On optimal spatial subsample size for variance estimation

Daniel J. Nordman
Dept. of Statistics
University of Dortmund
44221 Dortmund, Germany

Soumendra N. Lahiri
Dept. of Statistics
Iowa State University
Ames, IA 50011, USA

Abstract

In this paper, we consider the problem of determining the optimal block size for a spatial subsampling method for spatial processes observed on regular grids. We derive expansions for the mean square error of the subsampling variance estimator, which yields an expression for the theoretical optimal block size. The theoretical optimal block size is shown to depend in an intricate way on the geometry of the spatial sampling region as well as on the characteristics of the underlying random field. Final expressions for the optimal block size make use of some nontrivial estimates of lattice point counts in shifts of convex sets. Examples are presented to illustrate the variations in optimal block size for sampling regions of different shapes.

AMS(2000) Subject Classification: Primary 62G09; Secondary 62M40, 60G60

Key Words: Block bootstrap, block size, lattice point count, mixing, nonparametric variance estimation, random fields, spatial statistics, subsampling method

1 Introduction

We examine the problem of choosing subsample sizes to maximize the performance of subsampling methods for variance estimation. The data at hand are viewed as realizations of a stationary, weakly dependent, spatial lattice process. We consider the common scenario of sampling from sites of regular distance (e.g., indexed by the integer lattice $\mathbb{Z}^d$), lying within some region R_n embedded in $\mathbb{R}^d$. Such lattice data appear often in time series, agricultural field trials, and remote sensing and image analysis (medical and satellite image processing).

For variance estimation via subsampling, the basic idea is to construct several "scaled-down" copies of the sampling region R_n (subsamples) that fit inside R_n, evaluate the analog of $\hat{\theta}_n$ on each of these subregions, and then compute a properly normalized sample variance from the resulting values. The R_n-sampling scheme is essentially recreated at the level of the subregions. Two subsampling designs are most typical: subregions can be maximally overlapping (OL) or devised to be non-overlapping (NOL). The accuracy (e.g. variance and bias) of subsample-based estimators depends crucially on the choice of subsample size.

To place our work into perspective, we briefly outline previous research in variance estimation with subsamples and theoretical size considerations. The concept of variance estimation through subsampling originated from analysis of weakly dependent time processes. Let $\hat{\theta}_n$ be an estimator of a parameter of interest θ based on $\{Z(1), \ldots, Z(n)\}$ from a stationary temporal process $\{Z(i)\}_{i \geq 1}$. To obtain subsamples from a stationary stretch $Z(1), ..., Z(n)$ in time, Carlstein (1986) first proposed the use of NOL blocks of length $m \leq n$:

$$\Big\{Z(1+(i-1)m), ..., Z(im)\Big\}, \quad \text{for } i = 1, \ldots, \lfloor n/m \rfloor,$$

while the sequence of subseries:

$$\Big\{Z(i), ..., Z(i+m-1)\Big\} \quad \text{for } i = 1, ..., n-m+1$$

provides OL subsamples of length m (cf. Künsch, 1989; Politis and Romano, 1993b). Here, $\lfloor x \rfloor$ denotes the integer part of a real number x. In each respective collection, evaluations of an analog statistic $\hat{\theta}_i$ are made for each subseries and a normalized sample variance is calculated to estimate the parameter $n\mathrm{Var}(\hat{\theta}_n)$:

$$S_n^2 = \sum_{i=1}^{J} \frac{m(\hat{\theta}_i - \tilde{\theta})^2}{J}, \quad \tilde{\theta} = \sum_{i=1}^{J} \frac{\hat{\theta}_i}{J},$$

where $J = \lfloor n/m \rfloor$ ($J = n-m+1$) for the NOL (OL) subsample-based estimator. Carlstein (1986) and Fukuchi (1999) established the L_2 consistency of the NOL and OL estimators, respectively, for the variance of a general (not necessarily linear) statistic. Politis and Romano (1993b) determined asymptotic orders of the variance $O(m/n)$ and bias $O(1/m)$ of the subsample variance estimators for linear statistics. For mixing time series, they found that a subsample size m proportional to $n^{1/3}$ is optimal in the sense of minimizing the Mean Square Error (MSE) of variance estimation, concurring also with optimal block order for the moving block bootstrap variance estimator [Hall *et al.* (1995), Lahiri (1996)].

Cressie (1991, p. 492) conjectured the recipe for extending Carlstein's variance estimator to the general spatial setting, obtaining subsamples by tiling the sample region R_n with disjoint "congruent" subregions. Politis and Romano (1993a, 1994) have shown the consistency of subsample-based variance estimators for rectangular sampling/subsampling regions in $\mathbb{R}^d$ when the sampling sites are observed on $\mathbb{Z}^d \cap \prod_{i=1}^{d}[1, n_i]$ and integer translates of $\prod_{i=1}^{d}[1, m_i]$ yield the subsamples. Garcia-Soidan and Hall (1997) and Possolo (1991) proposed similar estimators under an identical sampling scenario. For linear statistics, Politis and Romano (1993a) determined that a subsampling scaling choice $\prod_{i=1}^{d} m_i = C\{\prod_{i=1}^{d} n_i\}^{d/(d+2)}$, for some unknown C, minimizes the order of a variance estimator's asymptotic MSE. Sherman and Carlstein (1994) and Sherman (1996) proved the MSE-consistency of NOL and OL subsample estimators, respectively, for the variance of general statistics in $\mathbb{R}^2$. Their work allowed for a more flexible sampling scheme: the "inside" of a simple closed curve defines a set $D \subset [-1, 1]^2$, $\mathbb{Z}^2 \cap nD$ (using a scaled-up copy of D) constitutes the set of sampling sites, and translates of mD within nD form subsamples. Sherman (1996) minimized a bound on the asymptotic order of the OL estimator's MSE to argue that the best size choice for OL subsamples involves $m = O(n^{1/2})$ (coinciding with the above findings of Politis and Romano (1993a) for rectangular regions in $\mathbb{R}^2$). Politis and Sherman (2001) have developed consistent subsampling methods for variance estimation with marked point process data [cf. Politis *et al.* (1999), Chapter 6].

Few theoretical and numerical recommendations for choosing subsamples have been offered in the spatial setting, especially with the intent of variance estimation. As suggested in the literature, an explicit theoretical determination of optimal subsample "scaling" (size) requires calculation of an order and associated proportionality constant for a given sampling region R_n. Even for the few sampling situations where the order of optimal subsample size has been established, the exact adjustments to theses orders are unknown and, quoting Politis and Romano (1993a), "important (and difficult) in practice." Beyond the time series case with the univariate sample mean, the influence of the geometry and dimension of R_n, as well as the structure of $\hat{\theta}_n$, on precise subsample selection has not been explored. We attempt here to advance some ideas on the best size choice, both theoretically and empirically, for subsamples.

We work under the "smooth function" model of Hall (1992), where the statistic of interest $\hat{\theta}_n$ can be represented as a function of sample means. We formulate a framework for sampling in $\mathbb{R}^d$ where the sampling region R_n (say) is obtained by "inflating" a prototype set in the unit cube in $\mathbb{R}^d$ and the subsampling regions are given by suitable translates of a scaled down copy of the sampling region R_n. We consider both a non-overlapping version and a (maximal) overlapping version of the subsampling method. For each method, we derive expansions for the variance and the bias of the corresponding subsample estimator of $\mathrm{Var}(\hat{\theta}_n)$. The asymptotic variance of the spatial subsample estimator for the OL version turns out to be smaller than that of the NOL version by a constant factor K_1 (say) which depends solely on the geometry of the sampling region R_n. In the time series case, Meketon and Schmeiser (1984), Künsch (1989), Hall *et al.* (1995) and Lahiri (1996) have shown in different degrees of generality that the asymptotic variance under the OL subsampling scheme, compared to the NOL one, is $K_1 = \frac{2}{3}$ times smaller. Results of this paper show that for rectangular sampling regions R_n in d-dimensional space, the factor K_1 is given by $(\frac{2}{3})^d$. We list the factor K_1 for sampling regions of some common shapes in Table 1.

Table 1: Examples of K_1 for several shapes of the sampling region $R_n \subset \mathbb{R}^d$.

Shape of R_n	Rectangle in $\mathbb{R}^d$	Sphere in $\mathbb{R}^3$	Circle in $\mathbb{R}^2$	Right triangle in $\mathbb{R}^2$
K_1	$(2/3)^d$	$17\pi/315$	$\pi/4 - 4/(3\pi)$	$1/5$

In contrast, the bias parts of both the OL and NOL subsample variance estimators are (usually) asymptotically equivalent and depend on the covariance structure of the random field as well as on the geometry of the sampling region R_n. Since the bias term is typically of the same order as the number of lattice points lying near a subsample's boundary, determination of the leading bias term involves some nontrivial estimates of the lattice point counts over translated subregions. Counting lattice points in scaled-up sets is a hard problem and has received a lot of attention in Analytic Number Theory and in Combinatorics. Even for the case of the plane (i.e., $d = 2$), the counting results available in the literature are directly applicable to our problem only for a very restricted class of subregions (that have the so-called "smoothly winding border" [cf. van der Corput (1920), Huxley (1993, 1996)]. Here explicit expressions for the bias terms are derived for a more general class of sampling regions using some new estimates on the discrepancy between the number of lattice points and the volume of the *shifted* subregions in the plane and in three dimensional Euclidean space. In particular, our results are applicable to sampling regions that do not necessarily have "smoothly winding borders".

Minimizing the combined expansions for the bias and the variance parts, we derive explicit expressions for the theoretical optimal block size for sampling regions of different shapes. To briefly describe the result for a few common shapes, suppose the sampling region R_n is obtained by inflating a given set $R_0 \in (-\frac{1}{2}, \frac{1}{2}]^d$ by a scaling constant λ_n as $R_n = \lambda_n R_0$ and that the subsamples are formed by considering the translates of ${}_sR_n = {}_s\lambda_n R_0$. Then, the theoretically optimal choice of the block size ${}_s\lambda_n$ for the OL version is of the form

$$
{}_s\lambda_n^{\mathrm{opt}} = \left(\frac{\lambda_n^d B_0^2}{d K_0 \tau^4} \right)^{1/(d+2)} (1 + o(1)) \quad \text{as} \quad n \to \infty
$$

for some constants B_0 and K_0 (coming from the bias and the variance terms, respectively), where τ^2 is a population parameter that does not depend on the shape of the sampling region R_n (see Theorem 5.1 for details). The following table lists the constants B_0 and K_0 for some shapes of R_n. It follows from Table 2 that, unlike the time series case, in higher dimensions the

Table 2: Examples of B_0, K_0 for some sampling regions R_n. Cross and triangle shapes appear in Figure 1, Section 6; see Section 6 for further details. Autocovariances $\sigma(\cdot)$ and Euclidean, l^1, and l^∞ norms $\|\cdot\|$, $\|\cdot\|_1$, $\|\cdot\|_\infty$ are described in Section 2.3.

R_n	K_0	B_0
Sphere in $\mathbb{R}^3$	$34/105$	$3/2 \sum_{\mathbf{k} \in \mathbb{Z}^3} \|\mathbf{k}\| \sigma(\mathbf{k})$
Cross in $\mathbb{R}^2$	$4/9 \cdot 191/192$	$4/3 \sum_{\mathbf{k} \in \mathbb{Z}^2} \|\mathbf{k}\|_1 \sigma(\mathbf{k})$
Right triangle in $\mathbb{R}^2$	$2/5$	$2 \sum_{\substack{\mathbf{k}=(k_1,k_2)' \in \mathbb{Z}^2 \\ \mathrm{sign} k_1 = \mathrm{sign} k_2}} \|\mathbf{k}\|_1 \sigma(\mathbf{k}) + 2 \sum_{\substack{\mathbf{k} \in \mathbb{Z}^2 \\ \mathrm{sign} k_1 \neq \mathrm{sign} k_2}} \|\mathbf{k}\|_\infty \sigma(\mathbf{k})$

optimal block size critically depends on the shape of the spatial sampling region R_n. It simplifies only slightly for the NOL subsampling scheme as the constant K_0 is unnecessary for computing optimal NOL subsamples, but the bias constant B_0 is often the same for both estimators from each version of subsampling. These expressions may be readily used to obtain 'plug-in' estimates of the theoretical optimal block lengths for use in practice.

The rest of the paper is organized as follows. In Section 2, we describe the spatial subsampling method and state the Assumptions used in the paper. In Sections 3 and 4, we respectively derive expansions for the variance and the bias parts of the subsampling estimators. Theoretical optimal block lengths are derived in Section 5. The results are illustrated with some common examples in Section 6. Section 7 addresses data-based subsample size selection. Proofs of variance and bias results are separated into Sections 8 and 9, respectively.

2 Variance estimators via subsampling

In Section 2.1, we frame the sampling design and the structure of the sampling region. Two methods of subsampling are presented in Section 2.2 along with corresponding nonparametric variance estimators. Assumptions and Conditions used in the paper are given in Section 2.3.

2.1 The sampling structure

To describe the sampling scheme used, we first assume all potential sampling sites are located on a translate of the rectangular integer lattice in $\mathbb{R}^d$. For a fixed (chosen) vector $\mathbf{t} \in [-1/2, 1/2)^d$, we identify the $\mathbf{t}$-translated integer lattice as $\mathbf{Z}^d \equiv \mathbf{t} + \mathbb{Z}^d$. Let $\{Z(\mathbf{s}) \mid \mathbf{s} \in \mathbf{Z}^d\}$ be a stationary weakly dependent random field (hereafter r.f.) taking values in $\mathbb{R}^p$. (We use bold font as a standard to denote vectors in the space of sampling $\mathbb{R}^d$ and normal font for vectors in $\mathbb{R}^p$, including $Z(\cdot)$.) We suppose that the process $Z(\cdot)$ is observed at sampling sites lying within the sampling region $R_n \subset \mathbb{R}^d$. Namely, the collection of available sampling sites is

$$\{Z(\mathbf{s}) \,:\, \mathbf{s} \in R_n \cap \mathbf{Z}^d\}.$$

To obtain the results in the paper, we assume that the sampling region R_n becomes unbounded as the sample size increases. This will provide a commonly used "increasing domain" framework for studying asymptotics with spatial lattice data [cf. Cressie (1991)]. We next specify the structure of the regions R_n and employ a formulation similar to that of Lahiri (1999a,b).

Let R_0 be a Borel subset of $(-1/2,\, 1/2]^d$ containing an open neighborhood of the origin such that for any sequence of positive real numbers $a_n \to 0$, the number of cubes of the scaled lattice $a_n\mathbb{Z}^d$ which intersect the closures $\overline{R_0}$ and $\overline{R_0^c}$ is $O((a_n^{-1})^{d-1})$ as $n \to \infty$. Let Δ_n be a sequence of $d \times d$ diagonal matrices, with positive diagonal elements $\lambda_1^{(n)}, \ldots, \lambda_d^{(n)}$, such that each $\lambda_i^{(n)} \to \infty$ as $n \to \infty$. We assume that the sampling region R_n is obtained by "inflating" the template set R_0 by the directional scaling factors Δ_n; that is,

$$R_n = \Delta_n R_0.$$

Because the origin is assumed to lie in R_0, the sampling region R_n grows outward in all directions as n increases. Furthermore, if the scaling factors are all equal ($\lambda_1^{(n)} = \cdots = \lambda_d^{(n)}$), the shape of R_n remains the same for different values of n.

The formulation given above allows the sampling region R_n to have a large variety of fairly irregular shapes with the boundary condition on R_0 imposed to avoid pathological cases. Some common examples of such regions are convex subsets of $\mathbb{R}^d$, such as spheres, ellipsoids, polyhedrons, as well as certain non-convex subsets with irregular boundaries, such as star-shaped regions. Sherman and Carlstein (1994) and Sherman (1996) consider a similar class of such regions in the plane (i.e. $d = 2$) where the boundaries of the sets R_0 are delineated by simple rectifiable curves with finite lengths. The border requirements on R_0 ensure that the number of observations near the boundary of R_n is negligible compared to the totality of data values. This assumption is crucial to the derivation of results to follow. In addition, Perera (1997) demonstrates that the border geometry of R_0 can significantly influence the asymptotic distribution of sample means taken with observations from "expanding" sets, like R_n.

2.2 Subsampling designs and variance estimators

We suppose that the relevant statistic, whose variance we wish to estimate, can be represented as a function of sample means. Let $\hat{\theta}_n = H(\bar{Z}_{N_n})$ be an estimator of the population parameter of interest $\theta = H(\mu)$, where $H : \mathbb{R}^p \to \mathbb{R}$ is a smooth function, $\mathrm{E}Z(\mathbf{t}) = \mu \in \mathbb{R}^p$ is the mean of the stationary r.f. $Z(\cdot)$, and $\bar{Z}_{N_n}$ is the sample mean of the N_n observations within R_n. We can write

$$\bar{Z}_{N_n} = N_n^{-1} \sum_{\mathbf{s} \in \mathbf{Z}^d \cap R_n} Z(\mathbf{s}). \tag{2.1}$$

This parameter and estimator formulation is what Hall (1992) calls the "smooth function" model and it has been used in other scenarios, such as with the moving block bootstrap (MBB) and empirical likelihood, for studying approximately linear functions of a sample mean [Lahiri (1996), DiCiccio *et al.* (1991)]. By considering suitable functions of the $Z(\mathbf{s})$'s, one can represent a wide range of estimators under the present framework. In particular, these include means, products and ratios of means, autocorrelation estimators, sample lag cross-correlation estimators [Politis and Romano (1993b)], and Yule-Walker estimates for autoregressive processes [cf. Guyon (1995)].

The quantity which we seek to estimate nonparametrically is the variance of the normalized statistic $\sqrt{N_n}\hat{\theta}_n$, say, $\tau_n^2 = N_n \mathrm{E}(\hat{\theta}_n - \mathrm{E}\hat{\theta}_n)^2$. In our problem, this goal is equivalent to consistently estimating the limiting variance $\tau^2 = \lim_{n\to\infty} \tau_n^2$.

2.2.1 Overlapping subsamples

Variance estimation with OL subsampling regions has been presented previously in the literature, though in more narrow sampling situations. Sherman (1996) considered an OL subsample-based estimator for sampling regions in $\mathbb{R}^2$; Politis and Romano (1994) extended a similar estimator for rectangular regions in $\mathbb{R}^d$ with faces parallel to the coordinate axes (i.e. $R_0 = (-1/2, 1/2]^d$); and a host of authors, in a variety of contexts, have examined OL subsample estimators applied to time series data [cf. Song and Schmeiser (1988), Politis and Romano (1993a), Fukuchi (1999)].

We first consider creating a smaller version of R_n, which will serve as a template for the OL subsampling regions. To this end, let ${}_s\Delta_n$ be a $d \times d$ diagonal matrix with positive diagonal elements, $\{{}_s\lambda_1^{(n)}, \ldots, {}_s\lambda_d^{(n)}\}$, such that ${}_s\lambda_i^{(n)}/\lambda_i^{(n)} \to 0$ and ${}_s\lambda_i^{(n)} \to \infty$, as $n \to \infty$, for each $i = 1, \ldots, d$. (The matrix Δ_n represents the determining scaling factors for R_n and ${}_s\Delta_n$ shall be factors used to define the subsamples.) We make the "prototype" subsampling region,

$$ {}_sR_n = {}_s\Delta_n R_0, \tag{2.2} $$

and identify a subset of $\mathbb{Z}^d$, say J_{OL}, corresponding to all integer translates of ${}_sR_n$ lying within R_n. That is,

$$ J_{OL} = \left\{ \mathbf{i} \in \mathbb{Z}^d \, : \, \mathbf{i} + {}_sR_n \subset R_n \right\}. $$

The desired OL subsampling regions are precisely the translates of ${}_sR_n$ given by: $R_{\mathbf{i},n} \equiv \mathbf{i} + {}_sR_n$, $\mathbf{i} \in J_{OL}$. Note that the origin belongs to J_{OL} and, clearly, some of these subregions overlap.

Let ${}_sN_n = |\mathbf{Z}^d \cap {}_sR_n|$ be the number of sampling sites in ${}_sR_n$ and let $|J_{OL}|$ denote the number of available subsampling regions. (The number of sampling sites within each OL subsampling region is the same, namely for any $\mathbf{i} \in J_{OL}$, ${}_sN_n = |\mathbf{Z}^d \cap R_{\mathbf{i},n}|$.) For each $\mathbf{i} \in J_{OL}$, compute $\hat{\theta}_{\mathbf{i}} = H(Z_{\mathbf{i},n})$, where

$$ Z_{\mathbf{i},n} = {}_sN_n^{\;-1} \sum_{\mathbf{s} \in \mathbf{Z}^d \cap R_{\mathbf{i},n}} Z(\mathbf{s}) $$

denotes the sample mean of observations within the subregion. We then have the OL subsample variance estimator of τ_n^2 as

$$ \hat{\tau}_{n,OL}^2 = |J_{OL}|^{-1} \sum_{\mathbf{i} \in J_{OL}} {}_sN_n \left(\hat{\theta}_{\mathbf{i},n} - \tilde{\theta}_n \right)^2, \quad \tilde{\theta}_n = |J_{OL}|^{-1} \sum_{\mathbf{i} \in J_{OL}} \hat{\theta}_{\mathbf{i},n}. $$

2.2.2 Non-overlapping subsamples

Carlstein (1986) first proposed a variance estimator involving NOL subsamples for time processes. Politis and Romano (1993a) and Sherman and Carlstein (1994) demonstrated the consistency of variance estimation, via NOL subsampling, for certain rectangular regions in $\mathbb{R}^d$ and some sampling regions in $\mathbb{R}^2$, respectively. Here we adopt a formulation similar to those of Sherman and Carlstein (1994) and Lahiri (1999a).

The sampling region R_n is first divided into disjoint "cubes". Let ${}_s\Delta_n$ be the previously described $d \times d$ diagonal matrix from (2.2), which will determine the "window width" of the partitioning cubes. Let

$$J_{NOL} = \left\{ \mathbf{i} \in \mathbf{Z}^d \,:\, {}_s\Delta_n \left(\mathbf{i} + (-1/2,\, 1/2]^d \right) \subset R_n \right\}$$

represent the set of all "inflated" subcubes that lie inside R_n. Denote its cardinality as $|J_{NOL}|$. For each $\mathbf{i} \in J_{NOL}$, define the subsampling region $\tilde{R}_{\mathbf{i},n} = {}_s\Delta_n(\mathbf{i} + R_0)$ by inscribing the translate of ${}_s\Delta_n R_0$ such that the origin is mapped onto the midpoint of the cube ${}_s\Delta_n(\mathbf{i} + (-1/2,\, 1/2]^d)$. This provides a collection of NOL subsampling regions, which are smaller versions of the original sampling region R_n, that lie inside R_n.

For each $\mathbf{i} \in J_{NOL}$, the function $H(\cdot)$ is evaluated at the sample mean, say $\tilde{Z}_{\mathbf{i},n}$ for a corresponding subsampling region $\tilde{R}_{\mathbf{i},n}$ to obtain $\hat{\theta}_{\mathbf{i},n} = H(\tilde{Z}_{\mathbf{i},n})$. The NOL subsample estimator of τ_n^2 is again an appropriately scaled sample variance:

$$\hat{\tau}^2_{n,NOL} = |J_{NOL}|^{-1} \sum_{\mathbf{i} \in J_{NOL}} {}_sN_{\mathbf{i},n} \left(\hat{\theta}_{\mathbf{i},n} - \tilde{\theta}_n \right)^2, \quad \tilde{\theta}_n = |J_{NOL}|^{-1} \sum_{\mathbf{i} \in J_{NOL}} \hat{\theta}_{\mathbf{i},n}$$

where ${}_sN_{\mathbf{i},n} = |\mathbf{Z}^d \cap \tilde{R}_{\mathbf{i},n}|$ denotes the number of sampling sites within a given NOL subsample.

We note that ${}_sN_{\mathbf{i},n}$ may differ between NOL subsamples, but all such subsamples will have exactly ${}_sN_{\mathbf{i},n} = {}_sN_n$ sites available if the diagonal elements of ${}_s\Delta_n$ are integers.

2.3 Assumptions

For stating the assumptions, we need to introduce some notation. For a vector $\mathbf{x} = (x_1, ..., x_d)' \in \mathbb{R}^d$, let $\|\mathbf{x}\|$ and $\|\mathbf{x}\|_1 = \sum_{i=1}^d |x_i|$ denote the usual Euclidean and l^1 norms of $\mathbf{x}$, respectively. Denote the l^∞ norm as $\|\mathbf{x}\|_\infty = \max_{1 \le k \le d} |x_k|$. Define $\text{dis}(E_1,\, E_2) = \inf\{\|\mathbf{x} - \mathbf{y}\|_\infty : \mathbf{x} \in E_1, \mathbf{y} \in E_2\}$ for two sets $E_1,\, E_2 \subset \mathbb{R}^d$. We shall use the notation $|\cdot|$ also in two other cases: for a countable set B, $|B|$ would denote the cardinality of the set B; for an uncountable set $A \subset \mathbb{R}^d$, $|A|$ would refer to the volume (i.e., the $\mathbb{R}^d$ Lebesgue measure) of A.

Let $\mathcal{F}_Z(T) = \sigma\langle Z(\mathbf{s}) \,:\, \mathbf{s} \in T \rangle$ be the σ-field generated by the variables $\{Z(\mathbf{s}) \,:\, \mathbf{s} \in T\}$, $T \subset \mathbf{Z}^d$. For $T_1,\, T_2 \subset \mathbf{Z}^d$, write

$$\tilde{\alpha}(T_1, T_2) = \sup\{|P(A \cap B) - P(A)P(B)| \,:\, A \in \mathcal{F}_Z(T_1), B \in \mathcal{F}_Z(T_2)\}.$$

Then, the strong mixing coefficient for the r.f. $Z(\cdot)$ is defined as

$$\alpha(k, l) = \sup\{\tilde{\alpha}(T_1, T_2) \,:\, T_i \subset \mathbf{Z}^d,\, |T_i| \le l,\, i = 1, 2;\, \text{dis}(T_1, T_2) \ge k\} \tag{2.3}$$

Note that the supremum in the definition of $\alpha(k, l)$ is taken over sets T_1, T_2 which are bounded. For $d > 1$, this is important. A r.f. on the (rectangular) lattice $\mathbb{Z}^d$ with $d \ge 2$ that satisfies a strong mixing condition of the form

$$\lim_{k \to \infty} \sup\{\tilde{\alpha}(T_1, T_2) \,:\, T_1, T_2 \subset \mathbf{Z}^d,\, \text{dis}(T_1, T_2) \ge k\} = 0 \tag{2.4}$$

with supremum taken over possibly unbounded sets necessarily belongs to the more restricted class of ρ-mixing r.f.'s [cf. Bradley (1989)]. Politis and Romano (1993a) used moment inequalities based on the mixing condition in (2.4) to determine the orders of the bias and variance of $\hat{\tau}^2_{n,OL}$, $\hat{\tau}^2_{n,NOL}$ for rectangular sampling regions.

For proving the subsequent theorems, the assumptions below are needed along with two conditions stated as functions of a positive argument $r \in \mathbb{Z}_+ = \{0, 1, 2, \ldots\}$. In the following, $\det(\Delta)$ represents the determinant of a square matrix Δ. For $\alpha = (\alpha_1, ..., \alpha_p)' \in (\mathbb{Z}_+)^p$, let D^α denote the αth order partial differential operator $\partial^{\alpha_1+...+\alpha_p}/\partial x_1^{\alpha_1}...\partial x_p^{\alpha_p}$ and $\nabla = (\partial H(\mu)/\partial x_1, \ldots, \partial H(\mu)/\partial x_p)'$ be the vector of first order partial derivatives of H at μ. Limits in order symbols are taken letting n tend to infinity.

Assumptions:

(A.1) There exists a $d \times d$ diagonal matrix Δ_0, $\det(\Delta_0) > 0$, such that

$$\frac{1}{{}_s\lambda_1^{(n)}} \, {}_s\Delta_n \;\rightarrow\; \Delta_0.$$

(A.2) For the scaling factors of the sampling and subsampling regions:

$$\sum_{i=1}^{d} \frac{1}{{}_s\lambda_i^{(n)}} + \sum_{i=1}^{d} \frac{{}_s\lambda_i^{(n)}}{\lambda_i^{(n)}} + \frac{[\det({}_s\Delta_n)]^{(d+1)/d}}{\det(\Delta_n)} = o(1), \quad \max_{1\le i\le d} \lambda_i^{(n)} = O\big(\min_{1\le i\le d} \lambda_i^{(n)}\big).$$

(A.3) There exist nonnegative functions $\alpha_1(\cdot)$ and $g(\cdot)$ such that $\lim_{k\to\infty} \alpha_1(k) = 0$, $\lim_{l\to\infty} g(l) = \infty$ and the strong-mixing coefficient $\alpha(k, l)$ from (2.3) satisfies the inequality

$$\alpha(k, l) \le \alpha_1(k)g(l) \quad k > 0,\ l > 0.$$

(A.4) $\sup\{\, \tilde{\alpha}(T_1, T_2) \,:\, T_1, T_2 \subset \mathbf{Z}^d,\ |T_1| = 1,\ \mathrm{dis}(T_1, T_2) \ge k\} = o(k^{-d})$.

(A.5) $\tau^2 > 0$, where $\tau^2 = \sum_{\mathbf{k}\in\mathbb{Z}^d} \sigma(\mathbf{k}),\quad \sigma(\mathbf{k}) = \mathrm{Cov}\big(\nabla' Z(\mathbf{t}), \nabla' Z(\mathbf{t}+\mathbf{k})\big)$.

Conditions:

D_r: $H : \mathbb{R}^p \to \mathbb{R}$ is r-times continuously differentiable and, for some $a \in \mathbb{Z}_+$ and real $\mathcal{C} > 0$,

$$\max\{|D^\nu H(\mathbf{x})| \,:\, \|\nu\|_1 = r\} \le \mathcal{C}(1 + \|\mathbf{x}\|^a), \quad \mathbf{x} \in \mathbb{R}^p.$$

M_r: For some $0 < \delta \le 1$, $0 < \kappa < (2r - 1 - 1/d)(2r + \delta)/\delta$, and $\mathcal{C} > 0$,

$$\mathrm{E}\|Z(\mathbf{t})\|^{2r+\delta} < \infty, \quad \sum_{x=1}^{\infty} x^{(2r-1)d-1} \alpha_1(x)^{\delta/(2r+\delta)} < \infty, \quad g(x) \le \mathcal{C}x^\kappa.$$

Some comments about the assumptions and the conditions are in order.

Assumption A.5 implies a positive, finite asymptotic variance τ^2 for the standardized estimator, $\sqrt{N_n}\hat{\theta}_n$. We would like $\tau^2 \in (0, \infty)$ for a purposeful variance estimation procedure.

In Assumption A.3, we formulate a conventional bound on the mixing coefficient $\alpha(k, l)$ from (2.3) that is applicable to many r.f.s and resembles the mixing assumption of Lahiri (1999a,b).

For r.f.s satisfying Assumption A.3, the "distance" component of the bound, $\alpha_1(\cdot)$, often decreases at an exponential rate while the function of "set size," $g(\cdot)$, increases at a polynomial rate [cf. Guyon (1995)]. Examples of r.f.s that meet the requirements of A.3 and M_r include Gaussian fields with analytic spectral densities, certain linear fields with a moving average or autoregressive (AR) representation (like m-dependent fields), separable AR(1)xAR(1) lattice processes suggested by Martin (1990) for modeling in $\mathbb{R}^2$, many Gibbs and Markov fields, and important time series models [cf. Doukhan (1994)]. Condition M_r combined with A.3 also provides useful moment bounds for normed sums of observations (see Lemma 8.2).

Assumption A.4 permits the CLT in Bolthausen (1982) to be applied to sums of $Z(\cdot)$ on sets of increasing domain, in conjunction with the boundary condition on R_0, Assumption A.3, and Condition M_r. This version of the CLT (Stein's method) is derived from α-mixing conditions which ensure asymptotic independence between a single point and observations in arbitrary sets of increasing distance [cf. Perera (1997)].

Assumptions A.1 and A.2 set additional guidelines for how sampling and subsampling design parameters, Δ_n and ${}_s\Delta_n$, may be chosen. The assumptions provide a flexible framework for handling "increasing domains" of many shapes. For $d = 1$, A.1-A.2 are equivalent to the requirements of Lahiri (1999c) who provides variance and bias expansions for the MBB variance estimator with weakly dependent time processes.

3 Variance expansions

We now give expansions for the asymptotic variance of the OL/NOL subsample variance estimators $\hat{\tau}^2_{n,OL}$ and $\hat{\tau}^2_{n,NOL}$ of $\tau^2_n = N_n \mathrm{Var}(\hat{\theta}_n)$.

Theorem 3.1 *Suppose that Assumptions* A.1 - A.5 *and Conditions* D_2 *and* M_{5+2a} *hold, then*

$$(a) \quad \mathrm{Var}(\hat{\tau}^2_{n,OL}) \;=\; K_0 \cdot \frac{\det({}_s\Delta_n)}{\det(\Delta_n)}\left[2\tau^4\right](1+o(1)),$$

$$(b) \quad \mathrm{Var}(\hat{\tau}^2_{n,NOL}) \;=\; \frac{1}{|R_0|} \cdot \frac{\det({}_s\Delta_n)}{\det(\Delta_n)}\left[2\tau^4\right](1+o(1)),$$

where $K_0 = \dfrac{1}{|R_0|} \cdot \displaystyle\int_{\mathbb{R}^d} \frac{|(\mathbf{x}+R_0)\cap R_0|^2}{|R_0|^2} d\mathbf{x}$ *is an integral with respect to the* $\mathbb{R}^d$ *Lebesgue measure.*

The constant K_0 appearing in the variance expansion of the estimator $\hat{\tau}^2_{n,OL}$ is a property of the *shape* of the sampling template R_0 but not of its exact embedding in space $\mathbb{R}^d$ or even the scale of the set. Namely, K_0 is invariant to invertible affine transformations applied to R_0 and hence can be computed from either R_0 or $R_n = \Delta_n R_0$. Values of K_0 for some template shapes are given in Table 3 and Section 6.

A stationary time sequence $Z_1, \ldots, Z_n$ can be obtained within our sampling formulation by choosing $R_0 = (-1/2, 1/2]$ and $\lambda_1^{(n)} = n$ on the untranslated integer lattice $\mathbf{Z} = \mathbb{Z}$. In this special sampling case, an application of Theorem 3.1 yields

$$\mathrm{Var}(\hat{\tau}^2_{n,OL}) = 2/3 \cdot \mathrm{Var}(\hat{\tau}^2_{n,NOL}), \quad \mathrm{Var}(\hat{\tau}^2_{n,NOL}) = {}_s\lambda_1^{(n)}/\lambda_1^{(n)} \cdot [2\tau^4](1+o(1)),$$

a result which is well-known for "nearly" linear functions $\hat{\theta}_n$ of a time series sample mean [cf. Künsch (1989)]. Theorem 3.1 implies that, under the "smooth" function model, the asymptotic

Table 3: Examples of K_0 from Theorem 3.1 for several shapes of $R_0 \subset \mathbb{R}^d$. *The trapezoid has a 90° interior $\angle$ and parallel sides $b_2 \geq b_1$; $c = (b_2/b_1+1)^{-2}[1+2(b_2/b_1-1)/(b_2/b_1+1)]$.

R_0 **Shape**	$\mathbb{R}^d$ Rectangle	$\mathbb{R}^3$ Ellipsoid	$\mathbb{R}^3$ Cylinder	$\mathbb{R}^2$ Ellipse	$\mathbb{R}^2$ Trapezoid*
K_0	$(2/3)^d$	$34/105$	$2/3\,(1-16/(3\pi^2))$	$1-16/(3\pi^2)$	$2/5\,(1+4c/9)$

variance of the OL subsample-based variance estimator is always strictly less than the NOL version because

$$K_1 \;=\; \lim_{n\to\infty} \frac{\mathrm{Var}(\hat{\tau}^2_{n,OL})}{\mathrm{Var}(\hat{\tau}^2_{n,NOL})} \;=\; K_0|R_0| \;<\; 1. \tag{3.1}$$

If both estimators have the same bias (which is often the case), (3.1) implies that variance estimation with OL subsamples is more efficient than the NOL subsample alternative owing to a smaller asymptotic MSE.

Unlike K_0, K_1 does depend on the volume $|R_0|$, which in turn is constrained by the R_0-template's geometry. That is, K_1 is ultimately bounded, through $|R_0|$ in (3.1), by the amount of space that an object of R_0's *shape* can possibly occupy within $(-1/2,1,2]^d$; or by how much volume can be filled by a given geometrical body (e.g. circle) compared to a cube. The constants K_1 in Table 1 are computed with templates of prescribed shape and largest possible volume in $(-1/2,1/2]^d$. These values most accurately reflect the influence of R_0's (or R_n's) geometry on the relative performance (ie. variance, efficiency) of the estimators $\hat{\tau}^2_{n,OL}$ and $\hat{\tau}^2_{n,NOL}$.

To conclude this section, we remark that both subsample-based variance estimators can be shown to be (MSE) consistent under Theorem 3.1 conditions, allowing for more general spatial sampling regions, in both shape and dimension, than previously considered. Inference on the parameter θ can be made through the limiting standard normal distribution of $\sqrt{N_n}(\hat{\theta}_n-\theta)/\hat{\tau}_n$ for $\hat{\tau}_n = \hat{\tau}_{n,OL}$ or $\hat{\tau}_{n,NOL}$.

4 Bias expansions

We now try to capture and precisely describe the leading order terms in the asymptotic bias of each subsample-based variance estimator, similar to the variance determinations from the previous section. We first establish and note the order of the dominant component in the bias expansions of $\hat{\tau}^2_{n,OL}$ and $\hat{\tau}^2_{n,NOL}$, which is the subject of the following lemma.

Lemma 4.1 *With Assumptions* A.1 - A.5*, suppose that Conditions D_2 and M_{2+a} hold for $d \geq 2$; and D_3 and M_{3+a} for $d=1$. Then, the subsample estimators of $\tau^2_n = N_n \mathrm{Var}(\hat{\theta}_n)$ have expectations:*

$$\mathrm{E}(\hat{\tau}^2_{n,OL}) = \tau^2_n + O(1/\,{}_s\lambda_1^{(n)}) \quad \textit{and} \quad \mathrm{E}(\hat{\tau}^2_{n,NOL}) = \tau^2_n + O(1/\,{}_s\lambda_1^{(n)}).$$

The lemma shows that, under the smooth function model, the asymptotic bias of each estimator is $O(1/\,{}_s\lambda_1^{(n)})$ for all dimensions of sampling. Politis and Romano (1993a, p.323) and Sherman (1996) showed this same *size* for the bias of $\hat{\tau}^2_{n,OL}$ with sampling regions based on rectangles $R_0 = (-1/2,1/2]^d$ or simple closed curves in $\mathbb{R}^2$, respectively. Lemma 4.1 extends these results to a broader class of sampling regions. However, we would like to precisely identify the

$O(1/{}_s\lambda_1^{(n)})$ bias component for $\hat{\tau}^2_{n,OL}$ or $\hat{\tau}^2_{n,NOL}$ to later obtain theoretically correct proportionality constants associated with optimal subsample scaling.

To achieve some measure of success in determining the exact bias of the subsampling estimators, we reformulate the subsampling design slightly so that ${}_s\lambda_n \equiv {}_s\lambda_1^{(n)} = \cdots = {}_s\lambda_d^{(n)}$. That is, a common scaling factor in all directions is now used to define the subsampling regions (as in Sherman and Carlstein (1994); Sherman (1996)). This constraint will allow us to deal with the counting issues at the heart of the bias expansion.

Adopting a common scaling factor ${}_s\lambda_n$ for the subsamples also is sensible for a few other reasons at this stage:

- "Unconstrained" optimum values of ${}_s\Delta_n$ cannot always be found by minimizing the asymptotic MSE of $\hat{\tau}^2_{n,OL}$ or $\hat{\tau}^2_{n,NOL}$, even for variance estimation of some desirable statistics on geometrically "simple" sampling and subsampling regions. Consider estimating the variance of a real-valued sample mean over a rectangular sampling region in $\mathbb{R}^d$ based on $R_0 = (-1/2, 1/2]^d$, with observations on $\mathbf{Z}^d = \mathbb{Z}^d$. If Assumptions A.1-A.5 and Condition M_1 hold, the leading term in the bias expansion can be shown to be:

$$\text{Bias of } \hat{\tau}^2_{n,OL} = \left(-\sum_{i=1}^{d} \frac{L_i}{{}_s\lambda_i^{(n)}} \right)(1 + o(1)); \quad L_i = \sum_{\substack{\mathbf{k}\in\mathbb{Z}^d \\ \mathbf{k}=(k_1,\ldots,k_d)'}} |k_i| \text{Cov}\big(Z(\mathbf{0}), Z(\mathbf{k})\big).$$

  In using the parenthetical sum above to expand the MSE of $\hat{\tau}^2_{n,OL}$, one finds that the resulting MSE cannot be minimized over the permissible, positive range of ${}_s\Delta_n$ if the signs of the L_i values are unequal. That is, for $d > 1$, the subsample estimator MSE cannot always be globally minimized to obtain optimal subsample factors ${}_s\Delta_n$ by considering just the leading order bias terms. An effort to determine and incorporate (into the asymptotic MSE) second or third order bias components quickly becomes intractable, even with rectangular regions.

- The diagonal components of ${}_s\Delta_n$ are asymptotically scalar multiples of each other by Assumption A.1. If so desired, a template choice for R_0 could be used to scale the expansion of the subsampling regions in each direction.

In the continued discussion, we assume

$${}_sR_n = {}_s\lambda_n R_0. \tag{4.1}$$

We frame the components necessary for determining the biases of the spatial subsample variance estimators in the next theorem. Let $C_n(\mathbf{k}) \equiv |\mathbf{Z}^d \cap {}_sR_n \cap (\mathbf{k} + {}_sR_n)|$ denote the number of pairs of observations in the subsampling region ${}_sR_n$ separated by a translate $\mathbf{k} \in \mathbb{Z}^d$.

Theorem 4.1 *Suppose that* $d \geq 2$, ${}_sR_n = {}_s\lambda_n R_0$ *and Assumptions* A.1 - A.5, *Conditions* D_3 *and* M_{3+a} *hold. If, in addition,* ${}_s\lambda_n \in \mathbb{Z}_+$ *for NOL subsamples and*

$$\lim_{n\to\infty} \frac{{}_sN_n - C_n(\mathbf{k})}{({}_s\lambda_n)^{d-1}} = C(\mathbf{k}) \tag{4.2}$$

exists for all $\mathbf{k} \in \mathbb{Z}^d$, *then*

$$\text{E}(\hat{\tau}^2_n) - \tau^2_n = \frac{-1}{{}_s\lambda_n |R_0|} \left(\sum_{\mathbf{k}\in\mathbb{Z}^d} C(\mathbf{k})\sigma(\mathbf{k}) \right)(1 + o(1))$$

where $\sigma(\mathbf{k}) = \mathrm{Cov}\big(\nabla' Z(\mathbf{t}), \nabla' Z(\mathbf{t}+\mathbf{k})\big)$ *and where* $\hat{\tau}_n^2$ *is either* $\hat{\tau}_{n,OL}^2$ *or* $\hat{\tau}_{n,NOL}^2$.

Note that the numerator on the left side of (4.2) is the number of integer grid points that lie in the subregion ${}_sR_n$, but not in the translate $\mathbf{k} + {}_sR_n$. Hence, computing the bias above actually requires counting the number of lattice points inside intersections like ${}_sR_n \cap \mathbf{k} + {}_sR_n$, which is difficult in general. To handle the problem, one may attempt to estimate the count $C_n(\mathbf{k})$ with the corresponding Lebesgue volume, $|{}_sR_n \cap \mathbf{k} + {}_sR_n|$, and then quantify the resulting approximation error. The determination of volumes or areas may not be easy either but hopefully more manageable. For example, if R_0 is a circle, the area of ${}_s\lambda_n R_0$ can be readily computed, but the number of $\mathbb{Z}^2$ integers inside ${}_s\lambda_n R_0$ is not so simple and was in fact a famous consideration of Gauss [cf. Krätzel (1988, p. 141)].

We first note that the boundary condition on R_0 provides a general (trivial) bound on the discrepancy between the count $C_n(\mathbf{k})$ and the volume $|{}_sR_n \cap \mathbf{k} + {}_sR_n|$: $O({}_s\lambda_n^{d-1})$. However, the size of the numerator in (4.2) is also $O({}_s\lambda_n^{d-1})$, corresponding to the order of $\mathbf{Z}^d$ lattice points "near" the boundary of ${}_sR_n$. Consequently, a standard $O({}_s\lambda_n^{d-1})$ bound on the volume-count approximation error is too large to immediately justify the exchange of volumes $|{}_sR_n|$, $|{}_sR_n \cap \mathbf{k} + {}_sR_n|$ for counts ${}_sN_n$, $C_n(\mathbf{k})$ in (4.2).

Bounds on the difference between lattice point counts and volumes have received much attention in analytic number theory, which we briefly mention. Research has classically focused on sets outlined by "smooth" simple closed curves in the plane $\mathbb{R}^2$ and on one question in particular [Huxley (1996)]: When a curve with interior area A is 'blown up' by a factor b, how large is the difference between the number of $\mathbb{Z}^2$ integer points inside the new curve and the area b^2A? For convex sets with a *smoothly winding border*, van der Corput's (1920) answer to the posed question above is $O(b^{46/69+\epsilon})$, while the best answer is $O(b^{46/73+\epsilon})$ for curves with sufficiently *differentiable radius of curvature* [Huxley (1993, 1996)]. These types of bounds, however, are invalid for many convex polygonal templates R_0 in $\mathbb{R}^2$ such as triangles, trapezoids, etc., where often the difference between number of $\mathbb{Z}^2$ integer points in ${}_sR_n = {}_s\lambda_n R_0$ and its area is of exact order $O({}_s\lambda_n)$ (set also by the boundary condition on R_0 or the perimeter length of ${}_sR_n$). The problem above, as considered by number theorists, does not directly address counts for intersections between an expanding region and its vector translates, e.g. ${}_sR_n \cap \mathbf{k} + {}_sR_n$.

To eventually compute "closed-form" bias expansions for $\hat{\tau}_{n,OL}^2$, we use approximation techniques for *subtracted* lattice point counts. For each $\mathbf{k} \in \mathbb{Z}^d$, we

1. replace the numerator of (4.2) with the difference of corresponding Lebesgue volumes.
2. show the following error term is of sufficiently small order $o({}_s\lambda_n^{d-1})$:

$$\Big({}_sN_n - C_n(\mathbf{k})\Big) - \Big({}_s\lambda_n^d|R_0| - |{}_sR_n \cap \mathbf{k} + {}_sR_n|\Big) = \Big({}_sN_n - {}_s\lambda_n^d|R_0|\Big) - \Big(C_n(\mathbf{k}) - |{}_sR_n \cap \mathbf{k} + {}_sR_n|\Big).$$

We do approximate the number of lattice points in ${}_sR_n$ and ${}_sR_n \cap \mathbf{k} + {}_sR_n$ by set volumes, though the Lebesgue volume may not adequately capture the lattice point count in either set. However, the *difference* between approximation errors ${}_sN_n - {}_s\lambda_n^d|R_0|$ and $C_n(\mathbf{k}) - |{}_sR_n \cap \mathbf{k} + {}_sR_n|$ can be shown to be asymptotically small enough, for some templates R_0, to justify replacing counts with volumes in (4.2) (see Lemma 9.4). That is, these two volume-count estimation errors can cancel to a sufficient extent when subtracted. The above approach becomes slightly more complicated for NOL subsamples, $\tilde{R}_{\mathbf{i},n} = {}_s\Delta_n(\mathbf{i} + R_0)$, which may vary in number of sampling sites ${}_sN_{\mathbf{i},n}$. In this case, errors incurred by approximating counts $|\mathbf{Z}^d \cap \tilde{R}_{\mathbf{i},n} \cap \mathbf{k} + \tilde{R}_{\mathbf{i},n}|$ with volumes $|\tilde{R}_{\mathbf{i},n} \cap \mathbf{k} + \tilde{R}_{\mathbf{i},n}|$ are shown to be asymptotically negligible, uniformly in $\mathbf{i} \in J_{NOL}$.

In the following theorem, we use this technique to give bias expansions for a large class of sampling regions in $\mathbb{R}^d$, $d \le 3$, which are "nearly" convex. That is, the sampling region R_n differs from a convex set possibly only at its boundary, but sampling sites on the border may be arbitrarily included or excluded from R_n.

Some notation is additionally required. For $\alpha = (\alpha_1, ..., \alpha_p)' \in (\mathbb{Z}_+)^p$, $\mathbf{x} \in \mathbb{R}^p$, write $\mathbf{x}^\alpha = \prod_{i=1}^p x_i^{\alpha_i}$, $\alpha! = \prod_{i=1}^p (\alpha_i!)$, and $c_\alpha = D^\alpha H(\mu)/\alpha!$. Let Z_∞ denote a random vector with a normal $\mathcal{N}(0, \Sigma_\infty)$ distribution on $\mathbb{R}^p$, where Σ_∞ is the limiting covariance matrix of the scaled sample mean $\sqrt{N_n}(\bar{Z}_{N_n} - \mu)$ from (2.1). Let B°, $\overline{B}$ denote the interior and closure of $B \in \mathbb{R}^d$, respectively.

Theorem 4.2 *Suppose* ${}_sR_n = {}_s\lambda_n R_0$ *and there exists a convex set* B *such that* $B^\circ \subset R_0 \subset \overline{B}$. *With Assumptions* A.2 - A.5, *assume Conditions* D_{5-d} *and* M_{5-d+a} *hold for* $d \in \{1, 2, 3\}$. *Then,*

$$C(\mathbf{k}) = V(\mathbf{k}) \equiv \lim_{n\to\infty} \frac{|{}_sR_n| - |{}_sR_n \cap (\mathbf{k} + {}_sR_n)|}{({}_s\lambda_n)^{d-1}}, \quad \mathbf{k} \in \mathbb{Z}^d$$

whenever $V(\mathbf{k})$ *exists and the biases* $\mathrm{E}(\hat{\tau}^2_{n,OL}) - \tau^2_n$, $\mathrm{E}(\hat{\tau}^2_{n,NOL}) - \tau^2_n$ *are equal to:*

for $d = 1$,
$$\frac{-1}{{}_s\lambda_n |R_0|} \left(\sum_{\mathbf{k} \in \mathbb{Z}} |\mathbf{k}| \sigma(\mathbf{k}) \ + \ C_\infty \right) (1 + o(1));$$

for $d = 2$ *or* 3,
$$\left(- \sum_{\mathbf{k} \in \mathbb{Z}^d} \frac{|{}_sR_n| - |{}_sR_n \cap (\mathbf{k} + {}_sR_n)|}{|{}_sR_n|} \sigma(\mathbf{k}) \right) (1 + o(1))$$

$$\text{or } \frac{-1}{{}_s\lambda_n |R_0|} \left(\sum_{\mathbf{k} \in \mathbb{Z}^d} V(\mathbf{k}) \sigma(\mathbf{k}) \right) (1 + o(1)), \ \textit{provided each } V(\mathbf{k}) \textit{ exists;}$$

where $\sigma(\mathbf{k}) = \mathrm{Cov}\big(\nabla' Z(\mathbf{t}), \nabla' Z(\mathbf{t}+\mathbf{k})\big)$ *and* $C_\infty = \mathrm{Var}\left(\sum_{\|\alpha\|_1 = 2} \frac{c_\alpha}{\alpha!} Z_\infty^\alpha \right) + 2 \sum_{\substack{\|\alpha\|_1=1,\\ \|\beta\|_1=3}} \frac{c_\alpha c_\beta}{\beta!} \mathrm{E}\left(Z_\infty^\alpha Z_\infty^\beta \right)$

$$+ 2 \sum_{\mathbf{k}_1, \mathbf{k}_2 \in \mathbb{Z}} \sum_{\substack{\|\alpha\|_1=1,\\ \|\beta\|_1=1, \|\gamma\|_1=1}} \frac{c_\alpha c_{(\beta+\gamma)}}{(\beta+\gamma)!} \mathrm{E}\left(\big[Z(\mathbf{t}) - \mu\big]^\alpha \big[Z(\mathbf{t}+\mathbf{k}_1) - \mu\big]^\beta \big[Z(\mathbf{t}+\mathbf{k}_2) - \mu\big]^\gamma \right).$$

Remark 1: If Condition D_x holds with $\mathcal{C} = 0$ for $x \in \{2, 3, 4\}$, then Condition M_{x-1} is sufficient in Theorem 4.2.

Remark 2: For each $\mathbf{k} \in \mathbb{Z}^d$, the numerator in $V(\mathbf{k})$ is of order $O({}_s\lambda_n^{d-1})$ by the R_0-boundary condition which holds for convex templates. We may then expand the bias of the estimators through the limiting, scaled volume differences $V(\mathbf{k})$. For $d = 1$, with samples and subsamples based on intervals, it can be easily seen that $V(\mathbf{k}) = |\mathbf{k}|$ which appears in Theorem 4.2.

The function $H(\cdot)$ needs to be increasingly "smoother" to determine the bias component of $\hat{\tau}^2_{n,OL}$ or $\hat{\tau}^2_{n,NOL}$ in lower dimensional spaces $d = 1$ or 2. For a real-valued time series sample mean $\hat{\theta}_n = \bar{Z}_n$, the well-known bias of the subsample variance estimators follows from Theorem 4.2 under our sampling framework $R_0 = (-1/2, 1/2]$, $\mathbf{Z} = \mathbb{Z}$:

$$\frac{-1}{{}_s\lambda_n}\left(\sum_{\mathbf{k}\in\mathbb{Z}} |\mathbf{k}| \mathrm{Cov}\left(\nabla' Z(\mathbf{0}), \nabla' Z(\mathbf{k})\right)\right) \tag{4.3}$$

with $\nabla = 1$. In general though, terms in the Taylor expansion of $\hat{\theta}_{\mathbf{i},n}$ (around μ) up to fourth order can contribute to the bias of $\hat{\tau}^2_{n,OL}$ and $\hat{\tau}^2_{n,NOL}$ when $d = 1$. In contrast, the asymptotic bias of the time series MBB variance estimator with "smooth" model statistics is very different from its subsample-based counterpart. The MBB variance estimator's bias is given by (4.3), determined only by the linear component from the Taylor's expansion of $\hat{\theta}_{\mathbf{i},n}$ [cf. Lahiri (1996)].

5 Asymptotically optimal subsample sizes

In the following, we consider "size" selection for the subsampling regions to maximize the large-sample accuracy of the subsample variance estimators. For reasons discussed in Section 4, we examine a theoretically optimal scaling choice ${}_s\lambda_n$ for subregions in (4.1).

5.1 Theoretical optimal subsample sizes

Generally speaking, there is a trade off in the effect of subsample size on the bias and variance of $\hat{\tau}^2_{n,OL}$ or $\hat{\tau}^2_{n,NOL}$. For example, increasing ${}_s\lambda_n$ reduces the bias but increases the variance of the estimators. The best value of ${}_s\lambda_n$ optimizes the over-all performance of a subsample variance estimator by balancing the contributions from both the estimator's variance and bias. An optimal ${}_s\lambda_n$ choice can be found by minimizing the asymptotic order of a variance estimator's MSE under a given OL or NOL sampling scheme.

Theorem 4.1 implies that the bias of the estimators $\hat{\tau}^2_{n,OL}$ and $\hat{\tau}^2_{n,NOL}$ is of exact order $O(1/{}_s\lambda_n)$. For a broad class of sampling regions R_n, the leading order bias component can be determined explicitly with Theorem 4.2. We bring these variance and bias expansions together to obtain an optimal subsample scaling factor ${}_s\lambda_n$.

Theorem 5.1 *Let ${}_sR_n = {}_s\lambda_n R_0$. With Assumptions* A.2 - A.5, *assume Conditions D_2 and M_{5+2a} hold if $d \geq 2$; and Conditions D_3 and D_{7+2a} if $d = 1$. If $B_0 \equiv |R_0|^{-1}\sum_{\mathbf{k}\in\mathbb{Z}^d} C(\mathbf{k})\sigma(\mathbf{k}) \neq 0$, then*

$${}_s\lambda^{\mathrm{opt}}_{n,OL} = \left(\frac{\det(\Delta_n)(B_0)^2}{dK_0\tau^4}\right)^{1/(d+2)}(1+o(1)) \quad \textit{and} \quad {}_s\lambda^{\mathrm{opt}}_{n,NOL} = \left(\frac{\det(\Delta_n)|R_0|(B_0)^2}{d\tau^4}\right)^{1/(d+2)}(1+o(1)).$$

Remark 3: If Condition D_x holds with $\mathcal{C} = 0$ for $x \in \{2, 3\}$, then Condition M_{2x-1} is sufficient.
Remark 4: Theorem 5.1 suggests that optimally scaled OL subsamples should be *larger* than the NOL ones by a scalar: $(K_1)^{-1/(d+2)} > 1$ where $K_1 = K_0|R_0|$ the limiting ratio of variances from (3.1).

It is well-known in the time series case that the OL subsampling scheme produces an asymptotically more efficient variance estimator than its NOL counterpart. We can now quantify the relative efficiency of the two subsampling procedures in d-dimensional sampling space. With each variance estimator respectively optimized using (4.1), $\hat{\tau}^2_{n,OL}$ is more efficient than $\hat{\tau}^2_{n,NOL}$ and the asymptotic relative efficiency of $\hat{\tau}^2_{n,NOL}$ to $\hat{\tau}^2_{n,OL}$ depends solely of the geometry of R_0:

$$ARE_d \;=\; \lim_{n\to\infty} \frac{\mathrm{E}(\hat{\tau}^2_{n,OL} - \tau^2_n)^2}{\mathrm{E}(\hat{\tau}^2_{n,NOL} - \tau^2_n)^2} \;=\; (K_1)^{2/(d+2)} \;<\; 1.$$

Possolo (1991), Politis and Romano (1993a, 1994), Hall and Jing (1996), and Garcia-Soidan and Hall (1997) have examined subsampling with rectangular regions based essentially on $R_0 = (-1/2, 1/2]^d$. Using the geometrical characteristic $K_1 = (\frac{2}{3})^d$ for rectangles, we can now examine the effect of the sampling dimension on the ARE_d of $\hat{\tau}^2_{n,NOL}$ to $\hat{\tau}^2_{n,OL}$ for these sampling regions. Although the ARE_d decreases as the dimension d increases, we find the relative improvement of $\hat{\tau}^2_{n,OL}$ over $\hat{\tau}^2_{n,NOL}$ is ultimately limited and the ARE_d has a lower bound of 4/9 for all $\mathbb{R}^d$-rectangular regions.

6 Examples

We now provide some examples of the important quantities K_0, K_1, B_0 associated with optimal scaling ${}_s\lambda_n^{\mathrm{opt}}$ with some common sampling region templates, determined from Theorems 3.1 and 4.2. For subsamples from (4.1), the theoretically best ${}_s\lambda_n^{\mathrm{opt}}$ can also be formulated in terms of $|R_n| = \det(\Delta_n)|R_0|$ (sampling region volume), K_1, and B_0.

6.1 Examples in $\mathbb{R}^2$

example 1. Rectangular regions in $\mathbb{R}^2$ (potentially rotated):
If $R_0 = \Big\{\big((l_1\cos\theta, l_2\sin\theta)\mathbf{x}, (-l_1\sin\theta, l_2\cos\theta)\mathbf{x}\big)' : \mathbf{x} \in (-1/2, 1/2]^2\Big\}$ for $\theta \in [0,\pi]$, $0 < l_1, l_2$, then

$$K_0 = \frac{4}{9}, \qquad B_0 \;=\; \sum_{\substack{\mathbf{k}\in\mathbb{Z}^2 \\ \mathbf{k}=(k_1,k_2)'}} \left(\frac{|k_1\cos\theta - k_2\sin\theta|}{l_1} + \frac{|k_1\sin\theta + k_2\cos\theta|}{l_2}\right)\sigma(\mathbf{k}).$$

The characteristics K_1, B_0 for determining optimal subsamples based on two rectangular templates are further described in Table 4.

example 2. If R_0 is a circle of radius $r \le 1/2$ centered at the origin, then K_0 appears in Table 3 and $B_0 = 2/(r\pi)\sum_{\mathbf{k}\in\mathbb{Z}^2}\|\mathbf{k}\|\sigma(\mathbf{k})$.

example 3. For *any* triangle, $K_0 = 2/5$. Two examples are provided in Tables 2 and 4.

example 4. For any parallelogram in $\mathbb{R}^2$ with interior angle γ and adjacent sides of ratio $b \ge 1$, $K_0 = 4/9 + 2/15 \cdot b^{-2}|\cos\gamma|(1 - |\cos\gamma|)$. In particular, if a parallelogram R_0 is formed by two vectors $(0, l_1)'$, $(l_2\cos\gamma, l_2\sin\gamma)'$ extended from a point $\mathbf{x} \in (-1/2, 1/2]^2$, then

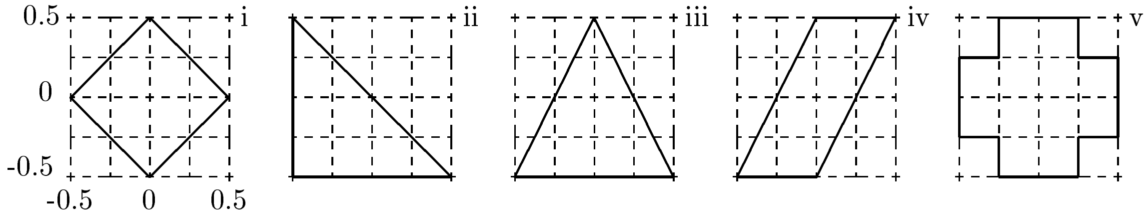
$$B_0 = \frac{1}{|\sin\theta|}\sum_{\mathbf{k}\in\mathbb{Z}^2}\left(\frac{\big|\,k_1\cdot|\cos\theta| - k_2\cdot|\sin\theta|\,\big|}{\max\{l_1,l_2\}} + \frac{|k_2|}{\min\{l_1,l_2\}}\right)\sigma(\mathbf{k}), \quad \gamma\in(0,\pi),\ l_1, l_2 > 0.$$

For further bias term B_0 calculation tools with more general (non-convex) sampling regions and templates R_0 (represented as the union of two approximately convex sets), see Nordman (2002).

Table 4: Examples of several shapes of $R_0 \subset \mathbb{R}^2$ and associated K_1, B_0 for ${}_s\lambda_n^{\text{opt}}$

R_0	K_1	B_0
$(-1/2, 1/2]^2$	$4/9$	$\sum_{\mathbf{k}\in\mathbb{Z}^2} \|\mathbf{k}\|_1 \sigma(\mathbf{k})$
Circle of radius 1/2 at origin	$\pi/4 - 4/(3\pi)$	$4/\pi \sum_{\mathbf{k}\in\mathbb{Z}^2} \|\mathbf{k}\| \sigma(\mathbf{k})$
"Diamond" in Figure 1(i)	$2/9$	$2\sum_{\mathbf{k}\in\mathbb{Z}^2} \|\mathbf{k}\|_\infty \sigma(\mathbf{k})$
Right triangle in Figure 1(ii)	$1/5$	Table 2
Triangle in Figure 1(iii)	$1/5$	$\sum_{\substack{\mathbf{k}\in\mathbb{Z}^2,\\ \lvert k_1\rvert \le \lvert k_2\rvert/2}} 2\lvert k_2\rvert\sigma(\mathbf{k}) + \sum_{\substack{\mathbf{k}\in\mathbb{Z}^2,\\ \lvert k_1\rvert > \lvert k_2\rvert/2}} \big(\lvert k_2\rvert + 2\lvert k_1\rvert\big)\sigma(\mathbf{k})$
Parallelogram in Figure 1(iv)	$2/9 + (\sqrt{5}-1)/375$	$4/\sqrt{5}\sum_{\mathbf{k}\in\mathbb{Z}^2}\big(\lvert k_1 - 2k_2\rvert/5 + \lvert k_2\rvert\big)\sigma(\mathbf{k})$

Figure 1. Examples of templates $R_0 \subset (-1/2, 1/2]^2$ are outlined by solid lines. Cross-shaped sampling regions R_n described in Table 2 are based on R_0 in (v).



6.2 Examples in $\mathbb{R}^d$, $d \geq 3$

example 5. For any sphere, K_0 is given in Table 3. The properties B_0, K_1 of the sphere described in Table 1 and 2 correspond to the template sphere R_0 of radius 1/2 (and maximal volume in $(-1/2, 1/2]^3$).

example 6. The K_0 value for any $\mathbb{R}^3$ cylinder appears in Table 3. If R_0 is a cylinder with circular base (parallel to x-y plane) of radius r and height h, then

$$B_0 = \sum_{\substack{\mathbf{k}\in\mathbf{Z}^3\\ \mathbf{k}=(k_1,k_2,k_3)'}} \left(\frac{|k_3|}{h} + \frac{2\sqrt{k_1^2 + k_2^2}}{\pi r} \right) \sigma(\mathbf{k}).$$

The results of Theorem 4.2 for determining the bias B_0 also seem plausible for convex sampling regions in $\mathbb{R}^d$, $d \geq 4$, but require further study of lattice point counting techniques in higher dimensions. However, bias expansions of the OL and NOL subsample variance estimators are relatively straightforward for an important class of rectangular sampling regions based on the prototype $R_0 = (-1/2, 1/2]^d$, which can then be used in optimal subsample scaling. These hypercubes have "faces" parallel to the coordinate axes which simplifies the task of counting sampling sites, or lattice points, within such regions. We give precise bias expansions in the following theorem, while allowing for potentially missing sampling sites at the borders, or faces, of the sampling region R_n.

Theorem 6.1 *Let* $(-1/2,1/2)^d \subset \Lambda_\ell^{-1}R_0 \subset [-1/2,1/2]^d$, $d \geq 3$ *for a* $d \times d$ *diagonal matrix* Λ_ℓ *with entries* $0 < \ell_i \leq 1$, $i = 1, \ldots, d$. *Suppose* ${}_sR_n = {}_s\lambda_n R_0$ *and Assumptions* A.2 - A.5, *Conditions* D_2 *and* M_{2+a} *hold. Then, the biases* $\mathrm{E}(\hat{\tau}^2_{n,OL}) - \tau^2_n$, $\mathrm{E}(\hat{\tau}^2_{n,NOL}) - \tau^2_n$ *are equal to* $-{}_s\lambda_n^{-1}B_0(1+o(1))$ *where*

$$B_0 = \sum_{\mathbf{k} \in \mathbb{Z}^d} \left(\sum_{i=1}^{d} \frac{|k_i|}{\ell_i} \right) \sigma(\mathbf{k}), \quad \sigma(\mathbf{k}) = \mathrm{Cov}\big(\nabla' Z(\mathbf{t}), \nabla' Z(\mathbf{t}+\mathbf{k})\big).$$

example 7. For rectangular sampling regions $R_n = \Delta_n(-1/2,1/2]^d$, optimal subsamples may be chosen with

$${}_s\lambda^{\mathrm{opt}}_{n,NOL} = \left(\frac{|R_n|}{d\tau^4} \left(\sum_{\mathbf{k} \in \mathbb{Z}^d} \|\mathbf{k}\|_1 \sigma(\mathbf{k}) \right)^2 \right)^{1/(d+2)} (1+o(1)) \quad \text{or} \quad {}_s\lambda^{\mathrm{opt}}_{n,OL} = {}_s\lambda^{\mathrm{opt}}_{n,NOL} \left(\frac{3}{2} \right)^{1/(d+2)}$$

(common directional scaling) using the template $R_0 = (-1/2,1/2]^d$.

7 Empirical subsample size determination

In this section we consider data-based estimation of the theoretical optimal subsample sizes. In particular, we propose methods for estimating the scaling factor ${}_s\lambda_n^{\mathrm{opt}}$ for subsamples in (4.1), which are applicable to both OL and NOL subsampling schemes. Inference on "best" subsample scaling closely resembles the problem of empirically gauging the theoretically optimal block size with the MBB variance estimator, which has been much considered for time series. We first frame the techniques used in this special setting, which can be modified for data-based subsample selection.

Two distinct techniques have emerged for estimating the (time series) MBB block size. One approach involves "plug-in" type estimators of optimal block length [cf. Bühlman and Künsch (1999)]. For subsample inference on $n\mathrm{Var}(\bar{Z}_n)$, Carlstein (1986) (with AR(1) models) and Politis *et al.* (1999) have described "plug-in" estimators for subsample length. The second MBB block selection method, suggested by Hall *et al.* (1995) and Hall and Jing (1996), uses subsampling to create a data-based version of the MSE of the MBB estimator (as a function of block size) which then is minimized to obtain an estimate of optimal block length. With spatial data, Garcia-Soidan and Hall (1997) extended this empirical MSE selection procedure with subsample-based distribution estimators. For variance estimation of a univariate time series sample mean, Lèger *et al.* (1992) proposed a subsample length estimate based on asymptotic MSE considerations; namely, selecting a size by simultaneously solving $({}_s\lambda_n^{\mathrm{opt}})^3 = (3/2)(\hat{B}/\hat{\tau}^2_{n,OL})^2 n$ for ${}_s\lambda_n^{\mathrm{opt}}$ and the associated estimator $\hat{\tau}^2_{n,OL}$, where $\hat{B}$ estimates the parenthetical quantity in (4.3).

A "plug-in" procedure for inference on the theoretical, optimal subsample size involves constructing, and subsequently substituting, (consistent) estimates of unknown population parameters in ${}_s\lambda_n^{\mathrm{opt}}$ from Theorem 5.1. The limiting variance τ^2 appears in the formulation of ${}_s\lambda_n^{\mathrm{opt}}$ and could be approximated with a subsample variance estimate based on some preliminary size choice. The value K_0 can be determined from the available sampling region R_n, but selection of a template R_0 is also required in the "plug-in" approach. The best solution may be to pick a sampling template R_0 to be the largest set of the form $\Lambda_l^{-1}R_n$ for a positive diagonal matrix Λ that fits within $(-1/2,1/2]^d$; this choice appears to reduce the magnitude of the bias of $\hat{\tau}^2_{n,OL}$, $\hat{\tau}^2_{n,NOL}$, which contributes heavily to the MSE of an estimator.

In a spirit similar to the "plug-in" approach, one could empirically select ${}_s\lambda_n^{\text{opt}}$ as in Lèger *et al.* (1992) and Politis and Romano (1993b). After evaluating K_0 and an estimate $\hat{B}_0$ of the leading bias component B_0, compute $\hat{\tau}^2_{n,OL}$ (or analogously $\hat{\tau}^2_{n,NOL}$) for a series of ${}_s\lambda_n$ values and simultaneously solve the asymptotical MSE-based formulation ${}_s\lambda_n^{d+2} = \det(\Delta_n)(\hat{B}_0)^2/\{dK_0(\hat{\tau}^2_{n,OL})^2\}$ for ${}_s\lambda_n$ and an associated $\hat{\tau}^2_{n,OL}$.

8 Proofs for variance expansions

For the proofs, we use $\mathcal{C}$ to denote generic positive constants that do not depend on n or any $\mathbb{Z}^d$ integers (or $\mathbf{Z}^d$ lattice points). The real number r, appearing in some proofs, always assumes the value stated under Condition M_r with respect to the lemma or theorem under consideration. Unless otherwise specified, limits in order symbols are taken letting n tend to infinity.

In the following, we denote the indicator function as $I_{\{\cdot\}}$ (ie. $I_{\{\cdot\}} \in \{0,1\}$ and $I_{\{A\}} = 1$ if and only if an event A holds). For two sequences $\{r_n\}$ and $\{t_n\}$ of positive real numbers, we write $r_n \sim t_n$ if $r_n/t_n \to 1$ as $n \to \infty$. We write $\lambda_n^{\max}$ and ${}_s\lambda_n^{\max}$ for the largest diagonal entries of Δ_n and ${}_s\Delta_n$, respectively, while ${}_s\lambda_{\min}^{(n)} \geq 1$ will denote the smallest diagonal entry of ${}_s\Delta_n$.

We require a few lemmas for the proofs.

Lemma 8.1 *Suppose* $T_1, T_2 \subset \mathbf{Z}^d \equiv \mathbf{t} + \mathbb{Z}^d$ *are bounded. Let* $p, q > 0$ *where* $1/p + 1/q < 1$. *If* X_1, X_2 *are random variables, with* X_i *measurable with respect to* $\mathcal{F}_Z(T_i), i = 1,2$, *then,*

$$|\mathrm{Cov}(X_1, X_2)| \leq 8(\mathrm{E}|X_1|^p)^{1/p}(\mathrm{E}|X_2|^q)^{1/q}\alpha\big(\mathrm{dis}(T_1, T_2); \max_{i=1,2}|T_i|\big)^{1-1/p-1/q},$$

provided expectations are finite and $\mathrm{dis}(T_1, T_2) > 0$.

Proof. Follows from Theorem 3, Doukhan (1994, p. 9). □

Lemma 8.2 *Let* $r \in \mathbb{Z}_+$. *Under Assumption* A.3 *and Condition* M_r, *for* $1 \leq m \leq 2r$ *and any* $T \subset \mathbf{Z}^d \equiv \mathbf{t} + \mathbb{Z}^d$,

$$\mathrm{E}\Big\|\sum_{\mathbf{s}\in T}\big(Z(\mathbf{s}) - \mu\big)\Big\|^m \leq \mathcal{C}(\alpha)|T|^{m/2};$$

$\mathcal{C}(\alpha)$ *is a constant that depends only on the coefficients* $\alpha(k,l)$, $l \leq 2r$, *and* $\mathrm{E}\|Z(\mathbf{t})\|^{2r+\delta}$.

Proof: This follows from Theorem 1, Doukhan (1994, p. 26-31) and Jensen's inequality. □

We next determine the asymptotic sizes of important sets relevant to the sampling or subsampling designs.

Lemma 8.3 *Under Assumptions* A.1-A.2, *the number of sampling sites within*
(a) the sampling region R_n: $N_n = |R_n \cap \mathbf{Z}^d| \sim |R_0| \cdot \det(\Delta_n)$;
(b) an OL subsample, $R_{\mathbf{i},n}$, $\mathbf{i} \in J_{OL}$: ${}_sN_n \sim |R_0| \cdot \det({}_s\Delta_n)$;
(c) a NOL subsample, $\tilde{R}_{\mathbf{i},n}$, $\mathbf{i} \in J_{NOL}$: ${}_sN_{\mathbf{i},n} \sim |R_0| \cdot \det({}_s\Delta_n)$.
The number of
(d) OL subsamples within $R_n : |J_{OL}| \sim |R_0| \cdot \det(\Delta_n)$;
(e) NOL subsamples within $R_n : |J_{NOL}| \sim |R_0| \cdot \det(\Delta_n) \cdot \det({}_s\Delta_n)^{-1}$.
(f) sampling sites near the border of a subsample, $R_{\mathbf{i},n}$ *or* $\tilde{R}_{\mathbf{i},n}$, *is less than*
$\sup_{\mathbf{i}\in\mathbb{Z}^d}|\{\mathbf{j} \in \mathbb{Z}^d : T^{\mathbf{j}} \cap \overline{R_{\mathbf{i},n}} \neq \varnothing, T^{\mathbf{j}} \cap \overline{R^c_{\mathbf{i},n}} \neq \varnothing$ *for* $T^{\mathbf{j}} = \mathbf{j} + [-2,2]^d\}| \leq \mathcal{C}({}_s\lambda_n^{\max})^{d-1}$.

Proof: Results follow from the boundary condition on R_0; see Nordman (2002) for more details.

We require the next lemma for counting the number of subsampling regions which are separated by an appropriately "small" integer translate; we shall apply this lemma in the proof of Theorem 3.1. For $\mathbf{k} = (k_1, \ldots, k_d)' \in \mathbb{Z}^d$, define the following sets,

$$J_n(\mathbf{k}) = |\{\mathbf{i} \in J_{OL} : \mathbf{i} + \mathbf{k} + {}_s\Delta_n R_0 \subset R_n\}|, \quad E_n = \{\mathbf{k} \in \mathbb{Z}^d : |k_j| \le {}_s\lambda_j^{(n)},\ j = 1, \ldots, d\}.$$

Lemma 8.4 *Under Assumption* A.2*,*

$$\max_{\mathbf{k} \in E_n} \left| 1 - \frac{J_n(\mathbf{k})}{|J_{OL}|} \right| \ = \ o(1).$$

Proof: For $\mathbf{k} \in E_n$, write the set $J_n^*(\mathbf{k})$ and bound its cardinality:

$$\begin{aligned} J_n^*(\mathbf{k}) \ &= \ |\{\mathbf{i} \in J_{OL} : (\mathbf{i} + \mathbf{k} + {}_s\Delta_n R_0) \cap \Delta_n R_0^c \neq \varnothing\}| \\ &\le \ |\{\mathbf{i} \in \mathbb{Z}^d : T^{\mathbf{i}} \cap \Delta_n \overline{R_0^c} \neq \varnothing, T^{\mathbf{i}} \cap \Delta_n \overline{R_0} \neq \varnothing;\ T^{\mathbf{i}} = \mathbf{i} + {}_s\lambda_n^{\max}[-2,2]^d\}| \ \le \ \mathcal{C}_s\lambda_n^{\max}(\lambda_n^{\max})^{d-1}, \end{aligned}$$

by the boundary condition on R_0. We have then that for all $\mathbf{k} \in E_n$,

$$|J_{OL}| \ge J_n(\mathbf{k}) = |J_{OL}| - J_n^*(\mathbf{k}) \ge |J_{OL}| - \mathcal{C}_s\lambda_n^{\max}(\lambda_n^{\max})^{d-1}.$$

By Assumption A.2 and the growth rate of $|J_{OL}|$ from Lemma 8.3, the proof is complete. □

We now provide a theorem which captures the main contribution to the asymptotic variance expansion of the OL subsample variance estimator $\hat{\tau}^2_{n,OL}$ from Theorem 3.1.

Theorem 8.1 *For* $\mathbf{i} \in \mathbb{Z}^d$*, let* $Y_{\mathbf{i},n} = \nabla'(Z_{\mathbf{i},n} - \mu)$*. Under the Assumptions and Conditions of Theorem* 3.1*,*

$${}_sN_n \sum_{\mathbf{k} \in E_n} \mathrm{Cov}(Y^2_{\mathbf{0},n}, Y^2_{\mathbf{k},n}) \ = \ K_0 \cdot [2\tau^4](1 + o(1)),$$

where the constant K_0 *is defined in Theorem* 3.1.

Proof: We give only a sketch of the important features; for more details, see Nordman (2002). For a set $T \subset \mathbb{R}^d$, define the function $\overline{\Sigma}(\cdot)$ as

$$\overline{\Sigma}(T) = \sum_{\mathbf{s} \in \mathbf{Z}^d \cap T} \nabla'\big(Z(\mathbf{s}) - \mu\big).$$

With the set intersection $R^{(\mathrm{I})}_{\mathbf{k},n} = {}_sR_n \cap (\mathbf{k} + {}_sR_n)$, $\mathbf{k} \in \mathbb{Z}^d$, write functions:

$$H_{1n}(\mathbf{k}) \ = \ \overline{\Sigma}(R_{\mathbf{k},n} \setminus R^{(\mathrm{I})}_{\mathbf{k},n}), \quad H_{2n}(\mathbf{k}) \ = \ \overline{\Sigma}(R_{\mathbf{0},n} \setminus R^{(\mathrm{I})}_{\mathbf{k},n}), \quad H_{3n}(\mathbf{k}) \ = \ \overline{\Sigma}(R^{(\mathrm{I})}_{\mathbf{k},n}).$$

These represent, respectively, sums over sites in: $R_{\mathbf{k},n}$ but not $R_{\mathbf{0},n} = {}_sR_n$; $R_{\mathbf{0},n}$ but not $R_{\mathbf{k},n}$; and both $R_{\mathbf{0},n}$ and $R_{\mathbf{k},n}$. Then define $h_n(\cdot) : \mathbb{Z}^d \to \mathbb{R}$ as

$$h_n(\mathbf{k}) = \mathrm{E}[H^2_{1n}(\mathbf{k})]\mathrm{E}[H^2_{2n}(\mathbf{k})] + \mathrm{E}[H^2_{1n}(\mathbf{k})]\mathrm{E}[H^2_{3n}(\mathbf{k})] + \mathrm{E}[H^2_{2n}(\mathbf{k})]\mathrm{E}[H^2_{3n}(\mathbf{k})] + \mathrm{E}[H^4_{3n}(\mathbf{k})] - ({}_sN_n)^4[\mathrm{E}(Y^2_{\mathbf{0},n})]^2.$$

We will make use of the following proposition.

Proposition 8.1 *Under the Assumptions and Conditions of Theorem* 3.1*,*

$$\max_{\mathbf{k} \in E_n} \left| ({}_sN_n)^2 \mathrm{Cov}(Y^2_{\mathbf{0},n}, Y^2_{\mathbf{k},n}) - ({}_sN_n)^{-2} h_n(\mathbf{k}) \right| = o(1).$$

The proof of Proposition 8.1 can be found in Nordman (2002) and involves cutting out $\mathbf{Z}^d$ lattice points near the borders of $R_{\mathbf{0},n}$ and $R_{\mathbf{k},n}$, (say) $B_n^{\mathbf{0}}$ and $B_n^{\mathbf{k}}$ with

$$\ell_n = \lfloor ({}_s\lambda_{\min}^{(n)})^e \rfloor, \quad B_{\mathbf{j},n} = \big\{\mathbf{i} \in \mathbf{Z}^d : \mathbf{i} \in \overline{R_{\mathbf{j},n}}, (\mathbf{i} + \ell_n(-1,1]^d) \cap \overline{R_{\mathbf{j},n}^c} \neq \varnothing\big\}, \quad \mathbf{j} \in \mathbb{Z}^d,$$

where $e = (\kappa\delta/\{(2r+\delta)(2r-1-1/d)\}+1)/2 < 1$ from Condition M_r. Here $\ell_n \to \infty$, $\ell_n = o({}_s\lambda_{\min}^{(n)})$ is chosen so that the remaining observations in $R_{\mathbf{0},n}, R_{\mathbf{k},n}, R_{\mathbf{k},n}^{(\mathrm{I})}$ are nearly independent upon removing $B_{\mathbf{0},n}, B_{\mathbf{k},n}$ points and, using the R_0-boundary condition, the set cardinalities $|B_{\mathbf{0},n}|, |B_{\mathbf{k},n}| \le \mathcal{C}\ell_n({}_s\lambda_n^{\max})^{d-1}$ are of smaller order than ${}_sN_n$ (namely, these sets are asymptotically negligible in size).

By Proposition 8.1 and $|E_n| = O({}_sN_n)$, we have

$$\Big|{}_sN_n \sum_{\mathbf{k}\in E_n} \mathrm{Cov}(Y_{\mathbf{0},n}^2, Y_{\mathbf{k},n}^2) - ({}_sN_n)^{-3} \sum_{\mathbf{k}\in E_n} h_n(\mathbf{k})\Big| = o(1). \tag{8.1}$$

Consequently, we need only focus on $({}_sN_n)^{-3}\sum_{\mathbf{k}\in E_n} h_n(\mathbf{k})$ to complete the proof of Theorem 8.1.

For measurability reasons, we create a set defined in terms of the $\mathbb{R}^d$ Lebesgue measure:

$$\mathrm{E}^+ \equiv (0,1) \cap \Big\{\epsilon < \frac{\det(\Delta_0)|R_0|}{2} : \Big|\Big\{\mathbf{x} \in \mathbb{R}^d : |(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0| = \epsilon \textit{ or } \det(\Delta_0)|R_0| - \epsilon\Big\}\Big| = 0\Big\}.$$

Note the set $\{0 < \epsilon < \min\{1, (\det(\Delta_0)|R_0|)/2\} : \epsilon \notin \mathrm{E}^+\}$ is at most countable [cf. Billingsley (1986), Theorem 10.4]. For $\epsilon \in \mathrm{E}^+$, define a new set as a function of ϵ and n:

$$\tilde{R}_{\epsilon,n} = \{\mathbf{k} \in \mathbb{Z}^d : |R_{\mathbf{k},n}^{(\mathrm{I})}| > \epsilon({}_s\lambda_1^{(n)})^d, \, |{}_sR_n \setminus R_{\mathbf{k},n}^{(\mathrm{I})}| > \epsilon({}_s\lambda_1^{(n)})^d\}.$$

Here $\tilde{R}_{\epsilon,n} \subset E_n$ because $\mathbf{k} \notin E_n$ implies $R_{\mathbf{k},n}^{(\mathrm{I})} = \varnothing$.

We now further simplify $({}_sN_n)^{-3}\sum_{\mathbf{k}\in E_n} h_n(\mathbf{k})$ using the following proposition involving $\tilde{R}_{\epsilon,n}$.

Proposition 8.2 *There exist* $\mathrm{N} \in \mathbb{Z}_+$ *and a function* $b(\cdot) : \mathrm{E}^+ \to (0,\infty)$ *such that* $b(\epsilon) \downarrow 0$ *as* $\epsilon \downarrow 0$ *and*

$$({}_sN_n)^{-3}\Big|\sum_{\mathbf{k}\in E_n} h_n(\mathbf{k}) - \sum_{\mathbf{k}\in \tilde{R}_{\epsilon,n}} h_n(\mathbf{k})\Big| \le \mathcal{C}\Big(\epsilon + ({}_s\lambda_1^{(n)})^{-1} + [b(\epsilon)]^d\Big), \tag{8.2}$$

where $\mathcal{C} > 0$ *does not depend on* $\epsilon \in \mathrm{E}^+$ *or* $n \ge \mathrm{N}$.

The proof of Proposition 8.2 is tedious and given in Nordman (2002). The argument involves bounding the sum of $h_n(\cdot)$ over two separate sets in E_n: those integers in E_n that are either "too large" or "too small" in magnitude to be included in $\tilde{R}_{\epsilon,n}$.

To finish the proof, our approach (for an arbitrary $\epsilon \in \mathrm{E}^+$) will be to write $({}_sN_n)^{-3}\sum_{\mathbf{k}\in\tilde{R}_{\epsilon,n}} h_n(\mathbf{k})$ as an integral of a step function with respect to the Lebesgue measure of a step function, say $f_{\epsilon,n}(\mathbf{x})$, then show $\lim_{n\to\infty} f_{\epsilon,n}(\mathbf{x})$ exists almost everywhere (*a.e.*) on $\mathbb{R}^d$, and apply the Lebesgue Dominated Convergence Theorem (LDCT). By letting $\epsilon \downarrow 0$, we will obtain the limit of ${}_sN_n \sum_{\mathbf{k}\in E_n} \mathrm{Cov}(Y_{\mathbf{0},n}^2, Y_{\mathbf{k},n}^2)$.

Fix $\epsilon \in \mathrm{E}^+$. With counting arguments based on the boundary condition of R_0 and the definition of $\tilde{R}_{\epsilon,n}$, it holds that for some $\mathrm{N}_\epsilon \in \mathbb{Z}_+$ and all $\mathbf{k} \in \tilde{R}_{\epsilon,n}$: $|R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d| \ge 1$ and ${}_sN_n - |R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d| \ge 1$ when $n \ge \mathrm{N}_\epsilon$. We can rewrite $({}_sN_n)^{-2}h_n(\mathbf{k})$, $\mathbf{k} \in \tilde{R}_{\epsilon,n}$, in the well-defined form (for $n \ge \mathrm{N}_\epsilon$):

$$\begin{aligned}
\frac{h_n(\mathbf{k})}{({}_sN_n)^2} &= \mathrm{E}\left[\frac{H_{1n}^2(\mathbf{k})}{{}_sN_n - |R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d|}\right]\mathrm{E}\left[\frac{H_{2n}^2(\mathbf{k})}{{}_sN_n - |R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d|}\right]\left(1 - \frac{|R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d|}{{}_sN_n}\right)^2 \\
&\quad + \sum_{j=1}^{2}\mathrm{E}\left[\frac{H_{jn}^2(\mathbf{k})}{{}_sN_n - |R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d|}\right]\mathrm{E}\left[\frac{H_{3n}^2(\mathbf{k})}{|R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d|}\right]\left(1 - \frac{|R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d|}{{}_sN_n}\right)\left(\frac{|R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d|}{{}_sN_n}\right) \\
&\quad + \mathrm{E}\left[\frac{H_{3n}^4(\mathbf{k})}{|R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d|^2}\right]\left(\frac{|R_{\mathbf{k},n}^{(\mathrm{I})} \cap \mathbf{Z}^d|}{{}_sN_n}\right)^2 - [{}_sN_n\mathrm{E}(Y_{\mathbf{0},n}^2)]^2.
\end{aligned}$$

For $\mathbf{x} = (x_1, ..., x_d)' \in \mathbb{R}^d$, write $\lfloor \mathbf{x} \rfloor = (\lfloor x_1 \rfloor, \ldots, \lfloor x_d \rfloor)' \in \mathbb{Z}^d$. Let $f_{\epsilon,n}(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}$ be the step function defined as

$$f_{\epsilon,n}(\mathbf{x}) \;=\; ({}_sN_n)^{-2} I_{\{\lfloor {}_s\lambda_1^{(n)} \mathbf{x} \rfloor \in \tilde{R}_{\epsilon,n}\}} h_n(\lfloor {}_s\lambda_1^{(n)} \mathbf{x} \rfloor).$$

We have then that (with the same fixed $\epsilon \in \mathrm{E}^+$):

$$\frac{1}{{}_sN_n} \sum_{\mathbf{k} \in \tilde{R}_{\epsilon,n}} ({}_sN_n)^{-2} h_n(\mathbf{k}) = \frac{({}_s\lambda_1^{(n)})^d}{{}_sN_n} \int_{\mathbb{R}^d} f_{\epsilon,n}(\mathbf{x}) d\mathbf{x}. \tag{8.3}$$

We focus on showing:

$$\lim_{n\to\infty} f_{\epsilon,n}(\mathbf{x}) \;=\; f_\epsilon(\mathbf{x}) \;\equiv\; I_{\{\mathbf{x} \in \tilde{R}_\epsilon\}} \left[2\tau^4\right] \left(\frac{|(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0|}{\det(\Delta_0)|R_0|} \right)^2 \quad a.e.\ \mathbf{x} \in \mathbb{R}^d, \tag{8.4}$$

with $\tilde{R}_\epsilon = \{\mathbf{x} \in \mathbb{R}^d : |(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0| > \epsilon, |\Delta_0 R_0 \setminus (\mathbf{x} + \Delta_0 R_0)| > \epsilon\}$, a Borel measurable set.

Write $\mathbf{x}_n = \lfloor {}_s\lambda_1^{(n)} \mathbf{x} \rfloor$, $\mathbf{x} \in \mathbb{R}^d$, to ease the notation. To establish (8.4), we begin by showing convergence of indicator functions:

$$I_{\{\mathbf{x}_n \in \tilde{R}_{\epsilon,n}\}} \to I_{\{\mathbf{x} \in \tilde{R}_\epsilon\}} \quad a.e.\ \mathbf{x} \in \mathbb{R}^d. \tag{8.5}$$

Define the sets $A_n(\mathbf{x}) = ({}_s\lambda_1^{(n)})^{-1}\{(\mathbf{x}_n + {}_sR_n) \cap {}_sR_n\}$, $\tilde{A}_n(\mathbf{x}) = \{({}_s\lambda_1^{(n)})^{-1} {}_sR_n\} \setminus A_n(\mathbf{x})$ as a function of $\mathbf{x} \in \mathbb{R}^d$. The LDCT can be applied to show: for each $\mathbf{x} \in \mathbb{R}^d$, $|A_n(\mathbf{x})| \to |(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0|$ and $|\tilde{A}_n(\mathbf{x})| \to |\Delta_0 \mathrm{R}_0 \setminus (\mathbf{x} + \Delta_0 R_0)|$. Thus, if $\mathbf{x} \in \tilde{R}_\epsilon$, then

$$|A_n(\mathbf{x})| \;\to\; |(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0| \;>\; \epsilon, \quad |\tilde{A}_n(\mathbf{x})| \to |\Delta_0 R_0 \setminus (\mathbf{x} + \Delta_0 R_0)| \;>\; \epsilon, \tag{8.6}$$

implying further that $1 = I_{\{\mathbf{x}_n \in \tilde{R}_{\epsilon,n}\}} \to I_{\{\mathbf{x} \in \tilde{R}_\epsilon\}} = 1$. Now consider $\tilde{R}_\epsilon^c$. If $\mathbf{x} \notin \tilde{R}_\epsilon$ such that $|(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0| < \epsilon$ (or $|\Delta_0 R_0 \setminus (\mathbf{x} + \Delta_0 R_0)| < \epsilon$), then $|A_n(\mathbf{x})| < \epsilon$ (or $|\tilde{A}_n(\mathbf{x})| < \epsilon$) eventually for large n and $0 = I_{\{\mathbf{x}_n \in \tilde{R}_{\epsilon,n}\}} \to I_{\{\mathbf{x} \in \tilde{R}_\epsilon\}} = 0$ in this case. Finally, $\epsilon \in \mathrm{E}^+$ implies that a last possible subset of $\tilde{R}_\epsilon^c$ has Lebesgue measure zero; namely, $\left|\{\mathbf{x} \in \tilde{R}_\epsilon^c : |(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0| = \epsilon \textit{ or } |\Delta_0 R_0 \setminus (\mathbf{x} + \Delta_0 R_0)| = \epsilon\}\right| = 0$. We have now proven (8.5).

We next establish a limit for $({}_sN_n)^{-2} h_n(\mathbf{x}_n)$, $\mathbf{x} \in \tilde{R}_\epsilon$. We wish to show:

$$\frac{|R^{(\mathrm{I})}_{\mathbf{x}_n,n} \cap \mathbf{Z}^d|}{{}_sN_n} \;\to\; \frac{|(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0|}{\det(\Delta_0)|R_0|}, \quad \mathbf{x} \in \tilde{R}_\epsilon. \tag{8.7}$$

Using the bound $\big|\,|R^{(\mathrm{I})}_{\mathbf{x}_n,n}| - |R^{(\mathrm{I})}_{\mathbf{x}_n,n} \cap \mathbf{Z}^d|\,\big| \le \mathcal{C}({}_s\lambda_n^{\max})^{d-1}$ from the R_0-boundary condition and noting the limit in (8.6), we find: $({}_s\lambda_1^{(n)})^{-d}|R^{(\mathrm{I})}_{\mathbf{x}_n,n} \cap \mathbf{Z}^d| \;\to\; |(\mathbf{x} + \Delta_0 R_0) \cap \Delta_0 R_0|$, $\mathbf{x} \in \tilde{R}_\epsilon$. By this and $({}_s\lambda_1^{(n)})^d / {}_sN_n \to (\det(\Delta_0)|R_0|)^{-1}$, (8.7) follows.

We can also establish: for each $\mathbf{x} \in \tilde{R}_\epsilon$, $j = 1$ or 2,

$$\mathrm{E}\left[\frac{H_{3n}^{2j}(\mathbf{x}_n)}{|R^{(\mathrm{I})}_{\mathbf{x}_n,n} \cap \mathbf{Z}^d|^j}\right] \to \mathrm{E}([\nabla' Z_\infty]^{2j}), \quad \mathrm{E}\left[\frac{H_{jn}^2(\mathbf{x}_n)}{{}_sN_n - |R^{(\mathrm{I})}_{\mathbf{x}_n,n} \cap \mathbf{Z}^d|}\right], {}_sN_n \mathrm{E}(Y^2_{\mathbf{0},n}) \to \mathrm{E}([\nabla' Z_\infty]^2) \tag{8.8}$$

where $\nabla' Z_\infty$ is a normal $\mathcal{N}(0, \tau^2)$ random variable and so it follows $\mathrm{E}([\nabla' Z_\infty]^{2j}) = (2j-1)\tau^{2j}$, $j = 1, 2$. The limits in (8.8) follow essentially from the Central Limit Theorem (CLT) of Bolthausen (1982), after verifying that the CLT can be applied; see Nordman (2002) for more details.

Putting (8.5), (8.7), and (8.8) together, we have shown the ($a.e.$) convergence of the univariate functions $f_{\epsilon,n}(\mathbf{x})$ as in (8.4). For $\mathbf{k} \in E_n$ and $n \ge \mathrm{N}_\epsilon$, Lemma 8.2 ensures: $({}_sN_n)^{-2}|h_n(\mathbf{k})| \;\le\; \mathcal{C}$, implying that for $\mathbf{x} \in \mathbb{R}^d$: $|f_{\epsilon,n}(\mathbf{x})| \le \mathcal{C} I_{\{\mathbf{x} \in [-c,c]^d\}}$ for some $c > 0$ by Assumption A.1. With this uniform bound on $f_{\epsilon,n}(\cdot)$ and the limits in (8.4), we can apply the LDCT to get

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} f_{\epsilon,n}(\mathbf{x}) d\mathbf{x} \;=\; \int_{\mathbb{R}^d} f_\epsilon(\mathbf{x}) d\mathbf{x}, \quad \epsilon \in \mathrm{E}^+. \tag{8.9}$$

Let $\{\epsilon_m\}_{m=1}^{\infty} \subset \mathrm{E}^+$ where $\epsilon_m \downarrow 0$. Then, $\tilde{R}_{\epsilon_m} \subset \Delta_0[-1,1]^d$ and $\lim_{m\to\infty} I_{\{\mathbf{x}\in\tilde{R}_{\epsilon_m}\}} \to I_{\{\mathbf{x}\in\tilde{R}_0\}}$ for $\mathbf{x} \neq \mathbf{0} \in \mathbb{R}^d$, with $\tilde{R}_0 = \{\mathbf{x} \in \mathbb{R}^d : 0 < |(\mathbf{x}+\Delta_0 R_0) \cap \Delta_0 R_0| < \det(\Delta_0)|R_0|\}$. Hence, by the LDCT,

$$\lim_{m\to\infty} \int_{\mathbb{R}^d} f_{\epsilon_m}(\mathbf{x})d\mathbf{x} \;=\; \int_{\mathbb{R}^d} f_0(\mathbf{x})d\mathbf{x}, \quad f_0(\mathbf{x}) \equiv I_{\{\mathbf{x}\in\tilde{R}_0\}}[2\tau^4]\left(\frac{|(\mathbf{x}+\Delta_0 R_0)\cap\Delta_0 R_0|}{\det(\Delta_0)|R_0|}\right)^2. \tag{8.10}$$

From (8.1)-(8.3), (8.9)-(8.10) and $({}_s\lambda_1^{(n)})^d/{}_sN_n \to (\det(\Delta_0)|R_0|)^{-1}$, we have that:

$$\begin{aligned}
&\overline{\lim}_{n\to\infty}\left|{}_sN_n\sum_{\mathbf{k}\in E_n}\mathrm{Cov}(Y_{\mathbf{0},n}^2, Y_{\mathbf{k},n}^2) - \frac{1}{\det(\Delta_0)|R_0|}\int_{\mathbb{R}^d} f_0(\mathbf{x})d\mathbf{x}\right| \\
\leq\;& \overline{\lim}\left|{}_sN_n\sum_{\mathbf{k}\in E_n}\mathrm{Cov}(Y_{\mathbf{0},n}^2, Y_{\mathbf{k},n}^2) - \frac{({}_s\lambda_1^{(n)})^d}{{}_sN_n}\int_{\mathbb{R}^d} f_{\epsilon_m,n}(\mathbf{x})d\mathbf{x}\right| + \frac{1}{\det(\Delta_0)|R_0|}\left|\int_{\mathbb{R}^d} f_{\epsilon_m}(\mathbf{x}) - f_0(\mathbf{x})d\mathbf{x}\right| \\
&+ \overline{\lim}\left|\frac{({}_s\lambda_1^{(n)})^d}{{}_sN_n}\int_{\mathbb{R}^d} f_{\epsilon_m,n}(\mathbf{x})d\mathbf{x} - \frac{1}{\det(\Delta_0)|R_0|}\int_{\mathbb{R}^d} f_{\epsilon_m}(\mathbf{x})d\mathbf{x}\right| \\
\leq\;& \mathcal{C}\Big(\epsilon_m + [b(\epsilon_m)]^d\Big) + \frac{1}{\det(\Delta_0)|R_0|}\left|\int_{\mathbb{R}^d} f_{\epsilon_m}(\mathbf{x}) - f_0(\mathbf{x})d\mathbf{x}\right| \;\to\; 0 \quad \text{as } \epsilon_m \downarrow 0.
\end{aligned}$$

Finally,

$$\frac{1}{\det(\Delta_0)|R_0|}\int_{\mathbb{R}^d} f_0(\mathbf{x})d\mathbf{x} \;=\; \frac{2\tau^4}{|R_0|}\int_{\mathbb{R}^d}\frac{|(\mathbf{y}+R_0)\cap R_0|^2}{|R_0|^2}d\mathbf{y},$$

using a change of variables $\mathbf{y} = \Delta_0^{-1}\mathbf{x}$. This completes the proof of Theorem 8.1. □

For clarity of exposition, we will prove Theorem 3.1, parts (a) and (b), separately for the OL and NOL subsample variance estimators.

8.1 Proof of Theorem 3.1(a)

For $\mathbf{i} \in J_{OL}$, we use a Taylor's expansion of $H(\cdot)$ (around μ) to rewrite the statistic $\hat{\theta}_{\mathbf{i},n} = H(Z_{\mathbf{i},n})$:

$$\begin{aligned}
\hat{\theta}_{\mathbf{i},n} \;&=\; H(\mu) + \sum_{\|\alpha\|_1=1} c_\alpha (Z_{\mathbf{i},n}-\mu)^\alpha + 2\sum_{\|\alpha\|_1=2}\frac{(Z_{\mathbf{i},n}-\mu)^\alpha}{\alpha!}\int_0^1 (1-\omega)D^\alpha H(\mu+\omega(Z_{\mathbf{i},n}-\mu))d\omega \\
&\equiv\; H(\mu) \,+\, Y_{\mathbf{i},n} \,+\, Q_{\mathbf{i},n}.
\end{aligned} \tag{8.11}$$

We also have: $\tilde{\theta}_n \;=\; H(\mu) + |J_{OL}|^{-1}\sum_{\mathbf{i}\in J_{OL}} Y_{\mathbf{i},n} + |J_{OL}|^{-1}\sum_{\mathbf{i}\in J_{OL}} Q_{\mathbf{i},n} \;\equiv\; H(\mu) + \bar{Y}_n + \bar{Q}_n$. Then,

$$\hat{\tau}_{n,OL}^2 = {}_sN_n\left[\frac{1}{|J_{OL}|}\sum_{\mathbf{i}\in J_{OL}} Y_{\mathbf{i},n}^2 \,+\, \frac{1}{|J_{OL}|}\sum_{\mathbf{i}\in J_{OL}} Q_{\mathbf{i},n}^2 \,+\, \frac{2}{|J_{OL}|}\sum_{\mathbf{i}\in J_{OL}} Y_{\mathbf{i},n}Q_{\mathbf{i},n} \,-\, \bar{Y}_n^2 \,-\, \bar{Q}_n^2 \,-\, 2(\bar{Y}_n)(\bar{Q}_n)\right].$$

We establish Theorem 3.1(a) in two parts by showing

$$\begin{aligned}
&(a)\quad \mathrm{Var}\left(\frac{{}_sN_n}{|J_{OL}|}\sum_{\mathbf{i}\in J_{OL}} Y_{\mathbf{i},n}^2\right) = K_0\cdot\frac{\det({}_s\Delta_n)}{\det(\Delta_n)}\cdot[2\tau^4](1+o(1)), \\
&(b)\quad \left|\mathrm{Var}(\hat{\tau}_{n,OL}^2) - \mathrm{Var}\left(\frac{{}_sN_n}{|J_{OL}|}\sum_{\mathbf{i}\in J_{OL}} Y_{\mathbf{i},n}^2\right)\right| \;=\; o\left(\frac{\det({}_s\Delta_n)}{\det(\Delta_n)}\right).
\end{aligned} \tag{8.12}$$

We will begin with proving (8.12)(a) above. For $\mathbf{k} \in \mathbb{Z}^d$, let $\sigma_n(\mathbf{k}) = \mathrm{Cov}(Y_{\mathbf{0},n}^2, Y_{\mathbf{k},n}^2)$. We write

$$\frac{({}_sN_n)^2}{|J_{OL}|^2}\mathrm{Var}\Big(\sum_{\mathbf{i}\in J_{OL}} Y_{\mathbf{i},n}^2\Big) \;=\; \frac{({}_sN_n)^2}{|J_{OL}|^2}\Big(\sum_{\mathbf{k}\in E_n} J_n(\mathbf{k})\sigma_n(\mathbf{k}) \,+\, \sum_{\mathbf{k}\in\mathbb{Z}^d\setminus E_n} J_n(\mathbf{k})\sigma_n(\mathbf{k})\Big) \;\equiv\; W_{1n} \,+\, W_{2n}.$$

By stationarity and Lemma 8.2, we bound: $|\sigma_n(\mathbf{k})| \le \mathrm{E}(Y_{\mathbf{0},n}^4) \le \mathcal{C}({}_sN_n)^{-2}$, $\mathbf{k} \in \mathbb{Z}^d$. Using this covariance bound, Lemmas 8.3-8.4, and $|E_n| \le 3^d \det({}_s\Delta_n)$:

$$\left|\frac{({}_sN_n)^2}{|J_{OL}|}\sum_{\mathbf{k}\in E_n}\sigma_n(\mathbf{k}) \;-\; W_{1n}\right| \;\le\; \mathcal{C}\cdot\frac{|E_n|}{|J_{OL}|}\cdot\max_{\mathbf{k}\in E_n}\left|1-\frac{J_n(\mathbf{k})}{|J_{OL}|}\right| \;=\; o\left(\frac{\det({}_s\Delta_n)}{\det(\Delta_n)}\right). \tag{8.13}$$

Then applying Theorem 8.1 and Lemma 8.3,

$$\frac{({}_sN_n)^2}{|J_{OL}|}\sum_{\mathbf{k}\in E_n}\sigma_n(\mathbf{k}) \;=\; K_0\cdot\frac{\det({}_s\Delta_n)}{\det(\Delta_n)}\cdot[2\tau^4](1+o(1)). \tag{8.14}$$

By (8.13)-(8.14), we need only show that $W_{2n} = o\left(\det({}_s\Delta_n)/\det(\Delta_n)\right)$ to finish the proof of (8.12)(a).

For $\mathbf{i}\in\mathbb{Z}^d$, denote a set of lattice points within a translated rectangular region:

$$F_{\mathbf{i},n} = \left(\mathbf{i} + \textstyle\prod_{j=1}^d\left(-\lceil {}_s\lambda_j^{(n)}\rceil/2,\, \lceil {}_s\lambda_j^{(n)}\rceil/2\right]\right)\cap\mathbf{Z}^d,$$

where $\lceil\cdot\rceil$ represents the "ceiling" function. Note that for $\mathbf{k} = (k_1,\ldots,k_d)' \in \mathbb{Z}^d\setminus E_n$, there exists $j\in\{1,\ldots,d\}$ such that $|k_j| > {}_s\lambda_j^{(n)}$, implying $\mathrm{dis}(R_{\mathbf{0},n}\cap\mathbf{Z}^d,\, R_{\mathbf{k},n}\cap\mathbf{Z}^d) \;\ge\; \mathrm{dis}(F_{\mathbf{0},n},\, F_{\mathbf{k},n}) \;\ge\; 1$. Hence, sequentially using Lemmas 8.1-8.2, we may bound the covariances $\sigma_n(\mathbf{k})$, $\mathbf{k}\in\mathbb{Z}^d\setminus E_n$, with the mixing coefficient $\alpha(\cdot,\cdot)$:

$$\begin{aligned}
|\sigma_n(\mathbf{k})| \;&\le\; 8\Big[\mathrm{E}(Y_{\mathbf{0},n}^{2(2r+\delta)/r})\Big]^{2r/(2r+\delta)}\alpha(\mathrm{dis}(R_{\mathbf{0},n}\cap\mathbf{Z}^d, R_{\mathbf{k},n}\cap\mathbf{Z}^d),\, {}_sN_n)^{\delta/(2r+\delta)}\\
&\le\; \mathcal{C}({}_sN_n)^{-2}\alpha(\mathrm{dis}(F_{\mathbf{0},n}, F_{\mathbf{k},n}),\, {}_sN_n)^{\delta/(2r+\delta)}.
\end{aligned}$$

From the above bound and $J_n(\mathbf{k})/|J_{OL}| \le 1$, $\mathbf{k}\in\mathbb{Z}^d$, we have

$$\begin{aligned}
|W_{2n}| \;&\le\; \mathcal{C}|J_{OL}|^{-1}\sum_{x=1}^{\infty}\Big(\sum_{j=1}^{d}C_{(x,j,n)}\Big)\alpha(x,{}_sN_n)^{\delta/(2r+\delta)}, \qquad (8.15)\\
C_{(x,j,n)} \;&=\; \Big|\Big\{\mathbf{i}\in\mathbb{Z}^d : \mathrm{dis}(F_{\mathbf{0},n},F_{\mathbf{i},n}) = x = \inf\{|v_j - w_j| : \mathbf{v}\in F_{\mathbf{0},n}, \mathbf{w}\in F_{\mathbf{i},n}\}\Big\}\Big|.
\end{aligned}$$

The function $C_{(x,j,n)}$ counts the number of translated rectangles $F_{\mathbf{i},n}$ that lie a distance of $x\in\mathbb{Z}_+$ from the rectangle $F_{\mathbf{0},n}$, where this distance is realized in the jth coordinate direction for $j = 1,\ldots,d$. For $\mathbf{i}\in\mathbb{Z}^d$, $x\ge 1$ and $j\in\{1,\ldots,d\}$, if $\mathrm{dis}(F_{\mathbf{0},n},F_{\mathbf{i},n}) = x = \inf\{|v_j-w_j| : \mathbf{v}\in F_{\mathbf{0},n}, \mathbf{w}\in F_{\mathbf{i},n}\}$, then $|i_j| = \lceil {}_s\lambda_j^{(n)}\rceil + x - 1$ with the remaining components of $\mathbf{i}$, namely i_m for $m\in\{1\ldots d\}\setminus\{j\}$, constrained by $|i_m| \le {}_s\lambda_m^{(n)} + x$. We use this observation to further bound the right hand side of (8.15) by:

$$\begin{aligned}
&\mathcal{C}|J_{OL}|^{-1}\sum_{x=1}^{\infty}\Big(\sum_{j=1}^{d}\prod_{m=1,j\neq m}^{d}3({}_s\lambda_m^{(n)}+x)\Big)\alpha(x,{}_sN_n)^{\delta/(2r+\delta)}\\
\le\;& \mathcal{C}\cdot\frac{\det({}_s\Delta_n)}{|J_{OL}|}\sum_{j=1}^{d}({}_s\lambda_{\min}^{(n)})^{-j}\Big[\sum_{x=1}^{\ell_n}x^{j-1} \;+\; \sum_{x=\ell_n+1}^{\infty}x^{j-1}[\alpha_1(x)g({}_sN_n)]^{\delta/(2r+\delta)}\Big]\\
\le\;& \mathcal{C}\cdot\frac{\det({}_s\Delta_n)}{|J_{OL}|}\Big[\frac{d\ell_n}{{}_s\lambda_{\min}^{(n)}} \;+\; \frac{\{\ell_n^{1/e}\}^{d\kappa\delta/(2r+\delta)}}{\ell_n^{2rd-d}}\sum_{x=\ell_n+1}^{\infty}x^{2rd-d-1}\alpha_1(x)^{\delta/(2r+\delta)}\Big] = o\left(\frac{\det({}_s\Delta_n)}{\det(\Delta_n)}\right),
\end{aligned}$$

using Assumptions A.1, A.3, Condition M_r, and $\ell_n = o({}_s\lambda_{\min}^{(n)})$. This completes the proof of (8.12)(a).

To establish (8.12)(b), first note that

$$\left|\mathrm{Var}(\hat{\tau}_{n,OL}^2) - \mathrm{Var}\left(\frac{{}_sN_n}{|J_{OL}|}\sum_{\mathbf{i}\in J_{OL}}Y_{\mathbf{i},n}^2\right)\right| \le 4\left(\sum_{j=1}^{5}A_{jn}^{1/2}\right)\left(\sum_{j=1}^{5}A_{jn}^{1/2} + \mathrm{Var}^{1/2}\left(\frac{{}_sN_n}{|J_{OL}|}\sum_{\mathbf{i}\in J_{OL}}Y_{\mathbf{i},n}^2\right)\right),$$

where $A_{1n} = \mathrm{Var}({}_sN_n\bar{Y}_n^2)$, $A_{2n} = \mathrm{Var}(|J_{OL}|^{-1}{}_sN_n\sum_{\mathbf{i}\in J_{OL}} Q_{\mathbf{i},n}^2)$, $A_{3n} = \mathrm{Var}({}_sN_n\bar{Q}_n^2)$, $A_{4n} = \mathrm{Var}(|J_{OL}|^{-1}{}_sN_n\sum_{\mathbf{i}\in J_{OL}} Y_{\mathbf{i},n}Q_{\mathbf{i},n})$, $A_{5n} = \mathrm{Var}({}_sN_n\bar{Y}_n\bar{Q}_n)$.

By (8.12)(a), it suffices to show that $A_{jn} = o(\det({}_s\Delta_n)/\det(\Delta_n))$ for each $j = 1,\ldots,5$. We handle only two terms for illustration: A_{1n}, A_{4n}.

Consider A_{1n}. For $\mathbf{s} \in R_n \cap \mathbf{Z}^d$, let $\omega(\mathbf{s}) = [2^d\det({}_s\Delta_n)]^{-1}|\{\mathbf{i} \in J_{OL} : \mathbf{s} \in \mathbf{i} + {}_s\Delta_n R_0\}|$ so that $0 \le \omega(\mathbf{s}) \le 1$. By Condition M_r and Theorem 3, Doukhan (1994, p.31) (similar to Lemma 8.2),

$$A_{1n} \le \mathrm{E}(\bar{Y}_n^4) = \frac{(2^d\det({}_s\Delta_n))^4}{|J_{OL}|^4({}_sN_n)^2}\mathrm{E}\bigg(\Big[\sum_{\mathbf{s}\in R_n\cap\mathbf{Z}^d}\omega(\mathbf{s})\nabla'\big(Z(\mathbf{s})-\mu\big)\Big]^4\bigg) \le \mathcal{C}\cdot\frac{(N_n)^2(\det({}_s\Delta_n))^4}{|J_{OL}|^4({}_sN_n)^2}. \tag{8.16}$$

Then $A_{1n} = o(\det({}_s\Delta_n)/\det(\Delta_n))$ follows from Lemma 8.3.

To handle A_{4n}, write $\sigma_{1n}(\mathbf{k}) = \mathrm{Cov}(Y_{\mathbf{0},n}Q_{\mathbf{0},n}, Y_{\mathbf{k},n}Q_{\mathbf{k},n})$, $\mathbf{k} \in \mathbb{Z}^d$. Then,

$$A_{4n} = \frac{({}_sN_n)^2}{|J_{OL}|^2}\sum_{\mathbf{k}\in\mathbb{Z}^d} J_n(\mathbf{k})\sigma_{1n}(\mathbf{k}) \le \frac{({}_sN_n)^2}{|J_{OL}|}\bigg(\sum_{\mathbf{k}\in E_n}|\sigma_{1n}(\mathbf{k})| + \sum_{\mathbf{k}\in\mathbb{Z}^d\setminus E_n}|\sigma_{1n}(\mathbf{k})|\bigg) = A_{4n}(E_n) + A_{4n}(E_n^c).$$

For $\mathbf{k} \in E_n$, note $|\sigma_{1n}(\mathbf{k})| \le \mathcal{C}({}_sN_n)^{-3}$ using $|Y_{\mathbf{0},n}Q_{\mathbf{0},n}| \le \mathcal{C}\|Z_{\mathbf{0},n}-\mu\|^3(1+\|Z_{\mathbf{0},n}-\mu\|^a)$ (from Condition D) with Lemmas 8.1-8.2. From this bound, Lemma 8.3, and $|E_n| \le 3^d\det({}_s\Delta_n)$, we find $A_{4n}(E_n) = o(\det({}_s\Delta_n)/\det(\Delta_n))$. We next bound the covariances $\sigma_{1n}(\mathbf{k})$, $\mathbf{k} \in \mathbb{Z}^d \setminus E_n$,

$$\begin{aligned}|\sigma_{1n}(\mathbf{k})| &\le 8\Big[\mathrm{E}(|Y_{\mathbf{0},n}Q_{\mathbf{0},n}|^{(2r+\delta)/r})\Big]^{2r/(2r+\delta)}\alpha(\mathrm{dis}(R_{\mathbf{0},n}\cap\mathbf{Z}^d, R_{\mathbf{k},n}\cap\mathbf{Z}^d), {}_sN_n)^{\delta/(2r+\delta)}\\ &\le \mathcal{C}({}_sN_n)^{-3}\alpha(\mathrm{dis}(F_{\mathbf{0},n}, F_{\mathbf{k},n}), {}_sN_n)^{\delta/(2r+\delta)}\end{aligned}$$

by the stationarity of the random field $Z(\cdot)$ and Lemmas 8.1-8.2. Using this inequality and repeating the same steps used to majorize 'W_{2n}' from the proof of (8.12)(a) (see (8.15)), we have $A_{4n}(E_n^c) = o(\det({}_s\Delta_n)/\det(\Delta_n))$. The proof of Theorem 3.1(a) is finished. □

8.2 Proof of Theorem 3.1(b)

To simplify the counting arguments, we assume here that ${}_s\Delta_n \in \mathbb{Z}_+$, implying ${}_sN_{\mathbf{i},n} = {}_sN_n$, $\mathbf{i} \in \mathbb{Z}^d$. The more general case, in which the NOL subregions may differ in the number of sampling sites, is treated in Nordman (2002).

For each NOL subregion $\tilde{R}_{\mathbf{i},n}$, we denote the corresponding sample mean $\tilde{Z}_{\mathbf{i},n} = ({}_sN_{\mathbf{i},n})^{-1}\sum_{\mathbf{s}\in\tilde{R}_{\mathbf{i},n}\cap\mathbf{Z}^d} Z(\mathbf{s})$. The subsample evaluations of the statistic of interest, $\hat{\theta}_{\mathbf{i},n}$, $\mathbf{i} \in J_{NOL}$, can be expressed through a Taylor's expansion of $H(\cdot)$ around μ, substituting $\tilde{Z}_{\mathbf{i},n}$ for $Z_{\mathbf{i},n}$ in (8.11): $\hat{\theta}_{\mathbf{i},n} = H(\tilde{Z}_{\mathbf{i},n}) = H(\mu) + \tilde{Y}_{\mathbf{i},n} + \tilde{Q}_{\mathbf{i},n}$.

We will complete the proof of Theorem 3.1(b) in two parts by showing

$$\begin{aligned}&(a)\quad \mathrm{Var}\bigg(\frac{{}_sN_n}{|J_{NOL}|}\sum_{\mathbf{i}\in J_{NOL}}\tilde{Y}_{\mathbf{i},n}^2\bigg) = \frac{\det({}_s\Delta_n)}{\det(\Delta_n)|R_0|}\cdot[2\tau^4](1+o(1)),\\ &(b)\quad \bigg|\mathrm{Var}(\hat{\tau}_{n,NOL}^2) - \mathrm{Var}\bigg(\frac{{}_sN_n}{|J_{NOL}|}\sum_{\mathbf{i}\in J_{NOL}}\tilde{Y}_{\mathbf{i},n}^2\bigg)\bigg| = o\bigg(\frac{\det({}_s\Delta_n)}{\det(\Delta_n)}\bigg).\end{aligned} \tag{8.17}$$

We will begin with showing (8.17)(a). For $\mathbf{k} \in \mathbb{Z}^d$, let $\tilde{J}_n(\mathbf{k}) = \{\mathbf{i} \in J_{NOL} : \mathbf{i}+\mathbf{k} \in J_{NOL}\}$ and $\tilde{\sigma}_n(\mathbf{k}) = \mathrm{Cov}(\tilde{Y}_{\mathbf{0},n}, \tilde{Y}_{\mathbf{k},n})$. Then we may express the variance:

$$\begin{aligned}\mathrm{Var}\bigg(\frac{{}_sN_n}{|J_{NOL}|}\sum_{\mathbf{i}\in J_{NOL}}\tilde{Y}_{\mathbf{i},n}^2\bigg) &= \frac{({}_sN_n)^2}{|J_{NOL}|^2}\bigg(\sum_{\substack{\mathbf{k}\in\mathbb{Z}^d\\ 0<\|\mathbf{k}\|_\infty\le 1}}\tilde{J}_n(\mathbf{k})\tilde{\sigma}_n(\mathbf{k}) + \sum_{\substack{\mathbf{k}\in\mathbb{Z}^d\\ \|\mathbf{k}\|_\infty>1}}\tilde{J}_n(\mathbf{k})\tilde{\sigma}_n(\mathbf{k}) + |J_{NOL}|\tilde{\sigma}_n(\mathbf{0})\bigg)\\ &\equiv U_{1n} + U_{2n} + |J_{NOL}|^{-1}({}_sN_n)^2\tilde{\sigma}_n(\mathbf{0}).\end{aligned} \tag{8.18}$$

We first prove $U_{2n} = o(|J_{NOL}|^{-1})$, noting that $\det({}_s\Delta_n)/\det(\Delta_n) = O(|J_{NOL}|^{-1})$ from Lemma 8.3.

When $\mathbf{k} = (k_1, \ldots, k_d)' \in \mathbb{Z}^d$, $\|\mathbf{k}\|_\infty > 1$, then for some $1 \leq m_k \leq d$:

$$\begin{aligned} \mathrm{dis}(\tilde{R}_{\mathbf{0},n} \cap \mathbf{Z}^d,\ \tilde{R}_{\mathbf{k},n} \cap \mathbf{Z}^d) &\geq \mathrm{dis}\Big({}_s\Delta_n[-1/2,\, 1/2]^d,\ {}_s\Delta_n(\mathbf{k} + [-1/2, 1/2]^d)\Big) \\ &= \max_{1 \leq j \leq d} (|k_j| - 1)_s\lambda_j^{(n)} \equiv (|k_{m_k}| - 1)_s\lambda_{m_k}^{(n)}. \end{aligned}$$

If $j \in \{1 \ldots d\}, j \neq m_k$, we have $|k_j| \leq ({}_s\lambda_j^{(n)})^{-1}(|k_{m_k}| - 1)_s\lambda_{m_k}^{(n)} + 1$. Note also if $\mathbf{k} \in \mathbb{Z}^d$, $\|\mathbf{k}\|_\infty > 1$, then $|\tilde{\sigma}(\mathbf{k})| \leq \mathcal{C}({}_sN_n)^{-2}\alpha\big((|k_{m_k}| - 1)_s\lambda_{m_k}^{(n)}, {}_sN_n\big)$ by Lemmas 8.1-8.2. Hence, we have

$$\begin{aligned} |U_{2n}| &\leq \frac{\mathcal{C}}{|J_{NOL}|}\sum_{x=1}^{\infty}\left[\sum_{j=1}^{d}({}_s\lambda_j^{(n)})^{-(d-1)}\prod_{i=1, i\neq j}^{d}({}_s\lambda_i^{(n)}x + {}_s\lambda_j^{(n)})\right]\alpha({}_s\lambda_{\min}^{(n)}x, {}_sN_n)^{\delta/(2r+\delta)} \\ &\leq \frac{\mathcal{C}}{|J_{NOL}|}\sum_{x=1}^{\infty}x^{d-1}\left(\frac{{}_s\lambda_n^{\max}}{{}_s\lambda_{\min}^{(n)}}\right)^{d-1}\alpha({}_s\lambda_{\min}^{(n)}x, {}_sN_n)^{\delta/(2r+\delta)} \\ &\leq \frac{\mathcal{C}}{|J_{NOL}|}\frac{\{\det({}_s\Delta_n)\}^{\kappa\delta/(2r+\delta)}}{({}_s\lambda_{\min}^{(n)})^{2rd-d-1}}\sum_{x=1}^{\infty}({}_s\lambda_{\min}^{(n)}x)^{2rd-d-1}\alpha_1({}_s\lambda_{\min}^{(n)}x)^{\delta/(2r+\delta)} = o\left(\frac{1}{|J_{NOL}|}\right), \end{aligned}$$

by Assumptions A.1, A.3 and Condition M_r.

We now show that $U_{1n} = o(|J_{NOL}|^{-1})$. For $\mathbf{k} \in \mathbb{Z}^d$, $0 < \|\mathbf{k}\|_\infty \leq 1$, define the trimmed set:

$$T_{\mathbf{k},n}^j =: \begin{cases} \{\mathbf{x} \in \mathbb{R}^d : 1/2 \cdot {}_s\lambda_j^{(n)} < x_j \leq 1/2 \cdot {}_s\lambda_j^{(n)} + \ell_n\} & \text{if } k_j = 1, \\ \{\mathbf{x} \in \mathbb{R}^d : -1/2 \cdot {}_s\lambda_j^{(n)} - \ell_n < x_j \leq -1/2 \cdot {}_s\lambda_j^{(n)}\} & \text{if } k_j = -1, \\ \varnothing & \text{if } k_j = 0, \end{cases}$$

for each coordinate direction $j = 1, \ldots, d$. Let $T_{\mathbf{k},n} = \cup_{j=1}^d T_{\mathbf{k},n}^j$. We decompose the sum: ${}_sN_n\tilde{Y}_{\mathbf{k},n} = \overline{\Sigma}(\tilde{R}_{\mathbf{k},n} \setminus T_{\mathbf{k},n}) + \overline{\Sigma}(\tilde{R}_{\mathbf{k},n} \cap T_{\mathbf{k},n}) \equiv \tilde{S}_{\mathbf{k},n} + \tilde{S}_{\mathbf{k},n}^*$. Then, $U_{1n} = o(|J_{NOL}|^{-1})$ follows from (1)-(4) below:

(1) $|\mathrm{E}(\tilde{Y}_{\mathbf{0},n}^2\tilde{S}_{\mathbf{k},n}\tilde{S}_{\mathbf{k},n}^*)| \leq \Big[\mathrm{E}(\tilde{Y}_{\mathbf{0},n}^6)\mathrm{E}(|\tilde{S}_{\mathbf{k},n}|^3)\mathrm{E}(|\tilde{S}_{\mathbf{k},n}^*|^3)\Big]^{1/3} = o(1)$, using Lemma 8.2 and

$$|\tilde{R}_{\mathbf{k},n} \cap T_{\mathbf{k},n} \cap \mathbf{Z}^d| \leq \sum_{j=1}^{d}|\tilde{R}_{\mathbf{k},n} \cap T_{\mathbf{k},n}^j \cap \mathbf{Z}^d| \leq \ell_n \det({}_s\Delta_n)\sum_{j=1}^{d}({}_s\lambda_j^{(n)})^{-1}.$$

(2) Likewise, $\mathrm{E}(\tilde{Y}_{\mathbf{0},n}^2\tilde{S}_{\mathbf{k},n}^{*2}) \leq [\mathrm{E}(\tilde{Y}_{\mathbf{0},n}^4)\mathrm{E}(\tilde{S}_{\mathbf{k},n}^{*4})]^{1/2} = o(1)$.

(3) $|{}_sN_n\mathrm{E}(\tilde{Y}_{\mathbf{k},n}^2) - ({}_sN_n)^{-1}\mathrm{E}(\tilde{S}_{\mathbf{k},n}^2)| \leq 4({}_sN_n)^{-1}\max\Big\{\Big[\mathrm{E}(\tilde{S}_{\mathbf{k},n}^2)\mathrm{E}(\tilde{S}_{\mathbf{k},n}^{*2})\Big]^{1/2}, \mathrm{E}(\tilde{S}_{\mathbf{k},n}^{*2})\Big\} = o(1)$.

(4) $|\mathrm{Cov}(\tilde{Y}_{\mathbf{0},n}^2,\ \tilde{S}_{\mathbf{k},n}^2)| \leq \mathcal{C}\alpha(\ell_n, {}_sN_n)^{\delta/(2r+\delta)} = o(1)$ by applying Lemmas 8.1-8.2, Assumption A.3, and Condition M_r with $\mathrm{dis}(\tilde{R}_{\mathbf{k},n} \cap \mathbf{Z}^d \setminus T_{\mathbf{k},n},\ {}_sR_n \cap \mathbf{Z}^d) \geq \ell_n$.

Since $\tilde{\sigma}_n(\mathbf{0}) = \mathrm{Var}(Y_{\mathbf{0},n}^2)$, the remaining quantity in (8.18) can be expressed as

$$\frac{({}_sN_n)^2}{|J_{NOL}|}\tilde{\sigma}_n(\mathbf{0}) = \frac{1}{|J_{NOL}|}\mathrm{Var}\Big([\nabla' Z_\infty]^2\Big)(1 + o(1)) = \frac{\det(\Delta_n)}{\det(\Delta_n)|R_0|}[2\tau^4](1 + o(1))$$

by applying the CLT (as in (8.8)) and Lemma 8.3. We have now established (8.17)(a).

We omit the proof of (8.17)(b), which resembles the one establishing (8.12)(b) and incorporates arguments used to bound U_{1n}, U_{2n}; Nordman (2002) provides more details. □

9 Proofs for bias expansions

We will use the following lemma concerning $\tau_n^2 = N_n\mathrm{Var}(\hat{\theta}_n)$ to prove the theorems pertaining to bias expansions of $\hat{\tau}_{n,OL}^2$ and $\hat{\tau}_{n,NOL}^2$.

Lemma 9.1 *Under the Assumptions and Conditions of Theorem* 3.1,

$$\tau_n^2 \;=\; \tau^2 \;+\; O\left([\det(\Delta_n)]^{-1/2}\right).$$

Proof: By a Taylor's expansion around μ: $\hat{\theta}_n = H(\bar{Z}_{N_n}) = H(\mu) + \bar{Y}_{N_n} + \bar{Q}_{N_n}$ (replacing $\bar{Z}_{N_n}$ for $Z_{\mathbf{i},n}$ in (8.11)) and so $N_n \mathrm{Var}(\hat{\theta}_n) = N_n \mathrm{Var}(\bar{Y}_{N_n} + \bar{Q}_{N_n})$. For $\mathbf{k} \in \mathbb{Z}^d$, let $N_n(\mathbf{k}) = |\{\mathbf{i} \in R_n \cap \mathbf{Z}^d : \mathbf{i} + \mathbf{k} \in R_n\}|$. It holds that $N_n(\mathbf{k}) \leq N_n$ and

$$\begin{aligned} N_n \;&\leq\; N_n(\mathbf{k}) + |\{\mathbf{i} \in \mathbb{Z}^d : T^{\mathbf{i}} \cap \overline{R_n} \neq \varnothing, T^{\mathbf{i}} \cap \overline{R_n^c} \neq \varnothing;\; T^{\mathbf{i}} = \mathbf{i} + \|\mathbf{k}\|_\infty[-1,1]^d\}| \\ &\leq\; N_n(\mathbf{k}) + \mathcal{C}\|\mathbf{k}\|_\infty^d (\lambda_n^{\max})^{d-1}, \end{aligned} \tag{9.1}$$

by the boundary condition on R_0. Also, by Lemma 8.1 and stationarity, for each $\mathbf{k} \neq \mathbf{0} \in \mathbb{Z}^d$:

$$|\sigma(\mathbf{k})| \;\leq\; \mathcal{C}\alpha_1(\|\mathbf{k}\|_\infty)^{\delta/(2r+\delta)}, \quad \mathbf{k} \in \mathbb{Z}^d. \tag{9.2}$$

Using $|\{\mathbf{k} \in \mathbb{Z}^d : \|\mathbf{k}\|_\infty = x\}| \leq \mathcal{C}x^{d-1}$, $x \geq 1$, the covariances are absolutely summable over $\mathbb{Z}^d$:

$$\sum_{\mathbf{k} \in \mathbb{Z}^d} |\sigma(\mathbf{k})| \;\leq\; |\sigma(\mathbf{0})| + \mathcal{C}\sum_{x=1}^{\infty} x^{d-1}\alpha_1(x)^{\delta/(2r+\delta)} \;<\; \infty. \tag{9.3}$$

From (9.1)-(9.3), we find

$$N_n \mathrm{Var}(\bar{Y}_{N_n}) \;=\; \frac{1}{N_n}\sum_{\mathbf{k} \in \mathbb{Z}^d} N_n(\mathbf{k})\sigma(\mathbf{k}) \;=\; \tau^2 \;+\; I_n; \tag{9.4}$$

$$\begin{aligned} |I_n| \;&\leq\; \frac{1}{N_n}\sum_{\mathbf{k} \in \mathbb{Z}^d} |N_n - N_n(\mathbf{k})| \cdot |\sigma(\mathbf{k})| \\ &\leq\; \mathcal{C} \cdot \frac{(\lambda_n^{\max})^{d-1}}{N_n}\sum_{x=1}^{\infty} x^{2d-1}\alpha_1(x)^{\delta/(2r+\delta)} \;=\; O\left([\det(\Delta_n)]^{-1/d}\right). \end{aligned} \tag{9.5}$$

By Condition D and Lemma 8.2, it follows that $N_n \mathrm{Var}(\bar{Q}_{N_n}) = O([\det(\Delta_n)]^{-1})$.

Finally, with bounds on the variance of $\bar{Y}_{N_n}$ and $\bar{Q}_{N_n}$, we apply the Cauchy-Schwartz inequality to the covariance: $N_n|\mathrm{Cov}(\bar{Y}_{N_n}, \bar{Q}_{N_n})| = O([\det(\Delta_n)]^{-1/2})$, setting the order on the difference $|N_n \mathrm{Var}(\hat{\theta}_n) - \tau^2|$. □

We give a few lemmas which help compute the bias of the estimators $\hat{\tau}^2_{n,OL}$ and $\hat{\tau}^2_{n,NOL}$.

Lemma 9.2 *Let* $\tilde{Y}_{\mathbf{i},n} = ({}_sN_{\mathbf{i},n})^{-1}\sum_{\mathbf{s} \in \mathbf{Z}^d \cap \tilde{R}_{\mathbf{i},n}} \nabla'(Z(\mathbf{s}) - \mu)$, $\mathbf{i} \in \mathbb{Z}^d$. *Suppose Assumptions* A.1 - A.5 *and Conditions* D_2 *and* M_{2+a} *hold with* $d \geq 2$. *Then,*

$$\begin{array}{r} \mathrm{E}(\hat{\tau}^2_{n,OL}) - {}_sN_{\mathbf{0},n}\mathrm{E}(\tilde{Y}^2_{\mathbf{0},n}) \\ \mathrm{E}(\hat{\tau}^2_{n,NOL}) - |J_{NOL}|^{-1}\sum_{\mathbf{i} \in J_{NOL}} {}_sN_{\mathbf{i},n}\mathrm{E}(\tilde{Y}^2_{\mathbf{i},n}) \end{array} =: O\left([\det({}_s\Delta_n)]^{-1/2}\right) \;+\; o\left([\det({}_s\Delta_n)]^{-1/d}\right).$$

Proof: We consider here only $\mathrm{E}(\hat{\tau}^2_{n,OL})$. For integer ${}_s\Delta_n$, the arguments for $\mathrm{E}(\hat{\tau}^2_{n,NOL})$ are essentially the same; more details are provided in Nordman (2002).

By stationarity and an algebraic expansion:

$$\mathrm{E}(\hat{\tau}^2_{n,OL}) = {}_sN_n\Big[\mathrm{E}(Y^2_{\mathbf{0},n}) + \mathrm{E}(Q^2_{\mathbf{0},n}) + 2\mathrm{E}(Y_{\mathbf{0},n}Q_{\mathbf{0},n}) - \mathrm{E}(\bar{Y}^2_n) - \mathrm{E}(\bar{Q}^2_n) - 2\mathrm{E}(\bar{Y}_n\bar{Q}_n)\Big].$$

With the moment arguments based on Lemma 8.2 and Condition D_r, we have

$${}_sN_n\mathrm{E}(Y^2_{\mathbf{0},n}) \leq \mathcal{C}, \quad {}_sN_n\mathrm{E}(\bar{Y}^2_n) \leq \mathcal{C}\,{}_sN_n(N_n)^{-1}, \quad {}_sN_n\mathrm{E}(Q^2_{\mathbf{0},n}),\, {}_sN_n\mathrm{E}(\bar{Q}^2_n) \leq \mathcal{C}({}_sN_n)^{-1}, \tag{9.6}$$

where bound on ${}_sN_n\mathrm{E}(\bar{Y}_n^2)$ follows from (8.16). By Holder's inequality and Assumption A.2:

$$\mathrm{E}(\hat{\tau}^2_{n,OL}) = {}_sN_n\mathrm{E}(Y^2_{\mathbf{0},n}) + O\Big(({}_sN_n)^{-1/2}\Big) + O\Big({}_sN_n(N_n)^{-1}\Big).$$

Note that $\mathrm{E}(Y^2_{\mathbf{0},n}) = \mathrm{E}(\tilde{Y}^2_{\mathbf{0},n})$, ${}_sN_n = {}_sN_{\mathbf{0},n}$. Hence, applying Lemma 8.3 and Assumption A.2, we establish Lemma 9.2 for $\hat{\tau}^2_{n,OL}$. □

The next lemma provides a small refinement to Lemma 9.2 made possible when the function $H(\cdot)$ is smoother. We shall make use of this lemma in bias expansions of $\hat{\tau}^2_{n,OL}$ and $\hat{\tau}^2_{n,NOL}$ in lower sampling dimensions, namely $d=1$ or 2.

Lemma 9.3 *Assume $d=1$ or 2. In addition to Assumptions* A.1 - A.5, *suppose that Conditions D_3 and M_{3+a} hold. Then,*

$$\begin{array}{r} \mathrm{E}(\hat{\tau}^2_{n,OL}) - {}_sN_{\mathbf{0},n}\mathrm{E}(\tilde{Y}^2_{\mathbf{0},n}) \\ \mathrm{E}(\hat{\tau}^2_{n,NOL}) - |J_{NOL}|^{-1}\sum_{\mathbf{i}\in J_{NOL}} {}_sN_{\mathbf{i},n}\mathrm{E}(\tilde{Y}^2_{\mathbf{i},n}) \end{array} =: \begin{cases} O\Big([\det({}_s\Delta_n)]^{-1}\Big) & \text{if } d=1, \\ o\Big([\det({}_s\Delta_n)]^{-1/2}\Big) & \text{if } d=2. \end{cases}$$

Proof: We again consider only $\hat{\tau}^2_{n,OL}$. For $\mathbf{i}\in J_{OL}$, we use a third-order Taylor's expansion of each subsample statistic around μ: $\hat{\theta}_{\mathbf{i},n} = H(\mu) + Y_{\mathbf{i},n} + Q_{\mathbf{i},n} + C_{\mathbf{i},n}$, where $Y_{\mathbf{i},n} = \nabla'(Z_{\mathbf{i},n}-\mu)$;

$$Q_{\mathbf{i},n} = \sum_{\|\alpha\|_1=2}\frac{c_\alpha}{\alpha!}(Z_{\mathbf{i},n}-\mu)^\alpha;\ C_{\mathbf{i},n} = 3\sum_{\|\alpha\|_1=3}\frac{c_\alpha}{\alpha!}(Z_{\mathbf{i},n}-\mu)^\alpha\int_0^1(1-\omega)^2D^\alpha H(\mu+\omega(Z_{\mathbf{i},n}-\mu))d\omega.$$

Here $C_{\mathbf{i},n}$ denotes the remainder term in the Taylor's expansion and $Q_{\mathbf{i},n}$ is defined a little differently. Write the sample means for the Taylor terms: $\bar{Y}_n$, $\bar{Q}_n$ as before, $\bar{C}_n = |J_{OL}|^{-1}\sum_{\mathbf{i}\in J_{OL}}C_{\mathbf{i},n}$. The moment inequalities in (9.6) are still valid and, by Lemma 8.2 and Condition D, we can produce bounds: ${}_sN_n\mathrm{E}(C^2_{\mathbf{0},n}), {}_sN_n\mathrm{E}(\bar{C}_n^2) \le \mathcal{C}({}_sN_n)^{-2}$. By Holder's inequality and the scaling conditions from Assumptions A.1-A.2, we then have

$$\mathrm{E}(\hat{\tau}^2_{n,OL}) \ = \ {}_sN_n\Big[\mathrm{E}(Y^2_{\mathbf{0},n}) \ + \ 2\mathrm{E}(Y_{\mathbf{0},n}Q_{\mathbf{0},n})\Big] \ + \ \begin{cases} O\left([\det({}_s\Delta_n)]^{-1}\right) & \text{if } d=1, \\ o\left([\det({}_s\Delta_n)]^{-1/2}\right) & \text{if } d=2. \end{cases}$$

Since ${}_sN_n\mathrm{E}(Y^2_{\mathbf{0},n}) = {}_sN_{\mathbf{0},n}\mathrm{E}(\tilde{Y}^2_{\mathbf{0},n})$, Lemma 9.3 for $\hat{\tau}^2_{n,OL}$ will follow by showing

$${}_sN_n\mathrm{E}(Y_{\mathbf{0},n}Q_{\mathbf{0},n}) = {}_sN_n\sum_{i,j,k=1}^{p}c_ia_{j,k}\mathrm{E}\Big[(Z_{i,\mathbf{0},n}-\mu)(Z_{j,\mathbf{0},n}-\mu)(Z_{k,\mathbf{0},n}-\mu)\Big] = O\left([\det({}_s\Delta_n)]^{-1}\right), \tag{9.7}$$

where $Z_{\mathbf{0},n} = (Z_{1,\mathbf{0},n},\ldots,Z_{p,\mathbf{0},n})' \in \mathbb{R}^p$ is a vector of coordinate sample means; $c_i = \partial H(\mu)/\partial x_i$; $a_{j,k} = 1/2\cdot\partial^2H(\mu)/\partial x_j\partial x_k$.

Denote the observation $Z(\mathbf{s}) = (Z_1(\mathbf{s}),\ldots,Z_p(\mathbf{s}))' \in \mathbb{R}^d$, $\mathbf{s}\in\mathbf{Z}^d$. Fix $i,j,k\in\{1,\ldots,p\}$ and WLOG assume $\mu=0$. Then, ${}_sN_n\big|\mathrm{E}(Z_{i,\mathbf{0},n}Z_{j,\mathbf{0},n}Z_{k,\mathbf{0},n})\big| = \big|({}_sN_n)^{-1}\mathrm{E}(Z_i(\mathbf{t})Z_j(\mathbf{t})Z_k(\mathbf{t})) + L^{ijk}_{1n} + L^{ijk}_{2n}\big|$ where

$$\begin{aligned} L^{ijk}_{1n} &= ({}_sN_n)^{-2}\sum_{\substack{\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbf{Z}^d\cap{}_sR_n,\\ \mathbf{u}\neq\mathbf{v}\neq\mathbf{w}}}\mathrm{E}\Big[Z_i(\mathbf{u})Z_j(\mathbf{v})Z_k(\mathbf{w})\Big], \\ L^{ijk}_{2n} &= ({}_sN_n)^{-2}\sum_{\substack{\mathbf{u},\mathbf{v}\in\mathbf{Z}^d\cap{}_sR_n,\\ \mathbf{u}\neq\mathbf{v}}}\mathrm{E}\Big[Z_i(\mathbf{u})Z_j(\mathbf{u})Z_k(\mathbf{v}) + Z_i(\mathbf{u})Z_j(\mathbf{v})Z_k(\mathbf{u}) + Z_i(\mathbf{v})Z_j(\mathbf{u})Z_k(\mathbf{u})\Big]. \end{aligned}$$

By Lemma 8.1, Assumption A.3, and Condition M_r,

$$|L^{ijk}_{2n}| \ \le \ \frac{\mathcal{C}}{{}_sN_n}\sum_{x=1}^{\infty}x^{d-1}\alpha(x,1)^{\delta/(2r+\delta)} \ = \ O\left([\det({}_s\Delta_n)]^{-1}\right),$$

similar to (9.3). For $\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3 \in \mathbb{R}^d$, define $\mathrm{dis}_3(\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}) = \max_{1 \le i \le 3} \mathrm{dis}(\{\mathbf{y}_i\}, \{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\} \setminus \{\mathbf{y}_i\})$. If $x \geq 1 \in \mathbb{Z}_+$, then $|\{(\mathbf{y}_1, \mathbf{y}_2) \in (\mathbb{Z}^d)^2 : \mathrm{dis}_3(\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{0}\}) = x\}| \leq \mathcal{C} x^{2d-1}$ from Theorem 4.1, Lahiri (1999a). Thus,

$$|L_{1n}^{ijk}| \;\leq\; \frac{\mathcal{C}}{{}_sN_n} \sum_{x=1}^{\infty} x^{2d-1} \alpha(x,2)^{\delta/(2r+\delta)} \;=\; O\left([\det({}_s\Delta_n)]^{-1}\right).$$

This establishes (9.7), completing the proof of Lemma 9.3 for $\hat{\tau}^2_{n,OL}$. □

We use the next lemma in the proof of Theorem 4.2. It allows us to approximate lattice point counts with Lebesgue volumes, in $\mathbb{R}^2$ or $\mathbb{R}^3$, to a sufficient degree of accuracy.

Lemma 9.4 *Let $d = 2, 3$ and $R_0 \subset (-1/2, 1/2]^d$ such that $B^\circ \subset R_0 \subset \overline{B}$ for a convex set B. Let $\{b_n\}_{n=1}^{\infty}$ be a sequence of positive real numbers such that $b_n \to \infty$. If $\mathbf{k} \in \mathbb{Z}^d$, then there exists $\mathrm{N}_{\mathbf{k}} \in \mathbb{Z}_+$ and $\mathcal{C}_d > 0$ such that for $n \geq \mathrm{N}_{\mathbf{k}}$, $\mathbf{i} \in \mathbb{Z}^d$,*

$$\begin{aligned} &\Big| \Big(|b_n R_0| - |\mathbf{Z}^d \cap b_n(\mathbf{i} + R_0)| \Big) - \Big(|b_n R_0 \cap \mathbf{k} + b_n R_0| - |\mathbf{Z}^d \cap b_n(\mathbf{i} + R_0) \cap \mathbf{k} + b_n(\mathbf{i} + R_0)| \Big) \Big| \\ &\leq \begin{cases} \mathcal{C}_2 \|\mathbf{k}\|_\infty^2 & \text{if } d = 2 \\ \mathcal{C}_3 \|\mathbf{k}\|_\infty^4 \left(b_n^{5/3} + \xi_{\mathbf{k},n} b_n^2 \right) & \text{if } d = 3, \end{cases} \end{aligned}$$

where $\{\xi_{\mathbf{k},n}\}_{n=1}^{\infty} \subset \mathbb{R}$ is a nonnegative sequence (possibly dependent on $\mathbf{k}$) such that $\xi_{\mathbf{k},n} \to 0$.

Proof: Provided in Nordman and Lahiri (2002). □

To establish Lemma 4.1, we require some additional notation. For $\mathbf{i}, \mathbf{k} \in \mathbb{Z}^d$, let ${}_sN_{\mathbf{i},n}(\mathbf{k}) = |\mathbf{Z}^d \cap \tilde{R}_{\mathbf{i},n} \cap \mathbf{k} + \tilde{R}_{\mathbf{i},n}|$ denote the number of sampling sites (lattice points) in the intersection of a NOL subregion with its $\mathbf{k}$-translate. Note ${}_sN_{\mathbf{i},n}(\mathbf{k})$ is a subsample version of $N_n(\mathbf{k})$ from (9.1).

Proof of Lemma 4.1: We start with bounds:

$$\begin{aligned} \sup_{\mathbf{i} \in \mathbb{Z}^d} |{}_sN_n - {}_sN_{\mathbf{i},n}| &\leq \mathcal{C}({}_s\lambda_n^{\max})^{d-1}, && (9.8) \\ |{}_sN_{\mathbf{i},n} - {}_sN_{\mathbf{i},n}(\mathbf{k})| &\leq |\{\mathbf{j} \in \mathbb{Z}^d : T^{\mathbf{j}} \cap \overline{{}_sR_n} \neq \varnothing, T^{\mathbf{j}} \cap \overline{{}_sR_n^c} \neq \varnothing;\ T^{\mathbf{j}} = \mathbf{j} + \|\mathbf{k}\|_\infty[-2,2]^d\}| \\ &\leq \mathcal{C}\|\mathbf{k}\|_\infty^d ({}_s\lambda_n^{\max})^{d-1}, && (9.9) \end{aligned}$$

by the boundary condition on R_0 (cf. Lemma 8.3) and $\inf_{\mathbf{j} \in \mathbb{Z}^d} \|{}_s\Delta_n \mathbf{i} - \mathbf{j}\|_\infty \leq 1/2$.

Modify (9.4) by replacing $N_n, N_n(\mathbf{k}), \bar{Y}_{N_n}$ with ${}_sN_{\mathbf{i},n}, {}_sN_{\mathbf{i},n}(\mathbf{k}), \tilde{Y}_{\mathbf{i},n} = \nabla'(\tilde{Z}_{\mathbf{i},n} - \mu)$ (i.e., use a NOL subsample version in place of the sample one); and replace $N_n, \Delta_n, \lambda_n^{\max}$ with the subsample analogs ${}_sN_{\mathbf{i},n}, {}_s\Delta_n, {}_s\lambda_n^{\max}$ in (9.5). We then find using (9.3): for each $\mathbf{i} \in \mathbb{Z}^d$,

$$\begin{aligned} {}_sN_{\mathbf{i},n} \mathrm{E}(\tilde{Y}_{\mathbf{i},n}^2) - \tau^2 \;&=\; \frac{1}{{}_sN_{\mathbf{i},n}} \sum_{\mathbf{k} \in \mathbb{Z}^d} ({}_sN_{\mathbf{i},n}(\mathbf{k}) - {}_sN_{\mathbf{i},n}) \sigma(\mathbf{k}) \;\equiv\; {}_sI_{\mathbf{i},n}, && (9.10) \\ \sup_{\mathbf{i} \in \mathbb{Z}^d} |{}_sI_{\mathbf{i},n}| \;&\leq\; \sup_{\mathbf{i} \in \mathbb{Z}^d} \left\{ \frac{1}{{}_sN_{\mathbf{i},n}} \sum_{\mathbf{k} \in \mathbb{Z}^d} |{}_sN_{\mathbf{i},n}(\mathbf{k}) - {}_sN_{\mathbf{i},n}| \cdot |\sigma(\mathbf{k})| \right\} \\ &\leq\; \mathcal{C} \cdot \frac{({}_s\lambda_n^{\max})^{d-1}}{{}_sN_n - \mathcal{C}({}_s\lambda_n^{\max})^{d-1}} \sum_{x=1}^{\infty} x^{2d-1} \alpha_1(x)^{\delta/(2r+\delta)} \;=\; O\left([\det({}_s\Delta_n)]^{-1/d} \right), && (9.11) \end{aligned}$$

from (9.8)-(9.9) and Assumption A.1. Now applying Lemma 9.1 and Assumption A.2 with Lemma 9.2 for $d \geq 2$ or Lemma 9.3 for $d = 1$, Lemma 4.1 follows. □

Proof of Theorem 4.1: Here ${}_sN_{\mathbf{i},n} = {}_sN_n$, ${}_sN_{\mathbf{i},n}(\mathbf{k}) = C_n(\mathbf{k})$, $\mathrm{E}(\tilde{Y}_{\mathbf{i},n}) = \mathrm{E}(\tilde{Y}_{\mathbf{0},n})$ for each $\mathbf{i},\mathbf{k} \in \mathbb{Z}^d$ (since ${}_s\lambda_n \in \mathbb{Z}_+$) and $\det({}_s\Delta_n) = {}_s\lambda_n^d$. Applying Lemma 9.2 for $d \geq 3$ and Lemma 9.3 for $d = 2$, Lemma 9.1, Assumption A.2, and (9.11):

$$\mathrm{E}(\hat{\tau}_n^2) - \tau_n^2 \;=\; \frac{-1}{{}_s\lambda_n|R_0|}\sum_{\mathbf{k}\in\mathbb{Z}^d} g_n(\mathbf{k}) + o({}_s\lambda_n^{-1}), \quad g_n(k) \equiv \frac{{}_sN_n - C_n(\mathbf{k})}{{}_s\lambda_n^{d-1}} \cdot \frac{{}_s\lambda_n^d|R_0|}{{}_sN_n}\cdot\sigma(\mathbf{k}).$$

From (9.11) and Lemma 8.3, it follows that $\sum_{\mathbf{k}\in\mathbb{Z}^d}|g_n(\mathbf{k})| \leq \mathcal{C}$, $n \in \mathbb{Z}_+$ and that $g_n(\mathbf{k}) \to C(\mathbf{k})\sigma(\mathbf{k})$ for $\mathbf{k} \in \mathbb{Z}^d$. By the LDCT, the proof of Theorem 4.1 is complete. □

To establish Theorem 4.2, we require some additional notation. For $\mathbf{i},\mathbf{k} \in \mathbb{Z}^d$, denote the difference between two Lebesgue volume-for-lattice-point-count approximations as:

$$D_{\mathbf{i},n}(\mathbf{k}) = \Big(|\tilde{R}_{\mathbf{i},n}| - {}_sN_{\mathbf{i},n}\Big) - \Big(|\tilde{R}_{\mathbf{i},n}\cap\mathbf{k}+\tilde{R}_{\mathbf{i},n}| - {}_sN_{\mathbf{i},n}(\mathbf{k})|\Big) = \Big(|{}_sR_n| - {}_sN_{\mathbf{i},n}\Big) - \Big(|{}_sR_n\cap\mathbf{k}+{}_sR_n| - {}_sN_{\mathbf{i},n}(\mathbf{k})|\Big).$$

Proof of Theorem 4.2: We handle here the cases $d = 2$ or 3. Details on the proof for $d = 1$ are given in Nordman (2002). We note first that if $V(\mathbf{k})$ exists for each $\mathbf{k} \in \mathbb{Z}^d$, then Lemma 9.4 implies $C(\mathbf{k}) = V(\mathbf{k})$.

Consider $\hat{\tau}^2_{n,NOL}$. Applying Lemma 9.2 for $d = 3$, and Lemma 9.3 for $d = 2$, with (9.8), (9.10), and (9.11) gives

$$\mathrm{E}(\hat{\tau}^2_{n,NOL}) - \tau_n^2 \;=\; |J_{NOL}|^{-1}\sum_{\mathbf{i}\in J_{NOL}} \frac{{}_sN_{\mathbf{i},n}}{|{}_sR_n|}{}_sI_{\mathbf{i},n} \;+\; o({}_s\lambda_n^{-1}).$$

Then, using (9.3), we can arrange terms to write

$$|J_{NOL}|^{-1}\sum_{\mathbf{i}\in J_{NOL}} \frac{{}_sN_{\mathbf{i},n}}{|{}_sR_n|}{}_sI_{\mathbf{i},n} \;=\; \Psi_n \;+\; \sum_{\mathbf{k}\in\mathbb{Z}^d}\frac{G_n(\mathbf{k})}{{}_s\lambda_n|R_0|}; \qquad G_n(\mathbf{k}) = \sum_{\mathbf{i}\in J_{NOL}}\frac{D_{\mathbf{i},n}(\mathbf{k})\cdot\sigma(\mathbf{k})}{{}_s\lambda_n^{d-1}|J_{NOL}|},$$

for $\Psi_n = -\sum_{\mathbf{k}\in\mathbb{Z}^d}|{}_sR_n|^{-1}(|{}_sR_n| - |{}_sR_n\cap\mathbf{k}+{}_sR_n|)\sigma(\mathbf{k})$. Since R_0 is convex, the boundary condition is valid and it holds that for all $\mathbf{i},\mathbf{k} \in \mathbb{Z}^d$:

$$\Big|{}_sN_{\mathbf{i},n}(\mathbf{k}) - |{}_sR_n\cap\mathbf{k}\cap{}_sR_n|\Big| \;\leq\; \mathcal{C}({}_s\lambda_n^{\max})^{d-1}, \qquad \Big||{}_sR_n| - |{}_sR_n\cap\mathbf{k}+{}_sR_n|\Big| \;\leq\; \mathcal{C}\|\mathbf{k}\|_\infty^d({}_s\lambda_n^{\max})^{d-1} \tag{9.12}$$

from Lemma 8.3 and (9.9). Then (9.3), Lemma 9.4, and (9.12) give: $\sum_{\mathbf{k}\in\mathbb{Z}^d}|G_n(\mathbf{k})| \leq \mathcal{C}$, $n \in \mathbb{Z}_+$; $G_n(\mathbf{k}) \to 0$ for $\mathbf{k} \in \mathbb{Z}^d$; and ${}_s\lambda_n\Psi_n = O(1)$. By the LDCT, we establish

$$\sum_{\mathbf{k}\in\mathbb{Z}^d}\frac{G_n(\mathbf{k})}{{}_s\lambda_n|R_0|} \;=\; o\left({}_s\lambda_n^{-1}\right), \qquad \mathrm{E}(\hat{\tau}^2_{n,NOL}) - \tau_n^2 \;=\; \Psi_n(1+o(1)),$$

representing the formulation of Theorem 4.2 in terms of Ψ_n. If $V(\mathbf{k})$ exists for each $\mathbf{k} \in \mathbb{Z}^d$, then (9.3) and (9.12) imply that we can use the LDCT again to produce

$$\Psi_n \;=\; \frac{-1}{{}_s\lambda_n|R_0|}\bigg(\sum_{\mathbf{k}\in\mathbb{Z}^d} V(\mathbf{k})\sigma(\mathbf{k})\bigg)(1+o(1)). \tag{9.13}$$

The proof of Theorem 4.2 for $\hat{\tau}^2_{n,NOL}$ is now finished.

Consider $\hat{\tau}^2_{n,OL}$. We can repeat the same steps as above to find

$$\mathrm{E}(\hat{\tau}^2_{n,OL}) - \tau_n^2 \;=\; \Psi_n + \sum_{\mathbf{k}\in\mathbb{Z}^d}\frac{G_n^*(\mathbf{k})}{{}_s\lambda_n|R_0|} \;+\; o\left({}_s\lambda_n^{-1}\right), \qquad G_n^*(\mathbf{k}) = \frac{D_{\mathbf{0},n}(\mathbf{k})\cdot\sigma(\mathbf{k})}{{}_s\lambda_n^{d-1}}.$$

The same arguments for G_n apply to G_n^* and (9.13) remains valid when each $V(\mathbf{k})$ exists, $\mathbf{k} \in \mathbb{Z}^d$, establishing Theorem 4.2 for $\hat{\tau}^2_{n,OL}$. Note as well that if $V(\mathbf{k})$ exists for each $\mathbf{k} \in \mathbb{Z}^d$, then Lemma 9.4 and

Lemma 4.1 also imply the second formulation of the bias in Theorem 4.2. □

Proof of Theorem 5.1: Follows from Theorems 3.1 and 4.1 and simple arguments from Calculus involving minimization of a smooth function of a real variable. □.

Proof of Theorem 6.1: We develop some tools to facilitate the counting arguments required for expanding the biases of $\hat{\tau}^2_{n,OL}$ and $\hat{\tau}^2_{n,NOL}$. We first define a count for the number of $\mathbf{Z}^d$ lattice points lying in a "face" of a d-dimensional rectangle. For a rectangle T, where $\prod_{j=1}^d (c_j, \tilde{c}_j) \subset T \subset \prod_{j=1}^d [c_j, \tilde{c}_j]$, $c_j, \tilde{c}_j \in \mathbb{R}$, define the "border" point set: $\mathcal{B}\{T\} = \bigcup_{j=1}^d \left\{\mathbf{s} = (s_1, \ldots, s_d)' \in \mathbf{Z}^d \cap T : \; s_j \in \{c_j, \tilde{c}_j\}\right\}$. With $\mathbf{k} \in \mathbb{Z}^d$, it can be shown that there exists $\mathrm{N}_{\mathbf{k}} \in \mathbb{Z}_+$ such that: for $n \geq \mathrm{N}_{\mathbf{k}}$,

$$\left|\mathcal{B}\{{}_s\lambda_n(\mathbf{i} + \overline{R_0})\}\right| - \left|\mathcal{B}\{{}_s\lambda_n(\mathbf{i} + \overline{R_0}) \cap \mathbf{k} + {}_s\lambda_n(\mathbf{i} + \overline{R_0})\}\right| \leq 3^{2d(d+1)} \|\mathbf{k}\|_\infty {}_s\lambda_n^{d-2}, \quad \mathbf{i} \in \mathbb{Z}^d. \tag{9.14}$$

See Nordman (2002) for more details.

Fix $\mathbf{k} \neq \mathbf{0} \in \mathbb{Z}^d$ and assume $\min_{1 \leq j \leq d} {}_s\lambda_n \ell_j > 3 + \|\mathbf{k}\|_\infty$. We then write, for $\mathbf{i} \in \mathbb{Z}^d$: ${}_sN_{\mathbf{i},n} = P_{5,\mathbf{i},n} - P_{4,\mathbf{i},n} + P_{3,\mathbf{i},n}$, ${}_sN_{\mathbf{i},n}(\mathbf{k}) = P_{6,\mathbf{i},n} - P_{7,\mathbf{i},n} + P_{8,\mathbf{i},n}$, and

$$|D_{\mathbf{i},n}(\mathbf{k})| \;=\; \left|\sum_{j=1}^4 (P_{2j-1,\mathbf{i},n} \;-\; P_{2j,\mathbf{i},n})\right|;$$

$P_{1,\mathbf{i},n} = \prod_{j=1}^d {}_s\lambda_n \ell_j$, $P_{2,\mathbf{i},n} = \prod_{j=1}^d ({}_s\lambda_n \ell_j - |k_j|)$, $P_{3,\mathbf{i},n} = |\mathcal{B}\{{}_s\lambda_n(\mathbf{i} + R_0)\}|$, $P_{4,\mathbf{i},n} = |\mathcal{B}\{{}_s\lambda_n(\mathbf{i} + \overline{R_0})\}|$, $P_{5,\mathbf{i},n} = \prod_{j=1}^d |{}_s\lambda_n(i_j + \ell_j[-1/2, 1/2]) - t_j \cap \mathbb{Z}|$, $P_{6,\mathbf{i},n} = \prod_{j=1}^d (|{}_s\lambda_n(i_j + \ell_j[-1/2, 1/2]) - t_j \cap \mathbb{Z}| - |k_j|)$, $P_{7,\mathbf{i},n} = |\mathcal{B}\{{}_s\lambda_n(\mathbf{i} + \overline{R_0}) \cap \mathbf{k} + {}_s\lambda_n(\mathbf{i} + \overline{R_0})\}|$, $P_{8,\mathbf{i},n} = |\mathcal{B}\{{}_s\lambda_n(\mathbf{i} + R_0) \cap \mathbf{k} + {}_s\lambda_n(\mathbf{i} + R_0)\}|$.

We show now that, for $\mathbf{k} \neq \mathbf{0}$, there exists $\mathcal{C} > 0$, $\mathrm{N}_{\mathbf{k}} \in \mathbb{Z}_+$, such that: $n \geq \mathrm{N}_{\mathbf{k}}$,

$$|D_{\mathbf{i},n}(\mathbf{k})| \;=\; \left|\sum_{j=1}^4 (P_{2j-1,\mathbf{i},n} \;-\; P_{2j,\mathbf{i},n})\right| \;\leq\; \mathcal{C}\|\mathbf{k}\|_\infty^{d-1} {}_s\lambda_n^{d-2}, \qquad \mathbf{i} \in \mathbb{Z}^d. \tag{9.15}$$

By (9.14), for large $n \geq \mathrm{N}_{\mathbf{k}}$, we have that $0 \leq (P_{3,\mathbf{i},n} - P_{8,\mathbf{i},n}) \leq (P_{4,\mathbf{i},n} - P_{7,\mathbf{i},n}) \leq \mathcal{C}\|\mathbf{k}\|_\infty {}_s\lambda_n^{d-2}$. Note that: $\big||{}_s\lambda_n(i + \ell_j[-1/2,\, 1/2]) - t_j \cap \mathbb{Z}| - {}_s\lambda_n \ell_j\big| \leq 3$ for $i \in \mathbb{Z}$, $j \in \{1, \ldots, d\}$. Hence,

$$\begin{aligned} &|(P_{1,\mathbf{i},n} - P_{5,\mathbf{i},n}) - (P_{2,\mathbf{i},n} - P_{6,\mathbf{i},n})| \\ \leq\; &\sum_{\substack{A \subset \{1,\ldots,d\}, \\ 1 \leq |A| < d}} \|\mathbf{k}\|_\infty^{|A|} \left|\prod_{j \in A} {}_s\lambda_n \ell_j - \prod_{j \in A} |{}_s\lambda_n(i_j + \ell_j[-1/2, 1/2]) - t_j \cap \mathbb{Z}|\right| \;\leq\; \mathcal{C}\|\mathbf{k}\|_\infty^{d-1} {}_s\lambda_n^{d-2}, \end{aligned}$$

which holds uniformly in $\mathbf{i} \in \mathbb{Z}^d$. With these bounds, we have (9.15).

Applying (9.15) in place of Lemma 9.4, the same proof for Theorem 4.2 establishes Theorem 6.1. □

Acknowledgments

Financial support from the Deutsche Forschungsgemeinshaft (SFB 475) and the National Science Foundation (NSF grants DMS 98-16630 and DMS 0072571) is gratefully acknowledged.

References

Billingsley, P. (1986). *Probability and Measure*, 2nd Edition. John Wiley & Sons, New York.

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

Bradley, R. C. (1989). A caution on mixing conditions for random fields. *Statist. Probab. Letters* **8**, 489-491.

BÜHLMAN, P. and KÜNSCH, H. R. (1999). Block length selection in the bootstrap for time series. *CmpStDA* **31**, 295-310.

CARLSTEIN, E. (1986). The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *Ann. Statist.* **14**, 1171-1179.

CORPUT, J. G. van der (1920). Über Gitterpunkte in der Ebene. *Math. Annalen* **81**, 1-20.

CRESSIE, N. (1991). *Statistics for Spatial Data*, 1st Edition. John Wiley & Sons, New York.

DICICCIO, T., HALL, P., and ROMANO, J. P. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.* **19**, 1053-1061.

DOUKHAN, P. (1994). Mixing: properties and examples. *Lecture Notes in Statistics* **85**. Springer-Verlag, New York.

FUKUCHI, J.-I. (1999). Subsampling and model selection in time series analysis. *Biometrika* **86**, 591-604.

GARCIA-SOIDAN, P. H. and HALL, P. (1997). In sample reuse methods for spatial data. *Biometrics* **53**, 273-281.

GUYON, X. (1995). *Random Fields on a Network*. Springer-Verlag, New York.

HALL, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

HALL, P., HOROWITZ, J. L., and JING, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561-574.

HALL, P. and JING, B.-Y. (1996). On sample reuse methods for dependent data. *J. Roy. Statist. Soc. Ser. B* **58**, 727-737.

HUXLEY, M. N. (1993). Exponential sums and lattice points II. *Proc. London Math. Soc.* **66**, 279-301.

HUXLEY, M. N. (1996). *Area, Lattice Points, and Exponential Sums*. Oxford University Press, New York.

KRÄTZEL, E. (1988). *Lattice Points*. Deutscher Verlag Wiss., Berlin.

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**, 1217-1241.

LAHIRI, S. N. (1996). On empirical choice of the optimal block size for block bootstrap methods. *Preprint.* Department of Statistics, Iowa State University, Ames, IA.

LAHIRI, S. N. (1999a). Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method. *Probab. Theory Related Fields* **114**, 55-84.

LAHIRI, S. N. (1999b). Central limit theorems for weighted sums under some spatial sampling schemes. *Preprint.* Department of Statistics, Iowa State University, Ames, IA.

LAHIRI, S. N. (1999c). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27**, 386-404.

MARTIN, R. J. (1990). The use of time-series models and methods in the analysis of agricultural field trials. *Commun. Statist.-Theory Meth.* **19**, 55-81.

MEKETON, M. S. and SCHMEISER, B. (1984). Overlapping batch means: something for nothing? *Proceedings Winter Simulation Conference*, 227-230, S. Sheppard, U. Pooch, and D. Pegden (editors). IEEE, Piscataway, New Jersey.

NORDMAN, D. J. (2002). On optimal spatial subsample size for variance estimation. Dissertation paper. Department of Statistics, Iowa State University, Ames, IA.

NORDMAN, D. J. and LAHIRI, S. N. (2002) On the approximation of differenced lattice point counts with application to statistical bias expansions. Dissertation/working paper. Department of Statistics, Iowa State University, Ames, IA.

PERERA, G. (1997). Geometry of $\mathbb{Z}^d$ and the central limit theorem for weakly dependent random fields. *J. Theor. Probab.* **10**, 581-603.

POLITIS, D. N. and ROMANO, J. P. (1993a). Nonparametric resampling for homogeneous strong mixing random fields. *J. Multivariate Anal.* **47**, 301-328.

POLITIS, D. N. and ROMANO, J. P. (1993b). On the sample variance of linear statistics derived from mixing sequences. *Stochst. Proc. Appl.* **45**, 155-167.

POLITIS, D. N. and ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Ann. Statist.* **22**, 2031-2050.

POLITIS, D. N. and SHERMAN, M. (2001). Moment estimation for statistics from marked point processes. *J. Roy. Statist. Soc. Ser. B* **63**, 261-275.

POLITIS, D. N., ROMANO, J. P., and WOLF, M. (1999). *Subsampling*. Springer-Verlag, New York.

POSSOLO, A. (1991). Subsampling a random field. *IMS Lecture Notes* **20**, 286-294.

SHERMAN, M. (1996). Variance estimation for statistics computed from spatial lattice data. *J. Roy. Statist. Soc. Ser. B* **58**, 509-523.

SHERMAN, M. and CARLSTEIN, E. (1994). Nonparametric estimation of the moments of a general statistic computed from spatial data. *J. Amer. Statist. Assoc.* **89**, 496-500.

SONG, W. M. T. and SCHMEISER, B. (1988). Estimating standard errors: empirical behavior of asymptotic MSE-optimal batch sizes. In *Computer Science and Statistics: Proc. 20th Symp. Interface*, 575-580, E. J. Wegman, D. T. Gantz, and J. J. Miller (editors). American Statistical Association.