Universidad de Cantabria

Facultad de Ciencias

Estimadores Eficientes de Objetos

Efficient Object Sampling

Trabajo de fin de grado
para acceder al

GRADO EN MATEMÁTICAS

Autor: Javier Cobo Rodríguez

Director: Domingo Gómez Pérez

Septiembre 2023

Acknowledgements

I would like to thank my director, Domingo Gómez Pérez, for all the time and help he has given me during the development of this work. His patience, friendlyness and commitment made the time we spent together really enjoyable and thanks to him I learned a lot about an unknown and interesting science for me which is stereology. I would also like to thank Luis María Cruz-Orive for the time he spent answering my questions about stereology and Ana Isabel Gómez for the help she's given me.

I would also like to thank my friends from the Bachelor's Double Degree in Physics and Mathematics, from the Bachelor's Degree in Physics and the Bachelor's Degree in Mathematics for all their help and for being there with me during all these five years. I'm really happy I shared that time with them and I hope there's more to come.

Finally, I would like to thank my family for all the love and support they've given me. All their work and effort made it possible for me to get here, and I'll always be grateful for that.

Estimadores Eficientes de Objetos

Resumen

palabras clave: Estereología, muestreo, eficiencia

La estereología es la ciencia del muestreo geométrico y permite estimar cantidades utilizando fórmulas que suelen incluir esperanzas. En este trabajo se han intentado mejorar los estimadores usados en estereología aproximando dichas esperanzas mediante el uso de integración QMC con N puntos óptimos que minimizan el *worst-case-error* para el *reproducing kernel Hilber space* H^1_{mix}. Este proceso debería dar estimadores con poca varianza, mejorando la calidad de la cantidad estimada.
Para comprobar el funcionamiento de este proceso se han seleccionado tres *test systems* conocidos, el *test system de quadrats*, el *Buffon test system* y el *Buffon-Steinhaus test system*, y se ha implementado en un código python para cada uno.
Los resultados calculados por ordenador apuntan a que este proceso puede que funcione mejor o peor dependiendo de la isotropía, anisotropía o simetría que tiene el objeto de cuya cantidad se quiere estimar.

Efficient Object Sampling

Abstract

keywords: Stereology, sampling, efficiency

Stereology is the science of geometric sampling and it allows to estimate quantities using formulas that usually include expectations. In this work we tried to improve stereology estimators by approximating the expectations using QMC integration with N optimal points that minimize the worst case error in the reproducing kernel Hilbert space H^1_{mix}. This process should provide estimators with low variances, improving the quality of the estimated quantity.
In order to test this process we selected three known test systems, namely the test system of quadrats, the Buffon test system and the Buffon-Steinhaus test system, and implemented it in python code for each one of them.
The computed results led us to think that this process might perform differently depending on the isotropy, anisotropy or symmetry that the object whose quantity we estimate has.

Contents

1 Introduction **1**
- 1.1 History . . . 2
- 1.2 Estimation Problems . . . 4
  - 1.2.1 Test System of Quadrats . . . 4
  - 1.2.2 Buffon-Steinhaus Test System . . . 5

2 Optimal point sets for quasi–Monte Carlo integration **7**
- 2.1 Mean and Variance . . . 7
- 2.2 Variance prediction formula of particle number estimation in the plane with a test system of quadrats . . . 11
- 2.3 Variance prediction formula of curve length in the plane with a Buffon-Steinhaus test system . . . 12
- 2.4 Quasi-Monte Carlo Integration (QMC Integration) . . . 13
- 2.5 Optimal Point Sets for QMC Integration . . . 18

3 Experimental results **27**
- 3.1 Test system of quadrats . . . 27
  - 3.1.1 Counting people . . . 29
  - 3.1.2 Counting birds . . . 30
- 3.2 Buffon and Buffon-Steinhaus test systems . . . 31
- 3.3 Estimating length of Pausinger's curves . . . 32
- 3.4 Estimating length of DNA molecules . . . 33
  - 3.4.1 Curve 5 . . . 33
  - 3.4.2 Curve 6 . . . 34

4 Conclusions and open problems **37**

Bibliography **39**

A Appendix **41**

1 Introduction

A general problem is to estimate a quantity from an object, relaying only in a finite number of samples and the method used to acquire them. When it comes to object sampling, Stereology is the quintessential science since it makes it able to estimate quantitative properties of spacial objects. According to Cruz-Orive [2017], "Stereology is the science of geometric sampling" whose applications extend to disciplines such as biology and materials science.

In traditional sampling on discrete populations, sampling units had to be accessible to observation. However, the target object in geometric sampling is a subset of space and the sample is obtained via the intersection between the object and what's called a *test probe*. A *test probe* is basically a regular arrangement of test points, lines, planes, or slabs whose size and shape are known and has some sort of randomness mechanism relative to the object (see Cruz-Orive [2017]).

There is another area of research which has been popular for its application to sampling: Quasi-Monte-Carlo algorithms. Among other differences, it can be said that Stereology has traditionally focus on estimations using a small number of samples, whereas Quasi-Monte-Carlo algorithms study methods with large number of samples and deterministic error bounds. The goal of this bachelor thesis is to show common areas of interest and explore the applicability of sampling strategies in well-known estimators used in Stereology. Our findings show that these new estimators are competitive candidates and their performance seem to correlate with geometric properties of the object. These observations open new directions on how to tailor estimators for specific problems.

This document is written with the aim to be accessible to all people with a degree in mathematics, trying to hint the deep ideas behind the proofs. Although this work tries to be self-contained, some results and concepts like dimension, measure, etc. are not formally defined inside the document and we operate with them in an intuitive fashion.

The structure of the thesis is the following. Chapter 1 is devoted to give an ample introduction to Stereology and its fundamentals onto the estimators that will be used in the thesis. Chapter 2 recalls the theory of continuous probability and Hilbert spaces of functions, leading towards the Quasi-Monte Carlo integration method and optimal point sets for its use. Chapter 3 is focused on the evaluation and interpretation of computed results obtained with our python code for our new estimation methods. Finally, Chapter 4 end the thesis with some conclusions and future work.

Remark. During Chapters **1** and **2**, a lot of history, theory and results have been gathered together in order to provide a simple understanding of the procedure and lead towards what we want to accomplish in this work. Therefore, the structure and information from these chapters will resemble to the ones from Cruz-Orive [2017], Cruz-Orive [2022], Leobacher y Pillichshammer [2014] and Hinrichs y Oettershagen [2016].

1.1 History

The underlying theory of Stereology is a blend of integral geometry, probability and statistics expected to be used in biomedical and material sciences. Since its development was focused on these two disciplines, Stereology can be divided in two branches depending on the application:

- *Design based Stereology*: this area does not make any assumption about the data and the inference process is based on the randomness of the samples. It is more centered on biomedical applications, thus the object is assumed to be fixed and bounded and sampling is done with randomized test probes.
- *Model based Stereology*: in this area there are some assumptions about the samples, specially some homogeneity in the object is assumed. It is more centered on material science applications where the object is huge, thus dealing with small portions of practically unbounded spacial structures and randomness being incorporated by means of a random set model (relies on model shapes, thereby usually leads to biased methods).

Another distinction can be done regarding stereology that splits it into *global stereology* and *particle stereology*:

- *Global stereology*: In the design context it deals with total quantities (*e.g.* total number, length, area, volume...). In the model context it deals with ratios (*e.g.* relative volume).
- *Particle stereology*: It deals with mean properties of individual particles (*e.g.* mean volume).

Just as an introduction to this science, let's consider an equivalent problem to Buffon's needle problem, which is said to encapsulate the art and spirit of Stereology:

> A needle of length ℓ is arbitrarily fixed inside a disk of diameter $h > \ell$ in the plane. A straight line in the same plane hits the disk at random. Calculate the probability that the straight line hits the needle

Buffon's problem incorporates both a uniform random location of the center of the needle and a uniform random orientation (isotropic orientation) of the needle. We sketch the proof for the close formula for this probability using simple arguments. An experimental approach would be to place "random lines" and count the proportion to estimate the probability. Now, if the number of random lines that hit the needle and the number of lines that miss the needle are known, then the ratio of those numbers would solve the problem. However, here it is necessary to use some sort of continuous geometric measure rather than traditional counting.

The following is an sketch with the main ideas needed to find the solution to Buffon's problem. Let L_1^2 be a straight line used as a *test probe*. This straight line can be defined with two coordinates (p, ϕ) in the plane, where $p \in (-\infty, \infty)$ is the distance of the line from a fixed origin O, and $\phi \in [0, \pi)$ is the orientation angle. Furthermore, the line can be associated with a density $\mathrm{d}L_1^2 = \mathrm{d}p\mathrm{d}\phi$. Using this density, the total number of all straight lines hitting a convex set of boundary length B, is exactly B.

For a needle of length ℓ, the hitting measure (namely the measure of all the probes hitting the set) is 2ℓ. The reason is because the needle has to be regarded as a flattened convex set of perimeter 2ℓ. For the same reason, the hitting measure is πh for a disk of diameter h. Therefore, the probability is $2\ell/(\pi h)$.

This result aroused controversy, resulting on the appearance of paradoxes such as "Bertrand's paradoxes". A few years later it was discovered that measure densities had to be motion invariant,

that is, invariant with respect to translations and rotations, therefor making them independent of the reference frame.

All these events made integral geometry emerge as a mathematical discipline used to obtain motion invariant densities for geometric objects, thus establishing a foundation of geometrical probability. Some typical integral geometry results are the Crofton formulas, which relate intersections between two objects and their properties (number, length, area, volume...). For example, let Y be a bounded planar curve of finite length B. The measure of the number $I(Y \cap L_1^2)$ of intersections determined by all the motion invariant straight lines hitting the curve is

$$\int_Y I(Y \cap L_1^2)\, \mathrm{d}L_1^2 = 2B \tag{1}$$

If Y is the boundary of a convex set, then $I(Y \cap L_1^2) = 2$, therefore the preceding formula yields the solution of Buffon's problem, noting the hitting measure is B.

The minimal requirement to sample by intersecting with a probe is that we obtain information. For that, Stereology also takes into account dimensional considerations in order to obtain the desired quantity. Depending on the dimension of the object and the probe, different procedures can be used and different parameters have to be taken into account. To be more specific, for an object Y hit by a probe T in $\mathbb{R}^d$,

$$dim(Y \cap T) = dim(Y) + dim(T) - d$$

holds up to a set of positions of zero measure. Since $dim(Y \cap T) \geq 0$, the following inequality must always hold,

$$dim(Y) + dim(T) \geq d. \tag{2}$$

If equation (2) holds and either $dim(Y) = d$ or $dim(T) = d$, then relating Y and T by a Crofton formula is viable as long as T has a fixed orientation relative to Y. On the other hand, if $dim(Y) < d$ and $dim(T) < d$, then the density of T has to be motion invariant.

The next breakthrough in history was that Buffon's problem and the development of sampling and statistics in the 20th century inspired estimation problems. Furthermore, proper probability measures for random probes were defined, which was fundamental in the development of stereology. For instance, the probability element associated with a test line hitting a disk is

$$\mathbb{P}(\mathrm{d}p, \mathrm{d}\phi) = \frac{\mathrm{d}p\mathrm{d}\phi}{h\pi}, \quad p \in [-h/2, h/2], \quad \phi \in [0, \pi),$$

namely the motion invariant density normalized by the measure of all test lines hitting the disk. This probability element implies that $\phi \in [0, \pi)$ is *uniform random* (UR), that is, *isotropic random* (IR) in the unit semicircle, whereas $p \in [-h/2, h/2]$ is independent and UR. Therefore, the test line is said to be *isotropic uniform random* (IUR) hitting the disk.

To show an application of the formula in equation (2) to an estimation problem, suppose Y is a planar curve with unknown length B contained in a disk of diameter h. We suppose that the curve is not possible to manipulate (straighten the curve, cut into pieces, etc) and it is necessary to estimate the parameter B. In order to do that, we design the following sampling experiment or estimator. The sampling experiment consists of a disk being hit by a IUR test line L_1^2. Considering equation (1), since $I(Y \cap L_1^2)$ is an integer valued random variable, its expectation (mean value) is

$$\mathbb{E}[I(Y \cap L_1^2)] = \int I(Y \cap L_1^2)\, \mathbb{P}(dp, d\phi) = \frac{2B}{h\pi}.$$

Therefore, $\widehat{B} = \frac{h\pi}{2} \cdot I(Y \cap L_1^2)$ is an *unbiased estimator* (UE) of B.

On one hand, the mean of different unbiased estimations taken at random and independently can be used to approximate the length better. On the other hand, one inconvenience of the following estimator is that some of those estimations could give wrong results, e.g. certain estimations can be zero if the test line misses the curve, which results on time and effort spent uselessly. In practice, *estimators* which sample in a strategic way, v.g. adding more test lines, is more convenient and generally more efficient than independent sampling because estimations are rarely zero and variances tend to be significantly lower (see Cruz-Orive [2017]).

These gives two basic strategies for sampling:

- Random sampling: which estimates the measures using different results obtained for random test probes.
- Systematic sampling: which estimates the measures using specific test probes that may get better results. This usually implies some kind of dependence between samples.

1.2 Estimation Problems

Now that some of the basic concepts and ideas about stereology have been introduced, lets focus on the estimation problems that will be studied in this work.
Notation: Through the rest of the thesis, we use the following conventions,

1. Y will denote a compact set with $dim(Y) = q$ and q-measure $\gamma(Y)$. For example, if $Y \subset \mathbb{R}^3$, then $\gamma(Y)$ may represent number of subsets, curve length, surface area or volume, depending on whether $q = 0, 1, 2, 3$ respectively.
2. $L_r \equiv L_r^d,\ \ r = 0, 1, ..., d$ will denote an r-plane in $\mathbb{R}^d$, for instance, a test point, line, plane or slab according to whether $r = 0, 1, 2, 3$ respectively.
3. $\alpha(X)$ will denote the geometric measure of a compact set X, thus $\alpha(\emptyset) = 0$. For example, if $Y \subset \mathbb{R}^3$, then $\alpha(Y \cap L_r)$ will represent number, length, area or volume, depending on whether $q + r - d = 0, 1, 2, 3$, respectively.
4. The idea behind this work can be applied to most or all of the different estimation problems. Here we test its performance in two of them.

1.2.1 Test System of Quadrats

Suppose $Y = \{y_1, y_2, ..., y_N\} \subset D \subset \mathbb{R}^2$ is a finite set of N disjoint point particles where y_i denotes the i-th point particle and D is a reference disk of diameter H. Consider $T_2^2(x, \omega)$, a test quadrat (square) of area a and random fixed orientation $\omega \in [0, 2\pi)$ that's free to move parallel to itself with translation invariant density $dT_2^2(x, \omega) = dx$. We have

$$\int_{\mathbb{R}^2} N(Y \cap T_2^2(x, \omega))\, \mathrm{d}x = \int_{\mathbb{R}^2} \sum_{i=1}^{N} 1_{T_2^2(x,\omega)}(y_i), dx = \int_{\mathbb{R}^2} \sum_{i=1}^{N} 1_{T_2^2(0,\omega)}(y_i - x), dx$$
$$= \sum_{i=1}^{N} \int_{\mathbb{R}^2} 1_{T_2^2(0,\omega)}(y_i - x), dx = Na \quad (3)$$

The probability element for a quadrat with fixed orientation $T_2^2(x)$ hitting a disk D is

$$\mathbb{P}(dx) = \frac{dx}{A(D_\oplus)},\ \ x \in D_\oplus.$$

where $A(D_{\oplus})$ is the area of the geometric loci of the AP as the test probe hits D. Therefore, the number of point particles is

$$N = \frac{A(D_{\oplus})}{a} \cdot \mathbb{E}(Q(x))$$

where $Q(x)$ is the number of point particles captured by the quadrat. Similar results can be obtained for $\mathbb{R}^d$.

Since the quadrat's orientation is fixed, it is usually called a *FUR quadrat* (Fixed Uniform Random). Let's consider a FUR quadrat $T_0 \equiv T_2^2(0,\omega) \subset J_0 \subset \mathbb{R}^2$ of area $a_0 > 0$ and fixed orientation $\omega \in \mathbb{S}$ where T_0 is what's called a *fundamental probe* contained in a *fundamental tile* J_0 of the test system. A *FUR test system of quadrats* can be defined as

$$\Lambda_x := \{T_2^2(x + t_k, \omega), k \in \mathbb{Z}\}, \;\; x \sim UR(J_0)$$

Considering equation (3) and Santaló's formula for FUR test systems with bounded fundamental test probes $T_0 \subset J_0 \subset \mathbb{R}^d$,

$$\gamma(Y)\nu(T_0) = \int_{J_0} \alpha(Y \cap \Lambda_x)\, \mathrm{d}x$$

where $\gamma(Y)$ is the target measure, $\nu(T_0)$ is the known measure for the fundamental probe and $\alpha(Y \cap \Lambda_x)$ is the geometric measure of $Y \cap \Lambda_x$, we have

$$\gamma(Y) \cdot a_0 = \int_{J_0} \alpha(Y \cap \Lambda_x)\, \mathrm{d}x = a \cdot \mathbb{E}[\alpha(Y \cap \Lambda_x)]$$

where $a \geq a_0$ is the area of J_0. Therefore, the basic identity for a FUR test system of quadrats is

$$\gamma(Y) = \frac{a}{a_0} \cdot \mathbb{E}[\alpha(Y \cap \Lambda_x)] \tag{4}$$

where $\gamma = N$ (number), B (length), A (area) depending on whether $dim(Y) = 0, 1, 2$ respectively (Cruz-Orive [2022]).

In this work, Y will be a set of point particles representing people in different photos. This means that $dim(Y) = 0$, thus the identity that will be used reads

$$N = \frac{a}{a_0} \cdot \mathbb{E}[N(Y \cap \Lambda_x)] \tag{5}$$

Later in this work we will use the aforementioned identity to estimate the amount of points present in photos and compare the efficiency of the usual programming methods with a new method that will be explained.

1.2.2 Buffon-Steinhaus Test System

Suppose $Y \subset \mathbb{R}^2$ is a curve of finite length $B > 0$. The *Buffon test system* consists of IUR parallel test lines separated by a constant distance $T > 0$, that is,

$$\Lambda_{x,\omega} := \{L_1^2(x + kT, \omega), \;\; k \in \mathbb{Z}\}, \;\; x \sim UR[0,T], \;\; \omega \sim IR[0,\pi].$$

In this case, a straight line $L_1^2(0,\omega)$ is the fundamental probe, and the fundamental tile for the auxiliary test system of points on an axis normal to the test lines is the segment $J_0 = [0,T)$. Thus, the joint probability element is

$$\mathbb{P}(\mathrm{d}x, \mathrm{d}\omega) = \frac{\mathrm{d}x}{T} \cdot \frac{\mathrm{d}\omega}{\pi}, \;\; x \in [0,T), \;\; \omega \in [0,\pi)$$

Now, considering Santaló's formula for an IUR test system of unbounded r-planes in $\mathbb{R}^d$,

$$c_1 \cdot \gamma(Y) = \int_{G_{r,d-r}} \mathrm{d}\omega \int_{J_0} \alpha(Y \cap \Lambda_{x,\omega})\, \mathrm{d}x,$$

where $G_{r,d-r}$ is *the Grassmannian* (*i.e.* the space of all non oriented r-subspaces $L^d_{r[0]} \equiv L^d_r(0,\omega)$ acted upon by the group of rotations of $\mathbb{R}^d$), the corresponding identity reads

$$\gamma(Y) = \frac{c_{10}}{c_1} \cdot V(J_0) \cdot \mathbb{E}[\alpha(Y \cap \Lambda_{x,\omega})],$$

where $V(J_0)$ represents the 'volume' of J_0, $q = dim(Y)$,

$$\frac{c_{10}}{c_1} = \frac{O_q O_r}{O_d O_{q+r-d}}, \quad \text{and } O_k = \frac{2\pi^{(k+1)/2}}{\Gamma((k+1)/2)}, \;\; k = 0, 1, ..., d,$$

is the surface area of the k-dimensional unit sphere $\mathbb{S}^k$.

With the aforementioned test system, $\gamma(Y) = B$, $c_{10}/c_1 = \pi/2$ and $V(J_0) = T$. Therefore, the basic identity for the Buffon test system is

$$B = \frac{\pi}{2} \cdot T \cdot \mathbb{E}[I(Y \cap \Lambda_{x,\omega})]. \tag{6}$$

In this work we will also use the *Buffon-Steinhaus test system*, also called the *square grid*. This system is the union of two perpendicular Buffon test systems, thus, the test length per unit area is doubled and T should be replaced with $T/2$. Whereby, the basic identity for the Buffon-Steinhaus test system is

$$B = \frac{\pi}{4} \cdot T \cdot \mathbb{E}[I(Y \cap \Lambda_{x,\omega})]. \tag{7}$$

Alternatively, one may keep the test system fixed and the target set $Y \equiv Y(x,\omega)$ mobile, in which case

$$B = \frac{\pi}{4} \cdot T \cdot \mathbb{E}[I(Y(x,\omega) \cap \Lambda_{0,0})],$$

which gives the same estimator (Cruz-Orive [2022]).

That being said, later in this work we will use the aforementioned identity to estimate the length of specific curves and compare the efficiency of the usual programming methods with the new method.

2 Optimal point sets for quasi–Monte Carlo integration

This chapter introduces the relationship between probability theory and Hilbert spaces. We will see that better estimators relate to sampling points that minimize some quantity. The chapter aims to connect all these subject in a rigorous way.

2.1 Mean and Variance

As it was shown in the previous chapter, estimation problems in Stereology generally make use of the expectation (namely the mean) in order to get a better estimation of the desired quantity. If one wants to know how good of an estimation was obtained, the variance comes into play. However, since the real quantity that we want to estimate is unknown, it is necessary to predict the error as well with *variance predictors*, which are usually employed to fulfill the same role as the usual variance.

We follow the naming convention in Stereology, stating that the mean of a random variable is called a *first order property* and the variance is a *second order property*. We remark that formulas for the variance of independent estimations are known. However, under systematic sampling the sampled items are correlated to unknown degrees depending on the population pattern, thus the problem is non trivial unless the population is a random permutation (Cruz-Orive [2017]).

In probability, we have the following definitions for the mean and the variance of real valued random variables.

Definition 1. *Given a sample space Ω, an event space σ and a probability function $\mathbb{P}$. If X is a discrete, real valued random variable defined in a probability space $(\Omega, \sigma, \mathbb{P})$, the mathematical expectation of X and the variance of X are defined as*

$$\mathbb{E}[X] = \sum_n x_n \cdot \mathbb{P}_X[x_n] = \sum_n x_n \cdot \mathbb{P}[X = x_n], \quad Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

given that the sums exist.

Definition 2. *Given a sample space Ω, an event space σ and a probability function $\mathbb{P}$. If X is a real valued random variable defined in a probability space $(\Omega, \sigma, \mathbb{P})$ such that it's density function f_X exists, the mathematical expectation of X can be defined as*

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} t \cdot f_X(t)\, \mathrm{d}t,$$

given that this expression makes sense.

Furthermore, a very useful proposition says as follows.

Proposition 1. *Given a sample space Ω, an event space σ and a probability function $\mathbb{P}$. If X is a real valued random variable defined in a probability space $(\Omega, \sigma, \mathbb{P})$ such that $\mathbb{E}[|X|^2] < \infty$, then*

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

Proof. By definition, $Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$. What's more,

$$(X - \mathbb{E}[X])^2 = X^2 + (\mathbb{E}[X])^2 - 2X\mathbb{E}[X].$$

Given that $\mathbb{E}[X]$ is a real value, applying the expectation we have

$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] + \mathbb{E}[(\mathbb{E}[X]^2)] - \mathbb{E}[2X\mathbb{E}[X]] \\ = \mathbb{E}[X^2] + (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.\end{aligned}$$

This finishes the proof. □

Regarding the error variance estimator, we will now show how to obtain one for the 1-dimensional case. Let f be a predictor function such that it is periodic in $[0, 1)$ and define the *covariogram* as

$$g(h) := \int_0^1 f(x) \cdot f(x + h)\, \mathrm{d}x. \tag{8}$$

This concept will be of utter importance, so first lets give some basic properties.

Proposition 2. *If f is a periodic function in $[0, 1)$ and g is its covariogram, then the following properties apply:*

1. *g is symmetric about $h = 0$, namely: $g(h) = g(-h)$*
2. *$|g(h)| \leq g(0)$ (i.e. $\sup_h g(h) = g(0)$)*
3. *$\int_0^1 g(h)\, \mathrm{d}h = \left\{ \int_0^1 f(x)\, \mathrm{d}x \right\}^2 =: V^2$*

Proof. We start by proving the first item

$$g(-h) = \int_0^1 f(x) \cdot f(x - h)\, \mathrm{d}x = \{y = x - h; dy = dx\} = \int_0^1 f(y + h) \cdot f(y)\, \mathrm{d}y = g(h).$$

For the second item

$$\begin{aligned}0 \leq \int_0^1 [f(x + h) - f(x)]^2\, \mathrm{d}x = \int_0^1 f^2(x + h)\, \mathrm{d}x \; + \int_0^1 f^2(x)\, \mathrm{d}x \; - 2 \cdot \int_0^1 f(x) \cdot f(x + h)\, \mathrm{d}x \\ = g(0) + g(0) - 2 \cdot g(h)\end{aligned}$$

For the last item,

$$\int_0^1 \mathrm{d}h \int_0^1 f(x) \cdot f(x + h)\, \mathrm{d}x = \int_0^1 f(x)\, \mathrm{d}x \int_0^1 f(x + h)\, \mathrm{d}h = \left\{ \int_0^1 f(x)\, \mathrm{d}x \right\}^2.$$

This finishes the proof. □

Note: We remark that if f is integrable, then its covariogram is derivable. This will be important when considering the Hilbert space of functions defined later in this chapter.

We apply this to design an estimator for the following parameter of interest:

$$V = \int_0^1 f(x)\mathrm{d}x,$$

where f is a periodic function, with period 1. Lets imagine that we take observations at $z + j/N$ with $j = 0, 1, 2, \ldots, N-1$. An estimator for V reads

$$\widehat{V} = \frac{1}{N} \cdot \sum_{j=0}^{N-1} f\left(z + \frac{j}{N}\right) \equiv v(z)$$

which is a periodic function of z with period $1/N$, thus it suffices to consider $z \in (0, 1/N)$. As a result, assuming $z \sim U(0, 1/N)$ we have

$$\mathbb{E}[v(z)] = \int_0^{1/N} \frac{\mathrm{d}z}{1/N} \cdot \frac{1}{N} \cdot \sum_j f(z + j/N) = \sum_j \int_0^{1/N} f(z + j/N)\,\mathrm{d}z = \sum_j V_j = V,$$

that is, $v(z)$ is an UE of V.

Now, by definition,

$$Var(v(z)) = \frac{1}{1/N} \int_0^{1/N} v^2(z)\,\mathrm{d}z - V^2 = N \int_0^{1/N} v^2(z)\,\mathrm{d}z - V^2$$

Since $v(z)$ is periodic of period $1/N$, we can write it as a *Fourier series*,

$$v(z) = \sum_{k=-\infty}^{\infty} c_k \cdot e^{2\pi i k z N},$$

where

$$c_k = \begin{cases} N \int_0^{1/N} v(z) \cdot e^{(-2\pi i k z N)}\,\mathrm{d}z & (k > 0) \\ N \int_0^{1/N} v(z) \cdot e^{(2\pi i k z N)}\,\mathrm{d}z & (k < 0). \end{cases}$$

Substituting $v(z)$, for $k > 0$ we can write

$$c_k = N \int_0^{1/N} \frac{1}{N} \cdot \sum_{j=0}^{N-1} f(z + j/N) \cdot e^{(-2\pi i k z N)}\,\mathrm{d}z = \int_0^1 f(z) \cdot e^{(-2\pi i k z N)}\,\mathrm{d}z$$

and the conjugate for $k < 0$.

Using *Parseval's theorem* and $v(z)$'s periodicity we get

$$N \int_0^{1/N} v^2(z)\,\mathrm{d}z = \sum_{k=-\infty}^{\infty} c_k \overline{c_k},$$

therefore

$$Var(v(z)) = \sum_{k=-\infty}^{\infty} c_k \overline{c_k} - \int_0^1 g(h)\,\mathrm{d}h,$$

where $\overline{c_k}$ is the conjugate of c_k.

We introduce the following notation

$$F(u) := \int_0^1 f(z) \cdot e^{(-2\pi iuz)} \, \mathrm{d}z,$$

then

$$Var(v(z)) = \sum_{k=-\infty}^{\infty} F\left(kN\right) \overline{F}\left(kN\right) - \int_0^1 g(h) \, \mathrm{d}h$$

where $\overline{F}(\cdot)$ is the conjugate of $F(\cdot)$.

Consider now the transform of $g(h)$, namely

$$\begin{aligned} G(u) := \int_0^1 g(h) \cdot e^{(-2\pi iuh)} \, \mathrm{d}h = \int_0^1 f(x) \, \mathrm{d}x \int_0^1 f(x+h) \cdot e^{(-2\pi iuh)} \, \mathrm{d}h \\ = \int_0^1 f(x) \, \mathrm{d}x \int_0^1 f(r) \cdot e^{(-2\pi iu(r-x))} \, \mathrm{d}x = F(u) \cdot \overline{F}(u), \end{aligned}$$

then the Fourier coefficients of the Variance have to be symmetric, giving,

$$Var(v(z)) = \sum_{-\infty}^{\infty} G\left(kN\right) - G(0) = 2 \cdot \sum_{k=1}^{\infty} G\left(kN\right).$$

In order to have a predictor, we are going to model the variance and suppose that the covariogram can be approximated by a polynomial, namely

$$g(h) := \sum_{j=0}^{r} a_j \cdot |h|^j.$$

We substitute the model for the covariogram and calculate the Fourier coefficients:

$$G(u) = 2 \cdot \int_0^1 \left(\sum_{j=0}^{r} a_j \cdot h^j \right) \cdot e^{(-2\pi iuh)} \, \mathrm{d}h = 2 \cdot \sum_{j=0}^{r} a_j \int_0^1 h^j \cdot e^{(-2\pi iuh)} \, \mathrm{d}h.$$

Because we only need real value of the variance, that is, the real part of $G(u)$, we define

$$\Gamma_{inc}(m) := \int_0^1 y^{m-1} \cdot e^{-y} \, \mathrm{d}y \approx \Gamma(m),$$

thus, taking $y = 2\pi iuh$ we get

$$\int_0^1 h^j \cdot e^{(-2\pi iuh)} \, \mathrm{d}h = \frac{1}{(2\pi iu)^{j+1}} \int_0^1 y^j \cdot e^{(-y)} \, \mathrm{d}y = \frac{\Gamma(j+1)}{(2\pi iu)^{j+1}},$$

and because we want the real part, we only want the terms given by odd j, namely $j = 2p-1$, $p = 1, 2, \ldots$. In consequence,

$$Var(v(z)) = 2 \cdot \sum_{k=1}^{\infty} Re\left[G\left(kN\right)\right] = 2 \cdot \sum_{p=1}^{j} a_{2p-1} \cdot \frac{1}{N^{2p}} \cdot \left\{ 2 \cdot \frac{\Gamma(2p)}{(-1)^p \cdot (2\pi)^{2p}} \cdot \sum_{k=1}^{\infty} \frac{1}{k^{2p}} \right\}$$

Furthermore, since the *Bernoulli number* $B_{2p} \equiv B_{2p}(0)$ can be expressed

$$B_{2p}(0) = \frac{(-1)^{p-1} \cdot 2\Gamma(2p+1)}{(2\pi)^{2p}} \cdot \sum_{k=1}^{\infty} \frac{1}{k^{2p}},$$

we can rewrite the variance as

$$Var(v(z)) = -\sum_{p=1}^{\lfloor \frac{r+1}{2} \rfloor} a_{2p-1} \cdot \frac{B_{2p}}{p \cdot N^{2p}} = -a_1 \frac{B_2}{N^2} - a_3 \frac{B_4}{2 \cdot N^4} - ...$$

Considering that $a_1 = g'(0)$ under the polynomial model and that $B_2 = 1/6$, we can approximate the variance as follows,

$$Var(v(z)) \approx -\frac{g'(0)}{6N^2}.$$

Finally, lets fit the parable $g(h) = a_0 + a_1 \cdot h + a_2 \cdot h^2$ through the sample points $\{(j/N, \widehat{g}(j/N)),\ j = 0, 1, 2\}$, where

$$\widehat{g}(j/N) = \frac{1}{N} \cdot \sum_{r=1}^{S} f_r \cdot f_{r+j}$$

with S the total number of sections and f_r the r-th section area observed at the r-th abscissa. More briefly, write

$$g_0 \equiv \sum f_i^2, \quad g_1 \equiv \sum f_i \cdot f_{i+1}, \quad \text{and } g_2 \equiv \sum f_i \cdot f_{i+2},$$

where $f_i \equiv f(z + i/N)$ and thus $\widehat{g}(j/N) \equiv \frac{1}{N} \cdot g_j$. Then we have

$$\frac{1}{N} \cdot g_0 = a_0, \quad \frac{1}{N} \cdot g_1 = a_0 + a_1 \cdot \frac{1}{N} + a_2 \cdot \frac{1}{N^2}, \quad \frac{1}{N} \cdot g_2 = a_0 + a_1 \cdot \frac{2}{N} + a_2 \cdot \left(\frac{2}{N}\right)^2$$

from which we get

$$a_1 = \frac{4g_1 - g_2 - 3g_0}{2}.$$

Therefore, we can get the following variance estimator,

$$\widehat{Var(v(z))} \approx -\frac{a_1}{6N^2} = \frac{3g_0 + g_2 - 4g_1}{12} \cdot \frac{1}{N^2}$$

This is the general method to approximate the variance. Next, we will present the error variance estimators that have been developed for the two estimation procedures that will be used in this work. The method to discover them follows the same principles but the proofs are more involved due to the number of variables. After that we will explain how the expectations are calculated.

2.2 Variance prediction formula of particle number estimation in the plane with a test system of quadrats

Consider $Y \subset \mathbb{R}^2$ a bounded and finite set of $N \equiv N(Y)$ particles. We want to estimate N using a test system of quadrats Λ_x whose fundamental tile J_0 is a square with side length T and whose fundamental probe T_0 is a square quadrat with side length $0 < t \leq T$. Let Q represent the number of particles captured by Λ_x, $x \sim UR(J_0)$. Since Y contains point particles, Q is the total number of particles within the quadrats, otherwise, the particles should be captured using unbiased sampling rules such as the *forbidden line rule*.

As it was shown in the first chapter, an UE for N is given by

$$\widehat{N} = \frac{a}{a_0} \cdot Q = \frac{T^2}{t^2} \cdot Q$$

In an effort to interpret the error variance predictor for $\widehat{N}$, the test system is regarded as a two stage sample:

- *First stage sample:* It consists of a set of n Cavalieri stripes of thickness t and period T encompassing the particle population Y.
- *Second stage sample:* It consists of another set of Cavalieri stripes with the same period and thickness but perpendicular to the other one.

Using this two stage sample, we get something similar to Λ_x providing new data such as Q_{oi} and Q_{ei}, namely the number of particles captured by the odd and even quadrats within the ith stripe respectively. This means that $Q_i = Q_{oi} + Q_{ei}$ is the number of particles captured by all quadrats within the i-th stripe, therefore $Q = \sum_{i=1}^{n} Q_i$. As a result, a variance predictor for $\widehat{N}$ can be obtained as follows,

$$Var(\widehat{N}) = \frac{\alpha(0,\tau)}{\tau^4} \cdot [3(C_0 - \widehat{\nu}_n) - 4C_1 + C_2] + \frac{\widehat{\nu}_n}{\tau^4}, \quad n \geq 3$$

where $\tau = t/T \in (0, 1]$,

$$C_k = \sum_{i=1}^{n-k} Q_i Q_{i+k}, \quad k = 0, 1, 2,$$

$$\widehat{\nu}_n = \sum_{i=1}^{n} Var(Q_i) = \frac{(1-\tau)^2}{3-2\tau} \cdot \sum_{i=1}^{n} (Q_{oi} - Q_{ei})^2$$

and

$$\alpha(0,\tau) = \frac{1}{6} \cdot \frac{(1-\tau)^2}{2-\tau}.$$

Note: This variance predictor must first be evaluated for a given direction of the stripes, and then for the perpendicular direction. The *real variance predictor* is the average of both results (Cruz-Orive [2022]).

2.3 Variance prediction formula of curve length in the plane with a Buffon-Steinhaus test system

Consider $Y \subset \mathbb{R}^2$ a bounded, piecewise smooth curve with finite length $B > 0$. We want to estimate B using the Buffon-Steinhaus test system, namely a square grid whose fundamental tile J_0 is a square with side length $T > 0$. Similarly to what's been done in the last section, consider the square grid as two different sets of Cavalieri lines (namely *set 1* and *set 2*) of period T, making angles ω and $\omega + \pi/2$ with a fixed axis, where $\omega \sim IR[0, \pi/2)$. Now, for the set i, let

$$\{I_{i1}, I_{i2}, ..., I_{in_i}\}, \quad i = 1, 2,$$

denote the number of intersections determined by each one of the n_i Cavalieri lines hitting the curve. As one would expect,

$$I_i = \sum_{j=1}^{n_i} I_{ij}, \quad i = 1, 2,$$

denotes the number of intersections determined by the set i hitting the curve. Whereby, an UE for B is given by

$$\widehat{B} = \frac{\pi}{4} \cdot T \cdot (I_1 + I_2).$$

An equivalent design is obtained if the curve is equipped with an associated vector (x, ω), with $x \sim UR(J_0)$ and $\omega \sim IR[0, 2\pi)$.

That being said, a variance predictor for $\widehat{B}$ can be obtained as follows,

$$Var(\widehat{B}) = \frac{\pi^2}{96} \cdot [(\widehat{l_1} - \widehat{l_2})^2 - \widehat{\nu}_2] + \frac{\pi^2}{16} \cdot \widehat{\nu}_2,$$

where

$$\widehat{l_i} = T \cdot I_i,$$

$$\widehat{\nu}_2 = \sum_{i=1}^{2} Var(\widehat{l_i}|l_i),$$

$$Var(\widehat{l_i}|l_i) = \frac{T^2}{12} \cdot (3C_{0i} - 4C_{1i} + C_{2i}), \;\; n_i \geq 3,$$

$$Var(\widehat{l_i}|l_i) = \frac{T^2}{6} \cdot (C_{0i} - C_{1i}), \;\; n_i = 2,$$

$$C_{ki} = \sum_{i=1}^{n_i - k} I_{ij} I_{i,j+k}, \;\; k = 0, 1, ..., n_i - 1, \;\; i = 1, 2.$$

The results and discussions for these two latter sections can be found in [Cruz-Orive, 2022].

2.4 Quasi-Monte Carlo Integration (QMC Integration)

One problem that has to be dealt with in order to compute the target measure estimate, is the numerical integration of multivariate functions. This problem arises because of the need to obtain the expectation, in this case for the number of intersections in both estimation problems.

To simplify things, lets normalize the integration domain to be the compact unit cube $[0,1]^d$, that is,

$$\int_{[0,1]^d} f(\mathbf{x})\, \mathrm{d}\mathbf{x} = \int_0^1 \cdots \int_0^1 f(x_1, ..., x_d)\, \mathrm{d}x_1 \ldots \mathrm{d}x_d$$

The goal is to approximate these integrals using *QMC rules* with fixed integration nodes $\mathbf{x}_0, ..., \mathbf{x}_{N-1} \in [0,1)^d$, namely

$$\int_{[0,1]^d} f(\mathbf{x})\, \mathrm{d}\mathbf{x} \approx \mathcal{Q}_{N,d}(f) := \frac{1}{N} \sum_{n=0}^{N-1} f(\mathbf{x}_n). \tag{9}$$

The crux of this method is the choice of underlying nodes. What's more, we also need some global information on the functions to be integrated, since one can find two functions $f, g : [0,1]^d \longrightarrow \mathbb{R}$ such that $f(\mathbf{x}_n) = g(\mathbf{x}_n) \;\; \forall n = 0, 1, ..., N-1$, but $\int_{[0,1]^d} f(\mathbf{x})\, \mathrm{d}\mathbf{x} - \int_{[0,1]^d} g(\mathbf{x})\, \mathrm{d}\mathbf{x}$ can be any number. So as to avoid this problem, function classes with certain smoothness properties are considered.

Lets start with *univariate QMC integration*, namely QMC integration of univariate real valued functions $f : [0,1] \longrightarrow \mathbb{R}$ with continuous first derivative on $[0,1]$. From the *fundamental theorem of calculus* we have

$$f(x) = f(1) - \int_x^1 f'(y)\, \mathrm{d}y, \;\; \forall x \in [0,1].$$

For the error of a QMC rule based on sample nodes $\mathcal{P} = \{x_0, ..., x_{N-1}\} \subset [0,1)$ we then get

$$\begin{aligned} e(f, \mathcal{P}) &= \int_0^1 f(x)\, \mathrm{d}x - \frac{1}{N} \sum_{n=0}^{N-1} f(x_n) = -\int_0^1 \int_x^1 f'(y)\, \mathrm{d}y\, \mathrm{d}x + \frac{1}{N} \sum_{n=0}^{N-1} \int_{x_n}^1 f'(y)\, \mathrm{d}y \\ &= -\int_0^1 \int_0^y f'(y)\, \mathrm{d}x\, \mathrm{d}y + \int_0^1 \frac{1}{N} \sum_{n=0}^{N-1} 1_{(x_n,1]}(y) f'(y)\, \mathrm{d}y = \int_0^1 f'(y) \left[\frac{1}{N} \sum_{n=0}^{N-1} 1_{(x_n,1]}(y) - y \right] \mathrm{d}y \end{aligned}$$

Since the number of indices $n \in \{0, ..., N-1\}$ for which $x_n \in [0, y)$ is

$$\sum_{n=0}^{N-1} 1_{(x_n,1]}(y) = \sum_{n=0}^{N-1} 1_{(0,y]}(x_n)$$

we find that

$$\frac{1}{N}\sum_{n=0}^{N-1} 1_{(x_n,1]}(y) - y = \Delta_{\mathcal{P},N}(y),$$

that is, the local discrepancy of $\mathcal{P}$ in y. Therefore,

$$e(f,\mathcal{P}) = \int_0^1 f'(y)\Delta_{\mathcal{P},N}(y)\,\mathrm{d}y.$$

Taking the absolute value and applying the *triangle inequality for integrals* and the *Hölder inequality* we get

$$|e(f,\mathcal{P})| \leq \int_0^1 |f'(y)||\Delta_{\mathcal{P},N}(y)|\,\mathrm{d}y \leq \left(\int_0^1 |f'(y)|^r\,\mathrm{d}y\right)^{1/r}\left(\int_0^1 |\Delta_{\mathcal{P},N}(y)|^s\,\mathrm{d}y\right)^{1/s} = \|f'\|_{L_r}\|\Delta_{\mathcal{P},N}\|_{L_s},$$

where $r,s \in [0,\infty]$ such that $\frac{1}{r}+\frac{1}{s}=1$. For QMC integration of functions f for which $\|f'\|_{L_r} < \infty$, sample nodes $\mathcal{P}$ with low L_s discrepancy $L_{s,N}(\mathcal{P}) = \|\Delta_{\mathcal{P},N}\|_{L_s}$ should be chosen.

Note: In order to develop a similar theory for multivariate functions, the notion of *reproducing kernel Hilbert space* will be used. For an integrable function $f : [0,1]^d \longrightarrow \mathbb{C}$ and a point set $\mathcal{P} = \{\mathbf{x}_0, ..., \mathbf{x}_N - 1\} \subset [0,1)^d$, the notation:

$$e(f,\mathcal{P}) = \int_{[0,1]^d} f(\mathbf{x})\,\mathrm{d}\mathbf{x} - \frac{1}{N}\sum_{n=1}^{N-1} f(\mathbf{x}_n),$$

will be used in analogy to the univariate case. What's more, the inner product in a *Hilbert Space* $\mathcal{H}$ will be denoted by $\langle\cdot,\cdot\rangle$ and the corresponding norm will be denoted by $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$.

Now lets give some useful definitions used with Hilbert spaces in order to develop a similar theory for multivariate functions.

Definition 3. *The worst-case error of a QMC rule based on a point set* $\mathcal{P} = \{\boldsymbol{x}_0, ..., \boldsymbol{x}_N - 1\} \subset [0,1)^d$ *in a Hilbert space* $\mathcal{H}$ *of integrable functions on* $[0,1]^d$ *is defined as*

$$e(\mathcal{H},\mathcal{P}) = \sup_{f\in\mathcal{H},\|f\|\leq 1} |e(f,\mathcal{P})|.$$

Also, the notion of *reproducing kernel* will be important for our purposes.

Definition 4. *A Hilbert space* $\mathcal{H}$ *of functions on* $[0,1]^d$ *is a reproducing kernel Hilbert space on* $[0,1]^d$ *if there exists a function* $K : [0,1]^d \times [0,1]^d \longrightarrow \mathbb{C}$ *such that*

K1: $K(\cdot,\boldsymbol{y}) \in \mathcal{H}\ \ \forall \boldsymbol{y} \in [0,1]^d$

K2: $\langle f, K(\cdot,\boldsymbol{y})\rangle = f(\boldsymbol{y})\ \ \forall \boldsymbol{y} \in [0,1]^d,\ \forall f \in \mathcal{H}.$

The function K *is the reproducing kernel of* $\mathcal{H}$.

Note: the reproducing kernel K has been considered as a function of the first variable, denoted by $\cdot$, thus, in $\langle f, K(\cdot,\mathbf{y})\rangle$ the inner product is applied with respect to the first variable of K.

Furthermore, reproducing kernels have the following properties.

Corollary 1. *A function satisfying K1 and K2 is symmetric, uniquely defined and positive semi-definite, that is,*

K3: $K(\boldsymbol{x},\boldsymbol{y}) = \overline{K(\boldsymbol{y},\boldsymbol{x})}\ \ \forall \boldsymbol{x},\boldsymbol{y} \in [0,1]^d$ *(symmetry)*

K4: If a function $\tilde{K}$ *satisfies K1 and K2, then* $\tilde{K} = K$ *(uniqueness)*

K5: For any $a_0, ..., a_{N-1} \in \mathbb{C}$ *and* $\boldsymbol{x}_0, ..., \boldsymbol{x}_{N-1} \in [0,1]^d$*, we have* $\sum_{m,n=0}^{N-1} \tilde{a}_m a_n K(\boldsymbol{x}_m, \boldsymbol{x}_n) \geq 0$ *(positive semi-definiteness)*

Proof. K3: $K(\mathbf{x},\mathbf{y}) = \langle K(\cdot,\mathbf{y}), K(\cdot,\mathbf{x})\rangle = \overline{\langle K(\cdot,\mathbf{x}), K(\cdot,\mathbf{y})\rangle} = \overline{K(\mathbf{y},\mathbf{x})}$.
K4: $\tilde{K}(\mathbf{x},\mathbf{y}) = \langle \tilde{K}(\cdot,\mathbf{y}), K(\cdot,\mathbf{x})\rangle = \overline{\langle K(\cdot,\mathbf{x}), \tilde{K}(\cdot,\mathbf{y})\rangle} = \overline{K(\mathbf{y},\mathbf{x}) = K(\mathbf{x},\mathbf{y})}$.
K5:

$$\sum_{m,n=0}^{N-1} \tilde{a}_m a_n K(\mathbf{x}_m,\mathbf{x}_n) = \sum_{m,n=0}^{N-1} \tilde{a}_m a_n \langle K(\cdot,\mathbf{x}_n), K(\cdot,\mathbf{x}_m)\rangle = \langle \sum_{n=0}^{N-1} a_n K(\cdot,\mathbf{x}_n), \sum_{m=0}^{N-1} a_m K(\cdot,\mathbf{x}_m)\rangle$$
$$= \left\| \sum_{m=0}^{N-1} a_m K(\cdot,\mathbf{x}_m) \right\|^2 \geq 0.$$

This concludes the proof. □

Additionally, a function K satisfying K3 and K5 uniquely determines a Hilbert space of functions for which K1, K2 and K4 hold. Whereby, it makes sense to speak of a reproducing kernel without specifying the corresponding Hilbert space.

That being said, let $\mathcal{H}$ be a Hilbert space of integrable functions $f : [0,1]^d \longrightarrow \mathbb{C}$ equipped with inner product $\langle\cdot,\cdot\rangle$ and corresponding norm $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$.

Definition 5. *A linear functional* T *on* $\mathcal{H}$ *is bounded if there exists* $M < \infty$ *such that* $|T(f)| \leq M$ *for all* f *with* $\|f\| \leq 1$.

Also, it is known that linear functionals boundedness is equivalent to continuity.

Lets consider $e(\mathcal{H},\mathcal{P})$ (*i.e.* the worst-case error of a QMC rule based on a point set $\mathcal{P} \subset [0,1)^d$ in $\mathcal{H}$). It's not clear if $e(\mathcal{H},\mathcal{P})$ is finite, that is to say, if the linear functional $e(\cdot,\mathcal{P})$ is continuous. However, conditions exist such that this is the case.

Let T_y be the linear functional that evaluates $f \in \mathcal{H}$ at $\mathbf{y} \in [0,1]^d$, namely

$$T_y(f) = f(\mathbf{y}) \ \text{ for } f \in \mathcal{H}.$$

T_y is called the *evaluation functional* in y. It turns out that if T_y is continuous $\forall \mathbf{y} \in [0,1]^d$, then so is every QMC rule. What's more, continuity of the evaluation functional is equivalent to the existence of a reproducing kernel.

Theorem 1. *Let* $\mathcal{H}$ *be a Hilbert space of functions on* $[0,1]^d$*.* $\mathcal{H}$ *is a reproducing kernel Hilbert space on* $[0,1]^d$ *if and only if the evaluation functionals*

$$T_y(f) = f(\boldsymbol{y}) \ \text{ for } f \in \mathcal{H}, \boldsymbol{y} \in [0,1]^d$$

are continuous.

Proof. $\Leftarrow$) We assume that the evaluation functionals are continuous, therefore, the *Fréchet-Riesz Representation Theorem* guarantees that, for every $\mathbf{y}$, a uniquely determined function $k_y \in \mathcal{H}$ exists such that

$$T_y(f) = \langle f, k_y \rangle \ \ \forall f \in \mathcal{H}.$$

Defining $K(\mathbf{x}, \mathbf{y}) := k_y(\mathbf{x})$ for $\mathbf{x}, \mathbf{y} \in [0,1]^d$, properties K1 and K2 are satisfied by K, thus $\mathcal{H}$ is a reproducing kernel Hilbert space whose reproducing kernel is K.

$\Rightarrow$) Assume K is a reproducing kernel for $\mathcal{H}$ and let $\mathbf{y} \in [0,1]^d$. Using the Cauchy-Schwarz inequality, we get

$$|T_y(f)| = |f(\mathbf{y})| = |\langle f, K(\cdot, \mathbf{y}) \rangle| \leq \|f\| \|K(\cdot, \mathbf{y})\| \ \ \forall f \in \mathcal{H}.$$

Now, using K2 we have $\|K(\cdot, \mathbf{y})\|^2 = \langle K(\cdot, \mathbf{y}), K(\cdot, \mathbf{y}) \rangle = K(\mathbf{y}, \mathbf{y})$, whereby $|T_y(f)| \leq M := \sqrt{K(\mathbf{y}, \mathbf{y})}$ for every f with $\|f\| \leq 1$. This means that T_y is continuous, which concludes the proof. □

Continuing with the situation above, lets consider the integration functional $I(f) = \int_{[0,1]^d} f(\mathbf{x}) \, \mathrm{d}\mathbf{x}$. If $\mathcal{H}$ has a reproducing kernel K, then for any $f \in \mathcal{H}$ with $\|f\| \leq 1$ we have

$$|\int_{[0,1]^d} f(\mathbf{y}) \, \mathrm{d}\mathbf{y}| = |\int_{[0,1]^d} T_y(f) \, \mathrm{d}\mathbf{y}| \leq \int_{[0,1]^d} |T_y(f)| \, \mathrm{d}\mathbf{y} \leq \int_{[0,1]^d} \sqrt{K(\mathbf{y}, \mathbf{y})} \, \mathrm{d}\mathbf{y}.$$

Therefore, if the reproducing kernel K satisfies

C: $\int_{[0,1]^d} \sqrt{K(\mathbf{y}, \mathbf{y})} \, \mathrm{d}\mathbf{y} < \infty$,

then the integration functional I is continuous.

In conclusion, if $\mathcal{H}$ has a reproducing kernel K that satisfies **C**, then function evaluation and integration are continuous linear functionals, and so is $e(\cdot, \mathcal{P})$ for any point set $\mathcal{P}$. What's more, under these conditions $e(\mathcal{H}, \mathcal{P})$ is a well-defined finite number.

A very important property for the error analysis follows.

Lemma 1. *Let $\mathcal{H}$ be a reproducing kernel Hilbert space with reproducing kernel K and inner product $\langle \cdot, \cdot \rangle$. If the mapping*

$$I(f) = \int_{[0,1]^d} f(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \ \textit{for} \ f \in \mathcal{H}$$

is a continuous linear functional on $\mathcal{H}$, then

$$\int_{[0,1]^d} \langle f, K(\cdot, \boldsymbol{y}) \rangle \, \mathrm{d}\boldsymbol{y} = \langle f, \int_{[0,1]^d} K(\cdot, \boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \rangle.$$

Proof. We assume that I is continuous, therefore, the Fréchet-Riesz representation theorem guarantees the existence of a uniquely determined function $R \in \mathcal{H}$ such that

$$\int_{[0,1]^d} f(\mathbf{y}) \, \mathrm{d}\mathbf{y} = I(f) = \langle f, R \rangle \ \ \forall f \in \mathcal{H}.$$

Given that $R \in \mathcal{H}$, we have

$$R(\mathbf{x}) = \langle R, K(\cdot, \mathbf{x}) \rangle = \overline{\langle K(\cdot, \mathbf{x}), R \rangle} = \overline{\int_{[0,1]^d} K(\mathbf{y}, \mathbf{x}) \, \mathrm{d}\mathbf{y}}.$$

Whereby

$$\int_{[0,1]^d} \langle f, K(\cdot,\mathbf{y})\rangle \,\mathrm{d}\mathbf{y} = \int_{[0,1]^d} f(\mathbf{y})\,\mathrm{d}\mathbf{y} = \langle f, R\rangle = \langle f, \overline{\int_{[0,1]^d} K(\mathbf{y},\cdot)\,\mathrm{d}\mathbf{y}}\rangle$$
$$= \langle f, R\rangle = \langle f, \int_{[0,1]^d} K(\cdot,\mathbf{y})\,\mathrm{d}\mathbf{y}\rangle.$$

This concludes the proof. □

Note: From now on, K will be assumed to satisfy the condition C.

Lets continue the worst-case error analysis. We have

$$I(f) = \int_{[0,1]^d} \langle f, K(\cdot,\mathbf{y})\rangle \,\mathrm{d}\mathbf{y} = \langle f, \int_{[0,1]^d} K(\cdot,\mathbf{y})\,\mathrm{d}\mathbf{y}\rangle.$$

On the other hand, we have

$$\mathcal{Q}_{N,d}(f) = \frac{1}{N}\sum_{n=0}^{N-1} f(\mathbf{x}_n) = \frac{1}{N}\sum_{n=0}^{N-1} \langle f, K(\cdot,\mathbf{x}_n)\rangle = \langle f, \frac{1}{N}\sum_{n=0}^{N-1} K(\cdot,\mathbf{x}_n)\rangle.$$

Therefore, the QMC rule $\mathcal{Q}_{N,d}$ in $\mathcal{H}$ has an integration error that can be expressed as an inner product, namely

$$e(f,\mathcal{P}) = \langle f, h\rangle,$$

where

$$h(\mathbf{x}) = \int_{[0,1]^d} K(\mathbf{x},\mathbf{y})\,\mathrm{d}\mathbf{y} - \frac{1}{N}\sum_{n=0}^{N-1} K(\mathbf{x},\mathbf{x}_n).$$

The function above is called the *representer* of the integration error. Taking the absolute value and applying the Cauchy-Schwarz inequality, we get

$$|e(f,\mathcal{P})| \le \|f\| \cdot \|h\|.$$

What's more, among all functions in the unit ball of $\mathcal{H}$, the normalized representer (*i.e.* $h/\|h\|$) is the hardest to integrate. Thus, the worst-case error reads

$$e(\mathcal{H},\mathcal{P}) = \|h\|.$$

As a result, the squared worst-case error can be written as

$$e^2(\mathcal{H},\mathcal{P}) = \langle h, h\rangle.$$

Finally, these results derive in the following theorem, which states the maximum integration error in terms of $\mathcal{P}$.

Theorem 2. *Let $\mathcal{H}$ be a reproducing kernel Hilbert space with reproducing kernel K such that condition C is satisfied, and let $\mathcal{P} = \{\boldsymbol{x}_0, ..., \boldsymbol{x}_{N-1}\} \subset [0,1)^d$ be a point set with N elements. Then*

$$e^2(\mathcal{H},\mathcal{P}) = \int_{[0,1]^d}\int_{[0,1]^d} K(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\boldsymbol{y} - \frac{2}{N}\sum_{n=0}^{N-1}\int_{[0,1]^d} K(\boldsymbol{x}_n,\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} + \frac{1}{N^2}\sum_{n,m=0}^{N-1} K(\boldsymbol{x}_n,\boldsymbol{x}_m) \qquad (10)$$

On the other hand, there is the problem of choosing the nodes that will be used for the QMC rule. One way of solving this problem is to use what's called a *lattice point set*.

Definition 6. *Let $d, N \in \mathbb{N}$, $N \geq 2$ and let $\boldsymbol{g} \in \mathbb{Z}^d$. A point set $\mathcal{P}(\boldsymbol{g}, N) = \{\boldsymbol{x}_0, ..., \boldsymbol{x}_{N-1}\}$ with*

$$\boldsymbol{x}_n := \left\{ \frac{n}{N} \boldsymbol{g} \right\} \ \forall n = 0, 1, ..., N-1$$

is called a lattice point set and $\boldsymbol{g}$ is its generating vector.

These lattice point sets have the following property,

$$\sum_{n=0}^{N-1} e^{2\pi i n \mathbf{h}\cdot\mathbf{g}/N} = \begin{cases} N & \text{if } \mathbf{h}\cdot\mathbf{g} \equiv 0 \pmod{N} \\ 0 & \text{if } \mathbf{h}\cdot\mathbf{g} \not\equiv 0 \pmod{N}, \end{cases}$$

which allow them to get low values for the worst-case error (see eq. 10). In order to do so, it is enough to find a generating vector $\mathbf{g} \in \{0, ..., N-1\}^d$. Since checking N^d integer vectors is not feasible in practice for big N and d values, an algorithm was developed that makes it able to create the generating vector component-by-component (CBC).

Algorithm 1. ***(CBC Algorithm).*** *Let $d, N \in \mathbb{N}$.*

1. *Choose $g_1 = 1$.*
2. *For $k = 2, 3, ..., d$, choose $g_k \in \{1, ..., N-1\}$ that minimizes $R_N((g_1, ..., g_{k-1}, z))$ as a function of $z \in \{1, ..., N-1\}$,*

where

$$R_N(\mathbf{g}) := \sum_{\mathbf{h} \in C_d^*(N) \cap \mathcal{L}(\mathbf{g}, N)} \frac{1}{r_1(\mathbf{h})},$$

$$r_1(\mathbf{h}) = \prod_{i=1}^{d} r_1(h_i),$$

$$r_1(h) := \max\{1, |h|\},$$

$$C_d^*(N) := [(-N/2, N/2] \cap \mathbb{Z}]^d \setminus \{0\}$$

and

$$\mathcal{L}(\mathbf{g}, N) := \{\mathbf{h} \in \mathbb{Z}^d : \mathbf{h}\cdot\mathbf{g} \equiv 0 \pmod{N}\}$$

is the *dual lattice* of the lattice point set $\mathcal{P}(\mathbf{g}, N)$ (Leobacher y Pillichshammer [2014]).

Having said that, several QMC constructions exist for which the optimal rate of convergence $\mathcal{O}(N^{-1}\log(N)^{1/2})$ yields asymptotically, but it's yet unknown what the optimal point set is (in the sense that it's a global minimizer of the worst case integration error). However, for small Fibonacci numbers N, it is known that the unique minimizer of the QMC worst-case error in the reproducing kernel Hilbert space H^1_{mix} of 1-periodic functions with mixed smoothness is the Fibonacci lattice. What's more, for $N = 1, 2, 3, 5, 7, 8, 12, 13$ the optimal point sets are integration lattices (Hinrichs y Oettershagen [2016]).

2.5 Optimal Point Sets for QMC Integration

Let's consider the reproducing kernel Hilbert space H^1_{mix} of 1-periodic functions with mixed smoothness. It's reproducing kernel is a tensor product kernel that reads

$$K_{d,\gamma}(\mathbf{x}, \mathbf{y}) := \prod_{i=1}^{d} K_{1,\gamma}(x_i, y_i),$$

where $\mathbf{x} = (x_1, ..., x_d), \mathbf{y} = (y_1, ..., y_d) \in [0,1)^d$,

$$K_{1,\gamma}(x,y) = 1 + \gamma k(|x-y|),$$

$$k(t) = \frac{1}{2}(t^2 - t + \frac{1}{6})$$

and $\gamma > 0$ is a parameter.

As a result, if we consider a point set $\mathcal{P} = \{\mathbf{x}_0, ..., \mathbf{x}_{N-1}\}$, then minimizing the worst-case error $e(H^1_{mix}, \mathcal{P})$ with respect to the Hilbert space norm corresponding to the reproducing kernel $K_{d,\gamma}$ is equivalent to minimizing the double sum

$$G_\gamma(\mathbf{x}_0, ..., \mathbf{x}_{N-1}) = \sum_{n,m=0}^{N-1} K_{d,\gamma}(\mathbf{x}_n, \mathbf{x}_m).$$

It is known that the Fibonacci lattice yields the optimal point configuration for integrating periodic functions, which may suggest that the optimal point configurations are integration lattices or at least lattice point sets, however, that's not always true. Nonetheless, integration lattices are always local minima of $e(H^1_{mix}, \mathcal{P})$. What's more, for small γ values the optimal points are always close to a lattice point set, which for $d = 2$ gives point sets of the form

$$\left\{ \left(\frac{i}{N}, \frac{\sigma(i)}{N} \right) : i = 0, ..., N-1 \right\},$$

where σ is a permutation of $\{0, 1, ..., N-1\}$.

In order to obtain these optimal point sets various considerations have to be contemplated. First, lets consider univariate 1-periodic functions $f : \mathbb{R} \longrightarrow \mathbb{R}$ given by their values on the torus $\mathbb{T} = [0,1)$.

Definition 7. *Let $k \in \mathbb{Z}$, the k-th Fourier coefficient of a function $f \in L_2(\mathbb{T})$ is defined as:*

$$\widehat{f}_k = \int_0^1 f(x) \cdot e^{2\pi i k x} \, \mathrm{d}x$$

In the univariate *Sobolev space* $H^1(\mathbb{T}) = W^{1,2}(\mathbb{T}) \subset L_2(\mathbb{T})$ of functions with first weak derivatives bounded in L_2, the following definition gives a Hilbert space norm on $H^1(\mathbb{T})$ depending on $\gamma > 0$,

$$\|f\|^2_{H^{1,\gamma}(\mathbb{T})} = \widehat{f}_0^2 + \gamma \cdot \sum_{k \in \mathbb{Z}} |2\pi k|^2 \widehat{f}_k^2 = \left(\int_{\mathbb{T}} f(x) \, \mathrm{d}x \right)^2 + \gamma \cdot \int_{\mathbb{T}} f'(x)^2 \, \mathrm{d}x,$$

where $f \in H^1(\mathbb{T})$ is a function.

Moreover, the corresponding inner product is defined as

$$\langle f, g \rangle_{H^{1,\gamma}(\mathbb{T})} = \left(\int_0^1 f(x) \, \mathrm{d}x \right) \cdot \left(\int_0^1 g(x) \, \mathrm{d}x \right) + \gamma \cdot \int_0^1 f'(x) \cdot g'(x) \, \mathrm{d}x$$

where $H^{1,\gamma}(\mathbb{T})$ denotes the Hilbert space $H^1(\mathbb{T})$ equipped with this inner product.

What's more, $H^{1,\gamma}(\mathbb{T})$ is continuously embedded in $C^0(\mathbb{T})$, thus, $H^{1,\gamma}(\mathbb{T})$ is a reproducing kernel Hilbert space whose reproducing kernel $K_{1,\gamma} : \mathbb{T} \times \mathbb{T} \longrightarrow \mathbb{R}$ is *positive definite* and reads

$$K_{1,\gamma}(x,y) := 1 + \gamma \cdot \sum_{k \in \mathbb{Z} \setminus \{0\}} |2\pi k|^{-2} \cdot e^{2\pi i k (x-y)} = 1 + \gamma \cdot k(|(x-y)|),$$

where $k(t) = \frac{1}{2}(t^2 - t + \frac{1}{6})$.

Also, this kernel has the following property,

$$f(x) = \langle f(\cdot), K(\cdot, x)\rangle_{H^{1,\gamma(\mathbb{T})}}, \ \ \forall f \in H^1(\mathbb{T}).$$

That being said, the tensor product space $H^{1,\gamma}_{mix}(\mathbb{T}^2) := H^1(\mathbb{T}) \otimes H^1(\mathbb{T}) \subset C(\mathbb{T}^2)$ has a reproducing kernel given by

$$K_{2,\gamma}(\mathbf{x}, \mathbf{y}) = K_{1,\gamma}(x_1, y_1) \cdot K_{1,\gamma}(x_2, y_2) = 1 + \gamma \cdot k(|x_1 - y_1|) + \gamma \cdot k(|x_2 - y_2|) + \gamma^2 \cdot k(|x_1 - y_1|) \cdot k(|x_2 - y_2|)$$

Furthermore, the squared worst-case error of a QMC rule based on a point set $\mathcal{P} = \{\mathbf{x}_0, ..., \mathbf{x}_{N-1}\}$ in $H^{1,\gamma}_{mix}(\mathbb{T}^2)$ is

$$e(H^{1,\gamma}_{mix}(\mathbb{T}^2), \mathcal{P})^2 = -1 + \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} K_{2,\gamma}(\mathbf{x}_n, \mathbf{x}_m).$$

In order to acquire an optimal point set, we have to minimize the worst-case error.

Note: So as to not confuse elements $\mathbf{x}_n \in \mathcal{P}$ with their coordinates $\mathbf{x}_n = (x_{n,1}, x_{n,2})$ and simplify the notation, a given point $\mathbf{x}_n \in \mathcal{P}$ will have coordinates $\mathbf{x}_n = (x_n, y_n)$, that is, the first coordinate of a point in $\mathcal{P}$ will be denoted with an 'x', whereby the second coordinate will be denoted with a 'y'.

For the squared worst-case error we have

$$e(H^{1,\gamma}_{mix}(\mathbb{T}^2), \mathcal{P})^2 = -1 + \frac{1}{N^2} \sum_{n,m=0}^{N-1} K_{2,\gamma}(\mathbf{x}_n, \mathbf{x}_m) = -1 + \frac{1}{N^2} \sum_{n,m=0}^{N-1} K_{1,\gamma}(x_n, x_m) K_{1,\gamma}(y_n, y_m)$$

$$= -1 + \frac{1}{N^2} \sum_{n,m=0}^{N-1} (1 + \gamma \cdot k(|x_n - x_m|) + \gamma \cdot k(|y_n - y_m|) + \gamma^2 \cdot k(|x_n - x_m|) \cdot k(|y_n - y_m|))$$

$$= \frac{\gamma}{N^2} \cdot \sum_{n,m=0}^{N-1} (k(|x_n - x_m|) + k(|y_n - y_m|) + \gamma \cdot k(|x_n - x_m|) \cdot k(|y_n - y_m|))$$

$$= \frac{\gamma \cdot (2 \cdot k(0) + \gamma \cdot k(0)^2)}{N} + \frac{2\gamma}{N^2} \cdot \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} (k(|x_n - x_m|) + k(|y_n - y_m|) + \gamma \cdot k(|x_n - x_m|) \cdot k(|y_n - y_m|))$$

Therefore, minimizing the worst-case error is equivalent to minimizing either

$$F_\gamma(\mathbf{x}, \mathbf{y}) := \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} (k(|x_n - x_m|) + k(|y_n - y_m|) + \gamma \cdot k(|x_n - x_m|) \cdot k(|y_n - y_m|))$$

or

$$G_\gamma(\mathbf{x}, \mathbf{y}) := \sum_{n,m=0}^{N-1} (1 + \gamma \cdot k(|x_n - x_m|)) \cdot (1 + \gamma \cdot k(|y_n - y_m|)).$$

G_γ is sometimes used for theoretical considerations, however, F_γ has less summands, thus it's better for numerical implementation.

Now, let $\tau, \sigma \in S_N$ be two permutations of $\{0, 1, ..., N-1\}$ and define the set

$$D_{\tau,\sigma} = \{\mathbf{x} \in [0,1)^N, \mathbf{y} \in [0,1)^N : x_{\tau(0)} \leq x_{\tau(1)} \leq ... \leq x_{\tau(N-1)}, \ y_{\sigma(0)} \leq y_{\sigma(1)} \leq ... \leq y_{\sigma(N-1)}\}$$

on which $|x_n - x_m| = s_{n,m}(x_n - x_m)$ holds for $s_{n,m} \in \{-1, 1\}$. It turns out that the restriction of F_γ to $D_{\tau,\sigma}$, namely $F_\gamma(\mathbf{x}, \mathbf{y})_{|D_{\tau,\sigma}}$, is a convex polynomial of degree 4 in $(\mathbf{x}, \mathbf{y})$ for sufficiently small γ.

Proposition 3. *$F_\gamma(\boldsymbol{x},\boldsymbol{y})_{|D_{\tau,\sigma}}$ and $G_\gamma(\boldsymbol{x},\boldsymbol{y})_{|D_{\tau,\sigma}}$ are convex if $\gamma \in [0,6]$.*

Proof. Because of how $F_\gamma(\mathbf{x},\mathbf{y})_{|D_{\tau,\sigma}}$ and $G_\gamma(\mathbf{x},\mathbf{y})_{|D_{\tau,\sigma}}$ have been defined, we only need to prove the claim above for one of them. Lets prove it for $G_\gamma(\mathbf{x},\mathbf{y})_{|D_{\tau,\sigma}}$.

The sum of convex functions is convex and $f(x-y)$ is convex if f is convex, thus, we have to prove that $f(r,s) = (1+\gamma\cdot k(r))\cdot(1+\gamma\cdot k(s))$ is convex for $r,s\in[0,1]$. To prove it, we will see if the Hesse matrix $\mathcal{H}(f)$ is positive definite for $0\le\gamma<6$. We have $k'(r) = (2r-1)/2$ and $k''(r)=1$, thus $f_{rr}(r,s) = \gamma\cdot(1+\gamma\cdot k(s))$ is positive if $\gamma<24$ because $\min_{s\in[0,1]} k(s) = -1/24$. Therefore, it's enough to check if $\mathcal{H}(f)$ is positive. Since $f_{rs} = \gamma^2\cdot(2r-1)/2\cdot(2s-1)/2$, we have

$$\begin{aligned}|\ \mathcal{H}(f)\ | &= (1+\gamma\cdot k(r))\cdot(1+\gamma\cdot k(s)) - \gamma^2\cdot\left(r-\frac{1}{2}\right)^2\cdot\left(s-\frac{1}{2}\right)^2 > 0\\ &\iff (1+\gamma\cdot k(r))\cdot(1+\gamma\cdot k(s)) > \gamma^2\cdot\left(r-\frac{1}{2}\right)^2\cdot\left(s-\frac{1}{2}\right)^2.\end{aligned}$$

Hence we can check if

$$1+\gamma\cdot k(r) = 1+\frac{\gamma}{2}\cdot\left(r^2-r+\frac{1}{6}\right) > \gamma\cdot\left(r-\frac{1}{2}\right)^2,$$

which is the case for $0\le\gamma\le 6$ and $r\in[0,1]$. For $\gamma=6$ extra arguments are necessary, so we omit this case. □

What's more, since

$$[0,1)^N\times[0,1)^N = \cup_{(\tau,\sigma)\in S_N\times S_N} D_{\tau,\sigma},$$

the global minimum of F_γ on $[0,1)^N\times[0,1)^N$ can be obtained by computing

$$\arg\min_{(\mathbf{x},\mathbf{y})\in D_{\tau,\sigma}} F_\gamma(\mathbf{x},\mathbf{y})$$

for all $(\tau,\sigma)\in S_N\times S_N$ and choosing the smallest of all the local minima.

Also, symmetries of the torus $\mathbb{T}^2$ allow to reduce the number of regions $D_{\tau,\sigma}$ for which the optimization problem has to be solved. However, these symmetries don't change the worst-case error for the considered classes of periodic functions. Since implementing all of the torus symmetries is difficult, only some of them have been considered. Nevertheless, using the torus symmetries it can always be arranged that $\tau = id$ and $\sigma(0)=0$ together with fixing the point $\mathbf{x}_0 = (x_0,y_0) = (0,0)$ leads to the sets

$$D_\sigma = \{\mathbf{x}\in[0,1)^N, \mathbf{y}\in[0,1)^N : 0 = x_0\le x_1\le ...\le x_{N-1},\ 0 = y_0\le y_{\sigma(1)}\le ...\le y_{\sigma(N-1)}\},$$

where $\sigma\in S_{N-1}$ denotes a permutation of $\{1,2,...,N-1\}$.

Lets define the symmetrized metric

$$d(n,m) := \min\{|n-m|, N-|n-m|\}\quad \text{for } 0\le n,m\le N-1$$

Definition 8. *We define the set of semi-canonical permutations as a set $\mathcal{C}_N\subset S_N$ whose permutations σ satisfy*

1. $\sigma(0)=0$
2. $d(\sigma(1),\sigma(2))\le d(0,\sigma(N-1))$

3. $\sigma(1) = \min\{d(\sigma(n), \sigma(n+1)) \mid n = 0, 1, ..., N-1\}$

4. σ *is lexicographically smaller than* σ^{-1}

$\sigma(N)$ *is identified with* $0 = \sigma(0)$.

As a result, we get the following lemma.

Lemma 2. *For any* $\sigma \in S_N$ *such that* $\sigma(0) = 0$*, there exists a semi-canonical permutation* σ' *such that* D_σ *and* $D_{\sigma'}$ *are equivalent (up to torus symmetry).*

Thus, only semi-canonical σ need to be considered.

Moreover, considering the objective function only in domains D_σ makes it strictly convex if $\mathbf{x}_0 = (x_0, y_0) = (0, 0)$.

Proposition 4. $F_\gamma(\boldsymbol{x}, \boldsymbol{y})_{|D_\sigma}$ *and* $G_\gamma(\boldsymbol{x}, \boldsymbol{y})_{|D_\sigma}$ *are strictly convex if* $\gamma \in [0, 6]$.

Proof. Similarly to the last Proposition, it's enough to prove the result for $G_\gamma(\mathbf{x}, \mathbf{y})_{|D_\sigma}$.

The sum of a convex and a strictly convex function is strictly convex, thus we only need to check if

$$f(x_1, ..., x_{N-1}, y_1, ..., y_{N-1}) = \sum_{n=1}^{N-1} (1 + \gamma \cdot k(|x_n - x_0|))(1 + \gamma \cdot k(|y_n - y_0|))$$
$$= \sum_{n=1}^{N-1} (1 + \gamma \cdot k(x_n))(1 + \gamma \cdot k(y_n))$$

is a strictly convex function on $[0, 1]^{N-1} \times [0, 1]^{N-1}$. In the last Proposition we actually proved that $f_n(x_n, y_n) = (1 + \gamma \cdot k(x_n))(1 + \gamma \cdot k(y_n))$ is strictly convex if $(x_n, y_n) \in [0, 1]^2$ for a fixed n. Therefore, the strict convexity of f follows from the following lemma. □

Lemma 3. *Let* $D_i \in \mathbb{R}^{d_i}$ *be convex domains and let* $f_i : D_i \longrightarrow \mathbb{R}$, $i = 1, ..., p$ *be strictly convex functions. Then*

$$\begin{array}{rccc} f: & D = D_1 \times ... \times D_p & \longrightarrow & \mathbb{R} \\ & (z_1, ..., z_p) & \longmapsto & \sum_{i=1}^{p} f_i(z_i) \end{array}$$

is strictly convex.

Whereby each D_σ has a unique point where the minimum of F_γ is attained.

That being said, we want to find the smallest of all the local minima of F_γ on each region $D_\sigma \subset [0, 1)^N \times [0, 1)^N$ for all semi-canonical permutations $\sigma \in \mathcal{C}_N \subset S_N$ to determine the global minimum. That gives for each $\sigma \in \mathcal{C}_N$ the following constrained optimization problem,

$$\min_{(\mathbf{x},\mathbf{y}) \in D_\sigma} F_\gamma(\mathbf{x}, \mathbf{y}) \text{ subject to } v_n(\mathbf{x}) \geq 0 \text{ and } w_n(\mathbf{y}) \geq 0, \ \forall n = 1, ..., N-1,$$

where

$$v_n(\mathbf{x}) = x_n - x_{n-1} \text{ and } w_n(\mathbf{y}) = y_{\sigma(n)} - y_{\sigma(n-1)} \text{ for } n = 1, ..., N-1.$$

What's more, we know that the necessary and sufficient (due to local strict convexity) conditions for $(\mathbf{x}, \mathbf{y}) \in D_\sigma$ to be local minima are

$$\frac{\partial}{\partial x_k} F_\gamma(\mathbf{x}, \mathbf{y}) = 0 \ \ \forall k = 1, ..., N-1$$

and

$$\frac{\partial}{\partial y_k} F_\gamma(\mathbf{x}, \mathbf{y}) = 0 \ \ \forall k = 1, ..., N-1,$$

where the partial derivatives of F_γ are given by the formulas from the following proposition.

Proposition 5. *Given a permutation $\sigma \in \mathcal{C}_N$, the partial derivative of $F_{\gamma|D_\sigma}$ with respect to the second component $\boldsymbol{y}$ is given by*

$$\frac{\partial}{\partial y_k} F_\gamma(\boldsymbol{x}, \boldsymbol{y})_{|D_\sigma} = y_k \left(\sum_{n=0, n\neq k}^{N-1} c_{n,k} \right) - \sum_{n=0, n\neq k}^{N-1} c_{n,k} \cdot y_n + \frac{1}{2} \cdot \left(\sum_{n=0}^{k-1} c_{n,k} \cdot s_{n,k} - \sum_{m=k+1}^{N-1} c_{k,m} \cdot s_{k,m} \right),$$

where $s_{n,m} := sign(y_n - y_m)$ and $c_{n,m} := 1 + \gamma \cdot k(|x_n - x_m|) = c_{m,n}$. The analogue for the partial derivatives with respect to $\boldsymbol{x}$ is obtained by interchanging $\boldsymbol{x}$ with $\boldsymbol{y}$ and using $c_{n,m} = 1 + \gamma \cdot k(|y_n - y_m|)$ and $s_{n,m} = -1$. The second order derivatives with respect to $\boldsymbol{y}$ are given by

$$\frac{\partial^2}{\partial y_k \partial y_j} F_\gamma(\boldsymbol{x}, \boldsymbol{y})_{|D_\sigma} = \begin{cases} \sum_{n=0}^{k-1} c_{n,k} + \sum_{n=k+1}^{N-1} c_{n,k} & \text{for } j = k \\ -c_{k,j} & \text{for } j \neq k, \end{cases} \quad k, j \in \{1, ..., N-1\}.$$

The analogue for $\frac{\partial^2}{\partial x_k \partial x_j} F_\gamma(\boldsymbol{x}, \boldsymbol{y})_{|D_\sigma}$ is obtained by interchanging $\boldsymbol{x}$ with $\boldsymbol{y}$ and using $c_{n,m} = 1 + \gamma \cdot k(|y_n - y_m|)$.

Proof. The partial derivative of $F_{\gamma|D_\sigma}$ with respect to $\mathbf{y}$ is

$$\begin{aligned}
\frac{\partial}{\partial y_k} F_\gamma(\mathbf{x}, \mathbf{y})_{|D_\sigma} &= \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} \frac{\partial}{\partial y_k} k(|y_n - y_m|)(1 + \gamma \cdot k(|x_n - x_m|)) + \frac{\partial}{\partial y_k} k(|x_n - x_m|) \\
&= \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} \frac{\partial}{\partial y_k} k(|y_n - y_m|) c_{n,m} + \frac{\partial}{\partial y_k} k(|x_n - x_m|) = \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} c_{n,m} \frac{\partial}{\partial y_k} k(|y_n - y_m|) \\
&= \sum_{n=0}^{N-2} \sum_{m=n+1}^{N-1} c_{n,m} k'(s_{n,m}(y_n - y_m)) \cdot \begin{cases} s_{n,m} & \text{for } n = k \\ -s_{n,m} & \text{for } m = k \\ 0 & \text{else} \end{cases} \\
&= \sum_{m=k+1}^{N-1} c_{k,m} s_{k,m} \left(s_{k,m}(y_k - y_m) - \frac{1}{2} \right) - \sum_{n=0}^{k-1} c_{n,k} s_{n,k} \left(s_{n,k}(y_n - y_k) - \frac{1}{2} \right) \\
&= y_k \left(\sum_{n=0, n\neq k}^{N-1} c_{n,k} \right) - \sum_{n=0, n\neq k}^{N-1} c_{n,k} y_n + \frac{1}{2} \left(\sum_{n=0}^{k+1} c_{n,k} s_{n,k} - \sum_{m=k+1}^{N-1} c_{k,m} s_{k,m} \right).
\end{aligned}$$

From this expression the second derivative of $F_{\gamma|D_\sigma}$ with respect to $\mathbf{y}$ is trivially obtained. The derivatives for $\mathbf{x}$ are obtained analogously. □

Now we can approximate local minima of F_γ on a given D_σ. Doing this for all $\sigma \in \mathcal{C}$ we can obtain a candidate for the global minimum, however, due to the finite precision of floating point arithmetic we can't be sure to be close to the actual global minimum. Luckily, it is possible to compute a lower

bound for the optimal point set for each D_σ. This is because a lower bound on a function F can be obtained for convex problems with linear inequality constraints using the *Lagrangian*

$$\mathcal{L}_F(\mathbf{x},\mathbf{y},\boldsymbol{\lambda},\boldsymbol{\mu}) \;:=\; F(\mathbf{x},\mathbf{y}) - \boldsymbol{\lambda}^T\cdot\mathbf{v}(\mathbf{x}) - \boldsymbol{\mu}^T\cdot\mathbf{w}(\mathbf{y}) \;=\; F(\mathbf{x},\mathbf{y}) - \sum_{n=1}^{N-1}(\lambda_n\cdot v_n(\mathbf{x}) + \mu_n\cdot w_n(\mathbf{y})).$$

As a result, we get

$$\min_{(\mathbf{x},\mathbf{y})\in D_\sigma} F(\mathbf{x},\mathbf{y}) \geq \mathcal{L}_F(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\boldsymbol{\lambda},\boldsymbol{\mu})$$

for all $(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\boldsymbol{\lambda},\boldsymbol{\mu})$ that satisfy

$$\nabla_{(\mathbf{x},\mathbf{y})}\mathcal{L}_F(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\boldsymbol{\lambda},\boldsymbol{\mu}) = 0 \;\text{ with } \boldsymbol{\lambda},\boldsymbol{\mu}\geq 0,$$

where $\nabla_{(\mathbf{x},\mathbf{y})} = (\nabla_{\mathbf{x}},\nabla_{\mathbf{y}})$ and $\nabla_{\mathbf{x}}$ denotes the gradient of a functions with respect to the variables in $\mathbf{x}$.

Therefore, we have to find an admissible point $(\tilde{\mathbf{x}},\tilde{\mathbf{y}},\boldsymbol{\lambda},\boldsymbol{\mu})$ for each D_σ which yields a lower bound that is larger than the candidate for the global minimum. The following theorem is also relevant.

Theorem 3. *For $\sigma\in\mathcal{C}_N$ and $\delta>0$ let $(\tilde{\boldsymbol{x}}_\sigma,\tilde{\boldsymbol{y}}_\sigma)\in D_\sigma$ be a point such that*

$$\frac{\partial}{\partial x_k}F(\tilde{\boldsymbol{x}}_\sigma,\tilde{\boldsymbol{y}}_\sigma) = \delta \;\; \forall k = 1,...,N-1 \tag{11}$$

and

$$\frac{\partial}{\partial y_k}F(\tilde{\boldsymbol{x}}_\sigma,\tilde{\boldsymbol{y}}_\sigma) = \delta \;\; \forall k = 1,...,N-1. \tag{12}$$

Then the following inequality holds,

$$F(\boldsymbol{x},\boldsymbol{y}) \geq F(\tilde{\boldsymbol{x}}_\sigma,\tilde{\boldsymbol{y}}_\sigma) - \delta\cdot\sum_{n=1}^{N-1}((N-n)\cdot v_n(\tilde{\boldsymbol{x}}_\sigma) + \sigma(N-n)\cdot w_n(\tilde{\boldsymbol{y}}_\sigma))$$
$$> F(\tilde{\boldsymbol{x}}_\sigma,\tilde{\boldsymbol{y}}_\sigma) - \delta\cdot N^2 \;\; \forall(\boldsymbol{x},\boldsymbol{y})\in D_\sigma. \tag{13}$$

Proof. Let $\mathbf{P}_\sigma\in\{-1,0,1\}^{(N-1)\times(N-1)}$ be the permutation matrix associated to $\sigma\in S_{N-1}$ and let

$$\mathbf{B} := \begin{pmatrix} 1 & -1 & 0 & ... & 0 & 0 \\ 0 & 1 & -1 & ... & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & & ... & 0 & 1 & -1 \\ 0 & & ... & & 0 & 1 \end{pmatrix} \in \mathbb{R}^{(N-1)\times(N-1)}.$$

The partial derivatives of $\mathcal{L}_F$ with respect to $\mathbf{x}$ and $\mathbf{y}$ are given by

$$\nabla_{\mathbf{x}}\mathcal{L}_F(\mathbf{x},\mathbf{y},\boldsymbol{\lambda},\boldsymbol{\mu}) = \nabla_{\mathbf{x}}F(\mathbf{x},\mathbf{y}) - \begin{pmatrix} \lambda_1-\lambda_2 \\ \vdots \\ \lambda_{N-2}-\lambda_{N-1} \\ \lambda_{N-1} \end{pmatrix} = \nabla_{\mathbf{x}}F(\mathbf{x},\mathbf{y}) - \mathbf{B}\boldsymbol{\lambda}$$

and

$$\nabla_{\mathbf{y}}\mathcal{L}_F(\mathbf{x},\mathbf{y},\boldsymbol{\lambda},\boldsymbol{\mu}) = \nabla_{\mathbf{y}}F(\mathbf{x},\mathbf{y}) - \begin{pmatrix} \mu_{\sigma(1)}-\mu_{\sigma(2)} \\ \vdots \\ \mu_{\sigma(N-2)}-\mu_{\sigma(N-1)} \\ \mu_{\sigma(N-1)} \end{pmatrix} = \nabla_{\mathbf{y}}F(\mathbf{x},\mathbf{y}) - \mathbf{B}\mathbf{P}_\sigma\boldsymbol{\mu}.$$

Therefore, by choosing

$$\boldsymbol{\lambda} = \mathbf{B}^{-1}\nabla_{\mathbf{x}}F(\tilde{\mathbf{x}}_\sigma, \tilde{\mathbf{y}}_\sigma) \quad \text{and} \quad \boldsymbol{\mu} = \mathbf{P}_\sigma^{-1}\mathbf{B}^{-1}\nabla_{\mathbf{y}}F(\tilde{\mathbf{x}}_\sigma, \tilde{\mathbf{y}}_\sigma)$$

we get

$$\nabla_{\mathbf{x}}F(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \mathbf{B}\boldsymbol{\lambda} \quad \text{and} \quad \nabla_{\mathbf{y}}F(\tilde{\mathbf{x}}, \tilde{\mathbf{y}}) = \mathbf{B}\mathbf{P}_\sigma\boldsymbol{\mu}.$$

Since

$$\mathbf{B}^{-1} := \begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 1 \\ \vdots & 0 & \ddots & \vdots \\ 0 & \dots & 0 & 1 \end{pmatrix} \in \mathbb{R}^{(N-1)\times(N-1)},$$

it turns out that $\mathbf{y}, \boldsymbol{\lambda} > 0$, which together with *Wolfe duality* gives the first inequality of equation (13). In order to get the second inequality it has to be noted that $0 \leq |v_n(\mathbf{x})|, |w_n(\mathbf{y})| \leq 1$ and that $2 \cdot \sum_{n=1}^{N-1} \sigma(N-n) = 2 \cdot \sum_{n=1}^{N-1} n = (N-1)(N-2) < N^2$. □

Having said that, suppose that $(\mathbf{x}^*, \mathbf{y}^*) \in D_{\sigma^*}$ is a candidate for an optimal point set. If we can find points $(\tilde{\mathbf{x}}_\sigma, \tilde{\mathbf{y}}_\sigma) \in D_\sigma$ that fulfill equations (11) and (12) for all other $\sigma \in \mathcal{C}_N$ and

$$F(\tilde{\mathbf{x}}_\sigma, \tilde{\mathbf{y}}_\sigma) - \delta N^2 \geq F_\gamma(\mathbf{x}^*, \mathbf{y}^*) \text{ for } \delta > 0,$$

we can be sure that D_{σ^*} is the unique domain D_σ that contains the globally optimal point set.

In conclusion, to find the global minimum $(\mathbf{x}^*, \mathbf{y}^*)$ of F_γ, one has to first compute

$$\sigma^* := \arg\min_{\sigma\in\mathcal{C}_N} \min_{(\mathbf{x},\mathbf{y})\in D_\sigma} F_\gamma(\mathbf{x}, \mathbf{y}),$$

in order to obtain a candidate point set $(\mathbf{x}^*, \mathbf{y}^*) \in D_{\sigma^*}$ for the global minimum of F_γ. After that, one has to compute lower bounds for all the other domains D_σ with $\sigma \in \mathcal{C}_N$. That way, if we obtain for each σ a point $(\tilde{\mathbf{x}}_\sigma, \tilde{\mathbf{y}}_\sigma)$ such that

$$\min_{(\mathbf{x},\mathbf{y})\in D_{\sigma^*}} F_\gamma(\mathbf{x}, \mathbf{y}) \approx \theta_N := F_\gamma(\mathbf{x}^*, \mathbf{y}^*) < \mathcal{L}_F(\tilde{\mathbf{x}}_\sigma, \tilde{\mathbf{y}}_\sigma) - 2N^2\delta \leq F_\gamma(\mathbf{x}, \mathbf{y}),$$

we could be sure that the global optimum is indeed located in D_{σ^*} and that $(\mathbf{x}^*, \mathbf{y}^*)$ is a good approximation to it (Hinrichs y Oettershagen [2016]).

The optimal point sets for a parameter value $\gamma = 1$ that have been obtained by Hinrichs y Oettershagen [2016] using this method are shown in Figure 21. In this work we will use those point sets so as to test the efficiency of the new method.

3 Experimental results

Given the results mentioned in Section 2, in this work we have tried to use the optimal point sets for QMC integration in order to compute the best possible estimators for different examples of the two stereology problems described in Section 1. We have used point sets whose cardinality goes from $N = 3, ..., 16$, but only the results for $N = 5, ..., 16$ have been presented here. Also, the results are expressed in number or length in $\mathbb{R}^2$, including the results that can be compared to the ones from other works. As a result, the direct comparison between the methods used in this work and in the other works is done via the *sample squared coefficient of error*.

Notes:

- The sample squared coefficient of error for an estimated value $\widehat{X}$, reads

$$ce(\widehat{X}) = \frac{\widehat{Var(\widehat{X})}}{\widehat{X}^2} \tag{14}$$

  where $\widehat{Var(\widehat{X})}$ is the error variance predictor. For the computed results however, we have used the sample variance instead of the error variance predictor, thus substituting $\widehat{Var(\widehat{X})}$ for $Var(\widehat{X})$. This is because several computations for each experiment are done, thus, there's no need for a predictor and the sample variance is easier to implement.

- We have decided to use the sample squared coefficient or error instead of the regular squared coefficient of error (in which the denominator of the preceding expression is the squared real value to be estimated, X^2) because the estimators used are known to work well, meaning that if they apparently don't give a sufficiently good estimator at least we can check whether the error variance is small (in which case the procedure might be good but the estimation is not implemented correctly).

3.1 Test system of quadrats

In this subsection we will discuss the results obtained for the estimated number of points $\widehat{N}$ that appear in different photos, but first we will explain the procedure used to count the number of points captured by the test system of quadrats needed for equation (5). To this end, we used a *Scipy* function called *scipy.spatial.KDTree.query_ball_point* from which we needed to specify the parameters C, point or points to search neighbours of, r, the radius of points to return, p, which indicates what *Minkowski p-norm* is used and *return_Length* which will be set to *True* so the function returns the number of points inside the radius. Because of how this function is implemented, in order to "acquire a test system of quadrats" we have to set $p = \infty$, $r = L/2$, where L is the length of one side of the quadrat, and $C = \{c \in \mathbb{R}^2 : \text{c is the center point of a quadrat from the test system}\}$. Now, similarly to what will happen with the Buffon and Buffon-Steinhaus test systems, we don't *create* the test system, but rather check if the points in $\mathbb{R}^2$ from the photos are contained in certain subspaces of $\mathbb{R}^2$. Furthermore, the *Scipy* function mentioned above uses what's called *KDTrees*, that is, binary trees where every node is a k-dimensional point, which provides more efficiency.

Having said that, lets go over the creation of the set of points C used in the aforementioned function. Considering the information available in Section 1, we need a test system of quadrats whose fundamental probe is a quadrat with side length t and area $a_0 = t^2$ and whose fundamental tile is a square with side length T and area $a = T^2$ containing the quadrat, therefore, contiguous points are separated a distance T apart. So as to change the position and orientation of the whole test system, we attach what's called an *associated vector* (AV) $(\widehat{x}, \omega)$ to an *associated point* (AP) $\widehat{x} = (x_1, x_2) \in \mathbb{R}^2$ of the test system. Since the test system will be "superimposed onto a photo", it will be contained inside a rectangle $[0, a] \times [0, b]$, thus we will take the point of the set C with lower values of the first and second coordinates as the AP of the test system. That way, the whole system moves depending on $\widehat{x}$ and rotates an angle $\omega \in [0, 2\pi)$ with respect to the horizontal axis. However, the AP $\widehat{x}$ will be randomly located inside the rectangle $[0, a] \times [0, b]$, whereby in order to maintain all the points from C inside the rectangle, we will apply *modulo* a to the first coordinate of the points in C and *modulo* b to the second coordinate of the points in C. With respect to the rotation of the test system, for simplicity we will instead rotate the points in the photo. Since this rotation can modify the rectangle $[0, a] \times [0, b]$, all the points in the photo will be moved on the horizontal axis so that we get a new rectangle $[0, A] \times [0, B]$ containing all the points from the photo.

Finally, we will explain how we implemented the optimal point sets for QMC integration in the code for this subsection. Each test system created for the code has an AP $\widehat{x}$ with AV $(\widehat{x}, \omega) \equiv (x_1, x_2, \omega) \equiv (x, y, \omega)$. Also, considering equation (9), we have

$$\int_{\mathbb{R}^2} N(Y \cap \Lambda_{\mathbf{x}})\, \mathrm{d}\mathbf{x} \approx \frac{1}{N} \sum_{n=0}^{N-1} N(Y \cap \Lambda_{\mathbf{x}_n}),$$

where $\Lambda_{\mathbf{x}}$ is the test system of quadrats and $\mathbf{x}_n$, $n = 0..., N-1$ are the optimal points for QMC integration. As a result, what we will do is calculate the number of points captured by the test system of quadrats when the AP $\widehat{x} = (x, y)$ is located in each of the *Hinrichs' optimal points* for bivariate periodic functions (namely the optimal points for QMC integration obtained in Hinrichs y Oettershagen [2016], see Figure 21 in the appendix). After that, those numbers calculated for each test system will be added and divided by the total number of Hinrichs' optimal points used, N. In consequence, this should give us the value $\mathbb{E}[N(Y \cap \Lambda_{\mathbf{x}})]$ with minimum worst case error. From now on, we will refer to this method as *Hinrichs* (x, y).

What's more, we have also tried a different approach to this method. Apart from using the Hinrichs' optimal points in the AP $\widehat{x} = (x, y)$, we have also computed results where the first coordinate x and the angle ω are modified by the optimal points for each system, that is, instead of locating the test system in the AP $\widehat{x} = (x, y)$ with (x, y) being optimal points, we have located the test system in the AP $\widehat{x} = (x, y)$ with x being the first coordinate of the optimal point and rotated it an angle $\omega + h \pmod{\pi}$ with h being the second coordinate of the optimal point. From now on, we will refer to this method as *Hinrichs* (x, ω).

Note: The Hinrichs' optimal points in Figure 21 are shown for the space $[0, 1] \times [0, 1]$. In order to use them for this work we escalated them depending on the method used. For example, for Hinrichs (x, y) we use the space $[0, T] \times [0, T]$ to locate the AP because of the symmetry of the test system of quadrats, thus the optimal points p_n, $n = 0, ..., N-1$ from Figure 21 are multiplied by T. On the other hand, for Hinrichs (x, ω) we use the space $[0, T] \times [0, \pi]$, thus the first coordinate of the optimal points p_n, $n = 0, ..., N-1$ is multiplied by T and the second coordinate is multiplied by π.

We will now resume the computed results acquired for the estimated number of points $\widehat{N}$ that appear in different photos and show how these methods' variances behave (via the squared coefficient of error (see equation (14))) compared to other methods. The other methods that have been included in order to test the behaviour of the new methods developed in this work are:

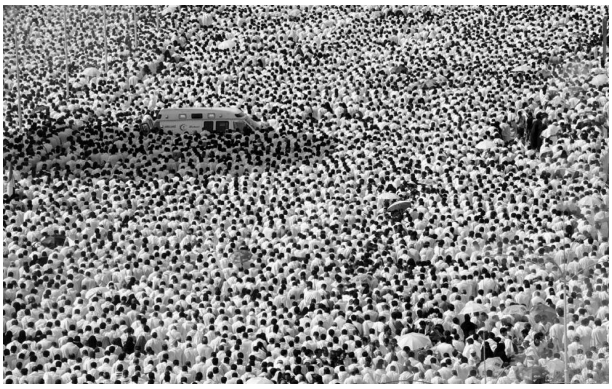
- Original method: the basic process of superimposing the test system onto the image and acquiring the results.
- Fictional method: a new method we invented whose process is similar to that from Hinrichs', but substituting the optimal points for other points which only move the test system on the horizontal axis, making it so we get the same number of evaluations but without being optimal points.
- CBC method: a method that applies the same process than that of Hinrichs' but substituting the optimal points for points from a lattice point set obtained using the CBC algorithm to create its generating vector.

These methods have also been tested for both the (x, y) process and the (x, ω) process that were used for Hinrichs (x, y) and Hinrichs (x, ω) respectively.

Notes:

- The results we show in this subsection have been obtained averaging the estimators computed for every possible rotation.
- The results obtained for the original method have been computed but were not included in the Figures so as to get a better visualization of the other results. For the record, the estimators obtained with the original method are similar to those obtained with the other methods, but the error variance is much bigger.

The photos that have been analyzed can be seen in Figure 1, and the images of the points from those photos that represent people and birds respectively can be seen in Figure 2.



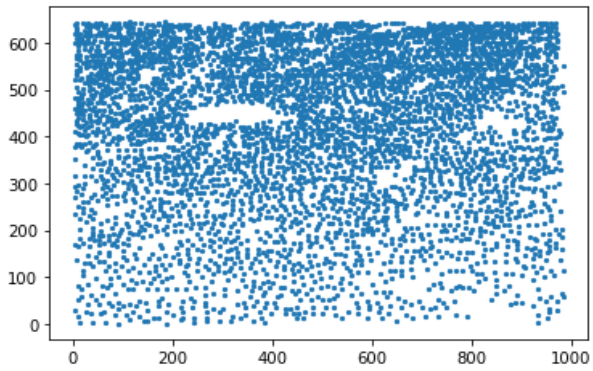
(a) Photo of a big group of people. The total number of people in the photo is $N = 4633$. (Data used by Idrees et al. [2013])

(b) Photo of a flock of birds. The total number of birds in the photo is $N = 13744$. (Photo obtained from: DRYAD open data repository)

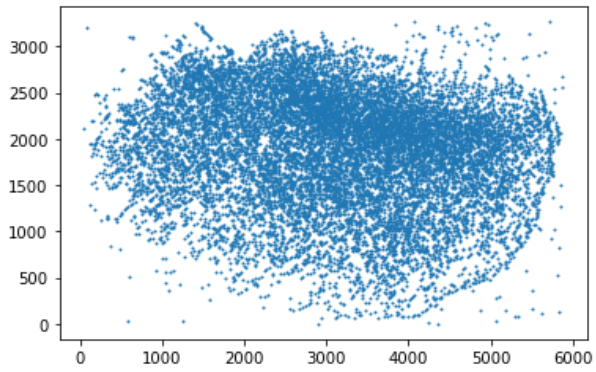
Figure 1: Photos from which the data that we analyze is obtained. The left photo is analyzed in the subsection 'Counting people' and the right photo is analyzed in the subsection 'Counting birds'.

3.1.1 Counting people

First, we will discuss the results obtained for the image in Figure 1a, which can be seen in Figure 5. Said image has a total amount of $N = 4633$ people, whereas the mean estimated number of people using a test system of quadrats with $t = 40$ and $T = 120$ was $\widehat{N} \sim 4400$ considering every method



(a) Points representing the group of people from Figure 1a. (Data used by Idrees et al. [2013])



(b) Points representing the flock of birds from Figure 1b. (Data obtained from: DRYAD open data repository)

Figure 2: Data points obtained from Figures 1a and 1b. The left image is analyzed in the subsection 'Counting people' and the right image is analyzed in the subsection 'Counting birds'.

and every set of points used.

If we now take a look at the different coefficient errors obtained, which can be seen in Figure 6, they are all quite small with around $\sim 4\%$ coefficient of error.

Taking a better look at the histograms from Figure 6, it looks like the CBC and Hinrichs methods have lower coefficient of error values when used with the (x, y) process (with a few exceptions). Also, these values don't seem to vary that much when changing the number of optimal points considered. The method with the lowest coefficient of error value in this case is CBC (x, y), however the Hinrichs (x, y) method gets a similar value with less points considered (more efficiency).

3.1.2 Counting birds

Now we will discuss the results obtained for the image in Figure 1b, which can be seen in Figure 7. Said image has a total amount of $N = 13744$ birds, whereas the mean estimated number of birds using a test system of quadrats with $t = 44$ and $T = 440$ was $\widehat{N} \sim 12700$ considering every method and every set of points used.

If we now take a look at the different coefficient errors shown in Figure 8, they are all somewhat small with around $\sim 7\%$ coefficient of error. However, González-Villa y Cruz [2020] obtained results for the same image, with the same parameters (t, T) but only for $\omega = 60^{\circ}$, giving a mean estimator $\widehat{N} = 13700$ and a 0.3% mean coefficient error.

Taking a better look at the histograms from Figure 8, it looks like this time only the CBC method has lower coefficient of error values when used with the (x, y) process for the majority of sets of points. Now the process (x, ω) has better results (lower coefficient of error values) for more sets of points when used with the Hinrichs method. Once again these values don't seem to vary that much when changing the number of optimal points considered. In regards to the best method, CBC has reached with two sets of points the lowest coefficient of error value with both processes ((x, y) and (x, ω)).

3.2 Buffon and Buffon-Steinhaus test systems

In this subsection we will discuss the results obtained for the estimated curve length $\widehat{B}$ of different curves. Since the Buffon-Steinhaus test system is the union of two perpendicular Buffon test systems, we will explain the procedure used to count the intersections needed for equation (6) by means of the Buffon test system. The same procedure applies for the Buffon-Steinhaus test system separately for each set of parallel lines that create the square grid.

The most important part of the procedure is the creation of the Buffon test system, namely the set of parallel lines that is "superimposed onto the curve". Now, in order to count the intersections we don't need to *create* the test system, but rather *check* if the points that conform the curve go past certain limits (meaning that we don't really superimpose the test system, but superimpose an imaginary test system instead). That being said, lets imagine that we want to create a test system of vertical lines separated a distance T apart. So as to change the position and orientation of the whole test system, we attach an AV $(\widehat{x}, \omega)$ to an AP $\widehat{x} = (x_1, x_2) \in \mathbb{R}^2$ of the test system. That way, the whole system moves depending on $\widehat{x}$ and rotates an angle $\omega \in [0, 2\pi)$ with respect to the horizontal axis.

In order to obtain the number of intersections with the test system, the first step was to rotate the curve (which is equivalent to rotating the test system) an angle $\omega \in [0, \pi)$. We don't need to consider the full interval $[0, 2\pi)$ since the same results are obtained in both cases due to the symmetry of the test system. The second step is to identify the AP of the test system, which for simplicity reasons was chosen to be the point $\widehat{x} = (m_x, m_y)$ with m_x the lowest value of the first coordinate from all points conforming the curve and m_y the lowest value of the second coordinate from all points conforming the curve. That way, for the purpose of maintaining randomness, a random value $r = (r_1, 0) \pmod{\mathrm{T}} \in \mathbb{R}^2$ was added to $\widehat{x}$ (there is no need to consider a random value for the second coordinate since the vertical lines are infinite). Finally, to count the number of intersections between the curve and the test system, we checked whether two consecutive points conforming the curve happen to have the first coordinate lower and greater (or vice versa) than that of a specific line from the test system, respectively. If that is the case for some line, then an intersection is counted, otherwise no intersection is counted. Furthermore, this process was vectorized so that several values of r and ω can be introduced respectively as *arrays*, making the code more efficient.

Lastly, we will explain how we implemented the optimal point sets for QMC integration in the code for this subsection. Similarly to last subsection, each test system created for the code has an AP $\widehat{x}$ with AV $(\widehat{x}, \omega) \equiv (x_1, x_2, \omega) \equiv (x, y, \omega)$. Also, considering equation (9), we have

$$\int_{\mathbb{R}^2} I(Y \cap \Lambda_{\mathbf{x},\omega})\, \mathrm{d}\mathbf{x} \approx \frac{1}{N} \sum_{n=0}^{N-1} I(Y \cap \Lambda_{\mathbf{x}_n,\omega}),$$

where $\Lambda_{\mathbf{x},\omega}$ is the Buffon test system and $\mathbf{x}_n$, $n = 0..., N-1$ are the optimal points for QMC integration. As a result, what we will do is calculate the number of intersections between the curve and the test system when the AP $\widehat{x} = (x, y)$ is located in each of the *Hinrichs' optimal points* for bivariate periodic functions (namely the optimal points for QMC integration obtained in Hinrichs y Oettershagen [2016], see Figure 21 in the appendix). After that, those numbers calculated for each test system will be added and divided by the total number of Hinrichs' optimal points used, N. In consequence, this should give us the value $\mathbb{E}[I(Y \cap \Lambda_{\mathbf{x},\omega})]$ with minimum worst case error. From now on, we will refer to this method as *Hinrichs* (x, y).

Again, we have also tried a different approach to this method. Apart from using the Hinrichs' optimal points in the AP $\widehat{x} = (x, y)$, we have also computed results where the first coordinate x and the angle ω are modified by the optimal points for each system (see explanation in the latter subsection). From now on, we will refer to this method as *Hinrichs* (x, ω).

Notes:

- In this subsection, we have also escalated the Hinrichs' optimal points depending on the method used.
- For the Buffon-Steinhaus test system we only have to use two perpendicular Buffon test systems and carry on the same procedure for each one of them.

We will now resume the computed results acquired for the estimated curve length $\widehat{B}$ of different curves and show how these methods' variances behave (via the squared coefficient of error (see equation (14))) compared to the other methods that have been presented in the last subsection. Again, these methods have also been tested for both the (x, y) process and the (x, ω) process.

Notes:

- The results we show in this subsection have been obtained averaging the estimators computed for every possible translation and rotation.
- The results obtained for the original method have been computed but were not included in the Figures so as to get a better visualization of the other results. For the record, the estimators obtained with the original method are similar to those obtained with the other methods, but the error variance is much bigger.

The curves whose length we tried to estimate can be seen in Figures 3 and 4.

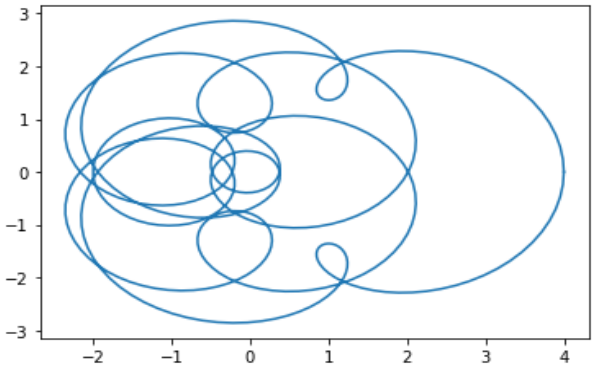
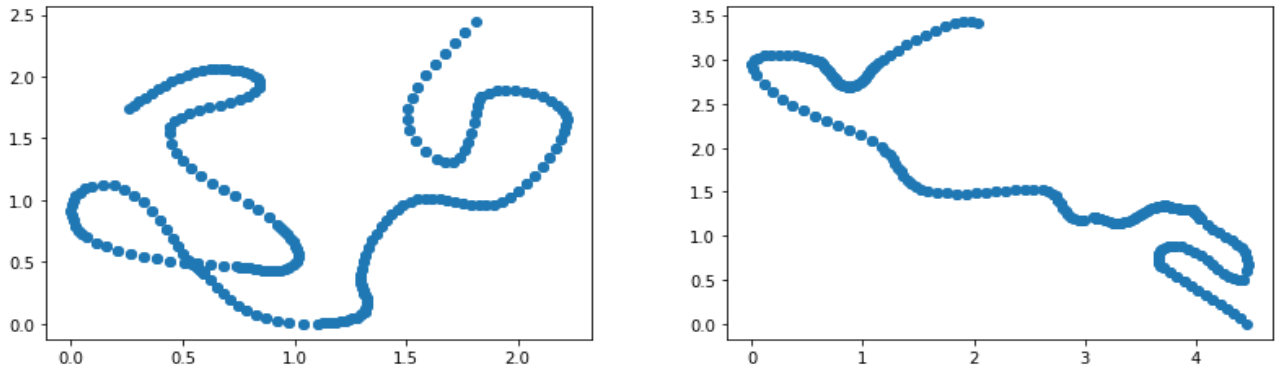


Figure 3: Image of a special symmetric curve obtained using the equation $\gamma_{\mathbf{a}}(t) = \sum_{j=0}^{m} e^{2\pi i \mathbf{a} t}$ from Pausinger y Vartziotis [2021] with $\mathbf{a} = (1, 3, 4, 11)$ $(m = 4)$. This curve is analyzed in the subsection 'Estimating length of Pausinger's curves'.

3.3 Estimating length of Pausinger's curves

First, we will discuss the results obtained for the image in Figure 3, which can be seen in Figures 9 and 10. The curve from that image has a total length in $\mathbb{R}^2$ $B = 72.86$, accordingly, the mean estimated curve length obtained using the Buffon and Buffon-Steinhaus test systems with $T = 1$ is $\widehat{B} \sim 72.87$ considering every method and every set of points used.

Looking at the different coefficient errors shown in Figures 11 and 12 they are all very small, with the biggest one being around $\sim 0.06\%$ and the lowest one being around $\sim 0.001\%$. The estimator used is known to work well and these results seem really good with low error.



(a) DNA molecule referred to as *curve 5*.

(b) DNA molecule referred to as *curve 6*.

Figure 4: DNA molecules from Gomez et al. [2016] whose curve length is estimated. The left molecule is analyzed in the subsection 'Curve 5' and the right molecule is analyzed in the subsection 'Curve 6'.

Taking a better look at the histograms from Figures 11 and 12, it looks like the (x, y) process (which in Buffon's test system it's called (x) because changing the component y doesn't modify the test system) is better (meaning that if gives a lower coefficient of error value) for all methods with the Buffon test system for all sets of points, however, with the Buffon-Steinhaus test system the process (x, y) is only better for all point sets with the Fictional and Hinrichs methods. The CBC method with a Buffon-Steinhaus test system is mostly better when used with the process (x, y), however, there are some exceptions. About the values that the coefficient of error takes, there seems to be a relation between these values and the number of optimal points considered for all methods with the (x, y) process except the CBC method with a Buffon-Steinhaus test system, which could mean that more points equals more precision. It is worth mentioning that the fictional (x, y), fictional (x, ω) and Hinrichs (x, y) methods give the same values when used with the Buffon and Buffon-Steinhaus test systems for any set of points considered. This is because the symmetry of those test systems makes it so fictional (x, y) and fictional (x, ω) are equal to one another and the process considered to obtain the AP of the test systems gives equivalent sets of points for the fictional and Hinrichs methods (modulo the test system's symmetry). All the methods have reached the lowest coefficient of error value with the (x, y) process using both the Buffon and the Buffon-Steinhaus test systems.

3.4 Estimating length of DNA molecules

Now we will discuss the results obtained for the images in Figure 4.

3.4.1 Curve 5

We will start talking about the DNA molecule in Figure 4a, which we will refer to as *Curve 5* because of how it's numerated by Gomez et al. [2016]. The results for curve 5 can be seen in Figures 13 and 14. Curve 5 has a total length in $\mathbb{R}^2$ $B = 12.78$, accordingly, the mean estimated curve length obtained using the Buffon and Buffon-Steinhaus test systems with $T = 0.7$ are $\widehat{B} \sim 12.76$ and $\widehat{B} \sim 12.80$ respectively considering every method and every set of points used.

Taking a look at the different coefficient errors shown in Figures 15 and 16 they are all very small, with the biggest one being around $\sim 2\%$ and the lowest one being around $\sim 0.02\%$, making the results seem really good with low error. Gomez et al. [2016] analyzed the same curve and obtained a curve length $B = 393$ nm and estimated curve length $\widehat{B} = 333$ nm with a coefficient error 11.6% using a Buffon-Steinhaus test system (square grid) with $T = 25$ nm. On their work they take $T = 25$ nm so

as to get an intersection expectation $\mathbb{E}[I] \sim 20$ intersections. In our case $T = 0.7$ was chosen, leading to $\mathbb{E}[I] \sim 10$ intersections (per optimal point) for the Buffon test system and $\mathbb{E}[I] \sim 20$ intersections (per optimal point) with the Buffon-Steinhaus test system. It seems like we are getting better results in this case for curve length estimation.

Contrary to what we saw in the counting results, taking a better look at the histograms from Figures 15 and 16 we see that the (x) process for the Buffon test system is worse in this case for all methods but the fictional (for any set of points considered). However, just like last time, coefficient of error values tend to lower when more optimal points are considered using the CBC and Hinrichs methods, but this time with the (x, ω) process. About the histograms for the Buffon-Steinhaus test system, the coefficient of error values seem random (no relation between number of points and coefficient of error value) and there doesn't seem to be a better process between the (x, y) and the (x, ω) in general. For the majority of sets of points the (x, y) process is better with the Hinrichs method (with exceptions), but with the CBC method the (x, ω) process looks better for the vast majority of sets of points (with few exceptions). Also, we can see again that the fictional (x, y), fictional (x, ω) and Hinrichs (x, y) methods give the same values when used with the Buffon and Buffon-Steinhaus test systems for any set of points considered. The method with the lowest coefficient of error value in this case is CBC with the (x, ω) process for both the Buffon and the Buffon-Steinhaus test systems.

3.4.2 Curve 6

Lastly, we will discuss the results obtained for the DNA molecule in Figure 4b, which we will refer to as *Curve 6* because of how it's numerated by Gomez et al. [2016]. The results for curve 6 can be seen in Figures 17 and 18. Curve 6 has a total length in $\mathbb{R}^2$ $B = 15.16$, accordingly, the mean estimated curve length obtained using the Buffon and Buffon-Steinhaus test systems with $T = 0.7$ are $\widehat{B} \sim 15.24$ and $\widehat{B} \sim 15.19$ respectively considering every method and every set of points used.

Taking a look at the different coefficient errors shown in Figures 19 and 20 they are all quite small, with the biggest one being around $\sim 7\%$ and the lowest one being around $\sim 0.009\%$, making the results seem really good with low error. Gomez et al. [2016] analyzed the same curve and obtained a curve length $B = 392$ nm and estimated curve length $\widehat{B} = 351$ nm with a coefficient error 11.2% using a Buffon-Steinhaus test system (square grid) with $T = 25$ nm. This time $T = 0.7$ led to $\mathbb{E}[I] \sim 14$ intersections (per optimal point) for the Buffon test system and $\mathbb{E}[I] \sim 28$ intersections (per optimal point) with the Buffon-Steinhaus test system. Again, it seems like we get better results for curve length estimation.

Once again, contrary to what we saw in the first length estimation results, taking a better look at the histograms from Figures 15 and 16 we see that the (x) process for the Buffon test system is worse for all methods but the fictional (for any set of points considered). Also, similarly to the results for curve 5, coefficient of error values tend to slightly lower when more optimal points are considered using the CBC and Hinrichs methods with the (x, ω) process. However, about the histograms for the Buffon-Steinhaus test system, even though the coefficient of error values still seem random (no relation between number of points and coefficient of error value) it's clear that the (x, ω) process is better for all methods but the fictional. Also, we can see again that the fictional (x, y), fictional (x, ω) and Hinrichs (x, y) methods give the same values when used with the Buffon and Buffon-Steinhaus test systems for any set of points considered. Once more, the method with the lowest coefficient of error value is CBC with the (x, ω) process for both the Buffon and the Buffon-Steinhaus test systems.

Finally we will mention a few other results that were obtained during the development of the code for the Buffon-Steinhaus test system. We checked that the Hinrichs (x, y) method provides an estimator with the same variance as the original method when considering the same number of evaluations.

This means that, for example, if one chooses a Buffon-Steinhaus test system with $T = 0.1$ and uses the original method then the estimator's variance will be similar to the one obtained for a Buffon-Steinhaus test system with $T = 1$ used with the Hinrichs (x, y) method considering $N = 10$ optimal points. On the other hand, the Hinrichs (x, ω) and CBC methods give worse results in this case. We also checked that the estimator's variance decreases when the value of the parameter T decreases, this makes sense because doing so we are lowering the chances for test system to intersect the curve in places where the number of intersections changes significantly. Moreover, we can say that, from all the experiments (for all test systems used in this work) that have been carried out, the computing time didn't seem to exceed the five minute mark. Considering that González-Villa y Cruz [2020] and Gomez et al. [2016] had computing times of around five minutes or more with less function evaluations, this indicates that our code is very efficient.

In conclusion, at least one of the CBC and Hinrichs methods seems to give the lowest coefficient of error values when used with the (x, ω) and (x, y) processes respectively. However, it is possible that one method is better than the other depending on the isotropy, anisotropy or symmetry that the analyzed object has (in this case the curve or the group of points from the photo). It is also possible that changing the value of the parameter γ mentioned in Hinrichs y Oettershagen [2016] provides better results for the Hinrichs method in every experiment. Nonetheless, the Hinrichs method seems to give low coefficient of error values at least for one of the processes (x, y) or (x, ω), which makes it a and easy-to-implement estimation method for stereology.

4 Conclusions and open problems

We have tried to improve stereology estimators. In order to do so, we tried to approximate expectations using QMC integration with N optimal points that minimize the worst case error in the reproducing kernel Hilbert space H^1_{mix}. Since stereology estimators usually require to obtain expectations, this idea should provide estimators with low variances, improving the quality of the estimated quantity. So as to apply this idea, we selected an associated point (AP) for the test system used in the estimation and carried out the whole estimation process repeatedly, putting the AP of the test system in a different optimal point every time and averaging the results.

Knowing this, we selected three known test systems, namely the test system of quadrats, the Buffon test system and the Buffon-Steinhaus test system, and implemented the process above in python code for each one of them in order to test its performance and compare it to other known methods. What's more, we not only tested it for associated points $(x, y) \in \mathbb{R}^2$ with (x, y) being optimal points, but also with associated points $(x, y) \in \mathbb{R}^2$ with x being the first coordinate of the optimal point and a modified rotation angle $\omega + h \pmod{\pi} \in [0, \pi)$ with h being the second coordinate of the optimal point.

The computed results for every experiment carried out with all test systems led us to make two hypotheses:

- At least one of the CBC and Hinrichs methods gives the lowest coefficient of error values when used with the (x, ω) and (x, y) processes respectively. Also, one method might be better than the other depending on the isotropy, anisotropy or symmetry that the analyzed object has.
- Changing the value of the parameter γ mentioned in Hinrichs y Oettershagen [2016] may provide better results for the Hinrichs method in every experiment.

Nevertheless, the Hinrichs method seems to give low coefficient of error values at least for one of the processes (x, y) or (x, ω), which makes it a good and easy-to-implement estimation method for stereology.

Some open problems that arise due to the development of this work are:

- How does the Hinrichs method behave depending on the type of curve (isotropic, anisotropic, symmetry)?
- How much does changing γ influence on the results?
- Are there other ways to implement the idea behind this work in stereology?

Bibliography

L. M. Cruz-Orive. Stereology: a historical survey. *Image Analysis & Stereology*, 36(3):153–177, 2017.

L. M. Cruz-Orive. *Stereology Theory and Applications*. 2022.

A. I. Gomez, M. Cruz y L. M. Cruz-Orive. On the precision of curve length estimation in the plane. *Image Analysis & Stereology*, 35(1):1–14, 2016.

J. González-Villa y M. Cruz. The countem software: simple, efficient and unbiased population size estimation. *Ecography*, 43(2):251–255, 2020.

A. Hinrichs y J. Oettershagen. Optimal point sets for quasi-monte carlo integration of bivariate periodic functions with bounded mixed derivatives. In *Monte Carlo and Quasi-Monte Carlo Methods: MCQMC, Leuven, Belgium, April 2014*, pages 385–405. Springer, 2016.

H. Idrees, I. Saleemi, C. Seibert, y M. Shah. Multi-source multi-scale counting in extremely dense crowd images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2547–2554, 2013.

G. Leobacher y F. Pillichshammer. *Introduction to quasi-Monte Carlo integration and applications*. Springer, 2014.

F. Pausinger y D. Vartziotis. On the symmetry of finite sums of exponentials. *Elemente der Mathematik*, 76(2):62–73, 2021.

A Appendix

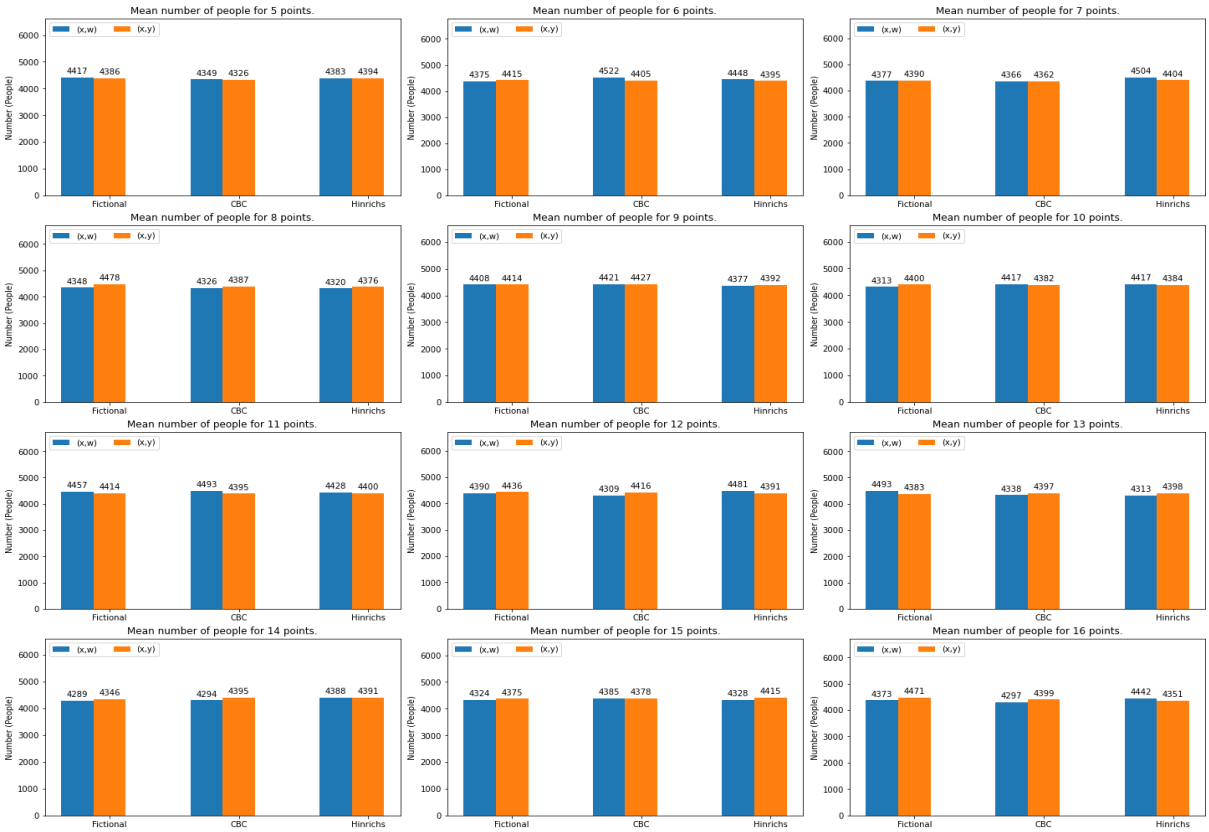


Figure 5: Histograms for sets of $N = 5, ..., 16$ points with the mean estimated number of people/points in the vertical axis and the method used in the horizontal axis when counting people/points from Figure 2a using a test system of quadrats with $t = 40$ and $T = 120$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

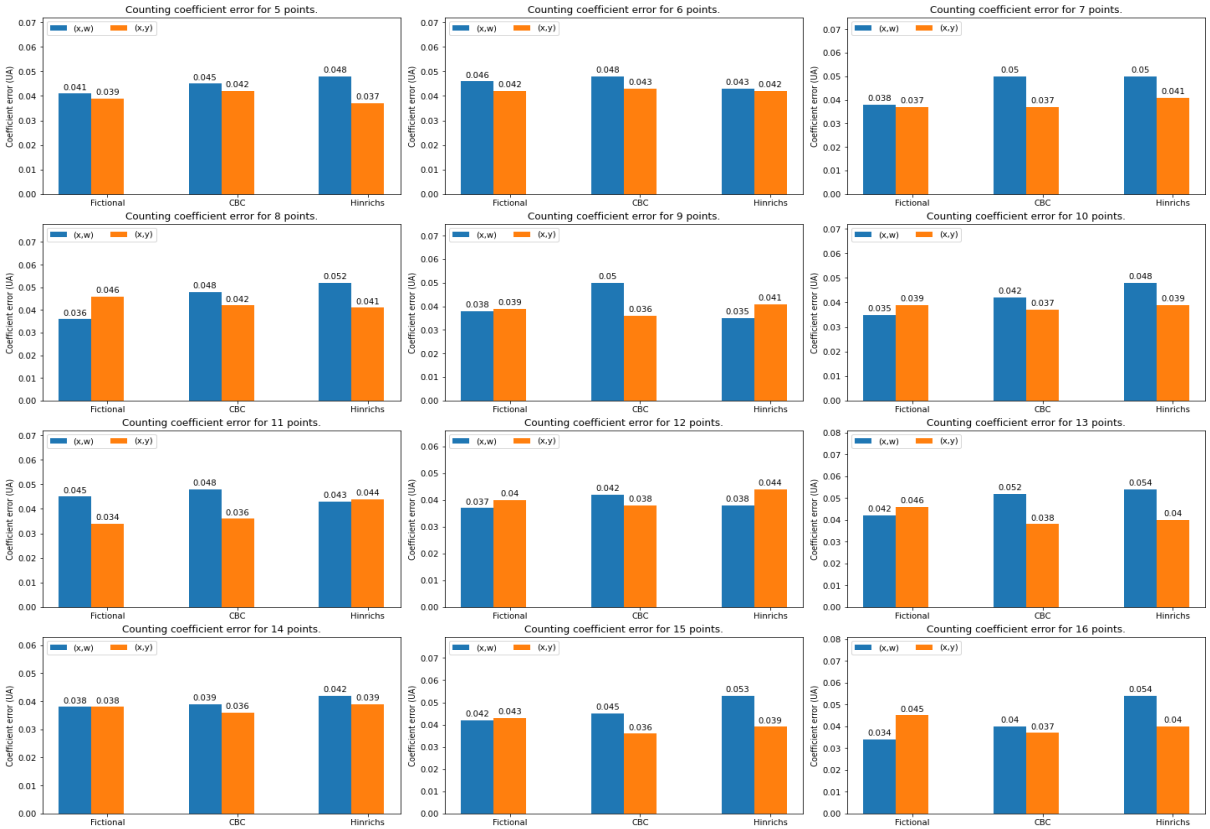


Figure 6: Histograms for sets of $N = 5, ..., 16$ points with the mean coefficient of error in the vertical axis and the method used in the horizontal axis when counting people/points from Figure 2a using a test system of quadrats with $t = 40$ and $T = 120$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

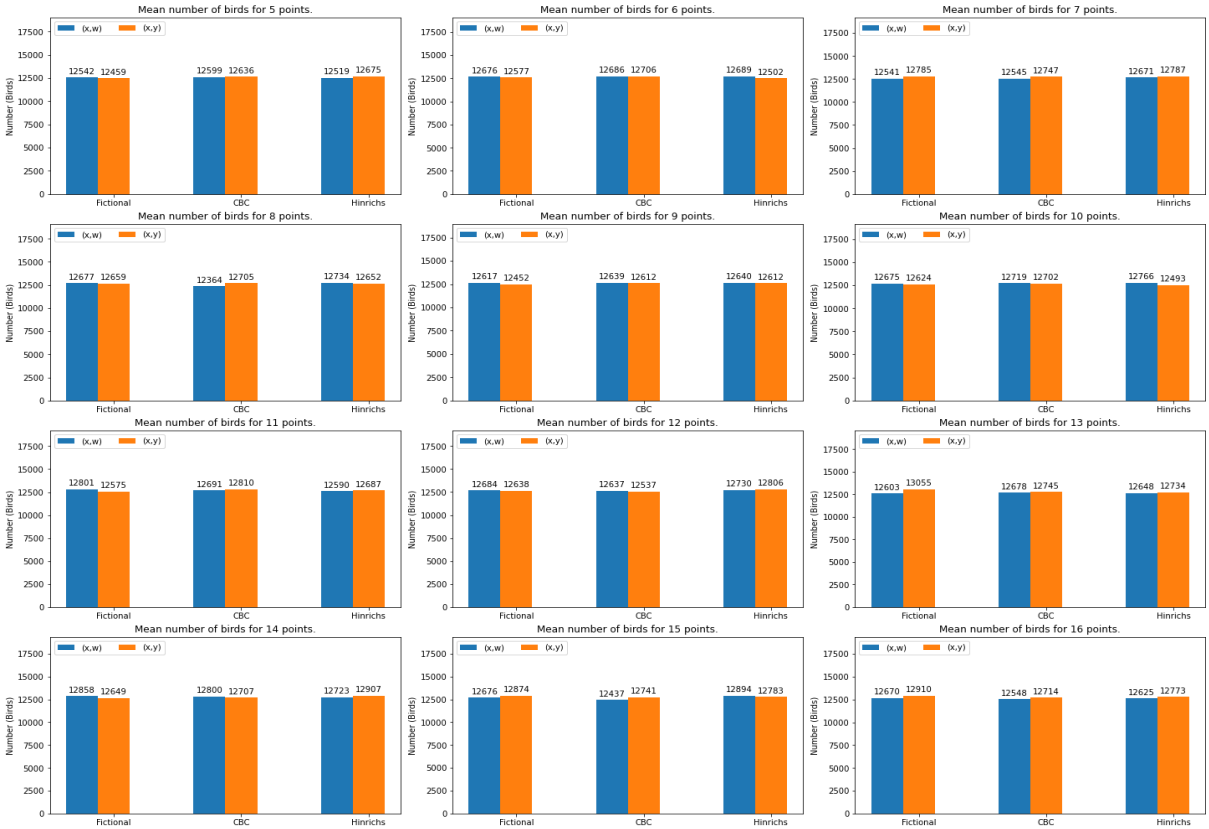


Figure 7: Histograms for sets of $N = 5, ..., 16$ points with the mean estimated number of birds/points in the vertical axis and the method used in the horizontal axis when counting birds/points from Figure 2b using a test system of quadrats with $t = 44$ and $T = 440$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

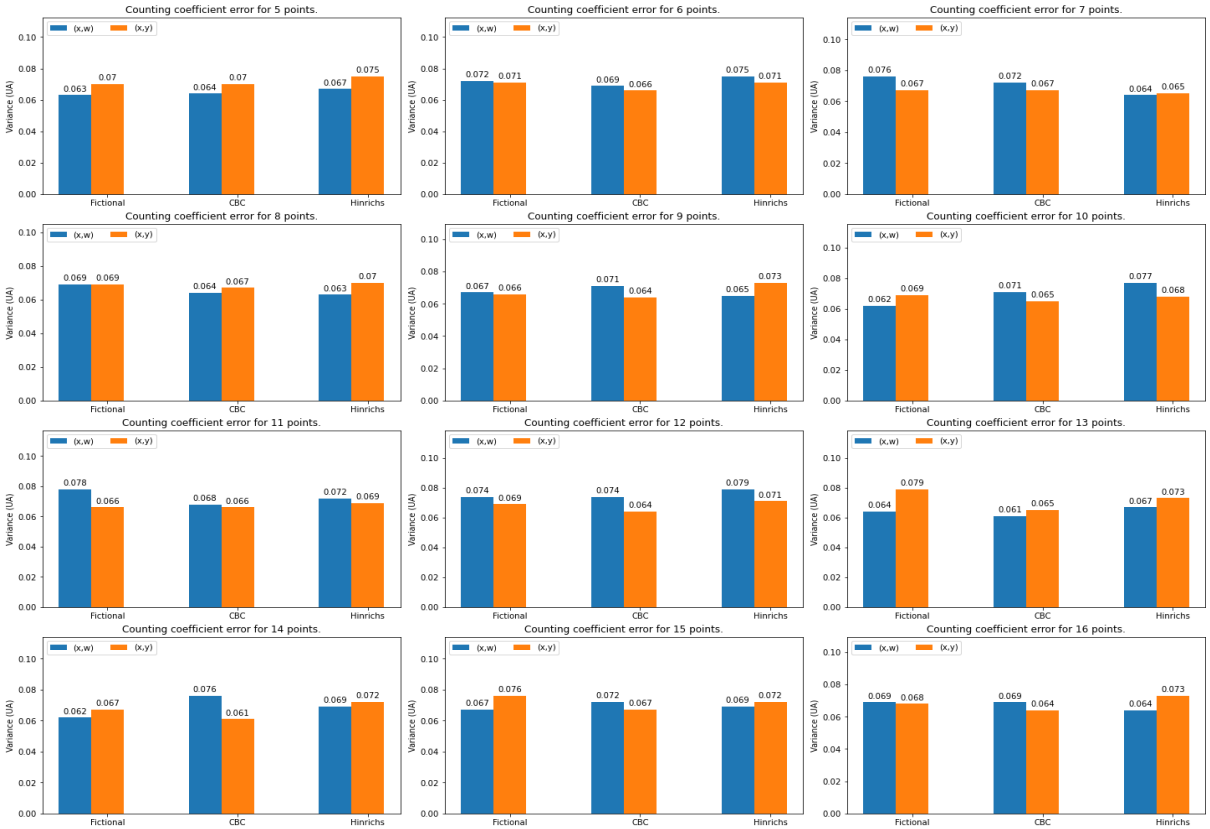


Figure 8: Histograms for sets of $N = 5, ..., 16$ points with the mean coefficient of error in the vertical axis and the method used in the horizontal axis when counting birds/points from Figure 2b using a test system of quadrats with $t = 44$ and $T = 440$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

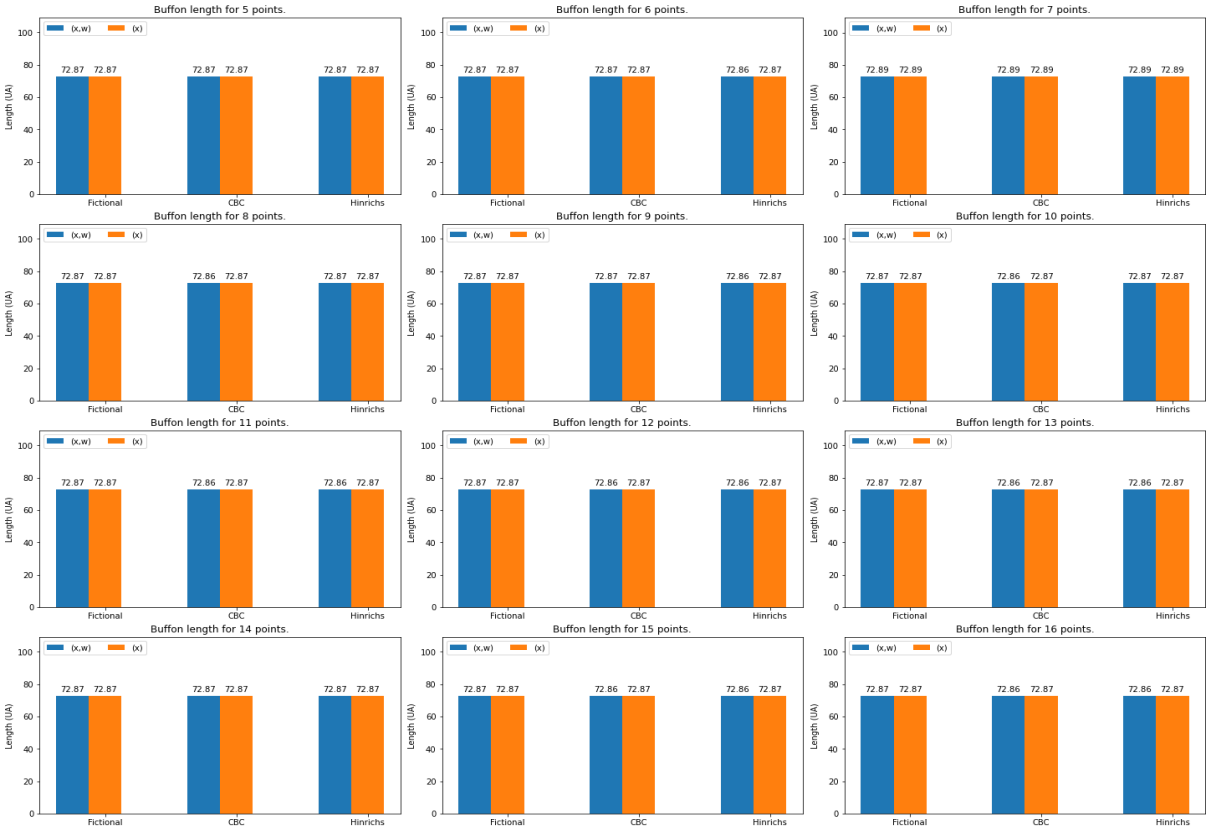


Figure 9: Histograms for sets of $N = 5, ..., 16$ points with the mean curve length in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 3 using a Buffon test system with $T = 1$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

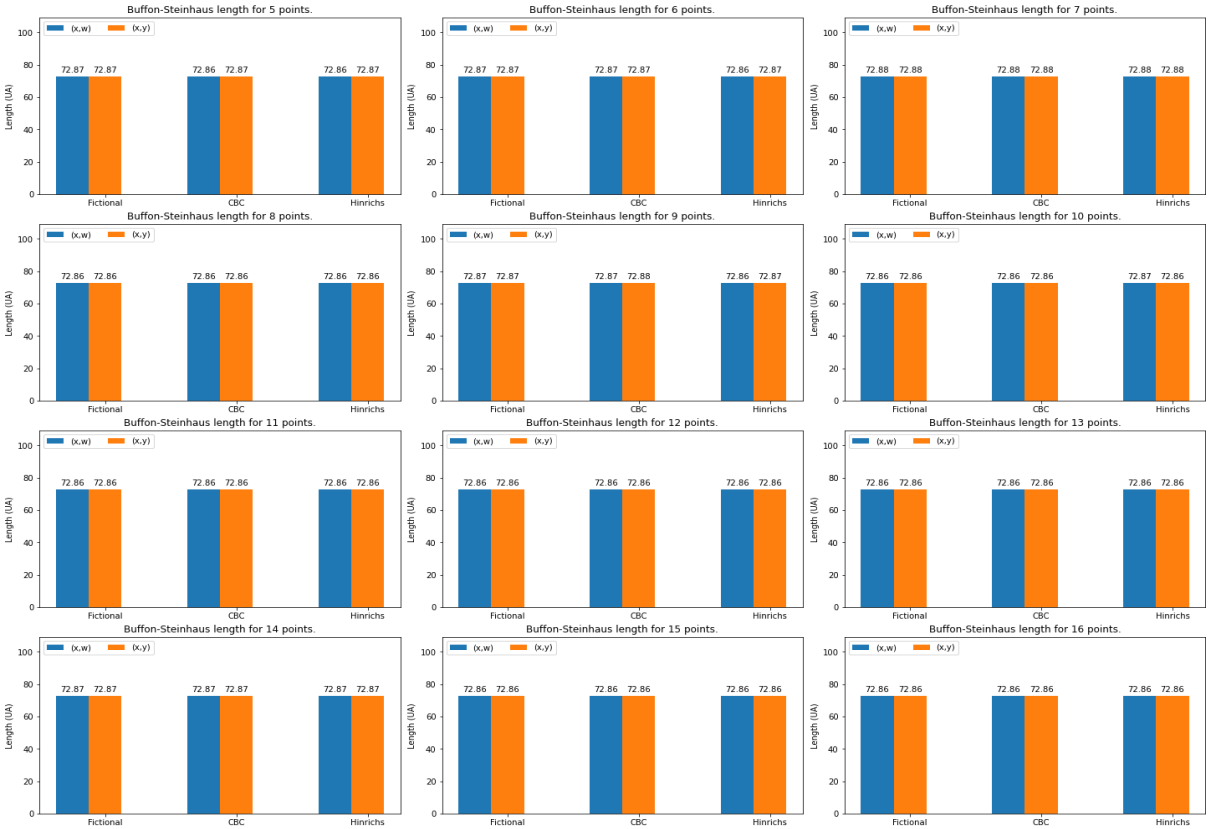


Figure 10: Histograms for sets of $N = 5, ..., 16$ points with the mean curve length in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 3 using a Buffon-Steinhaus test system with $T = 1$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

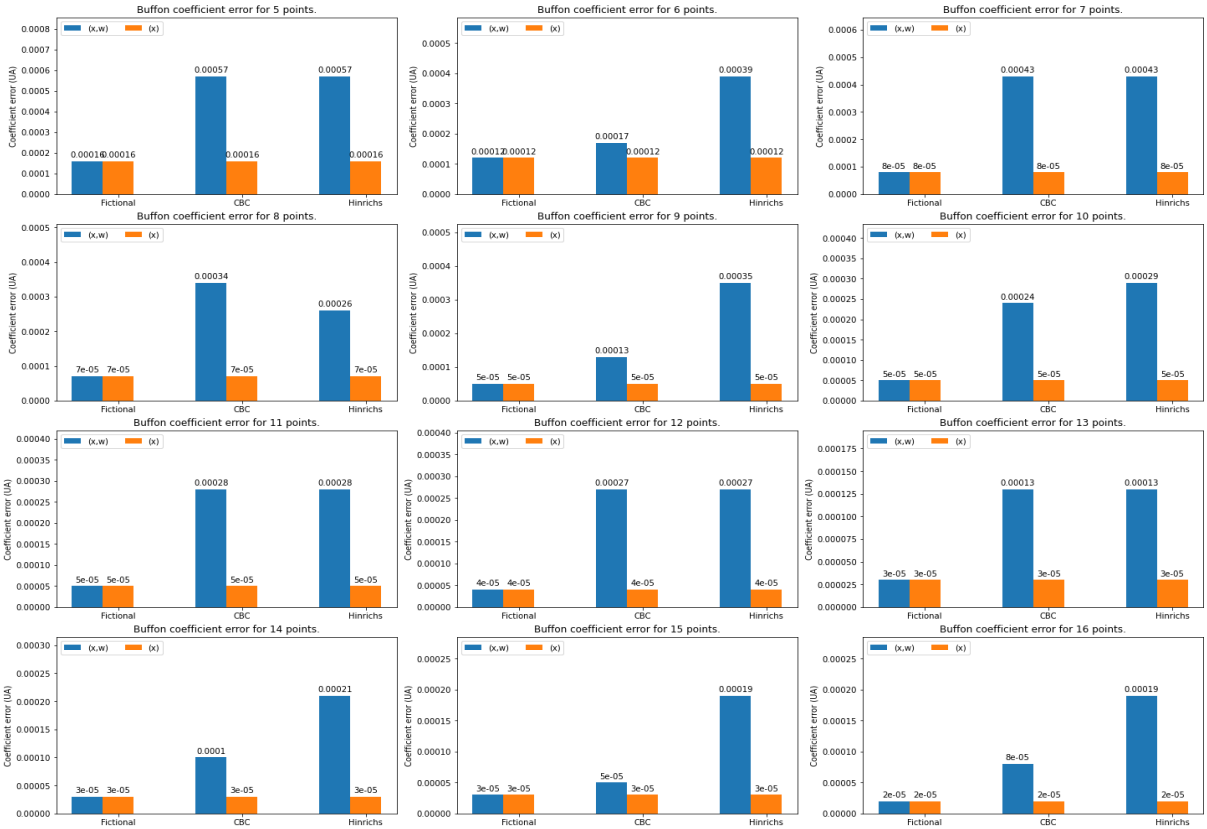


Figure 11: Histograms for sets of $N = 5, ..., 16$ points with the mean coefficient of error in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 3 using a Buffon test system with $T = 1$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

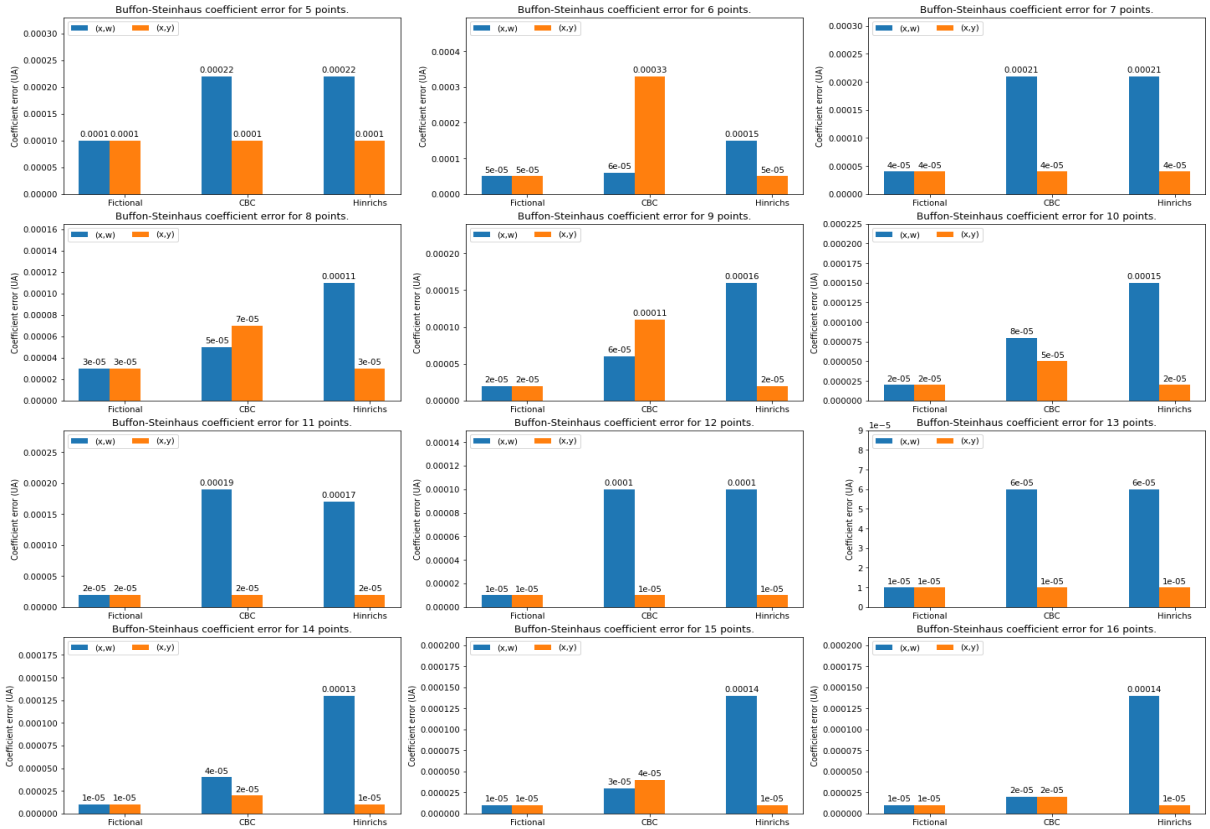


Figure 12: Histograms for sets of $N = 5, ..., 16$ points with the mean coefficient of error in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 3 using a Buffon-Steinhaus test system with $T = 1$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

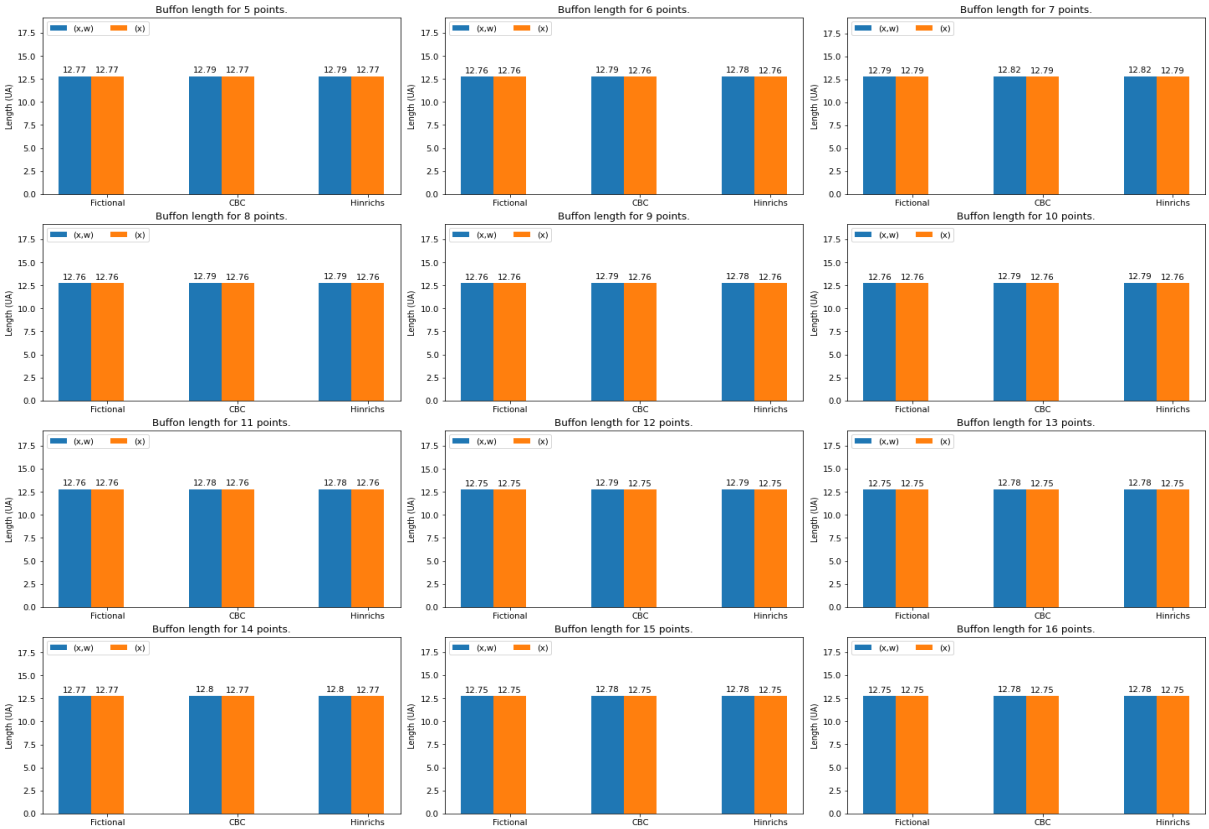


Figure 13: Histograms for sets of $N = 5, ..., 16$ points with the mean curve length in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 4a using a Buffon test system with $T = 0.7$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

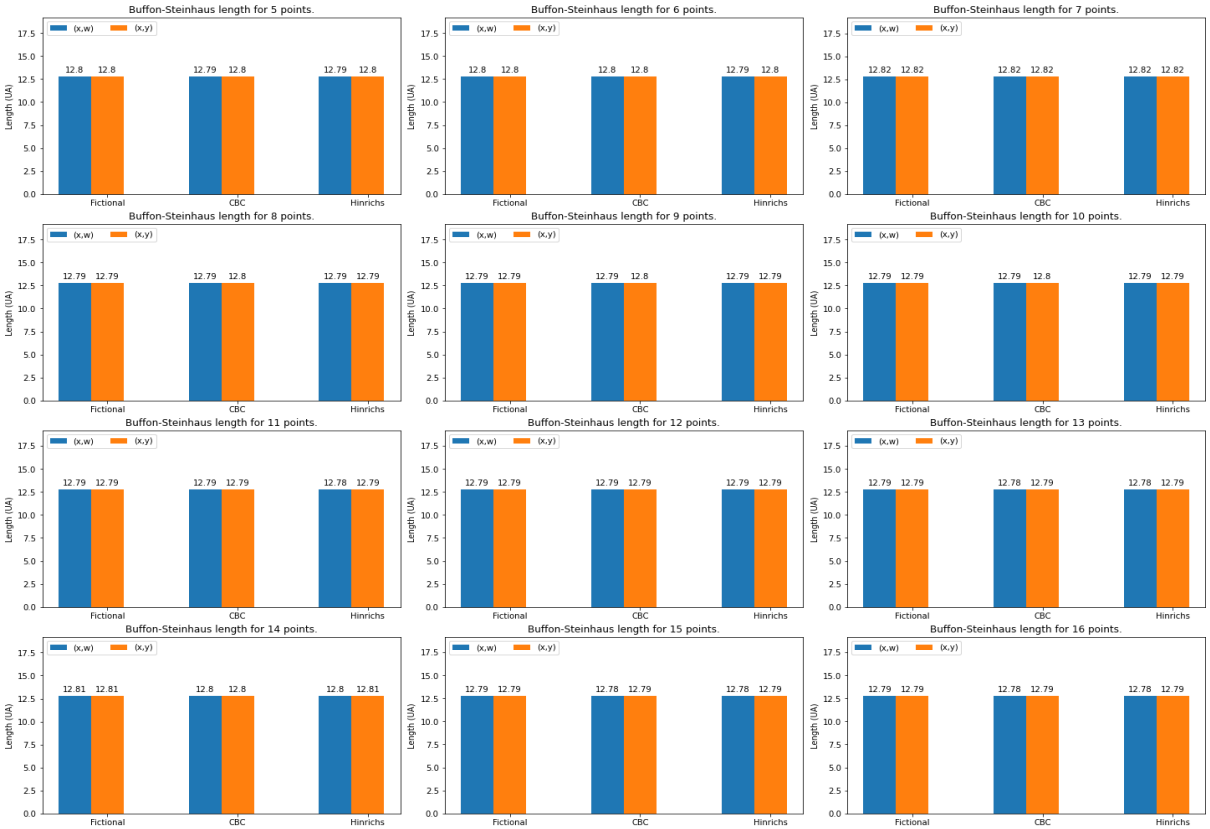


Figure 14: Histograms for sets of $N = 5, ..., 16$ points with the mean curve length in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 4a using a Buffon-Steinhaus test system with $T = 0.7$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

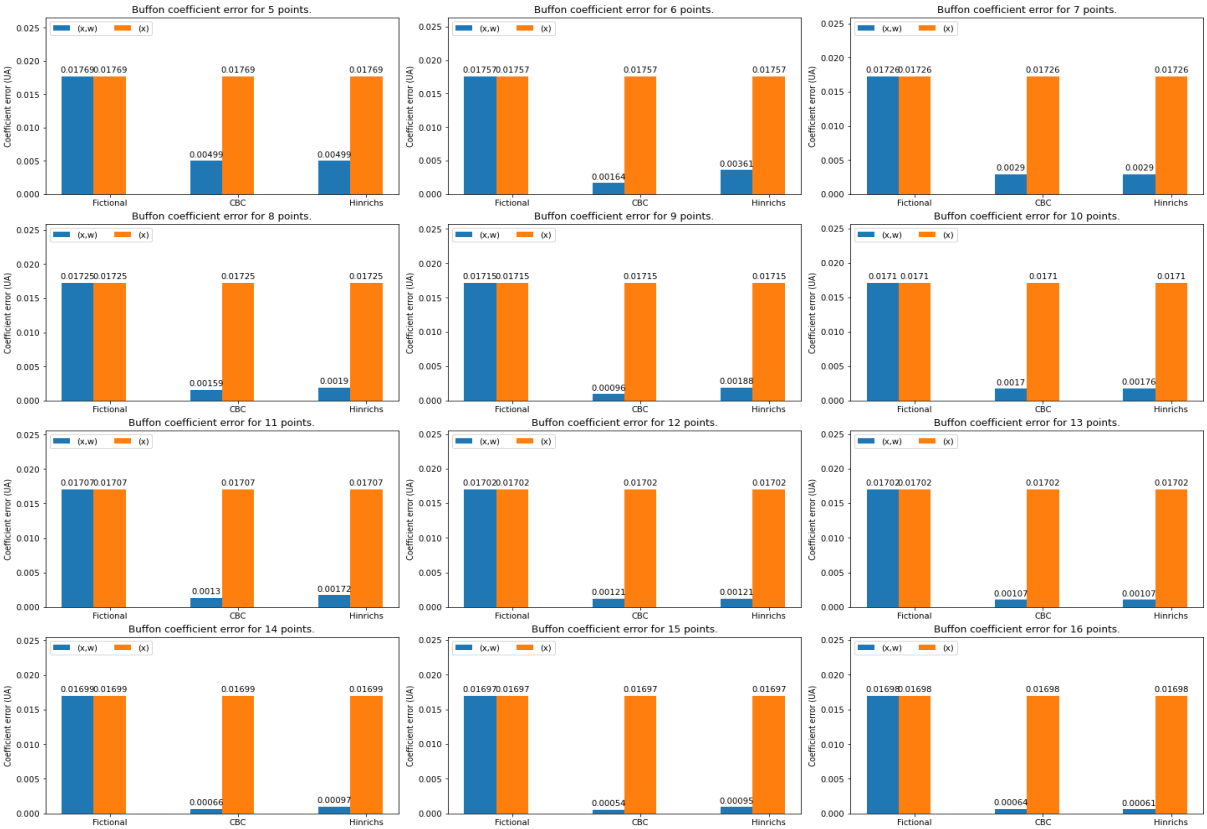


Figure 15: Histograms for sets of $N = 5, ..., 16$ points with the mean coefficient of error in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 4a using a Buffon test system with $T = 0.7$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

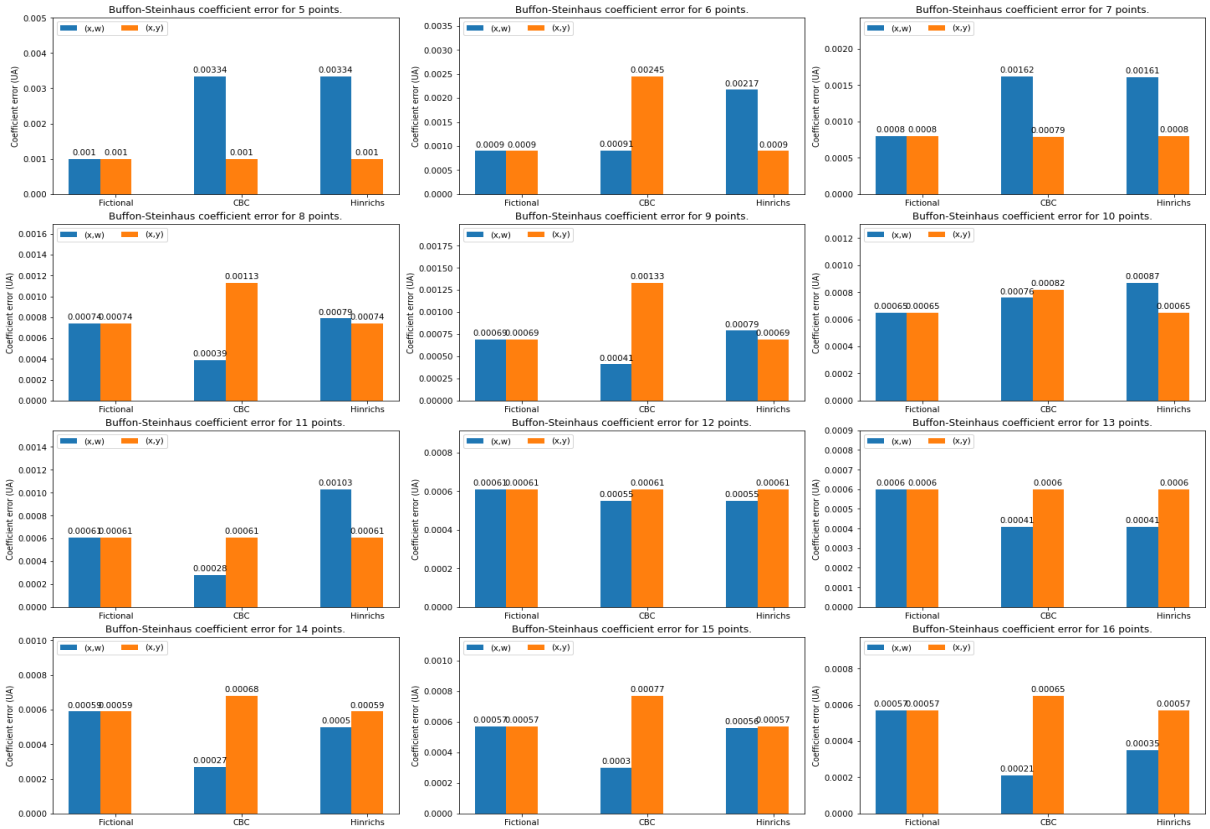


Figure 16: Histograms for sets of $N = 5, ..., 16$ points with the mean coefficient of error in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 4a using a Buffon-Steinhaus test system with $T = 0.7$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

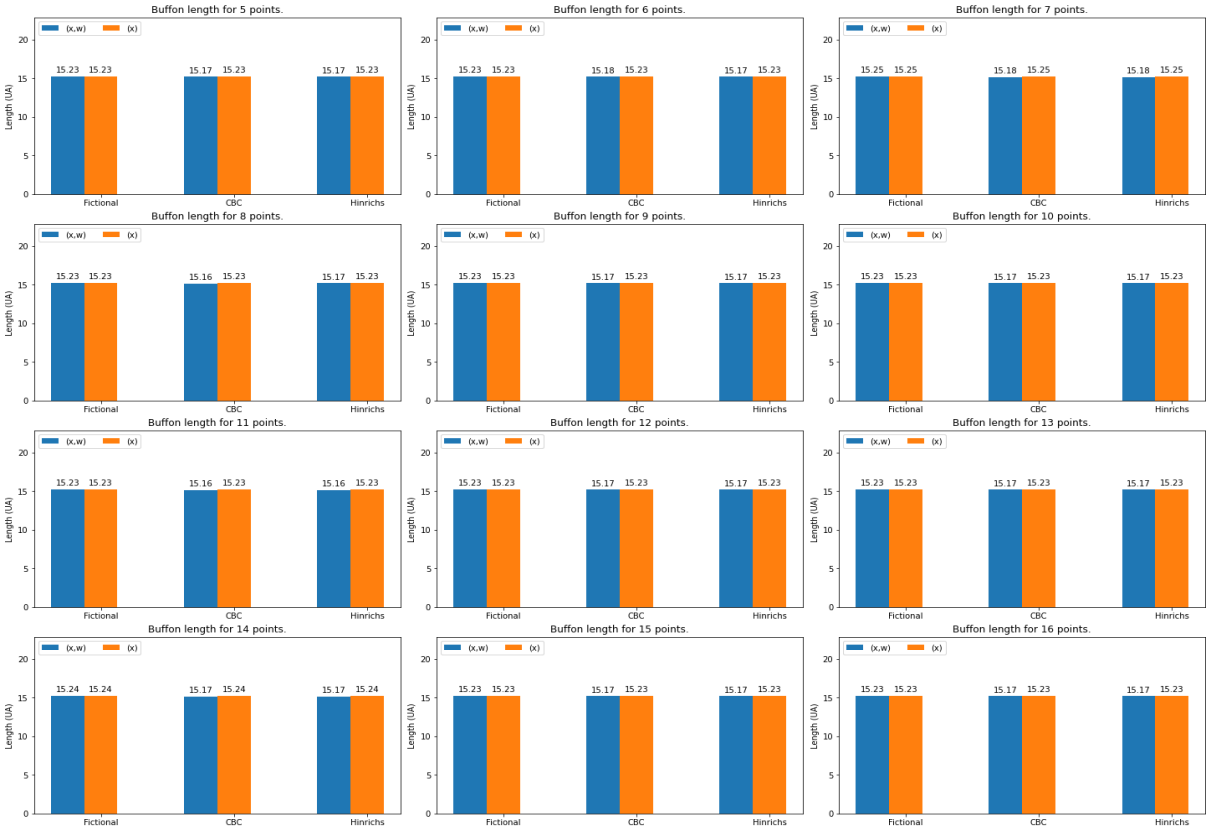


Figure 17: Histograms for sets of $N = 5, ..., 16$ points with the mean curve length in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 4b using a Buffon test system with $T = 0.7$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

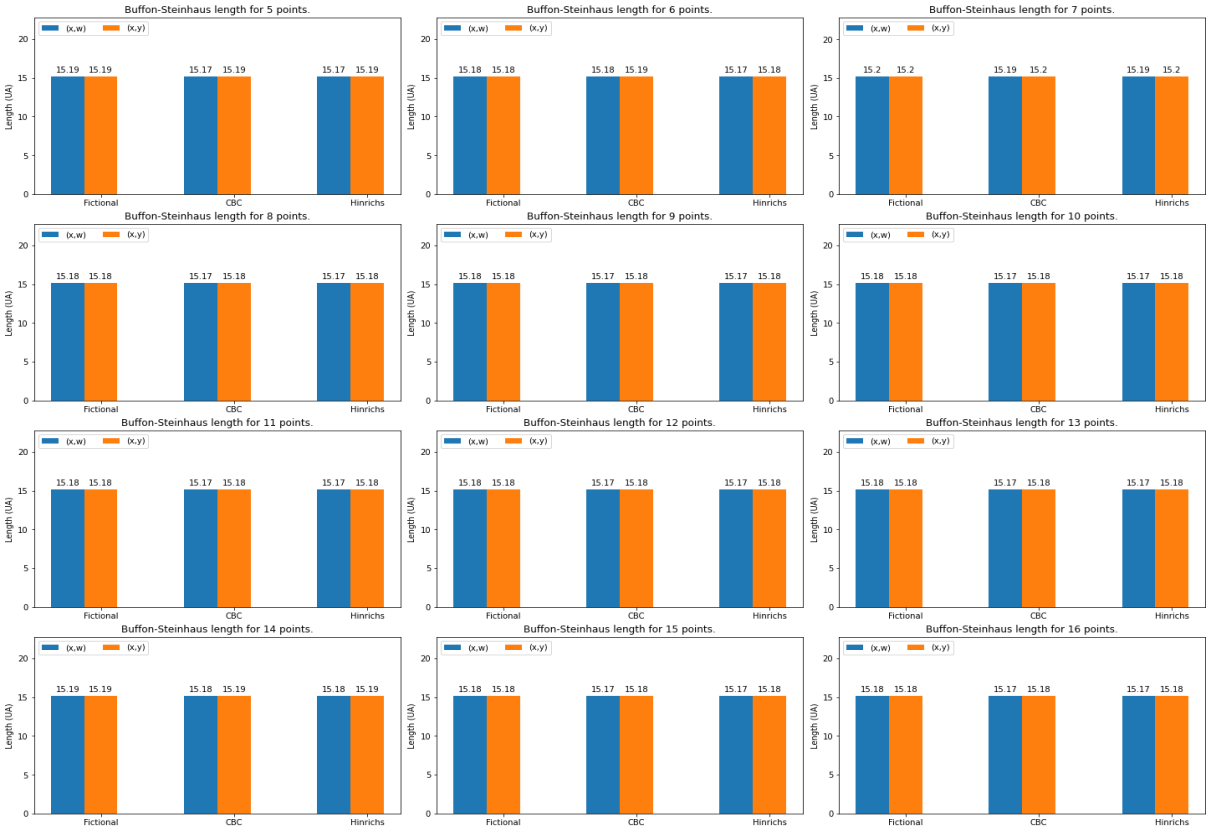


Figure 18: Histograms for sets of $N = 5, ..., 16$ points with the mean curve length in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 4b using a Buffon-Steinhaus test system with $T = 0.7$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

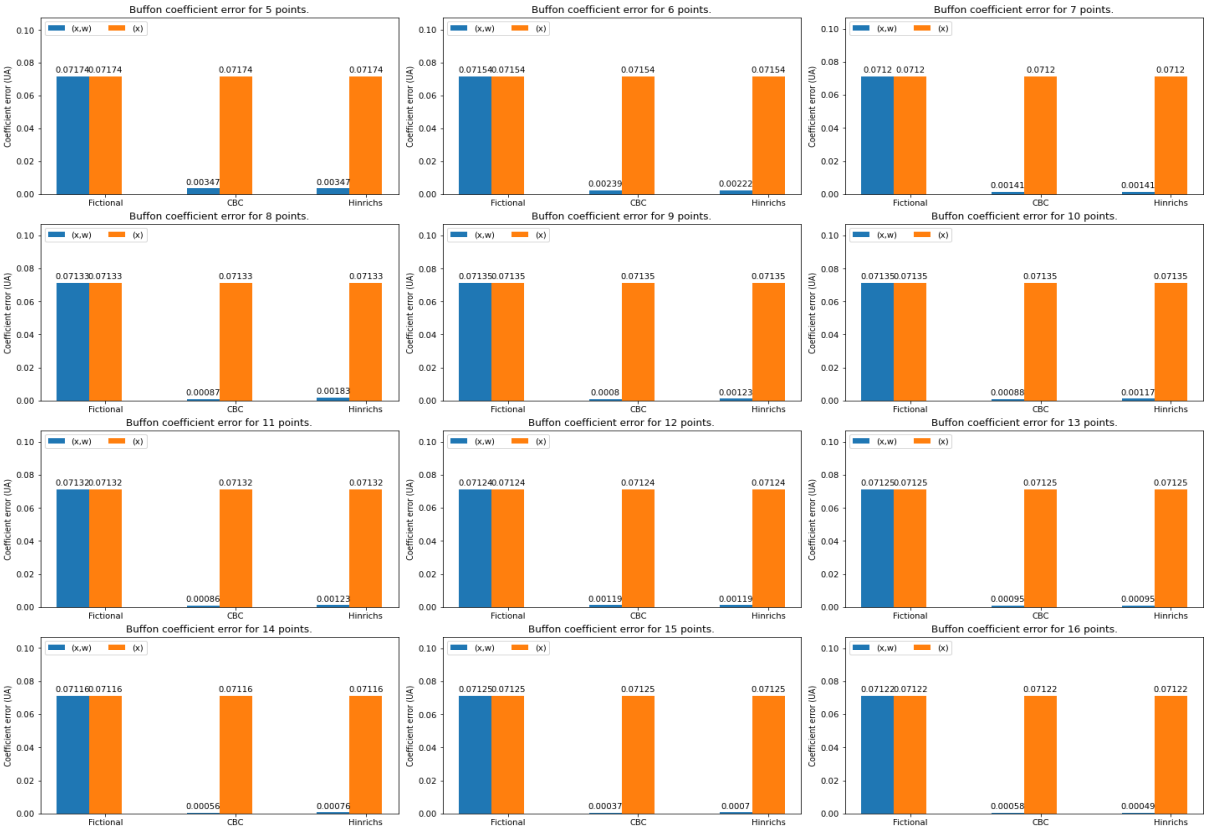


Figure 19: Histograms for sets of $N = 5, ..., 16$ points with the mean coefficient of error in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 4b using a Buffon test system with $T = 0.7$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

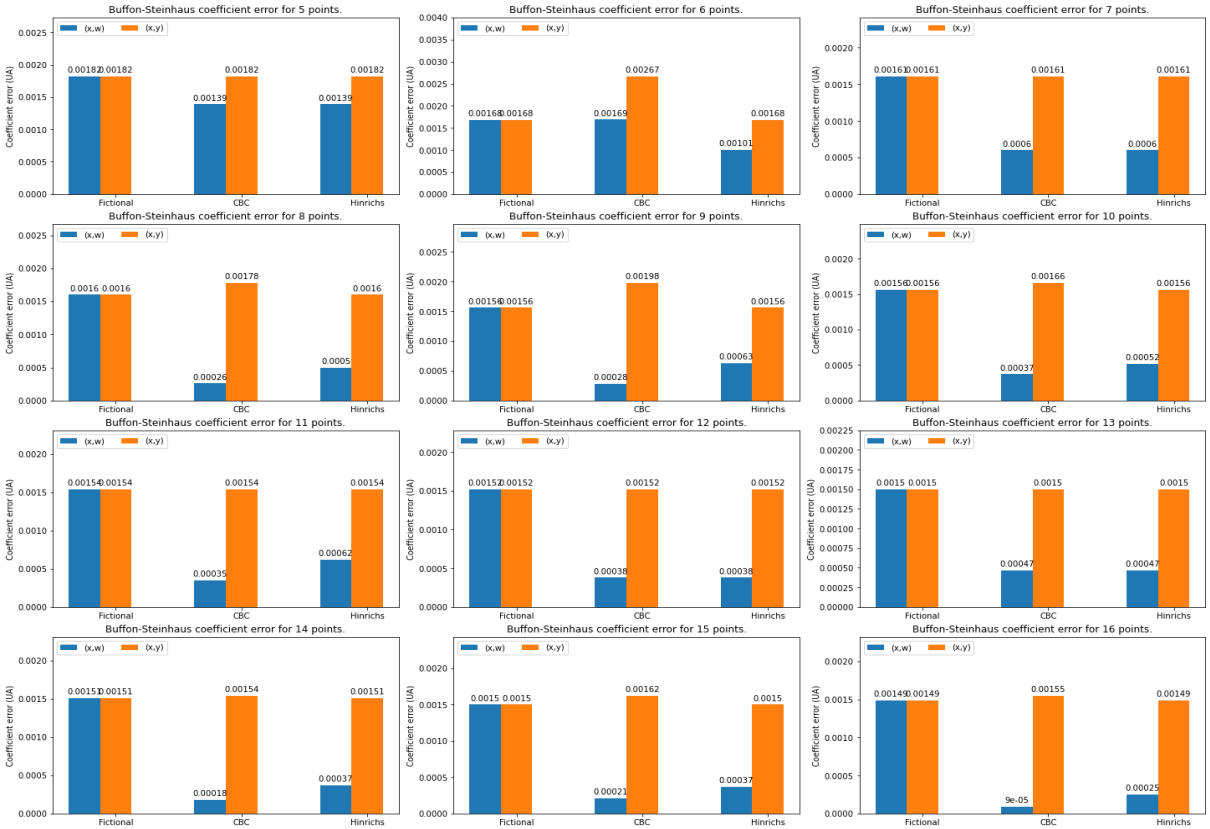


Figure 20: Histograms for sets of $N = 5, ..., 16$ points with the mean coefficient of error in the vertical axis and the method used in the horizontal axis when estimating the curve length from Figure 4b using a Buffon-Steinhaus test system with $T = 0.7$. The blue bars correspond to the (x, ω) process and the orange bars correspond to the (x, y) process of the corresponding method.

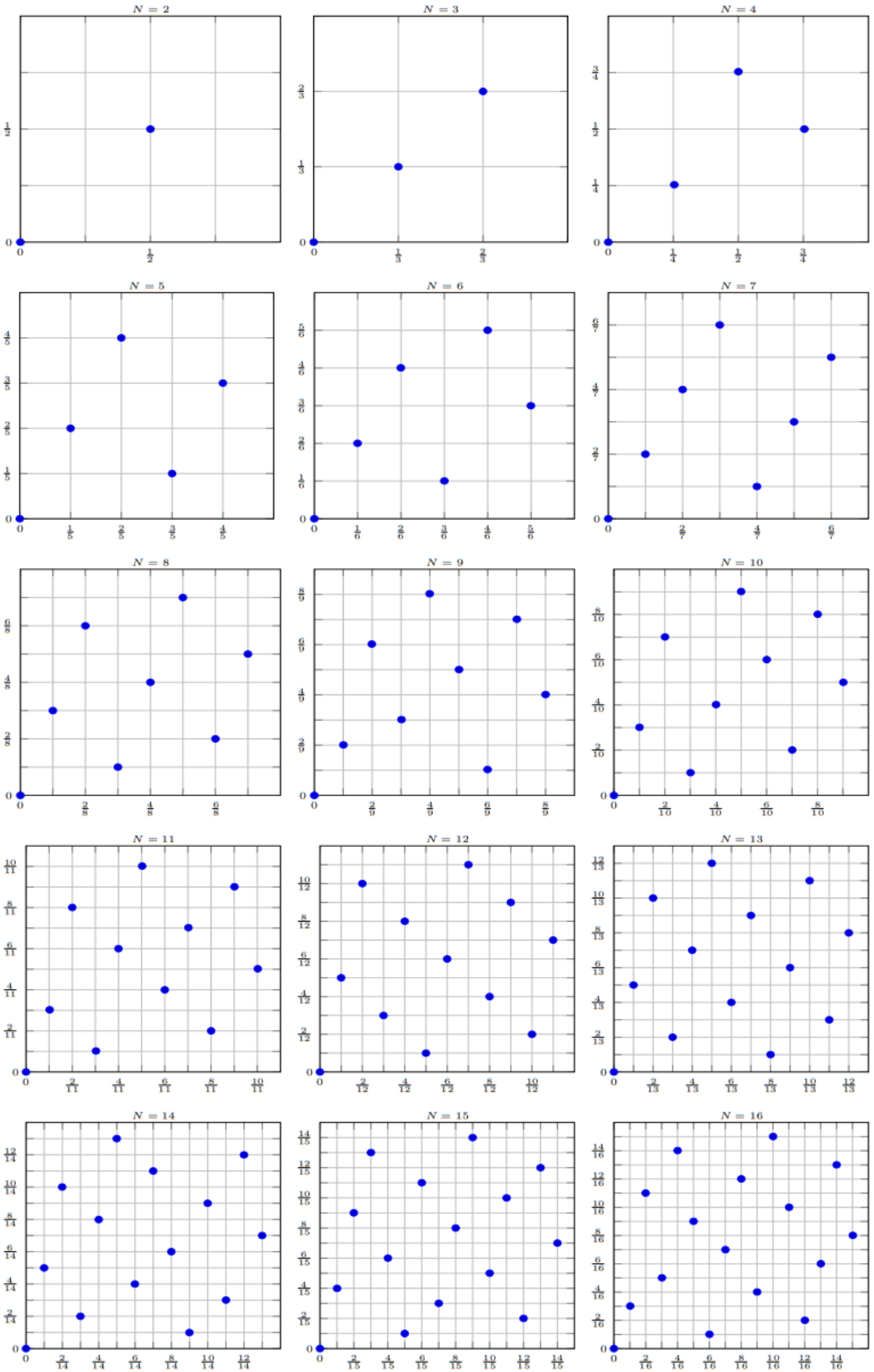


Figure 21: Optimal point sets for Quasi-Monte Carlo integration of bivariate periodic functions. Optimal point sets from $N = 2$ to $N = 16$ points and $\gamma = 1$ are shown (Hinrichs y Oettershagen [2016]). These are the points used for the Hinrichs method.