Unbiased estimation of the volume of a convex body

Nikolay Baldin
Institute of Mathematics
Humboldt-Universität zu Berlin
baldin.np@gmail.com

Markus Reiβ*
Institute of Mathematics
Humboldt-Universität zu Berlin
mreiss@math.hu-berlin.de

February 20, 2015

Abstract

Based on observations of points uniformly distributed over a convex set in $\mathbb{R}^d$, a new estimator for the volume of the convex set is proposed. The estimator is minimax optimal and also efficient non-asymptotically: it is nearly unbiased with minimal variance among all unbiased oracle-type estimators. Our approach is based on a Poisson point process model and as an ingredient, we prove that the convex hull is a sufficient and complete statistic. No hypotheses on the boundary of the convex set are imposed. In a numerical study, we show that the estimator outperforms earlier estimators for the volume. In addition, an improved set estimator for the convex body itself is proposed.

Keywords: volume estimation, convex sets, Poisson point process, UMVU, gift-wrapping algorithm, exact oracle inequality, missing volume

MSC code: 60G55, 62G05, 62M30

*Financial support by the DFG-research unit FOR1735 Structural Inference in Statistics is gratefully acknowledged
1 Introduction

Driven by applications in image analysis and signal processing, the estimation of the support of a density attracts a lot of statistical activity. In many cases it is natural to assume a convex shape for the support set. First fundamental results for convex support estimation have been achieved by Korostelev and Tsybakov (1993, 1994) who prove minimax-optimal rates of convergence in Hausdorff distance for a set estimator. In particular, Korostelev and Tsybakov (1993) prove that the convex hull of the points \( \hat{C}_n \), which is a maximum likelihood estimator for the set \( C \), is rate-optimal. Interestingly, the volume \( |\hat{C}_n| \) of the convex hull is not rate-optimal for estimation of the volume \( |C| \) of the convex set and an alternative two-step estimator, optimal up to a logarithmic factor, was proposed. A fully rate-optimal estimator for the volume of a convex set with smooth boundary was then constructed by Gayraud (1997) based on three-fold sample splitting. For various extensions and applications of convex support estimation, let us refer to Mammen and Tsybakov (1995); Guntuboyina (2012); Brunel (2014) and the literature cited there. Related ideas under Hölder and monotonicity constraints, respectively, have been adopted by Reiß and Selk (2014) for a one-sided regression model.

The contribution of this paper is the construction of a very simple volume estimator which is not only rate-optimal over all convex sets without boundary restrictions, but even adaptive in the sense that it attains almost the parametric rate if the convex set is a polytope. Our approach is non-asymptotic and provides much more precise properties. The analysis is based on a Poisson point process (PPP) observation model with intensity \( \lambda > 0 \) on the convex set \( C \subseteq \mathbb{R}^d \). We thus observe

\[
X_1, \ldots, X_N \overset{i.i.d.}{\sim} U(C), \quad N \sim \text{Pois}(\lambda |C|),
\]

where \( X_1, \ldots, X_N, N \) are independent, see Section 2 below for a concise introduction to the PPP model. Using Poissonisation and de-Poissonisation techniques, this model exhibits the same asymptotic properties as a sample of \( n = \lambda |C| \) uniformly on \( C \) distributed random variables \( X_1, \ldots, X_n \). The beautiful geometry of the PPP model, however, allows for much more concise ideas and proofs, see also Meister and Reiß (2013) for connections between PPP and regression models with irregular error distributions. From an applied perspective, PPP models are often natural, e.g. for spatial count data of photons or other emissions.
For known intensity $\lambda$ of the PPP, we construct in Section 3 an oracle estimator $\hat{\vartheta}_{oracle}$. Then, Theorem 3.1 shows that this estimator is UMVU (uniformly of minimum variance among unbiased estimators) and rate-optimal. To this end, moment bounds from stochastic geometry for the missing volume of $\hat{C}$, obtained by Bárényi and Larman (1988) and Dwyer (1988), as well as the result of independent interest that $\hat{C}$ forms a sufficient and complete statistic are essential. For the more realistic case of unknown intensity $\lambda$, we analyse in Section 4 our final estimator

$$\hat{\vartheta} \overset{\text{def}}{=} \frac{N + 1}{N_0 + 1} |\hat{C}|,$$

where $\hat{C} = \text{conv}\{X_1, \ldots, X_N\}$ denotes the convex hull of the observed points and $N_0$ denotes the number of points in the interior of $\hat{C}$. We are able to prove a sharp oracle inequality, comparing the risk of this estimator to that of $\hat{\vartheta}_{oracle}$. Here, very recent and advanced results by Reitzner (2003); Pardon (2011); Beermann and Reitzner (2015) on the variance of the number of points $N_\hat{C}$ and the missing volume $|C \setminus \hat{C}|$ are of key importance. This fascinating interplay between stochastic geometry and statistics prevails throughout the work.

The lower bound showing that $\hat{\vartheta}$ is indeed minimax-optimal is proved in Section 5 by constructing an asymptotically least-favourable Bayesian prior. A small simulation study is presented in Section 6. Moreover, we propose to enlarge the convex hull set by the factor $(N + 1)/(N_0 + 1)$ and we study its error as an estimator of the set $C$ itself. The proof of Lemma 4.1 is deferred to the appendix.

## 2 Theoretical digression

Most of the notations and results are adapted from Karr (1991). We fix a compact convex set $E$ in $\mathbb{R}^d$ with non-empty interior as a state space and denote by $\mathcal{E}$ its Borel $\sigma$-algebra. We define the family of convex subsets $C = \{C \subseteq E, \text{convex, closed}\}$ (this implies that all sets in $C$ are compact). It is natural to equip the space $C$ with the Hausdorff-metric $d_H$ and its Borel $\sigma$-algebra $\mathcal{B}_C$. Then $(C, d_H)$ is a compact and thus separable space and the mapping $(x_1, \ldots, x_k) \mapsto \text{conv}\{x_1, \ldots, x_k\}$, which generates the convex hull of points $x_i \in E$, is continuous from $E^k$ to $(C, d_H)$. 

On \((\mathbb{E}, \mathcal{E})\) we define the set of point measures \(\mathcal{M} = \{ m \text{ measure on } \mathbb{E} : m(A) \in \mathbb{N}, \forall A \in \mathcal{E} \}\) equipped with the \(\sigma\)-algebra \(\mathcal{M} = \sigma(m \mapsto m(A), A \in \mathcal{E})\). Fix a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and a convex set \(C \in \mathbb{C}\). We call a measurable mapping \(\mathcal{N} : \Omega \to \mathcal{M}\) a Poisson point process (PPP) of intensity \(\lambda > 0\) on \(C\) if

- for any \(A \in \mathcal{E}\), we have \(\mathcal{N}(A) \sim \text{Poiss}(\lambda |A \cap C|)\), where \(|A \cap C|\) denotes the Lebesgue measure of \(A \cap C\);
- for all mutually disjoint sets \(A_1, \ldots, A_n \in \mathcal{E}\), the random variables \(\mathcal{N}(A_1), \ldots, \mathcal{N}(A_n)\) are independent.

A more constructive and intuitive representation of the PPP \(\mathcal{N}\) is \(\mathcal{N} = \sum_{i=1}^{\infty} \delta_{X_i}\) for \(\mathcal{N} \sim \text{Poiss}(\lambda|C|)\) and i.i.d. random variables \((X_i)\), independent of \(\mathcal{N}\) and distributed uniformly \(\mathbb{P}(X_i \in A) = |A \cap C|/|C|\), so that \(\mathcal{N}(A) = \sum_{i=1}^{\infty} 1(X_i \in A)\) for any \(A \in \mathcal{E}\).

We shall consider the convex hull of the PPP points \(\hat{C} = \text{conv}\{X_1, \ldots, X_N\}\), which by the above continuity property of the convex hull is a random element with values in the Polish space \((\mathbb{C}, d_H)\). In the sequel, conditional expectations and probabilities with respect to \(\hat{C}\) are thus well defined. We can also evaluate the probability

\[
\mathbb{P}_C(\hat{C} \in A) = \sum_{k=0}^{\infty} \frac{e^{-\lambda|C|} \lambda^k}{k!} \int_{C^k} 1(\text{conv}\{x_1, \ldots, x_k\} \in A)d(x_1, \ldots, x_k)
\]

for \(A \in \mathcal{B}_C\). Usually, we only write the subscript \(C\) or sometimes \((C, \lambda)\) when different probability distributions are considered simultaneously. The likelihood function \(d\mathbb{P}_{C, \lambda} / d\mathbb{P}_{E, \lambda_0}\) for \(C \in \mathbb{C}\) and \(\lambda, \lambda_0 > 0\) is then given by, cf. Thm. 1.3 in Kutoyants (1998):

\[
d\mathbb{P}_{C, \lambda} / d\mathbb{P}_{E, \lambda_0}(X_1, \ldots, X_N) = e^{\lambda_0 \mathbb{E}[\lambda|C|]} (\lambda/\lambda_0)^N 1(\forall i = 1, \ldots, N : X_i \in C)
\]

\[
= e^{\lambda_0 \mathbb{E}[\lambda|C|]} (\lambda/\lambda_0)^N 1(\hat{C} \subseteq C). \quad (2.1)
\]

For the last line, we have used that a point set is in \(C\) if and only if its convex hull is contained in \(C\).

Finally, for asymptotic bounds we write \(f(x) = O(g(x))\) or \(f(x) \lesssim g(x)\) to say that \(f(x)\) is bounded by a constant multiple of \(g(x)\) and \(f(x) \sim g(x)\) if \(f(x) \lesssim g(x)\) as well as \(g(x) \lesssim f(x)\).
3 Oracle case: intensity $\lambda$ is known

For a PPP on $C \in \mathcal{C}$ with intensity $\lambda > 0$, we know $N \sim \text{Poiss}(\lambda|C|)$. In the oracle case, when the intensity $\lambda$ is known, $N/\lambda$ estimates $|C|$ without bias and yields for $\lambda \to \infty$ the classical parametric rate $\lambda^{-1/2}$:

$$\mathbb{E}[(N/\lambda - |C|)^2] = \lambda^{-2}\text{Var}(N) = \frac{|C|}{\lambda}.$$

Another natural idea might be to use the plug-in estimator $|\hat{C}|$ whose error is given by the missing volume and satisfies

$$\mathbb{E}[(|\hat{C}| - |C|)^2] = \mathbb{E}[|C \setminus \hat{C}|^2] = O(|C|^{2(d-1)/(d+1)}\lambda^{-4/(d+1)}),$$

where the bound is obtained similarly to (3.4) and (3.5) below. This means that its error is of smaller order than $\lambda^{-1}$ for $d \leq 2$, but larger for $d \geq 4$. For any $d \geq 2$, however, both convergence rates are worse than the minimax-optimal rate $\lambda^{-(d+3)/(d+1)}$, established below.

The way to improve these estimators is to observe that by the likelihood representation (2.1) for $\lambda = \lambda_0$ and the Neyman factorisation criterion the convex hull is a sufficient statistic. Consequently, by the Rao-Blackwell theorem, the conditional expectation of $N/\lambda$ given the convex hull $\hat{C}$ is an estimator with smaller mean squared error (MSE).

The number of points $N$ can be split into the number of points on the convex hull $N_{\hat{C}}$, which is measurable with respect to the convex hull $\hat{C}$, and the number of points in the interior of the convex hull $N_0$. Using independence (the convex hull can be constructed by sweeping the space from the outside, the so called gift-wrapping algorithm, see Jarvis (1973)), $N_0$ is, conditionally on the convex hull $\hat{C}$, Poisson-distributed:

$$N_0 \mid \hat{C} \sim \text{Poiss}(\lambda_0) \text{ with } \lambda_0 \overset{\text{def}}{=} \lambda|\hat{C}|.$$

Thus, we obtain the oracle estimator

$$\hat{\theta}_{\text{oracle}} \overset{\text{def}}{=} \mathbb{E}\left[\frac{N}{\lambda} \mid \hat{C}\right] = \mathbb{E}\left[\frac{N_0 + N_{\hat{C}}}{\lambda} \mid \hat{C}\right] = |\hat{C}| + \frac{N_{\hat{C}}}{\lambda}. \quad (3.2)$$

**Theorem 3.1.** For known intensity $\lambda > 0$, the oracle estimator $\hat{\theta}_{\text{oracle}}$ is unbiased and of minimal variance among all unbiased estimators (UMVU).
It satisfies
\[
\text{Var}(\hat{\vartheta}_{\text{oracle}}) = \frac{1}{\lambda} \mathbb{E}[|C \setminus \hat{C}|].
\]
Its worst case mean squared error over \( C \) decays as \( \lambda \uparrow \infty \) like \( \lambda^{-(d+3)/(d+1)} \)
in dimension \( d \):
\[
\limsup_{\lambda \to \infty} \lambda^{(d+3)/(d+1)} \sup_{C \in \mathcal{C}} \left\{ |C|^{-(d-1)/(d+1)} \mathbb{E}[(\hat{\vartheta}_{\text{oracle}} - |C|)^2] \right\} < \infty.
\]

**Remark 3.2.** The theorem implies that the rate of convergence for the RMSE (root mean-squared error) of the estimator \( \hat{\vartheta}_{\text{oracle}} \) is \( \lambda^{-(d+3)/(2d+2)} \). In Theorem 5.1 below, we prove that the lower bound on the minimax risk in the PPP model is of the same order implying that the rate is minimax-optimal. Even more, the oracle estimator is adaptive to the class of polytopes: as it was shown in Bárány and Larman (1988) and independently in Dwyer (1988), \( \mathbb{E}[|C \setminus \hat{C}|] \sim \lambda^{-1} (\log(\lambda |C|))^{d-1} \) when the true set \( C \) is a polytope, which implies a faster (almost parametric) rate of convergence for the RMSE of the oracle estimator.

**Proof.** The unbiasedness follows immediately from the definition (3.2). By the law of total variance, we obtain
\[
\text{Var}(\hat{\vartheta}_{\text{oracle}}) = \text{Var} \left( \frac{N}{\lambda} \right) - \mathbb{E} \left[ \text{Var} \left( \frac{N}{\lambda} \mid \hat{C} \right) \right] = \frac{|C|}{\lambda} - \mathbb{E} \left[ \text{Var} \left( \frac{N_0}{\lambda} \mid \hat{C} \right) \right]
\]
\[
= \frac{|C|}{\lambda} - \mathbb{E} \left[ \frac{\lambda_0}{\lambda^2} \right] = \frac{1}{\lambda} \mathbb{E}[|C \setminus \hat{C}|]. \quad (3.3)
\]

Proposition 3.3 below affirms that the convex hull \( \hat{C} \) is not only a sufficient, but also a complete statistic such that by the Lehmann-Scheffé theorem, the estimator \( \hat{\vartheta}_{\text{oracle}} \) has the UMVU property.

Finally, we bound the expectation of the missing volume \( |C \setminus \hat{C}| \) by Poissonisation, i.e. using that the convex hull \( \hat{C} \) in the PPP model conditionally on the event \{\( N = k \)\} is distributed as the convex hull \( \hat{C}_k = \text{conv}\{X_1, \ldots, X_k\} \) in the model with \( k \) uniform observations on \( C \), for which the following upper bound is known (e.g., Bárány and Larman (1988)):
\[
\sup_{C \in \mathcal{C}} \mathbb{E} \left[ \frac{|C \setminus \hat{C}_k|}{|C|} \right] = O(k^{-2/(d+1)}). \quad (3.4)
\]
Thus, it follows by a Poisson moment bound

\[
\sup_{C \subseteq \mathcal{C}} \mathbb{E}\left[\frac{|C \setminus \hat{C}|}{|C|^{(d-1)/(d+1)}}\right] = \sup_{C \subseteq \mathcal{C}} \sum_{k=0}^{\infty} \frac{e^{-\lambda|C|} (\lambda|C|)^k}{|C|^{-2/(d+1)} k!} \mathbb{E}\left[\frac{|C \setminus \hat{C}_k|}{|C|}\right]
\]

\[= O\left(\lambda^{-2/(d+1)}\right). \tag{3.5}\]

This bound, together with (3.3), yields the assertion. \hfill \Box

**Proposition 3.3.** For known intensity \( \lambda > 0 \), the convex hull \( \hat{C} = \text{conv}\{X_1, \ldots, X_N\} \) is a complete statistic.

**Proof.** We need to show the implication

\[\forall C \in \mathcal{C} : \mathbb{E}_C[T(\hat{C})] = 0 \implies T(\hat{C}) = 0 \text{ } \mathbb{P}_E - \text{a.s.}\]

for any \( \mathcal{B}_C \)-measurable function \( T : \mathcal{C} \to \mathbb{R} \). From the likelihood in (2.1) for \( \lambda = \lambda_0 \), we derive

\[\mathbb{E}_C[T(\hat{C})] = \mathbb{E}_E[T(\hat{C}) \exp(\lambda|\mathbb{E} \setminus C|) 1(\hat{C} \subseteq C)].\]

Since \( \exp(\lambda|\mathbb{E} \setminus C|) \) is deterministic, \( \mathbb{E}_C[T(\hat{C})] = 0 \) for all \( C \in \mathcal{C} \) implies

\[\forall C \in \mathcal{C} : \mathbb{E}_E[T(\hat{C}) 1(\hat{C} \subseteq C)] = 0.\]

For \( C \in \mathcal{C} \), define the family of convex subsets of \( C \) as \( [C] = \{A \in \mathcal{C} \mid A \subseteq C\} \) such that \( \hat{C} \subseteq C \iff \hat{C} \in [C] \). Splitting \( T = T^+ - T^- \) with non-negative \( \mathcal{B}_C \)-measurable functions \( T^+ \) and \( T^- \), we infer that the measures

\[\mu^\pm(B) = \mathbb{E}_E[T^\pm(\hat{C}) 1(\hat{C} \in B)], \quad B \in \mathcal{B}_C, \text{ agree on } \{[C]\mid C \in \mathcal{C}\}.\]

Note that the brackets \( \{[C]\mid C \in \mathcal{C}\} \) are \( \cap \)-stable due to \( [A] \cap [C] = [A \cap C] \) and \( A \cap C \in \mathcal{C} \). If the \( \sigma \)-algebra \( \mathcal{C} \) generated by \( \{[C]\mid C \in \mathcal{C}\} \) contains \( \mathcal{B}_C \), the uniqueness theorem asserts that the measures \( \mu^+, \mu^- \) agree on all Borel sets in \( \mathcal{B}_C \), in particular on \( \{T > 0\} \) and \( \{T < 0\} \), which entails \( \mathbb{E}_E[T^+(\hat{C})] = \mathbb{E}_E[T^-(\hat{C})] = 0 \). Thus, in this case, \( T(\hat{C}) = 0 \) holds \( \mathbb{P}_E \)-a.s.

It remains to show that the brackets \( [C] \) generate the Borel \( \sigma \)-algebra \( \mathcal{B}_C \). This can be derived as a non-trivial consequence of Choquet’s theorem, see Thm. 7.8 in Molchanov (2006), but we propose a short self-contained proof here. Let us define the family \( \langle C \rangle = \{B \in \mathcal{C} \mid C \subseteq B\} \) of convex sets
containing \( C \). Then the closed Hausdorff ball with center \( C \) and radius \( \varepsilon > 0 \) has the representation

\[
B_\varepsilon(C) \overset{\text{def}}{=} \{ A \in C \mid d_H(A, C) \leq \varepsilon \} = \{ A \in C \mid U_{-\varepsilon}(C) \subseteq A \subseteq U_\varepsilon(C) \},
\]

with \( U_\varepsilon(C) = \{ x \in E \mid \text{dist}(x, C) \leq \varepsilon \} \), \( U_{-\varepsilon}(C) = \{ x \in C \mid \text{dist}(x, E \setminus C) \leq \varepsilon \} \). Noting that \( U_\varepsilon(C), U_{-\varepsilon}(C) \) are closed and convex and thus in \( C \), we obtain

\[
B_\varepsilon(C) = \langle U_{-\varepsilon}(C) \rangle \cap [U_\varepsilon(C)].
\]

Since \( (C, d_H) \) is separable, our problem is reduced to proving that all angle sets \( \langle C \rangle \) for \( C \in C \) are in \( C \). A further reduction is achieved by noting \( \langle C \rangle = \cap_{x \in C} \langle x \rangle = \cap_{x \in C \cap \mathbb{Q}^d} \langle x \rangle \) setting \( \langle x \rangle = \langle \{ x \} \rangle \) for short such that it suffices to prove \( \langle x \rangle \in C \) for all \( x \in E \).

Now, \( x \notin C \) implies by the Hahn-Banach theorem that there are \( \delta > 0, \nu \in \mathbb{R}^d \) such that \( \langle \nu, c-x \rangle \geq \delta \) holds for all \( c \in C \). By a density argument, we may choose \( \delta \in \mathbb{Q}^+ \) and \( \nu \in \mathbb{Q}^d \). Denoting the corresponding hyperplane intersected with \( E \) by \( H_{\delta, \nu} = \{ \xi \in E \mid \langle \nu, \xi - x \rangle \geq \delta \} \), see Figure 1, we conclude

\[
\langle x \rangle^C = \bigcup_{\delta \in \mathbb{Q}^+} \bigcup_{\nu \in \mathbb{Q}^d} [H_{\delta, \nu}]_{\xi \in C} \in C.
\]

Consequently, \( \langle x \rangle \in C \) and thus \( B_C \subseteq C \) hold.

\[ \square \]
4 Unknown intensity $\lambda$: nearly unbiased estimation

In the case, when the intensity $\lambda$ is unknown and the oracle estimator $\hat{\vartheta}_{\text{oracle}}$ in (3.2) is inaccessible, the maximum-likelihood approach suggests to use $N/|\hat{C}|$ as an estimator for $\lambda$ in (2.1). This yields the plug-in estimator for the volume,

$$\hat{\vartheta}_{\text{plugin}} \overset{\text{def}}{=} |\hat{C}| + \frac{N_{\hat{C}}}{N} |\hat{C}|.$$ 

In the unlikely event $N = |\hat{C}| = 0$, we define $\hat{\vartheta}_{\text{plugin}} = 0$. This estimator has a significant bias due to the following result, which is proved in the appendix.

**Lemma 4.1.** For the bias of the plug-in MLE estimator $\hat{\vartheta}_{\text{plugin}}$, it follows with some universal constant $c > 0$

$$|C| - \mathbb{E}[\hat{\vartheta}_{\text{plugin}}] \geq c \mathbb{E}[|\hat{C} \setminus C|^2], \quad \forall C \in \mathcal{C}. \quad (4.1)$$

The maximal bias over $C \in \mathcal{C}$ is thus of the order $\lambda^{-4/(d+1)}$, which is worse than the minimax rate $\lambda^{-(d+3)/(2d+2)}$ for $d > 5$. Yet, in the two-dimensional finite sample study of Section 6 below, its performance is quite convincing. We surmise that $\hat{\vartheta}_{\text{plugin}}$ is rate-optimal for $d \leq 5$, but we leave that question aside because the final estimator we propose will be nearly unbiased and will satisfy an exact oracle inequality. In particular, it is rate-optimal in any dimension. The new idea is to exploit that the number of interior points of $\hat{C}$ satisfies $N_0 |\hat{C} \sim \text{Poiss}(\lambda_0)$, see (3.1).

**Remark 4.2.** There is no conditionally unbiased estimator for $\lambda_0^{-1}$ based on observing $N_0 |\hat{C} \sim \text{Poiss}(\lambda_0)$ for $\lambda_0$ ranging over some open (non-empty) interval. Otherwise, an estimator $\tilde{\mu}(N_0)$ for $\lambda_0^{-1}$ would satisfy $\mathbb{E}[\tilde{\mu}(N_0)|\hat{C}] = \lambda_0^{-1}$ implying

$$\sum_{k=0}^{\infty} \frac{\lambda_0^k}{k!} \tilde{\mu}(k) e^{-\lambda_0} = \lambda_0^{-1} \Rightarrow \sum_{k=0}^{\infty} \frac{\lambda_0^{k+1}}{k!} \tilde{\mu}(k) = \sum_{k=0}^{\infty} \frac{\lambda_0^k}{k!}.$$ 

The coefficient for the constant term in the left and right power series would thus differ (0 versus 1), in contradiction with the uniqueness theorem for power series.
We provide an almost unbiased estimator for $\lambda_0^{-1}$ by noting that the first jump time of a Poisson process with intensity $\lambda_0$ is Exp($\lambda_0$)-distributed and thus has expectation $\lambda_0^{-1}$. Taking conditional expectation with respect to the Poisson process at time 1, we conclude that

$$\widehat{\mu}(N_0, \lambda_0) \overset{\text{def}}{=} \begin{cases} (N_0 + 1)^{-1}, & \text{for } N_0 \geq 1, \\ 1 + \lambda_0^{-1}, & \text{for } N_0 = 0 \end{cases}$$

satisfies $\mathbb{E}[\widehat{\mu}(N_0, \lambda_0)|\widehat{C}] = \lambda_0^{-1}$. Omitting the term depending on $\lambda_0$ in the unlikely case $N_0 = 0$, we define our final estimator

$$\widehat{\vartheta} \overset{\text{def}}{=} |\widehat{C}| + \frac{N_0 \widehat{C}}{N_0 + 1} |\widehat{C}|.$$

For the proofs, we also define the pseudo-estimator

$$\widehat{\vartheta}_{\text{pseudo}} \overset{\text{def}}{=} |\widehat{C}| + |\widehat{C}| N_0 \widehat{C} \left( \frac{1}{N_0 + 1} + \frac{e^{-\lambda_0}}{\lambda_0} \right).$$

**Theorem 4.3.** The pseudo-estimator $\widehat{\vartheta}_{\text{pseudo}}$ is unbiased and the estimator $\widehat{\vartheta}$ is asymptotically unbiased in the sense that with constants $c_1, c_2 > 0$ whenever $\lambda|C| \geq 1$:

$$0 \leq |C| - \mathbb{E}[\widehat{\vartheta}] \leq c_1 |C| \exp(-c_2(\lambda|C|)^{(d-1)/(d+1)}), \quad \forall C \in C.$$

**Proof.** We have

$$\mathbb{E} \left[ \frac{1}{N_0 + 1} + \frac{e^{-\lambda_0}}{\lambda_0} \mid \widehat{C} \right] = e^{-\lambda_0} \lambda_0^{-1} \left( \sum_{k=0}^{\infty} \frac{\lambda_0^{k+1}}{(k+1)!} + 1 \right) = \lambda_0^{-1},$$

which by $|\widehat{C}|\lambda_0^{-1} = \lambda^{-1}$ and $\mathbb{E}[\widehat{\vartheta}_{\text{oracle}}] = |C|$ implies unbiasedness of $\widehat{\vartheta}_{\text{pseudo}}$. Thus, it follows that

$$|C| - \mathbb{E}[\widehat{\vartheta}] = \mathbb{E} \left[ |\widehat{C}| N_0 e^{-\lambda_0} \lambda_0^{-1} \right] = \lambda^{-1} \mathbb{E} \left[ N_0 e^{-\lambda|\widehat{C}|} \right].$$

We exploit the deviation inequality from Thm. 1 in Brunel (2013) and derive the bound for the exponential moment of the missing volume in the model with fixed number of points

$$\mathbb{E} \left[ \exp \left( \lambda |C \setminus \widehat{C}_k| \right) \right] \leq b_1 \exp \left( b_2 \lambda |C| k^{-2/(d+1)} \right), \quad k \geq 2,$$
for positive constants $b_1, b_2$ depending on the dimension according to Brunel (2013). For the cases $k = 0, 1$, we have the identity $\mathbb{E}[\exp (\lambda |C \setminus \hat{C}|)] = \exp (\lambda |C|)$. By Poissonisation, similarly to (3.5), we derive

$$
\exp(-\lambda |C|)\mathbb{E}[\exp (\lambda |C \setminus \hat{C}|)] \leq b_3 \exp \left(-c_2(\lambda |C|)^{(d-1)/(d+1)}\right),
$$

for positive constants $b_3, c_2$ depending on the dimension. Hence, using the Cauchy-Schwarz inequality and the bound for the moments of the points on the convex hull,

$$
\mathbb{E}[N^q_C] = O\left((\lambda |C|)^{q(d-1)/(d+1)}\right), \quad q \in \mathbb{N},
$$

see e.g. Section 2.3.2 in Brunel (2014), we derive for a constant $c_1 > 0$

$$
\lambda^{-1}\mathbb{E}\left[N_C e^{-\lambda |\hat{C}|}\right] \leq \lambda^{-1}e^{-\lambda |C|}\mathbb{E}\left[N_C^2 e^{2\lambda |C \setminus \hat{C}|}\right]^{1/2} \\
\leq c_1 \lambda^{-2(d+1)} |C|^{(d-1)/(d+1)} \exp \left(-c_2(\lambda |C|)^{(d-1)/(d+1)}\right) \\
\leq c_1 |C| \exp \left(-c_2(\lambda |C|)^{(d-1)/(d+1)}\right).
$$

The next step of the analysis is to compare the variance of the pseudo-estimator $\hat{\theta}_{\text{pseudo}}$ with the variance of the oracle estimator $\hat{\theta}_{\text{oracle}}$, which is UMVU.

**Theorem 4.4.** The following oracle inequality holds with constants $c, c_1, c_2 > 0$ for all $C \in \mathcal{C}$ with $\lambda |C| \geq 1$:

$$
\text{Var}(\hat{\theta}_{\text{pseudo}}) \leq (1 + c\alpha(\lambda, C)) \text{Var}(\hat{\theta}_{\text{oracle}}) + r(\lambda, C),
$$

where

$$
\alpha(\lambda, C) = \frac{1}{|C|} \left(\frac{1}{\lambda + \frac{\text{Var}(|C \setminus \hat{C}|)}{\mathbb{E}[|C \setminus \hat{C}|]} + \mathbb{E}[|C \setminus \hat{C}|]}\right),
$$

$$
r(\lambda, C) = c_1(\lambda |C|)^{2(d-1)/(d+1)} \exp \left(-c_2(\lambda |C|)^{(d-1)/(d+1)}\right).
$$

**Proof.** By the law of total variance, we obtain

$$
\text{Var}(\hat{\theta}_{\text{pseudo}}) = \text{Var}(\mathbb{E}[\hat{\theta}_{\text{pseudo}}|\hat{C}]) + \mathbb{E}[\text{Var}(\hat{\theta}_{\text{pseudo}}|\hat{C})] \\
= \text{Var}(\hat{\theta}_{\text{oracle}}) + \mathbb{E}\left[\left(\mathcal{N}_{\hat{C}}|\hat{C}|\right)^2 \text{Var}\left(\frac{1}{\mathcal{N}_{\hat{C}}^2 + 1}|\hat{C}|\right)\right].
$$
In view of $N_o \mid \hat{C} \sim \text{Poiss}(\lambda_o)$, a power series expansion gives

\[
\mathbb{E}[(N_o + 1)^{-2} \mid \hat{C}] = \lambda_o^{-1} e^{-\lambda_o} \int_0^{\lambda_o} (e^t - 1) / t \, dt.
\]

The conditional variance can for $\lambda \rightarrow \infty$ thus be bounded by

\[
\text{Var}((1 + N_o)^{-1} \mid \hat{C}) \leq \lambda_o^{-1} e^{-\lambda_o} \int_{\lambda_o/2}^{\lambda_o} e^t / t \, dt - (\lambda_o)^{-2} + O(e^{-\lambda_o/4})
\]

\[
= (\lambda_o)^{-1} \int_0^{\lambda_o/2} e^{-s} \left( \frac{1}{\lambda_o - s} - \frac{1}{\lambda_o} \right) \, ds + O(e^{-\lambda_o/4})
\]

\[
= \lambda_o^{-3}(1 + o(1)),
\]

where we have used $(\lambda_o - s)^{-1} - \lambda_o^{-1} = s \lambda_o^{-1}(\lambda_o - s)^{-1}$, $\int_0^\infty se^{-ds} = 1$ and dominated convergence. Thanks to $(N_o + 1)^{-1} \in [0, 1]$ we conclude for some constant $c \geq 1$

\[
\text{Var}((1 + N_o)^{-1} \mid \hat{C}) \leq c(1 \wedge \lambda_o^{-3}).
\]

Consequently, we have

\[
\text{Var}(\hat{\theta}_{\text{pseudo}}) \leq \text{Var}(\hat{\theta}_{\text{oracle}}) + \mathbb{E}[(N_{\hat{C}} \mid \hat{C})^2 c(1 \wedge (\lambda \mid \hat{C})^{-3})]
\]

\[
= \text{Var}(\hat{\theta}_{\text{oracle}}) + c\mathbb{E}[(N_{\hat{C}} \mid \hat{C})^2 \wedge \lambda^{-3}(N_{\hat{C}})^2 |\hat{C}|^{-1}],
\]

and with (3.3)

\[
\frac{\text{Var}(\hat{\theta}_{\text{pseudo}})}{\text{Var}(\hat{\theta}_{\text{oracle}})} \leq 1 + c \frac{\mathbb{E}[(N_{\hat{C}} \lambda \mid \hat{C})^2 \wedge (N_{\hat{C}})^2 (\lambda \mid \hat{C})^{-1}]}{\lambda \mathbb{E}|C \setminus \hat{C}|}
\]

\[
= 1 + c \frac{\mathbb{E}[(N_{\hat{C}})^2 ((\lambda \mid \hat{C})^2 \wedge (\lambda \mid \hat{C})^{-1})]}{\mathbb{E}|N_{\hat{C}}|}. \quad (4.5)
\]

Define the ‘good’ event $\mathcal{G} = \{ \hat{C} > |C|/2 \}$, on which $( (\lambda \mid \hat{C})^2 \wedge (\lambda \mid \hat{C})^{-1} ) \leq 2(\lambda |C|)^{-1}$. On the complement $\mathcal{G}^c$, we infer from $A^2 \wedge A^{-1} \leq 1$ for $A > 0$

\[
\mathbb{E}[(N_{\hat{C}})^2 ((\lambda \mid \hat{C})^2 \wedge (\lambda \mid \hat{C})^{-1}) 1_{\mathcal{G}^c}] \leq \mathbb{E}[N_{\hat{C}}^2 1_{\mathcal{G}^c}]
\]

\[
\leq \mathbb{E}[N_{\hat{C}}^4]^{1/2} \mathbb{P}(|C \setminus \hat{C}| \geq |C|/2)^{1/2}
\]

\[
\leq c_1 (\lambda |C|)^{2(d-1)/(d+1)} \exp \left( - c_2 (\lambda |C|)^{(d-1)/(d+1)} \right), \quad (4.6)
\]
for some positive constant \( c_1 \) and \( c_2 \), using (4.2) and (4.3). It remains to estimate the upper bound (4.5) on \( G \)

\[
\frac{2c}{\lambda|C|} \mathbb{E}[N_\hat{C}^2] = \frac{2c}{\lambda|C|} \left( \frac{\text{Var}(N_\hat{C})}{\mathbb{E}[N_\hat{C}]} + \mathbb{E}[N_\hat{C}] \right). 
\] (4.7)

Using the identity (17) in Beermann and Reitzner (2015) for the factorial moments for the number of vertices \( N_\hat{C} \), we derive \( \text{Var}(N_\hat{C}) \leq \lambda^2 \text{Var}(|C \setminus \hat{C}|) + \lambda \mathbb{E}[|C \setminus \hat{C}|] \) in view of \( \mathbb{E}[N_\hat{C}] = \lambda \mathbb{E}[|C \setminus \hat{C}|] \). Thus, (4.7) is bounded by

\[
\frac{2c}{\lambda|C|} \mathbb{E}[N_\hat{C}^2] \leq \frac{2c}{|C|} \left( \frac{1}{\lambda} + \frac{\text{Var}(|C \setminus \hat{C}|)}{\mathbb{E}[|C \setminus \hat{C}|]} + \mathbb{E}[|C \setminus \hat{C}|] \right),
\]

which yields the assertion. \( \square \)

As a result, we obtain an oracle inequality for the estimator \( \hat{\vartheta} \).

**Theorem 4.5.** It follows for the risk of the estimator \( \hat{\vartheta} \) for all \( C \in \mathcal{C} \) whenever \( \lambda|C| \geq 1 \):

\[
\mathbb{E}[(\hat{\vartheta} - |C|)^2]^{1/2} \leq (1 + c\alpha(\lambda, C))\mathbb{E}[(\hat{\vartheta}_{\text{oracle}} - |C|)^2]^{1/2} + r(\lambda, C),
\]

with constant \( c > 0 \) and \( \alpha(\lambda, C), r(\lambda, C) \) from Theorem 4.4. For any \( C \in \mathcal{C} \) and \( \lambda > 0 \) we have \( \alpha(\lambda, C) \leq 1 + \frac{1}{\lambda|C|} \).

**Proof.** In view of \( \lambda_c = \lambda|\hat{C}| \), we have \( \hat{\vartheta} = \hat{\vartheta}_{\text{pseudo}} - \lambda^{-1}N_\hat{C}e^{-\lambda|\hat{C}|} \) and we derive as in (4.4) and (4.6) with some constants \( c_1, c_2 > 0 \)

\[
\mathbb{E}[(\hat{\vartheta} - \hat{\vartheta}_{\text{pseudo}})^2] \leq \lambda^{-2}\mathbb{E}[N_\hat{C}^4]^{1/2}\mathbb{E}[e^{-\lambda|\hat{C}|}]^{1/2} \leq c_1^2 \exp \left( -2c_2(\lambda|C|)^{(d-1)/(d+1)} \right).
\]

To establish the oracle inequality, we apply the triangle inequality in \( L^2 \)-norm together with Theorems 3.1 and 4.4.

The universal bound on \( \alpha(\lambda, C) \) follows from the rough bound \( \mathbb{E}[|C \setminus \hat{C}|^2] \leq |C|\mathbb{E}[|C \setminus \hat{C}|] \).

\( \square \)

Note that the remainder term \( r(\lambda, C) \) is exponentially small in \( \lambda|C| \). Therefore, an immediate implication of Theorem 4.5 is that asymptotically our estimator \( \hat{\vartheta} \) is minimax rate-optimal in all dimensions, where the lower bound is proved in the next section. Yet, even more is true: the oracle inequality is in all well studied cases exact in the sense that \( \alpha(\lambda, C) \to 0 \) holds for \( \lambda \to \infty \) such that the the UMVU risk of \( \hat{\vartheta}_{\text{oracle}} \) is attained asymptotically.

13
Lemma 4.6. We have tighter bounds on $\alpha(\lambda, C)$ from Theorem 4.4 in the following cases:

1. for $d = 1, 2$ and $C \in C$ arbitrary: $\alpha(\lambda, C) \lesssim (\lambda|C|)^{-2/(d+1)}$,

2. for $d \geq 2$, $C$ with $C^2$-boundary of positive curvature: $\alpha(\lambda, C) \lesssim (\lambda|C|)^{-2/(d+1)}$,

3. for $d \geq 2$ and $C$ a polytope: $\alpha(\lambda, C) \lesssim \lambda^{-1} (\log(\lambda|C|))^{d-1}$.

Proof. Let us restrict to $|C| = 1$, the case of general volume follows by rescaling. In view of the expectation upper bound (3.5), the main issue is to bound $\operatorname{Var}(\cdot C \setminus \hat{C})/\mathbb{E}[\cdot C \setminus \hat{C}]$ uniformly. Case (1) follows from Pardon (2011), where $\lambda \operatorname{Var}(\cdot C \setminus \hat{C}) \sim \mathbb{E}[\cdot C \setminus \hat{C}]$ is established.

For case (2) with smooth boundary, the upper bound for the variance, $\operatorname{Var}(\cdot C \setminus \hat{C}) \lesssim \lambda^{-(d+3)/(d+1)}$, was obtained in Reitzner (2003), while the lower bound for the first moment, $\mathbb{E}[\cdot C \setminus \hat{C}] \gtrsim \lambda^{-2/(d+1)}$, is due to Schütt (1994).

For the case (3) of polytopes, the upper bound $\operatorname{Var}(\cdot C \setminus \hat{C}) \lesssim \lambda^{-2/(d+1)}$ was obtained in Bárany and Reitzner (2010), while the lower bound for the first moment, $\mathbb{E}[\cdot C \setminus \hat{C}] \gtrsim \lambda^{-1} (\log(\lambda))^{d-1}$, was proved in Bárany and Larman (1988). The expectation upper bound from Remark 3.2 thus yields the result.

We conjecture that $\lambda \operatorname{Var}(\cdot C \setminus \hat{C}) \sim \mathbb{E}[\cdot C \setminus \hat{C}]$ holds universally for all convex sets in arbitrary dimensions and thus that the oracle inequality is always exact. Proving such a universal bound is a challenging open problem in stochastic geometry, strongly connected to the discussion on universal variance asymptotics in terms of the floating body by Bárany and Reitzner (2010).

5 Lower bound and rate optimality

The lower bound in the PPP framework requires a specific treatment although the result for the lower bound in the model with a fixed number of observations is not new, cf. Gayraud (1997).

Theorem 5.1. For estimating $|C|$ in the PPP model with parameter class $C$, the following asymptotic lower bound holds

$$\liminf_{\lambda \to \infty} \lambda^{(d+3)/(d+1)} \inf_{\tilde{\theta}} \sup_{C \in C} \mathbb{E}_C[(|C| - \tilde{\theta})^2] > 0,$$

14
where the infimum extends over all estimators \( \hat{\vartheta} \) in the PPP model with intensity \( \lambda \).

Proof. The steps of the proof follow similar schemes in Gayraud (1997) and Reiß and Selk (2014). First, we reduce the class \( C \) to a properly constructed smaller class of convex sets. Then, we bound the supremum of the mean squared risk by the Bayes risk and carefully bound the Bayes risk at its minimum from below. For simplicity, we consider the case \((\mathbb{E}, \mathcal{E}) = ([0,1]^d, \mathcal{B}_{[0,1]^d})\). Following the scheme, define the class \( \bar{C} \) of closed convex sets

\[
G = \{ X = (x, y) \in [0,1]^d : x \in [0,1]^{d-1}, y \in [0,1], 0 \leq y \leq g(x) \}
\]

for a concave smooth function \( g : \mathbb{R}^{d-1} \to \mathbb{R} \) with support in \([0,1]^{d-1}\) satisfying \( 0 < g(x) < 1, \forall x \in [0,1]^{d-1} \). Then it follows,

\[
\sup_{C \in \mathcal{C}} \lambda^{(d+3)(d+1)} E_C[(|C| - \hat{\vartheta}_\lambda)^2] \geq \sup_{C \in \mathcal{C}} \lambda^{(d+3)(d+1)} E_C[(|C| - \hat{\vartheta}_\lambda)^2].
\]

Choosing \( a > 0 \) sufficiently small, the Hessian

\[
\nabla^2 g(x) = \sum_{k=1}^{M} 4a \xi_k \nabla^2 u(v)|_{v=2(x-q_k)/h_\lambda} - 2I_{d-1},
\]

is negative-definite for all \( x \in [0,1]^{d-1} \) because the function \( \nabla^2 u \) is bounded. This implies that the corresponding sets \( C(\xi) \) with boundaries \( g \) are convex and

\[
\sup_{C \in \mathcal{C}} \lambda^{(d+3)(d+1)} E_C[(|C| - \hat{\vartheta}_\lambda)^2] \geq \lambda^{(d+3)(d+1)} E_{\xi} E_{C(\xi)}[(|C| - \hat{\vartheta}_\lambda)^2] \quad (5.2)
\]
follows. Note that the volume of any convex set $C \in \tilde{\mathcal{C}}$ with the boundary (5.1) can be written as
\[
|C| = \int_{[0,1]^{d-1}} g(x) dx = \sum_{k=1}^{M} \xi_k \int_{[0,1]^{d-1}} g_k(x) dx - \int_{[0,1]^{d-1}} \|x\|^2.
\]
The Bayes-optimal estimator for (5.2) is the conditional expectation $\hat{\vartheta}_B = \mathbb{E}[|C| | (X_1, ..., X_N)] = \int_{[0,1]^{d-1}} \hat{g}(x) dx$ with
\[
\hat{g}(x) = \mathbb{E}[g(x) | (X_1, ..., X_N)] = \sum_{k=1}^{M} \mathbb{E}[\xi_k | (X_1, ..., X_N)] g_k(x) + \|x\|^2.
\]
Using the Bayes formula and the likelihood function (2.1), we derive
\[
\hat{\xi}_k \overset{\text{def}}{=} \mathbb{E}[\xi_k | (X_1, ..., X_N)] = \mathbb{P}(\xi_k = 1 | (X_1, ..., X_N))
= \frac{(d\mathbb{P}_{\mathbb{P}_k}/d\mathbb{P}_0)(d\mathbb{P}_{\mathbb{P}_0}/d\mathbb{P}_0)^{-1}p}{1 - p + (d\mathbb{P}_{\mathbb{P}_k}/d\mathbb{P}_0)(d\mathbb{P}_{\mathbb{P}_0}/d\mathbb{P}_0)^{-1}p}
= \frac{p e^{\lambda I_k \mathbb{P}_k} g_k dx}{1 - p + p e^{\lambda I_k \mathbb{P}_k} g_k dx} 1 \left( \forall X_i^{(1:d-1)} \in I_k : X_i^{(d)} \geq \mathbb{P}_k(X_i^{(1:d-1)}) \right),
\]
where the superscript in $X_i^{(1:d-1)}$ and $X_i^{(d)}$ denotes the corresponding components of the vector $X_i$, $\mathbb{P}_k(x) = g_k(x) - \|x\|^2$, $\mathbb{P}_0(x) = 0$ and $\mathbb{P}_{\mathbb{P}_k}$, $\mathbb{P}_{\mathbb{P}_0}$ stand for the measures associated with the convex sets whose boundaries are $\mathbb{P}_k(x)$ and $\mathbb{P}_0(x)$ in the k-th cube $I_k$. Observing $\hat{\vartheta}_B = \sum_{k=1}^{M} \hat{\xi}_k \int_{[0,1]^{d-1}} g_k(x) dx - \int_{[0,1]^{d-1}} \|x\|^2$ and $\text{Var}(\xi_k | (X_1, ..., X_N)) = \hat{\xi}_k(1 - \hat{\xi}_k)$, the Bayes risk can be calculated as
\[
\mathbb{E}_{\hat{\xi}} \mathbb{E}_{C(\xi)}[(|C| - \hat{\vartheta}_B)^2] = \sum_{k=1}^{M} \text{Var}(\hat{\xi}_k - \xi_k) \left( \int_{I_k} g_k(x) dx \right)^2
= \sum_{k=1}^{M} \mathbb{E} \left[ \text{Var}(\xi_k | (X_1, ..., X_N)) \right] \left( \int_{I_k} g_k(x) dx \right)^2
= \sum_{k=1}^{M} \frac{p(1 - p)}{1 - p + p e^{\lambda I_k \mathbb{P}_k} g_k dx} \left( \int_{I_k} g_k(x) dx \right)^2.
\]
Note, that $\lambda \int_{I_k} g_k(x)dx \leq \lambda \int_{I_k} g_k(x)dx \leq \lambda b_1 h^{d+1}_\lambda \leq b_2$ for some constants $b_1, b_2 > 0$ if we choose $h_\lambda \sim \lambda^{-1/(d+1)}$. This implies the bound for $b_3, b_4, b_5 > 0$

$$E_\xi E_C[|C| - \widehat{\vartheta}_B]^2 \geq Mb_3 (\int_{I_k} g_k(x)dx)^2 \geq Mb_3 h^4_\lambda a^2 (\int_{I_k} u(x - q_k)h_\lambda dx)^2 \geq b_4 h^{d+3}_\lambda = b_5 \lambda^{-(d+3)(d+1)},$$

which completes the proof. \hfill \Box

6 Finite sample behaviour and dilated hull estimator

In this section, we demonstrate the performance of the main estimator $\widehat{\vartheta}$ numerically and compare it to other estimators including the naive estimator $|\hat{C}|$, the naive oracle estimator $N/\lambda$, the UMVU oracle estimator $\widehat{\vartheta}_{oracle}$ and the plug-in MLE estimator $\widehat{\vartheta}_{plugin} = |\hat{C}|(1 + N\hat{C}/N)$. The main competitor from the literature is a rate-optimal estimator proposed in Gayraud (1997). In their construction, the whole sample is divided into three equal parts $X, X'$ and $X''$ of sizes $N^*_\star$ (without loss of generality $N^*_\star \in \mathbb{N}$) and the estimator is given by

$$\widehat{\vartheta}_G = |\hat{C}| + |\hat{C}|_{N^*_\star} \sum_{i=1}^{N^*_\star} 1(X'_i \notin \hat{C}),$$
Figure 3: Monte Carlo RMSE estimates for the studied estimators for the volume of two convex sets: a polygon and an ellipse.

where \( \hat{C}'' \) is the convex hull of the third sample \( X'' \). The data points are simulated for two convex sets: an ellipse and a polygon; see Figure 2.

The RMSE estimate normalised by the area of the true set is based on \( M = 500 \) Monte Carlo iterations in each case. The results of the simulations with normalized intensity \( \lambda = n/|C| \) are depicted in Figure 3. The worst convergence rate of \( N/\lambda \) is clearly visible. More importantly, we see that the RMSE of \( \hat{\vartheta} \) approaches the oracle risk for larger \( n \) (i.e. \( \lambda \)) as the oracle inequality predicts. It is also conspicuous that in the studied cases the plug-in estimator \( \hat{\vartheta}_{\text{plugin}} \) and the estimator \( \hat{\vartheta} \) perform rather similarly. This is explained by the fact that the number of points \( N_{\hat{C}} \) on the convex hull increases with a moderate speed in the two-dimensional case, \( \mathbb{E}[N_{\hat{C}}] = O(\lambda^{1/3}) \), which results in a small difference between the multiplication factors \( N_{\hat{C}}/N \) and \( N_{\hat{C}}/(N_0 + 1) \).

The empirical bias of the proposed estimator \( \hat{\vartheta} \) is always of smaller order than \( 10^{-3} \). In fact, the case \( N_0 = 0 \) almost never appeared during the simulations such that \( \hat{\vartheta} = \hat{\vartheta}_{\text{pseudo}} \) in almost all cases. The other estimators exhibited a significant bias.

As an application of the obtained results, we propose a new estimator for
Figure 4: Monte Carlo error ratio for the convex hull and its dilation when the true set is a polygon.

a convex set itself:

\[ \tilde{C} \overset{\text{def}}{=} \left\{ \hat{x}_0 + \frac{\hat{\theta}}{|C|} (x - \hat{x}_0) \left| x \in \hat{C} \right\} = \left\{ \hat{x}_0 + \frac{N + 1}{N_0 + 1} (x - \hat{x}_0) \left| x \in \hat{C} \right\}, \]

which is just the dilation of the convex hull \( \hat{C} \) from its centre of gravity \( \hat{x}_0 \) (the convex hull is enlarged by the factor \( (N + 1)/(N_0 + 1) \) around its centre of gravity), see the dashed polygons in Figure 2. The alternative idea, to dilate the set as seen from the sample mean of the points instead of the centre of gravity does not improve the estimator. Since the convex hull is a sufficient statistic (for known \( \lambda \)), the points in its interior do not bear any information on the shape of \( C \) itself.

In a small simulation study, we investigate the behaviour of the new estimator for the polygon. The error ratio \( \mathbb{E}[|C \Delta \tilde{C}|]/\mathbb{E}[|C \Delta \hat{C}|] \) in terms of the symmetric difference \( A \Delta B = A \setminus B \cup B \setminus A \) is approximated in \( M = 500 \) Monte Carlo iterations and shown in Figure 4. It turns out, the dilation significantly improves the convex hull as an estimator for \( C \). A profound analysis of this estimator and its asymptotic properties is beyond the present scope, but seems very desirable.
Proof of Lemma 4.1. Using that the bias of the oracle estimator $\hat{\theta} = |\hat{C}| + N_{\hat{C}}/(N_o + 1)|\hat{C}|$ is exponentially small, it remains to compare its expectation with the expectation of the plug-in estimator $\hat{\theta}_{\text{plugin}}$ to show (4.1):

$$E[\hat{\theta} - \hat{\theta}_{\text{plugin}}] = E \left[ |\hat{C}| \left( \frac{N_{\hat{C}}}{N_o + 1} - \frac{N\hat{C}}{N} \right) \right] = E \left[ |\hat{C}| \left( \frac{N^2_{\hat{C}} - N\hat{C}}{(N_o + 1)(N_o + N_{\hat{C}})} \right) \right]$$

$$\geq \frac{d}{d+1} E \left[ \frac{|\hat{C}| N^2_{\hat{C}}}{(N_o + 1)2\lambda|C|} \mathbf{1}(N \leq 2\lambda|C|) \right],$$

where in the last line we have used $|\hat{C}| > 0$ only if $N_{\hat{C}} \geq d + 1$ and in this case $N^2_{\hat{C}} - N\hat{C} \geq \frac{d}{d+1} N^2_{\hat{C}}$. Using $E[(N_o + 1)^{-1} |\hat{C}|] = \lambda_o^{-1}(1 - e^{-\lambda_o})$ from above, we obtain after writing $1(N \leq 2\lambda|C|) = 1 - 1(N > 2\lambda|C|)$

$$E[\hat{\theta} - \hat{\theta}_{\text{plugin}}] \geq \frac{d}{d+1} \left( E \left[ \frac{N^2_{\hat{C}} |\hat{C}| (1 - e^{-\lambda_o})}{2\lambda_o |\hat{C}|} \right] - E \left[ \frac{N^2_{\hat{C}} |\hat{C}|}{2\lambda |C|} \mathbf{1}(N > 2\lambda|C|) \right] \right)$$

$$\geq \frac{d}{d+1} \left( \frac{E[N^2_{\hat{C}} (1 - e^{-\lambda_o})]}{2\lambda^2 |C|} - \frac{E[N^2_{\hat{C}} \mathbf{1}(N > 2\lambda|C|)]}{2\lambda} \right).$$

By Cauchy-Schwarz inequality and large deviations similarly to (4.4), the first term is bounded from below by a constant multiple of $E[|C \setminus \hat{C}|^2]/|C|$ in view of $E[N^2_{\hat{C}}] \geq \lambda^2 E[|C \setminus \hat{C}|^2]$, see e.g. Section 2.3.2 in Brunel (2014). Because of $N \sim \text{Poiss}(\lambda|C|)$, the second term is of order $\lambda |C|^2 e^{-\lambda|C|}$ and thus asymptotically of much smaller order.

\[ \square \]

References

Bárány, I., and D. Larman (1988): “Convex bodies, economic cap coverings, random polytopes,” Mathematika, 35(2), 274–291.

Bárány, I., and M. Reitzner (2010): “On the variance of random polytopes,” Advances in Mathematics, 225(4), 1986–2001.
Beermann, M., and M. Reitzner (2015): “Beyond the Efron-Buchta identities: Distributional Results for Poisson Polytopes,” *Discrete & Computational Geometry*, 53(1), 226–244.

Brunel, V.-E. (2013): “A universal deviation inequality for random polytopes,” arXiv:1311.2902.

——— (2014): “Non-parametric estimation of convex bodies and convex polytopes,” Theses, Université Pierre et Marie Curie - Paris VI; University of Haifa, https://tel.archives-ouvertes.fr/tel-01066977.

Dwyer, R. (1988): “On the convex hull of random points in a polytope,” *Journal of Applied Probability*, 25(4), 688–699.

Gayraud, G. (1997): “Estimation of functionals of density support,” *Mathematical Methods of Statistics*, 6(1), 26–46.

Guntuboyina, A. (2012): “Optimal rates of convergence for convex set estimation from support functions,” *The Annals of Statistics*, 40(1), 385–411.

Jarvis, R. (1973): “On the identification of the convex hull of a finite set of points in the plane,” *Information Processing Letters*, 2(1), 18–21.

Karr, A. (1991): *Point processes and their statistical inference*, Probability: Pure and Applied. CRC press.

Korostelev, A., and A. Tsybakov (1993): *Minimax theory of image reconstruction*. Springer.

——— (1994): “Asymptotic efficiency in estimation of a convex set,” *Problemy Peredachi Informatsii*, 30(4), 33–44.

Kutoyants, Y. (1998): *Statistical Inference for Spatial Poisson Processes*, Lecture Notes in Statistics. Springer.

Mammen, E., and A. Tsybakov (1995): “Asymptotical Minimax Recovery of Sets with Smooth Boundaries,” *The Annals of Statistics*, 23(2), 502–524.

Meister, A., and M. Reiss (2013): “Asymptotic equivalence for nonparametric regression with non-regular errors,” *Probability Theory and Related Fields*, 155(1-2), 201–229.
Molchanov, I. (2006): *Theory of random sets*. Springer.

Pardon, J. (2011): “Central limit theorems for random polygons in an arbitrary convex set,” *The Annals of Probability*, 39(3), 881–903.

Reiß, M., and L. Selk (2014): “Estimating nonparametric functionals efficiently under one-sided errors,” arXiv:1407.4229.

Reitzner, M. (2003): “Random polytopes and the Efron–Stein jackknife inequality,” *The Annals of Probability*, 31(4), 2136–2166.

Schütt, C. (1994): “Random polytopes and affine surface area,” *Mathematische Nachrichten*, 170(1), 227–249.