf_{n\to\infty} \sup_{|z|<\frac{b}{\sqrt{n}}} E_{f+z\psi} \left[\sqrt{n}\left(T_n - \theta - z\beta_f''(0)\right)\right]^2 \geq \beta_f''(0).$$

The proof may be found in [2] (cf. [2] Theorem 12.1 and Remark 12.2).

Proof of Theorem 4.3.

To verify (*), note that for any $\varepsilon > 0$, $\delta > 0$ and $f_0 \in \mathcal{F}_\rho^+$ by Lemma 4.12 there exists $N \in \mathbb{N}^+$ such that

$$E_f\left[\sqrt{n}\left(\theta(\hat{\mu}_n) - \theta\right)\right]^2 \leq \beta_f''(0) + \varepsilon$$

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

$$\liminf_{n\to\infty} \sup_{f\in\overline{U}_{f_0}(\delta)} E_f\left[\sqrt{n}(\theta(\hat{\mu}_n) - \theta)\right]^2 \leq \sup_{f\in\overline{U}_{f_0}(\delta)} \beta_f''(0).$$

The inequality "$\geq$" can be derived similarly. Therefore,

$$\liminf_{n\to\infty} \sup_{f\in\overline{U}_{f_0}(\delta)} E_f\left[\sqrt{n}\left(\theta(\hat{\mu}_n) - \theta\right)\right]^2 = \sup_{f\in\overline{U}_{f_0}(\delta)} \beta_f''(0).$$

Let $\delta \to 0$, then by Lemma 4.10 $\theta(\hat{\mu}_n)$ satisfies (*).

To verify (**), we use an argument similar to that in [5]. For any $b > 0$, $\delta > 0$ and any statistic T_n based on $X_0, \ldots, X_{n-1}$, when n is sufficiently large we obtain that

$$\begin{aligned}
\sup_{f\in\overline{U}_{f_0}(\delta)} E_f\left[\sqrt{n}(T_n - \theta)\right]^2 &\geq \sup_{z:|z|<\frac{b}{\sqrt{n}}} E_{f_0+z\psi}\left[\sqrt{n}(T_n - \theta(\mu_{f_0+z\psi}))\right]^2 \\
&= \sup_{|z|<\frac{b}{\sqrt{n}}} E_{f_0+z\psi}\left[\sqrt{n}\left(T_n - \theta(\mu_{f_0}) - z\beta_{f_0}''(0) + o(z)\right)\right]^2 \\
&= \sup_{|z|<\frac{b}{\sqrt{n}}} E_{f_0+z\psi}\left[\sqrt{n}\left(T_n - \theta(\mu_{f_0}) - z\beta_{f_0}''(0)\right) + o(1)\right]^2 \\
&= \sup_{|z|<\frac{b}{\sqrt{n}}} E_{f_0+z\psi}\left[\sqrt{n}\left(T_n - \theta(\mu_{f_0}) - z\beta_{f_0}''(0)\right)\right]^2 + o(1).
\end{aligned}$$

Hence

$$\liminf_{n\to\infty} \sup_{f\in\overline{U}_{f_0}(\delta)} E_f[\sqrt{n}(T_n-\theta)]^2$$

$$\geq \liminf_{n\to\infty} \sup_{|z|<\frac{b}{\sqrt{n}}} E_{f_0+z\psi}\left[\sqrt{n}\left(T_n-\theta(\mu_{f_0})-z\beta''_{f_0}(0)\right)\right]^2.$$

Since $b>0$ is arbitrary, by Lemma 4.13.

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

Now (**) follows by letting $\delta\to 0$. □

§5. Appendix: Proof of Lemma 4.12

Recall Theorem 1.1: *For each real-valued $f\in\mathcal{F}_\rho^+$, associated with $\mathcal{L}_f:\mathcal{F}_\rho^+\to\mathcal{F}_\rho^+$ there exist $\lambda_f\in(0,\infty)$, $h_f\in\mathcal{F}_\rho^+$ and $\nu_f\in(\mathcal{F}_\rho^+)^*$ such that*

(A1) $\mathcal{L}_f h_f=\lambda_f h_f$, λ_f *is a simple eigenvalue of* $\mathcal{L}_f$;

(A2) $\mathcal{L}_f^*\nu_f=\lambda_f\nu_f$;

(A3) Spectrum $\mathcal{L}_f\backslash\{\lambda_f\}\subset\{z\in\mathbb{C}:|z|<\lambda_f-\varepsilon\}$ *for some* $\varepsilon>0$;

(A4) $\lim_{n\to\infty}\|\mathcal{L}_f^n g/\lambda_f^n-\nu_f(g)\cdot h_f\|_\infty=0$, $\forall g\in\mathcal{F}_\rho^+$.

Note that (A1) implies that if $h\in\mathcal{F}_\rho^+$, $\mathcal{L}_f h=\lambda_f h$, then $h=ch_f$ for some constant c.

The next lemma shows that the measure satisfying (A2) is essentially unique as well.

Lemma A1. *If $\nu\in(\mathcal{F}_\rho^+)^*$, $\mathcal{L}_f^*\nu=\lambda_f\nu$, then $\nu=c\nu_f$ for some constant c.*

Proof. By (A4),

$$\lim_{n\to\infty}\nu\big(\mathcal{L}_f^n(g)/\lambda_f^n\big)=\nu\big(\nu_f(g)\cdot h_f\big),\quad \forall\, g\in\mathcal{F}_\rho^+.$$

Since $\nu\big(\mathcal{L}_f^n(g)/\lambda_f^n\big)=\nu(g)$, $\forall\, n\in\mathbb{N}^+$, we have

$$\nu(g)=\nu(h_f)\cdot\nu_f(g),\quad \forall\, g\in\mathcal{F}_\rho^+.$$

Therefore, $\nu=c\nu_f$ with $c=\nu(h_f)$. □

For each real-valued $f\in\mathcal{F}_\rho^+$, define

$$\mathcal{L}'_f\in(\mathcal{F}_\rho^+)^* \text{ by } \mathcal{L}'_f h=\lambda_f\cdot\nu_f(h)\cdot h_f$$

and

$$\mathcal{L}''_f \in (\mathcal{F}_\rho^+)^* \text{ by } \mathcal{L}''_f h = \mathcal{L}_f h - \mathcal{L}'_f h \quad \text{for } h \in \mathcal{F}_\rho^+;$$

and let

$$F'_f = \{z h_f : z \in \mathbb{C}\},$$

$$F''_f = \{h \in \mathcal{F}_\rho^+ : \nu_f(h) = 0\}.$$

Lemma A2.

(i) $\mathcal{F}_\rho^+ = F'_f \oplus F''_f$;

(ii) $\mathcal{L}_f = \mathcal{L}'_f + \mathcal{L}''_f$, where $\mathcal{L}'_f : F'_f \to F'_f$, $\mathcal{L}''_f : F''_f \to F''_f$,

$$\mathcal{L}'_f : F''_f \to 0, \quad \mathcal{L}''_f : F'_f \to 0;$$

and

(iii)

$$\textit{Spectrum } \mathcal{L}'_f = \{\lambda_f\},$$

$$\textit{Spectrum } \mathcal{L}''_f = \textit{Spectrum } \mathcal{L}_f \backslash \{\lambda_f\}.$$

The proofs of (i), (ii), are straightforward. For the proof of (iii), see Kato [3], III, §6.4 and §6.5.

Now let

$$H = \{h \in \mathcal{F}_\rho^+ : \nu_{f_0}(h) = 1\},$$

$$H' = \{\nu \in (\mathcal{F}_\rho^+)^* : \nu(h_{f_0}) = 1\}.$$

Observe that H and H' are translates of the Banach spaces $\{h \in \mathcal{F}_\rho^+ : \nu_{f_0}(h) = 0\}$ and $\{\nu \in (\mathcal{F}_\rho^+)^* : \nu(h_{f_0}) = 0\}$ respectively.

Lemma A3. *There exists $\delta > 0$ such that the maps*

(i) $\mathcal{F}_\rho^+ \to \mathbb{C} : f \to \lambda_f$

(ii) $\mathcal{F}_\rho^+ \to H : f \to h_f$

(iii) $\mathcal{F}_\rho^+ \to H' : f \to \nu_f$

are all C^1 (i.e. have continuous Frechet derivatives) in $U_{f_0}(\delta) = \{f \in \mathcal{F}_\rho^+ : \|f - f_0\|_\rho < \delta\}$, and (A1), (A2) hold for all $f \in U_{f_0}(\delta)$.

Proof. We first check (i) and (ii). It is verified that the map

$$\mathcal{F}_\rho^+ \times \mathbb{C} \times H \to \mathcal{F}_\rho^+ : (f, \lambda, h) \to \mathcal{L}_f h - \lambda h$$

is C^1. For given $f_0 \in \mathcal{F}_\rho^+$, the map

$$\pi : (\lambda, h) \to \mathcal{L}_{f_0} h - \lambda h$$

has its Frechet derivative at (λ_{f_0}, h_{f_0}), say $D\pi$, given by

$$D\pi(\lambda, h) = (\mathcal{L}_{f_0} - \lambda_{f_0} I)h - \lambda h_{f_0},$$

where I is the identity map on $\mathcal{F}_\rho^+$. Obviously $D\pi$ is a continuous linear map. We claim that $D\pi$ is also $1-1$ and onto.

In fact, if $(\lambda, h), (\lambda', h') \in \mathbb{C} \times H$ satisfy

$$(\mathcal{L}_{f_0} - \lambda_{f_0} I)h - \lambda h_{f_0} = (\mathcal{L}_{f_0} - \lambda_{f_0} I)h' - \lambda' h_{f_0},$$

then by letting ν_{f_0} act on both sides we obtain $\lambda = \lambda'$. Moreover, $h = h'$ also follows from the standard argument.

To check $D\pi$ is onto, for every $g \in \mathcal{F}_\rho^+$ consider the equation

$$(\mathcal{L}_{f_0} - \lambda_{f_0} I)h - \lambda h_{f_0} = g \tag{A5}$$

with the variable $(\lambda, h) \in \mathbb{C} \times H$. By Lemma A2, $\lambda_{f_0} \notin \mathit{Spectrum}\ \mathcal{L}''_{f_0}$, so $\mathcal{L}''_{f_0} - \lambda_{f_0} I$ is invertible. There exist unique $g' \in F'_{f_0}$, $g'' \in F''_{f_0}$ such that $g = g' + g''$. Define (λ, h) by

$$\lambda = -\nu_{f_0}(g),$$

$$h = h' + h'' \quad \text{with} \quad h' = h_{f_0}$$

$$h'' = (\mathcal{L}''_{f_0} - \lambda_{f_0} I)^{-1} g''.$$

We claim that (λ, h) satisfies (A5) and $\nu_{f_0}(h) = 1$. Actually, by Lemma A2,

$$\begin{aligned}(\mathcal{L}_{f_0} - \lambda_{f_0} I)h - \lambda h_{f_0} &= \mathcal{L}'_{f_0} h' - \lambda_{f_0} h' + \nu_{f_0}(g) \cdot h_{f_0} + (\mathcal{L}''_{f_0} - \lambda_{f_0} I)h'' \\ &= \lambda_{f_0} \cdot \nu_{f_0}(h_{f_0}) \cdot h_{f_0} - \lambda_{f_0} h_{f_0} + g' + g'' \\ &= g' + g'' \\ &= g.\end{aligned}$$

Thus $D\pi$ is onto.

Now the Implicit Mapping Theorem (cf. [8], p17) implies that there exist $\delta > 0$ and a unique map $\tau : U_{f_0}(\delta) \to \mathbb{C} \times H$, $\tau(f) = \big(\lambda(f),\, h(f)\big)$, $f \in U_{f_0}(\delta)$, such that τ is a C^1-map, and

$$\tau(f_0) = (\lambda_{f_0}, h_{f_0}),$$

$$\mathcal{L}_f h(f) = \lambda(f) \cdot h(f), \quad \forall f \in U_{f_0}(\delta). \tag{A6}$$

It follows from the uniqueness of τ and the fact that the map $f \to \lambda_f$ is a real-analytic that

$$\lambda(f) = \lambda_f, \quad \forall f \in U_{f_0}(\delta).$$

Since λ_f is a simple eigenvalue of $\mathcal{L}_f : \mathcal{F}_\rho^+ \to \mathcal{F}_\rho^+$,

$$h(f) = h_f.$$

So (A6) and (A1) coincide. Therefore, (i) and (ii) hold.

A similar argument implies that there exist $\delta > 0$ and a unique map $\tau' : U_{f_0}(\delta) \to \mathbb{C} \times H'$, $\tau'(f) = \big(\lambda'(f), \nu(f)\big)$, $f \in U_{f_0}(\delta)$, such that τ' is a C^1-map, and

$$\tau'(f_0) = (\lambda_{f_0}, \nu_{f_0}),$$

$$\mathcal{L}_f^*\big(\nu(f)\big) = \lambda'(f) \cdot \nu(f), \quad \forall\, f \in U_{f_0}(\delta). \tag{A7}$$

By the uniqueness of such a map,

$$\lambda'(f) = \lambda_f.$$

Furthermore, by Lemma A1,

$$\nu(f) = \nu_f.$$

So (A7) and (A2) coincide, hence (iii) holds. □

Note. $f \to \nu_f$ is a continuous map from $\mathcal{F}_\rho^+$ into $(\mathcal{F}_\rho^+)^*$, but not from $\mathcal{F}_\rho^+$ into $\big(C(\Sigma_A^+)\big)^*$ under the total variation norm.

Lemma A4. *For every real-valued $f_0 \in \mathcal{F}_\rho^+$, there exist $\delta > 0$ and $\varepsilon \in (0, \frac{\lambda_{f_0}}{3})$, such that for all $f \in U_{f_0}(\delta)$*

(i)
$$\lambda_f > \lambda_{f_0} - \varepsilon;$$

and

(ii)
$$\textit{Spectrum } \mathcal{L}_f'' \subset \{z \in \mathbb{C} : |z| < \lambda_{f_0} - 3\varepsilon\}.$$

Proof. (i) simply follows from Lemma A3 (i).

To show (ii), first notice that by Theorem 1.1 and Lemma A2 (iii), there exists $\varepsilon \in (0, \frac{\lambda_{f_0}}{3})$ such that

$$\textit{Spectrum } \mathcal{L}_{f_0}'' \subset \{z \in \mathbb{C} : |z| < \lambda_{f_0} - 3\varepsilon\}.$$

Thus $(zI - \mathcal{L}_{f_0}'')^{-1}$ exists and is continuous for all $z \in \mathbb{C}$ satisfying $|z| > \lambda_{f_0} - 3\varepsilon$. Since $\|(zI - \mathcal{L}_{f_0}'')^{-1}\| \to 0$ as $|z| \to \infty$,

$$0 < \alpha \triangleq \sup_{|z| \geq \lambda_{f_0} - 3\varepsilon} \|(zI - \mathcal{L}_{f_0}'')^{-1}\| < \infty.$$

There exists $\delta > 0$ such that

$$\|\mathcal{L}_f'' - \mathcal{L}_{f_0}''\| < \frac{1}{2\alpha}, \quad \forall\, f \in U_{f_0}(\delta).$$

So $(zI - \mathcal{L}_f'')^{-1}$ exists for all $f \in U_{f_0}(\delta)$, $z \in \mathbb{C}$ satisfying $|z| > \lambda_{f_0} - 3\varepsilon$, and

$$(zI - \mathcal{L}_f'')^{-1} = (zI - \mathcal{L}_{f_0}'')^{-1}[I + (zI - \mathcal{L}_{f_0}'')^{-1}(\mathcal{L}_{f_0}'' - \mathcal{L}_f'')]^{-1}.$$

Since $\|(zI - \mathcal{L}_{f_0}'')^{-1}(\mathcal{L}_{f_0}'' - \mathcal{L}_f'')\| < \frac{1}{2}$, this proves (ii). □

Proposition A5. *There exists $\delta > 0$ such that*

(*A8*)
$$\lim_{n\to\infty} \|\mathcal{L}_f^n h / \lambda_f^n - \nu_f(h) \cdot h_f\|_\infty = 0$$

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

Proof. By the Spectral Radius Formula (cf. Kato [3], III, §3, §6) and Lemma A4 (ii),

$$\|(\mathcal{L}''_{f_0})^k\| \le (\lambda_{f_0} - \frac{5}{2}\varepsilon)^k, \text{ for some } k \in \mathbb{N}^+.$$

Since $f \to \mathcal{L}''_f$ is continuous, there exists $\delta > 0$ such that

$$\|(\mathcal{L}''_f)^k\| \le (\lambda_{f_0} - 2\varepsilon)^k, \quad \forall\, f \in \overline{U}_{f_0}(\delta).$$

By Lemma A4 (i),

$$\|(\mathcal{L}''_f)^k / \lambda_f^k\| \le \left(\frac{\lambda_{f_0} - 2\varepsilon}{\lambda_{f_0} - \varepsilon}\right)^k < 1.$$

So we have

$$\lim_{m\to\infty} \|(\mathcal{L}''_f)^{km} / \lambda_f^{km}\| = 0 \tag{A9}$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$. Since Lemma A2 implies

$$\mathcal{L}_f^n = (\mathcal{L}'_f)^n + (\mathcal{L}''_f)^n,$$

we obtain

$$\frac{\mathcal{L}_f^n h}{\lambda_f^n} = \frac{(\mathcal{L}''_f)^n h}{\lambda_f^n} + \nu_f(h) \cdot h_f, \quad \forall\, n \in \mathbb{N},\ f \in \overline{U}_{f_0}(\delta),\ h \in \mathcal{F}_\rho^+.$$

Hence (A8) follows from (A9). □

Corollary A6. *Let* $Y_n(f) = \frac{S_n\psi - n\beta'_f(0)}{\sqrt{n}}$. *Then there exists* $\delta > 0$ *such that*

$$\lim_{n\to\infty} E_f e^{zY_n(f)} = e^{z^2\beta''_f(0)/2} \tag{A10}$$

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

Proof. Taylor expansion implies that

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

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

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

uniformly for all $f \in \overline{U}_{f_0}(\delta)$, $|z| \leq \delta$, when $\delta > 0$ is sufficiently small. Moreover, by (A8) and Lemma A3 (ii), (iii),

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

and

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

uniformly for all $f \in \overline{U}_{f_0}(\delta)$, $|z| \leq \delta$. Therefore (A10) holds. □

Corollary A7. *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 (A10) 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 $\overline{U}_\delta$,

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

Therefore,

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

By the Dominated Convergence Theorem,

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

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

$$(A11) \qquad \lim_{n\to\infty} E_f[Z_n(f) - Y_n(f)]^2 = 0$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$, then Corollary A7 plus the Cauchy-Schwarz inequality implies that

$$(A12) \qquad \lim_{n\to\infty} E_f Z_n^2(f) = \beta_f''(0)$$

uniformly for all $f \in \overline{U}_{f_0}(\delta)$, which is just Lemma 4.12.

The verification of (A11) needs two steps.

Step 1. Following the proof of Theorem 3.1 (b′) - (e′), we have

$$|Z_n(f) - Y_n(f)| \le C(\frac{1}{\sqrt{l_n}} + |\frac{n-l_n}{\sqrt{n}}|),$$

for some $C > 0$ which does not depend on f.

By the Cauchy-Schwarz inequality,

$$E_f[Z_n(f) - Y_n(f)]^2 \le C^2\{E_f(\frac{1}{l_n}) + E_f(\frac{n-l_n}{\sqrt{n}})^2 + 2[E_f(\frac{1}{l_n}) \cdot E_f(\frac{n-l_n}{\sqrt{n}})^2]^{\frac{1}{2}}\}.$$

The next step is to show that

$$(A13) \qquad \lim_{n\to\infty} E_f\left(\frac{1}{l_n}\right) = 0$$

and

$$(A14) \qquad \lim_{n\to\infty} E_f\left(\frac{n-l_n}{\sqrt{n}}\right)^2 = 0$$

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

Step 2. In what follows, all probabilities $P(\cdot)$ corresponds to the Gibbs state μ_f.

Since $l_n \ge 1$ for all $n \in \mathbb{N}^+$, for every $\varepsilon \in (0,1)$

$$E_f(\frac{1}{l_n}) \le \int_0^1 P(\frac{1}{l_n} > t)dt \le \varepsilon + C\gamma^{[n^{\frac{1}{4}}]},$$

for some $C > 0$, $\gamma \in (0,1)$ and for all $n > N$. Notice that the constants C, γ, N can be so chosen that they do not depend on any particular $f \in \overline{U}_{f_0}(\delta)$ by reviewing the proof of Lemma 3.4. Let $n \to \infty$ and $\varepsilon \to 0$, then (A13) follows.

Furthermore, for every $\varepsilon \in (0,1)$

$$\begin{aligned} E_f\left(\frac{n-l_n}{\sqrt{n}}\right)^2 &= \int_0^\infty tP\left(|\frac{n-l_n}{\sqrt{n}}| > t\right) dt \\ &= \int_0^\varepsilon tP\left(|\frac{n-l_n}{\sqrt{n}}| > t\right) dt + \int_\varepsilon^{\sqrt{n}} tP\left(|\frac{n-l_n}{\sqrt{n}}| > t\right) dt \\ &\leq \frac{\varepsilon^2}{2} + P\left(|\frac{n-l_n}{\sqrt{n}}| > \varepsilon\right) \cdot \left[\frac{t^2}{2}\right]_\varepsilon^{\sqrt{n}}. \end{aligned}$$

By reviewing the proof of Lemma 3.4 we conclude that

$$P\left(|\frac{n-l_n}{\sqrt{n}}| > \varepsilon\right) \leq C\gamma^{[n^{\frac{1}{5}}]},$$

for some $C > 0$, $\gamma \in (0,1)$ and for all $n > N$. Here the constants C, γ, N again do not depend on any particular $f \in \overline{U}_{f_0}(\delta)$. Let $n \to \infty$ and $\varepsilon \to 0$, then (A14) follows.

Thus we have proved Lemma 4.12.

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

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

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

[5] 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.

[6] 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.

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

[8] Lang, S. (1972). Differential manifolds. Addison-Wesley, Reading, Massachusetts.

[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.