IMPROVING THE CAVALIERI ESTIMATOR UNDER NON-EQUIDISTANT SAMPLING AND DROPOUTS

MADS STEHR[⊠,1] AND MARKUS KIDERLEN[2]

[1]Center for Statistics, Dpt. of Finance, CBS, Campus Solbjerg Plads 3, 2000 Frederiksberg, Denmark, [2]Centre for Stochastic Geometry and Advanced Bioimaging, Institute for Mathematics, Aarhus University, 8000 Aarhus C, Denmark
e-mail: mast.fi@cbs.dk, kiderlen@math.au.dk

ABSTRACT

Motivated by the stereological problem of volume estimation from parallel section profiles, the so-called Newton-Cotes integral estimators based on random sampling nodes are analyzed. These estimators generalize the classical Cavalieri estimator and its variant for non-equidistant sampling nodes, the generalized Cavalieri estimator, and have typically a substantially smaller variance than the latter. The present paper focuses on the following points in relation to Newton-Cotes estimators: the treatment of dropouts, the construction of variance estimators, and, finally, their application in volume estimation of convex bodies.

Dropouts are eliminated points in the initial stationary point process of sampling nodes, modeled by independent thinning. Among other things, exact representations of the variance are given in terms of the thinning probability and increments of the initial points under two practically relevant sampling models.

The paper presents a general estimation procedure for the variance of Newton-Cotes estimators based on the sampling nodes in a bounded interval. Finally, the findings are illustrated in an application of volume estimation for three-dimensional convex bodies with sufficiently smooth boundaries.

Keywords: Cavalieri estimator, dropouts, Newton-Cotes quadrature, numerical integration with random nodes, stationary point process, variance estimation, weakly (m, p)-piecewise smooth function.

INTRODUCTION

Before proceeding to a more detailed introduction to the paper, we give an informal description of the original motivation and the main findings of this work. Using Cavalieri's principle, it is well known that the volume of a d-dimensional solid can be approximated from $(d - 1)$-dimensional volume measurements on parallel hyperplanes intersecting the object with equidistant spacing. In particular, the volume $\mathrm{Vol}(Y)$ of a compact object $Y \subset \mathbb{R}^3$ can be approximated from sections with equidistant and parallel planes positioned along some fixed direction $v \in S^2$, if the area of each intersection profile is accessible; see (Baddeley and Jensen, 2004, Chapter 7). Formally, if $f(x)$ is the area of the intersection of Y with the plane positioned at a signed distance $x \in \mathbb{R}$ from the origin along v, the integral $\int f \mathrm{d}x = \mathrm{Vol}(Y)$ of $f : \mathbb{R} \to \mathbb{R}$ can be approximated by a Riemann sum

$$\hat{V}(f) = t \sum_{x \in X} f(x), \qquad (1)$$

where $X = t(U + \mathbb{Z})$ is a regular standard grid in $\mathbb{R}$ shifted with $U \in \mathbb{R}$ and scaled by $t > 0$. In design-based sampling the set of nodes is randomized by choosing U uniform in the interval $(0, 1)$, in which case the estimator (1) is commonly known as the

classical Cavalieri estimator. This choice of U turns X into a stationary point process (random locally finite collection of points in $\mathbb{R}$ with a translation invariant distribution) with intensity (expected number of points per unit interval) $1/t$; see (Baddeley and Jensen, 2004). Naturally, the properties of (1) do not rely on the fact that f represents an area function. Thus, from now on we assume that f is a compactly supported and integrable function and that it is known at all points of a stationary point process $X \subseteq \mathbb{R}$, and we refer to f as the *measurement function*. The estimator (1) can also be used if the points of the process X with intensity $1/t$ are not equidistant, in which case it is referred to as the *generalized Cavalieri estimator* as defined initially in (Baddeley *et al.*, 2006). Both estimators are unbiased for the integral $\int_{\mathbb{R}} f(x)\mathrm{d}x$ (Baddeley *et al.*, 2006, Theorem 1), however, as indicated in (Baddeley *et al.*, 2006) and quantified in (Ziegel *et al.*, 2010), the variance of the generalized Cavalieri estimator may be substantially higher than in the equidistant case. Ziegel *et al.* (2010; 2011) also consider the practically relevant situation where the f-values at some of the points of X are unavailable (*dropouts*). Ziegel *et al.* (2011) suggests a better alternative to (1) in this case where discarded function values are approximated by a weighted average of the closest two neighboring known measurements before (1) is applied. In either

case, the variance is higher than in the equidistant case.

As shown in (Stehr and Kiderlen, 2020), the variance inflation imposed by non-equidistant sampling in the generalized Cavalieri estimator can in fact be avoided by using the so-called Newton-Cotes estimators defined in (Kiderlen and Dorph-Petersen, 2017). More precisely, it is shown that these estimators are unbiased for $\int_{\mathbb{R}} f(x)\mathrm{d}x$ and that their variances decrease (with increasing intensity of X) at a rate equivalent to that of the classical Cavalieri estimator as long as the typical increment of X satisfies a mild integrability assumption. In the present paper we continue the study on variance properties and variance approximations of the Newton-Cotes estimators, and we relate these to an application to volume estimation. In particular, we consider sampling processes obtained from an initial process with dropouts, and, under practically relevant sampling models, we give explicit variance representations. Furthermore, we derive estimates for the variance of the Newton-Cotes estimators based on sampling in a bounded interval of $\mathbb{R}$.

Given that the increments of X are available, Kiderlen and Dorph-Petersen (2017) define the Newton-Cotes estimators using quadrature rules as follows. For fixed $n \in \mathbb{N}$, on the interval from $x_0 \in X$ to its nth neighbor $x_n \in X$, the function f is approximated by a piecewise polynomial of degree at most n passing through the points $\{x_j, f(x_j)\}_{j=0}^n$, where $x_1 < \cdots < x_{n-1}$ are the ordered points in $X \cap (x_0, x_n)$. The *nth (order) Newton-Cotes estimator* $\hat{V}_n(f)$ is then obtained as the sum of integrals of such function approximations averaged with respect to the starting point. By this construction, $\hat{V}_n(f)$ is given as the weighted sum

$$\hat{V}_n(f) = \sum_{x \in X} \alpha_n(x)f(x), \qquad (2)$$

where $\alpha_n(x) = \alpha_n(x; X)$ is a rational function of n point increments to the left and right of $x \in X$; see (Stehr and Kiderlen, 2020, Eq.'s (3.3) and (3.4)). In particular, $\alpha_1(x) = (h_1(x) + h_0(x))/2$, where $h_0(x)$ and $h_1(x)$ denotes the increment to the left and right of x, respectively. This gives the *trapezoidal estimator*

$$\hat{V}_1(f) = \sum_{x \in X} \frac{h_1(x) + h_0(x)}{2} f(x),$$

which will be of particular interest throughout this paper. As for the classical and generalized Cavalieri estimators, the nth Newton-Cotes estimator (2) is unbiased for the integral $\int_{\mathbb{R}} f(x)\mathrm{d}x$ as long as the typical increments satisfy certain integrability conditions; see (4) below and (Stehr and Kiderlen,

2020, Theorem 2.1). Interestingly, Newton-Cotes estimators of any order coincide with the classical Cavalieri estimator when the points in X are equidistant, and moreover, the trapezoidal estimator applied to the equidistant process combined with dropouts coincides with the correction method of (Ziegel *et al.*, 2011). This situation is modeled (and will in the remainder be referred to) as independent p-thinning, where independently for each point its corresponding function value is discarded with probability $p \in [0, 1)$. The latter claim will be shown in the section on independent p-thinning. One great advantage of Newton-Cotes estimators is that they can be applied to any stationary point process satisfying (4) with a variance-order independent of the underlying point process. This will be clarified in the next paragraph. In particular, the variances of estimators applied to a stationary point process or a thinning hereof are of the same order; see the section on independent p-thinning.

The order of the variance depends not only on the order n of $\hat{V}_n$ but also on the smoothness of the measurement function f. The smoothness concepts described below are given in terms of jumps of a function, which we define as follows: The function $h : \mathbb{R} \to \mathbb{R}$ jumps at the point $a \in \mathbb{R}$ if the limits and the difference in

$$J_h(a) = h(a^+) - h(a^-) = \lim_{x \downarrow a} h(x) - \lim_{x \uparrow a} h(x)$$

are defined on $\mathbb{R} \cup \{-\infty, \infty\}$ and $J_h(a) \neq 0$. We refer to J_h as the jump-function of h and we let D_h denote the set of jump-points of h. The classical smoothness concept is that of $(m, 1)$-piecewise smoothness, and this is used first and foremost in (Kiêu, 1997) and subsequently in (Stehr and Kiderlen, 2020). For a given $m \in \mathbb{N}_0$, a compactly supported function is said to be $(m, 1)$-piecewise smooth if it is $(m-1)$-times continuously differentiable and if the mth and $(m+1)$st derivatives exist and are continuous except in at most finitely many points, where they may have finite jumps. However, the condition on the $(m+1)$st derivative turns out to be rather restrictive from a practical point of view: In the section on stereological volume estimation below, we give an example of a practically relevant measurement function whose second derivative has infinite jumps at the boundary of its support. It turns out that this less restrictive smoothness property actually suffices for the variance results presented in this paper to hold; in particular, the entire paper (Stehr and Kiderlen, 2020) could have been formulated under this milder smoothness property. We refer to the appendix for a justification. For this reason, from now on we consider what we call *weakly $(m, 1)$-piecewise smooth functions*, which

differ from $(m,1)$-piecewise smooth functions by the less restrictive property that the $(m+1)$st derivative is allowed to have finitely many (possibly infinite) jumps. As the variance of $\hat{V}_n$ is essentially given by the order of the first non-continuous derivative of the measurement function, we assume throughout the following that m is chosen largest possible for a given f, that is, we assume that the mth derivative has at least one jump. It is not difficult to see that the $(m+1)$st derivative of a weakly $(m,1)$-piecewise smooth function is integrable.

It should be mentioned that a third smoothness concept for measurement functions has been introduced in the literature by García-Fiñana and Cruz-Orive (2004). In that paper $(q,1)$-piecewise smoothness has been extended to allow for arbitrary real $q \geq 0$, and functions with this property are called q-smooth. Their theory is based on fractional derivatives. For $m \in \mathbb{N}_0$ an $(m,1)$-piecewise smooth function is also m-smooth, but the converse is false. The function $f(x) = 1_{[0,1]}x\log(x)$ is weakly $(0,1)$-piecewise smooth, but neither $(0,1)$-piecewise smooth nor 0-smooth, showing that our new notion covers previously intractable measurement functions. On the other hand, even when applying the theory to the problem of estimating the volume of a compact convex object, one cannot avoid fractional smoothness; see (García-Fiñana and Cruz-Orive, 2004). Although smoothness assumptions – like those we will impose in the section on volume estimation – can avoid fractional smoothness of the measurement function, a unifying approach is still missing but beyond the scope of the present paper.

The variance of $\hat{V}_n(f)$ is given in terms of the jumps of the $(2m+1)$st derivative of the covariogram $g(z) = \int_{\mathbb{R}} f(x)f(x+z)\mathrm{d}x$, $z \in \mathbb{R}$, associated to f. Here we make use of the fact that g is weakly $(2m+1,1)$-piecewise smooth by Corollary 17 in the appendix (an adaption of (Kiêu, 1997, Corollary 5.8) to weakly piecewise smooth functions) and, since the mth derivative of f has non-zero jumps, the fact that $(2m+1)$ is the order of the first discontinuous derivative of g. If the measurement function is weakly $(m,1)$-piecewise smooth, a decomposition of the variance of the classical Cavalieri estimator (1), known for $(m,1)$-piecewise smooth functions, holds:

$$\mathbb{V}\mathrm{ar}(\hat{V}(f)) = \mathbb{V}\mathrm{ar}_E(\hat{V}(f)) + Z(t) + r(t). \quad (3)$$

Here, the *extension term* $\mathbb{V}\mathrm{ar}_E(\hat{V}(f))$ explains the general behavior of the variance and is given by $\mathbb{V}\mathrm{ar}_E(\hat{V}(f)) = t^{2m+2}g^{(2m+1)}(0^+)c_m$ for a non-zero constant c_m independent of t and the measurement function f. For instance, $c_0 = -1/6$ and $c_1 = 1/360$. The *Zitterbewegung* $Z(t)$, which is also of order t^{2m+2},

is a finite sum of terms oscillating around 0 with the oscillation given in terms of the jumps $a \in D_{g^{(2m+1)}}$, $a \neq 0$, of $g^{(2m+1)}$ away from the origin, and $r(t)$ is a lower-order remainder. The main message here is that the variance decreases at a rate of t^{2m+2} when f is weakly $(m,1)$-piecewise smooth and this rate of decrease is not achievable with the generalized Cavalieri estimator in the non-equidistant case; see e.g. (Ziegel *et al.*, 2010, Proposition 1). However, as shown in (Stehr and Kiderlen, 2020, Theorem 2.3), if X is stationary with intensity $1/t$, applying a Newton-Cotes estimator of order m also yields a decrease rate of t^{2m+2} for the variance. Moreover, under certain assumptions on the covariance structure of X, the variances of Newton-Cotes estimators decompose similar to (3) but with the Zitterbewegung not necessarily being oscillating. This is the result presented in Lemma 2 below.

Two point models will receive particular interest in this paper as they did in (Ziegel *et al.*, 2010; 2011) and (Stehr and Kiderlen, 2020), namely the perturbed and cumulative models. Note that the perturbed model is formulated in terms of perturbations from the equidistant model, and hence it includes the equidistant model if the perturbations have a degenerate distribution concentrated at 0.

Definition 1 (*Perturbed model*). A stationary point process $X = \{x_k\}_{k \in \mathbb{Z}}$ with intensity $1/t$ is from the perturbed model if $x_k = t(U + k + E_k)$, where U is a uniform random variable on $(0,1)$ and the perturbations $\{E_k\}_{k \in \mathbb{Z}}$ are independent and identically distributed with $\mathbb{E}E_k = 0$ and $|E_k| < 1/2$ almost surely. Moreover, U and $\{E_k\}$ are assumed to be stochastically independent.

Definition 2 (*Cumulative model*). A stationary point process X with intensity $1/t$ is from the cumulative model if it has independent and identically distributed increments $\{\omega_k\}_{k \in \mathbb{Z}}$, where ω_k (necessarily) has expectation t. Furthermore, the increments are assumed to have a continuous distribution and a moment-generating function $\eta \mapsto \mathbb{E}e^{\eta \omega_k}$ which is finite in a neighborhood of 0.

Note that the cumulative model is defined slightly differently than the *model with cumulative errors* in (Stehr and Kiderlen, 2020), hence the different naming. In contrast to that paper, we always require the condition on the moment-generation function. In fact, all variance results in (Stehr and Kiderlen, 2020) for the model with cumulative errors also require the latter condition.

The purpose of this paper is to extend the results of (Stehr and Kiderlen, 2020) on Newton-Cotes estimation. The first goal is to allow for independently p-thinned point processes; in particular, we will state

explicit variance expressions in terms of the thinning probability p for the trapezoidal estimator when sampling from the perturbed or cumulative models. By letting $p = 0$ the exact expressions (Stehr and Kiderlen, 2020, Theorems 2.4 & 2.5) for the original models are recovered. Secondly, we derive and present statistical estimates for the variance of Newton-Cotes estimators applied to any point process. In the special cases of sampling from the perturbed or cumulative model with potential thinning alternative estimates are suggested which appear to be particularly robust. Finally, we substantiate our findings by application to the stereological problem of estimating the volume of a compact convex set $Y \subset \mathbb{R}^3$ from parallel section profiles. We will state variance relations for $\hat{V}_n$ in terms of principal curvatures of the boundary of Y. We will also discuss the case where the joint direction of the section planes is isotropically randomized and show among other things that the variance of the Newton-Cotes estimator then is essentially proportional to the surface area of the object. This is well-known for the classical Cavalieri estimator based on equidistant points (see (Cruz-Orive, 1989; Matheron, 1965)) but exact conditions on Y for this statement to hold have not been specified before, not even in the classical case.

The paper is organized as follows. First we introduce relevant notation and rephrase two important results on the variance of Newton-Cotes estimators in general and in particular of the trapezoidal estimator. We introduce independent p-thinning to model dropouts, show that the results for Newton-Cotes estimators also apply to a thinned process and give explicit variance expressions for $\hat{V}_n$ when sampling from the perturbed or cumulative model with potential thinning. Next, we give an overview of techniques to estimate the variance of Newton-Cotes estimators, concluding with a section on stereological volume estimation.

VARIANCE OF NEWTON-COTES ESTIMATORS

In this section we introduce the notation used throughout the paper, and we give a brief overview of the results on Newton-Cotes estimators presented in (Stehr and Kiderlen, 2020).

For a function $f : \mathbb{R} \to \mathbb{R}$, we let $f^{(k)}$ denote its kth derivative whenever it is defined. At times, we will also use the notation f' and f'' for the first and second derivative, respectively.

For any finite-intensity stationary point process X on $\mathbb{R}$ we define points and increments in X relative to $x \in X$ as follows: For any $j \in \mathbb{Z}$, $s_j(x) = s_j(x; X)$ denotes the jth successor (predecessor for $j < 0$) of x in X with $s_0(x) = x$ by definition, and $h_j(x) = h_j(x; X) = s_j(x) - s_{j-1}(x)$ denotes the distance from the jth successor (predecessor) of x to its left neighbor in X.

Throughout the paper, a subscript u on a point process indicates a scaling to unit intensity, i.e. $X_u = X/t$ has intensity 1 when X has intensity $1/t$, $t > 0$.

We let $\mathbb{P}_X^0$ denote the Palm distribution of X, that is, the conditional distribution of X given that it contains the origin (see e.g. (Schneider and Weil, 2008, Section 3.3)), and we let $\mathbb{E}_X^0$ denote the expectation with respect to the Palm measure. When considering a point process under its Palm distribution, we suppress the dependence on the origin 0, writing for instance $s_j(0) = s_j$ and $h_j(0) = h_j$, $j \in \mathbb{Z}$.

The results in (Stehr and Kiderlen, 2020) are formulated under the overall assumption of finite positive and negative moments of the typical point increment, and for completeness we include it here. Note that (Stehr and Kiderlen, 2020, Assumption 2.1) is slightly stronger than Assumption 1 below, but the arguments and comments in that paper show that (4) is actually sufficient for all results to hold.

Assumption 1. *For fixed Newton-Cotes order $n \in \mathbb{N}$, we have*

$$\mathbb{E}_X^0 h_1^j < \infty \qquad \text{for all} \quad \begin{cases} j \in \mathbb{N} & \text{if} \quad n = 1, \\ j \in \mathbb{Z} & \text{if} \quad n \geq 2. \end{cases} \tag{4}$$

Before proceeding to the results on Newton-Cotes estimators, we introduce a stochastic process $K_m(\cdot) = K_m(\cdot; X)$ on $\mathbb{R}$, $m \in \mathbb{N}_0$. It is commonly referred to as the *(mth) Peano kernel*, as it appears as an integration kernel in a Peano-type error representation of the Newton-Cotes estimator for a weakly $(m, 1)$-piecewise smooth function f (adaption of (Stehr and Kiderlen, 2020, Theorem 2.2)):

$$\hat{V}_n(f) - \int_{\mathbb{R}} f(x) \mathrm{d}x$$
$$= \int_{\mathbb{R}} f^{(m+1)}(r) K_m(r) \mathrm{d}r + \sum_{a \in D_{f^{(m)}}} J_{f^{(m)}}(a) K_m(a) \tag{5}$$

for $m \leq n$. For details on K_m and the summarized properties below, see Sections 3 & 4 in (Stehr and Kiderlen, 2020). The Peano kernel is a piecewise polynomial of order at most $(m+1)$ with coefficient given in terms of the underlying point process X and the order n of the Newton-Cotes estimator. More specifically, $K_m(r)$ depends on the n points in X to the left and right of $r \in \mathbb{R}$. It is translation covariant, i.e. $K_m(r + s; X + s) = K_m(r; X)$ for all $s, r \in \mathbb{R}$. When X is a stationary point process satisfying (4), K_m is

a stationary stochastic process with finite absolute moments of any positive order, i.e. $\mathbb{E}|K_m(r)|^j = \mathbb{E}|K_m(0)|^j < \infty$ for all $j \in \mathbb{N}$ and $r \in \mathbb{R}$. We let $H_m(\cdot) = H_m(\cdot;X)$ denote the covariance function of K_m associated to X, that is, $H_m(s) = \text{Cov}(K_m(r), K_m(r + s))$. It is independent of $r \in \mathbb{R}$ and, by Cauchy-Schwarz inequality, $H_m(s) \leq H_m(0)$ for all $s \in \mathbb{R}$ where equality cannot hold on non-degenerate intervals; see Lemma 1 below. Moreover, K_m is co-homogeneous of degree $m+1$, and, thus

$$H_m(tr;tX) = t^{2m+2}H_m(r;X) = t^{2m+2}H_m(r) \quad (6)$$

for all $r \in \mathbb{R}$ and $0 < t < \infty$.

Lemma 1. *Let $n \in \mathbb{N}$, $m \leq n$ and a compact interval $I \subset \mathbb{R}$ be given. If $H_m(s) = H_m(0)$ for all $s \in I$, then I is a singleton.*

Proof. For contradiction, assume that $H_m(s) = H_m(0)$ for all $s \in I$, where I is a compact interval of positive length $r > 0$. Without loss of generality we may assume that I has rational endpoints. For $j \in \mathbb{Z}$ and $s \in \mathbb{Q} \cap I$, we have $K_m(jr + s) = K_m(jr)$ for all $s \in \mathbb{Q} \cap I$, almost surely. Here we used the assumed equality, stationarity and the fact that $\mathbb{Q}$ is countable. As the intervals in $\{jr + I : j \in \mathbb{Z}\}$ cover $\mathbb{R}$, and any two neighboring intervals have rational endpoints in common, we conclude that

$$K_m(s) = K_m(0) \text{ for all } s \in \mathbb{Q}, \quad (7)$$

almost surely. As K_0 is linear on each connected component of $\mathbb{R} \setminus X$ by (Stehr and Kiderlen, 2020, Eq. (3.7)), Equation (7) implies $K_m \equiv K_m(0)$ on $\mathbb{R} \setminus X$ almost surely when $m = 0$. For $m \geq 1$, the process $K_m(\cdot)$ has continuous paths by (Stehr and Kiderlen, 2020, Lemma 3.1), so (7) implies $K_m \equiv K_m(0)$ on $\mathbb{R}$ almost surely. In either case (5) yields $\hat{V}_n(f) = \int_{\mathbb{R}} f(x)dx$ for almost all realization of X and all weakly $(m,1)$-piecewise smooth functions f. Taking two such functions f_1, f_2 coinciding on all points of X but with $\int_{\mathbb{R}} f_1(x)dx \neq \int_{\mathbb{R}} f_2(x)dx$ we obtain a contradiction. □

An important property introduced in (Stehr and Kiderlen, 2020) is that of admissibility of a point process. Strong admissibility improves the order of the variance when the degree of smoothness of the measurement function exceeds the order of the estimator (Stehr and Kiderlen, 2020, Theorem 2.3), and weak admissibility ensures a variance decomposition similar to (3) as seen in Lemma 2 below. Note that the class of admissible point processes is closed under scaling.

Definition 3. Let $n \in \mathbb{N}$ be given and let X be a stationary point process satisfying (4). X is called *strongly n-admissible* if $\int_0^z H_n(s)ds$ is uniformly bounded in $z \geq 0$. X is called *weakly n-admissible* if $\lim_{z \to \infty} \frac{1}{z} \int_0^z H_n(s)ds = 0$.

Lemma 2. *Let $n \in \mathbb{N}$ be given and assume that X is a stationary point process satisfying (4) and with intensity $1/t > 0$. If f is weakly $(m,1)$-piecewise smooth with $m \leq n$ and covariogram g, then*

$$\mathbb{V}\text{ar}(\hat{V}_n(f)) = \mathbb{V}\text{ar}_E(\hat{V}_n(f)) + Z_m(t) + r(t), \quad (8)$$

where the extension term $\mathbb{V}\text{ar}_E(\hat{V}_n(f))$ is of order t^{2m+2} and is given by

$$\mathbb{V}\text{ar}_E(\hat{V}_n(f)) = (-1)^{m+1} 2g^{(2m+1)}(0^+) H_m(0), \quad (9)$$

the Zitterbewegung $Z_m(t)$ is of order $O(t^{2m+2})$ and satisfies

$$Z_m(t) = (-1)^{m+1} \sum_{a \in D_{g^{(2m+1)}} \setminus \{0\}} J_{g^{(2m+1)}}(a) H_m(a), \quad (10)$$

and the remainder $r(t)$ is of order $O(t^{2m+2})$. If $m < n$ or X is weakly n-admissible, $r(t)$ is of order $o(t^{2m+2})$.

Proof. Applying (6), we may write

$$H_m(r) = H_m(r;X) = t^{2m+2}H_m(r/t;X_u) \quad (11)$$

for all $r \in \mathbb{R}$, where $X_u = X/t$ is a unit-intensity point process. By properties of the Peano kernel, $H_m(0;X_u)$ is independent of t and $H_m(r/t;X_u)$ is uniformly bounded in t. The result now follows from an adaption of (Stehr and Kiderlen, 2020, Proposition 6.1) to weakly piecewise smooth functions. □

Both the perturbed and cumulative models are strongly n-admissible for all $n \in \mathbb{N}$ and, in particular, the remainder $r(t)$ in Lemma 2 is of order $o(t^{2m+2})$. Even stronger, if X is from the perturbed model then $r \mapsto H_m(r;X_u)$ is periodic with period 1 for all $m \in \mathbb{N}_0$ and sufficiently large $|r|$, and $\int_r^{r+1} H_m(s;X_u)ds = 0$ (Stehr and Kiderlen, 2020, Lemma 7.1). Hence, if X has intensity $1/t$, the Zitterbewegung (10) is of order t^{2m+2} and it is a finite sum of terms each oscillating around 0. If X is from the cumulative model there is $\varepsilon > 0$ such that $H_m(r/t;X_u) = O(e^{-\varepsilon r/t})$ for all $r \in \mathbb{R}$ as $t \downarrow 0$, due to (Stehr and Kiderlen, 2020, Lemma 7.3). Hence, by (11), it follows that the Zitterbewegung $Z(t)$ is $o(t^{2m+2})$ and thus (8) reads

$$\mathbb{V}\text{ar}(\hat{V}_n(f)) = \mathbb{V}\text{ar}_E(\hat{V}_n(f)) + o(t^{2m+2}) \quad (12)$$

as $t \downarrow 0$ in this case.

In (Stehr and Kiderlen, 2020, Eq.'s (7.8) and (7.9)) $H_m(0;X_u)$ is simplified for $n = 1$ and $m \in \{0,1\}$. Using these representations, the following corollary to Lemma 2 is a consequence of the fact that $\mathbb{E}_X^0 h_1^j = t^j \mathbb{E}_{X_u}^0 h_1^j$ for all $j \in \mathbb{N}$.

Corollary 3. *Let X be a stationary point process satisfying* (4) *and with intensity* $1/t > 0$, *and let f be weakly* $(m, 1)$-*piecewise smooth with covariogram g. For* $m = 0$ *we have*

$$\begin{aligned} \mathbb{V}\text{ar}_E(\hat{V}_1(f)) &= -g^{(1)}(0^+)\tfrac{1}{6}\tfrac{1}{t}\mathbb{E}_X^0 h_1^3 \\ &= -t^2 g^{(1)}(0^+)\tfrac{1}{6}\mathbb{E}_{X_u}^0 h_1^3, \end{aligned} \quad (13)$$

and for $m = 1$ *we have*

$$\begin{aligned} \mathbb{V}\text{ar}_E(\hat{V}_1(f)) &= g^{(3)}(0^+)\tfrac{1}{360}\big(6\tfrac{1}{t}\mathbb{E}_X^0 h_1^5 - 5(\tfrac{1}{t}\mathbb{E}_X^0 h_1^3)^2\big) \\ &= t^4 g^{(3)}(0^+)\tfrac{1}{360}\big(6\mathbb{E}_{X_u}^0 h_1^5 - 5(\mathbb{E}_{X_u}^0 h_1^3)^2\big). \end{aligned}$$

INDEPENDENT p-THINNING

In this section we consider the point process X to be obtained by so-called independent p-thinning, where $0 \le p < 1$ denotes the probability of an initial point being removed. We let $\tilde{X} = \{\tilde{x}_k\}_{k \in \mathbb{Z}}$ denote the underlying stationary point process on $\mathbb{R}$ with finite intensity $1/\tilde{t} > 0$, and we define X by $X = \{\tilde{x}_k \in \tilde{X} : U_k > p\}$, where $\{U_k\}_{k \in \mathbb{Z}}$ denotes an i.i.d. sequence of uniform $(0, 1)$ variables independent of $\tilde{X}$. Hence, U_k is the thinning variable associated to the point $\tilde{x}_k \in \tilde{X}$, and therefore $\tilde{x}_k \in X$ with probability $\mathbb{P}(U_k > p) = 1 - p$. We refer to $\tilde{X}$ as the *initial* process and X as the *observed* or *thinned* process. If $p = 0$ no dropouts occur and naturally $\tilde{X} = X$. If X has intensity $1/t$, it is not difficult to see that $t = \tilde{t}/(1-p)$. As in the previous section we let $X_u = X/t$ denote the scaled unit-intensity process of X.

As mentioned previously, the correction method introduced in (Ziegel *et al.*, 2011, Section 4) coincides with the trapezoidal estimator applied to the thinned process X if the initial process $\tilde{X}$ is equidistant. To see this, recall that $\alpha_1(x) = (h_1(x) + h_0(x))/2$ for all $x \in X$, where, for some $k \in \mathbb{Z}$, $x = \tilde{x}_k$. If $\tilde{X}$ has intensity $1/\tilde{t}$ and $x \in X$ we have $h_1(x) = \tilde{t}\min\{j \ge 1 : U_{k+j} > p\}$ and $h_0(x) = \tilde{t}\min\{j \ge 1 : U_{k-j} > p\}$. Consequently, the trapezoidal estimator reads

$$\hat{V}_1(f) = \tilde{t}\sum_{k \in \mathbb{Z}} \psi_k f(\tilde{x}_k),$$

with the weight

$$\begin{aligned} \psi_k = \tfrac{1}{2}\mathbb{1}_{(U_k > p)}\big(&\min\{j \ge 1 : U_{k+j} > p\} \\ &+ \min\{j \ge 1 : U_{k-j} > p\}\big). \end{aligned}$$

This is exactly the estimator introduced in (Ziegel *et al.*, 2011, Section 4). Note that the estimators only coincide when the underlying point process is equidistant, and if this is not the case, the trapezoidal estimator is superior. For instance, the correction

method of (Ziegel *et al.*, 2011) is of order $\tilde{t}^3$ when the underlying process is from the perturbed model and the measurement function is weakly $(1, 1)$-piecewise smooth, whereas the trapezoidal estimator is always of order t^4 (and hence also of order $\tilde{t}^4$) for weakly $(1, 1)$-piecewise smooth functions; see (Ziegel *et al.*, 2011, Proposition 4) and Lemma 2 above.

We now aim for variance expressions of Newton-Cotes estimators when applied to the observed process X. As the results in the previous section rest on Assumption 1, we have to assure that this condition is satisfied by the observed process X. Lemma 4 below shows that this is in fact the case if the initial process satisfies Assumption 1. Subsequently, we find explicit variance expressions for the trapezoidal estimator applied to the observed process X, when the initial process $\tilde{X}$ is from the perturbed or cumulative model. We show that X is strongly n-admissible in either case, and hence it satisfies (8) with the remainder $r(t)$ of order $o(t^{2m+2})$. $\mathbb{V}\text{ar}_E(\hat{V}_n(f))$ is given by (9) for arbitrary n and explicitly in Corollary 3 for the trapezoidal estimator.

Lemma 4. *If the initial process $\tilde{X}$ satisfies* (4), *then so does the thinned process X.*

Proof. Between two consecutive points in the thinned process X, the number of removed points from the initial process $\tilde{X}$ follows a geometric distribution on $\mathbb{N}_0$ with success probability $(1-p)$. By independence, this implies

$$\mathbb{E}_X^0 h_1^j = \sum_{k=1}^{\infty}(1-p)p^{k-1}\mathbb{E}_{\tilde{X}}^0 s_k^j. \quad (14)$$

As $\mathbb{E}_X^0 h_1^j \le \mathbb{E}_{\tilde{X}}^0 h_1^j < \infty$ for all $j < 0$ we may assume $j \ge 0$. By an application of the multinomial theorem and Hölder's inequality, $\mathbb{E}_{\tilde{X}}^0 s_k^j$ has the upper bound

$$\begin{aligned} \mathbb{E}_{\tilde{X}}^0(h_1 + \cdots + h_k)^j &\le \sum_{j_1 + \cdots + j_k = j}\binom{j}{j_1, \ldots, j_k}\mathbb{E}_{\tilde{X}}^0[h_1^j] \\ &= k^j\mathbb{E}_{\tilde{X}}^0[h_1^j], \end{aligned}$$

and we conclude the proof as $\mathbb{E}_X^0 h_1^j \le \mathbb{E}_{\tilde{X}}^0[h_1^j](1-p)\sum_{k=1}^{\infty} k^j p^{k-1} < \infty$, due to the assumption on $\tilde{X}$. $\square$

As mentioned, a point process from the perturbed model is strongly n-admissible for any $n \in \mathbb{N}$, and this is also the case under independent thinning.

Lemma 5. *Let the initial process $\tilde{X}$ be from the perturbed model and let X be the observed process. For all $n, j \in \mathbb{N}$ and $m \in \mathbb{N}_0$, X satisfies*

(i) $H_m(r; X_u) = H_m(r+1; X_u) + o(r^{-j})$,

(ii) $\int_r^{r+1} H_m(s; X_u)\mathrm{d}s = o(r^{-j})$,

as $r \to \pm\infty$. Furthermore,

(iii) X is strongly n-admissible.

Proof. As $r \mapsto H_m(r)$ is an even function due to stationarity, it is enough to show the claims (i) and (ii) for positive $r \to \infty$.

Before proving (i)–(iii) we introduce some notation which will be used throughout the proof. First, to simplify notation, we will write $K_m(r) = K_m(r; X_u)$ and $H_m(r) = H_m(r; X_u)$, as we do not consider the observed process but only the associated unit-scaled process. For $s \in \mathbb{R}$ and $r \geq 0$ let

$$A_-(s) = A_-(s; X_u) = \left\{ \#\left(X_u \cap \left[s - \tfrac{r}{3}, s\right)\right) \geq n \right\}, \quad \text{and}$$
$$A_+(s) = A_+(s; X_u) = \left\{ \#\left(X_u \cap \left[s, s + \tfrac{r}{3}\right)\right) \geq n \right\}$$

be the events that X_u has at least n points to the left and right of s, respectively, within a distance of at most $r/3$. Furthermore define $I_\pm(s) = I_\pm(s; X_u) = 1_{A_\pm(s; X_u)}$ as the associated indicator functions, and note that they are vanishing for small r. We also define the shifted process $X^* = X_u - U = \{k + E_k : U_k > p, k \in \mathbb{Z}\}$ and correspondingly $K_m^*(s) = K_m(s; X^*)$ and $I_\pm^*(s) = I_\pm(s; X^*)$. Due to the i.i.d. property of perturbations and thinning variables, the distributional equivalence

$$K_m^*(s)I_\pm^*(s) \overset{\mathscr{D}}{=} K_m^*(s+1)I_\pm^*(s+1) \qquad (15)$$

holds for all $s \in \mathbb{R}$, where the translation covariance of $K_m(s; X)$ and $I_\pm(s; X)$ was used. Furthermore, by construction, $K_m^*(s)I_-^*(s)$ and $K_m^*(s)I_+^*(s)$ only depend on the variables

$$\{(E_i, U_i) : i + E_i \geq s - r/3\}$$
$$\subseteq \{(E_i, U_i) : i \geq s - r/3 - 1/2\},$$

and

$$\{(E_i, U_i) : i + E_i < s + r/3\}$$
$$\subseteq \{(E_i, U_i) : i < s + r/3 + 1/2\},$$

respectively. In particular, $K_m^*(s)I_-^*(s)$ and $K_m^*(s')I_+^*(s')$ are stochastically independent for $s - s' \geq 2r/3 + 1$.

Let $n \in \mathbb{N}$ and $m \in \mathbb{N}_0$ be fixed and consider $r \geq 0$. To prove (i) we observe by stationarity that $\mathbb{P}((A_-(s) \cap A_+(0))^c) \leq 2\mathbb{P}(A_+(0)^c) = 2\mathbb{P}(\#(X_u \cap [0, r/3)) < n)$, and thus

$$\mathbb{P}\left((A_-(s) \cap A_+(0))^c\right) = o(r^{-j}) \qquad (16)$$

for all $j \in \mathbb{N}$ and $s \in \mathbb{R}$, as $r \to \infty$. This is seen as follows: With $\lfloor \cdot \rfloor$ denoting integer part, there are at least $N_r = \lfloor r/3 \rfloor - 2$ points in $\tilde{X}_u \cap [0, r/3)$, and the number B of these points also in X_u is a binomial

random variable with N_r trials and success probability $1 - p$. As $p < 1$ it is not difficult to see that $\mathbb{P}(B < n)$ is of order $o(N_r^{-j})$ for any $j \in \mathbb{N}$, and hence also of order $o(r^{-j})$ as $r \to \infty$. Equation (16) now follows as $\mathbb{P}(\#(X_u \cap [0, \tfrac{r}{3})) < n) \leq \mathbb{P}(B < n)$.

By the Cauchy-Schwarz inequality and the fact that $\mathbb{E}K_m^4(r) = \mathbb{E}K_m^4(0) < \infty$, (16) implies that

$$\mathbb{E}[K_m(r)K_m(0)] = S(r) + o(r^{-j}) \qquad (17)$$

for all $j \in \mathbb{N}$, where

$$S(r) = \mathbb{E}[K_m(r)I_-(r)K_m(0)I_+(0)]$$

is defined for notational convenience. Conditioning on U and using the translation covariance of K_m and $I_\pm$, $S(r)$ reads

$$S(r) = \int_0^1 \mathbb{E}[K_m^*(r-u)I_-^*(r-u)K_m^*(-u)I_+^*(-u)]\mathrm{d}u$$

$$= \int_0^1 \mathbb{E}[K_m^*(r-u)I_-^*(r-u)]\mathbb{E}[K_m^*(-u)I_+^*(-u)]\mathrm{d}u$$

for $r \geq 3$, where it has been used that $K_m^*(r-u)I_-^*(r-u)$ and $K_m^*(-u)I_+^*(-u)$ are independent for such r. Equation (15) now implies that

$$S(r) = S(r+1)$$

for all $j \in \mathbb{N}$ and $r \geq 3$. Using this and (17) with r and $r+1$ yields (i).

Defining $R_\pm^*(s) = \mathbb{E}[K_m^*(s)I_\pm^*(s)]$, we note that

$$\int_0^1 R_\pm^*(s-u)\mathrm{d}u = \mathbb{E}[K_m(s)I_\pm(s)] \qquad (18)$$
$$= \mathbb{E}[K_m(0)I_\pm(0)] = \mathbb{E}[K_m(0)] + o(r^{-j})$$

for all $s \in \mathbb{R}$, as $r \to \infty$. The first equality is seen by conditioning on U, the second equality is due to stationarity, and the third equality is obtained by Cauchy-Schwarz' inequality using the fact that $\mathbb{P}(A_\pm(0)^c) = o(r^{-j})$ and $\mathbb{E}K_m^2(0) < \infty$. Equations (17)–(18), Fubini's theorem and a substitution now yield

$$\int_r^{r+1} \mathbb{E}[K_m(0)K_m(s)]\mathrm{d}s$$

$$= \int_0^1 R_+^*(-u)\int_0^1 R_-^*(r+1-u-s)\mathrm{d}s\mathrm{d}u + o(r^{-j})$$

$$= (\mathbb{E}[K_m(0)])^2 + o(r^{-j}),$$

which proves (ii).

That X_u (and consequently X by (6)) is strongly n-admissible is easily seen by (ii) as

$$\left| \int_0^z H_n(r)\mathrm{d}r \right| = \left| \sum_{i=0}^{\lfloor z \rfloor - 1} \int_i^{i+1} H_n(s)\mathrm{d}s + \int_{\lfloor z \rfloor}^z H_n(s)\mathrm{d}s \right|$$

$$\leq \sum_{i=0}^\infty \left| \int_i^{i+1} H_n(s)\mathrm{d}s \right| + \mathbb{V}\mathrm{ar}(K_n(0)) < \infty.$$

This is exactly assertion (iii). $\qquad \square$

Theorem 6. *Let the initial process $\tilde{X}$ be from the perturbed model with intensity $1/\tilde{t}$ and let X be the observed process with intensity $1/t$. Let $\theta_k = \mathbb{E}(E_1 - E_0)^k$ be the kth moment of the difference in perturbations. If f is weakly $(m,1)$-piecewise smooth with covariogram g, then $\mathbb{V}\mathrm{ar}_E(\hat{V}_1(f))$ coincides with*

$$-\tilde{t}^2 g^{(1)}(0^+)\frac{1}{6}\Big(1 + 3\theta_2 + \frac{6p}{(1-p)^2}\Big), \quad (19)$$

for $m = 0$, and

$$\tilde{t}^4 g^{(3)}(0^+)\frac{1}{360}\Big(1 + \frac{120p^3 + 300p^2 + 120p}{(1-p)^4} \\ + 30\theta_2\frac{p^2 + 4p + 1}{(1-p)^2} - 45\theta_2^2 + 30\theta_4\Big), \quad (20)$$

for $m = 1$. Moreover, the Zitterbewegung $Z_m(t)$ is of order t^{2m+2} (hence $\tilde{t}^{2m+2}$) and it is a finite sum of terms each oscillating around 0.

Before proving this result, we make a few comments on the extension term. As mentioned, the trapezoidal estimator and the correction method of (Ziegel *et al.*, 2011) coincide when the initial point process is equidistant. In particular, when $\theta_k = 0$ for all $k \in \mathbb{N}$, the result above should match (Ziegel *et al.*, 2011, Proposition 3). However, according to (Ziegel *et al.*, 2011), the second summand in (20) is erroneously claimed to be $(120p^3 + 390p^2 + 120p)/(1-p)^4$.

Considering perturbed sampling without thinning ($p = 0$), how much do we underestimate the extension term if we use the classical variance, that is, if we disregard the perturbations by setting $\theta_k = 0$ above? In the rather extreme case where E_1 is uniform on $(-1/2, 1/2)$ the extension term equals $3/2$ and $27/4$ times the classical extension term for $m = 0, 1$, respectively. By comparison, the extension term of the generalized Cavalieri estimator equals 2 times the classical extension term when $m = 0$; see (Ziegel *et al.*, 2011, p. 190). In the worst case where the errors E_i are arbitrarily close to $\pm 1/2$, the extension term approaches $5/2$ and $79/4$ times the classical extension term for $m = 0, 1$, respectively.

Proof. As processes from the perturbed model satisfy (4), also X satisfies (4) due to Lemma 4. Recall that $\theta_k = \mathbb{E}(E_j - E_0)^k = \mathbb{E}(E_1 - E_0)^k$ for all $j \neq 0$, with $\theta_k = 0$ for all odd k. It is not difficult to see that $\mathbb{E}_{\tilde{X}}^0 s_j^5 = \tilde{t}^5\mathbb{E}(j + E_j - E_0)^5 = \tilde{t}^5(j^5 + 10\theta_2 j^3 + 5\theta_4 j)$ for all $j \in \mathbb{N}$, and consequently, using (14),

$$\mathbb{E}_X^0 h_1^5 = \sum_{j=1}^{\infty} p^{j-1}(1-p)\mathbb{E}_{\tilde{X}}^0 s_j^5$$

$$= \tilde{t}^5 \sum_{j=1}^{\infty} p^{j-1}(1-p)(j^5 + 10\theta_2 j^3 + 5\theta_4 j)$$

$$= \tilde{t}^5\Big(\frac{p^4 + 26p^3 + 66p^2 + 26p + 1}{(1-p)^5} \\ + 10\theta_2\frac{p^2 + 4p + 1}{(1-p)^3} + 5\theta_4\frac{1}{1-p}\Big).$$

By similar arguments it is shown that $\mathbb{E}_X^0 h_1^3 = \tilde{t}^3((1 + 3\theta_2)/(1-p) + 6p/(1-p)^3)$. Since $t = \tilde{t}/(1-p)$, (19) and (20) follow by applying Corollary 3 to X. The claim on the Zitterbewegung is a consequence of Lemma 5 and (10). $\qquad\square$

Next we turn to the cumulative model from Definition 2. Interestingly, this model class is closed under independent thinning. In the following, $\{\tilde{\omega}_k\}$ and $\{\omega_k\}$ denotes the increments of the initial process $\tilde{X}$ and the thinned process X, respectively.

Lemma 7. *If the initial process $\tilde{X}$ is from the cumulative model, then so is the thinned process X. In this case, for all $n \in \mathbb{N}$ and $m \in \mathbb{N}_0$, there is some $\varepsilon > 0$ such that $H_m(s) = O(e^{-\varepsilon s})$ as $s \to \infty$. In particular, X is strongly n-admissible.*

Proof. Let $\tilde{F}$ be the continuous distribution function of the initial increments. The independence of these increments and the thinning variables show that X has independent and identically distributed increments with continuous distribution function F given by $F(x) = \sum_{k=1}^{\infty}(1-p)p^{k-1}\tilde{F}^{*k}(x)$. Here $\tilde{F}^{*k}$ denotes the k-fold convolution of $\tilde{F}$, that is, the distribution of $\tilde{\omega}_1 + \cdots + \tilde{\omega}_k$. Furthermore, $\mathbb{E}e^{\eta\omega_1} = \sum_{k=1}^{\infty}(1 - p)p^{k-1}(\mathbb{E}e^{\eta\tilde{\omega}_1})^k$, which is finite if $\mathbb{E}e^{\eta\tilde{\omega}_1} < 1/p$. Since there is $\tilde{\eta} > 0$ such that $\mathbb{E}e^{\tilde{\eta}\tilde{\omega}_k} < \infty$, a dominated convergence argument shows that $\mathbb{E}e^{\eta\tilde{\omega}_1} < 1/p$ if we choose $\eta > 0$ small enough. The claim on the convergence rate of H_m is now a consequence of (Stehr and Kiderlen, 2020, Lemma 7.3). $\qquad\square$

Theorem 8. *Let the initial process $\tilde{X}$ be from the cumulative model with intensity $1/\tilde{t}$ and let X be the observed process with intensity $1/t$. Define $\tilde{v}_k = \mathbb{E}\tilde{\omega}_1^k/\tilde{t}^k$ as the kth moment of the increments in the initial unit-intensity scaled process $\tilde{X}_u$. If f is weakly $(m,1)$-piecewise smooth with covariogram g, then $\mathbb{V}\mathrm{ar}_E(\hat{V}_1(f))$ coincides with*

$$-\tilde{t}^2 g^{(1)}(0^+)\frac{1}{6}\Big(\tilde{v}_3 + \tilde{v}_2\frac{6p}{1-p} + \frac{6p^2}{(1-p)^2}\Big),$$

for m = 0, and

$$\tilde{t}^4 g^{(3)}(0^+) \frac{1}{360} \left(6\tilde{v}_5 - 5\tilde{v}_3^2 + \frac{60p(\tilde{v}_4 + \tilde{v}_3\tilde{v}_2)}{1-p} \right.$$
$$\left. + \frac{60p^2(5\tilde{v}_3 + 6\tilde{v}_2^2)}{(1-p)^2} + \frac{\tilde{v}_2 1080p^3}{(1-p)^3} + \frac{540p^4}{(1-p)^4} \right),$$

for m = 1. Moreover, $Z_m(t)$ is of order $o(t^{2m+2})$ (hence $o(\tilde{t}^{2m+2})$) for $m = 0, 1$.

Proof. By definition, the cumulative model has finite exponential expectation and in particular (4) is satisfied for both the initial and the observed model. By the multinomial theorem, the independence and identical law of the increments, and by the fact that $\tilde{v}_1 = 1$,

$$\mathbb{E}(\tilde{\omega}_1 + \cdots + \tilde{\omega}_j)^5 / \tilde{t}^5$$

$$= \sum_{k_1 + \cdots + k_j = 5} \binom{5}{k_1, \ldots, k_j} \prod_{i=1}^{j} \tilde{v}_{k_i}$$

$$= \binom{j}{1} \tilde{v}_5 + \binom{j}{2}\binom{2}{1}(5\tilde{v}_4 + 10\tilde{v}_3\tilde{v}_2)$$

$$+ \binom{j}{3}\binom{3}{2}(20\tilde{v}_3 + 30\tilde{v}_2^2)$$

$$+ \binom{j}{4}\binom{4}{3}60\tilde{v}_2 + \binom{j}{5}120$$

$$= j\tilde{v}_5 + j(j-1)(5\tilde{v}_4 + 10\tilde{v}_3\tilde{v}_2)$$
$$+ j(j-1)(j-2)(10\tilde{v}_3 + 15\tilde{v}_2^2)$$
$$+ j(j-1)(j-2)(j-3)10\tilde{v}_2$$
$$+ j(j-1)(j-2)(j-3)(j-4)$$

for all $j \in \mathbb{N}$. Consequently,

$$\mathbb{E}_X^0 h_1^5 = \sum_{j=1}^{\infty} p^{j-1}(1-p)\mathbb{E}_{\tilde{X}}^0 s_j^5$$

$$= \sum_{j=1}^{\infty} p^{j-1}(1-p)\mathbb{E}(\tilde{\omega}_1 + \cdots + \tilde{\omega}_j)^5$$

$$= \tilde{t}^5 \left(\frac{\tilde{v}_5}{1-p} + \frac{(10\tilde{v}_4 + 20\tilde{v}_3\tilde{v}_2)p}{(1-p)^2} \right.$$
$$+ \frac{(60\tilde{v}_3 + 90\tilde{v}_2^2)p^2}{(1-p)^3} + \frac{240\tilde{v}_2 p^3}{(1-p)^4}$$
$$\left. + \frac{120p^4}{(1-p)^5} \right).$$

Similarly, it is seen that $\mathbb{E}(\tilde{\omega}_1 + \cdots + \tilde{\omega}_j)^3 / \tilde{t}^3 = j\tilde{v}_3 + 3j(j-1)\tilde{v}_2 + j(j-1)(j-2)$ for all $j \in \mathbb{N}$ and hence $\mathbb{E}_X^0 h_1^3 = \tilde{t}^3 (\tilde{v}_3/(1-p) + \tilde{v}_2 6p/(1-p)^2 + 6p^2/(1-p)^3)$. Since the relation $t = \tilde{t}/(1-p)$ holds, the lemma is concluded by Corollary 3 (applied to X), Lemma 7 and (12). $\qquad\square$

VARIANCE ESTIMATION

Traditionally, the variance of the classical Cavalieri estimator is approximated by its extension term, and hence an estimation of the extension term serves as an estimation of the variance as a whole. It is debatable if this is appropriate, as a particular sample may result in a large Zitterbewegung. Hence, using the extension term only, one risks actually underestimating the variance. However, at least in the stereological application that is discussed at the end of this paper, we see that the Zitterbewegung never can exceed the extension term, see (33), so $2\mathbb{V}\text{ar}_E(\hat{V}_n)$ is an upper bound for the variance, only neglecting the remainder term in this application. The extension term is (relatively) easy to estimate and is traditionally used as an approximation for the unknown variance of the classical Cavalieri estimator (1). In view of the variance decomposition (8) we can follow these lines also for the new Newton-Cotes estimators and focus therefore on the estimation of the extension term. The estimation naturally depends on the available information on the sampling points. In this section we discuss a general estimation approach based on the observed process, and, if we sample from the perturbed or cumulative models with information on the initial process available, we mention another approach for the trapezoidal estimator exploiting the exact representations from Theorems 6 and 8.

Throughout this section we assume that the measurement function f is weakly $(m, 1)$-piecewise smooth with known $m \in \mathbb{N}_0$, and we estimate the variance based on the points $X \cap L$ of the observed process X falling inside a bounded interval $L \subseteq \mathbb{R}$. We assume that f is known at the points $X \cap L$ and that L contains the support of f. We let the observed process X have intensity $1/t$.

Before proceeding, we introduce an estimation procedure for the Palm expectation of a function of increments as this will be of great relevance below. More precisely, for the point process X and a fixed $n \in \mathbb{N}$, we aim to estimate $\Theta = t^{-1}\mathbb{E}_X^0 F(h_1, \ldots, h_n)$, where $F : \mathbb{R}^n \to \mathbb{R}$ is an integrable function. For the n largest points in $X \cap L$, some of the subsequent n point increments are not accessible from the information of X in L. To correct for this censoring close to the boundary of L, we use the following Hanisch estimator, where we only consider points x in $X \cap L$ for which all of the subsequent n increments $h_1(x), \ldots, h_n(x)$ are observed:

$$\hat{\Theta} = \sum_{\substack{x \in X \cap L \\ s_n(x) \in L}} \frac{F(h_1(x), \ldots, h_n(x))}{\mathcal{H}^1(L \cap (L - s_n(x) + x))}. \qquad (21)$$

Here $\mathscr{H}^1$ denotes the 1-dimensional Hausdorff measure, that is, the length measure. By the refined Campbell theorem (Schneider and Weil, 2008, Theorem 3.5.3) it is seen that $\hat{\Theta}$ is unbiased for Θ. To estimate the intensity $1/t$ we simply use the unbiased estimator $\#(X \cap L)/\mathscr{H}^1(L)$.

As the extension term (9) factorizes similar to the classical case, we can estimate the contribution from the point process (through H_m) and from the measurement function (through g) separately. We first estimate the variance $H_m(0)$ depending on the point process X only. More specifically, if $n \in \mathbb{N}$ is the order of the Newton-Cotes estimator, using (Stehr and Kiderlen, 2020, Eq.'s (3.6) and (3.7)) and the refined Campbell theorem (Schneider and Weil, 2008, Theorem 3.5.3), it is not difficult to see that there are (known) rational functions $p_m, q_m : \mathbb{R}^n \to \mathbb{R}$ such that

$$H_m(0) = \mathbb{E}K_m^2(0) - (\mathbb{E}K_m(0))^2$$
$$= \tfrac{1}{t}\mathbb{E}_X^0 p_m(h_1, \ldots, h_n) - (\tfrac{1}{t}\mathbb{E}_X^0 q_m(h_1, \ldots, h_n))^2,$$

where the latter term vanishes for all $m < n$; see (Stehr and Kiderlen, 2020, Lemma 4.1). For $n = 1$, comparing with Corollary 3, we see that $p_0(h_1) = h_1^3/12$, $p_1(h_1) = h_1^5/120$ and $q_1(h_1) = h_1^3/12$. Applying (21) with F substituted by p_m and q_m, $H_m(0)$ can be estimated by the information of X in L.

Secondly, we estimate the derivative $g^{(2m+1)}(0^+)$. To do so, we apply similar techniques as in (Kiêu, 1997, Section 6.2), where we explicitly use the fact that the covariogram g is weakly $(2m+1, 1)$-piecewise smooth. An induction argument using Lemma 15 in the appendix (an adaption of the refined partial integration formula (Kiêu, 1997, Lemma 4.1) to weakly piecewise smooth functions) yields

$$g(y) = \sum_{j=0}^{2m+1} \frac{g^{(j)}(0^+)}{j!} y^j + R_{2m+1}(y;g) \qquad (22)$$

for all $y > 0$, where the remainder R_{2m+1} satisfies

$$R_{2m+1}(y;g) = \sum_{a \in D_{g^{(2m+1)}} \cap (0,y)} \frac{(y-a)^{2m+1}}{(2m+1)!} J_{g^{(2m+1)}}(a)$$
$$+ \frac{1}{(2m+1)!} \int_0^y g^{(2m+2)}(t)(y-t)^{2m+1}\mathrm{d}t.$$

For $k \in \mathbb{N}_0$, we estimate the covariogram g using

$$\hat{g}(k,t) = t \sum_{x \in X \cap L} f(x)f(s_k(x)), \qquad (23)$$

which, by the refined Campbell theorem (Schneider and Weil, 2008, Theorem 3.5.3) and the fact that L contains the support of f, satisfies $\mathbb{E}\hat{g}(k,t) = \mathbb{E}_X^0 g(s_k)$.

Let β_k be given by $\beta_k = g^{(k)}(0^+)/k!$. As $g^{(j)}$ is odd and continuous for all odd $j \le 2m$, we find using (22) that

$$\hat{g}(k,t) = \sum_{j=0}^{m} \mathbb{E}_X^0[s_k^{2j}]\beta_{2j} + \mathbb{E}_X^0[s_k^{2m+1}]\beta_{2m+1} + \varepsilon_{k,t},$$
$$(24)$$

where the error $\varepsilon_{k,t}$ arises from a Taylor expansion and the covariogram estimation. Consequently, $\mathbb{E}\varepsilon_{k,t} = \mathbb{E}_X^0 R_{2m+1}(s_k; g)$, and hence

$$\mathbb{E}\varepsilon_{k,t} = \mathbb{E}_X^0 \Big[\sum_{\substack{a \in D_{g^{(2m+1)}} \\ a > 0}} \frac{(s_k - a)_+^{2m+1}}{(2m+1)!} J_{g^{(2m+1)}}(a)$$
$$+ \frac{1}{(2m+1)!} \int_0^\infty g^{(2m+2)}(r)(s_k - r)_+^{2m+1}\mathrm{d}r \Big],$$

where $(x - r)_+^k = 1_{x>r}(x - r)^k$. By definition, $s_k^j(0) = t^j s_k^j(0; X_u)$, and as $\mathbb{E}_{X_u}^0 s_k^j < \infty$ for all $k, j \in \mathbb{N}_0$ and $g^{(2m+2)}$ is integrable by assumption, we conclude by dominated convergence that $\mathbb{E}\varepsilon_{k,t} = o(t^{2m+1})$ as $t \downarrow 0$. Applying this convergence rate, the proposition below follows from (24).

Proposition 9. *Let X be the observed process with intensity $1/t$ and let f be weakly $(m,1)$-piecewise smooth with covariogram g. For $N \ge m+1$ define the $(N+1) \times (m+2)$ matrix $Q_X = (\mathbb{E}_X^0 s_i^j)_{i \in \{0,\ldots,N\}, j \in \{0,2,\ldots,2m,2m+1\}}$. If the linear system $Q_X \eta = (0,\ldots,0,1)^\top$ has a solution $\eta = \{\eta_k\}_{k=0}^N$, then, with $\hat{g}(k,t)$ given by (23),*

$$\hat{\beta}_{2m+1} = \sum_{k=0}^N \hat{g}(k,t)\eta_k \qquad (25)$$

estimates $g^{(2m+1)}(0^+)/(2m+1)!$ with a bias of order $o(1)$ as $t \downarrow 0$.

In the corollary below we present estimators for the extension term of the trapezoidal estimator when $m \in \{0,1\}$. To use Proposition 9 for $m = 1$, we require the denominator in (27) to be non-zero. This ensures that the linear system has a solution, and it is for instance always satisfied for the perturbed model. Note that for $m = 0$ we choose $N = 2$ larger than necessary since this is traditionally used in equidistant sampling; see e.g. (Kiêu, 1997, Formula (6.9)) and recall (13). If we instead choose $N = 1$ we estimate $g'(0^+)$ by $(-\hat{g}(0,t) + \hat{g}(1,t))/t$.

Corollary 10. *Let X be the observed process with intensity $1/t$ and let f be weakly $(m,1)$-piecewise smooth. Define $\tilde{g}(k,t) = \hat{g}(k,t)/t$, with $\hat{g}(k,t)$ given by (23), and $\gamma_{i,j} = \tfrac{1}{t}\mathbb{E}_X^0 s_i^j$. The extension term $\mathbb{V}\mathrm{ar}_E(\hat{V}_1(f))$ of the trapezoidal estimator is estimated with a bias of order $o(1)$ by:*

$$\left(\tfrac{1}{4}\tilde{g}(0,t) - \tfrac{1}{3}\tilde{g}(1,t) + \tfrac{1}{12}\tilde{g}(2,t)\right)\gamma_{1,3} \qquad (26)$$

for m = 0, and

$$\frac{\tilde{g}(0,t)(\gamma_{2,2} - \gamma_{1,2}) - \tilde{g}(1,t)\gamma_{2,2} + \tilde{g}(2,t)\gamma_{1,2}}{\gamma_{1,2}\gamma_{2,3} - \gamma_{2,2}\gamma_{1,3}} \quad (27)$$
$$\times \left(\tfrac{1}{10}\gamma_{1,5} - \tfrac{1}{12}(\gamma_{1,3})^2\right)$$

for m = 1, if the denominator in (27) is non-zero. All $\gamma_{i,j}$ can be unbiasedly estimated using $X \cap L$ by the Hanish estimator (21).

We have described a general estimation approach where the only requirement on the observed process X (other than stationarity) is the existence of a solution to the linear system in Proposition 9. However, if the model of X is known this can also be used to construct other estimates for $H_m(0)$ and the weights $\{\eta_k\}$ in (25). In particular, if both the observed process X and its underlying initial process $\tilde{X}$ are accessible, a decomposition of the extension term similar to Theorems 6 and 8 can be constructed, representing the Palm expectations in terms of the thinning probability p and increments of the initial process. Preliminary simulations for the trapezoidal estimator under perturbed and cumulative sampling indicate that such a model-specific estimation has a smaller variance than the general approach, and furthermore it appears to be much more robust for varying intensity, especially when thinning is present. This will be investigated further in a future paper.

AN APPLICATION TO VOLUME ESTIMATION

We now specialize the above results to stereological volume estimation. This had been the original starting point of our research and extends the settings of the well-established Cavalieri estimator and its generalizations. The target is the volume of a compact set $Y \subset \mathbb{R}^d$. For a unit vector $\omega \in S^{d-1}$ let $\omega^\perp$ be the hyperplane with normal ω and let $\mathscr{H}^k$ be the k-dimensional Hausdorff-measure in $\mathbb{R}^d$. We assume that the measurement function

$$f(x) = \mathscr{H}^{d-1}\big(Y \cap (x\omega + \omega^\perp)\big) \quad (28)$$

is available at all points x of a stationary point process X in $\mathbb{R}$. In order to apply the theory of Newton-Cotes estimators we assume throughout the following that Assumption 1 is satisfied for a given $n \in \mathbb{N}$. Then, the Newton-Cotes estimator of order n is unbiased for $\int_{\mathbb{R}} f(x)\mathrm{d}x$, which is equal to $\mathscr{H}^d(Y)$ by Fubini's theorem. In the special case where the points in X are equidistant, Newton-Cotes estimators of any order coincide with the classical Cavalieri volume-estimator.

We recall a number of fundamental notions from convex geometry; see e.g. (Schneider, 2014). A set $Y \subset \mathbb{R}^d$ is called a *convex body*, if it is non-empty, compact and convex. We say that a hyperplane H in $\mathbb{R}^d$ *supports* the convex body Y, if $H \cap Y \neq \emptyset$ and one of the two open half-spaces generated by H does not contain any points of Y. If $H = x\omega + \omega^\perp$ supports Y and ω points in the direction of the open half-space that is disjoint from Y, the unit vector ω is called an *outer normal vector* of Y at the *support set* $H \cap Y$. We say that Y is *strictly convex* if its boundary does not contain any non-degenerate line-segments. A convex body is *of class C_+^2* if its boundary is a regular submanifold of $\mathbb{R}^d$ that is twice continuously differentiable (in the sense of differential geometry) and all principal curvatures are positive at all boundary points. In particular, if Y is of class C_+^2, it is strictly convex.

If the convex body $Y \subset \mathbb{R}^d$ and a unit vector $\omega \in S^{d-1}$ are given, the support of the measurement function (28) is a compact interval $[x_-, x_+]$. The hyperplanes $H_- = x_-\omega + \omega^\perp$ and $H_+ = x_+\omega + \omega^\perp$ support Y at support sets Y_- and Y_+ with outer unit normals $-\omega$ and ω, respectively. The number $d(\omega) = x_+ - x_-$ is called the *width* of Y in direction ω. The function $\omega \mapsto d(\omega)$ is continuous on S^{d-1}. If the width $d(\omega)$ does not depend on ω, the set Y is called a *body of constant width*. Clearly, any ball is of constant width with $d = d(\omega)$ being its diameter, but there are other convex bodies of constant width in $\mathbb{R}^d$. However, among all *point symmetric bodies Y* (meaning that there is a point $z \in \mathbb{R}^d$ with $Y - z = \{-x : x \in Y - z\}$) balls can be characterized by having constant width. These and further results on bodies of constant width can be found in (Gardner, 2006, 3.2); see in particular Theorem 3.2.7 of this monograph.

The asymptotic variance behavior depends on the smoothness of f, which in turn reflects properties of the set Y. To illustrate how basic geometric regularity of a set Y yields smoothness properties of f, we restrict considerations to convex objects in three-dimensional space ($d = 3$), but note that generalizations to sets of higher dimension and with a boundary being a smooth manifold are also possible. As a consequence of the Brunn-Minkowski inequality (Gardner, 2002, p. 361), the measurement function $f^{1/2}$ for a given $Y \subset \mathbb{R}^3$ is concave on its support $[x_-, x_+]$. If the support set Y_+ (Y_-) is a singleton or a line-segment, the function f is continuous in x_+ (x_-). In particular, f is continuous when Y is strictly convex.

If Y is of class C_+^2 the principal radii $r_1(\omega), r_2(\omega)$ of curvature exist at the support point of Y with outer normal vector ω. The second normalized symmetric function $s_2(\omega) = r_1(\omega)r_2(\omega)$ of the principal radii is

continuous and has integral

$$\int_{S^2} s_2(\omega)\mathcal{H}^2(\mathrm{d}\omega) = \mathcal{H}^2(\partial Y), \qquad (29)$$

where the right side is the surface area of the boundary ∂Y of Y; see (Schneider, 2014, (4.26) and (4.32)). For later use, we also remark that the function s_2 determines the convex set Y of class C_+^2 uniquely up to translations; see (Schneider, 2014, (4.26) and Theorem 8.1.1).

Proposition 11. *Let $Y \subset \mathbb{R}^3$ be a convex body of class C_+^2. Then the measurement function f in (28) is twice continuously differentiable on $\mathbb{R} \setminus \{x_-, x_+\}$, and f' jumps exactly at the endpoints $x_- < x_+$ of the support of f. The jump at x_+ is $2\pi\sqrt{s_2(\omega)}$ and the jump at x_- is $2\pi\sqrt{s_2(-\omega)}$.*

If f'' has a right sided limit at x_- and a left sided limit at x_+ then f is weakly $(1,1)$-piecewise smooth.

Proof. We first set out to prove that f is twice continuously differentiable on $\mathbb{R} \setminus \{x_-, x_+\}$, and can restrict attention to the interior of its support. Translating Y appropriately, it is enough to show that f is twice continuously differentiable in a neighborhood of $x = 0$ when the origin is an interior point of Y. Assuming $0 \in \mathrm{int}\, Y$, we show first that the radial function $\rho_Y(u) = \max\{t \geq 0 : tu \in Y\}, u \in S^2$, is a twice continuously differentiable function on the sphere S^2. It is enough to show this claim in the neighborhood of one unit vector, which we may assume to coincide with the last standard basis vector e_3. As Y is of class C_+^2 and 0 is an interior point of Y, there is an open ball U in $e_3^\perp$ centered at 0 (and with a radius strictly smaller than 1) and a local C^2-parametrization $h : U \to \mathbb{R}$ such that $(x, h(x))$, $x \in U$, parametrizes a patch of the boundary of Y close to $(0, h(0))$. Any point u in a neighborhood of e_3 in S^2 can be written as $u = (x, \sqrt{1 - \|x\|^2}), x \in U$, and thus tu is a boundary point of Y if and only if $t\sqrt{1 - \|x\|^2} = h(tx)$. The implicit definition of $\rho_Y(u) = t$ through

$$F(x, t) = h(tx) - t\sqrt{1 - \|x\|^2} = 0$$

shows that ρ_Y is C^2 in a neighborhood of e_3 in S^2 by the implicit function theorem. As the origin is an interior point of Y, this implies that the Minkowski functional ρ_Y^{-1} of Y is in $C^2(S^2)$ and Y is therefore 2-smooth in the sense of Koldobsky (2005). Like in (Koldobsky, 2005, Lemma 2.4) one now shows that f is twice continuously differentiable in a neighborhood of $x = 0$, noting that the origin-symmetry required in the statement of the lemma is not needed for the proof.

We have already seen that strict convexity implies continuity of f. We now show that the first derivative of f has finite jumps at the endpoints x_- and x_+ of its support. Without loss of generality, we may now assume $\omega = e_3$, that Y is rescaled and that the origin $0 \in Y$ is chosen such that e_3 is the boundary point of Y where the support plane at position $x_+ = 1$ meets Y. As we have already seen, there is a local C^2-parametrization $h : U \to \mathbb{R}$ with U as above, such that $(x, h(x))$, $x \in U$, parametrizes a patch of the boundary of Y close to $(0, h(0)) = e_3$. A first order Taylor expansion with a second order remainder term, using the fact that the gradient of h must be zero at $x = 0$, shows

$$h(x) = 1 - \tfrac{1}{2}x^\top A(\xi_x)x \qquad (30)$$

in U, where $A(\cdot)$ is the Hessian matrix of $-h$ and ξ_x is a point on the line segment with endpoints 0 and x.

The eigenvalues λ_1 and λ_2 of $A(0)$ are the principal curvatures of Y at e_3 and thus coincide with $1/r_1(e_3)$ and $1/r_2(e_3)$ up to permutation (Schneider, 2014, Section 2.5) and are positive by assumption. As $A(\cdot)$ is continuous, there is a compact neighborhood V of 0 in U such that $x^\top A(\xi_x)x \geq \frac{\min\{\lambda_1, \lambda_2\}}{2}\|x\|^2$ for all $x \in V$. This and the convexity of Y imply the existence of a constant ε_0 such that $M_\varepsilon = \{x \in U : h(x) \geq 1 - \varepsilon\} \subset V$ for all $0 < \varepsilon < \varepsilon_0$. We claim that for such ε, the orthogonal projection of $Y_\varepsilon = Y \cap ((1 - \varepsilon)e_3 + e_3^\perp)$ onto $e_3^\perp$ coincides with M_ε, so

$$f(1 - \varepsilon) = \mathcal{H}^2(M_\varepsilon). \qquad (31)$$

In fact, if this was not the case, there would be a point $(x', t) \in Y_\varepsilon$ with $x' \notin U$. The convexity of Y implies that all points $x \in U$ of the line segment s between x' and 0 are in M_ε. But $s \cap (U \setminus V) \neq \emptyset$, so M_ε contains points outside V, a contradiction.

In view of (31) and (30) we have

$$\begin{aligned}
\frac{1}{\varepsilon}f(1 - \varepsilon) &= \mathcal{H}^2\left(\frac{1}{\sqrt{\varepsilon}}M_\varepsilon\right) \\
&= \mathcal{H}^2\left(\{x \in \tfrac{1}{\sqrt{\varepsilon}}U : x^\top A(\xi_{\sqrt{\varepsilon}x})x \leq 2\}\right) \\
&\to \mathcal{H}^2\left(\{x \in e_3^\perp : x^\top A(0)x \leq 2\}\right),
\end{aligned}$$

as $\varepsilon \downarrow 0$, where the continuity of $A(\cdot)$ was used again. The limit set in the last displayed formula is an ellipse in $e_3^\perp$ with half axes $\sqrt{2/\lambda_i}$, $i = 1, 2$ and area $2\pi/\sqrt{\lambda_1\lambda_2}$. This implies

$$f'(1-) = -\frac{2\pi}{\sqrt{\lambda_1\lambda_2}} = -2\pi\sqrt{s_2(e_3)},$$

which is the negative of the jump of f' at $x_+ = 1$, as $f'(x) = 0$ for $x > x_+$. Replacing ω by $-\omega$, the above

arguments show a corresponding equality for the jump at x_-.

As f'' exists and is continuous on $\mathbb{R} \setminus \{x_-, x_+\}$, this function can at most have two discontinuities. Under the assumption in the last claim of the theorem, these discontinuities are points where f'' has jumps, so f is weakly $(1,1)$-piecewise smooth. This concludes the proof. $\square$

As mentioned in the introduction, there are practically relevant examples of measurement functions with unbounded second derivative. In particular, there are convex bodies of class C_+^2 with a measurement function f which is not $(1,1)$-piecewise smooth but only weakly $(1,1)$-piecewise smooth. An example can be constructed by modifying the measurement function $\pi(1-(1-x)^2) = \pi(x(2-x))$, $0 \leq x \leq 2$, of the unit ball centered at the point ω. We consider the measurement function

$$f(x) = \pi(x + x^{3/2})(2-x) = \pi(2x + 2x^{3/2} - x^2 - x^{5/2}),$$

$0 \leq x \leq 2$ (and zero elsewhere), of a body of revolution Y with axis ω. Because $f''(0^+) = \infty$, the function f is not $(1,1)$-piecewise smooth, but only weakly $(1,1)$-piecewise smooth. For the corresponding surface of revolution, the principal curvatures at a boundary point with level x are known to be

$$\kappa_1(x) = \frac{1}{\rho(x)\sqrt{1 + \rho'(x)^2}}$$
$$= \frac{2}{\sqrt{(2\rho)^2(x) + (2 + 3x^{1/2} - 2x - \frac{5}{2}x^{3/2})^2}},$$

and

$$\kappa_2(x) = -\frac{\rho''(x)}{\sqrt{1 + \rho'(x)^2}^3} = -\rho^3(x)\rho''(x)\kappa_1^3(x),$$

where $\rho(x) = \sqrt{f(x)/\pi}$ is the radius of the section disk of Y at level x. This can be used to show that Y is a convex body of class C_+^2.

In the last statement of the proposition, we assumed that all discontinuity points of f'' are jump points. It is an open problem if this condition can be replaced by a more geometric assumption on Y. But it is well-known that the measurement function of any ellipsoid in $\mathbb{R}^3$ has a second derivative which satisfies this property for any direction ω. In addition, f'' is uniformly bounded in ω in this case; see, for instance, (Kiêu, 1997, Appendix B.1).

Corollary 12. *For* $\omega \in S^2$ *assume that* $Y \subset \mathbb{R}^3$ *is a convex body of class* C_+^2 *with measurement function* f

supported by $[x_-, x_+]$, *and such that* f'' *has one-sided limits at* x_- *and* x_+. *Then the covariogram* g *of the measurement function* (28) *is weakly* $(3,1)$-*piecewise smooth and* $g^{(3)}$ *has three jumps. These jumps have positions* $-d(\omega)$, 0 *and* $d(\omega) = x_+ - x_-$ *and their heights are*

$$J_{g^{(3)}}(0) = 4\pi^2(s_2(\omega) + s_2(-\omega)), \quad \text{and}$$
$$J_{g^{(3)}}(\pm d(\omega)) = 4\pi^2\sqrt{s_2(\omega)s_2(-\omega)}.$$

Proof. That g is weakly $(3,1)$-piecewise smooth follows from Corollary 17 of the appendix. The relation

$$J_{g^{(3)}}(c) = \sum_{b-a=c} J_{f'}(a)J_{f'}(b), \quad c \in \mathbb{R},$$

which is a special case of (40), in combination with Proposition 11 yields the remaining claims. $\square$

Note that $2g^{(3)}(0^+) = J_{g^{(3)}}(0)$ as $g^{(3)}$ is odd. Hence,

$$g^{(3)}(0^+) = 2\pi^2(s_2(\omega) + s_2(-\omega)).$$

When Y is point symmetric, we have $s_2(\omega) = s_2(-\omega)$ for all $\omega \in S^2$ and thus

$$J_{g^{(3)}}(0) = 2J_{g^{(3)}}(\pm d(\omega)) = 8\pi^2 s_2(\omega).$$

Theorem 13. *Let* $n \in \mathbb{N}$ *be given. Assume that* X_u *is a unit-intensity stationary point process satisfying Assumption 1, and that* $Y \subset \mathbb{R}^3$ *is a convex body of class* C_+^2 *whose measurement function supported by* $[x_-, x_+]$ *in direction* $\omega \in S^2$ *has a second derivative with one-sided limits at* x_- *and* x_+. *If* $V_n(Y) = V_{n,\omega}(Y)$ *is the nth Newton-Cotes estimator based on intersections of* Y *with the hyperplanes* $\{x\omega + \omega^\perp : x \in tX_u\}$ *with* $t > 0$, *then*

$$\mathbb{V}\mathrm{ar}(V_{n,\omega}(Y)) = \mathbb{V}\mathrm{ar}_E(V_{n,\omega}(Y)) + Z_\omega(t) + r_\omega(t), \quad (32)$$

where the extension term is given by

$$\mathbb{V}\mathrm{ar}_E(V_{n,\omega}(Y)) = 4\pi^2(s_2(\omega) + s_2(-\omega))H_1(0; X_u)t^4,$$

and the Zitterbewegung

$$Z_\omega(t) = 8\pi^2\sqrt{s_2(\omega)s_2(-\omega)}H_1\left(\frac{d(\omega)}{t}; X_u\right)t^4$$

satisfies

$$Z_\omega(t) \leq \mathbb{V}\mathrm{ar}_E(V_{n,\omega}(Y)). \quad (33)$$

The remainder $r_\omega(t)$ *is of order* $O(t^4)$. *If* X_u *is weakly 1-admissible,* $r_\omega(t)$ *is of order* $o(t^4)$.

We remark that (33) also holds for more general objects Y, as long as its measurement function f is weakly $(1,1)$-piecewise smooth and f' has exactly two jumps.

Proof. The explicit forms of $\mathbb{V}\mathrm{ar}_E(V_{n,\omega}(Y))$ and $Z_\omega(t)$ are obtained by inserting the jumps of the third derivative $g^{(3)}$ of the covariogram given in Corollary 12 into the corresponding equations of Lemma 2, taking (6) into account. The asymptotic behavior of $r_\omega(t)$ has also been established in Lemma 2. The bound (33) follows from $H_1(\cdot;X_u) \leq H_1(0;X_u)$ and the inequality of arithmetic and geometric means. $\qquad\square$

The explicit expression for the extension term in Theorem 13 also shows which directions $\omega \in S^2$ are best possible in terms of asymptotic variance when the Zitterbewegung is neglected. If possible, one should choose ω in the set of all unit vectors for which the average $(s_2 + \check{s}_2)/2$ attains its minimum, where $\check{s}_2$ denotes the reflection of s_2. When Y is point symmetric, this is the set where the second normalized symmetric curvature function s_2 is minimal. Of course, s_2 is not available in applications, but with this in mind, one might want to choose ω such that the corresponding support point has large curvature (that is, the vicinity is 'peaked').

A common strategy to determine the orientation of the hyperplane stack in applications is to randomize the direction ω in an isotropic way. We will write $\mathbb{E}_\omega$ for the expectation with respect to ω in this case. The assumptions of the following corollary are for instance satisfied if Y is an ellipsoid; see the comment right before Corollary 12.

Corollary 14. *Let the assumptions of Theorem 13 be satisfied for all $\omega \in S^2$. Assume in addition that the second derivative of the measurement function is uniformly integrable in $\omega \in S^2$. If ω is a uniform random unit vector which is independent of X_u, then (8) with $m = 1$ holds with*

$$\mathbb{V}\mathrm{ar}_E(V_n(Y)) = 2\pi\mathscr{H}^2(\partial Y)H_1(0;X_u)t^4. \qquad (34)$$

The term corresponding to the Zitterbewegung is

$$\overline{Z}(t) = 8\pi^2 \mathbb{E}_\omega\left[\sqrt{s_2(\omega)s_2(-\omega)}\,H_1\!\left(\tfrac{d(\omega)}{t};X_u\right)\right]t^4,$$

and satisfies

$$\overline{Z}(t) \leq \mathbb{V}\mathrm{ar}_E(V_n(Y)) \qquad (35)$$

with equality if and only if Y is a ball with diameter d_0, say, and $H_1(d_0/t;X_u) = H_1(0;X_u)$. The remainder is of order $O(t^4)$, and even of order $o(t^4)$ when X_u is weakly 1-admissible.

Proof. Due to the law of total variance and the fact that $V_{n,\omega}(f)$ is unbiased for all ω, we have $\mathbb{V}\mathrm{ar}(V_n(Y)) = \mathbb{E}_\omega \mathbb{V}\mathrm{ar}[V_{n,\omega}(Y)|\omega]$, where the latter variance has been described in Theorem 13. Hence, taking expectations in (32) shows

$$\mathbb{V}\mathrm{ar}(V_n(Y)) = \mathbb{E}_\omega \mathbb{V}\mathrm{ar}_E(V_{n,\omega}(Y)) + \mathbb{E}_\omega Z_\omega(t) + \mathbb{E}_\omega r_\omega(t).$$

The first term on the right equals (34) due to (29) and the second term is obviously $\overline{Z}(t)$.

Now let g_ω denote the covariogram of the measurement function $f = f_\omega$ with respect to the direction $\omega \in S^2$. By (Stehr and Kiderlen, 2020, Proposition 6.1)

$$r_\omega(t) = t^4 \int_\mathbb{R} g_\omega^{(4)}(s)H_1(\tfrac{s}{t};X_u)\mathrm{d}s,$$

and, for all ω, it is of order $o(t^4)$ if X_u is weakly 1-admissible. As $H_1(\cdot;X_u)$ is uniformly bounded and independent of ω, the asserted lower order properties of $r(t) = \mathbb{E}_\omega r_\omega(t)$ follow by dominated convergence if

$$\mathbb{E}_\omega \int_\mathbb{R} |g_\omega^{(4)}(s)|\mathrm{d}s < \infty. \qquad (36)$$

The function $g_\omega^{(4)}$ is uniformly integrable in $\omega \in S^2$ by (38) and the assumption that f_ω'' is uniformly integrable in ω. Thus (36) follows.

We now fix $t > 0$. Inequality (35) clearly holds due to (33) and (29). If the former holds with equality for a given t, we must have

$$H_1\!\left(\tfrac{d(\cdot)}{t};X_u\right) = H_1(0;X_u) \qquad (37)$$

almost surely, and $s_2 = \check{s}_2$, almost surely. The function s_2 is continuous by assumption, so this implies $s_2 = \check{s}_2$ on S^2, which shows that Y must be point symmetric, as Y is determined by s_2 up to translation. As the width $d(\cdot)$ is continuous and positive on S^2, its range is a compact interval in $(0,\infty)$, which, by the almost sure equality (37) and Lemma 1, implies that the range is degenerate and hence Y has constant width. This means that $d(\omega)$ does not depend on ω, so there is a constant $d_0 > 0$ such that $d = d_0$ on S^2. But the only origin symmetric convex bodies with constant width d_0 are balls with diameter d_0. This and (37), which now reads $H_1(d_0/t;X_u) = H_1(0;X_u)$, shows one of the implications of the characterization of equality in (35). The other is trivially satisfied. $\qquad\square$

Under the assumptions of Corollary 14 one can also show $\overline{Z}(t) \geq -\mathbb{V}\mathrm{ar}_E(V_n(Y))$. However, this lower bound is not sharp even in the equidistant case.

APPENDIX

In this appendix we list generalizations of important results from (Kiêu, 1997) such that they now apply not only to functions with finitely many finite jumps, but also to integrable functions with finitely many, possibly infinite jumps. With these generalizations in mind, it is easily seen that all the results of (Stehr and Kiderlen, 2020) involving $(m, 1)$-piecewise smooth functions also apply for *weakly* $(m, 1)$-piecewise smooth functions. One simply has to note that integrability of $f^{(m+1)}$ and $g^{(2m+2)}$ actually suffice where their boundedness was used in previous papers (here f denotes the measurement function and g its associated covariogram).

To state the results below, we follow the notation in (Kiêu, 1997), and let C_K be the set of all compactly supported piecewise continuous functions with finitely many finite jumps, and we let C_b be the set of all piecewise continuous bounded functions with locally finitely many and finite jumps. Furthermore, we define $\overline{C}_K$ to be the set of compactly supported piecewise continuous and integrable functions with finitely many (possibly infinite) jumps. In particular, if $h \in C_K$ such that h' is continuous in all but finitely many points with possibly infinite jumps, then h' is integrable by the first fundamental theorem of calculus, and thus $h' \in \overline{C}_K$.

The following generalization of (Kiêu, 1997, Lemma 4.1) follows simply by noticing that the integrability of h' is enough to guarantee the integrability of $h'\phi$ in its proof.

Lemma 15 (Generalization of (Kiêu, 1997, Lemma 4.1)). *Let $h \in C_K$ and $\phi \in C_b$ such that $h' \in \overline{C}_K$ and $\phi' \in C_b$. Then*

$$\int_{\mathbb{R}} h(x)\phi'(x)\mathrm{d}x + \int_{\mathbb{R}} h'(x)\phi(x)\mathrm{d}x = -\sum_{a \in D_{h\phi}} J_{h\phi}(a).$$

In the proposition below, $\check{f}$ means the reflection of f, i.e. $\check{f}(x) = f(-x)$, and $*$ denotes convolution. Similarly, $\check{f}^{(k)}$ is the kth derivative of $\check{f}$, and $J_{\check{f}^{(k)}}$ is the reflection of the jump-function associated to $f^{(k)}$.

Proposition 16 (Generalization of (Kiêu, 1997, Proposition 5.7)). *Let f be a function with all derivatives up to order m in C_K and derivative of order $m+1$ in $\overline{C}_K$. Then all derivatives up to order $2m+1$ of its covariogram g are in C_K, and the derivative of order $2m+2$ of g is in $\overline{C}_K$ with*

$$g^{(2k)} = f^{(k)} * \check{f}^{(k)} + \sum_{0 \le \ell < k} J_{f^{(\ell)}} * \check{f}^{(2k-\ell-1)} + \sum_{0 \le \ell < k} (-1)^{\ell+1} J_{\check{f}^{(\ell)}} * f^{(2k-\ell-1)}, \tag{38}$$

$$g^{(2k+1)} = f^{(k)} * \check{f}^{(k+1)} + \sum_{0 \le \ell < k} J_{f^{(\ell)}} * \check{f}^{(2k-\ell)} + \sum_{0 \le \ell \le k} (-1)^{\ell+1} J_{\check{f}^{(\ell)}} * f^{(2k-\ell)} \tag{39}$$

for all $2k$ and $2k+1$ less than or equal to $2m+2$.

Proof. The claim for $2k$ and $2k + 1$ less than or equal to $2m$ is exactly (Kiêu, 1997, Proposition 5.7). The expression (39) for $k = m$ and the claim that $g^{(2m+1)} \in C_K$ follow by differentiation from (Kiêu, 1997, Proposition 5.6) realizing that this proposition holds true even under the weaker assumption $f_1, f_2 \in C_K$ and $f_2' \in \overline{C}_K$, where the notation of this proposition has been adopted. One simply has to realize that $f_1 * f_2'$ is continuous also for $f_2' \in \overline{C}_K$.

The expression (38) for $k = m + 1$ follows by differentiation using another generalization of (Kiêu, 1997, Proposition 5.6): Let $f_1 \in \overline{C}_K$ and $f_2 \in C_K$ such that $f_2' \in \overline{C}_K$. Then $f_1 * f_2 \in C_K$ is continuous, $f_1 * f_2' \in \overline{C}_K$ and $(f_1 * f_2)' \in \overline{C}_K$ coincides with $f_1 * f_2' + J_{f_2} * f_1$ outside $D_{(f_1*f_2)'}$. Moreover, $J_{J_{f_2}*f_1} = J_{f_2} * J_{f_1}$. The proof follows along the same lines as that of (Kiêu, 1997, Proposition 5.6) using Lemma 15 and the appropriate generalizations of (Kiêu, 1997, Lemmas 5.4 and 5.5). $\square$

The following corollary generalizes the claims of (Kiêu, 1997, Corollary 5.8) relevant for this paper and for (Stehr and Kiderlen, 2020).

Corollary 17 (Generalization of (Kiêu, 1997, Corollary 5.8)). *If f is weakly $(m, 1)$-piecewise smooth then its covariogram g is weakly $(2m+1, 1)$-piecewise smooth with*

$$J_{g^{(2m+1)}} = (-1)^{m+1} J_{\check{f}^{(m)}} * J_{f^{(m)}}. \tag{40}$$

Proof. The fact that $g^{(k)}$ is continuous for $k \le 2m$ is shown in the original corollary. The fact that g is then weakly $(2m+1, 1)$-piecewise smooth with the given jumps of $g^{(2m+1)}$ follows easily from Proposition 16 and its proof. $\square$

Acknowledgments

REFERENCES

Baddeley AJ, Jensen EBV (2004). Stereology for Statisticians. Chapman & Hall/CRC Monographs on Statistics & Applied Probability, CRC Press, Boca Raton.

Baddeley AJ, Dorph-Petersen K-A, Jensen EBV (2006). A note on the stereological implications of irregular spacing of sections. J. Microsc 222:177–81.

Cruz-Orive LM (1989). On the precision of systematic sampling: a review of Matheron's transitive methods. J. Microsc 153:315–33.

García-Fiñana M, Cruz-Orive LM (2004). Improved variance prediction for systematic sampling on $\mathbb{R}$. Statistics 38:243–72.

Gardner RJ (2002). The Brunn–Minkowski inequality. Bull. Amer. Math. Soc. 39: 355–405.

Gardner RJ (2006). Geometric Tomography. (2nd edn.) Cambridge Univ. Press, Cambridge.

Kiderlen M, Dorph-Petersen K-A (2017). The Cavalieri estimator with unequal section spacing revisited. Image Anal. Stereol. 36:135–41.

Kiêu K (1997). Three Lectures on Systematic Geometric Sampling. Memoirs 13, Department of Theoretical Statistics, University of Aarhus, Aarhus.

Koldobsky A (2005). Fourier Analysis in Convex Geometry. Mathematical Surveys and Monographs, Am. Math. Soc., Providence RI.

Matheron G (1965). Les Variables régionalisées et leur Estimation – Une application de la théorie des Fonctions aléatoires aux Sciences de la Nature. Masson et Cie, Paris.

Schneider R (2014). Convex Bodies: The Brunn–Minkowski Theory. (2nd edn.) Cambridge Univ. Press, Cambridge.

Schneider R, Weil W (2008). Stochastic and Integral Geometry. Springer, Heidelberg.

Stehr M, Kiderlen M (2020). Asymptotic variance of Newton-Cotes quadratures based on randomized sampling points. *To appear in* Adv. Appl. Probab.

Ziegel J, Baddeley AJ, Dorph-Petersen K-A, Jensen EBV (2010). Systematic sampling with errors in sample locations. Biometrika 97:1–13.

Ziegel J, Jensen EBV, Dorph-Petersen K-A (2011). Variance estimation for generalized Cavalieri estimators. Biometrika 98:187–98.