THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

Variance estimation for generalized Cavalieri estimators

Johanna Ziegel and Eva B. Vedel Jensen

Research Report No. 10 I September 2009

Variance estimation for generalized Cavalieri estimators

Johanna Ziegel[1] and Eva B. Vedel Jensen[2]
[1]ETH Zurich, Switzerland,
[2]University of Aarhus, Denmark

Abstract

The precision of stereological estimators based on systematic sampling is a question of great practical importance. This paper presents methods of data-based variance estimation for generalized Cavalieri estimators where errors in sampling positions may occur. Variance estimators are derived under perturbed systematic sampling, systematic sampling with cumulative errors and systematic sampling with random dropouts.

Keywords: Asymptotic variance; Cavalieri estimator; Cumulative error; Moment measure; Perturbed systematic sampling; Point process; Spatial statistics; Stereology.

1 Introduction

The target of Cavalieri estimators are parameters Θ that can be expressed as an integral

$$\Theta = \int_{\mathbb{R}} f(x)dx,$$

where f is the so-called measurement function, an integrable function with bounded support. The generalized Cavalieri estimators are based on measurements at points constituting a second order stationary point process on the real line $\Phi = \{y_k\}_{k\in\mathbb{Z}}$ and take the following form

$$\widehat{\Theta} = \frac{1}{\mu}\sum_{k\in\mathbb{Z}} f(y_k),$$

where $\mu > 0$ is the intensity of Φ. This class of estimators was first defined in Baddeley *et al.* (2006) and studied further in Ziegel *et al.* (2009).

If $y_k = t(U + k)$, where U is uniform random in $[0, 1]$, then Φ is a systematic sample with spacing $t = 1/\mu > 0$. The estimator $\widehat{\Theta}$ is then the classical Cavalieri estimator. The asymptotic expansion of its variance as $t \to 0$ can be decomposed into a sum of an extension term, a Zitterbewegung or fluctuation term and higher order terms, cf. Baddeley and Jensen (2005, Chapter 13) and references therein. The extension term represents the overall trend of the variance, while the Zitterbewegung term oscillates around zero. It is common practice in stereology to estimate the variance by estimating the extension term which depends on the behaviour of the covariogram

$$g(y) = \int_{\mathbb{R}} f(x)f(x+y)dx, \quad y \in \mathbb{R},$$

near the origin.

The generalized Cavalieri estimators allow for errors in the placement of systematic sampling points. As shown in Ziegel *et al.* (2009), such errors can lead to substantial inflation of the variance compared to the variance of the classical Cavalieri estimator. In this paper, we derive asymptotic expansions of the variance of generalized Cavalieri estimators for which the leading term can easily be estimated from data. We address several different sampling procedures used in practice such as perturbed systematic sampling, systematic sampling with cumulative error and systematic sampling with random dropouts.

2 Perturbed systematic sampling

Suppose Φ follows the model of perturbed systematic sampling such that the intended equally spaced sampling points $x_k = t(U+k)$, where U is uniform random in $[0,1]$, are perturbed by independent and identically distributed random errors $(D_k)_{k\in\mathbb{Z}}$ with error density h_t, so that the actual locations are $y_k = x_k + D_k$. Under this model $\mu = 1/t$.

We consider the asymptotic variance as $t \to 0$ of the generalized Cavalieri estimator for an error density of the form $h_t(x) = (1/t)h_0(x/t)$, where h_0 is a probability density function with a finite number of jumps of finite size and compact support $\operatorname{supp}(h_0) \subseteq [-1/2, 1/2]$. For later use, we let

$$c_k = \int_{\mathbb{R}} |x|^k h_0 * \check{h}_0(x)dx,$$

where $\check{h}_0(x) = h_0(-x)$.

The covariogram estimator

$$\hat{g}(l,t) = t\sum_{k\in\mathbb{Z}} f(y_k)f(y_{k+l})$$

is unbiased for $g(lt)$ under systematic sampling without errors. Under perturbed systematic sampling, we still have that $G(0,t) = \mathbb{E}\{\hat{g}(0,t)\} = g(0)$, but for $l \neq 0$

$$G(l,t) = \mathbb{E}\{\hat{g}(l,t)\} = g * \check{h}_t * h_t(lt).$$

The asymptotic variance of $\widehat{\Theta}$ under perturbed systematic sampling depends on the smoothness of the measurement function. The function f is said to be $(m,1)$-piecewise smooth with $m \geq 0$ if $f^{(k)}$ is a measurable function with compact support and a finite number of jumps of finite size for all $k = 0,\ldots,m+1$, and $f^{(k)}$ is continuous for $k = 0,\ldots,m-1$.

The most commonly used measurement functions are $(m,1)$-piecewise smooth with $m = 0$ or $m = 1$. In Ziegel *et al.* (2009, Proposition 1), it is shown that if f

is $(m, 1)$-piecewise smooth, then the covariogram g is $(2m+1, 1)$-piecewise smooth. Furthermore, for $m = 0$ we have

$$\mathrm{Var}(\widehat{\Theta}) = -t^2 \left(c_2 + \frac{1}{6} \right) g'(0^+) + o(t^2), \text{ as } t \to 0, \tag{1}$$

while for $m = 1$

$$\mathrm{Var}(\widehat{\Theta}) = -t^3 \frac{c_2}{2} g''(0) + t^4 \left(\frac{1}{60} - \frac{c_2}{2} - \frac{c_4}{2} \right) \frac{1}{6} g^{(3)}(0^+) + o(t^4), \quad \text{as } t \to 0. \tag{2}$$

The main term of these expansions do not lend themselves easily to estimation based on data collected at perturbed sampling points. Propositions 2.1 and 2.2 below give equivalent asymptotic expansions of the variance for which the main term can be estimated directly if measurements of the errors in positioning are available. The proof of the two propositions can be found in the Appendix.

Proposition 2.1. *Let f be a $(0,1)$-piecewise smooth measurement function. Then,*

$$\mathrm{Var}(\widehat{\Theta}) = t \frac{1 + 6c_2}{12 - 6c_2} \sum_{i=0}^{2} \alpha_i G(i, t) + o(t^2), \quad \textit{as } t \to 0, \tag{3}$$

where $\alpha_0 = 3$, $\alpha_1 = -4 - c_2$, $\alpha_3 = 1 + c_2$.

Proposition 2.2. *Let f be a $(1,1)$-piecewise smooth measurement function. Then,*

$$\mathrm{Var}(\widehat{\Theta}) = t \frac{1}{3c_2^2 + c_2 + 4} \sum_{i=0}^{2} \alpha_i G(i, t) + o(t^4), \quad \textit{as } t \to 0, \tag{4}$$

where

$$\begin{aligned}
\alpha_0 &= \tfrac{1}{20} + \tfrac{11}{2} c_2 + 3c_2^2 - \tfrac{3}{2} c_4, \\
\alpha_1 &= -\tfrac{1}{15} - \tfrac{361}{60} c_2 - \tfrac{11}{2} c_2^2 + \tfrac{1}{2} c_2 c_4 + 2c_4, \\
\alpha_2 &= \tfrac{1}{60} + \tfrac{31}{60} c_2 + \tfrac{5}{2} c_2^2 - \tfrac{1}{2} c_2 c_4 - \tfrac{1}{2} c_4.
\end{aligned}$$

Remark. For exact systematic sampling, the main term of the asymptotic variance for $(0,1)$-piecewise smooth functions is usually approximated by

$$\tfrac{t}{12} \left\{ 3g(0) - 4g(t) + g(2t) \right\}, \tag{5}$$

see for example Baddeley and Jensen (2005, Paragraph 13.2.5). For $c_2 = 0$, the leading term in (3) is equal to (5). For $(1,1)$-piecewise smooth functions and $c_2 = c_4 = 0$, the main term of the asymptotic variance (4) simplifies to

$$\tfrac{t}{240} \left\{ 3g(0) - 4g(t) + g(2t) \right\}. \tag{6}$$

Substituting $\hat{g}$ for g in (6), we get the estimator which is given in Cruz-Orive (1999) and Gundersen *et al.* (1999) for exact systematic sampling.

For $(1,1)$-piecewise smooth measurement functions, the next proposition gives an alternative expression for the leading term of order t^3 in the asymptotic expansion (2). This result can be used to derive an upper bound for the leading term of order t^3, which can actually be estimated without use of measurements of the errors in positioning. The proof of the proposition may be found in the Appendix.

Proposition 2.3. *Let f be a $(1,1)$-piecewise smooth measurement function. Then,*

$$\mathrm{Var}(\widehat{\Theta}) = \tfrac{t}{3}\{3G(0,t) - 4G(1,t) + G(2,t)\} + t^3 g^{(3)}(0^+)\left(\tfrac{1}{2}c_2 - \tfrac{2}{3}c_1\right) + o(t^3),$$

as $t \to 0$.

Remark. The term $g^{(3)}(0^+)$ is always non-negative and $c_2 \leq c_1$. This can be seen as follows. Let X be a random variable with density $h_0 * \check{h}_0$. This density is supported in $[-1,1]$, hence $|X| \leq 1$. This implies $c_2 = \mathbb{E}(|X|^2) \leq E(|X|) = c_1$. Therefore we suggest to estimate an upper bound of the variance by

$$\tfrac{t}{3}\{3\hat{g}(0,t) - 4\hat{g}(1,t) + \hat{g}(2,t)\}.$$

Note that this is the same estimator up to a factor 1/80 that was given in Cruz-Orive (1999) and Gundersen *et al.* (1999) for the estimation of the variance of the classical Cavalieri estimator with a (1,1)-smooth measurement function. See (6) and Baddeley and Jensen (2005, Paragraph 13.2.5).

3 Systematic sampling with cumulative error

In this section we assume that Φ follows the model of systematic sampling with cumulative error. This means that the actual locations $\{y_k\}_{k\in\mathbb{Z}}$ of the sampling points are such that the increments $y_k - y_{k-1}$, $k \in \mathbb{Z}$, are i.i.d. with density $h_t : \mathbb{R}_+ \to \mathbb{R}_+$ with finite expectation $t > 0$. Furthermore the distribution for y_1 is chosen such that Φ is strictly stationary with finite intensity $\mu = 1/t$.

Under the further assumptions that g is continuous at 0 and bounded, and $h_t(x) = (1/t)h_0(x/t)$ for some absolutely continuous probability density h_0 on the positive half line with expected value 1, satisfying

$$\int_{\mathbb{R}} |x^l h_0'(x)| dx < \infty,$$

for some $l \geq 2$, Ziegel *et al.* (2009, Proposition 2) shows that

$$\mathrm{Var}(\widehat{\Theta}) = t g(0)\nu^2 + o(t), \quad \text{as } t \to 0,$$

where ν^2 is the variance of a random variable with density h_0.

The leading term of the expansion can be estimated by estimating $g(0)$ by the unbiased estimator $\hat{g}(0,t)$ and ν^2 from the actual locations of the sampling points. For section spacing $t > 0$ the parameter ν^2 can be estimated by the sample variance of the increments $(y_k - y_{k-1})_{k\in\{N_1,\dots,N_2\}}$ for some $N_1 < N_2$ divided by t^2.

Remark. In the case of perturbed systematic sampling, the estimators of $\mathrm{Var}(\widehat{\Theta})$ based on Propositions 2.1 and 2.2 behave robustly with respect to decreasing error, see the remark just after Proposition 2.2. In contrast to this, we have not been able to give an estimator for the variance in the case of systematic sampling with cumulative error that behaves robustly with respect to decreasing error, i.e. $h_0 \to \delta_1$.

4 Systematic sampling with random dropouts

In practice it is encountered that at some locations y_k the value of f cannot be determined. In this section, we study the variance of an estimator of Θ based on interpolation to approximate the missing values $f(y_k)$.

The sample locations are given by the process $\Phi = \{y_k\}_{k\in\mathbb{Z}}$ with intensity $\mu > 0$. We define

$$\tilde{\Theta} = \frac{1}{\mu}\sum_{k\in\mathbb{Z}} \omega_k f(y_k),$$

where ω_k are random weights. Given Φ, they are defined as

$$\begin{aligned}\omega_k|\Phi = \mathbb{1}\{U_k > p\}\bigl(1 + \tfrac{1}{2}\max\{l \geq 0 : U_{k+1} \leq p, \dots, U_{k+l} \leq p\}\\ + \tfrac{1}{2}\max\{l \geq 0 : U_{k-1} \leq p, \dots, U_{k-l} \leq p\}\bigr),\end{aligned}$$

where $(U_k)_{k\in\mathbb{Z}}$ is a sequence of independent and identically distributed uniform random variables on $[0,1]$, and $p > 0$ is the probability that the value of f cannot be determined at y_k. If $f(y_k)$ cannot be determined, then in $\tilde{\Theta}$ the value is replaced by the average of the nearest observation to the left and right of y_k.

It is a short calculation to check that $\mathbb{E}(\omega_k|\Phi) = 1$, hence

$$\mathbb{E}(\tilde{\Theta}) = \mathbb{E}\left\{\frac{1}{\mu}\sum_{k\in\mathbb{Z}} \mathbb{E}(\omega_k|\Phi) f(x_k)\right\} = \mathbb{E}(\widehat{\Theta}) = \Theta.$$

Furthermore, it is needed to obtain

$$\mathbb{E}(\tilde{\Theta}^2) = \frac{1}{\mu^2}\mathbb{E}\left\{\sum_{k,l\in\mathbb{Z}} \mathbb{E}(\omega_k\omega_l|\Phi) f(x_k) f(x_l)\right\}.$$

It is tedious, but not hard to check that $\mathbb{E}(\omega_k^2|\Phi) = 1 + 3p/(2-2p)$ and for $|k-l| = n \geq 1$

$$\mathbb{E}(\omega_k\omega_l|\Phi) = 1 + \frac{n(n-1)}{8}p^{n-1} - \frac{3+n^2}{4}p^n + \frac{n(n+1)}{8}p^{n+1}. \tag{7}$$

Therefore

$$\mathbb{E}(\tilde{\Theta}^2) = \frac{1}{\mu}\left(1 + \frac{3p}{2-2p}\right) G(0,\mu^{-1}) + \frac{2}{\mu}\sum_{n=1}^{\infty}(1+a_n)G(n,\mu^{-1}),$$

where a_n is given by the last three terms on the right hand side of (7), and $G(n,\mu^{-1}) = \mathbb{E}\{\hat{g}(n,\mu^{-1})\}$. For systematic sampling, perturbed systematic sampling and systematic sampling with cumulative error we have

$$\mathbb{E}(\widehat{\Theta}^2) = tG(0,t) + 2t\sum_{n=1}^{\infty} G(n,t),$$

where $t = \mu^{-1}$, see Moran (1950); Baddeley *et al.* (2006); Ziegel *et al.* (2009). Therefore we obtain for these models that

$$\operatorname{Var}(\tilde{\Theta}) = \operatorname{Var}(\widehat{\Theta}) + t\frac{3p}{2-2p}G(0,t) + 2t\sum_{n=1}^{\infty} a_n G(n,t). \tag{8}$$

Propositions 4.1, 4.2 and 4.3 below give the asymptotic variance of $\tilde{\Theta}$ under exact systematic sampling, perturbed systematic sampling and systematic sampling with cumulative errors, respectively. The proofs of the propositions can be found in the Appendix.

Proposition 4.1. *Let f be an $(m,1)$-piecewise smooth measurement function. Suppose that Φ follows the model of exact systematic sampling. If $m=0$, then*

$$\mathrm{Var}(\tilde{\Theta}) = -t^2 g'(0^+)\left\{\frac{1}{6} + \frac{p}{(1-p)^2}\right\} + Z(t) + o(t^2), \text{ as } t \to 0,$$

and if $m \geq 1$, then

$$\mathrm{Var}(\tilde{\Theta}) = t^4 g^{(3)}(0^+)\left\{\frac{1}{360} + \frac{3p(p^2+3p+1)}{2(1-p)^4}\right\} + Z(t) + o(t^4),$$

as $t \to 0$, where $Z(t)$ is the Zitterbewegung term.

Proposition 4.2. *Let f be an $(m,1)$-piecewise smooth measurement function. Suppose that Φ follows the model of perturbed systematic sampling with error density h_0 compactly supported in $[-1/2,1/2]$ with a finite number of jumps of finite size. If $m=0$, then*

$$\mathrm{Var}(\tilde{\Theta}) = -t^2 g'(0^+)\left\{c_2 + \frac{1}{6} + \frac{p}{(1-p)^2}\right\} + o(t^2), \text{ as } t \to 0,$$

and if $m \geq 1$

$$\mathrm{Var}(\tilde{\Theta}) = -t^3 c_2 g''(0)\left\{\frac{1}{2} + \frac{3p}{4(1-p)}\right\} + o(t^3), \text{ as } t \to 0.$$

Proposition 4.3. *Let the covariogram g be continuous at 0 and bounded. Suppose that Φ follows the model of systematic sampling with cumulative error with an increment density h_0 that fulfils the conditions of Ziegel* et al. *(2009, Proposition 2). Then* $\mathrm{Var}(\tilde{\Theta}) = tg(0)\nu^2 + o(t)$*, as $t \to 0$, where $\nu^2 < \infty$ is the variance of a random variable with density h_0.*

The Propositions 4.1, 4.2 and 4.3 can be used for variance estimation using the estimator

$$\tilde{g}(l,t) = t\sum_{k\in\mathbb{Z}} \omega_k\omega_{k+l} f(x_k) f(x_{k+l})$$

for the covariogram. We obtain in particular

$$\begin{aligned}
\mathbb{E}\{\tilde{g}(0,t)\} &= \left(1 + \frac{3p}{2-2p}\right) G(0,t) \\
\mathbb{E}\{\tilde{g}(1,t)\} &= \left(1 - p + \frac{1}{4}p^2\right) G(1,t) \\
\mathbb{E}\{\tilde{g}(2,t)\} &= \left(1 + \frac{1}{4}p - \frac{7}{4}p^2 + \frac{3}{4}p^3\right) G(2,t).
\end{aligned}$$

This shows that $\tilde{g}(l,t)$ can be scaled to yield an unbiased estimator of $G(l,t)$. We illustrate the estimation procedure in detail for the case when Φ follows the model of systematic sampling with a $(0,1)$-piecewise smooth measurement function f. It follows using Taylor expansion that

$$-t^2 g'(0^+) = \frac{t}{2}\{3G(0,t) - 4G(1,t) + G(2,t)\} + o(t^2), \text{ as } t \to 0.$$

This means that we can use Proposition 4.1 and the scaled versions of $\tilde{g}(0,t)$, $\tilde{g}(1,t)$ and $\tilde{g}(2,t)$ to estimate the leading term of the variance. The case of an $(m,1)$-piecewise smooth function with $m \geq 1$ works analogously. Similar results can also be obtained for perturbed systematic sampling following the proofs of Proposition 2.1 and 2.2. The case of systematic sampling with cumulative error is immediate.

Remark. In Ziegel *et al.* (2009) we did not use interpolation but instead used an estimator of the form $\tilde{\mu}^{-1}\sum_{k\in\mathbb{Z}} f(\tilde{y}_k)$ where $\{\tilde{y}_k\}_{k\in\mathbb{Z}}$ are the sample locations at which observations can be made and $\tilde{\mu} = (1-p)\mu$. The results in this section shows that interpolation is superior in the sense that the main term of the variance of the estimator, using interpolation, is of higher order under exact systematic sampling and perturbed systematic sampling. For systematic sampling with cumulative error the order of convergence remains the same, but the leading term is smaller using interpolation.

5 Exhaustive cut and storage of stacks

In practice the following sampling procedure is sometimes encountered. Sample locations are chosen according to a point process $\Phi = \{y_k\}_{k\in\mathbb{Z}}$. Then stacks of K consecutive sections are stored. For estimating Θ one chooses a uniform random section in each stack, determines $f(y_k)$ for this section and then calculates

$$\bar{\Theta} = \frac{K}{\mu}\sum_{y_k\in\Psi} f(y_k),$$

where Ψ is the set of all chosen sample locations. Let $(U_l)_{l\in\mathbb{Z}}, U^*$ be independent and identically distributed with $\mathbb{P}(U^* = n) = 1/K$, for $n \in \{0,\dots,K-1\}$. Define for all $k \in \mathbb{Z}$, $\tilde{U}_k = U_l$, if $K(l-1) \leq k < Kl$. The point process Ψ can be described as

$$\Psi = \{y_k \in \Phi \mid k \bmod K = \tilde{U}_{k-U^*}\}.$$

If Φ is second order stationary with intensity μ and second order reduced factorial density $\tilde{m}_{[2]}$, then Ψ is also second order stationary with intensity μ/K and second order reduced factorial density $\tilde{m}_{[2]}/K^2$. Therefore this sampling procedure has the same first and second order behaviour as systematic sampling with independent p-thinning with thinning probability $p = 1 - 1/K$.

If Φ follows the model of systematic sampling or perturbed systematic sampling we obtain by Ziegel *et al.* (2009, Proposition 3) that

$$\mathrm{Var}(\bar{\Theta}) = t(K-1)g(0) + o(t), \quad \text{as } t \to 0.$$

For systematic sampling with cumulative error we have by Ziegel *et al.* (2009, Proposition 4) that

$$\mathrm{Var}(\bar{\Theta}) = t\left(\nu^2 + K - 1\right) g(0) + o(t), \quad \text{as } t \to 0.$$

Defining $\bar{g}(0,t) = Kt\sum_{y_k \in \Psi} f(y_k)^2$, we obtain $\mathbb{E}\{\bar{g}(0,t)\} = g(0)$ in all three cases. As we will typically have $(K-1) \gg \nu^2$ we suggest to estimate the variance in all three cases by

$$t(K-1)\bar{g}(0,t).$$

Appendix

Proof of Proposition 2.1. The covariogram g of f is $(1,1)$-piecewise smooth. Therefore there exists $\varepsilon > 0$ such that g is continuously differentiable on $[0,\varepsilon]$ and twice differentiable on $(0,\varepsilon)$. Let t be so small that $[(k-1)t,(k+1)t] \in (-\varepsilon,\varepsilon)$ for $k = 1,2$. Using Taylor's theorem, $\mathrm{supp}(h_0) \subseteq [-1/2, 1/2]$, and the symmetry of g, we obtain

$$G(k,t) = g(0) + tkg'(0^+) + \frac{1}{2}\int g''(\xi_y)y^2\check{h}_t * h_t(kt-y)dy,$$

where $\xi_y \in [0,\varepsilon)$. Since $G(0,t) = g(0)$, we find

$$\begin{aligned}\sum_{i=0}^{2}\alpha_i G(i,t) = {} & t(-2+c_2)g'(0^+) \\ & + \frac{1}{2}\int g''(\xi_y)\{(-4-c_2)y^2\check{h}_t * h_t(t-y) + (1+c_2)y^2\check{h}_t * h_t(2t-y)\}dy.\end{aligned}$$

The last term of the above equation is of order $o(t)$, as $t \to 0$, which one can check using that g'' is bounded. Using (1), the result follows. □

Proof of Proposition 2.2. The covariogram g of f is $(3,1)$-piecewise smooth. Therefore there exists $\varepsilon > 0$ such that g is three times continuously differentiable on $[0,\varepsilon]$ and it is four times differentiable on $(0,\varepsilon)$. Let t be so small that $[(k-1)t,(k+1)t] \in (-\varepsilon,\varepsilon)$ for $k = 1,2$. Using Taylor's theorem, $\mathrm{supp}(h_0) \subseteq [-1/2,1/2]$, and the symmetry of g we obtain

$$\begin{aligned}G(k,t) = {} & g(0) + t^2\frac{1}{2}g''(0)(k^2+c_2) + t^3\frac{1}{6}g^{(3)}(0^+)(k^3+3kc_2) \\ & + \frac{1}{24}\int g^{(4)}(\xi_y)y^4\check{h}_t * h_t(kt-y)dy,\end{aligned}$$

where $\xi_y \in [0,\varepsilon)$. Since $G(0,t) = g(0)$, we find

$$\begin{aligned}\sum_{i=0}^{2}\alpha_i G(i,t) = {} & -t^2(3c_2^2+c_2+4)\frac{c_2}{2}g''(0) \\ & + t^3(3c_2^2+c_2+4)\left(\frac{1}{60} - \frac{c_2}{2} - \frac{c_4}{2}\right)\frac{1}{6}g^{(3)}(0^+) \\ & + \frac{1}{24}\int g^{(4)}(\xi_y)\{\alpha_1 y^4\check{h}_t * h_t(t-y) + \alpha_2 y^4\check{h}_t * h_t(2t-y)\}dy.\end{aligned}$$

The last term of the above equation is of order $o(t^3)$, as $t \to 0$, which one can check using that $g^{(4)}$ is bounded. Using (2), the result follows. □

Proof of Proposition 2.3. By Ziegel *et al.* (2009, Proposition 1), we have

$$\mathrm{Var}(\widehat{\Theta}) = t\left\{g(0) - g * \check{h}_t * h_t(0)\right\} + o(t^3), \quad \text{as } t \to 0, \tag{9}$$

and the term $t\{g(0) - g * \check{h}_t * h_t(0)\}$ is of order $o(t^2)$. The function $H_t := g * \check{h}_t * h_t$ is compactly supported, even, and three times continuously differentiable with

$$H_t^{(3)}(x) = g^{(3)} * \check{h}_t * h_t(x) + \sum_a s(a)\check{h}_t * h_t(x - a), \tag{10}$$

where $s(a) = \lim_{y \to a^+} g^{(3)}(y) - \lim_{y \to a^-} g^{(3)}(y)$. The function s is zero in all but finitely many points. A derivation of this formula can be found in a set of lecture notes by Kien Kiêu entitled Three Lectures on Systematic Geometric Sampling, which appeared as Memoirs at the Department of Theoretical Statistics at the University of Aarhus in 1997. Applying Taylor's theorem we obtain for $y \in [0, \infty)$

$$H_t(y) = H_t(0) + \tfrac{1}{2}H_t''(0)y^2 + R_2(y),$$

where $R_2(y) = (1/2)\int_0^y H_t^{(3)}(x)(y - x)^2 dx$ and therefore

$$4G(1, t) - G(2, t) = 4H_t(t) - H_t(2t) = 3H_t(0) + 4R_2(t) - R_2(2t). \tag{11}$$

Using (10) we obtain, as $t \to 0$,

$$\begin{aligned} 4R_2(t) - R_2(2t) = \sum_a s(a) \Bigg\{ & 2\int_{-a/t}^{1-a/t} \check{h}_0 * h_0(x)(t - a - tx)^2 dx \\ & - \frac{1}{2}\int_{-a/t}^{2-a/t} \check{h}_0 * h_0(x)(2t - a - tx)^2 dx \Bigg\} + o(t^2). \end{aligned} \tag{12}$$

For $a \neq 0$ and t sufficiently small, both integrals on the right hand side of the above equation are zero as h_0 is compactly supported. Therefore the right hand side of (12) simplifies to

$$\begin{aligned} 2g^{(3)}(0^+) \Bigg\{ & 2\int_0^1 \check{h}_0 * h_0(x)(t - tx)^2 dx \\ & - \frac{1}{2}\int_0^2 \check{h}_0 * h_0(x)(2t - tx)^2 dx \Bigg\} + o(t^2) \\ = {} & t^2 g^{(3)}(0^+)\left(\frac{3}{2}c_2 - 2c_1\right) + o(t^2). \end{aligned} \tag{13}$$

as $t \to 0$. Combining (11), (12) and (13) with (9) yields the claim. □

Proof of Proposition 4.1. The last two terms on the right hand side of (8) can be rewritten as

$$\frac{t}{4}(1 - p)\sum_{n=1}^{\infty} n(n + 1)p^n\{G(n + 1, t) - G(n, t)\} - \frac{3t}{2}\sum_{n=1}^{\infty} p^n\{G(n, t) - G(0, t)\}. \tag{14}$$

For exact systematic sampling we have $G(n,t) = g(nt)$. If the measurement function f is $(0,1)$-piecewise smooth we use Taylor expansion to obtain

$$g\{(n+1)t\} - g(nt) = tg'(0^+) + o(t), \quad \text{as } t \to 0.$$

Using the dominated convergence theorem, this implies

$$\begin{aligned}\sum_{n=1}^{\infty} n(n+1)p^n[g\{(n+1)t\} - g(nt)] &= tg'(0^+)\sum_{n=1}^{\infty} n(n+1)p^n + o(t)\\ &= tg'(0^+)\frac{2p}{(1-p)^3} + o(t),\end{aligned}$$

as $t \to 0$. Taylor expansion also yields

$$g(nt) - g(0) = g'(\xi)nt + o(t), \quad \text{as } t \to 0,$$

where $\xi \in (0, nt)$. By dominated convergence we therefore obtain

$$\sum_{n=1}^{\infty} p^n\{g(nt) - g(0)\} = tg'(0^+)\sum_{n=1}^{\infty} np^n + o(t) = tg'(0^+)\frac{p}{(1-p)^2} + o(t),$$

as $t \to 0$, and hence

$$\mathrm{Var}(\tilde{\Theta}) = \mathrm{Var}(\widehat{\Theta}) - t^2 g'(0^+)\frac{p}{(1-p)^2} + o(t^2), \quad \text{as } t \to 0. \tag{15}$$

If the measurement function f is $(m,1)$-piecewise smooth with $m \geq 1$, Taylor expansion yields

$$g\{(n+1)t\} - g(nt) = \tfrac{1}{2}t^2 g''(0)(2n+1) + \tfrac{1}{6}t^3 g^{(3)}(0^+)(3n^2+3n+1) + o(t^3),$$

as $t \to 0$, and

$$g(nt) - g(0) = \sum_{k=1}^{m} \frac{1}{(2k)!} g^{(2k)}(0) n^{2k} t^{2k} + \frac{1}{(2m+1)!} g^{(2m+1)}(\xi) n^{2m+1} t^{2m+1},$$

where $\xi \in (0, nt)$. Using the dominated convergence theorem, this implies

$$\begin{aligned}&\sum_{n=1}^{\infty} n(n+1)p^n[g\{(n+1)t\} - g(nt)]\\ &\quad= \frac{t^2}{2} g''(0)\sum_{n=1}^{\infty} n(n+1)(2n+1)p^n\\ &\qquad+ \frac{t^3}{6} g^{(3)}(0^+)\sum_{n=1}^{\infty} n(n+1)(3n^2+3n+1)p^n + o(t^3)\\ &\quad= t^2 g''(0)\frac{3p(1+p)}{(1-p)^4} + t^3 g^{(3)}(0^+)\frac{p(7p^2+22p+7)}{3(1-p)^5} + o(t^3),\end{aligned}$$

and

$$\sum_{n=1}^{\infty} p^n\{g(nt)-g(0)\} = \frac{t^2}{2}g''(0)\sum_{n=1}^{\infty} n^2p^n + \frac{t^3}{6}g^{(3)}(0^+)\sum_{n=1}^{\infty} n^3p^3 + o(t^3)$$
$$= t^2g''(0)\frac{p(1+p)}{2(1-p)^3} + t^3g^{(3)}(0^+)\frac{p(p^2+4p+1)}{6(1-p)^4} + o(t^3),$$

as $t \to 0$. This yields

$$\mathrm{Var}(\tilde{\Theta}) = \mathrm{Var}(\widehat{\Theta}) + t^4g^{(3)}(0^+)\frac{3p(p^2+3p+1)}{2(1-p)^4} + o(t^4), \quad \text{as } t \to 0. \tag{16}$$

Combining (15) and (16), respectively, with the asymptotic expansions for $\mathrm{Var}(\widehat{\Theta})$ as given in Baddeley and Jensen (2005, 13.2.2) yields the claim. □

Proof of Proposition 4.2. The proof works very similar to the proof of Proposition 4.1, so we only give the key steps. For perturbed systematic sampling we have $G(0,t) = g(0)$ and $G(n,t) = g * \check{h}_t * h_t(nt)$ for $n \neq 0$. Using Taylor expansion and dominated convergence, we obtain for $m = 0$

$$\sum_{n=1}^{\infty} n(n+1)p^n[g * \check{h}_t * h_t\{(n+1)t\} - g * \check{h}_t * h_t(nt)] = tg'(0^+)\frac{2p}{(1-p)^3} + o(t),$$

as $t \to 0$, and

$$\sum_{n=1}^{\infty} p^n\{g * \check{h}_t * h_t(nt) - g(0)\} = tg'(0^+)\frac{p}{(1-p)^2} + o(t), \quad \text{as } t \to 0.$$

For $m \geq 1$ we have

$$\sum_{n=1}^{\infty} n(n+1)p^n[g * \check{h}_t * h_t\{(n+1)t\} - g * \check{h}_t * h_t(nt)] = t^2g''(0)\frac{3p(1+p)}{(1-p)^4} + o(t^2),$$

as $t \to 0$, and

$$\sum_{n=1}^{\infty} p^n\{g * \check{h}_t * h_t(nt) - g(0)\} = t^2g''(0)\left\{\frac{p(1+p)}{2(1-p)^3} + c_2\frac{p}{2(1-p)}\right\} + o(t^2),$$

as $t \to 0$. Using (1), (2) and (14), we obtain the claim. □

Proof of Proposition 4.3. For systematic sampling with cumulative error we have $G(0,t) = g(0)$ and $G(n,t) = \int h_t^{n*}(x)g(x)dx$. Using the fact that the covariogram is continuous at 0 and bounded, we obtain that $G(n+1,t) - G(n,t) \to 0$ and $G(n,t) - G(0,t) \to 0$ as $t \to 0$ for all $n \geq 1$. Using (14) and Ziegel *et al.* (2009, Proposition 2) we therefore obtain $\mathrm{Var}(\tilde{\Theta}) = \mathrm{Var}(\widehat{\Theta}) + o(t) = tg(0)\nu^2 + o(t)$, as $t \to 0$. □

References

Baddeley, A. and Jensen, E. B. V. (2005). *Stereology for Statisticians*. Chapman & Hall/CRC, Boca Raton.

Baddeley, A., Dorph-Petersen, K. A., and Jensen, E. B. V. (2006). A note on the stereological implications of irregular spacing of sections. *J. Microsc.*, **222**, 177–181.

Cruz-Orive, L. M. (1999). Precision of Cavalieri sections and slices with local errors. *J. Microsc.*, **193**, 182–198.

Gundersen, H. J. G., Jensen, E. B. V., Kiêu, K., and Nielsen, J. (1999). The efficiency of systematic sampling - reconsidered. *J. Microsc.*, **193**, 199–211.

Moran, P. A. P. (1950). Numerical integration by systematic sampling. *Proc. Cambridge Philos. Soc.*, **46**, 111–115.

Ziegel, J., Baddeley, A., Dorph-Petersen, K. A., and Jensen, E. B. V. (2009). Systematic sampling with errors in sample locations. *To appear in Biometrika*.