Estimating the volume of a convex set
Salim Rao

To cite this version:
Salim Rao. Estimating the volume of a convex set. 2015. <hal-01144492>

HAL Id: hal-01144492
https://hal.archives-ouvertes.fr/hal-01144492
Submitted on 21 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Estimating the volume of a convex set

Salim Rao
Bengal Engineering and Science University, Shibpur

Abstract
A new estimator for the volume of the convex set in $\mathbb{R}^d$ is proposed. Our approach is based on a Poisson point process model. We also prove that the convex hull is a sufficient and complete statistic.

1 Introduction
Estimating the support of a density or a function is a central statistical question, for instance in image analysis. 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 [63, 62] who prove minimax-optimal rates of convergence in Hausdorff distance for a set estimator. In particular, [63] 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 based on three-fold sample splitting.

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 (P3) observation model with intensity $\lambda > 0$ on the convex set $C \subseteq \mathbb{R}^d$. We thus observe

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

(1.1)
where \(X_1, \ldots, X_N, N\) are independent, see Section 2 below for a concise introduction to the P3 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 P3 model, however, allows for much more concise ideas and proofs. From an applied perspective, P3 models are often natural, e.g. for spatial count data of photons or other emissions.

For known intensity \(\lambda\) of the P3, we construct in Section 3 an oracle estimator \(\hat{\vartheta}_{\text{oracle}}\). Then, Theorem 1 shows that this estimator is UMVAUE (uniformly of minimum variance among unbiased estimators) and rate-optimal. To this end, moment bounds from stochastic geometry for the missing volume of \(\hat{C}\) 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 \(\hat{\vartheta}_{\text{oracle}}\).

The lower bound showing that \(\hat{\vartheta}\) is indeed minimax-optimal is proved in Section 5 by constructing an asymptotically least-favourable Bayesian prior. The proof of Lemma 4 is deferred to the appendix.

2 Theoretical background

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 \(\mathcal{C} = \{C \subseteq E, \text{convex, closed}\}\) (this implies that all sets in \(\mathcal{C}\) are compact). It is natural to equip the space \(\mathcal{C}\) with the Hausdorff-metric \(d_H\) and its Borel \(\sigma\)-algebra \(\mathfrak{B}_C\). Then \((\mathcal{C}, d_H)\) 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 \((\mathcal{C}, d_H)\).

On \((E, \mathcal{E})\) we define the set of point measures \(\mathcal{M} = \{m\text{ measure on }\mathcal{E} : m(A) \in \mathbb{N}, \forall A \in \mathcal{E}\}\) 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 \(C \in \mathcal{C}\). We call a measurable mapping \(N : \Omega \to \mathcal{M}\) a Poisson point process (P3) 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 P3 \( \mathcal{N} \) is

\[
\mathcal{N} = \sum_{i=1}^{N} \delta_{X_i}
\]

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

We shall consider the convex hull of the P3 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 \((C, d_H)\). In the sequel, conditional expectations and probabilities with respect to \( \hat{C} \) are thus well defined. We can also evaluate the probability

\[
\mathbb{P}(C \in A) = \sum_{k=0}^{\infty} \frac{e^{-\lambda |C|/\lambda_0}}{k!} \int_{C^k} 1(\text{conv}\{x_1, \ldots, x_k\} \in A) d(x_1, \ldots, x_k)
\]

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

\[
\frac{d\mathbb{P}_{C, \lambda}}{d\mathbb{P}_{E, \lambda_0}}(X_1, \ldots, X_N) = \frac{d\mathbb{P}_{E, \lambda}}{d\mathbb{P}_{E, \lambda_0}}(X_1, \ldots, X_N) \left( \frac{d\mathbb{P}_{C, \lambda}}{d\mathbb{P}_{E, \lambda}}(X_1, \ldots, X_N) \right)
\]

\[
= e^{\lambda_0 |E|/\lambda_0} |\lambda/\lambda_0|^N 1(\forall i = 1, \ldots, N : X_i \in C) \quad (2.1)
\]

\[
= e^{\lambda_0 |E|/\lambda_0} |\lambda/\lambda_0|^N 1(\hat{C} \subseteq C) \quad (2.2)
\]

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

### 3 Known intensity

For a P3 on \( C \in \mathbb{C} \) 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 \( \lambda^{-1/2} \):

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

3
Another natural idea might be to use the plug-in estimator $|\widehat{C}|$ whose error is given by the missing volume and satisfies

$$\mathbb{E}[(|\widehat{C}| - |C|)^2] = \mathbb{E}[|C \setminus \widehat{C}|^2] = O(|C|^{2(d-1)/(d+1)} \lambda^{-4/(d+1)}), \quad (3.2)$$

where the bound is obtained similarly to (3.8) and (3.10) 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 (2.3) 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 $\widehat{C}$ 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 $N_{\widehat{C}}$, which is measurable with respect to the convex hull $\widehat{C}$, 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, $N_0$ is, conditionally on the convex hull $\widehat{C}$, Poisson-distributed:

$$N_0 | \widehat{C} \sim \text{Poiss}(\lambda_0) \text{ with } \lambda_0 \overset{\text{def}}{=} \lambda |\widehat{C}|. \quad (3.3)$$

The law can also be derived directly from the likelihood representation (2.3). Thus, we obtain the oracle estimator

$$\hat{\vartheta}_{\text{oracle}} \overset{\text{def}}{=} \mathbb{E} \left[ \frac{N}{\lambda} | \widehat{C} \right] = \mathbb{E} \left[ \frac{N_0 + N_{\widehat{C}}}{\lambda} | \widehat{C} \right] = |\widehat{C}| + \frac{N_{\widehat{C}}}{\lambda}. \quad (3.4)$$

**Theorem 1.** For known intensity $\lambda > 0$, the oracle estimator $\hat{\vartheta}_{\text{oracle}}$ is unbiased and of minimal variance among all unbiased estimators (UMVAUE). It satisfies

$$\text{Var}(\hat{\vartheta}_{\text{oracle}}) = \frac{1}{\lambda} \mathbb{E}[|C \setminus \widehat{C}|].$$

Its worst case mean squared error over $C$ decays as $\lambda \uparrow \infty$ like $\lambda^{-(d+3)/(d+1)}$ in dimension $d$:

$$\limsup_{\lambda \to \infty} \lambda^{(d+3)/(d+1)} \sup_{C \in \mathcal{C}} \left\{ |C|^{-(d-1)/(d+1)} \mathbb{E}[|(\hat{\vartheta}_{\text{oracle}} - |C|)^2]| \right\} < \infty. \quad (3.5)$$

**Remark 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 9 below, we prove that the lower bound on the minimax risk in
the P3 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: $\mathbb{E}[|C \setminus \hat{C}|] \sim \lambda^{-1}(\log(\lambda|C|))^{d-1}$ 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 (3.4). By the law of total variance, we obtain

$$\text{Var}(\hat{\vartheta}_{\text{oracle}}) = \text{Var}\left(\frac{N}{\lambda}\right) - \mathbb{E}\left[\text{Var}\left(\frac{N}{\lambda} \mid \hat{C}\right)\right] = \frac{|C|}{\lambda} - \mathbb{E}\left[\text{Var}\left(\frac{N_0}{\lambda} \mid \hat{C}\right)\right].$$

Proposition 3 below affirms that the convex hull $\hat{C}$ is not only a sufficient, but also a complete statistic such that by the Lehmann-Scheffe theorem, the estimator $\hat{\vartheta}_{\text{oracle}}$ has the UMVAUE 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 P3 model conditionally on the event $\{N = k\}$ is distributed as the convex hull $\hat{C}_k = \text{conv}\{X_1, ..., X_k\}$ in the model with $k$ uniform observations on $C$, for which the following upper bound is known:

$$\sup_{C \in \mathcal{C}} \mathbb{E}\left[\frac{|C \setminus \hat{C}_k|}{|C|}\right] = O\left(k^{-2/(d+1)}\right).$$

Thus, it follows by a Poisson moment bound

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

This bound, together with (3.7), yields the assertion. \square

Proposition 3. For known intensity $\lambda > 0$, the convex hull $\hat{C} = \text{conv}\{X_1, ..., X_N\}$ is a complete statistic.

Proof. We need to show the implication

$$\forall C \in \mathcal{C} : \mathbb{E}_C[T(\hat{C})] = 0 \implies T(\hat{C}) = 0 \quad \mathbb{P}_E - a.s. \quad (3.11)$$

for any $\mathcal{B}_C$-measurable function $T : \mathcal{C} \to \mathbb{R}$. From the likelihood in (2.3) for $\lambda = \lambda_0$, we derive

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

5
Since \( \exp(\lambda |E \setminus C|) \) is deterministic, \( \mathbb{E}_C[T(\hat{C})] = 0 \) for all \( C \in \mathcal{C} \) implies

\[
\forall C \in \mathcal{C} : \mathbb{E}_E[T(\hat{C})1(\hat{C} \subseteq C)] = 0. \tag{3.13}
\]

For \( C \in \mathcal{C} \), define the family of convex subsets of \( C \) as \( |C| = \{ A \in \mathcal{C} | A \subseteq C \} \) 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 \( \mu^\pm(B) = \mathbb{E}_E[T^\pm(\hat{C})1(\hat{C} \in B)], B \in \mathcal{B}_C \), agree on \( \{|C| \} \). Note that the brackets \( \{|C| \} \) 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 \( \{|C| \} \) 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 \( \mathcal{B}_C \). Let us define the family \( \langle C \rangle = \{ B \in \mathcal{C} | C \subseteq B \} \) of convex sets containing \( C \). Then the closed Hausdorff ball with center \( C \) and radius \( \varepsilon > 0 \) has the representation

\[
B_\varepsilon(C) = \{ A \in \mathcal{C} | d_H(A, C) \leq \varepsilon \} = \{ A \in \mathcal{C} | U_\varepsilon(C) \subseteq A \subseteq U_{-\varepsilon}(C) \}, \tag{3.14}
\]

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

\[
B_\varepsilon(C) = \langle U_{-\varepsilon}(C) \rangle \cap \langle U_\varepsilon(C) \rangle. \tag{3.15}
\]

Since \( (\mathcal{C}, d_H) \) is separable, our problem is reduced to proving that all angle sets \( \langle C \rangle \) for \( C \in \mathcal{C} \) are in \( \mathcal{C} \). A further reduction is achieved by noting \( \langle C \rangle = \bigcap_{x \in C} \langle x \rangle = \bigcap_{x \in \mathcal{C} \cap \mathbb{Q}^d} \langle x \rangle \) setting \( \langle x \rangle = \{ \{ x \} \} \) for short such that it suffices to prove \( \langle x \rangle \in \mathcal{C} \) for all \( x \in E \). Now, \( x \notin \mathcal{C} \) implies by the Hahn-Banach theorem that there are \( \delta > 0, v \in \mathbb{R}^d \) such that \( \langle v, c - x \rangle \geq \delta \) holds for all \( c \in \mathcal{C} \). By a density argument, we may choose \( \delta \in \mathbb{Q}^+ \) and \( v \in \mathbb{Q}^d \).

Denoting the corresponding hyperplane intersected with \( E \) by \( H_{\delta,v} = \{ \xi \in E | \langle v, \xi - x \rangle \geq \delta \} \), we conclude

\[
\langle x \rangle^\mathcal{C} = \bigcup_{\delta \in \mathbb{Q}^+} \bigcup_{v \in \mathbb{Q}^d} \bigcup_{c \in \mathcal{C}} [H_{\delta,v}] \in \mathcal{C}. \tag{3.16}
\]

Consequently, \( \langle x \rangle \in \mathcal{C} \) and thus \( \mathcal{B}_C \subseteq \mathcal{C} \) hold. \( \square \)
4 Unknown intensity

In the case, when the intensity $\lambda$ is unknown and the oracle estimator $\hat{\vartheta}_{\text{oracle}}$ in (3.4) is inaccessible, the maximum-likelihood approach suggests to use $N/|\hat{C}|$ as an estimator for $\lambda$ in (2.3). This yields the plug-in estimator for the volume,

$$\hat{\vartheta}_{\text{plugin}} \overset{\text{def}}{=} |\hat{C}| + \frac{N_{\hat{C}}}{N} |\hat{C}|. \quad (4.1)$$

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

$$|C| - E[\hat{\vartheta}_{\text{plugin}}] \geq cE[|\hat{C}\setminus C|^2], \quad \forall C \in \mathcal{C}. \quad (4.2)$$

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 $d > 5$. 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 $N_0 |\hat{C} \sim \text{Poisson}(\lambda_0)$, see (3.3).

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

$$\sum_{k=0}^{\infty} \frac{\lambda_0^k}{k!} \tilde{\mu}(k) e^{-\lambda_0} = \lambda_0^{-1} \Rightarrow \sum_{k=0}^{\infty} \frac{\lambda_0^{k+1}}{k!} \tilde{\mu}(k) = \sum_{k=0}^{\infty} \frac{\lambda_0^k}{k!}. \quad (4.3)$$

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 $\text{Exp}(\lambda_0)$-distributed and thus has expectation $\lambda_0^{-1}$. Taking conditional expectation with respect to the Poisson process at time 1, we conclude that

$$\tilde{\mu}(N_0, \lambda_0) \overset{\text{def}}{=} \begin{cases} (N_0 + 1)^{-1}, & \text{for } N_0 \geq 1, \\ 1 + \lambda_0^{-1}, & \text{for } N_0 = 0 \end{cases}$$
satisfies $\mathbb{E}[\hat{\mu}(N_0, \lambda_0)|\hat{C}] = \lambda_0^{-1}$. Omitting the term depending on $\lambda_0$ in the unlikely case $N_0 = 0$, we define our final estimator

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

For the proofs, we also define the pseudo-estimator

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

**Theorem 5.** The pseudo-estimator $\hat{\vartheta}_{\text{pseudo}}$ is unbiased and the estimator $\hat{\vartheta}$ is asymptotically unbiased in the sense that with constants $c_1, c_2 > 0$ whenever $\lambda|C| \geq 1$:

$$0 \leq |C| - \mathbb{E}[\hat{\vartheta}] \leq c_1|C|\exp(-c_2(\lambda|C|)^{(d-1)/(d+1)}), \quad \forall C \in \mathcal{C}. \quad (4.4)$$

**Proof.** We have

$$\mathbb{E}\left[\frac{1}{N_0 + 1} + \frac{e^{-\lambda_0}}{\lambda_0} |\hat{C}|\right] = e^{-\lambda_0}\lambda_0^{-1}\left(\sum_{k=0}^{\infty} \frac{\lambda_0^{k+1}}{(k+1)k!} + 1\right) = \lambda_0^{-1}, \quad (4.5)$$

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

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

We exploit the deviation inequality that can be written in our setting, using the conditioning argument similarly to (3.10), as and derive the bound for the exponential moment of the missing volume in the model with fixed number of points

$$\mathbb{E}[\exp(\lambda|C\setminus\hat{C}_k|)] \leq b_1 \exp\left(b_2\lambda|C|k^{-2/(d+1)}\right), \quad k \geq 2, \quad (4.6)$$

for positive constants $b_1, b_2$ depending on the dimension. For the cases $k = 0, 1$, we have the identity $\mathbb{E}[\exp(\lambda|C\setminus\hat{C}_k|)] = \exp(\lambda|C|)$. By Poissonisation, similarly to (3.10), we derive

$$\exp(-\lambda|C|)\mathbb{E}[\exp(\lambda|C\setminus\hat{C}|)] \leq b_3 \exp\left(-c_2(\lambda|C|)^{(d-1)/(d+1)}\right), \quad (4.7)$$

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,

$$\mathbb{E}[N_0^q] = O((\lambda|C|)^{(d-1)/(d+1)})', \quad q \in \mathbb{N}. \quad (4.8)$$
we derive for a constant $c_1 > 0$

$$
\lambda^{-1} \mathbb{E} \left[ N_C e^{-\lambda |\hat{C}|} \right] 
\leq \lambda^{-1} e^{-\lambda |C|} \mathbb{E} \left[ N_C^2 \right]^{1/2} \mathbb{E} \left[ e^{2\lambda |\hat{C}|} \right]^{1/2}
\leq c_1 \lambda^{2/(d+1)} |C|^{(d-1)/(d+1)} \exp \left( -c_2 (\lambda |C|)^{(d-1)/(d+1)} \right)
\leq c_1 |C| \exp \left( -c_2 (\lambda |C|)^{(d-1)/(d+1)} \right),
(4.9)
$$

$$
\lambda^{-1} \mathbb{E} \left[ N_C e^{-\lambda |\hat{C}|} \right] 
\leq \lambda^{-1} e^{-\lambda |C|} \mathbb{E} \left[ N_C^2 \right]^{1/2} \mathbb{E} \left[ e^{2\lambda |\hat{C}|} \right]^{1/2}
\leq c_1 \lambda^{2/(d+1)} |C|^{(d-1)/(d+1)} \exp \left( -c_2 (\lambda |C|)^{(d-1)/(d+1)} \right)
\leq c_1 |C| \exp \left( -c_2 (\lambda |C|)^{(d-1)/(d+1)} \right),
(4.10)
$$

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

**Theorem 6.** The following oracle inequality holds with constants $c, c_1, c_2 > 0$ for all $C \in C$ with $\lambda |C| \geq 1$:

$$
\text{Var}(\hat{\vartheta}_{\text{pseudo}}) \leq (1 + c \alpha(\lambda, C)) \text{Var}(\hat{\vartheta}_{\text{oracle}}) + r(\lambda, C),
(4.12)
$$

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

$$
\text{Var}(\hat{\vartheta}_{\text{pseudo}}) = \text{Var}(\mathbb{E}[\hat{\vartheta}_{\text{pseudo}}|\hat{C}]) + \mathbb{E} \left[ \text{Var}(\hat{\vartheta}_{\text{pseudo}}|\hat{C}) \right]
(4.13)
$$

$$
= \text{Var}(\hat{\vartheta}_{\text{oracle}}) + \mathbb{E} \left[ (N_C|\hat{C}|)^2 \text{Var} \left( \frac{1}{N_C+1} |\hat{C}| \right) \right].
(4.14)
$$

In view of $N_0 |\hat{C} \sim \text{Poiss}(\lambda_0)$, a power series expansion gives

$$
\mathbb{E}[(N_0 + 1)^{-2} |\hat{C}] = \lambda_0^{-1} e^{-\lambda_0} \int_0^{\lambda_0} (e^t - 1)/t dt.
$$

The conditional variance can for $\lambda_0 \to \infty$ thus be bounded by

$$
\text{Var}((1 + N_0)^{-1} |\hat{C}) \leq \lambda_0^{-1} e^{-\lambda_0} \int_0^{\lambda_0} e^t/t dt - (\lambda_0)^{-2} + O(e^{-\lambda_0/4})
= (\lambda_0)^{-1} \int_0^{\lambda_0/2} e^{-s} (\frac{1}{\lambda_0 - s} - \frac{1}{\lambda_0}) ds + O(e^{-\lambda_0/4})
= \lambda_0^{-3}(1 + o(1)),
$$
where we have used \((\lambda_0 - s)^{-1} - \lambda_0^{-1} = s\lambda_0^{-1}(\lambda_0 - s)^{-1}\), \(\int_0^\infty se^{-s}ds = 1\) and dominated convergence. Thanks to \((N_0 + 1)^{-1} \in [0, 1]\) we conclude for some constant \(c \geq 1\)

\[
\Var((1 + N_0)^{-1} | \hat{C}) \leq c(1 \wedge \lambda_0^{-3}).
\]

Consequently, we have

\[
\Var(\hat{\varpi}_{\text{pseudo}}) \leq \Var(\hat{\varpi}_{\text{oracle}}) + \mathbb{E}\left[ (N_{\hat{C}}|\hat{C})^2 c(1 \wedge (\lambda|\hat{C})^{-3}) \right] \tag{4.15}
\]

\[
= \Var(\hat{\varpi}_{\text{oracle}}) + c\mathbb{E}\left[ (N_{\hat{C}}|\hat{C})^2 \wedge \lambda^{-3}(N_{\hat{C}})^2|\hat{C}^{-1} \right] \tag{4.16}
\]

and with \((3.7)\)

\[
\frac{\Var(\hat{\varpi}_{\text{pseudo}})}{\Var(\hat{\varpi}_{\text{oracle}})} \leq 1 + \frac{c\mathbb{E}\left[ (N_{\hat{C}}|\hat{C})^2 \wedge (N_{\hat{C}})^2(\lambda|\hat{C})^{-1} \right]}{\lambda \mathbb{E}[|C| \setminus \hat{C}]} \tag{4.17}
\]

\[
= 1 + \frac{c\mathbb{E}\left[ (N_{\hat{C}})^2 (\lambda|\hat{C})^2 \wedge (\lambda|\hat{C})^{-1}) \right]}{\mathbb{E}[N_{\hat{C}}]} \tag{4.18}
\]

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

\[
\mathbb{E}\left[ (N_{\hat{C}})^2 (\lambda|\hat{C})^2 \wedge (\lambda|\hat{C})^{-1}) \right] \leq \mathbb{E}\left[ (N_{\hat{C}})^2 1_{\mathcal{G}^c} \right] \tag{4.19}
\]

\[
\leq \mathbb{E}[N_{\hat{C}}^{1/2} P(|C \setminus \hat{C}| \geq |C|/2)^{1/2} \tag{4.20}
\]

\[
\leq c_1 (\lambda|C|)^{(d-1)/(d+1)} \exp \left( - c_2 (\lambda|C|)^{(d-1)/(d+1)} \right), \tag{4.21}
\]

for some positive constant \(c_1\) and \(c_2\), using \((4.7)\) and \((4.8)\). It remains to estimate the upper bound \((4.18)\) on \(\mathcal{G}\)

\[
\frac{2c}{\lambda|C|} \frac{\mathbb{E}[N_{\hat{C}}^2]}{\mathbb{E}[N_{\hat{C}}]} = \frac{2c}{\lambda|C|} \left( \frac{\Var(N_{\hat{C}})}{\mathbb{E}[N_{\hat{C}}]} + \mathbb{E}[N_{\hat{C}}] \right). \tag{4.22}
\]

Using the identity for the factorial moments for the number of vertices \(N_{\hat{C}}\), we derive \(\Var(N_{\hat{C}}) \leq \lambda^2 \Var(|C \setminus \hat{C}|) + \lambda \mathbb{E}[|C \setminus \hat{C}|]\) in view of \(\mathbb{E}[N_{\hat{C}}] = \lambda \mathbb{E}[|C \setminus \hat{C}|]\). Thus, \((4.22)\) is bounded by

\[
\frac{2c}{\lambda|C|} \frac{\mathbb{E}[N_{\hat{C}}^2]}{\mathbb{E}[N_{\hat{C}}]} \leq \frac{2c}{|C|} \left( \frac{1}{\lambda} + \frac{\Var(|C \setminus \hat{C}|)}{\mathbb{E}[|C \setminus \hat{C}|]} + \mathbb{E}[|C \setminus \hat{C}|] \right), \tag{4.23}
\]

which yields the assertion. \(\square\)

As a result, we obtain an oracle inequality for the estimator \(\hat{\varpi}\).
Theorem 7. It follows for the risk of the estimator $\hat{\vartheta}$ for all $C \in C$ 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 6. For any $C \in C$ and $\lambda > 0$ we have $\alpha(\lambda, C) \leq 1 + \frac{1}{\lambda |C|}$.

Proof. In view of $\lambda = \lambda |\hat{C}|$, we have $\hat{\vartheta} = \hat{\vartheta}_{\text{pseudo}} - \lambda^{-1} N_{\hat{C}} e^{-\lambda |\hat{C}|}$ and we derive as in (4.11) and (4.21) with some constants $c_1, c_2 > 0$

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

To establish the oracle inequality, we apply the triangle inequality in $L^2$-norm together with Theorems 1 and 6.

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 7 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 UMVAUE risk of $\hat{\vartheta}_{\text{oracle}}$ is attained asymptotically.

Lemma 8. We have tighter bounds on $\alpha(\lambda, C)$ from Theorem 7 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 (3.10), the main issue is to bound $\text{Var}(|C \setminus \hat{C}|)/\mathbb{E}[|C \setminus \hat{C}|]$ uniformly. Case (1) follows from $\lambda \text{Var}(|C \setminus \hat{C}|) \sim \mathbb{E}[|C \setminus \hat{C}|]$.

For case (2) with smooth boundary, the upper bound for the variance, $\text{Var}(|C \setminus \hat{C}|) \lesssim \lambda^{-(d+3)/(d+1)}$, while the lower bound for the first moment, $\mathbb{E}[|C \setminus \hat{C}|] \gtrsim \lambda^{-2/(d+1)}$. 11
For the case (3) of polytopes, the upper bound \( \text{Var}(|C \setminus \hat{C}|) \lesssim \lambda^{-2}(\log \lambda)^{d-1} \), while the lower bound for the first moment, \( \mathbb{E}[|C \setminus \hat{C}|] \gtrsim \lambda^{-1}(\log \lambda)^{d-1} \). The expectation upper bound from Remark 2 thus yields the result.

## 5 Lower bound and rate optimality

The lower bound in the P3 framework requires a specific treatment although the result for the lower bound in the model with a fixed number of observations is not new.

**Theorem 9.** For estimating \(|C|\) in the P3 model with parameter class \( C \), the following asymptotic lower bound holds

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

where the infimum extends over all estimators \( \hat{\vartheta}_\lambda \) in the P3 model with intensity \( \lambda \).

**Proof.** 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 \((\mathcal{E}, \mathcal{F}) = ([0, 1]^d, \mathcal{B}_{[0,1]^d})\). Following the scheme, define the class \( \tilde{C} \) 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,

\[
\sup_{C \in C} \lambda^{(d+3)/(d+1)} \mathbb{E}_C[(|C| - \hat{\vartheta}_\lambda)^2] \gtrsim \sup_{C \in C} \lambda^{(d+3)/(d+1)} \mathbb{E}_C[(|C| - \hat{\vartheta}_\lambda)^2].
\]

To bound the mean squared risk by the Bayes risk, consider the partition of \([0, 1]^{d-1}\) into cubes with sizes of length \( h_\lambda \), denote by \( g_k \) the center of the \( k \)-th cube \( I_k \) and let \( M = (1/h_\lambda)^{d-1} \in \mathbb{N} \) be the number of cubes. Then, for the prior on \( \tilde{C} \), define the random variables \( \xi_k \overset{i.i.d.}{\sim} \text{Bernoulli}(p) \), i.e. \( \mathbb{P}(\xi_k = 0) = p, \mathbb{P}(\xi_k = 1) = 1 - p \) and set

\[
g(x) = \sum_{k=1}^{M} \xi_k g_k(x) - \|x\|^2, \quad g_k(x) = ah_\lambda^2 u\left(\frac{x - q_k}{h_\lambda/2}\right),
\]

where \( u \) is a concave function with support in \([0, 1]^{d-1}\). The Bayes risk can be carefully bounded from below to yield the claimed lower bound.
where \( a > 0, \ u : \mathbb{R}^{d-1} \to \mathbb{R}_+ \) is smooth and has support in \([-1,1]^{d-1}\).

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

\[
\nabla^2 g(x) = \sum_{k=1}^{M} 4a_ξ k \nabla^2 u(v)|_{v=2(x-q_k)/h_λ} - 2I_{d-1} , \tag{5.5}
\]

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 \( C(ξ) \) with boundaries \( g \) are convex and

\[
\sup_{C \in \tilde{C}} \lambda^{(d+3)(d+1)} \mathbb{E}_C[(|C| - \bar{J}_λ)^2] \geq \lambda^{(d+3)(d+1)} \mathbb{E}_{\tilde{C}(ξ)}[(|C| - \bar{J}_λ)^2] \tag{5.6}
\]

follows. Note that the volume of any convex set \( C \in \tilde{C} \) with the boundary \( 5.4 \) can be written as

\[
|C| = \int_{[0,1]^{d-1}} g(x) dx = \sum_{k=1}^{M} ξ_k \int_{[0,1]^{d-1}} g_k(x) dx - \int_{[0,1]^{d-1}} \|x\|^2 . \tag{5.7}
\]

The Bayes-optimal estimator for \( 5.6 \) is the conditional expectation \( \hat{J}_B = \mathbb{E}[|C| \mid (X_1,\ldots,X_N)] = \int_{[0,1]^{d-1}} \tilde{g}(x) dx \) with

\[
\tilde{g}(x) = \mathbb{E}[g(x) \mid (X_1,\ldots,X_N)] = \sum_{k=1}^{M} \mathbb{E}[ξ_k \mid (X_1,\ldots,X_N)] g_k(x) + \|x\|^2 . \tag{5.8}
\]

Using the Bayes formula and the likelihood function \( 2.3 \), we derive

\[
\hat{J}_k \overset{\text{def}}{=} \mathbb{E}[ξ_k \mid (X_1,\ldots,X_N)] = \mathbb{P}(ξ_k = 1 \mid (X_1,\ldots,X_N)) \tag{5.9}
\]

\[
= \frac{(d\mathbb{P}_k/d\mathbb{P}_0)(d\mathbb{P}_0/d\mathbb{P}_0)^{-1} p}{1 - p + (d\mathbb{P}_k/d\mathbb{P}_0)(d\mathbb{P}_0/d\mathbb{P}_0)^{-1} p} \tag{5.10}
\]

\[
= \frac{pe^λ \int_{I_k} \tilde{g}_k dx}{1 - p + pe^λ \int_{I_k} \tilde{g}_k dx} 1(\forall X^{(1:d-1)}_i \in I_k : X^{(d)}_i \geq \bar{g}_k(X^{(1:d-1)}_i) \tag{5.11}
\]

where the superscript in \( X^{(1:d-1)}_i \) and \( X^{(d)}_i \) denotes the corresponding components of the vector \( X_i \), \( \bar{g}_k(x) = g_k(x) - \|x\|^2 \), \( \bar{g}_0(x) = 0 \) and \( \mathbb{P}_k, \mathbb{P}_0 \) stand for the measures associated with the convex sets whose boundaries are \( \bar{g}_k(x) \) and \( \bar{g}_0(x) \) in the k-th cube \( I_k \). Observing
\[ \hat{\cal B} = \sum_{k=1}^{M} \xi_k g_k(x) dx - \int_{[0,1]}^2 \|x\|^2 \text{ and } \text{Var}(\xi_k \mid (X_1, \ldots, X_N)) = \xi_k(1 - \xi_k), \]

the Bayes risk can be calculated as

\[ \mathbb{E}[\xi \mathbb{E}[\xi \mid (X_1, \ldots, X_N)] \left( |\hat{\cal C}| - \hat{\cal B} \right)^2] = \sum_{k=1}^{M} \text{Var}(\xi_k \mid (X_1, \ldots, X_N)) \left( \int_{I_k} g_k(x) dx \right)^2, \]

(5.12)

(5.13)

(5.14)

Note, that \( \lambda \int_{I_k} \xi_k g_k(x) dx \leq \lambda \int_{I_k} g_k(x) dx \leq \lambda b_2 h^{d+1} \leq b_2 \) for some constants \( b_1, b_2 > 0 \) if we choose \( h_\lambda \sim \lambda^{-1/(d+1)} \). This implies the bound for \( b_3, b_4, b_5 > 0 \)

\[ \mathbb{E}[\xi \mathbb{E}[\xi \mid (X_1, \ldots, X_N)] \left( |\hat{\cal C}| - \hat{\cal B} \right)^2] \geq M b_3 \left( \int_{I_k} g_k(x) dx \right)^2 \geq M b_3 h^{d+1} \left( \int_{I_k} u \left( \frac{x - q_k}{h_\lambda} \right)^2 dx \right)^2, \]

(6.1)

\[ \geq b_4 h^{d+3} \geq b_5 \lambda^{-(d+3)(d+1)}, \]

(5.16)

which completes the proof. \( \square \)

6 Appendix

Proof of Lemma 4. Using that the bias of the oracle estimator \( \hat{\theta} = |\hat{\cal C}| + N_\cal C/(N_0 + 1)|\hat{\cal C}| \) is exponentially small, it remains to compare its expectation with the expectation of the plug-in estimator \( \hat{\theta}_{\text{plugin}} \) to show (4.2):

\[ \mathbb{E}[(\hat{\theta} - \hat{\theta}_{\text{plugin}})] = \mathbb{E} \left[ |\hat{\cal C}| \left( \frac{N_\cal C}{N_0 + 1} - \frac{N_\cal C}{N} \right) \right] = \mathbb{E} \left[ |\hat{\cal C}| \left( \frac{N_\cal C^2 - N_\cal C}{(N_0 + 1)(N_0 + N_\cal C)} \right) \right], \]

(6.1)

(6.2)

where in the last line we have used \( |\hat{\cal C}| > 0 \) only if \( N_\cal C \geq d + 1 \) and in this case \( N_\cal C^2 - N_\cal C \geq \frac{d}{d+1} N_\cal C^2 \). Using \( \mathbb{E}[(N_0 + 1)^{-1} |\hat{\cal C}|] = \lambda_0^{-1} (1 - e^{-\lambda_0}) \) from
above, we obtain after writing $1(N \leq 2\lambda|C|) = 1 - 1(N > 2\lambda|C|)$

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

By Cauchy-Schwarz inequality and large deviations similarly to (4.11), the first term is bounded from below by a constant multiple of $E[|C \setminus \hat{C}|^2]/|C|$ in view of $E[N^2_C] \geq \lambda^2 E[|C \setminus \hat{C}|]^2$. Because of $N \sim \text{Poiss}(\lambda|C|)$, the second term is of order $\lambda|C|^2e^{-\lambda|C|}$ and thus asymptotically of much smaller order.

References


