INFERENCE ABOUT THE SHAPE OF NEIGHBORING POINTS IN FIELDS

J. L. Denny and A. L. Wright
Department of Statistics
University of Arizona
Tucson, Arizona 85721
July 1988

ABSTRACT

One observes a finite sample from a discrete process indexed by a d-dimensional lattice, $d >= 1$. If $d = 1$ and it is assumed that the process is a Markov chain of order at most t, known, then procedures are available for estimating the order and obtaining the admissible confidence intervals. These procedures are (partially) extended to Markov fields, $d >= 2$, where it is necessary to consider shapes and where the "order" is described by the nearest neighbor potential. Bayes tests are obtained for testing independence against a dependence described by an arbitrary clique. The equivalence of a minimal sufficient statistic and a canonical nearest neighbor Gibbs potential is obtained for arbitrary Markov chains and for Markov fields which are exponential families.

1980 Mathematics Subject Classification (1985) revision). Primary 62M05, 62F05; Secondary 60J10.

1. Introduction

Imagine that you are looking at a map of southern Arizona and you see the locations of twenty rain-gauge stations. In addition to the map you have daily precipitation data for the past ten years at each station. (An extensive network of raingauge stations exists near Tombstone, Arizona, the location of Wyatt Earp's shootout at the OK Corral.) Next imagine you are looking at a black-and-white photograph, greatly enlarged. In both cases, the rain-gauge stations and the photograph, we can ask if there is a pattern of dependence in the data. We study this question, dependence, using techniques from statistical inference about Markov chains and the ideas in the pioneering papers of Besag (1972, 1974) and Pickard (1976, 1982, 1987) as our guide.

We assume the reader is familiar with the notions of Markov chains and their generalization to spatial variables, Markov fields. Because the literature on chains is enormous, we give no general references. The terminology we use for Markov fields is taken from the books by Kindermann and Snell (1980) and Preston (1974). We first briefly discuss statistical inference about the order of Markov chains, then proceed to Markov fields. Along the way, we establish a correspondence between minimal sufficient statistics and nearest neighbor Gibbs potentials, obtaining as a corollary a method for estimating the cliques in the canonical nearest neighbor potential.

Let $X_1, X_2, \ldots$ be a stationary Markov chain with $r+1$ states, $1 \leq r < \infty$. (With more notation the results of this section extend to countable-state chains.) If $x = (x_1, \ldots, x_N)$ is a sample of size N then the density is

$$P(x) = p_{x_1} p_{01}^{f_0} \cdots p_{or}^{f_{or}} p_{10}^{f_{10}} \cdots p_{1r}^{f_{1r}} \cdots p_{r1}^{f_{r1}} \cdots p_{rr}^{f_{rr}}, \tag{1}$$

where p_{x_1} is the probability of the initial state, the p_{ij} are the transition probabilities, and the f_{ij} count the number of times state i is followed by state j in the sample. The statistic $T = (x_1, \{f_{ij} : i,j = 0, \ldots, r\})$ is a minimal sufficient statistic if $r \geq 2$ (Wright (1980)). Two remarks are in order. If $X_1, X_2, \ldots$ is a stationary two-state Markov chain, there exists a complete sufficient statistic (Denny and Wright (1978), Denny and Yakowitz (1978)), but if the stationary chain has k states, $3 \leq k \leq \infty$, there exist no nontrivial optimal estimation and confidence interval methods, without imposing additional structure on the statistical decision set-up. To give an example, the sample mean from a stationary 3-state Markov chain is not a minimum variance unbiased estimate of the expected value of the chain. On the other hand, Pickard (1987) has observed that for some stationary two-state Markov fields, there does exist a complete sufficient statistic. Recall that the statistic $T = (x_1, f_{ij} : i,j = 0, \ldots r)$ is a minimal sufficient statistic for a stationary Markov chain with $k \geq 3$ states. The definitions of cliques and potentials are recalled in Section 2.

Theorem 1. *The minimal sufficient statistic T determines the cliques in the canonical nearest neighbor Gibbs potentials, and conversely.*

Proof. Let $P(x)$ be a probability determined by a nearest neighbor Gibbs potential V, so that

$$P(x) = Z^{-1} \exp\left\{ \sum_C V_C(x) \right\}, \tag{2}$$

where the summation is over all cliques V_C on $1, \ldots, N$. In general the V_C are not unique. We first show that the representation (1) determines a potential V so that (2) holds and also

(a) V is a nearest neighbor Gibbs potential: $V_C(x) = 0$ whenever C is not a clique;

(b) $V_C(x) = 0$ whenever $x_t = 0$ for at least one $t \in C$.

The representation (1) becomes

$$\begin{aligned} P(x) &= (p_{x_1}/p_0)p_0 p_{00}^{N-1-f_{01}-\ldots-f_{rr}} p_{01}^{f_{01}} \ldots p_{rr}^{f_{rr}} \\ &= P(\overline{0})(p_{x_1}/p_0)(p_{01}/p_{00})^{f_{01}} \ldots (p_{rr}/p_{00})^{f_{rr}}, \end{aligned}$$

where $\overline{0} = (0, \ldots, 0)$ and we use $f_{00} + \ldots + f_{rr} = N - 1$. Let $f_{\cdot j} = \sum_i f_{ij}$ and $f_{i\cdot} = \sum_j f_{ij}$. Then

$$\begin{aligned} P(x) &= P(\overline{0})(p_{x_1}/p_0)\left(\prod_{i,j\geq 1}(p_{ij}/p_{00})^{f_{ij}}\right)\cdot\left(\prod_{j=1}^{r}(p_{0j}/p_{00})^{f_{0j}}\right)\cdot\left(\prod_{i=1}^{r}(p_{i0}/p_{00})^{f_{i0}}\right) \\ &= P(\overline{0})(p_{x_1}/p_0)\left(\prod_{i,j\geq 1}(p_{ij}p_{00}/p_{i0}p_{0j})^{f_{ij}}\right)\cdot\left(\prod_{j=1}^{r}(p_{0j}/p_{00})^{f_{\cdot j}}\right)\cdot \\ &\qquad \left(\prod_{i=1}^{r}(p_{i0}/p_{00})^{f_{i\cdot}}\right). \end{aligned} \tag{3}$$

Let $\overline{f}_i$ equal the number of i's in $(x_2, \ldots, x_{N-1})$. Because $f_{i\cdot}$ is the number of i's in $(x_1, \ldots, x_{N-1})$ and $f_{\cdot j}$ is the number of j's in $(x_2, \ldots, x_N)$, then $\overline{f}_i$, $f_{i\cdot}$, and $f_{\cdot i}$ differ by at most one. Moreover, (x_1, x_N) determines whether $|f_{i\cdot} - \overline{f}_i| = 1$ or 0, and similarly for the other cases. But $(x_1, \{f_{ij}\})$ determines x_N (Billingsley (1961)) and since $T(= (x_1, \{f_{ij} : i, j \geq 0\}))$ is equivalent to $S = (x_1, \{f_{ij} : i, j \geq 1\}, \{f_{i\cdot}, f_{\cdot i} : i = 1, \ldots, r\})$, it follows that S determines x_N.

The preceding facts together with (3) gives us

$$P(x) = P(\overline{0})\exp\{H_1(x_1) + H_2(x_N) + \sum_{i,j\geq 1}^{r} f_{ij}\ell n(p_{ij}p_{00}/p_{i0}p_{0j}) + \sum_{i=1}^{r}\overline{f}_i\ell n(p_{i0}p_{0i}/p_{00}^2)\}, \tag{4}$$

where the functions H_1 and H_2 are obtained by enumerating the different cases for x_1 and x_N, and $H_1(0) = H_2(0) = 0$. Then (4) becomes (2) if we define $Z^{-1} = P(\overline{0}), V_{\{1\}}(x) = H_1(x_1), V_{\{N\}}(x) = H_2(x_N)$, $V_{\{s\}}(x) = \ell n(p_{i0}p_{0i}/p_{00}^2)$ if $x_s = i$ for $s = 2, \ldots, N-1$ and $i = 0, \ldots, r$, $V_{\{s,s+1\}}(x) = \ell n(p_{ij}p_{00}/p_{i0}p_{0j})$ if $x_s = i \geq 1$ and $x_{s+1} = j \geq 1$ or $V_{\{s,s+1\}}(x) = 0$ if either x_s or $x_{s+1} = 0$ for $s = 1, \ldots, N_1$, and $V_c(x) = 0$ in all other cases. Because T and S are equivalent, it follows that S determines the canonical potential V, above, and by reversing the argument, that V determines S. This completes the proof.

Corollary. *The maximum likelihood estimates of the cliques are obtained from the above V_C by replacing the p_{ij} by their maximum likelihood estimates.*

Remarks. Theorem 1 generalizes the discussion, referring to two-state chains, on pages 29 to 31 of Kindermann and Sell (1980). Tanemura (1988) considers questions related to the Corollary.

We recall briefly the problem of deciding the order of dependence in a Markov chain of order $s \leq t$, t known. Recall that a process $\{X_i : i = 1, 2, \ldots\}$ is a Markov chain of order s if the process $\{Y_i : i = 1, 2, \ldots\}$ is an ordinary Markov chain of order 1, where $Y_i = (X_i, X_{i+1}, \ldots x_{i+s-1})$. (Because each chain of order s is of order $s+1$, the order of a chain is of course the smallest such s.) Perhaps the earliest paper to consider testing independence (order 0) against an assumed order of at most 1 is that of David (1947), who developed the run test. David considers two-state stationary chains. Using the relation between similar tests and complete sufficient statistics, one can show that there are tests for independence against the alternative $P(X_i = 1|X_{i-1} = 1) > P(X_i = 1|X_{i-1} = 0)$ which are, for suitable levels of significance, uniformly most powerful among the class of unbiased tests (Lehmann (1986)). As soon as one tries to generalize this result to k-state chains or chains of order $s \geq 2$, one encounters the difficulty that there is no complete sufficient statistic for such stationary chains. (See above.) Nevertheless, by assuming the initial state is fixed (the initial state is $(x_1, \ldots, x_t)$ if the order is assumed to be less than or equal to t), one obtains the result that all admissible level α tests for the hypothesis that the order is $s \leq s_0$ against $s_0 < s \leq t$ are Bayes tests based on the transition count statistic. Other facts about tests for dependence in chains may be found in the survey article by Denny (1985).

2. Markov fields

In this section we recall definitions and properties of Markov fields, mostly following the paper of Besag (1974), but also touching on ideas in Kindermann and Snell (1980). We give a version of the Hammersley-Clifford theorem, again following Besag, which is useful for statistical problems. A reader familiar with the basic definitions may want to proceed to Section 3. Let S be a countable set and let $\{X_s : s \in S\}$ be a collection of random variables. We say that elements $s, t \in S$ are neighbors if the random variables X_s, X_t are *not* conditionally independent given $\{X_i : i \in S, i \neq s, t\}$. Recall the definition of a clique: a non-empty subset $C \subset S$ is a clique if either (a) C is a singleton $\{s\}$, or (b) each element of C is a neighbor of every other element of C. Assume that each $X_s, s \in S$, can take at most a countable set of values, say $\{0, 1, \ldots\}$. Because we appeal to the Hammersley-Clifford theorem (Besag (1974)) we assume that $P(X_s = 0, s \in S) > 0$. Let $x = (x_s, s \in S)$ denote a particular realization of $\{X_s, s \in S\}$ and assume that S is finite, relabeling $x = (x_1, \ldots, x_N)$. Let $\mathcal{X}$ denote the set of all realizations x, so that $\mathcal{X}$ can be identified with a product of N copies of the nonnegative integers. If Q is an arbitrary function on $\mathcal{X}$ with $Q(\overline{0}) = 0$, then it is elementary to show that Q has a unique representation

$$Q(x) = \sum_{A(1)} x_i G_i(x_i) + \sum_{A(2)} x_i x_j G_{i,j}(x_i x_j) + \ldots + \sum_{A(N)} x_1 x_2 \ldots x_N G_{1,2,\ldots,N}(x_1, x_2, \ldots, x_N),$$

where $A(k)$ is the set of all ordered k-tuples, $1 \leq j_1 < j_2 < \ldots, j_k \leq N$. Recall the definition of a Markov field: the process $\{X_s : s \in S\}$ is a Markov field if $P(X_s = x_s | X_t = x_t, t \in S, t \neq s) = P(X_s = x_s | X_t = x_t, t \in N_s)$, where N_s is the set of neighbors of s, this holding for all $s \in S$. Additional requirements may be needed if S is not a finite set, and to simplify we will hereafter assume S is finite. Here is a version of the Hammersley-Clifford theorem, not as Hammersley and Clifford stated it, but adequate for our purposes. We omit a proof; but see Besag (1974).

Theorem. *Let $\{X_s : s \in S\}$ be a process, with S a finite set. Assume a representation $P(x) = P(\overline{0}) e^{Q(x)}, Q(\overline{0}) = 0$, with $P(\overline{0}) > 0$, so that $\sum_x e^{Q(x)} < \infty, P(\overline{0}) =$*

$\left(\sum_x e^{Q(x)}\right)^{-1}$. *If* $\{X_s : s \in S\}$ *is a Markov field then (a) if a function* $G_{i,j\ldots,k}(x_i, x_j, \ldots, x_k)$ *in the representation of* Q *is not identically zero then* $(i, j, \ldots, k)$ *is a clique, and (b) for fixed* (i, j), *if* $G_{M(k)}$ *is identically zero for all* k*-tuples* $M(k), k \geq 2$, *containing* (i, j) *then* (i, j) *is not a clique. Conversely, assume* $\{X_s\}$ *is a process with the above representation. Then* $\{X_s\}$ *is a Markov field with a neighborhood system* $\{N_s : s \in S\}$ *defined as follows: for each* $s \in S, N_s = \{t :$ *there exists a* k*-tuple* $M(k)$, $k \geq 2$, *containing* $\{s, t\}$ *so that* $G_{M(t)}$ *is not identically zero*$\}$.

We digress to rephrase the above in the language of Kindermann and Snell. Assume that Gibbs measure representation

$$P(x) = Z^{-1} \exp\left\{\sum_C V_C(x)\right\}, \tag{5}$$

the summation over all non-empty subsets of S. The potential V, a function of C and s satisfies: if $C = \{i, j, \ldots, k\}$ then $V_C(x)$ is a function of $(x_i, x_j, \ldots, x_k)$. A potential V is called a nearest neighbor Gibbs potential if $V_C(x)$ is identically zero whenever C is not a clique, and is called the canonical potential if $V_C(x) = 0$ whenever $x_t = 0$ for at least one $t \in C$. Clearly, if each $V_C(x) = x_i x_j \ldots x_k G_{i,j,\ldots,k}(x_i, x_j, \ldots, x_k), C = \{i, j, \ldots, k\}$, then V is a canonical nearest neighbor Gibbs potential, is unique, and $Z^{-1} = P(\overline{0})$.

3. Markov fields – parametric Bayesian inference

In this section we propose some natural Bayes estimators of the extent (order) of the dependence in Markov fields, assuming the densities of the fields to be a k-parameter exponential family. One result uses the theorems of Pickard (1982, 1987), which prove consistency of the maximum likelihood estimators (avoiding phase transitions), and theorems of Le Cam (1958, 1986) about asymptotic properties of Bayes procedures.

Because we follow Pickard, we use his assumptions. To begin, let Z denote the integers and let Z^d, $d \geq 1$, denote the product of d copies of Z. We introduce a simple graph G_∞ on Z^d, requiring that there exists a finite upper bound on the distance between two vertices joined by an edge. Let $S \subset Z^d$ be a finite subset whose boundary δS satisfies $|\delta S| = O(|S|^{1-1/d})$, so that the dimension of δS is essentially smaller than that of S. Let G denote the subgraph of G_∞ inherited by S. Recall that a simple graph is complete if each pair of vertices is adjacent. If C_1 and C_2 are complete subgraphs of G, then C_1 and C_2 are said to be equivalent if C_1 is a translate of C_2. Let $\mathcal{C}_1, \ldots, \mathcal{C}_k$ denote the equivalence classes of complete subgraphs of G with the above equivalence relation, where $\mathcal{C}_1$ is the class of singletons. With each $s \in S$ associate a random variable X_s which takes values in $\{0, 1\}$. We define a statistic T on $\mathcal{X} = \{0, 1\}^S$ (the set of all functions from S into $\{0, 1\}$), $T' = (T_1, \ldots, T_k)$, by

$$T_i(x) = \sum_{C \in \mathcal{C}_i} \prod_{j \in C} x_j, \; x \in \mathcal{X}.$$

For each $a \in R^k, a' = (a_1, \ldots, a_k)$, define a probability P_a on $\mathcal{X}$ by

$$P_a(x) = c(a) e^{a'T}. \tag{6}$$

In particular, $P_a(\overline{0}) = c(a)$. If for $a \in R^k$ the distribution of $\{X_s\}$ is given by $P_a(X = x) = P_a(x)$ then by the Hammersley-Clifford theorem $\{X_s\}$ is a Markov field with

a neighborhood system $\{N_s : s \in S\}$ defined by $N_s = \{t : t$ belongs to a complete subgraph $C \in \bigcup_2^k c_i$ containing $s\}$.

The representation (6) of $P_a(x)$ gives a correspondence between complete sufficient statistics and the canonical nearest neighbor Gibbs potential. The situation is simpler than the case of Markov chains, for there is no transition count statistic. The following theorem, whose proof is immediate, holds for Markov fields with X_s taking values in $\{0, 1, \ldots\}$.

Theorem 2. *The representation (6) in terms of the complete sufficient statistic T implies the representation (5) with the canonical nearest neighbor Gibbs potential V_C defined by $V_C(x) = a_i \prod_{j \in C} x_j$, for a_i corresponding to T_i.*

We briefly describe the decision problem, our methods being standard. As already indicated, the parameter space is R^k, with $a' = (a_1, \ldots, a_k)$ denoting a parameter point. We will decompose R^k in order to decide on the extend of the dependence. Let $A(0) = \overline{0} \in R^k$, and for each non-empty subset $\{i_1, \ldots, i_p\} \subset \{1, \ldots, k\}$ let $A(i_1, \ldots, i_p) = \{(a_1, \ldots, a_k) : |a_{i_j}| > 0$ for only $i_1, \ldots, i_p\}$. We let the set Δ of decisions be the 2^k sets $A(i_1, \ldots, i_p)$. Choosing $A(0)$ is equivalent to agreeing that $a = (0, \ldots, 0)$ and that the random variables X_s are independent, and choosing the other $A(i_1, \ldots, c_p)$ is to be interpreted similarly. We introduce the remaining (familiar) elements of the decision problem. Let $W(a, \overline{s})$ denote a bounded loss function, $a \in R^k$, $\overline{s} \in \Delta$, and let a decision rule $\overline{s}$ associate with each $x \in \mathcal{X}$ a distribution F_x on Δ. Let

$$W(a, \overline{s}, x) = \int_\Delta W(a, \overline{s}) F_x(ds),$$

let

$$R(a, \overline{s}) = \int_x W(a, \overline{s}, x) P_a(dx);$$

and let

$$r(\mu, \overline{s}) = \int_{R^k} R(a, \overline{s}) \mu(da),$$

the Bayes risk, where μ is a probability on R^k, the prior probability. Let $\delta(\mu, \overline{s}) = r(\mu, \overline{s}) - \inf\{r(\mu, \overline{s}')\}$, the inf being taken over all decision rules $\overline{s}'$. If $\{\overline{s}_N\}$ is a sequence of decision rules, then $\{\overline{s}_N\}$ is said to be asymptotically Bayes with respect to μ if $\delta(\mu, \overline{s}_n) \to 0$. Let μ be a probability on R^k. Then there exist probabilities μ_i satisfying $\mu_i(A(i)) = 1$ and nonnegative $\pi_i, \sum \pi_i = 1$, so that $\mu(B) = \sum \pi_i \mu_i(B)$, Borel $B \subset R^k$. By the usual argument,

$$P(A(i)|x) = c(x) \pi_i \int c(a) e^{a' T(x)} \mu_i(da),$$

where P is the probability induced by μ and the P_a. Assume now that $W(a, s) = 0$ if a belongs to the subset of R^k identified with the decision, and $W(a, s) = 1$ otherwise. Then clearly the Bayes decision rule $\overline{s}$ minimizing the Bayes risk is: $\overline{s}(x) =$ that i maximizing $P(A(i)|x)$.

We turn to the question of the dependence of $\bar{s}$ on a particular prior μ. Theorem 3 shows that for a class of priors avoiding phase transition the above decision rule is asymptotically Bayes. This fact, a consequence of theorems of Pickard (1982) and Le Cam (1958, 1986), may turn out to be useful, for finding computationally feasible Bayes procedures has eluded us in the general case. We now follow the assumptions of Pickard (1982). Let $A \subset R^k$ be a compact set, disjoint from $\mathcal{F}$, the set of phase transitions. Recall that $c(a), a \in R^k$, denotes the constant term in the probability P_a on $\mathcal{X}$, and note that $\mathcal{X}$ and $c(a)$ depend on $S \subset Z^d$, the set of sites. Consider an increasing sequence of S, ordered by inclusion and satisfying the above boundary conditions, and let $N = |S|$. In this section we write $c(a) = c_N(a)$. Assume there is a function H defined on A so that uniformly for $a \in A$

(i) $N^{-1}\ell n c_N(a) \to H(a)$,

(ii) $$\frac{N^{-1}\partial^{r_1+\ldots+r_k}\ell n c_N(a)}{\partial a_1^{r_1}\ldots\partial a_k^{r_k}} \to \frac{\partial^{r_1+\ldots+r_k}H(a)}{\partial a_1^{r_1}\ldots\partial a_k^{r_k}},$$

for $r_1 + \ldots + r_k \leq 3$, and H satisfies

(iii) The Hessian of H is positive definite.

Theorem 3. *With the above assumption on A and $c_N(a)$, $a \in A$, let μ be a probability, $\mu(a) = 1$. If ν is a probability absolutely continuous with respect to μ and $\{\bar{s}_N\}$ is a sequence of decision rules asymptotically Bayes for μ, then $\{\bar{s}_N\}$ is asymptotically Bayes for ν.*

Proof. From Theorem 1 of Pickard (1982) it follows that the maximum likelihood estimates $\hat{a}$ of $a \in A$ are consistent. The assertion follows from Proposition 2 and Theorem 1 of Le Cam (1958).

We specialize the decision problem to obtain computationally feasible methods which are applicable in some generality. Looking at the parametric Markov fields defined by (6), one sees that the extent of the dependence involves the shape of the cliques, the size of the cliques, and the weighting of the cliques – the parameter $a \in R^k$. Notice that in such a parametric statistical problem, the statistician chooses the shape, size, and number of the cliques, and only the weighting remains unknown. We need not elaborate on the perils of the doubtful path. Nevertheless, that is the path we have taken.

We first consider only two $\mathcal{C}_i$, one being $\mathcal{C}_1$, the equivalence class of singletons. Let then C_2 denote the equivalence class of a specific complete subgraph on a finite subset $S \subset Z^d$, $\mathcal{C}_2 \neq \mathcal{C}_1$. In the notation of (6), we have $a = (a_1, a_2), T'(x) = (T_1(x), T_2(x))$, where $T_1(x) = \Sigma x_j, T_2(x) = \sum_{C \in e_2} \prod_{j \in C} x_j, x = (x_1, \ldots, x_N) \in \mathcal{X}$.

To simplify notation, we assume hereafter that the X_j take only the values $\{0, 1\}$. The results extend at once to countable-valued X_j provided that $E(\prod_j X_j)^2 < \infty$, $C \in \mathcal{C}_2$. In general $P_a(x) = c(\bar{o})e^{a_1 T_1(x) + a_2 T_2(x)}$, and the X_j are in independent if and only if $a_2 = 0$.

The uniformly most powerful unbiased tests of the hypotheses that $a_{20} \leq a_2 \leq a_{21}$ are given by: reject the hypotheses if $T_2(x) < c_1(T_1(x))$ or $T_2(x) > c_2(T_1(x))$, with

randomization on the boundary, where of course the constants $c_i(T_1(x))$ depend on a prescribed level of significance. If $a_{20} = a_{21} = 0$, then the above test is a test of independence against a dependence described by $\mathcal{C}_2$. If the parameter space is compact and $a_{20} = a_{21} = 0$, then such tests, being admissible, fall in the class of Bayes tests described earlier.

We assume hereafter that $a_2 = 0$. The remaining problem is to find the values $c_1(T_1(x))$ and $c_2(T_2(x))$, and to tie these values to the prior probability μ. But in this paper we sidestep the choice of the actual prior and will simply point out a choice of c_1 and c_2 based on hypothesis testing considerations. (Thus, the actual decision rule, the choice of the prior or the level of significance, is left to "the reader", as is often the case.) Because $a_2 = 0$, the statistic T, suitably normalized, has the asymptotic distribution of a normal vector Z, as $N = |S| \to \infty$. The asymptotic normality is immediate since T_1 is a sum of independent random variable and T_2 is a sum of k-dependent random variables; the asymptotic normality is also a consequence of Theorem 1 of Pickard (1982). So if $(\mu(a_1), \Sigma(a_1))$ is the mean and covariance of Z, familiar methods approximate the constants $c_1(T_1(x))$ and $c_2(T_1(x))$.

It may be suggestive to describe the problem in terms of pixels. Imagine a rectangle $S \subset Z^2$ with m rows and n columns, where each site in the rectangle corresponds to a pixel which is either on or off. Let us agree that 1 corresponds to a pixel being on, 0 to a pixel being off. Then $T_1(x)/mn$ is the proportion of pixels which are on. In addition , $T_2(x)$ is the sum of $\ell(m,n)$ products, where $\ell(m,n)$ is the number of cliques. A particular product is 1 only when all pixels in the corresponding clique are on. Therefore $T_2(x)/\ell(m,n)$ is the proportion of cliques which are totally lit up. We reject independence of the pixels against to an equivalence class of cliques if the proportion of totally lit up cliques is large or small relative to the proportion of turned on pixels.

As an example, assume that $\mathcal{C}_2$ is the class of cliques equivalent to a star – an element of $\mathcal{C}_2$ is a translation of $\{(0,0),(1,0)(0,-1)(-1,0),(0,1)\}$. We have that $T_1(x) = \Sigma x_i$ and $T_2(x) = \sum_C \prod_{j \in C} x_j$, where $C \in \mathcal{C}_2$. Assume the rectangle S has m rows and n columns, and that m and n are large. Recall that the conditional distribution of T_2 given T_1 does not depend on a_1, when $a_2 = 0$. with $N = mn$, $\overline{T}_2 = T_2/(m-2)(n-2)$, we have asymptotically that $N^{\frac{1}{2}}(\overline{t}_2 - \widehat{p}^5)$ is normal with mean 0 and variance $\widehat{p}^5(1 + 9\widehat{p}^3 + 4\widehat{p}^4 - 13\widehat{p}^5)$, where $\widehat{p} = T_1/N$.

Bibliography

Besag, J. E. (1972). Nearest-neighbor systems and the auto-logistic model for binary data. *Journal of Royal Statistical Society* B **34**, 75–83.

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society* B **36**, 192–236.

Billingsley, P. (1961). Statistical methods in Markov chains. *Ann. Math. Statist.* **32**, 12–40.

David. F. N. (1947). A power function for tests of randomness in a sequence. *Biometrika* **34**, 335–339.

Denny, J. L. (1985). *Markovian dependence.* Encyclopedia of Statistical Sciences, Volume 3. Wiley, New York.

Denny, J. L., & Yakowitz, S. (1978). Admissible run-contingency type tests for independence. *Journal of American Statistical Association* **72**, 177–181.

Denny, J. L., & Wright, A. L. (1978). On tests for Markov dependence. *Z. Wahrsch.* **43**, 331–338.

Kindermann, R., & Snell, J. L. (1980). *Markov random fields and their applications.* American Mathematical Society, Providence.

Le Cam, L. (1958). Les propriétés asymptotiques des solutions de Bayes. *Publ. Inst. Statist. Univ. Paris* **7**, 17–35.

Le Cam, L. (1986). *Asymptotic methods in statistical decision theory.* Springer-Verlag, New York.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses.* 2nd Ed., John Wiley, New York.

Pickard, D. K. (1976). Asymptotic inference for an Ising lattice. *Journal of Applied Probability* **13**, 486–497.

Pickard, D. K. (1982). Inference for general Ising models. *Journal of Applied Probability* **19** (A), 345–357.

Pickard, D. K. (1987). Inference for discrete Markov fields: the simplest nontrivial case. *Journal of American Statistical Association* **82**, 90–96.

Presont, C. J. (1974), *Gibbs states on countable sets.* Cambridge University Press, Cambridge, U.K.

Tanemura, M. (1988). Estimation of interaction potentials of Gibbs processes in one-dimensional space. *Conference on Spatial Statistics and Imaging.*

Wright, A. L. (1980). Nonexistence of complete sufficient statistics for stationary K-state Markov chains, $K \geq 3$. *Ann. Insti. Statist. Math.* **32** (A), 95–97.