Plug-in estimators for random fields with nearest neighbor interactions

Wolfgang Wefelmeyer
University of Cologne
http://www.mi.uni-koeln.de/~wefelm/

Joint with

Cindy Greenwood, University of British Columbia
Ian McKeague, Columbia University

Salerno, June 15, 2018

Empirical estimator. Let $\mathbb{Z}^d$ be the d-dimensional square lattice, E an arbitrary state space, X a stationary random field on $E^{\mathbb{Z}^d}$. Observe X on the window $W = [-n, n]^d$.

Let f_V on $E^{\mathbb{Z}^d}$ be *local*, i.e., dependent only on a finite set $V \subset \mathbb{Z}^d$. A natural estimator fo the expectation $Ef_V(X_V)$ is the *empirical estimator*

$$\mathbb{P}_V f_V = \frac{1}{|i : V + i \subset W|} \sum_{V+i \subset W} f_V(X_{V+i}).$$

Under appropriate integrability conditions, the empirical estimator is asymptotically normal.
If LAN holds and no structural assumptions on the field are made, the empirical estimator is efficient.
Greenwood and W (1999), Janžura (2014).

The distribution of the random field X on $E^{\mathbb{Z}^d}$ is determined by its *local characteristic* at 0: the conditional distribution at 0 given the other sites.
Denote the *Manhattan metric* on $\mathbb{Z}^d$ by

$$|i| = \sum_{r=1}^{d} |i_r|, \qquad i \in \mathbb{Z}^d.$$

The *nearest neighbors* of 0 are the sites i with $|i| = 1$.
These are the $2d$ points in the unit sphere (rhombus) around 0.
The random field has *nearest neighbor interactions* if the local characteristic at 0 depends only on the $2d$ nearest neighbors.

A *clique* is a set of sites that are neighbors to each other. For nearest neighbor interactions, these are 0 and the d pairs $(0, e_r)$ with unit vectors e_r, and their shifts.

Gibbs representation. From now on the state space is $E = [0,1]$, and the random field on $[0,1]^{\mathbb{Z}^d}$ has Lebesgue density and nearest neighbor interactions. The *shift* of x by $-i \in \mathbb{Z}^d$ is $\vartheta_i(x)_j = x_{j+i}$.
Then for each finite $V \subset \mathbb{Z}^d$, the conditional density on V given the complement of V has the form

$$\frac{1}{Z_V} \exp\left[-\sum_{(C+i)\cap V \neq \emptyset} u_C \circ \vartheta_i\right]$$

with Z_V the norming constant and u_C functions depending on one of the $d+1$ cliques 0 and $(0, e_r)$, $r = 1, \ldots, d$.
In principle, the Gibbs representation can be used to construct efficient estimators. For *parametric* random fields, in particular finite state space, one can use a maximum likelihood estimator. The norming constant Z_V is a problem.
The empirical estimator for the expectation $Ef_V(X_V)$ of a local function f_V is efficient if and only if f_V is sum of functions each depending on a single clique. Greenwood and W (1999).

Factoring the distribution. For arbitrary random field and configurations A, B, C, we say that A *splits* B and C if B and C are conditonally independent given A.
W say that B *factors* given A if the sites in B are conditionally independent given A.

Consider a random field on $\mathbb{Z}^d$. Call a site i *even* if $\sum_{r=1}^{d} i_r$ is even; otherwise *odd*.
Assume nearest neighbor interactions. Then the odd sites factor given the even sites, and conversely. Besag (1974).

Let B_m and S_m be the ball and the sphere of radius m around 0.
Then S_m consists of even sites if m is even, and conversely.
Hence S_{m-1} splits B_{m-2} and B^c_{m-1}.
Hence the conditional distribution of B^c_{m-1} given B_{m-1} equals the conditional distribution of B^c_{m-1} given S_{m-1}.
Also, by Besag's observation, S_m factors given S_{m-1}.

The distribution on the ball B_k around 0 is now factored starting at 0 and going outward through the spheres $S_1, \ldots, S_k$.
Take distribution P_0 at 0. Now S_1 factors given $S_0 = \{0\}$ into $2d$ one-dimensional conditional distributions given 0.
Given $s \in S_m$ write (s) for the neighbors of s in S_{m-1}.
Write $Q(x_{(s)}, dx_s)$ for the corresponding conditional distribution.

The distribution on the ball B_k factors as

$$P_0(dx_0) \prod_{m=1}^{k} \prod_{s \in S_m} Q(x_{(s)}, dx_s).$$

Note: Configuration (s) is the smaller the more "exposed" s is:
If s is in the interior of a side of S_m of dimension $r = 2, \ldots, d-1$, then s has r neighbors in S_{m-1}.
Also, the "corners" of S_m have single neighbor in S_{m-1}.

Plug-in estimator. Let X be a stationary nearest-neighbor random field on $[0,1]^{\mathbb{Z}^d}$.
For simplicity, let the finite-dimensional densities of X be *quasi-uniform*, i.e., bounded and bounded away from 0.

Write $[s]$ for the union of s and (s).
Factor the density of $X_{[s]}$ into density of $X_{(s)}$ and conditional density of X_s given $X_{(s)}$:

$$p_{[s]}(x_{[s]}) = p_{(s)}(x_{(s)}) q_{s|(s)}(x_{(s)}, x_s).$$

Then the density on B_k factors as

$$p_0(x_0) \prod_{m=1}^{k} \prod_{s \in S_m} \frac{p_{[s]}(x_{[s]})}{p_{(s)}(x_{(s)})}.$$

Plug in density estimators $\widehat{p}_0$, $\widehat{p}_{(s)}$, $\widehat{p}_{[s]}$ for p_0, $p_{(s)}$, $p_{[s]}$ based on observations in the window $W = [-n, n]^d$ if the densities are smooth enough to allow for density estimators that have uniform rate $o(n^{-1/4})$ in probability, and a *plug-in property*:

$$\int f_V(x_V)\widehat{p}_V(x_V)\, dx_V = \mathbb{P}_V f_V + o_p(n^{-1/2}),$$

where $\mathbb{P}_V f_V$ is again the empirical estimator

$$\mathbb{P}_V f_V = \frac{1}{|i : V + i \subset W|} \sum_{V+i \subset W} f_V(X_{V+i}).$$

The *plug-in estimator* for an expectation $Ef_{B_k}(X_{B_k})$ is

$$\widehat{P}_{B_k} f_{B_k} = \int f(x_{B_k})\widehat{p}_0(x_0) \prod_{m=1}^{k} \prod_{s \in S_m} \frac{\widehat{p}_{[s]}(x_{[s]})}{\widehat{p}_{(s)}(x_{(s)})}\, dx_{B_k}.$$

It has smaller asymptotic variance than the empirical estimator

$$\mathbb{P}_{B_k} f_{B_k} = \frac{1}{|i : B_k + i \subset W|} \sum_{B_k+i \subset W} f(X_{B_k+i}).$$

Explicitly: The plug-in estimator is *asymptotically linear*,

$$\begin{aligned}\widehat{P} f_{B_k} &= \mathbb{P}_0 E(f_{B_k} | X_0) \\ &+ \sum_{m=1}^{k} \sum_{s \in S_m} \Big(\mathbb{P}_{[s]} E(f_{B_k} | X_{[s]}) - \mathbb{P}_{(s)} E(f_{B_k} | X_{(s)}) \Big) + o_p(n^{-1/2}),\end{aligned}$$

and the influence function is a projection of the influence function of the empirical estimator.

Degenerate case: Dimension $d = 1$.

Then the nearest neighbor random field is a first-order Markov chain.

The local characteristic at 0 is the conditional distribution at 0 given sites -1 and 1. The cliques are 0, $(0, 1)$, and their shifts.

For notational convenience, consider the time interval $\{0, \ldots, k\}$ in place of the ball $B_k = \{-k, \ldots, k\}$ around 0.

Assume state space $[0, 1]$ and Lebesgue density.

We have a representation of the density on $0, \ldots, k$ in terms of (conditional) densities on cliques, as in a Gibbs representation:

$$\begin{aligned} p_{k+1}(x_0, \ldots, x_k) &= p(x_0)q(x_0, x_1) \cdots q(x_{m-1}, x_m) \\ &= \frac{p(x_0, x_1) \cdots p(x_{m-1}, x_m)}{p_2(x_1) \cdots p_2(x_{m-1})}. \end{aligned}$$

Here p and p_2 are the 1- and 2-dimensional densities of the Markov chain, and q is the transition density.

Under conditions, the plug-in estimator is always efficient.

Kwon (2000). Different construction: Schick and W (2002).