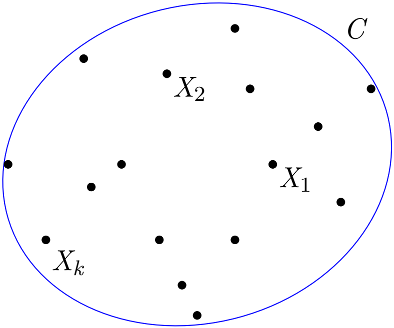
Unbiased estimation of the volume of a convex body

Nikolay Baldin
joint work with Markus Reiß

Institut für Mathematik
Humboldt-Universität zu Berlin

August 31, 2015. Frejus

Problem

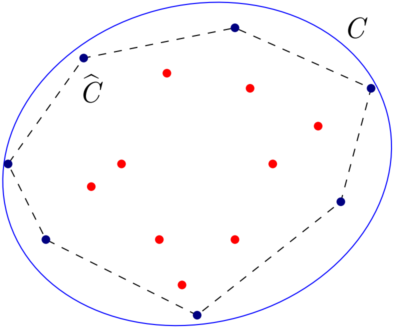


Observations: $X_1, ..., X_N \overset{i.i.d.}{\sim} U(C)$, $C \subseteq \mathbb{R}^d$ compact convex set

Target: Area/volume $\vartheta = |C|$ of C

Challenge: No estimators statistically or computationally attractive have been found

Take-home



$$\widehat{\vartheta} = \frac{N+1}{N_\circ + 1} |\widehat{C}|$$

- $\widehat{C} = \mathrm{conv}\{X_1, .., X_N\}$ convex hull of observations
- $N_\circ$ - number of observations in the interior of $\widehat{C}$ (red points)

“Poissonization” and oracle estimator

“Poissonization”: $N \sim \text{Poiss}(\lambda|C|)$ independently of $X_1, ..., X_N$

Equivalently, one can say $X_1, ..., X_N$ is an observation of a homogeneous *Poisson point process* (PPP) of intensity $\lambda > 0$ on C.

Oracle estimator for known intensity λ:

$$\boxed{\widehat{\vartheta}_{oracle} = |\widehat{C}| + N_{\widehat{C}}/\lambda}$$

where $N_{\widehat{C}}$ is the number of vertices of $\widehat{C}$.

Main oracle result (B, Reiß 2015)
$\widehat{\vartheta}_{oracle}$ is UMVU with $\text{Var}(\widehat{\vartheta}_{oracle}) = \mathbb{E}[|C \setminus \widehat{C}|]/\lambda$. For $\lambda \to \infty$ we have $\text{Var}(\widehat{\vartheta}_{oracle}) \lesssim \lambda^{-(d+3)/(d+1)}$ (missing volume asymptotics). This rate is minimax optimal.

Rate-improvement over $|\widehat{C}|$.
$d = 1 : \lambda^{-2} \to \lambda^{-2}$ $\quad d = 2 : \lambda^{-4/3} \to \lambda^{-5/3}$ $\quad d = 3 : \lambda^{-1} \to \lambda^{-3/2}$

Efficient estimation for unknown intensity

Idea: Use $N_\circ \,|\, \widehat{C} \sim \text{Poiss}(\lambda|\widehat{C}|)$ for number $N_\circ$ of interior points of $\widehat{C}$.
Almost unbiased estimator: $\widehat{\lambda^{-1}} = |\widehat{C}|/(N_\circ + 1)$

Final estimator:

$$\boxed{\widehat{\vartheta} = |\widehat{C}| + N_{\widehat{C}}\widehat{\lambda^{-1}} = \frac{N+1}{N_\circ + 1}|\widehat{C}|}$$

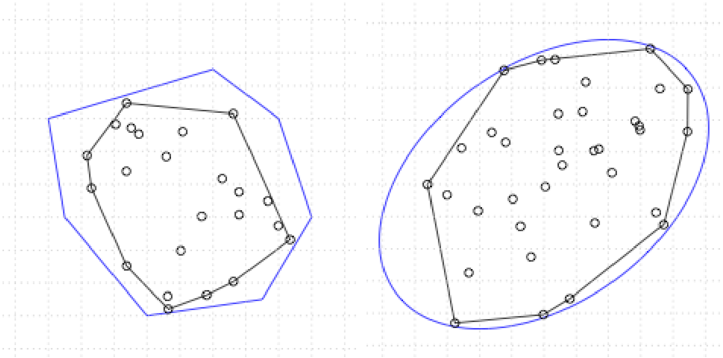
Main result (B, Reiß 2015)

$\widehat{\vartheta}$ has exponentially small bias and

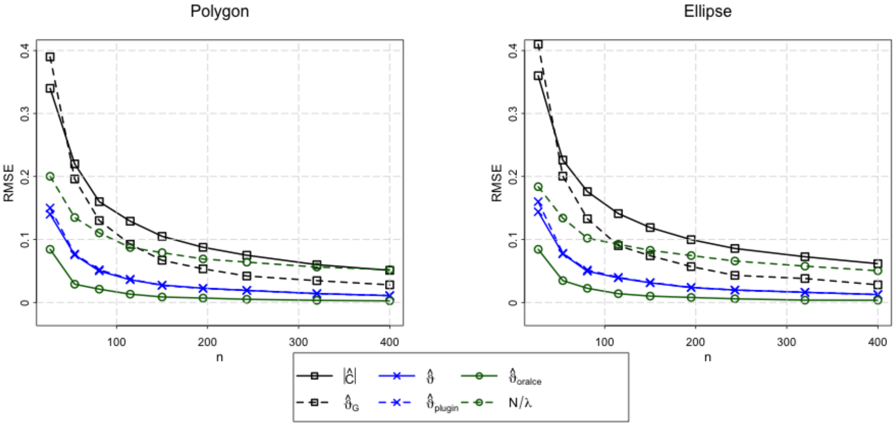
$$\text{Var}(\widehat{\vartheta}) = \Big(1 + \mathcal{O}\Big(\frac{\mathbb{E}[|C \setminus \widehat{C}|^2]}{\mathbb{E}[|C \setminus \widehat{C}|]}\Big)\Big)\,\text{Var}(\widehat{\vartheta}_{oracle})$$

In all known cases ($d \leq 2$, polytopes, smooth boundary):
$\text{Var}(\widehat{\vartheta}) = (1 + \mathcal{O}(\mathbb{E}[|C \setminus \widehat{C}|]))\,\text{Var}(\widehat{\vartheta}_{oracle})$, so the oracle inequality is *exact*.

Numerical study. Planar case, $n = \lambda|C|$



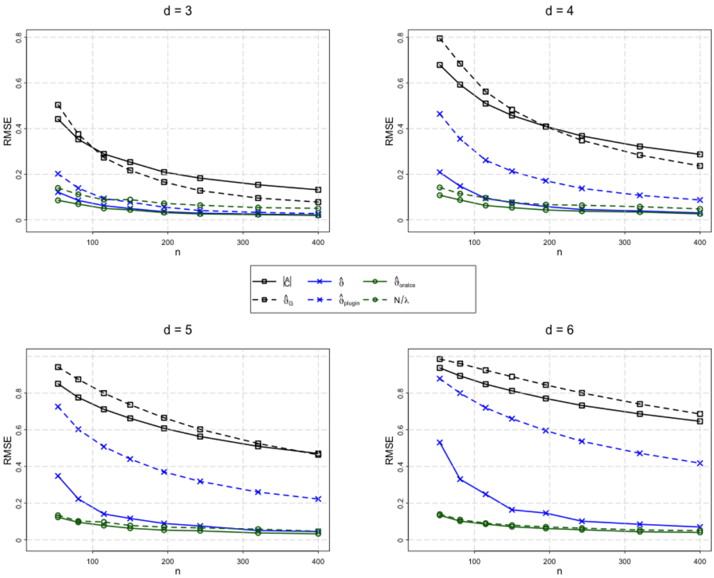
Numerical study. Planar case, $n = \lambda|C|$



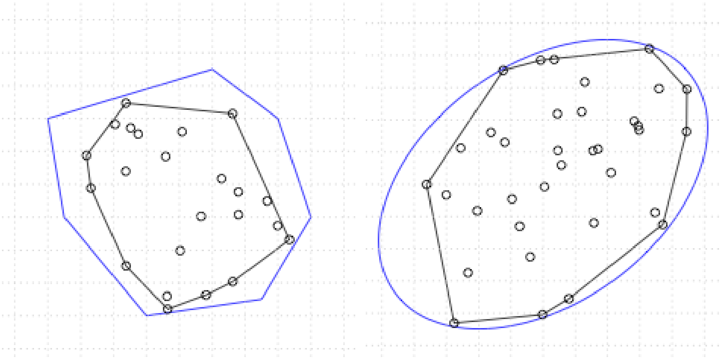
Competitors: $|\widehat{C}|$, the naive oracle estimator N/λ, the UMVU oracle estimator $\widehat{\vartheta}_{oracle}$ and the plug-in MLE estimator $\widehat{\vartheta}_{plugin} = |\widehat{C}|(1 + N_{\widehat{C}}/N)$.

Striking fact: The empirical bias of the proposed estimator $\widehat{\vartheta}$ s always of smaller order than 10^{-3}.

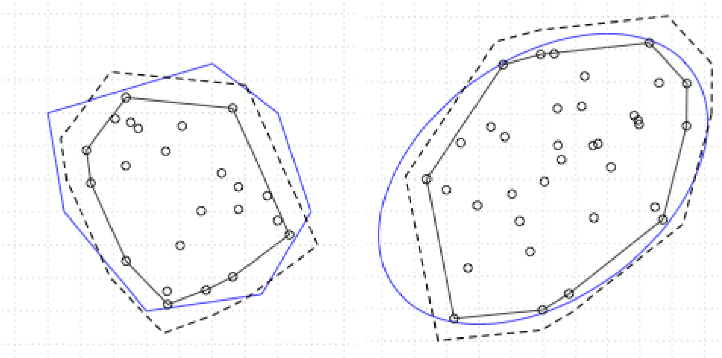
Higher dimensions. $C = [0,1]^d$, $d = 3, 4, 5, 6$



Applications



Applications



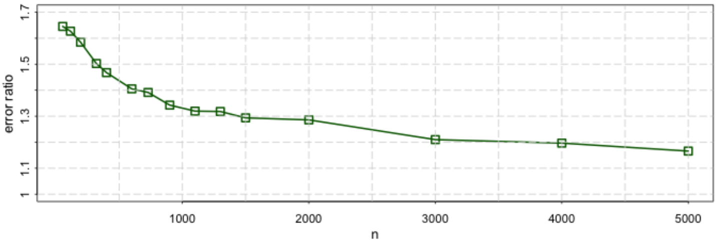
A new dilated-hull estimator for a convex set:

$$\widetilde{C} \stackrel{\text{def}}{=} \left\{ \widehat{x}_0 + \left(\frac{\widehat{\theta}}{|\widehat{C}|} \right)^{1/d} (x - \widehat{x}_0) \,\middle|\, x \in \widehat{C} \right\}$$

$$= \left\{ \widehat{x}_0 + \left(\frac{N+1}{N_\circ + 1} \right)^{1/d} (x - \widehat{x}_0) \,\middle|\, x \in \widehat{C} \right\},$$

Error ratio

$$\frac{\mathbb{E}[|C \Delta \widehat{C}|]}{\mathbb{E}[|C \Delta \widetilde{C}|]}$$



Statistical guarantees $d = 1$

$$\frac{\mathbb{E}[|C \Delta \widehat{C}|]}{\mathbb{E}[|C \Delta \widetilde{C}|]} = \frac{1}{2}(1 + 1/n)^n \to \frac{e}{2}$$

Wrap-up

- We have constructed the estimator for the volume of a convex body, which is minimax optimal and also efficient non asymptotically: it is almost unbiased and its variance is close to the minimal variance among all unbiased estimators.
- The results are non asymptotic and assumptions-free.
- The paper is available on arXiv.

Open questions in stochastic geometry:

1) Does $\lambda \operatorname{Var}(|C \setminus \widehat{C}|) \asymp \mathbb{E}(|C \setminus \widehat{C}|)$ hold for all $C \in \mathbf{C}$?
2) Can we get a uniform bound for the variance $\operatorname{Var}(|C \setminus \widehat{C}|) = O(\lambda^{-(d+3)/(d+1)})$?
3) Is the variance $\operatorname{Var}(|C \setminus \widehat{C}|)$ the largest for a unit ball for all C with $|C| = 1$?