Splitting Markov fields
and combining empirical estimators

Priscilla E. Greenwood        Ian W. McKeague
University of British Columbia Florida State University
Wolfgang Wefelmeyer
Universität-GH Siegen

Abstract

We have shown elsewhere that the empirical estimator for the expectation of a
local function on a Markov field over a lattice is efficient if and only if the function
is a sum of functions each of which depends only on the values of the field on
a clique of sites. For countable state space, the estimation of such expectations
reduces to the estimation of probabilities of configurations over finite subsets of
the lattice. The corresponding empirical estimator is efficient if and only if the
set is a clique. If the set is not a clique, we can construct better estimators.
They are rational functions of empirical estimators for configurations over subsets
of the given set. The construction is based on a factorization of probabilities of
configurations which makes use of the splitting property of Markov fields.


Key words and Phrases. Empirical estimator, improved estimator, Markov splitting,
local interactions, random field.

1 Introduction

We want to estimate the expectation of a function on a random field with known finite
range of interactions. The usual estimator is the empirical estimator. For consistency
see Burton and Steif (1995) and (1997) and Steif (1997); asymptotic normality follows
from the central limit theorem, see Künsch (1982), Bolthausen (1982) and Dedecker
(1998). The empirical estimator does not make use of the information about the range

\footnote{Work supported by NSERC, Canada, and the Crisis Points group of the Peter Wall Institute
for Advanced Studies at UBC. Research done while the third author was visiting the Department of
Mathematics at UBC.}
of interactions and is, in general, not efficient. We construct estimators using this information. This use of structural information about the field can be regarded as a semiparametric approach. Nevertheless, our improved estimator will be a function of certain empirical estimators.

Asymptotic normality of empirical estimators uses central limit theorems for centered random variables. On the other hand, estimators for parametric families are often obtained from estimating functions which are sums of conditionally centered random variables. The coding estimator of Besag (1974a) uses the local characteristics of a random field at a coding set, i.e., a set of sites which are conditionally independent given the other sites; see also Besag (1974b). This amounts to using a partial likelihood. The maximum pseudo-likelihood estimator of Besag (1975) and (1977) uses the local characteristics at all sites. It is asymptotically normal under weak conditions; see Jensen and Künsch (1994), Janžura and Lachout (1995) and Comets and Janžura (1998). For more information on estimation in parametric models, including the maximum likelihood estimator, we refer to Janžura (1988), Comets and Gidas (1992), Gidas (1993), Guyon (1995) and Janžura (1997).

Let \( X_i, i \in \mathbb{Z}^d \), be a homogeneous random field on the \( d \)-dimensional square lattice, with discrete state space \( E \), and let \( \mu \) denote its distribution. The expectation of a function \( f \) can be written \( \mu(f) = \sum \mu(a)f(a) \), where the sum extends over all configurations \( a \). The estimation of such an expectation reduces therefore to the estimation of probabilities of configurations. Suppose we observe the field in a window \([-n, n]^d\) and want to estimate a probability \( \mu(a_D) = \mu(X_D = a_D) \), where \( D \) is some finite subset of the lattice, and \( a_D = (a_i)_{i \in D} \) is a point in \( E^D \). If nothing is known about \( \mu \), we will use the empirical estimator
\[
E_n(a_D) = \frac{1}{|D_n|} |\{ j : D + j \subset [-n, n]^d, X_{D+j} = a_D \}|.
\]

Here \( |D_n| \) is the number of shifts of \( D \) which are contained in the window \([-n, n]^d\).

For Gibbs fields satisfying certain regularity conditions, Greenwood and Wefelmeyer (1998) have shown that if \( \mu \) has local interactions with known range, then the empirical estimator \( E_n(a_D) \) is efficient for \( \mu(a_D) \) if and only if \( D \) is a clique, i.e., a set such that each two points are neighbors; see also Section 2.

Related results are known in the i.i.d. case and for Markov chains. If \( X_1, \ldots, X_n \) are i.i.d. with unknown distribution \( \pi \), then the empirical estimator
\[
E_n(A) = \frac{1}{n} |\{ i : X_i \in A \}|
\]
is efficient for the probability \( \pi(A) \), but the empirical estimator
\[
E_n(A \times B) = \frac{1}{n-1} |\{ i : X_{i-1} \in A, X_i \in B \}|
\]
is not efficient for $\pi \otimes \pi (A \times B)$. Since $\pi \otimes \pi (A \times B) = \pi(A)\pi(B)$, an efficient estimator is the product of the two empirical estimators for $\pi(A)$ and $\pi(B)$, the von Mises statistic

$$E_n(A)E_n(B) = \frac{1}{n^2}|\{(i, j) : X_i \in A, X_j \in B\}|;$$

see Levit (1974) and Koshevnik and Levit (1976). The result translates to random fields with no interactions.

If $X_1, \ldots, X_n$ are observations of a homogeneous Markov chain with unknown transition distribution $Q(x, dy)$ and invariant distribution $\pi(dx)$, the the empirical estimator $E_n(A \times B)$ is efficient for $\pi \otimes Q(A \times B)$; see Greenwood and Wefelmeyer (1995). However, the empirical estimator

$$E_n(A \times B \times C) = \frac{1}{n-2} \left| \left\{ i : X_{i-2} \in A, X_{i-1} \in B, X_i \in C \right\} \right|$$

is not efficient for $\pi \otimes Q \otimes Q(A \times B \times C)$. If $A, B, C$ are one-point sets $\{a\}, \{b\}, \{c\}$, then $\pi \otimes Q \otimes Q(\{a\} \times \{b\} \times \{c\}) = \pi(\{a\})Q(a, \{b\})Q(b, \{c\})$, and an efficient estimator for this probability is

$$E_n(\{a\}) \frac{E_n(\{a\} \times \{b\})}{E_n(\{b\})} \frac{E_n(\{b\} \times \{c\})}{E_n(\{c\})}.$$

If we interpret the Markov chain as a one-dimensional random field, the sets $\{i - 1, i\}$ are the maximal cliques.

In this paper we address the question whether there is a similar way of improving the empirical estimator $E_n(a_D)$ for nearest neighbor random fields of dimension $d \geq 2$.

## 2 Construction of the estimator

Consider the $d$-dimensional square lattice $\mathbb{Z}^d$ and a countable state space $E$. We restrict attention to a stationary random field $\mu$ on $E^{\mathbb{Z}^d}$ with nearest neighbor interactions. The Manhattan norm on $\mathbb{Z}^d$ is defined by $|i| = \sum_{k=1}^d |i_k|$ for $i \in \mathbb{Z}^d$. Two points $i, j \in \mathbb{Z}^d$ are neighbors if $|i - j| = 1$. The neighborhood of $0$ is the sphere $\partial 0 = \{ i : |i| = 1 \}$. The random field is $\mu$ determined by its local characteristic at $0$, the conditional distribution $\mu(a_0|a_{-0}) = \mu(a_0|a_{00})$.

We observe the random field in a large window $[-n, n]^d$. We want to construct an estimator for the probability of a configuration on a finite subset of the lattice. The construction uses the following sets of sites: The sphere of radius $2^k$,

$$S_k = \{ i : |i| = 2^k \}.$$

The interior of $S_k$, the ball

$$B_k = \{ i : |i| < 2^k \}.$$
The union of the hyperplanes parallel to the faces of the sphere, intersected with \( B_k \), the ‘cross’

\[ C_k = \{ i \in B_k : i_1 + \cdots + i_r - i_{r+1} - \cdots - i_d = 0 \text{ for some } r = 1, \ldots, d \}. \]

The faces of the sphere and the hyperplanes are orthogonal to one of the \( d \) vectors \((1, \ldots, 1, -1, \ldots, -1)\) with \( r \) plus signs and \( d-r \) minus signs. It is convenient to consider first a configuration on the set

\[ D_m = B_m \cup S_m = \{ i : |i| \leq 2^k \}. \]

We also need the set of sites in \( D_m - (C_m \cup S_m) \) with Manhattan norm divisible by \( 2^k \) but not by \( 2^{k+1} \),

\[ D_{mk} = \{ i \in D_m - (C_m \cup S_m) : 2^k \text{ divides } |i|, 2^{k+1} \text{ does not divide } |i| \}. \]

Let \( a_{D_m} \) be a configuration on \( D_m \). We factor the probability \( \mu(a_{D_m}) \) using the splitting property of the nearest neighbor random field. Write

\[ \mu(a_{D_m}) = \mu(a_{S_m}) \mu(a_{B_m} | a_{S_m}). \tag{2.1} \]

The complement of the cross \( C_m \) in the ball \( B_m \) consists of \( 2^d \) disjoint balls \( B_{m-1} + i \) with centers \( i \in D_{m,m-1} \). The configurations \( a_{B_{m-1}+i}, i \in D_{m,m-1}, \) are conditionally independent given \( C_m \) and \( S_m \). Hence \( \mu(a_{B_m} | a_{S_m}) \) factors,

\[ \mu(a_{B_m} | a_{S_m}) = \mu(a_{C_m}) \mu(a_{B_m} | a_{S_m} a_{C_m}) = \mu(a_{C_m}) \prod_{i \in D_{m,m-1}} \mu(a_{B_{m-1}+i} | a_{S_{m-1}+i}). \tag{2.2} \]

Using (2.1) and (2.2), and factoring each conditional probability \( \mu(a_{B_{m-1}+i} | a_{S_{m-1}+i}) \) in the same way, we arrive at

\[ \mu(a_{D_m}) = \mu(a_{S_m}) \mu(a_{C_m} | a_{S_m}) \prod_{k=0}^{m-1} \prod_{i \in D_{mk}} \mu(a_{C_{k+i}+i} | a_{S_{k+i}}). \]

By definition of the conditional probabilities, this can be written

\[ \mu(a_{D_m}) = \mu(a_{C_m} a_{S_m}) \prod_{k=0}^{m-1} \prod_{i \in D_{mk}} \frac{\mu(a_{C_{k+i}+i} a_{S_{k+i}})}{\mu(a_{S_{k+i}})}. \tag{2.3} \]

It was convenient to describe the factorization from the highest order, the largest set, \( D_m \). One can, alternatively, begin with the lowest order, the one-point sets. In this case, the first step is

\[ \mu(a_{D_m}) = \mu(a_{D_m - D_{m0} a_{D_{m0}}}) = \mu(a_{D_m - D_{m0}}) \prod_{i \in D_{m0}} \mu(a_i | S_0 + i) = \mu(a_{D_m - D_{m0}}) \prod_{i \in D_{m0}} \mu(a_i | \partial \Omega). \]
Here $D_{m0}$ is the set of odd sites in $D_m$, and $D_m - D_{m0}$ the set of even sites. For parametric random fields, the coding estimator of Besag (1974a) and (1974b) is derived from the last product, the product of local characteristics at the odd sites. The top-down and bottom-up lattice partitions arise in the context of seismic models, see Newman, Gabriolov, Durand, Phoenix and Turcotte (1994) and Saleur, Sammis and Sornette (1996).

The empirical estimator for the probability of a configuration $a_D$ on a finite set $D \subset \mathbb{Z}^d$ is

$$E_n(a_D) = \frac{1}{|D_n|} \sum \{j : D + j \subset [-n, n]^d, X_{D+j} = a_D\},$$

where $|D_n|$ is the number of shifts of $D$ which are contained in $[-n, n]^d$. Instead of using $E_n(a_{D_m})$ to estimate $\mu(a_{D_m})$, we suggest estimating the probabilities in the factorization (2.3) by the corresponding empirical estimators,

$$T_n(a_D) = E_n(a_{C_m} a_{S_m}) \prod_{k=0}^{m-1} \prod_{i \in D_{mk}} E_n(a_{C_k+i} a_{S_k+i}) E_n(a_{S_k+i}).$$

(2.4)

Consider now an arbitrary finite set $D \subset \mathbb{Z}^d$ and a configuration $a_D$. Choose $m$ minimal so that $D$ is contained in (a shift of) $D_m$. We have

$$\mu(a_D) = \sum_{a_{D_m-D}a_D} \mu(a_{D_m-D}a_D).$$

From (2.4) we obtain an estimator for $\mu(a_D)$,

$$T_n(a_D) = \sum_{a_{D_m-D}a_D} T_n(a_{D_m-D}a_D).$$

(2.5)

To simplify the exposition, we have restricted attention to nearest neighbor random fields with respect to the Manhattan norm. The construction of the estimator (2.5) can be modified for other types of local interaction. For example, for the norm $|i| = \max\{i_k : k = 1, \ldots, d\}$, the sphere $\partial 0$ is a cube with faces parallel to the coordinate axes rather than to the vectors $(1, \ldots, 1, -1, \ldots, -1)$, and the factorization is analogous to (2.3).

References


Koshevnik, Y. A. and Levit, B. Y. (1976). On a non-parametric analogue of the infor-


Priscilla E. Greenwood
Department of Mathematics
University of British Columbia
Vancouver, B. C.
Canada V6T 1Z2
pgreenw@math.ubc.ca
http://www.math.ubc.ca/people/faculty/pgreenw/pgreenw.html

Ian W. McKeague
Department of Statistics
Florida State University
Tallahassee, Florida 32306- 4330
USA
mckeague@stat.fsu.edu
http://stat.fsu.edu/∼mckeague/

Wolfgang Wefelmeyer
Universität-GH Siegen
Fachbereich 6 Mathematik
Hölderlninstr. 3
57068 Siegen
Germany