Cavalieri-type estimation for perturbed systematic sampling

Mads Stehr (mads.stehr@math.au.dk) - Joint work with Markus Kiderlen
CSGB, Aarhus University, Denmark

Introduction

The **aim** is to estimate the volume of a solid $Y \subseteq \mathbb{R}^n$ from Lebesgue measurements on the $(n-1)$-dimensional intersections with random parallel hyperplanes. We will consider the case $n = 3$, see Figure 1.

Notation: If H_y is the hyperplane orthogonal to some reference unit vector ω with distance y from the origin, we define the *measurement function* f as

$$f(y) = \lambda_2(Y \cap H_y) = A(Y \cap H_y) .$$

Assumptions: For some fixed $t > 0$, U is uniform on $(0, t]$.

Classical Cavalieri estimator [1] based on the equidistant sampling positions $\{y_k\} = \{U + kt\}$ with section-spacing $t > 0$ is

$$\hat{V} = t \sum_{k \in \mathbb{Z}} f(y_k) , \qquad (1)$$

Properties: $\hat{V}$ is unbiased for $V(Y) = \int_{\mathbb{R}} f dy$.

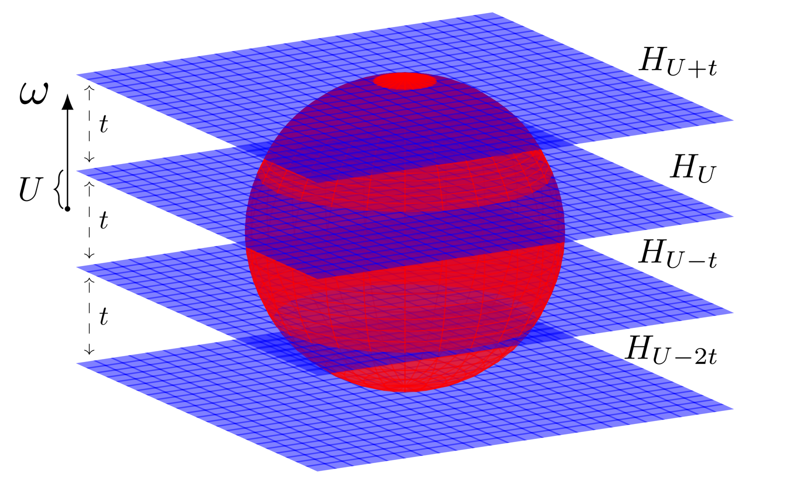

Figure 1: Classical set-up for estimation of volume of the unit ball in $\mathbb{R}^3$.

Generalized Cavalieri

Problem: Equidistant sampling positions are rarely realistic in applications.

New model: Sampling positions $X = \{x_k\}$ form a point process on $\mathbb{R}$.

Assumptions: X is a stationary point process with intensity $\gamma > 0$.

Generalized Cavalieri estimator [1]: With the average distance between consecutive points in X being $\bar{t} = \gamma^{-1}$,

$$\hat{V}_0 = \bar{t} \sum_{k \in \mathbb{Z}} f(x_k) , \qquad (2)$$

Properties: $\hat{V}_0$ is unbiased for $V(Y)$ with $\mathbb{V}\mathrm{ar}\hat{V}_0$ potentially considerably larger [1] than $\mathbb{V}\mathrm{ar}\hat{V}$. If $x_k = U + kt$ for all $k \in \mathbb{Z}$ then $\hat{V}_0 = \hat{V}$.

Quadrature rules & perturbed systematic sampling

Geometry

The classical Cavalieri estimator approximates the integral of the measurement function f by a Riemann sum. Using the same approximation rule when the sampling positions are not equidistant causes errors, see Figure 2.

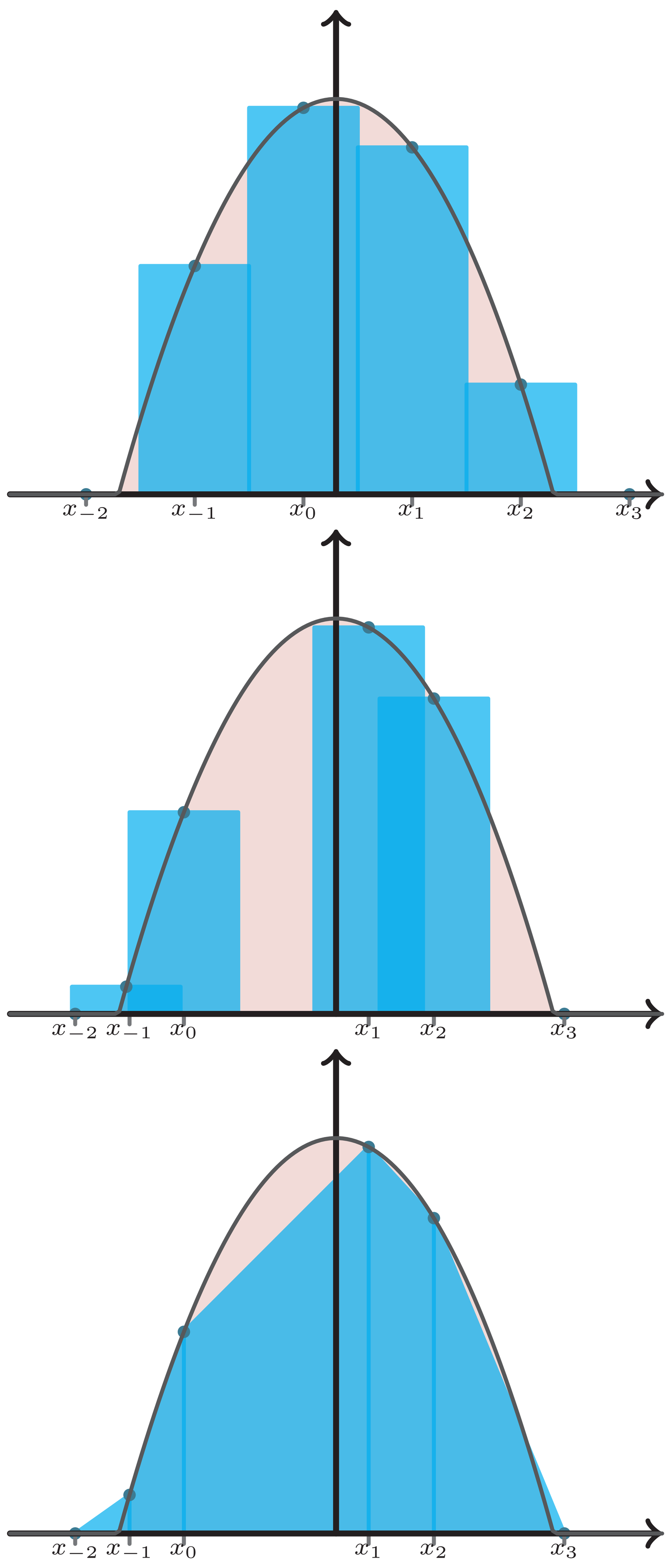

Figure 2: The rectangular quadrature rule for equidistant points $\{x_k\}$ as used in the classical Cavalieri (top), the corresponding rule when applied to points that are not equidistant (middle), and the trapezoidal rule (bottom) which approximates with a piecewise linear function. The measurement function is $f(x) = 2(1 - x^2)$ for $x \in [-1, 1]$ (Figures with courtesy from [2]).

Improvements [2]

Assumptions: Distances between consecutive sampling positions in the stationary point process $X = \{x_k\}$ are available.

Method: Use higher order quadrature rules to approximate measurement the function f, see Figure 2.

Trapezoidal rule: Approximate f by a piecewise linear function. Yields the estimator

$$\hat{V}_1 = \sum_{k \in \mathbb{Z}} \frac{x_{k+1} - x_{k-1}}{2} f(x_k) , \qquad (3)$$

which is unbiased for $V(Y)$.

Simpson's rule: Approximates f by a piecewise quadratic polynomial. Under mild integrability conditions on X, Simpson's rule leads to an unbiased estimator.

Perturbed Sampling

Sampling positions: Randomly perturbed from intended equidistant locations, i.e. $X = \{x_k\} = \{U + kt + D_k\}$

Error assumptions: Errors $\{D_k\}$ are iid and independent of U. $D_1 \sim h \cdot \lambda$, $\mathbb{E}D_1 = 0$ and $\mathbb{V}\mathrm{ar}D_1 = \sigma^2$.

Variance formula [1] generalized Cavalieri (2):

$$\mathbb{V}\mathrm{ar}\hat{V}_0 = tg(0) + t \sum_{k \neq 0} g * h * \check{h}(kt) - \int_{\mathbb{R}} g(z) dz \quad (4)$$

where $\check{h}(x) = h(-x)$ and $g = f * \check{f}$ is the *covariogram* of f.

Variance formula trapezoidal rule (3):

$$\begin{aligned}\mathbb{V}\mathrm{ar}\hat{V}_1 =& \mathbb{V}\mathrm{ar}\hat{V}_0 + \frac{1}{2t}(\mathrm{ID}\check{h}) * (\mathrm{ID}h) * g(t) \\ & - (\mathrm{ID}\check{h}) * h * g(t) - (\mathrm{ID}h) * \check{h} * g(t) \\ & + \frac{\sigma^2}{2t} g(0) - \frac{1}{2t} g * h * \check{h}(2t) , \qquad (5)\end{aligned}$$

with $\mathrm{ID}h$ defined as $\mathrm{ID}h(x) = xh(x)$.

Variance

Theoretical variance

Figure 3 shows that the trapezoidal rule indeed leads to a decrease in variance compared to the generalized Cavalieri estimator when used to determine the volume of the unit ball with randomly perturbed sampling.

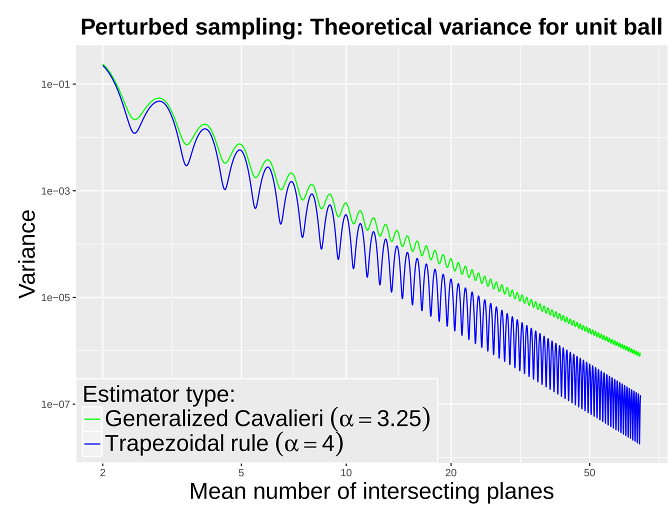

Figure 3: Theoretical variance of generalized Cavalieri and trapezoidal rule estimator found by (4) and (5) respectively. D_1 is assumed uniform on $(-0.1225t/2, 0.1225t/2)$ and α approximates the rate of decrease.

Sample variance

Figure 4 displays the sample variance for the generalized Cavalieri estimator, the trapezoidal rule and Simpson's rule for simulated perturbed sampling positions with D_1 uniform on $(-0.1225t/2, 0.1225t/2)$. Notice that the trapezoidal rule and Simpson's rule in both cases display a rate of decrease similar to the classical case with equidistant section-spacing.

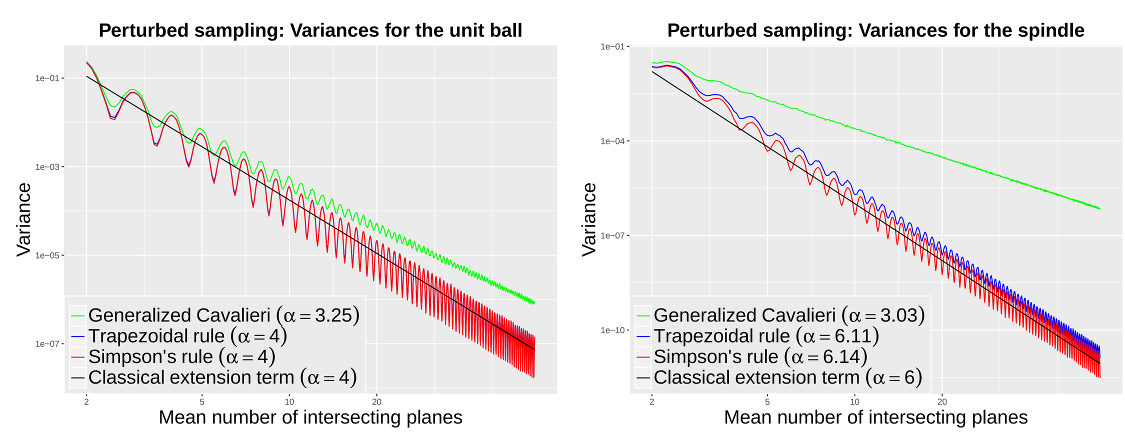

Figure 4: Sample variance for volume estimation of the unit ball (left) and a spindle shaped solid S with measurement function $f_S(x) = \frac{\pi}{2}(1 + \cos \pi x)$ for $x \in [-1, 1]$ (right).

References

[1] A.J. Baddeley, K.A. Dorph-Petersen, and E.B.V. Jensen. A note on the stereological implications of irregular spacing of sections. Journal of Microscopy, 222(3):177–181, 2006.

[2] M. Kiderlen and K.A. Dorph-Petersen. The cavalieri estimator with unequal section spacing revisited. CSGB Reasearch Report 2017, 04, 2017.