# Characterizing efficient empirical estimators for local interaction Gibbs fields

Priscilla E. Greenwood
University of British Columbia

Wolfgang Wefelmeyer
Universität-GH Siegen

**Abstract**

The expectation of a local function on a stationary random field can be estimated from observations in a large window by the empirical estimator, i.e., the average of the function over all shifts within the window. Under appropriate conditions, the estimator is consistent and asymptotically normal. Suppose that the field is a Gibbs field with known finite range of interactions but otherwise unknown potential. We show that the empirical estimator is efficient if and only if the function is (equivalent to) a sum of functions each of which depends only on the values of the field on a clique of sites.

*AMS 1991 subject classifications.* 62F12, 62M40.

*Key words and Phrases.* Efficient estimator, empirical estimator, Gibbs field, local interactions, random field.

## 1 Introduction

Suppose we observe a configuration $x = (x_i)_{i \in \mathbf{Z}^d}$ of a stationary random field $X = (X_i)_{i \in \mathbf{Z}^d}$ on a window $[-n, n]^d$. Call a function $f_D$ on the random field *local* if it depends only on the values $x_D = (x_i)_{i \in D}$ at sites in a finite set $D \subset \mathbf{Z}^d$. We want to estimate the expectation of such a local function. The usual estimator is the *empirical estimator*

$$E_n f_D = |D_n|^{-1} \sum_{j: D+j \subset [-n,n]^d} f_D \circ \vartheta_j,$$

where $\vartheta_j(x)_i = x_{i+j}$ is the shift of $x$ by $-j$, and $|D_n|$ is the number of shifts of $D$ which are contained in the window $[-n, n]^d$.

[1]Work supported by NSERC, Canada, and the Crisis Points group of the Peter Wall Institute for Advanced Studies at UBC. Research done while the second author was visiting the Department of Mathematics at UBC.

Conditions for asymptotic normality are given in Bolthausen (1982) and Künsch (1982a), (1982b). If nothing is known about the random field, the estimator is efficient; see the Remark in Section 3. If the field has only *local* interactions, with known range, is the empirical estimator for the expectation of a local function still efficient? If it is not efficient for all local functions, can we characterize the functions for which it is?

To begin let us recall what is known for i.i.d. sequences of observations. Suppose $X_1, \dots, X_n$ are independent with unknown distribution. Let $f_D$ be a function which depends only on $X = (X_i)_{i \in D}$ with $D \subset \mathbf{Z}$ finite, say $D = \{1, \dots, s\}$, and suppose that $f_D$ cannot be written as a sum of functions $f_E$ with $E$ a strict subset of $D$. The *empirical estimator* for the expectation of $f_D$ is

$$\frac{1}{n-s+1} \sum_{j=0}^{n-s} f_D \circ \vartheta_j = \frac{1}{n-s+1} \sum_{j=1}^{n-s+1} f_D(X_j, \dots, X_{j+s-1}).$$

It is well-known that this estimator is efficient if and only if $D$ consists of a single site, e.g., $D = \{1\}$.

For $s > 1$, an efficient estimator is the *von Mises statistic*

$$\frac{1}{n^s} \sum_{i_1, \dots, i_s = 1}^{n} f_D(X_{i_1}, \dots, X_{i_s});$$

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

There is an analogous result for Markov chains. Let $X_0, \dots, X_n$ be observations of a stationary $k$-order Markov chain $(X_i)_{i \in \mathbf{Z}}$. Let $f_D$ be a function which depends only on the values $(X_i)_{i \in D}$ with $D \subset \mathbf{Z}$ finite, say $D = \{0, \dots, s\}$, and suppose that $f_D$ cannot be written as a sum of functions $f_E$ with $E$ a strict subset of $D$. The *empirical estimator* for the expectation of $f_D$ is

$$\frac{1}{n-s+1} \sum_{j=0}^{n-s} f_D \circ \vartheta_j = \frac{1}{n-s+1} \sum_{j=0}^{n-s} f_D(X_j, \dots, X_{j+s}).$$

It follows easily from Greenwood and Wefelmeyer (1995) that this estimator is efficient if and only if $D$ is a set of $k+1$ adjacent sites, e.g., $D = \{0, \dots, k\}$. These are the sites involved in the transition distribution $Q(X_0, \dots, X_{k-1}, dx_k)$ which 'parametrizes' the distribution of the chain. Similar results for Markov step processes and for semi-Markov processes are in Greenwood and Wefelmeyer (1994), (1996).

For $s > k$ and *discrete* state space, the expectation of $f_D$ can be written

$$\sum_{x_0, \dots, x_s} \pi(x_0, \dots, x_{k-1}) \left( \prod_{j=0}^{s-k} q(x_j, \dots, x_{j+k-1}; x_{j+k}) \right) f_D(x_{s-k-1}, \dots, x_s)$$
$$= \sum_{x_0, \dots, x_s} \pi(x_0, \dots, x_{k-1}) \left( \prod_{j=0}^{s-k} \frac{\pi(x_j, \dots, x_{j+k})}{\pi(x_j, \dots, x_{j+k-1})} \right) f_D(x_{s-k-1}, \dots, x_s),$$

where $q(x_0, \ldots, x_{k-1}; x_k)$ is the transition probability from $(x_0, \ldots, x_{k-1})$ to $x_k$, and $\pi(x_0, \ldots, x_{k-1})$ is the corresponding stationary probability of $(x_0, \ldots, x_{k-1})$. Now an efficient estimator is obtained by replacing the probabilities by empirical estimators. For $s > k$ and *continuous* state space, the construction of an efficient estimator is considerably more involved; see Schick and Wefelmeyer (1998).

For a stationary random field, a role similar to the transition distribution is played by its *local characteristic* say at site 0, i.e., the conditional distribution of $X_0$ given the configuration on the other sites. If the field has local interactions, the local characteristic conditions only on a finite set of sites $C \subset \mathbf{Z}^d \backslash \{0\}$, the *neighborhood* of 0. A first guess might be that the empirical estimator for the expectation of $f_D$ is efficient if and only if $f_D$ is a sum of functions which depend only on shifts of $C \cup \{0\}$.

We will see that this guess is wrong. Indeed, the analogy with Markov chains is misleading. The local characteristic is not a freely varying parameter like the transition distribution. The transition distribution is an arbitrary Markov kernel, whereas the local characteristic must satisfy certain consistency conditions. These additional restrictions contain information which is not used by the empirical estimator and makes it inefficient, in general.

The restrictions can be seen by looking at the simplest case, a *one*-dimensional field in which the neighborhood of site 0 consists of the two adjacent sites, $C = \{-1, 1\}$. The local characteristic at 0 involves the three sites $-1$, 0, 1. But the field can also be viewed as a Markov chain and is then described by the transition distribution $Q(X_0, dx_1)$ which involves only the two sites 0, 1. The local characteristic has an expression in terms of $Q$, which constitutes a restriction; see Georgii (1988, Theorems 3.5 and 10.25). Indeed, as mentioned above, the empirical estimator is efficient if and only if $f_D$ is a sum of functions which depend only on *two* adjacent sites.

For higher-dimensional random fields, there is in general no description by simpler objects than the local characteristics, except when the latter are strictly positive. In that case, the field can be written as a Gibbs field and is described by the potential. Notice that the assumption of positivity does not change the efficiency question, which depends only on the structure of the model in a neighborhood of the distribution governing the observations.

Suppose that the field is stationary and has only local interactions (with known range), i.e., the local characteristic at 0 depends on a finite set $C$ of sites, the *neighbors* of 0. A *clique* is a set of sites such that each two points are neighbors of each other. The potential involves only functions which depend on cliques. This suggests that the sets which play the role of singletons in the i.i.d. case and of pairs of adjacent sites in the first-order Markov chain case are, in the case of Gibbs fields, the cliques. Suppose, for example, that the lattice is two-dimensional, and that the local characteristic at a site depends only on the four adjacent sites. Then the local characteristic involves five sites, whereas the cliques are the one-point sets and the pairs of adjacent sites. We show in Section 3 that indeed the cliques are the appropriate sets: The empirical estimator for the expectation of a local function is efficient if and only if the function is (equivalent to) a sum of functions each of which depends only on the values of the field on a clique of sites.

Improved estimators for other functions in the discrete state space case are constructed in Greenwood, McKeague and Wefelmeyer (1998).

We do not prove the result under minimal conditions. It remains true as long as the likelihood admits a stochastic expansion of the form (3.3), and the empirical estimators for expectations of local functions are asymptotically normal with variance of the form (3.6).

## 2 The model

In this section we introduce our model, the family of all stationary Gibbs measures on $E^{\mathbf{Z}^d}$ with general state space $E$, and with given range of interactions. We begin with notation for general Gibbs measures.

We consider the $d$-dimensional square lattice $S = \mathbf{Z}^d$. To each site we attach the *state space* $E$ which we take to be a complete separable metric space with Borel field $\mathcal{E}$. The corresponding *configuration space* is $(\Omega, \mathcal{F}) = (E^S, \mathcal{E}^S)$.

For $V \subset S$ let $\mathcal{F}_V \subset \mathcal{F}$ denote the $\sigma$-field generated by the projections $x \to x_i$, for each $i \in V$. If $S$ is partitioned into $V$, $W$, and $y \in E^V$, $z \in E^W$, then $yz$ denotes the configuration $x$ such that $x_i = y_i$ for $i \in V$ and $x_i = z_i$ for $i \in W$. In particular, $x = x_V x_{E\setminus V}$.

Call a function $f$ on $\Omega$ *local* if $f$ is $\mathcal{F}_V$-measurable for some finite set $V \subset S$. Let $\| \, \|$ denote the sup-norm of functions on $\Omega$. Call a function *quasilocal* if there is a sequence of local functions $f_n$ such that $\|f_n - f\| \to 0$ for $n \to \infty$.

A *specification* is a family $\gamma$ of Markov kernels $\gamma_V$ from $\mathcal{F}_{E\setminus V}$ to $\mathcal{F}$, with $V \subset S$ finite, which satisfies the *consistency condition*

$$\gamma_V \gamma_W = \gamma_V \quad \text{when } W \subset V.$$

A random field $\mu$ on $\mathcal{E}$ is *specified* by $\gamma$ if

$$\mu_V(F|\cdot) := \mu(F|\mathcal{F}_{E\setminus V}) = \gamma_V(F|\cdot) \quad \text{almost surely for all finite } V \subset S \text{ and } F \in \mathcal{F}.$$

For the *projection* of $\gamma_V$ we write

$$\pi_V(F|x) = \gamma_V(x_V \in F|x).$$

*Dobrushin's interdependence matrix* is

$$C_{ij} = \sup\{\|\pi_i(\cdot|x) - \pi_i(\cdot|y)\| : x_{E\setminus\{j\}} = y_{E\setminus\{j\}}\},$$

where $\|\eta\| = \sup_{A\in\mathcal{F}} \|\eta(A)\|$ is the sup-norm of a signed measure $\eta$ on $\mathcal{F}$.

A specification $\gamma$ is *quasilocal* if $\gamma_V(f|\cdot)$ is quasilocal for all finite $V$ and quasilocal $f$. A specification satisfies *Dobrushin's condition* if it is quasilocal and

$$\sup_{i\in S}\sum_{j\in S} C_{ij} < 1. \tag{2.1}$$

Then, by Dobrushin's uniqueness theorem, Georgii (1988, Theorem 8.7), there is exactly one random field specified by $\gamma$.

Here we restrict attention to stationary Gibbs fields. For $i \in S$ let $\vartheta_i$ denote the *shift* by $-i$, defined by $\vartheta_i(x)_j = x_{j+i}$, $j \in S$. Let $\lambda$ be a finite measure on $\mathcal{E}$, and denote the product measure on $\Omega = \mathcal{E}^S$ by $\lambda^S$. Let $\mathcal{A}$ be a finite collection of finite sets $A \subset S$. We assume without loss of generality that no set in $\mathcal{A}$ is a shift of one of the other sets. This is done so that each set represents a different equivalence class of shifts. Consider a collection $u = (u_A)_{A\in\mathcal{A}}$ of $\mathcal{F}_A$-measurable functions $u_A$ on $\Omega$ such that

$$\sum_{A\in\mathcal{A}} |A|^2\|u_A\| < 1. \tag{2.2}$$

Each $u$ generates a shift invariant *potential*

$$U^u = \{u_A \circ \vartheta_i : A \in \mathcal{A}, i \in S\}. \tag{2.3}$$

The corresponding *Hamiltonian* is

$$H_V^u = \sum_{\substack{A\in\mathcal{A}, i\in S\\ (A+i)\cap V\neq\emptyset}} u_A \circ \vartheta_i.$$

For $A \in \mathcal{A}$ and $i \in S$, the sets $A + i$ and their subsets are called *cliques*; see, e.g., Winkler (1995, p. 49). The potential is $\lambda$-*admissible*, i.e.,

$$Z_V^u(x) = \int \lambda^V(dy)e^{-H_V^u(yx_{E\setminus V})} < \infty \quad \text{for all finite } V \subset S \text{ and all } x \in \Omega.$$

The function $Z_V^u$ is the *partition function* in $V$, and the *Gibbs distribution* in $V$ is

$$\gamma_V^u(F|x) = \frac{1}{Z_V^u(x)}\int_F \lambda^V(dy)e^{-H_V^u(yx_{E\setminus V})}, \quad F \in \mathcal{F}.$$

The *Gibbs specification* corresponding to $u$ is

$$\gamma^u = \{\gamma_V^u : V \subset S, V \text{ finite}\}.$$

Any random field specified by $\gamma^u$ is called a *homogeneous Gibbs field.*

The *oscillation* of a function $f$ on $\Omega$ is

$$\rho(f) = \sup\{|f(x) - f(y)| : x, y \in \Omega\}.$$

Assumption (2.2) implies

$$\sup_{j\in S}\sum_{\substack{A\in\mathcal{A},i\in S\\ j\in A+i}}(|A|-1)\rho(u_A)=\sum_{A\in\mathcal{A}}|A|(|A|-1)\rho(u_A)<2.$$

Hence Dobrushin's condition (2.1) holds by Georgii (1988, Proposition 8.8), and by Dobrushin's uniqueness theorem there is exactly one Gibbs measure $\mu^u$ corresponding to $\gamma^u$.

Our model is the family of Gibbs measures $\mu^u$ determined by all possible collections $u=(u_A)_{A\in\mathcal{A}}$ of functions on $\Omega$ satisfying condition (2.2). Usually one describes the Gibbs measure by the shift invariant potential $U^u$ generated by $u$, see (2.3). Here we look at a collection of representatives, $u$, because it leads to a more convenient parametrization. Since the same Gibbs measure $\mu^u$ can result from several collections $u$, we parametrize by *equivalence classes*. Two collections $u$ and $v$ are equivalent, $u\sim v$, if

$$H_V^{u-v} \text{ is } \mathcal{F}_{E\setminus V}\text{-measurable for all finite } V\subset S,$$

see Georgii (1988, Theorem 2.34).

The linear structure of the collections $u$ is compatible with the equivalence relation: if $u\sim u'$ and $v\sim v'$, then $u+u'\sim v+v'$. This justifies working with representatives of equivalence classes rather than with the classes themselves.

# 3 The result

In this section we characterize the efficient estimators among the empirical estimators. Our efficiency concept is based on a nonparametric version of Hájek's (1970) convolution theorem for the asymptotic distribution of regular estimators. The theorem requires the model to be asymptotically normal. Hence our next step is to introduce a local parameter space, the linear space $K$ of all (equivalence classes) of collections $k=(k_A)_{A\in\mathcal{A}}$ of bounded $\mathcal{F}_A$-measurable functions $k_A$ on $\Omega$. Fix a parameter $u=(u_A)_{A\in\mathcal{A}}$ fulfilling (2.2). The perturbed parameter $u+tk$ also fulfills (2.2) for sufficiently small $t$: There is a $c>0$ such that

$$\sup_{|t|\le c}\sum_{A\in\mathcal{A}}|A|^2\|(u+tk)_A\|<1. \tag{3.1}$$

The proof of local asymptotic normality is based on the following Proposition 1, a perturbation expansion of Künsch (1982a, Proposition 5.3) for expectations of quasilocal functions, which generalizes a result of Gross (1981); see also Georgii (1988, Corollary 8.37). We use only the version for *stationary* random fields. Let

$$\rho_i(f)=\sup\{|f(x)-f(y)|:x_{E\setminus\{i\}}=y_{E\setminus\{i\}}\}.$$

**Proposition 1**. *For every $k \in K$ and every quasilocal $f$ with $\sum_{i \in S} \rho_i(f) < \infty$, the expectation $\mu^{u+tk}(f)$ is continuously differentiable at $t = 0$, and*

$$\frac{d}{dt}\mu^{u+tk}(f)|_{t=0} = -\mu^u(T^u Rk \cdot f)$$

*with*

$$\begin{aligned} Rk &= \sum_{A \in \mathcal{A}} \sum_{i \in A} \frac{1}{|A|} k_A \circ \vartheta_i, \\ T^u r &= \sum_{j \in S} (r \circ \vartheta_j - \mu^u(r)) \quad \textit{for local } r, \end{aligned}$$

*i.e.,*

$$T^u Rk = \sum_{j \in S} \sum_{A \in \mathcal{A}} (k_A \circ \vartheta_j - \mu^u(k_A)).$$

Künsch's proof of Proposition 1 starts from the corresponding result for *conditional* expectations,

$$\frac{d}{dt}\mu^{u+tk}(f|\mathcal{F}_V)|_{t=0} = -\sum_{j \in S} \mu^u \left( (Rk \circ \vartheta_j - \mu^u(Rk \circ \vartheta_j|\mathcal{F}_V)) f | \mathcal{F}_V \right), \tag{3.2}$$

which we will also use. A representation of the likelihood ratio $d\mu_V^{u+ck}/d\mu_V^u$ of the field in the window $V$ can be derived from Proposition 1.

**Proposition 2**. *For $k \in K$, and $c > 0$ small enough,*

$$\log \frac{d\mu_V^{u+ck}}{d\mu_V^u} = -c \int_0^1 \mu^{u+tck}(T^{u+tck} Rk|\mathcal{F}_V) dt.$$

The proof is in Section 4. We can now prove local asymptotic normality. Our proof is similar to that for a *parametric* Gibbs field model in Janžura (1989). Let $V_n = [-n, n]^d$ and $c_n = (2n+1)^d$.

**Proposition 3**. *For $k \in K$,*

$$\begin{aligned} \log \frac{d\mu_{V_n}^{u - c_n^{-1/2} k}}{d\mu_{V_n}^u} &= c_n^{-1/2} \sum_{j \in V_n} (Rk \circ \vartheta_j - \mu^u(Rk)) \\ &\quad - \frac{1}{2} \sum_{j \in S} \mathrm{cov}^u(Rk \circ \vartheta_j, Rk) + o_{\mu^u}(1), \end{aligned} \tag{3.3}$$

*and the stochastic term on the right is asymptotically normal under $\mu^u$ with mean zero and variance $\sum_{j\in S}\operatorname{cov}^u(Rk\circ\vartheta_j, Rk)$. The covariance is taken with respect to $\mu^u$.*

The proof is in Section 4. Our approximation to the likelihood is measurable with respect to a $\sigma$-field which is larger than $\mathcal{F}_{V_n}$. This is for notational convenience. An asymptotically equivalent $\mathcal{F}_{V_n}$-measurable approximation could be obtained by extending the sum over a smaller window than $V_n$.

Local asymptotic normality induces an inner product on $K$; for $k, k' \in K$,

$$(k, k') = \sum_{j\in S}\operatorname{cov}^u(Rk\circ\vartheta_j, Rk').$$

Since the field is stationary, the inner product simplifies,

$$\begin{aligned}(k, k') &= \sum_{j\in S}\sum_{A\in\mathcal{A}}\sum_{A'\in\mathcal{A}}\frac{1}{|A||A'|}\sum_{i\in A}\sum_{i'\in A'}\operatorname{cov}^u(k_A\circ\vartheta_i\circ\vartheta_j, k'_{A'}\circ\vartheta_{i'}) \\ &= \sum_{j\in S}\operatorname{cov}^u\left(\sum_{A\in\mathcal{A}}k_A\circ\vartheta_j, \sum_{A\in\mathcal{A}}k'_A\right). \end{aligned} \tag{3.4}$$

This means that the local parameter $k$ enters the inner product through (the equivalence class of shifts of) $\sum_{A\in\mathcal{A}}(k_A - \mu^u(k_A))$. Write

$$K_0 = \{\sum_{A\in\mathcal{A}}(k_A - \mu^u(k_A)) : k \in K\}.$$

Consider a bounded local function, say a bounded $\mathcal{F}_D$-measurable function $f_D$, with $D \subset S$ finite. The *empirical estimator* for the expectation $\mu^u(f_D)$ is

$$E_n f_D = |D_n|^{-1}\sum_{j\in D_n} f_D\circ\vartheta_j, \tag{3.5}$$

where $D_n = \{j : D + j \subset V_n = [-n, n]^d\}$ and $|D_n|$ is the number of sites in $D_n$. By a variant of the central limit theorem of Künsch (1982a), (1982b), applied for the local function $f_D$, as opposed to a continuous function, see also Bolthausen (1982), we have

$$\begin{aligned}|D_n|^{1/2}(E_n f_D - \mu^u(f_D)) &= |D_n|^{-1/2}\sum_{j\in D_n}(f_D\circ\vartheta_j - \mu^u(f_D)) \\ &\Rightarrow \left(\sum_{j\in S}\operatorname{cov}^u(f_D\circ\vartheta_j, f_D)\right)^{1/2} N, \end{aligned} \tag{3.6}$$

where $N$ is a standard normal random variable.

Consider $\mu^u(f_D)$ as a functional of $u$. By Proposition 1,

$$c_n^{1/2}(\mu^{u-c_n^{-1/2}k}(f_D) - \mu^u(f_D)) \to \mu^u(T^u Rk\cdot f_D) = \sum_{j\in S}\operatorname{cov}^u(Rk\circ\vartheta_j, f_D). \tag{3.7}$$

The right-hand side is an inner product of $k$ and $f_D$. We write it as the inner product (3.4) induced by local asymptotic normality,

$$\sum_{j\in S} \operatorname{cov}^u(Rk \circ \vartheta_j, f_D) = \sum_{j\in S} \operatorname{cov}^u \left( \sum_{A\in\mathcal{A}} k_A \circ \vartheta_j, f_D \right). \tag{3.8}$$

It depends on $f_D$ only through (the equivalence class of shifts of) $f_D - \mu^u(f_D)$. We call $f_D - \mu^u(f_D)$ a *gradient* of $\mu^u(f_D)$. The *canonical gradient* $g_D$ is the projection of $f_D - \mu^u(f_D)$ into $K_0$.

Call an estimator $T_n$ for $\mu^u(f_D)$ *regular* with *limit* $L$ if

$$c_n^{1/2}(T_n - \mu^{u-c_n^{-1/2}k}(f_D)) \Rightarrow L \quad \text{under } \mu^{u-c_n^{-1/2}k} \text{ for all } k \in K.$$

Regularity is a weak form of continuous convergence: For each continuous and bounded function $h$ and each $k \in K$,

$$E_{u-c_n^{-1/2}k} h(T_n - \mu^{u-c_n^{-1/2}k}(f_D)) \to Eh(L).$$

For the convolution theorem of Hájek (1970), see now Bickel, Klaassen, Ritov and Wellner (1993, p. 63, Theorem 2), regularity must be assumed in order to exclude *superefficient* estimators. The convolution says that if $T_n$ is regular with limit $L$, then

$$L = \left( \sum_{j\in S} \operatorname{cov}^u(g_D \circ \vartheta_j, g_D) \right)^{1/2} N + M \quad \text{in distribution},$$

with $M$ independent of $N$. If $M$ is independent of $N$, we have

$$P(|cN + M| < t) \le P(|cN| < t) \quad \text{for all } t,$$

i.e., $cN$ is more concentrated in symmetric intervals than $cN + M$. This justifies calling an estimator $T_n$ *efficient* if

$$c_n^{1/2}(T_n - \mu^u(f_D)) \Rightarrow \left( \sum_{j\in S} \operatorname{cov}^u(g_D \circ \vartheta_j, g_D) \right)^{1/2} N \quad \text{under } \mu^u.$$

Call an estimator $T_n$ for $\mu^u(f_D)$ *asymptotically linear* with *influence function* $h$ if $h$ is bounded and local with $\mu^u(h) = 0$, and

$$c_n^{1/2}(T_n - \mu^u(f_D)) = c_n^{-1/2} \sum_{j\in V_n} h \circ \vartheta_j + o_{\mu^u}(1).$$

By Proposition 3, the model is locally asymptotically normal, and by relations (3.7) and (3.8), the functional $\mu^u(f_D)$ is differentiable with gradient $f_D - \mu^u(f_D)$. Therefore, we have the following characterizations; see Bickel et al. (1993, p. 63).

**Proposition 4**. *An asymptotically linear estimator $T_n$ is regular for $\mu^u(f_D)$ if and only if its influence function is a gradient of $\mu^u(f_D)$.*

*An estimator $T_n$ is regular and efficient for $\mu^u(f_D)$ if and only if it is asymptotically linear with influence function equal to the canonical gradient,*

$$c_n^{1/2}(T_n - \mu^u(f_D)) = c_n^{-1/2} \sum_{j \in V_n} g_D \circ \vartheta_j + o_{\mu^u}(1). \tag{3.9}$$

Suppose that $f_D$ is a sum of functions $f_{D_1} + f_{D_2}$ with $D_2 + j \subset D_1$ for some $j$, say. Then it is preferable to use the empirical estimator $E_n \overline{f}_{D_1}$, where $\overline{f}_{D_1} = f_{D_1} + f_{D_2} \circ \vartheta_j$ is $\mathcal{F}_{D_1}$-measurable. The reason is that $|D_{1n}| > |D_n|$ unless $D_2 \subset D_1$. Hence $E_n \overline{f}_{D_1}$ is an average over more shifts than $E_n f_D$. Asymptotically, however, the estimators $E_n f_D$ and $E_n \overline{f}_{D_1}$ are equivalent. More generally, any local function admits a *minimal* representation by a constant or by a function

$$f_D = \sum_{r=1}^{m} f_{D_r}, \tag{3.10}$$

where $f_{D_r}$ is $\mathcal{F}_{D_r}$-measurable, the sets $D_r$ are finite and non-empty, none of them is a shift of one of the other sets, and none of the functions $f_{D_r}$ admits a non-trivial representation of the form (3.10).

If two local functions have the same minimal representation, the corresponding empirical estimators are asymptotically equivalent.

**Theorem.** *Let $f_D$ be bounded and $\mathcal{F}_D$-measurable, with $D \subset S$ finite. Then the empirical estimator $E_n f_D$ defined in (3.5) is regular for $\mu^u(f_D)$, and it is efficient if and only if $f_D$ has a minimal representation (3.10) which is a sum of functions each of which is $\mathcal{F}_C$-measurable, with $C$ a clique.*

The proof is in Section 4.

**Remark.** It follows from the Theorem that the empirical estimator $E_n f_D$ is always efficient in the model consisting of local interaction Gibbs fields of *arbitrary* range. In this model, any finite set $D \subset S$ is a clique. In fact, the empirical estimator will also be efficient in the larger model consisting of *all* Gibbs fields, provided that we have local asymptotic normality and a central limit theorem for empirical estimators under the field which generates the observations.

# 4 Proofs

**Proof of Proposition 2.** For $\mathcal{F}_V$-measurable $f$ and $s$ sufficiently small,

$$\mu^{u+sk}\left(\frac{d}{ds}\log\frac{d\mu_V^{u+sk}}{d\mu_V^u}\cdot f\right) = \mu^u\left(\frac{d\mu_V^{u+sk}}{d\mu_V^u}\frac{\frac{d}{ds}\frac{d\mu_V^{u+sk}}{d\mu_V^u}}{\frac{d\mu_V^{u+sk}}{d\mu_V^u}}\cdot f\right)$$

$$= \mu^u\left(\frac{d}{ds}\frac{d\mu_V^{u+sk}}{d\mu_V^u}\cdot f\right) = \frac{d}{ds}\mu^u\left(\frac{d\mu_V^{u+sk}}{d\mu_V^u}\cdot f\right)$$

$$= \frac{d}{ds}\mu^{u+sk}(f) = -\mu^{u+sk}(T^{u+sk}Rk\cdot f). \tag{4.1}$$

The last equality follows from Proposition 1, applied with $u+sk$ in place of $u$. It remains to check that the interchange of the derivative with the integral is justified. It is easy to check the following uniform version of Proposition 1. Uniformly for uniformly bounded $\mathcal{F}_V$-measurable $f$,

$$\mu_V^u\left(\left(\frac{d\mu_V^{u+tk}}{d\mu_V^u}-\frac{d\mu_V^{u+sk}}{d\mu_V^u}\right)f\right) = \mu_V^{u+tk}(f)-\mu_V^{u+sk}(f) = O(t-s).$$

This implies

$$\frac{d\mu_V^{u+tk}}{d\mu_V^u}(x_V)-\frac{d\mu_V^{u+sk}}{d\mu_V^u}(x_V) = O(t-s) \quad \text{uniformly for } \mu_V^u\text{-a.a. } x_V.$$

Relation (4.1) implies

$$\frac{d}{ds}\log\frac{d\mu_V^{u+sk}}{d\mu_V^u} = -\mu_V^{u+sk}(T^{u+sk}Rk|\mathcal{F}_V) \quad \mu^u\text{-a.s.}$$

Hence

$$\log\frac{d\mu_V^{u+ck}}{d\mu_V^u} = \int_0^1\frac{d}{ds}\frac{d\mu_V^{u+tck}}{d\mu_V^u}dt = -\int_0^1\mu^{u+tck}(T^{u+tck}Rk|\mathcal{F}_V)dt.$$

**Proof of Proposition 3.** (i) By Proposition 2,

$$\log\frac{d\mu_{V_n}^{u-c_n^{-1/2}k}}{d\mu_{V_n}^u} = c_n^{-1/2}\int_0^1\mu^{u-tc_n^{-1/2}k}(T^{u-tc_n^{-1/2}k}Rk|\mathcal{F}_{V_n})dt$$

$$= c_n^{-1/2}\int_0^1\sum_{j\in S}\mu^{u-tc_n^{-1/2}k}(Rk\circ\vartheta_j-\mu^{u-tc_n^{-1/2}k}(Rk)|\mathcal{F}_{V_n})dt$$

$$= c_n^{-1/2}\sum_{j\in V_n}(Rk\circ\vartheta_j-\mu^u(Rk))-c_n^{1/2}\int_0^1(\mu^{u-tc_n^{-1/2}k}(Rk)-\mu^u(Rk))dt$$

$$-\int_0^1 H^{u-tc_n^{-1/2}k}dt+\int_0^1 J^{u-tc_n^{-1/2}k}dt \tag{4.2}$$

with

$$H^u = c_n^{-1/2} \sum_{j \in V_n} (Rk \circ \vartheta_j - \mu^u(Rk \circ \vartheta_j | \mathcal{F}_{V_n})),$$
$$J^u = c_n^{-1/2} \sum_{j \notin V_n} (\mu^u(Rk \circ \vartheta_j | \mathcal{F}_{V_n}) - \mu^u(Rk)).$$

The first summand in (4.2) is the stochastic term of the expansion of the likelihood. The second summand contributes the deterministic term: By Proposition 1, applied with $u - sc_n^{-1/2}k$ in place of $u$,

$$\begin{aligned} & c_n^{1/2} \int_0^1 (\mu^{u-tc_n^{-1/2}k}(Rk) - \mu^u(Rk))dt \\ &= c_n^{1/2} \int_0^1 \int_0^t \frac{d}{ds} \mu^{u-sc_n^{-1/2}k}(Rk) ds dt \\ &= \int_0^1 \int_0^t \mu^{u-sc_n^{-1/2}k}(T^{u-sc_n^{-1/2}k} Rk \cdot Rk) ds dt. \end{aligned} \tag{4.3}$$

By Proposition 1, the integrand is the derivative of $\mu^{u+tk}$ at $t = -sc_n^{-1/2}$, and is continuous at $t = 0$. Hence it converges uniformly for $0 \leq s \leq 1$ to $\mu^u(T^u Rk \cdot Rk)$. Therefore, (4.3) converges to

$$\begin{aligned} \frac{1}{2}\mu^u(T^u Rk \cdot Rk) &= \frac{1}{2} \sum_{j \in S} \mu^u\big((Rk \circ \vartheta_j - \mu^u(Rk))(Rk - \mu^u(Rk))\big) \\ &= \frac{1}{2} \sum_{j \in S} \mathrm{cov}^u(Rk \circ \vartheta_j, Rk). \end{aligned}$$

It remains to show that the last two terms in (4.2) are negligible.

(ii) We show that $\int_0^1 H^{u-tc_n^{-1/2}k} dt$ is of order $o_{\mu^u}(1)$. By the dominated convergence theorem, it suffices to show that $H^{u-tc_n^{-1/2}k}$ is of order $o_{\mu^u}(1)$ for $0 \leq t \leq 1$. Write $H^{u-tc_n^{-1/2}k} = H_1 + H_2$ with

$$H_1 = c_n^{-1/2} \sum_{j \in V_n} \big(Rk \circ \vartheta_j - \mu^u(Rk \circ \vartheta_j | \mathcal{F}_{V_n})\big),$$
$$H_2 = c_n^{-1/2} \sum_{j \in V_n} \big(\mu^u(Rk \circ \vartheta_j | \mathcal{F}_{V_n}) - \mu^{u-tc_n^{-1/2}k}(Rk \circ \vartheta_j | \mathcal{F}_{V_n})\big).$$

By the intermediate value theorem and (3.2), for some $s$ between 0 and $t$,

$$\begin{aligned} H_2 &= -c_n^{-1/2} \sum_{j \in V_n} \frac{d}{ds} \mu^{u-sc_n^{-1/2}k}(Rk \circ \vartheta_j | \mathcal{F}_{V_n}) \\ &= c_n^{-1} t \sum_{j \in V_n} \sum_{i \in S} \mathrm{cov}^{u-sc_n^{-1/2}k}(Rk \circ \vartheta_i, Rk \circ \vartheta_j | \mathcal{F}_{V_n}). \end{aligned}$$

Here (3.2) is applied at $u = u - sc_n^{-1/2}k$ and for $f = Rk \circ \vartheta_j$. Because of assumption (3.1), condition (2.1) holds uniformly for $s \in [0,1]$. We may therefore, without loss of generality, restrict attention to $s = 0$, i.e., $u$ in place of $u - sc_n^{-1/2}k$.

A bound for the covariances will be described in terms of the following notation:

$$C = (C_{ij})_{i,j\in S}; \quad C_V = (C_{ij})_{i,j\in V};$$
$$\chi_{ij}^V = \textstyle\sum_{n=0}^{\infty}(C_V^n)_{ij} \text{ for } i,j \in V; \quad \chi_{ij} = \chi_{ij}^S.$$

Clearly, $\chi_{ij}^V \le \chi_{ij}$. For a local function $f$, we have $\rho_i(f \circ \vartheta_j) = 0$ for $i, j$ sufficiently far apart. Applying Künsch (1982a, Corollary 3.3) for $\mu^{u-sc_n^{-1/2}k}(\cdot|\mathcal{F}_{V_n})$ and with semimetric identically equal to zero, we obtain $c, c' > 0$ such that

$$\begin{aligned} |H_2| &\le cc_n^{-1}t \sum_{j\in V_n}\sum_{i\in S}\sum_{a\in S}\rho_a(Rk\circ\vartheta_i)\sum_{b\in S}\rho_b(Rk\circ\vartheta_j) \\ &\le cc'c_n^{-1}t\left(\sum_{i\in S}\rho_i(Rk)\right)^2 \to 0. \end{aligned}$$

Since $H_1$ is conditionally centered, it is easy to compute its variance:

$$\mu^u(H_1^2) = \mu^u\left(c_n^{-1}\sum_{i,j\in V_n}\mathrm{cov}^u(Rk\circ\vartheta_i, Rk\circ\vartheta_j|\mathcal{F}_{V_n})\right).$$

The integrand tends to zero as $H_2$. By the dominated convergence theorem, $\mu^u(H_1^2) \to 0$, and therefore $H_1 = o_{\mu^u}(1)$.

(iii) We show that $\int_0^1 J^{u-tc_n^{-1/2}k}dt$ is of order $o_{\mu^u}(1)$. By the dominated convergence theorem, it suffices to show that $J^{u-tc_n^{-1/2}k}$ is of order $o_{\mu^u}(1)$ for $0 \le t \le 1$. Write $J^{u-tc_n^{-1/2}k} = J_1 + J_2$ with

$$\begin{aligned} J_1 &= c_n^{-1/2}\sum_{j\notin V_n}\left(\mu^u(\mu^{u-tc_n^{-1/2}k}(Rk\circ\vartheta_j|\mathcal{F}_{V_n})) - \mu^{u-tc_n^{-1/2}k}(Rk)\right), \\ J_2 &= c_n^{-1/2}\sum_{j\notin V_n}\left(\mu^{u-tc_n^{-1/2}k}(Rk\circ\vartheta_j|\mathcal{F}_{V_n}) - \mu^u(\mu^{u-tc_n^{-1/2}k}(Rk\circ\vartheta_j|\mathcal{F}_{V_n}))\right). \end{aligned}$$

By Proposition 1,

$$\begin{aligned} J_1 &= \int_0^t c_n^{-1/2}\sum_{j\notin V_n}\frac{d}{ds}\mu^{u-sc_n^{-1/2}k}(\mu^{u-tc_n^{-1/2}k}(Rk\circ\vartheta_j|\mathcal{F}_{V_n}))ds \\ &= \int_0^t c_n^{-1}\sum_{j\notin V_n}\mu^{u-sc_n^{-1/2}k}\big(T^{u-sc_n^{-1/2}k}Rk\cdot\mu^{u-tc_n^{-1/2}k}(Rk\circ\vartheta_j|\mathcal{F}_{V_n})\big)ds. \end{aligned}$$

By the dominated convergence theorem, to show that $J_1$ tends to zero, it suffices to show that the right-hand integrand tends to zero. Applying Künsch (1982a, Corollary 3.3) for $\mu^{u-sc_n^{-1/2}k}(\cdot|\mathcal{F}_{V_n})$ and semimetric identically equal to zero, we obtain $c, c' > 0$ such that

$$\begin{aligned}
& c_n^{-1} \sum_{j \notin V_n} \mu^{u-sc_n^{-1/2}k}\big(T^{u-sc_n^{-1/2}k} Rk \cdot \mu^{u-tc_n^{-1/2}k}(Rk \circ \vartheta_j | \mathcal{F}_{V_n})\big) \\
&= c_n^{-1} \sum_{j \notin V_n} \sum_{i \in S} \operatorname{cov}^{u-sc_n^{-1/2}k}(Rk \circ \vartheta_i, \mu^{u-tc_n^{-1/2}k}(Rk \circ \vartheta_j | \mathcal{F}_{V_n})) \\
&\leq c c_n^{-1} \sum_{j \notin V_n} \sum_{i \in S} \sum_{a \in S} \rho_a(Rk \circ \vartheta_i) \sum_{b \in S} \rho_b(\mu^{u-tc_n^{-1/2}k}(Rk \circ \vartheta_j | \mathcal{F}_{V_n})) \\
&\leq c c' c_n^{-1} \sum_{i \in S} \rho_i(Rk) \sum_{j \notin V_n} \sum_{b \in S} \rho_b(\mu^{u-tc_n^{-1/2}k}(Rk \circ \vartheta_j | \mathcal{F}_{V_n})). \qquad (4.4)
\end{aligned}$$

By the argument in Künsch (1982a, proof of Corollary 2.4),

$$\begin{aligned}
\rho_b(\mu(f|\mathcal{F}_V)) &= \sup\{|\mu_V(f|x) - \mu_V(f|y)| : x_{E\setminus\{b\}} = y_{E\setminus\{b\}}\} \\
&\leq \sum_{r,s \notin V} \gamma_{br} \chi_{rs}^{E\setminus V} \rho_s(f) + \rho_b(f).
\end{aligned}$$

Hence, writing $u$ for $u - tc_n^{-1/2}k$, and using again $\rho_i(f \circ \vartheta_j) = 0$ for $i, j$ sufficiently far apart,

$$\begin{aligned}
& c_n^{-1} \sum_{j \notin V_n} \sum_{b \in S} \rho_b(\mu^u(Rk \circ \vartheta_j | \mathcal{F}_{V_n})) \\
&\leq c_n^{-1} \sum_{r,s \notin V_n} \left( \sum_{b \in S} C_{br} \right) \chi_{rs}^{-V_n} \sum_{j \notin V_n} \rho_s(Rk \circ \vartheta_j) \;+\; \sum_{j \notin V_n} \sum_{b \in S} \rho_b(Rk \circ \vartheta_j) \\
&\leq c c_n^{-1} \sum_{s \in S} \rho_s(Rk) \sum_{r \in S} \chi_{rs} \;+\; c c_n^{-1} \sum_{b \in S} \rho_b(Rk) \to 0. \qquad (4.5)
\end{aligned}$$

Inserting (4.5) into (4.4), we see that the integrand of $J_1$ converges pointwise to zero, and hence $J_1$ converges to zero.

The variance of $J_2$ is

$$\mu^u(J_2^2) = c_n^{-1} \sum_{i,j \notin V_n} \operatorname{cov}^u\big(\mu^{u-tc_n^{-1/2}k}(Rk \circ \vartheta_i | \mathcal{F}_{V_n}), \mu^{u-tc_n^{-1/2}k}(Rk \circ \vartheta_j | \mathcal{F}_{V_n})\big).$$

By the same arguments as for (4.4), we have $\mu^u(J_2^2) \to 0$, and therefore $J_2 = o_{\mu^u}(1)$.

(iv) Asymptotic normality of $c_n^{-1/2} \sum_{j \in V_n} (Rk \circ \vartheta_j - \mu^u(Rk))$ follows from the central limit theorem (3.6), applied for $Rk$ in place of $f_D$.

**Proof of the Theorem.** (i) By the first equation of (3.6), the empirical estimator $E_n f_D$ is asymptotically linear with influence function $f_D - \mu^u(f_D)$. By (3.7) and (3.8), this function is a gradient of $\mu^u(f_D)$. Hence $E_n f_D$ is regular by the characterization in Section 3.

(ii) The functional $\mu^u(f_D)$ is linear in $f_D$. Furthermore, $|D_n|^{1/2}(E_n f_D - \mu^u(f_D))$ is linear in $f_D$ up to $o_{\mu^u}(1)$ in the sense that we have homogeneity, and

$$|D_{1n}|^{1/2}(E_n f_{D_1} - \mu^u(f_{D_1})) + |D_{2n}|^{1/2}(E_n f_{D_2} - \mu^u(f_{D_2}))$$
$$= |(D_1 \cup D_2)_n|^{1/2}\big(E_n(f_{D_1} + f_{D_2}) - \mu^u(f_{D_1} + f_{D_2})\big) + o_{\mu^u}(1).$$

Hence it suffices to prove the assertion for functions $f_D$ which do not admit a non-trivial representation (3.10). For that, we must prove the following two statements:

1. If $D \subset A$ for some $A \in \mathcal{A}$, then $E_n f_D$ is efficient.
2. If some $A \in \mathcal{A}$ is a strict subset of $D$, then $E_n f_D$ is not efficient.

(iii) We prove statement 1. Let $A \in \mathcal{A}$ and $D \subset A$. By (i), the empirical estimator $E_n f_D$ is asymptotically linear with influence function $f_D - \mu^u(f_D)$, and this function is a gradient of $\mu^u(f_D)$. Since the gradient is in $K_0$, it is canonical. Hence $E_n f_D$ is regular and efficient by the characterization (3.9).

(iv) We prove statement 2. Let $A \in \mathcal{A}$ be a strict subset of $D$. Then the influence function $f_D - \mu^u(f_D)$ of $E_n f_D$ is not in $K_0$. The canonical gradient $g_D$ is the projection of $f_D - \mu^u(f_D)$ into $K_0$. Hence the asymptotic variance of $E_n f_D$ is larger than the variance bound by the squared length of the residual $h = f_D - \mu^u(f_D) - g_D$ with respect to the inner product (3.4) induced by local asymptotic normality,

$$\|h\|^2 = (h, h) = \sum_{j \in S} \mathrm{cov}^u(h \circ \vartheta_j, h).$$

# References


Bickel, P. J. , Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* Johns Hopkins University Press, Baltimore.

Bolthausen, E. (1982). On the central limit theorem for stationary mixing random fields. *Ann. Probab.* 10, 1047–1050.

Georgii, H.-O. (1988). *Gibbs measures and phase transitions.* De Gruyter Studies in Mathematics 9, De Gruyter, Berlin.

Greenwood, P. E., McKeague, I. W. and Wefelmeyer, W. (1998). Splitting Markov fields and combining empirical estimators. Unpublished manuscript.

Greenwood, P. E. and Wefelmeyer, W. (1994). Nonparametric estimators for Markov step processes. *Stochastic Process. Appl.* 52, 1–16.

Greenwood, P. E. and Wefelmeyer, W. (1995). Efficiency of empirical estimators for Markov chains. *Ann. Statist.* 23, 132–143.

Greenwood, P. E. and Wefelmeyer, W. (1996). Empirical estimators for semi-Markov processes. *Math. Meth. Statist.* 5, 299–315.

Gross, L. (1981). Absence of second-order phase transitions in the Dobrushin uniqueness region. *J. Statist. Phys.* 25, 57–72.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Z. Wahrsch. Verw. Gebiete* 14, 323–330.

Janžura, M. (1989). Local asymptotic normality for Gibbs random fields. In: *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics*, (P. Mandl, M. Huškova, eds.), 275–284, Charles University, Prague.

Koshevnik, Y. A. and Levit, B. Y. (1976). On a non-parametric analogue of the information matrix. *Theory Probab. Appl.* 21, 738–753.

Künsch, H. (1982a). Decay of correlations under Dobrushin's uniqueness condition and its applications. *Comm. Math. Phys.* 84, 207–222.

Künsch, H. (1982b). The central limit theorem for Gibbs fields with weak interaction. Unpublished manuscript.

Levit, B. Y. (1974). On optimality of some statistical estimates. In: *Proceedings of the Prague Symposium on Asymptotic Statistics*, (J. Hájek, eds.), 2, 215–238, Charles University, Prague.

Schick, A. and Wefelmeyer, W. (1998). Estimating joint distributions of Markov chains. Unpublished manuscript.

Winkler, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods.* Applications of Mathematics 27, Springer, Berlin.

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`

Wolfgang Wefelmeyer
Universität-GH Siegen
Fachbereich 6 Mathematik
Walter-Flex-Str. 3
57068 Siegen
Germany
`wefelmeyer@mathematik.uni-siegen.de`
`http://www.math.uni-siegen.de/statistik/wefelmeyer.html`