SOME ESTIMATION PROBLEMS FOR GIBBS STATES

by

Chuanshu Ji

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

ABSTRACT

The problems of estimating certain infinite-dimensional unknown parameters (such as interaction potentials, local characteristics, etc.) for Gibbs states are considered in this paper. We describe the construction of consistent estimators for one-dimensional Gibbs states and discuss the possibility of extending these results to multi-dimensional Gibbs states.

Key Words and Phrases: Gibbs state, consistent estimator

Running Title: Estimation for Gibbs States

§1. Introduction

Gibbs states have played an important role in recent developments of spatial statistics and image processing (cf. [1], [3], [11]). For instance, a two-dimensional picture may be described as a continuous region being partitioned into a fine rectangular array of sites or picture elements ("pixels"), each having a particular color, lying in a prescribed finite set. Most researchers assume there is a probability distribution on the set of pictures, and choose Gibbs states as models for this distribution.

Gibbs states were originally introduced as models in statistical mechanics. A particle on each site of a two-dimensional integer lattice may represent the "spins" of a magnet, and there usually are interactions between particles at different sites. A configuration consists of particles on all sites. Gibbs states are probability distributions on the set of all configurations. In image processing, particles are interpreted as colorings (pixel variables). Other intrinsic properties of Gibbs states are preserved.

In statistical context, Gibbs states are parametrized by certain unknown parameters which affect the interactions between pixels. One wants to estimate those parameters based on collected data. For our interest, the data are finite samples which are larger and larger pieces of an infinite configuration on a two-dimensional lattice from an infinite-volume Gibbs state (see [12] for its general definition).

The current state of art of parameter estimation for Gibbs states is reflected by a series of papers of Gidas (cf. [5], [6], [7], [8]). Assuming the Gibbs states are parametrized by elements in a finite-dimensional

Euclidean space, Gidas constructed the maximum likelihood estimators and showed that they are (i) consistent even at points of phase transitions; (ii) asymptotically normal and asymptotically efficient under appropriate conditions. He also investigated the asymptotic bias and variance, and studied the estimation problem based on partially observed data (degraded images).

As Gidas pointed out in his papers, one interesting open problem is to estimate certain infinite-dimensional unknown parameters which parametrize Gibbs states. In this case, consistency of maximum likelihood estimators fails in general. Therefore new ideas and methods need to be brought in. In Section 3 of this paper, we state a new result in estimating some infinite-dimensional unknown parameters for one-dimensional Gibbs states. We also give some heuristic arguments and make a few comments on the case of multi-dimensional Gibbs states in Section 4.

From the point of view of picture restoration, people prefer simple models such as Markov random fields with the nearest neighbor interactions, because the models with many parameters may be computationally too complicated. Nevertheless, it is still worthwhile considering models with infinite-dimensional parameters as the first step when we do not have much information about the true model. The estimators proposed in this paper may be more robust and easier to compute than estimators of maximum likelihood type. Further discussion in this respect will be in Section 4.

§2. Basic model and notation

We adopt the following *general (binary) Ising model* to illustrate the estimation problems under investigation. The general framework of multi-dimensional Gibbs states is given in [12].

Let $\mathbb{Z}^2$ be the two-dimensional integer lattice. With each pixel $i \in \mathbb{Z}^2$, we associate a random variable X_i taking values in the set $\{-1,1\}$, i.e. X_i represents the coloring of the pixel i. -1 and 1 may stand for "black" and "white" respectively.

A sequence of non-negative numbers $J = \{J(k), k \in \mathbb{Z}^2\}$ specifies the pair interactions so that $J(i-j)$ represents the interaction between the pixel i and pixel j. We assume that

(A1) $J(k) \geq 0, \quad \forall\, k \in \mathbb{Z}^2$;

(A2) $J(i-j) = J(j-i), \quad \forall\, i, j \in \mathbb{Z}^2$; (symmetry)

(A3) $\sum_{k\in\mathbb{Z}^2} J(k) < \infty$. (summability)

Apparently, the interaction function J so defined is also translation-invariant. An interaction J is said to have finite range if $J(k) = 0$ for all k with sufficiently large Euclidean norm $\|k\|$. For our interest, J should have infinite range.

For each configuration $x = (x_i, i \in \mathbb{Z}^2)$ and each particular pixel ℓ, define the energy

$$(2.1)\quad H_\ell(x) = h\, x_\ell + \beta \sum_{i\neq\ell} J(i-\ell)x_i x_\ell,$$

where $h \in \mathbb{R}$ is called the external field coefficient and $\beta > 0$ is called the inverse temperature. $H_\ell(x)$ can be interpreted as the contribution of the pixel ℓ to the total (pair interaction) energy associated with the configuration x.

The local characteristic at pixel ℓ is given by

$$(2.2)\quad P_\ell(x) = \frac{e^{H_\ell(x)}}{Z},$$

where $Z = Z(\ell;\ x_i,\ i \neq \ell)$ is the normalization such that

$$(2.3)\quad \sum_{x_\ell} P_\ell(x) = 1.$$

The two-dimensional Gibbs state μ is a probability measure on the space $\Omega = \{-1,1\}^{\mathbb{Z}^2}$ such that if $X = (X_i,\ i \in \mathbb{Z}^2)$ has the probability distribution μ, then for every $\ell \in \mathbb{Z}^2$ and $x \in \Omega$

$$(2.4)\quad \mu(X_\ell = x_\ell | X_j = x_j,\ j \neq \ell) = P_\ell(x).$$

So $P_\ell(x)$ is the conditional probability (under μ) that the coloring at pixel ℓ is x_ℓ given that the colorings at pixels other than ℓ are x_j, $j \neq \ell$. Note that the local characteristic $P \triangleq \{P_\ell(x),\ \ell \in \mathbb{Z}^2,\ x \in \Omega\}$ need not uniquely determine μ (possibility of phase transition).

For each finite subset S of $\mathbb{Z}^2$, let $x_S = (x_j,\ j \in S)$ be the subconfiguration of x restricted to S. And for each $i \in \mathbb{Z}^2$, let $x_{S+i} = (x_{j+i},\ j \in S)$ denote x_S translated by i. μ is said to be stationary if for every finite $S \subset \mathbb{Z}^2$, $i \in \mathbb{Z}^2$ and x_S

$$(2.5)\quad \mu(X_{S+i} = x_S) = \mu(X_S = x_S).$$

In this case the local characteristic P is also translation invariant in the sense that

$$(2.6)\quad \mu(X_\ell = x_0 | X_{\ell+j} = x_j, j \neq 0) = P_0(x),\ \forall\ \ell \in \mathbb{Z}^2 \text{ and } x \in \Omega.$$

Note that the translation invariance of P does not imply the stationarity of μ (possibility of symmetry breakdown).

The Gibbs state μ is parametrized by h, β and J. If h, β are unknown and J is known, that would be a special case of the model Gidas considered.

In this paper, we assume h, β are both known and J is unknown, which is an infinite-dimensional parameter. Let Λ_n be an n×n symmetric square in $\mathbb{Z}^2$, then X_{Λ_n} contains n^2 observations. We consider the following estimation problems:

(I) Estimate the local characteristic P = P(J);

(II) Estimate the interaction function J.

Since (2.5) implies (2.6), by assuming the true Gibbs state μ is stationary we only need to estimate the function $P_0(\cdot)$ in (I). It is observed that under our assumptions, for every finite $\Lambda \subset \mathbb{Z}^2$ the conditional probabilities $\mu(X_\Lambda = x_\Lambda | X_{\Lambda^c} = x_{\Lambda^c})$, $x \in \Omega$ can be derived from $P_0(\cdot)$.

A random function T_n on Ω constructed from X_{Λ_n} is called a (strongly) consistent estimator of $P_0(\cdot)$ if for every J

$$(2.7) \quad \sup_{y \in \Omega} |T_n(y) - P_0(y)| \to 0, \quad \text{a.s.} \quad \text{under } \mu \text{ as } n \to \infty.$$

Such an estimator T_n was constructed in [10] for one-dimensional Gibbs states when J satisfies certain conditions. The results will be sketched in Section 3.

The problem (II) is more subtle than (I) because J does not have a clear probabilistic interpretation as P has. In Section 4 we will give some ideas how to estimate J based on the estimator T_n for $P_0(\cdot)$.

§3. Estimating local characteristics for one-dimensional Gibbs states

One dimensional Gibbs states themselves are important in topological dynamics (cf. [2]). They may also be used as models in categorical time series. In this section some results for the problem (I) are stated in the context of one-dimensional Gibbs states. Hopefully, they could shed light on the multi-dimensional cases. All proofs are omitted. Further details can be found in [10].

For convenience we let the configuration space be

$$\Sigma^{+} = \{x = (x_0, x_1, \ldots) : x_i = -1 \text{ or } 1, \quad i=0,1,\ldots\}.$$

The results in this section can also be extended to other cases:

(i) $\Sigma = \{-1,1\}^{\mathbb{Z}^2}$, the set of all two-sided sequences;

(ii) Σ_A^{+} (resp. Σ_A), the set of all one-sided (resp. two-sided) sequences in which transitions between certain states of their coordinates are not allowed.

Let $X = (X_0, X_1, \ldots)$ be a stationary sequence with the probability distribution μ. Assume that the interaction J satisfies

(3.1) $J(n) \le C\,\rho^n$, $\forall\, n \in \mathbb{N}$ and some $C > 0$, $\rho \in (0,1)$.

Then for every h, β, J there uniquely exists a Gibbs state μ. Moreover, the following "uniform mixing" condition holds.

Lemma 3.1. *Let $\mathcal{A}_{m-1}$ be the σ-field generated by $(X_0, \ldots, X_{m-1})$; $\mathcal{A}_{m+n,\infty}$ be the σ-field generated by $(X_i,\ i \ge m+n)$. Then there exist constants $C > 0$ and $\alpha \in (0,1)$ such that*

(3.2) $$\left|\frac{\mu(A \cap B)}{\mu(A)\cdot\mu(B)} - 1\right| \le C\,\alpha^n$$

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

The local characteristics at 0 now are written as

(3.3) $P_0(x) = \mu(X_0 = x_0 \mid X_i = x_i,\ i \in \mathbb{N}) \triangleq \mu(x_0 \mid x_1, x_2, \ldots)$, $x \in \Sigma^{+}$.

$P_0(\cdot)$ is a unknown function when J is unknown. Given observations $X_0, \ldots, X_{n-1}$, we want to construct T_n such that

(3.4) $\sup_{y \in \Sigma^{+}} |T_n(y) - P_0(y)| \to 0$, a.s. under μ as $n \to \infty$.

This is just the one-dimensional version of (2.7). In (3.3) $P_0(\cdot)$ may be regarded as an infinite-step backward transition function, which suggests the following plan to construct T_n.

We first approximate $P_0(\cdot)$ by a sequence of finite-step (backward) transition functions $\{\mu(x_0|x_1,\dots,x_{m-1}),\ m \in \mathbb{N},\ x \in \Sigma^+\}$. Then at each stage m we estimate $\mu(x_0|x_1,\dots,x_{m-1})$ by "sample transition function" which is a ratio of two empirical measures. Based on convergence rate of estimators, we argue that the "step-length" m should grow like c log n, where the constant c is determined by certain bounds for the sequence J and its tail parts.

Define the forward shift operator $\sigma : \Sigma^+ \to \Sigma^+$ by $(\sigma x)_n = x_{n+1}$, $n = 0,1,\dots,$ for $x \in \Sigma^+$. Then we have

Construction of T_n

Given $X_0,\dots,X_{n-1}$, we define n periodic sequences $\sigma^j X(n)$, $j=0,1,\dots,n-1$, where

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

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

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

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

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

$$(3.5)\quad T_n(y) = \begin{cases} \dfrac{N_m^{(n)}(y)}{N_{m-1}^{(n)}(y)}, & \text{if } N_{m-1}^{(n)}(y) > 0 , \\ 0 , & \text{otherwise} \end{cases}$$

$T_n(y)$, also written as $\frac{N_m^{(n)}(y)}{n} / \frac{N_{m-1}^{(n)}(y)}{n}$, is the "sample conditional frequency" of y_0 appearing in the slot 0 given that $y_1, \ldots, y_{m-1}$ appear in the slots $1, \ldots, m-1$. The next two theorems show that under certain conditions T_n satisfies (3.4).

Theorem 3.2. *Suppose there are two known constants* $K > 0$ *and* $\rho \in (0,1)$ *satisfying*

$$(3.6) \quad \sup_{n \in \mathbb{N}} \frac{J(n)}{\rho^n} \leq K.$$

Then there exists $c \in (0,1)$ *which only depends on* K *and* ρ *such that* T_n *satisfies* (3.4) *if we choose*

$$(3.7) \quad m = [c \log n] \text{ (the integer part of } c \log n).$$

Theorem 3.3. *If we drop the assumption* (3.6) *and choose a sequence of constants* $\{c_n, n \in \mathbb{N}\}$ *such that* $c_n \downarrow 0$ *as* $n \to \infty$ *with an arbitrarily slow rate (e.g.* c_n *satisfies* $c_n \log n \to \infty$ *as* $n \to \infty$*). Set* $m = [c_n \log n]$. *Then* T_n *also satisfies* (3.4).

Note that (3.2) is a very strong mixing condition, under which X can be decomposed into two nearly iid sequences by the standard "blocking" and "splitting" technique. Hence some classical large deviation results hold. However, (3.2) need not be necessary for constructing consistent estimators. Future improvement is possible.

§4. Comments on the two-dimensional case

For the problem (I), we may just imitate the construction of T_n given in Section 3.

Without loss of generality, assume that the origin 0 is either at the center of the square Λ_n (when n is odd), or near the center of Λ_n (when n is even). Let B_m be an m×m symmetric square where

(4.1) m = the greatest odd integer less than or equal to $\sqrt{c \log n}$,

$c \in (0,1)$ is a constant. We let 0 also be the center of B_m. Λ_n can be partitioned into k subsquares $B^{(1)},\ldots,B^{(k)}$, plus some remainders near the edge of Λ_n. Each $B^{(j)}$ is of size m×m, j=1,...,k and $B^{(1)} = B_m$. Accordingly X_{Λ_n} consists of sub-configurations $X_{B^{(j)}}$, j=1,...,k plus the remainders.

For every $y \in \Omega$, let

$$N^{(n)}(y_{B_m}) = \sum_{j=1}^{k} I\{X_{B^{(j)}} = y_{B_m}\} \ ;$$

$$N^{(n)}(y_{B_m\setminus\{0\}}) = \sum_{j=1}^{k} I\{X_{B^{(j)}\setminus\{t_j\}} = y_{B_m\setminus\{0\}}\},$$

where t_j is the center of $B^{(j)}$, j=1,...k. Define

$$(4.2) \quad T_n(y) = \begin{cases} \dfrac{N^{(n)}(y_{B_m})}{N^{(n)}(y_{B_m\setminus\{0\}})} , & \text{if } N^{(n)}(y_{B_m\setminus\{0\}}) > 0 , \\ 0 , & \text{otherwise.} \end{cases}$$

We conjecture that T_n would satisfy (2.7) if J satisfies certain conditions. What kind of conditions should it be? Unlike the one-dimensional case, here it is not enough to just specify the tail behavior of J. For example, if h=0 and $\beta > \beta_c$ (β_c = inverse of critical temperature), then phase transition occurs even when J is the nearest neighbor interaction (i.e. J(k) = 0 for all k satisfying $\|k\| > 1$). In this case, the consistency of T_n would fail because μ is not unique and all mixing properties break down. Actually, the Gibbs random fields have long-range dependence. The similar construction might still provide consistent estimators if we assume that h, β and J correspond to an extremal Gibbs state and adjust the "grid size" m from

$[\sqrt{c \log n}]$ to something smaller (e.g. $m = [c(\log n)^\lambda]$, $\lambda > 0$ is small). However, all these speculations seem to be premature for the time being. Some work related to those open problems is in the process of development.

So far we have not discussed how to solve the problem (II). One way to do it is to derive the estimator $\hat{J}(k)$ of $J(k)$ $(k \in \mathbb{Z}^2,\ k \neq 0)$ from the estimator T_n of $P_0(\cdot)$. Notice that J can be determined by $P_0(\cdot)$ in the following way.

For every $x \in \Omega$, let $x' = (x_i,\ i \in \mathbb{Z}^2,\ i \neq 0)$ and

$$(4.3) \quad a(x') = h + \beta \sum_{k \neq 0} J(k)\, x_k .$$

Then $P_0(x) = \dfrac{e^{x_0 a(x')}}{e^{a(x')} + e^{-a(x')}}$.

Set $x_0 = -1$ and denote the corresponding x by x'_{-1}. We obtain

$$(4.4) \quad a(x') = \frac{1}{2} \log \left[\frac{1}{P_0(x'_{-1})} - 1\right].$$

Furthermore, let y, $z \in \Omega$ such that $y_0 = -1$, $y_i = 1$, $i \neq 0$; and $z_k = 1$, $z_i = -1$, $i \neq k$.

By setting $x'_{-1} = y$ and $x'_{-1} = z$ respectively in (4.4) we have for every $k \in \mathbb{Z}^2$, $k \neq 0$

$$(4.5) \quad J(k) = \frac{1}{\beta}\left\{\frac{1}{4}\left[\log\left[\frac{1}{P_0(y)} - 1\right] + \log\left[\frac{1}{P_0(z)} - 1\right]\right] - h\right\}.$$

If we have the estimates of $P_0(y)$ and $P_0(z)$, then (4.5) will produce the estimate of $J(k)$.

There could be a more direct way to estimate J based on Grenander's sieves method (cf. [9], [4]). If each sieve is generated by only finite number of parameters $J(k)$, then on each sieve we may try to find the maximum likelihood estimators and see if their limit yields a consistent estimator of

J when the sequence of sieves tends to the whole parameter space. A drawback of this approach is that usually the maximum likelihood estimators are computationally intractable if there are many parameters. While our previous method based on counting frequencies looks simpler. Actually, that is also one way of choosing sieves.

Acknowledgement

The author wishes to thank Professors Basilis Gidas and Steven Lalley for illuminating discussions and suggestions.

References

[1] Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). J. Roy. Stat. Soc., Series B, 48, 259-302.

[2] Bowen, R. (1975). *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*. Lecture Notes in Math. 470. Springer-Verlag, New York.

[3] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images. IEEE Trans., PAMI-6, 721-741.

[4] Geman, S. and Hwang, C. (1982). Nonparametric maximum likelihood estimation by the method of sieves. Annals of Statistics, 10, 401-414.

[5] Gidas, B. (1986). Consistency of maximum likelihood and pseudo-likelihood estimators for Gibbs distribution. Proceedings of the Workshop on *Stochastic Differential Systems with Applications in Electrical /Computer Engineering, Control Theory, and Operations Research*, IMA, University of Minnesota, 129-145.

[6] Gidas, B. (1987). Parameter estimation for Gibbs distributions, I: fully observed data. Tech. Report, Division of Appl. Math., Brown Univ.

[7] Gidas, B. (1988). Parameter estimation for Gibbs distributions, II: partially observed data. Tech. Report, Division of Appl. Math., Brown Univ.

[8] Gidas, B. (1988). Asymptotic bias and variance of maximum likelihood estimators at points of phase transitions. Tech. Report, Division of Appl. Math., Brown Univ.

[9] Grenander, U. (1981). *Abstract Inference*. Wiley, New York.

[10] Ji, C. (1988). Estimating potential functions of one-dimensional Gibbs states under constraints. Tech. Report 88-13, Dept. of Stat., Purdue Univ.

[11] Ripley, B. (1988). *Statistical Inference of Spatial Processes*. Cambridge Univ. Press, New York.

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