# Dependence concepts and selection criteria for lattice rules

by

Yoshihiro Taniguchi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Statistics

Waterloo, Ontario, Canada, 2013

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

**Abstract**

Lemieux recently proposed a new approach that studies randomized quasi-Monte Carlo through dependency concepts. By analyzing the dependency structure of a rank-1 lattice, Lemieux proposed a copula-based criterion with which we can find a "good generator" for the lattice. One drawback of the criterion is that it assumes that a given function can be well approximated by a bilinear function. It is not clear if this assumption holds in general.

In this thesis, we assess the validity and robustness of the copula-based criterion. We do this by working with bilinear functions, some practical problems such as Asian option pricing, and perfectly non-bilinear functions. We use the quasi-regression technique to study how bilinear a given function is. Beside assessing the validity of the bilinear assumption, we proposed the bilinear regression based criterion which combines the quasi-regression and the copula-based criterion. We extensively test the two criteria by comparing them to other well known criteria, such as the spectral test through numerical experiments. We find that the copula criterion can reduce the error size by a factor of 2 when the function is bilinear. We also find that the copula-based criterion shows competitive results even when a given function does not satisfy the bilinear assumption. We also see that our newly introduced BR criterion is competitive compared to well-known criteria.

**Acknowledgements**

First and foremost, I would like to thank my supervisor Dr. Christiane Lemieux for patient guidance and insightful comments about my research throughout this project. I would have never been able to complete this thesis without all the support. I would also like to thank my thesis committee members Dr. Adam Kolkiewicz and Dr. Carole Bernard for taking time to read my thesis and proving me with suggestions.

I am grateful to my office mates, especially Ethan and Yorker, for giving me a number of suggestions for my research. My special thanks goes to my best friends Raiyana Rahim and Taeho Kim. I would not have been able to make it through the difficult times without their warm encouragements.

Lastly, I would like to thank my parents for their unconditional love and support. I am really grateful for that they supported my decision to pursue a graduate degree.

## Dedication

This is dedicated to my sons and daughters in the future.

# Table of Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

The need to evaluate or approximate high dimensional integrals arises in various fields including statistics, physics, operations research and finance. For a very long time Monte Carlo (MC) had been considered the sole practical method capable of handling very high dimensional problems. Quasi-Monte Carlo (QMC) is now taking over MC in many fields as this method's integration error converges faster than MC.

The Korobov lattice is one of the most used construction of QMC due to its simplicity. In order to construct the Korobov lattice, we first need to select a parameter called the generator. As the quality of the point set produced by the Korobov lattice depends heavily on the choice of generator, it is crucial to select a 'good' generator. In order to know which one is a good generator, we usually select a search criterion that measures a certain quality of a given point set. Then we run a computer search to find the optimal generator with respect to the chosen search criterion. A number of search criteria have been proposed. In particular, the spectral test [3] and its variants are widely used criteria

In [13], Lemieux proposed a novel approach that studies RQMC through dependency concepts. Lemieux defined a copula-based criterion using dependency concepts and conducted numerical experiments that compared the performance of the Korobov lattice based on the generator from the copula-based and the ones from the spectral test. The numerical experiments in [13] showed favourable results for the copula-based criterion. However, the results of the numerical experiments were based on just one example. More numerical experiments are required to conclude whether nor not the copula-based criterion gives 'good' generators. Also, the copula-based criterion assumes that the function to be integrated can be well approximated by a bilinear function in ANOVA sense (see Section 3.2.2). That is, some function in a bilinear form can explain a large part of the variation of the function

of interest. It is not clear if such assumption holds for many problems.

The contribution of this thesis is two fold: numerical experiments are conducted to assess the robustness of the copula-based criterion and the newly proposed bilinear regression based (BR) criterion is introduced. We numerically examined four problems which consist of two parts. First, we assess the bilinear assumption of a given function. Then we compare the generators based on the copula-based criterion and the BR criterion to the generators based on other criteria. We use the technique called quasi-regression [15] to assess how much variance of a function is captured by a bilinear function. Also, by combining the copula-based criterion with quasi-regression, we propose a bilinear regression based (BR) criterion in this thesis. The BR criterion differs from other criteria in that it is tailored to the problem at hand.

Through numerical experiments, we found that some functions satisfy the bilinearity assumption, but others do not. We also saw that even if a function is bilinear based on a certain set of parameters, the function could be non-bilinear based on another set of parameters. However, we found that the copula criterion and the BR criterion are competitive to other criteria even when the bilinear assumption is not met. We believe that we have done enough numerical experiments to test the validity of the copula-based criterion and the BR criterion. Our conclusion is that the copula-based criterion provides generators that work well in the general setting. Also, the generators given by BR criterion that are designed to work well with one problem may work well with other problems.

The rest of this thesis is organized as follows. In Chapter 2, we provide the necessary background for MC, QMC and RQMC. The material introduced in this chapter is not meant to be a complete introduction to MC and QMC. We recommend readers to [14] for a comprehensive introduction to MC and QMC. In Chapter 3, we present dependency concepts for RQMC [13] and quasi-regression [15]. In the last section of the chapter, we discuss how we can use quasi-regression to assess the validity of the assumption made by the copula-based criterion. We also briefly provide the idea for the BR criterion in that section. In Chapter 4, we work specifically with bilinear functions and study whether the generators based on the copula-based criterion perform well under ideal cases. We also provide the detailed description of the BR criterion. In Chapter 5, we carry out numerical experiments with non-bilinear functions. In each problem, we construct bilinear regression of the function, assess how linear and bilinear the function is, obtain the BR generator in some cases, and run simulations to compare generators based on various criteria.

# Chapter 2

# Background

## 2.1 Monte Carlo

Suppose given a function $f(\mathbf{u}), \mathbf{u} \subset [0,1)^s$, we want to evaluate the integral

$$I(f) = \int_{[0,1)^s} f(\mathbf{u})\, d\mathbf{u}. \tag{2.1}$$

In general, the domain of a given function is not $[0,1)^s$. However, a change of variables can transform the domain into an $s$-dimensional unit hypercube without changing the resulting integral.

The key step of the Monte Carlo (MC) method is to look at the integral as an expectation in order to construct an estimator for $I(f)$. Suppose $\mathbf{U} \sim U[0,1)^s$. Let $g(\mathbf{u})$ denotes the probability density function of a uniform distribution of support $[0,1)^s$, then $g(\mathbf{u}) = 1$ for $\mathbf{u} \subset [0,1)^s$ and 0 elsewhere. Therefore, we have

$$E[f(\mathbf{U})] = \int_{[0,1)^s} f(\mathbf{u})g(\mathbf{u})\, d\mathbf{u} = \int_{[0,1)^s} f(\mathbf{u}) * 1\, d\mathbf{u} \tag{2.2}$$

$$= \int_{[0,1)^s} f(\mathbf{u})\, d\mathbf{u} = I(f). \tag{2.3}$$

The integration problem is equivalent to computing the expectation. Let $\mathbf{U}_1, \mathbf{U}_2, \cdots, \mathbf{U}_n$ be $n$ independent samples from $U[0,1)^s$. We can construct the estimator for $I(f)$ by taking

$\bar{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{U}_i)$. This is an unbiased estimator for $I(f)$ since

$$E[\frac{1}{n}\sum_{i=1}^{n} f(\mathbf{U}_i)] = \frac{1}{n} nE[f(\mathbf{U})] = I(f). \tag{2.4}$$

Furthermore, the Strong Law of Large Numbers assures that $\bar{\mu}_n$ converges to $I(f)$ as $n \to \infty$ with probability 1.

Computing MC estimates is straightforward although it does not mean it is computationally inexpensive. Typically, we follow the following three steps to estimate $I(f)$ by MC.

Step 1. Draw $s$ independent samples $u_1, \ldots, u_s$ from $U[0,1)$, and form an $s$-dimensional vector $\mathbf{u} = [u_1, \ldots, u_s]$. It is easy to see that $\mathbf{u}$ is a realization of $U[0,1)^s$.

Step 2. Evaluate $f(\mathbf{u})$.

Step 3. Repeat Step 1 and Step 2 $n$ times and take a sample mean of $f(\mathbf{u})$'s.

We usually use pseudo-random numbers instead of pure random numbers when sampling from $U[0,1)$. Pseudo-random numbers are a deterministic sequence of numbers which mimic the behaviour of samples from $U[0,1)^s$. There are a number of advantages to use pseudo-random numbers over pure random numbers. One of the advantages is speed. Obtaining pure random numbers requires observations of some physical processes. The problem is that we often need millions of samples for MC. Imagine how long it would take to get one million samples by observing physical processes. On the other hand, it takes a few seconds if not less to generate one million pseudo-random numbers on modern computers.

The next question is, how accurate is the MC estimate? The estimate is guaranteed to coincide with the true value only when the number of samples is infinite. In practice, as only finite samples are available, the MC estimates have some errors. It is very important to know the error bound when using any kind of numerical method, as the estimate itself does not mean anything. Suppose that the estimate of integral is 5. If the error bound is 1000, the estimate is probably far off the true value. If the error bound is 0.001, then we know that the estimate is much more accurate.

The Central Limit Theorem (CLT) allows us to derive a probabilistic error bound in the form of a confidence interval. The CLT states that

$$\frac{\bar{\mu}_n - I(f)}{\sigma_f/\sqrt{n}} \to N(0,1), \tag{2.5}$$

where $\sigma_f$ denotes the standard deviation of the function $f$. The variance (square of standard deviation) is calculated by

$$\sigma_f^2 = \int_{[0,1)^s} (f(\mathbf{u}) - I(f))^2 \, d\mathbf{u}. \tag{2.6}$$

We can show that the approximate $100(1-\alpha)\%$ confidence interval for $I(f)$ is

$$\{\hat{\mu}_n \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_f}{\sqrt{n}}\}, \tag{2.7}$$

$$\tag{2.8}$$

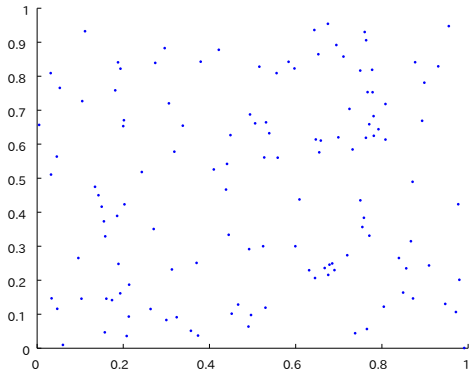
where $z_\alpha$ denotes the point at which $P(Z < z_\alpha) = \alpha$ for $Z \sim N(0,1)$. In general, $I(f)$ is unknown, so $\sigma_f$ is also unknown. We replace $\sigma_f$ by the sample standard deviation

$$\hat{\sigma}_f = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (f(\mathbf{u}_i) - \hat{\mu}_n)^2}. \tag{2.9}$$

As $\alpha$ is chosen prior to simulation and $\sigma_f$ is a constant, the error converges to 0 at rate of $O(\frac{1}{\sqrt{n}})$. It is worthwhile to notice that the convergence rate $O(\frac{1}{\sqrt{n}})$ is independent of $s$. This is one of the main justification why we prefer MC over deterministic numerical scheme for high dimensional problems. Deterministic methods suffer from what is called the *curse of dimensionality*, the phenomenon that the rate of convergence exponentially deteriorates with $s$. The convergence rate for MC is independent of $s$, but it is still slow. In order to improve the accuracy by an order of $k$, we need $k^2$ times as many sampling and evaluation of $f(\mathbf{u})$. If we want one more digit of accuracy, for instance, we need 100 times as many evaluations of $f$.

Suppose that all points happens to be sampled from a small subset of the entire domain. Then we know the estimate is unreliable. When computing an integral by MC, intuitively the estimate is more accurate if the sampled points cover the entire domain uniformly. Unfortunately, sampling that relies purely on pseudo-random numbers does not behave this way. Figure 2.1 shows 128 two-dimensional pseudo-random numbers. It shows that samples are not uniformly scattered. There are clusters of points while there are regions where no points are sampled. This poor coverage of the domain partially explains the slow convergence of MC. It would be advantageous if we had a sampling scheme where each sample somehow avoids each other. Quasi-Monte Carlo takes advantage of such sampling scheme, and it provides faster convergence.

Figure 2.1: Two dimensional Pseudo-Random Numbers




## 2.2 Quasi-Monte Carlo

Similar to the MC estimate, the QMC estimate is $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{u}_i)$, but the points $\mathbf{u}_i$ come from a low-discrepancy sequence instead of pseudo-random numbers. The precise definition of low-discrepancy sequence will be given in Section 2.2.2. A low-discrepancy sequence has a property of being more equidistributed than random numbers. That is, a low-discrepancy sequence is more uniformly scattered within the domain of interest.

When we use the term sequence, we mean an infinite sequence of points. In practice we use only a finite subset from such sequence. We refer to those subsets as point sets. A low-discrepancy point set is a finite set of points obtained from a low-discrepancy sequence.

Various ways to construct low-discrepancy sequences have been proposed; see [14] for a survey of low-discrepancy sequences. This thesis focuses on what is known as a Korobov lattice introduced by Korobov [22] and Hlawka [7]. A Korobov lattice is in turn a special case of a rank-1 lattice point set.

### 2.2.1 Lattice Point Set

We follow the notation introduced in [14] to describe a lattice point set. Given a generating vector $\mathbf{z}=(z_1, \ldots, z_s)$ of $s$ integers smaller than $n$, a rank-1 lattice point set is defined by

$$P_n = \{\tfrac{i}{n}(z_1, \ldots, z_s) \bmod 1, i = 0, \ldots, n-1\},$$

where the modulo operation is applied component-wise. A point set $P_n$ is said to be fully projection regular if for any non-empty subset $I \subseteq \{1, \ldots, s\}$, $P_n(I)$ contains as many distinct points as in $P_n$, where $P_n(I)$ is the projection of $f$ over the dimensions $j \in I$. In other words, if we look at one fixed coordinate of a fully projection-regular point set, each point has different value. A rank-1 point set is fully projection-regular if each $z_j$ is co-prime with $n$ [14], i.e. $\gcd(z_j, n) = 1$ for $j = 1, \ldots, s$.

A Korobov lattice assumes that the generating vector is of the form $\mathbf{z} = (1, a, a^2, \ldots, a^{s-1}) \bmod n$ for some integer $a$ smaller than $n$. Therefore, a Korobov point set is of the form

$$P_n = \{\tfrac{i}{n}(1, a \bmod n, a^2 \bmod n \ldots, a^{s-1} \bmod n) \bmod 1, i = 0, \ldots, n-1\}$$

It is trivial to see that as long as $\gcd(a, n) = 1$, a Korobov lattice is fully projection-regular. Generally, $n$ is prime or of the form $b^k$ for some prime $b$ and some positive integer $k$. In the former case, for any $1 \leq a \leq n-1$, the point set is projection-regular and for the latter case, all we need is $\gcd(a,\ b) = 1$.

Figure 2.2: Korobov Point Set

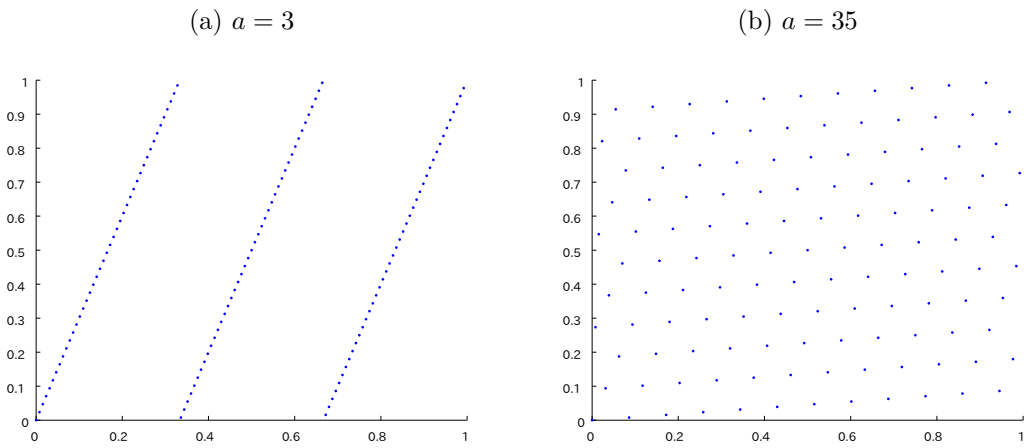



Fig 2.2 illustrates that the distribution property of a Korobov lattice depends on the choice of $a$. When $a = 3$, all the points lie on three lines and fail to fill most of the domain. When $a = 35$, the lattice structure of the point set is observable. The point set covers the entire domain uniformly. This suggests that depending on the choice of generator, a Korobov lattice could do better or worse than MC.

It is obvious that $a = 35$ is a better choice than $a = 3$ by looking at the plot. It may not be possible to tell whether the Korobov lattice with $a = 35$ is more uniformly

distributed than another one with $a = 51$ just by looking at the plot, especially when the dimensionality is high.

One may question if there is any way to quantify the uniformity of a given point set. Such measure will allow us to compare two point sets and choose the better one. The star-discrepancy is a widely used measure of uniformity.

### 2.2.2 Star-Discrepancy and Error bound of QMC estimate

Again, we follow the notations in [14]. The star discrepancy of a point set $P_n$ is given by

$$D^*(P_n) = \sup_{\mathbf{v}\in[0,1)^s} |v_1 \dots v_s - \alpha(P_n, \mathbf{v})/n|,$$

where $\alpha(P_n, \mathbf{v})$ is the number of points from $P_n$ that are in $\prod_{j=1}^{s}[0, v_j)$. Take a hyper-rectangle $H$ of the form $\prod_{j=1}^{s}[0, v_j)$. Suppose that the volume of $H$ is $V$. If the point set $P_n$ is truly equidistributed, exactly $V \cdot n$ of all points should lie in $H$, for all $H$. In that case, the star-discrepancy of $P_n$ is 0. On the other hand, if all the points lie in a small cluster, the star-discrepancy is close to 1. All sampling schemes are between the two cases. A sequence of points is called a low-discrepancy sequence if $D^*(P_n) \in O(n^{-1}(\log n)^s)$. There is a direct relationship between the star-discrepancy and the error bound of a QMC estimator. More precisely, the *Koksma-Halwaka Inequality* [7] states that

$$|\hat{\mu}_n - I(f)| \leq D^*(P_n)V(f),$$

provided $V(f)$, *the variation in the sense of Hardy and Krause*, is finite. If we use a low-discrepancy sequence for sampling, the error is in $O(n^{-1}(\log n)^s)$. Recall that the error bound of MC estimator is $O(\frac{1}{\sqrt{n}})$. For fixed $s$, the error of QMC estimate is asymptotically smaller than that of MC estimate. This justifies the use of QMC over MC. However, this error bound suggests that the accuracy of QMC deteriorates as the dimensionality of a problem increases.

The problem with this error bound is that it is virtually impossible to compute. It is very hard to calculate the $D^*(P_n)$ as well as $V(f)$ term. Even if we were able to compute the error bound, the bound is known to be too conservative to be useful. This problem can be solved by randomizing the underlying low-discrepancy sequence. This technique allows us to derive probabilistic error bounds and will be discussed in the next section.

## 2.3 Randomized Quasi-Monte Carlo

As mentioned in [14], "randomized quasi-Monte Carlo consists in choosing a deterministic low-discrepancy point set $P_n$ and applying a randomization such that (i) each point $\tilde{\mathbf{u}}_i$ in the randomized point set $\tilde{P}_n$ is $U[0,1)^s$ and (ii) the low-discrepancy of $P_n$ is preserved (in some sense) after the randomization". In order to construct an RQMC estimator, first note that

$$\hat{\mu}_{rqmc,l} = \frac{1}{n}\sum_{i=0}^{n-1} f(\tilde{\mathbf{u}}_i)$$

is an unbiased estimator for $I(f)$ since each $\tilde{\mathbf{u}}_i \sim U[0,1)^s$. Randomizing $P_n$ $m$ times gives $\tilde{P}_{n,1} \ldots, \tilde{P}_{n,m}$. With these, we obtain $\{\hat{\mu}_{rqmc,1} \ldots \hat{\mu}_{rqmc,m}\}$. Each estimator is an unbiased estimator of $I(f)$. Moreover, those estimators are independently and identically distributed (iid). This iid condition allows us to use the Central Limit Theorem and derive probabilistic error bounds. Let

$$\hat{\mu}_{m,rqmc} = \frac{1}{m}\sum_{i=1}^{m} \hat{\mu}_{rqmc,i}.$$

Note the slightly confusing notation. The estimator $\hat{\mu}_{rqmc,l}$ is obtained from one randomization of $P_n$. The subscript $l$ means that the estimator comes from the $l^{th}$ randomization. On the other hand, $\hat{\mu}_{m,rqmc}$ is the mean of those $m$ estimators. Clearly $\hat{\mu}_{m,rqmc}$ is an unbiased estimator of $I(f)$. The approximate $100(1-\alpha)\%$ confidence interval is

$$\{\hat{\mu}_{m,rqmc} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma_{m,rqmc}}{\sqrt{m}}\}.$$

As $\sigma_{m,rqmc}$ is unknown, we replace it with the sample standard deviation

$$\hat{\sigma}_{m,rqmc} = \sqrt{\frac{1}{m-1}\sum_{i=1}^{m}(\hat{\mu}_{rqmc,i} - \hat{\mu}_{m,rqmc})^2}.$$

The approximate $100(1-\alpha)\%$ confidence interval becomes

$$\{\hat{\mu}_{m,rqmc} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\hat{\sigma}_{m,rqmc}}{\sqrt{m}}\}.$$

### 2.3.1 Shifted Korobov Lattice

Random shift is a commonly used randomization technique for a rank-1 lattice. Suppose $P_n = \{\mathbf{u}_0, \ldots, \mathbf{u}_{n-1}\}$ is a rank-1 lattice point set and $\mathbf{v} \sim U[0,1)^s$. Let

$$\tilde{\mathbf{u}}_i = (\mathbf{u}_i + \mathbf{v}) \bmod 1 \text{ for } i = 0, \ldots, n-1.$$

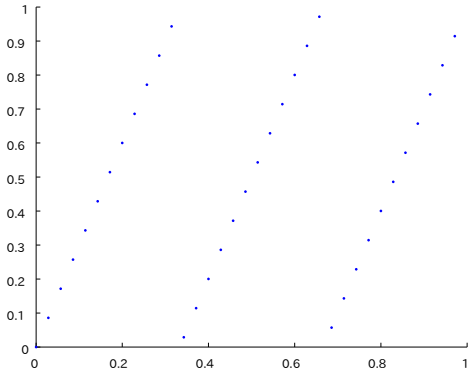
Then, $\tilde{P}_n = \{\tilde{\mathbf{u}}_0, \ldots, \tilde{\mathbf{u}}_{n-1}\}$ is a shifted Korobov lattice point set.
We need only $m$ iid samples $\mathbf{v}_1, \ldots \mathbf{v}_m$ from $U[0,1)^s$ to randomize $P_n$ $m$ times. The $l^{th}$ randomized point set is given by

$$\tilde{P}_{n,l} = \{(\mathbf{u}_i + \mathbf{v}_l) \bmod 1,\ i = 0, \ldots, n-1\}.$$

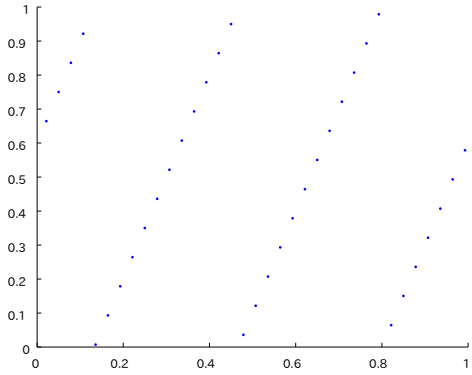
Figure 2.3 compares a Korobov lattice and a randomly shifted version of it. The lattice point set on the left is generated with parameters $n = 35$ and $a = 3$. Then the shift of (0.25,0.35) was applied to obtain the point set on the right. As we can see, a random shift shifts a lattice point set without breaking the lattice structure.

Figure 2.3: Korobov Point Set

(a) Korobov lattice




(b) Shifted Korobov lattice




As discussed earlier, the parameter we need to construct a lattice point set is the generating vector, an $s$-dimensional vector of integers less than $n$. For a Korobov lattice, the parameter is $a$, an integer less than $n$. We discussed conditions under which the lattice is fully projection-regular. However, that should not be the only consideration when choosing generators. As we saw, the choice of parameter has a significant effect on the uniformity

of a lattice point set. We usually select a search criterion that measures certain quality of a point set, and use a computer search to find the optimal parameters with respect to the criterion. The next chapter introduces the copula-based criterion proposed in [13].

# Chapter 3

# Dependence Concepts and Quasi-Regression

This chapter consists of three large sections. The first two sections introduce the necessary background for the rest of this thesis. The last section combines the material presented in the first two sections. Lemieux proposed a new approach in [13] that studies RQMC through analyzing its dependency structure instead of using the traditional discrepancy concepts. As the aim of this thesis is to expand on this approach, the first section of this chapter is dedicated to reintroduce important concepts in [13] which are relevant to this thesis. The second section is dedicated to [15] which introduces the quasi-regression approach to ANOVA decomposition. We will discuss how quasi-regression techniques can be used to assess the bilinearity assumption made for the copula-based search criterion defined in [13]. In the third section, we combine the copula-based criterion and quasi-regression to define a new search criterion.

## 3.1 Dependence concepts for randomized quasi-Monte Carlo methods

### 3.1.1 Overview

*Antithetic sampling* (AS) and *Latin hypercube sampling* (LHS) are two well-known variance reduction techniques. Both techniques use correlated sampling schemes that have the *negative quadrant dependence* (NQD) property. The fact that the NQD property is

preserved under monotone transformations combined with *Hoeffding's Lemma* can be used to show that the estimators based on AS and LHS have smaller variance than naive MC estimators for monotone functions, as shown in [13].

Randomized quasi-Monte Carlo (RQMC) also uses correlated sampling schemes. The question is whether RQMC satisfy the NQD property like AV and LHS do. Lemieux derives the copula for an one-dimensional *randomly shifted grid* (RSG), a particular construction of one-dimensional RQMC methods, and shows that RSG does not satisfy the NQD property. So, we cannot use the same argument for AV and LHS to show that a RSG gives better estimates than naive MC for monotone functions. Lemieux then turns to the derivation of the Spearman rho for one-dimensional RSG. This quantity plays an important role as we can think of the Spearman rho as measuring the average NQD. It can be shown that a one-dimensional RSG in fact has negative Spearman rho when averaged over all possible pairs of samples.

Then the one-dimensional analysis is extended to the multidimensional case. First, the covariance expression $\mathrm{Cov}(f(\mathbf{U}), f(\mathbf{V}))$, where $f$ is defined over $[0,1)^s$ is derived in terms of the sampling scheme's copula, which bears a close resemblance to the *Hlawka-Zaremba* identity. Then the copula and the Spearman rho for an $s$-dimensional rank-1 lattice point sets are derived. Finally, Lemieux proposes a search criterion for Korobov lattices based on the corresponding copula.

The rest of this section is organized as follows. In Section 3.1.2, the necessary background along with important definitions are introduced. In Section 3.1.3, the analysis for a one-dimenional RSG is presented. In Section 3.1.4, the analysis is extended to the $s$-dimensional case and the copula based search criterion is defined. This section ends with Section 3.1.5, which discusses possible combinations of the copula-based search criteria and quasi-regression techniques.

### 3.1.2 Background

Recall that the size of probabilistic error bounds of the MC method is proportional to

$$\sigma_{\hat{\mu}_n} = \frac{\sigma_f}{\sqrt{n}}.$$

So, reducing the error of MC estimators is equivalent to reducing the standard deviation. As discussed in the previous chapter, the size of the error bound from MC remains relatively large even if a large number of samples are taken.

We often combine naive MC with some kind of variance reduction techniques to obtain estimators with smaller variance. Variance reduction techniques generally transform $f$ into $\tilde{f}$ that has the same expectation as $f$, but for which $\sigma_{\tilde{f}} < \sigma_f$. The first property ensures that the estimator based on variance reduction techniques has a correct expectation, while the latter property means that the estimator has smaller variance than that from the naive estimator.

Two of the popular variance reduction techniques are *antithetic variates*(AV) and *Latin Hypercube Sampling*(LHS). Instead of drawing samples that are all independent, both techniques are designed in such a way that samples are correlated. Before describing what AV and LHS are, we first introduce two important definitions.

**Definition 3.1.1** A point set $P_n = \{\mathbf{U}_1, \dots \mathbf{U}_n\} \in [0,1)^s$ is said to be a *2-exchangeable uniform sampling scheme* if
(1) each $\mathbf{U}_i \sim U[0,1)^s$;
(2) every pair ($\mathbf{U}$,$\mathbf{V}$) of distinct points in $P_n$ has the same joint distribution.

The points from a typical construction of RQMC is ordered in a way that it fails to satisfy the conditions to be a 2-exchangeable uniform sampling scheme. Thus, we assume that a given RQMC point set is randomly permuted before constructing a RQMC estimator. This has no effect on the value of the estimator but the point set then becomes a 2-exchangeable uniform sampling scheme.

**Definition 3.1.2** Consider two random variables X and Y. We say that X and Y are *negatively quadrant dependent* (NQD) [12] if

$$P(X \le x, Y \le y) \le P(X \le x)P(Y \le y)$$

for all $x, y \in \mathbb{R}$.

Two random variables are NQD if they are likely to behave in opposite directions. That is, when one is large, the other tends to be small. Now we describe AV and LHS. AV was originally proposed by Hammersley and Handscomb [6]. Naive MC forms $P_n = \{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ by taking $\mathbf{U}_i \overset{iid}{\sim} U[0,1)^s$ for $i = 1, 2, \dots n$.
With AV, assuming $n$ is even, we take $\mathbf{U}_i \overset{iid}{\sim} U[0,1)^s$ for $i = 1, 3, \dots n-1$ and $\mathbf{U}_i = 1 - \mathbf{U}_{i-1}$ for $i = 2, 4, \dots n$, where the subtraction $1 - \mathbf{U}_{i-1}$ is applied component wise. It is obvious that under AV the pairs $(\mathbf{U}_{2i-1}, \mathbf{U}_{2i})$ are negatively correlated for $i = 1, 2, \dots, \frac{n}{2}$.

LHS was first introduced by Mckay et al. [18]. The idea is to randomly generate $s$ iid permutations $\pi_1, \dots, \pi_s$ of $[1, \dots, n]$ and let

$$U_{ij} = \frac{\pi_j[i]-1}{n} + V_{ij}, \text{ where } V_{ij} \overset{iid}{\sim} U[0, 1/n) \text{ for } i = 1, \dots n, j = 1, \dots s.$$

This construction is such that each $s$ one-dimensional projection of $P_n$ has exactly one point in each of the intervals $[(i - 1/n, i/n)], i = 1, \dots n$.

It can be shown that AV and LHS are *2-exchangeable uniform sampling schemes* that are NQD, even though the negative dependency is not obvious for LHS. Our interest is whether AV and LHS give estimators with smaller variance compared to naive MC. The following notation is useful to derive sufficient conditions for variance reduction.

**Definition 3.1.3** Let $f$ be square-integrable function and let $\hat{\mu}_n$ be a MC estimator for $I(f)$ based on *2-exchangeable uniform sampling scheme* $P_n$. We define $\sigma_{uv,f}$ as

$$\sigma_{uv,f} := \text{Cov}(f(\mathbf{U}), f(\mathbf{V})),$$

the (common) covariance term between a generic pair of points ($\mathbf{U}$,$\mathbf{V}$).

This covariance term plays an essential role when comparing MC and other sampling schemes as we have

$$\text{Var}(\hat{\mu}_n) = \text{Var}(\hat{\mu}_{mc,n}) + \frac{n-1}{n}\sigma_{uv,f}, \tag{3.1}$$

where $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{U}_i)$ and $\mathbf{U}_i \overset{iid}{\sim} U[0, 1)^s$.

So, if $\sigma_{uv,f} < 0$, the corresponding method gives estimators with smaller variance than naive MC does. Note that having an NQD sampling scheme does not always translate to $\text{Cov}(f(\mathbf{U}), f(\mathbf{V})) < 0$. That is, having $U$ and $V$ negatively correlated does not necessary mean that $f(\mathbf{U})$ and $f(\mathbf{V})$ are negatively correlated. We are interested in a class of functions that preserve the correlation among the sample. The following theorem [12] provides such class of functions.

**Theorem 3.1.4 (LEHMANN)** Let $\mathbf{U} = (U_1, \dots, U_s)$ and $\mathbf{V}= (V_1, \dots, V_s)$ be two vectors of $U(0, 1)$ rv's such that the s pairs $(U_j, V_j)$ are iid with common joint distribution that is NQD. Assume $f : [0, 1)^s \to \mathbb{R}$ is monotone in each of its $s$ arguments and let $X = f(U_1, \dots, U_s)$ and $Y = f(V_1, \dots, V_s)$. Then $(X, Y)$ is NQD.

Note that $(X, Y)$ being NQD implies that $\sigma_{uv,f} < 0$. Now we can present the well-known theorem.

**Theorem 3.1.5** Consider a *2-exchangeable uniform sampling scheme* $P_n$ with corresponding estimator $\hat{\mu}_n$ such that for any pair of distinct points $(\mathbf{U},\mathbf{V})$, $(U_j, V_j)$ is NQD for each $j = 1, \ldots, s$. Assume $f : [0,1)^s \to \mathbb{R}$ is monotone in each of its $s$ arguments and square-integrable. Then

$$\mathrm{Var}(\hat{\mu}_n) \leq \mathrm{Var}(\hat{\mu}_{mc,n}). \tag{3.2}$$

If $f$ is a monotone function, the sampling scheme based on AV or LHS, being NQD, gives estimators with smaller variance compared to naive MC.

We can write the conditions of being NQD in terms of the copula associated with the sampling sheme if the underlying random variables are $U[0,1)$.

**Definition 3.1.6** A copula $C(u_1, \ldots u_s)$ is a joint CDF defined over $[0,1)^s$ such that each of its $s$ marginal distributions are $U(0,1)$.

For$X, Y \sim U[0,1)$, the NQD condition is equivalent to $C_{X,Y}(u,v) \leq uv, \forall (u,v) \in \mathbb{R}^2$, where $C_{X,Y}(u,v) = P(X \leq u, Y \leq y)$.

Then Lemieux makes a connection with the *Spearman rho* of X and Y, which is defined in [19] to be

$$\rho_{X,Y} = 12 \int\limits_{[0,1)^2} C_{X,Y}(u,v)dudv - 3. \tag{3.3}$$

Suppose that $X, Y \sim U[0,1)$. Satisfying the NQD property is a stronger condition than having a negative Spearman rho. If the joint distribution of $(X,Y)$ is NQD, then $\rho_{X,Y} < 0$. The converse is not generally true. Note that $\rho_{X,Y} < 0$ if we have $C_{X,Y}(u,v) \leq uv$ when averaged over all $(u,v) \in [0,1)$. In other words, we can think of a negative Spearman rho as equivalent to being NQD on average.

### 3.1.3 One-Dimensional Case

When $s = 1$, a randomly shifted grid is given by

$$P_n = \left\{ U_i = \left( \frac{i-1}{n} + W \right) \bmod 1, i = 1, \ldots n \right\}, . \tag{3.4}$$

where $W \sim U[0,1)$.

The following Lemma and Proposition give the copula and the Spearman rho for a one-dimensional grid.

**Lemma 3.1.7** Assume $i < j$ and let $k = j - i$ and $\theta = k/n$. Then

$$C_k := P(U_i \leq u, U_j \leq v) = \begin{cases} 0 & \text{if } u \leq 1-\theta, v \leq \theta \\ \min(u, v-\theta) & \text{if } u \leq 1-\theta, v > \theta \\ \min(u-(1-\theta), v) & \text{if } u > 1-\theta, v \leq \theta \\ u+v-1 & \text{if } u > 1-\theta, v > \theta. \end{cases}$$

*Proof.* See the proof for LEMMA 3.1 in [13].

Figure 3.1: Joint distribution of $U_i$ and $U_j$ for $j > i$

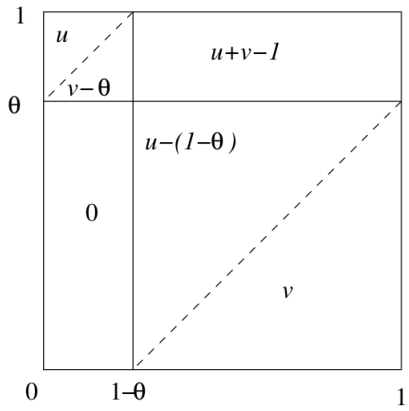



Figure 3.1, a direct copy of Fig 2 in [13], illustrates the joint distribution of $(U_i, U_j)$ specified by Lemma 3.1.7. As the figure shows, the joint distribution of $(U_i, U_j)$ is not NQD in the lower left corner and the upper right corner of the domain. For instance, if $u \leq 1-\theta, v > \theta$ and $v - \theta > u$, we have $P(U_i \leq u, U_j \leq v) = u > uv$. That is, $U_i$ and $U_j$ are positively correlated in some region of the domain, which violates the conditions for NQD.

The next question is whether $(U_i, U_j)$ satisfies the weaker condition of having a negative Spearman rho.

**Proposition 3.1.8** Assume $i < j$ and $k = j - i$. Then the Spearman rho of $(U_i, U_j)$ is given by

$$\rho_k = \frac{1}{12} - \frac{1}{2}\left[\frac{k}{n}\left(1 - \frac{k}{n}\right)\right]$$

and therefore $\rho_k \leq 0$ if and only if $1/2(1 - 1/\sqrt{3}) \leq k/n \leq 1/2(1 + 1/\sqrt{3})$.

*Proof.* See the proof for PROPOSITION 3.4 in [13].

So, the Spearman rho of $(U_i, U_j)$ could be positive for some pairs. However, the Spearman rho is negative when averaged over all pairs as shown in the next proposition. This proposition is not explicitly stated in [13] but implied by Proposition 3.7 in [13].

**Proposition 3.1.9** *Suppose we have two distinct randomly chosen points $U$ and $V$ from (3.4). Then the Spearman rho of the corresponding copula for the pair $(U, V)$ is*

$$\rho_{lr} = -\frac{1}{n}. \tag{3.5}$$

*Proof.*

First, we use the fact that

$$\rho_{lr} = \frac{12}{n-1} \sum_{k=1}^{n-1} \left( \frac{1}{3} - \frac{k}{2n} \left( 1 - \frac{1}{n} \right) \right) - 3.$$

It is easy to see that

$$\sum_{k=1}^{n-1} \left( \frac{1}{3} - \frac{k}{2n} \left( 1 - \frac{1}{n} \right) \right) = \frac{3n^2 - 4n + 1}{12n} = \frac{(3n-1)(n-1)}{12n}.$$

$$\text{So, } \rho_{lr} = \frac{12}{n-1} \frac{(3n-1)(n-1)}{12n} - 3 = -\frac{1}{n}.$$

The following proposition relates the copula, the Spearman rho and $\sigma_{uv,f}$ for a randomly shifted grid.

**Proposition 3.1.10** Let $f(u)$ be a continuous function over [0,1) such that $F := P(f(U) \leq x)$ is strictly increasing and absolutely continuous. Consider a 2-exchangeable uniform sampling scheme $P_n$ and let $C(u, v; 1)$ denote the copula for two distinct points $(U, V)$ in $P_n$. Then

$$\sigma_{uv,f} = \int_{[0,1)^2} (C(u, v; 1) - uv) dF^{-1}(u) dF^{-1}(v), \tag{3.6}$$

where the '1' listed in the argument of $C(\cdot)$ refers to the dimension of the lattice point set under consideration.

*Proof.* See the proof for PROPOSITION 3.9 in [13].

If $P_n$ is a NQD sampling scheme, $C(u, v; 1) < uv$ and so $\sigma_{uv,f} < 0$. This aligns with the discussion we had before. Lemieux discusses that having $\sigma_{uv,f} < 0$ is almost equivalent to having $\rho < 0$. The difference between (3.3) and (3.6) are the $dudv$ term in the former expression and the $dF^{-1}(u)dF^{-1}(v)$ term for the latter one. We can think of $dF^{-1}(\cdot)$ term as the change of measure induced by $f$, as Lemieux discusses.

### 3.1.4 Multidimensional case

This section generalizes the one-dimensional analysis of the previous section to the $s$-dimensional case. We start with an $s$-dimensional version of (3.6).

**Theorem 3.1.11** Let $f$ be a left-continuous function of bounded variation in the sense of Hardy and Krause. Let $P_n \in [0, 1)^s$ be a 2-exchangeable uniform sampling scheme. Then

$$\sigma_{uv,f} = \int_{[0,1)^{2s}} \sum_{\emptyset \neq K \in S} E_n(\mathbf{u}, \mathbf{v}; K) W(f; K) d\mathbf{u} d\mathbf{v}, \tag{3.7}$$

where

$$E_n(\mathbf{u}, \mathbf{v}; K) := C(\mathbf{u}, \mathbf{v}; K) - \prod_{i \in} u_i v_i$$

$$W(f; K) := \sum_{I,J \subseteq S : I \cap J = K} (-1)^{|I|+|J|} df_I(\mathbf{u}_I) df_J(\mathbf{u}_J) \prod_{i \in I \setminus K} u_i \prod_{j \in J \setminus K} v_j.$$

*Proof.* See the proof for Theorem 4.1 in [13].

The above expression bears close resemblance to the Hlawka-Zaremba identity [24] in that both of them separate the quantity of interest into two parts: one that depends only on the point set and the other only on $f$.

The following proposition gives the expression for the copula of a two-dimensional Korobov lattice.

**Proposition 3.1.12** *For a two-dimensional randomly-shifted Korobov lattice with generator $a$, we have*

$$C_{lr}(\boldsymbol{u}, \boldsymbol{v}; \{1, 2\}) = \frac{1}{n-1} \sum_{k=1}^{n-1} C_k(u_1, v_1) C_{\overline{ak}_n}(u_2, v_2),$$

where $\overline{ak}_n$ denotes $ak$ mod $n$.
*Proof.* See the proof for PROPOSITION 4.2 in [13].

The following two propositions generalize the copula and the Spearman rho of a one-dimensional grid to those of a $s$-dimensional rank-1 lattice.

**Proposition 3.1.13** For a randomly shifted rank-1 lattice based on the generating vector $(z_1, \dots, z_s)$ and for a subset $I \subseteq \{1, \dots s\}$, we have

$$C_{lr}(\mathbf{u}_I, \mathbf{v}_I; I) = \frac{1}{n(n-1)} \sum_{k=1}^{n-1} \prod_{j \in I} C_{(\overline{kz_j})_n}(u_j, v_j)$$

*and*

$$\rho_{lr,I} = \frac{2^d(d+1)}{2^d - (d+1)} \left( \frac{1}{n(n-1)} \sum_{k=1}^{n-1} \prod_{j \in I} \left( \frac{1}{3} - \frac{(\overline{kz_j})_n/n(1 - (\overline{kz_j})_n/n)}{2} \right) - 4^{-|I|} \right)$$

*where* $d = 2|I|$.

*Proof.* See the proof for PROPOSITION 4.5 in [13].

The following proposition becomes useful when defining a copula-based search criterion.

**Proposition 3.1.14** *Suppose we have a function that can be well approximated (in an ANOVA sense, see Section 3.2.1 or [2] for example) by a sum of the form*

$$\tilde{f}(\boldsymbol{u}) := \sum_{K \in \mathcal{I}} c_K \prod_{j \in K} u_j, \tag{3.8}$$

*where* $\mathcal{I}$ *contains different subsets* $K \subseteq \{1, \dots, s\}$. *Then the covariance term* $\sigma_{uv,\tilde{f}}$ *for a 2-exchangeable uniform sampling scheme* $P_n$ *is of the form*

$$\sigma_{uv,\tilde{f}} = \sum_{K \in \mathcal{I}} \gamma_K \rho_{lr,K}, \tag{3.9}$$

*where the* $\gamma_K = \frac{2^{2k} - (2k+1)}{(2k+1)2^{2k}}$ *are weights depending on the function* $\tilde{f}$, *and* $\rho_{lr,K}$ *is the Spearman rho associated with the copula* $C(\boldsymbol{u}, \boldsymbol{v}; K)$ *of* $P_n$.

*Proof.* See the proof for PROPOSITION 5.1 in [13].
As shown in the proof for PROPOSITION 5.1 in [13], the explicit expression for $\gamma_K$ is

$$\gamma_K = t_K(2^{2k} - (2k+1))((2k+1)2^{2k}) \text{ and}$$
$$t_K = \sum_{I,J \subseteq S: I \cap J = K} (-1)^{|I|+|J|} 2^{-l} \sum_{I \in G} d_G \sum_{J \in H} d_H,$$

where $l = |I| + |J| - 2|K|$.

In practice, we may have the intuition that $f$ can be well approximated by $\tilde{f}$ as in (3.8) but we may not know the coefficients $c_K$. In an effort to minimize (3.9) in such situation, Lemieux proposed a search criterion for Korobov lattice where the optimal generator $a$ minimizes

$$\gamma_M := \max_{1 \le m \le M} \rho_{lr,\{1,m\}} \tag{3.10}$$

for some positive integer $M$. The underlying assumption for the criterion is that a good approximation of $f$ can be obtained by a sum of univariate and bivariate terms, with the latter containing indices that are not too far apart. We also note that a Korobov point set is dimension-stationary [14], that is, we have that $P_n(I) = P_n(I + j)$ for any $I \subseteq \{1, 2, \dots s\}$. Hence, it is sufficient to consider two dimensional projections of the form $I = \{1, m\}$. Furthermore, a Korobov point set is *fully projection-regular* as long as $a$ is coprime with $n$.

### 3.1.5 Discussion

Lemieux's paper proposed a new framework that studies RQMC through dependency structure. The search criterion (3.10) assumes that $f$ can be well approximated by a bilinear function. In practice, it may not be clear if a bilinear function approximates $f$ well or not. The technique called quasi-regression allows us to evaluate the validity of the assumption. We describe quasi-regression in the next section.

## 3.2 Quasi-regression and the relative importance of the ANOVA component of a function

As discussed in the previous section, the copula-based criterion assumes that the function of interest $f$ can be well-approximated by a bilinear function $\tilde{f}$ in ANOVA sense. The

traditional ANOVA decomposition allows us to evaluate the contribution of a certain subset of variables to the overall variance of $f$. However, this approach does not allow us to evaluate how much variations of $f$ are due to the bilinear component $f$.

Fortunately, Lemieux and Owen's approach [15] to ANOVA decomposition that uses quasi-regression fulfills this purpose. As this technique is crucial to this thesis, we dedicate this section to introduce the relevant concepts in [15].

The rest of this section is organized as follows. Section 3.2.1 describes the traditional approach to ANOVA decomposition. Section 3.2.2 shows that one can indirectly obtain ANOVA decomposition through quasi-regression. In Section 3.2.3, we review unbiased and consistent estimators for quasi-regression. Section 3.2.4 discusses the uses of quasi-regression for the copula-based criterion (3.10).

### 3.2.1 ANOVA decomposition

The motivation of ANOVA decomposition arises from the fact that variables of $f$ are not equally important in many practical applications. ANOVA decomposition allows us to quantify relative importance of a certain subset of variables of $f$.

Assume that $\int_{[0,1)^s} f^2(\mathbf{u})d\mathbf{u} < \infty$. The ANOVA decomposition of $f$ is $f(\mathbf{u}) = \sum_{I\subseteq\{1,\dots,s\}} f_I(\mathbf{u})$, where

$$f_I(\mathbf{u}) := \int_{[0,1)^s} f(\mathbf{u})d\mathbf{u}_{I^c} - \sum_{J\subsetneq I} f_J(\mathbf{u}), \tag{3.11}$$

and $\mathbf{u}_{I^c} = (u_j)_{j\notin I}$ represents the vector of variables $u_j$ whose index $j$ is not in $I$. One can show that for any nonempty subset $I$, and $\int_{[0,1)^s} f_\emptyset(\mathbf{u})d\mathbf{u} = \mu$, where $\mu = \int_{[0,1)^s} f(\mathbf{u})d\mathbf{u}$.

The ANOVA decomposition is orthogonal, i.e., for two subsets $I \neq J \subseteq 1,\dots,s$ one has

$$\int_{[0,1)^s} f_I(\mathbf{u})f_J(\mathbf{u})d\mathbf{u} = 0. \tag{3.12}$$

From (3.12) we can decompose the variance of $f$ via

$$\sigma^2 := \int_{[0,1)^s} f^2(\mathbf{u})d\mathbf{u} - \mu^2 = \sum_{I\subseteq\{1,\dots,s\}} \sigma_I^2,$$

where $\sigma_I^2 = \int\limits_{[0,1)^s} f_I^2(\mathbf{u}) d\mathbf{u}$ for $I \neq \emptyset$, and $\sigma_\emptyset^2 = 0$.

The quantity $\sigma_I^2/\sigma^2$ represents the fraction of variance of $f$ due to component $f_I$ and we take this as a measure of the relative importance of $f_I$. The problem of this approach is that we cannot assess how linear or quadratic $f$ is. We turn to quasi-regression techniques in the next section.

Before moving to the next question, we define what is meant by "well approximated in ANOVA sense". Let $W$ be the set of all subsets of $I = \{1, \ldots, s\}$. Then we can write the ANOVA decomposition of $f$ as $f(\mathbf{u}) = \sum\limits_{I \in W} f_I(\mathbf{u})$, where $f_I(\mathbf{u})$ is defined previously. We can obtain an approximation of $f$ by summing over some $V \subseteq W$. More precisely, the approximation is $\tilde{f}(\mathbf{u}) = \sum\limits_{I \in V} f_I(\mathbf{u})$. Then $\sigma_f^2 = \sum\limits_{I \subseteq W} \sigma_I^2$ is the overall variance of $f$ and $\sigma_{\tilde{f}}^2 = \sum\limits_{I \subseteq V} \sigma_I^2$ is the variance of $f$ captured by $\tilde{f}$. We say $f$ is well approximated by $\tilde{f}$ in ANOVA sense if $\sigma_{\tilde{f}}^2/\sigma_f^2$ is close to 1.

### 3.2.2 Basis functions and Quasi-Regression

The basic idea of quasi-regression is to express an arbitrary function $f$ in a simpler form, so we have more insight on $f$. We often do this by expanding $f$ as a linear combination of functions in $s$-dimensional basis functions.

Given one-dimensional basis functions, we take tensor products of them to obtain a multidimensional basis as in [1]. We chose our basis to be a polynomial orthonormal basis.

The multidimensional basis is precisely the tensor products of normalized Legendre polynomials [4]. Write our univariate basis as $\{\phi_j(u) | j = 0, 1, 2 \ldots, \}$, where each $\phi_j(u)$ is the normalized Legendre polynomial of degree $j$. Since Legendre polynomials are defined on $[-1, 1]$, $\phi_j(u)$ represents the normalized Legendre polynomial of degree $j$ evaluated at $(2u - 1)$. Following are the first three basis functions:

$$\begin{aligned}
\phi_0(u) &= 1 \\
\phi_1(u) &= \sqrt{3}(2u - 1) \\
\phi_2(u) &= \frac{\sqrt{5}}{2}\left(3\left(2u - 1\right)^2 - 1\right).
\end{aligned}$$

For $\mathbf{r} = (r_1, \ldots, r_s)$, where each $r_j$ is a non-negative integer, define

$$\phi_{\mathbf{r}}(\mathbf{u}) = \prod_{j=1}^{s} \phi_{r_j}(u_j).$$

Then $\{\phi_{\mathbf{r}} | \mathbf{r} = (r_1, \ldots, r_s), r_j \geq 0, j = 1, \ldots s\}$ defines an $s$-dimensional basis. We can then expand $f$ as

$$f(\mathbf{u}) = \sum_{r_1=0}^{\infty} \cdots \sum_{r_s=0}^{\infty} \beta_{\mathbf{r}} \phi_{\mathbf{r}}(\mathbf{u}), \tag{3.13}$$

where the coefficients $\beta_{\mathbf{r}}$ in (3.13) are obtained via

$$\beta_{\mathbf{r}} = \int_{[0,1)^s} f(\mathbf{u}) \phi_{\mathbf{r}}(\mathbf{u}). \tag{3.14}$$

The integral in (3.14) can be approximated by MC or RQMC. We focus on constructing RQMC estimators of $\beta_{\mathbf{r}}$ in the next section as they are more accurate than MC estimators as shown in [15].

The following proposition connects ANOVA decomposition and quasi-regression.

**Proposition 3.2.1** *Let $f := [0,1)^s \to \mathbb{R}$ be a square-integrable function. For each subset $I \subseteq \{1, \ldots s\}$, define the component $f_I(\cdot)$ as in* (3.11)*. Let the coefficients $\beta_{\boldsymbol{r}}$ be defined as in* (3.14)*. Then for $I$ nonempty, one has*

$$\sigma_I^2 = \sum_{\boldsymbol{r} \in R(I)} \beta_{\boldsymbol{r}}^2, \tag{3.15}$$

*where the set $R(I)$ contains all vectors $\boldsymbol{r} = \{r_1, \ldots r_s\}$ whose entries are non-negative integers satisfying $r_j > 0$ if and only if $j \in I$.*

So, we can indirectly estimate $\sigma_I^2$ by a sum of $\beta_{\mathbf{r}}^2$. In practice, we have to truncate the infinite sum of the right-hand side of (3.14). One of the disadvantages of this approach is that the estimated $\sigma_I^2$ may not be accurate due to truncation, if the coefficients $\beta_{\mathbf{r}}$ decay slowly. One advantage of this approach is that we can sum certain subsets of $\beta_{\mathbf{r}}^2$ to gain insight in $f$.

For non-negative integers $d$ and $m \leq d$, define

$$R(I,d,m) := \{\mathbf{r} \in R(I) | \sum_{j=1}^{s} r_j \leq d, r_j \leq m \text{ for each } j \in I\}. \tag{3.16}$$

A truncated version of $\sigma_I^2$ is then

$$\sigma_{I,d,m}^2 = \sum_{\mathbf{r} \in R(I,d,m)} \beta_{\mathbf{r}}^2. \tag{3.17}$$

The parameter $d$ and $m$ are called the *degree* and *order*, respectively, in [1]. We can interpret $\sigma_{I,d,m}^2/\sigma^2$ as the fraction of the variance of $f$ explained by the approximation

$$\tilde{f} = \sum_{\mathbf{r} \in V(I,d,m)} \beta_{\mathbf{r}} \phi_{\mathbf{r}}(\mathbf{u}), \tag{3.18}$$

where $V(I,d,m) := \{\mathbf{r} \subseteq I | \sum_{j=1}^{s} r_j \leq d, r_j \leq m \text{ for each } j \in I\}$.

Given $R(I,d,m)$ and the corresponding $\tilde{f}$, if $\sigma_{I,d,m}^2/\sigma^2$ is close to 1, we say that $f$ is well approximated by $\tilde{f}$ in ANOVA sense.

Depending on the selection of $d$ and $m$, we can approximate $f$ by linear functions or quadratic functions and so on. For instance, we obtain a linear approximation of $f$ if we set $I = \{1, \ldots s\}$, $d = 1$, and $m = 1$. Also, we obtain a bilinear approximation (which includes the linear components) of $f$ if we set $I = \{1, \ldots s\}$, $d = 2$, and $m = 1$. We refer to the quasi-regression concerned with the linear and bilinear approximation (which includes the linear components) of $f$ as the linear and bilinear regression of $f$, respectively, and denoted by $\tilde{f}_L$ and $\tilde{f}_{LB}$, respectively. More explicitly, we have

$$\tilde{f}_L = \sum_{\mathbf{r} \in V(\{1,\ldots s\},1,1)} \beta_{\mathbf{r}} \phi_{\mathbf{r}}(\mathbf{u}) \text{ and} \tag{3.19}$$

$$\tilde{f}_{LB} = \sum_{\mathbf{r} \in V(\{1,\ldots s\},2,1)} \beta_{\mathbf{r}} \phi_{\mathbf{r}}(\mathbf{u}). \tag{3.20}$$

In relation to the copula-based criterion $\gamma_M$ defined in (3.10), it is relevant to consider the bilinear regression whose bivariate terms are included only if their indices are not

too far apart. We denote such bilinear regression of $f$ by $\tilde{f}_{LB,M}$. More precisely, let $V'(I,d,m,M) := \{\mathbf{r} \subseteq I | \sum_{j=1}^{s} r_j \leq d, r_j \leq m, r_i \text{ and } \max\{j : r_j > 0\} - \min\{j : r_j > 0\} - M\}$. Then we can write

$$\tilde{f}_{LB,M} = \sum_{\mathbf{r} \in V'(\{1,\ldots s\},2,1,M)} \beta_{\mathbf{r}} \phi_{\mathbf{r}}(\mathbf{u}).$$

Since $\phi_0(u) = 1$ and $\phi_1(u) = \sqrt{3}(2u-1)$, we can write

$$\begin{aligned}
\tilde{f}_L &= \mu + \sqrt{3}\sum_{i=1}^{s} c_i(2u_i - 1), \\
\tilde{f}_{LB} &= \mu + \sqrt{3}\sum_{i=1}^{s} c_i(2u_i - 1) + 3\sum_{i<j}^{s} d_{i,j}(2u_i - 1)(2u_i - 1), \\
\tilde{f}_{LB,M} &= \mu + \sqrt{3}\sum_{i=1}^{s} c_i(2u_i - 1) + 3\sum_{i<j, j-i<M}^{s} d_{i,j}(2u_i - 1)(2u_i - 1),
\end{aligned} \tag{3.21}$$

where:
$\mu$ is $\beta_{\mathbf{r}}$ with all the elements of $\mathbf{r}$ are zero,
$c_i$ is $\beta_{\mathbf{r}}$ with only the $i$th element of $\mathbf{r}$ is 1 and all the other elements are zero,
$d_{i,j}$ is $\beta_{\mathbf{r}}$ with only the $i$th an $j$th element of $\mathbf{r}$ are 1 and all the other elements are zero,

In the next section, we construct RQMC estimators for $\beta_{\mathbf{r}}$ and $\beta_{\mathbf{r}}^2$.

### 3.2.3 RQMC Estimators for Quasi-Regression

Suppose $P_n = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is a QMC point set. With $Q$ randomizations of $P_n$, we can construct $Q$ iid unbiased estimators for $\beta_{\mathbf{r}}$ via

$$\hat{\beta}_{\mathbf{r},QM,q} = \frac{1}{n}\sum_{k=1}^{n} f(\mathbf{v}_{k,q})\phi(\mathbf{v}_{k,q}), \tag{3.22}$$

where $\mathbf{v}_{k,q}$ denotes the $k^{th}$ sample point of the $q^{th}$ randomization of $P_n$.

Note that simply taking the square of (3.22), that is, $(\hat{\beta}_{\mathbf{r},QM,q})^2$, is a biased estimator of $\beta_{\mathbf{r}}^2$. Instead, we obtain an unbiased estimator of $\beta_{\mathbf{r}}^2$ as follows:

$$\hat{\beta}_{\mathbf{r},QM}^2 = \frac{Q}{Q-1}(\bar{\beta}_{\mathbf{r},QM})^2 - \frac{1}{Q(Q-1)}\sum_{q=1}^{Q}(\hat{\beta}_{\mathbf{r},QM,q})^2, \tag{3.23}$$

where $\bar{\beta}_{\mathbf{r},QM} := \frac{1}{Q}\sum_{q=1}^{Q}\hat{\beta}_{\mathbf{r},QM,q}$.

Using $\hat{\beta}^2_{\mathbf{r},QM}$, we now define an unbiased estimator ${}_{QM}\hat{\sigma}^2_{I,d,m}$ for $\sigma^2_{I,d,m}$:

$$ {}_{QM}\hat{\sigma}^2_{I,d,m} = \frac{Q}{Q-1}\sum_{\mathbf{r}\in R(I,d,m)}(\bar{\beta}_{\mathbf{r},QM})^2 - \frac{1}{Q(Q-1)}\sum_{\mathbf{r}\in R(I,d,m)}\sum_{q=1}^{Q}(\hat{\beta}_{\mathbf{r},QM,q})^2. \tag{3.24}$$

An unbiased estimator of $\sigma^2$ can be obtained by

$$\hat{\sigma}^2_{QM} = \frac{1}{nQ}\sum_{q=1}^{Q}\sum_{k=1}^{n} f^2(\mathbf{v}_{k,q}) - \frac{2}{Q(Q-1)}\sum_{1\le q_1\le q_2\le Q}\mu_{QM,q_1}\mu_{QM,q_2}, \tag{3.25}$$

and $\mu_{QM,q} = \sum_{k=1}^{n} f(\mathbf{v}_{k,q})/n$, $q = 1,\ldots Q$.

Let $\psi_{I,d,m} := \sigma^2_{I,d,m}/\sigma^2$. We can interpret $\psi_{I,d,m}$ as the fraction of the variance of $f$ explained by the corresponding $\tilde{f}$. Let $\hat{\psi}_{I,d,m} :=_{QM} \hat{\sigma}^2_{I,d,m}/\hat{\sigma}^2_{QM}$. Then, $\hat{\psi}_{I,d,m}$ is a consistent estimator of $\psi_{I,d,m}$. We then obtain an approximate confidence interval for the estimate of $\psi$ by constructing R iid copies of estimates and taking their sample standard deviation.

### 3.2.4 Discussion

In this section, we introduce several notations which we use frequently in subsequent chapters. The linear or bilinear regression related notations may have the subscript $L$, $B$, or $LB$. These subscripts indicate the type of underlying approximation of $f$. The subscript $L$ means the underlying quasi-regression is linear, $B$ means bilinear without linear components, and $LB$ means bilinear with linear components. Thus, the variance of $f$ explained by $\tilde{f}_{LB}$ defined in (3.21) is denoted by $\sigma^2_{LB}$ and the fraction of such variance over $\sigma^2$ is denoted by $\psi_{LB}$, that is, $\psi_{LB} := \sigma^2_{LB}/\sigma^2$. The notations $\sigma^2_L$, $\sigma^2_B$, $\psi_L$ and $\psi_B$ are defined similarly.

As discussed in the previous section, the copula-based criterion $\gamma_M$ introduces an additional parameter $M$. We denote the variance of $f$ explained by $\tilde{f}_{LB,M}$ defined in 3.21 by $\sigma^2_{LB,M}$. Also, we denote the fraction of the variance of $f$ explained by $\tilde{f}_{LB,M}$ by $\psi_{LB,M}$. Other notations such as $\sigma^2_{B,M}$ and $\psi_{B,M}$ are defined similarly.

If $\psi_{LB}$ is close to one, we can say that the assumption of the copula-based criterion is valid. We can assess this assumption by computing $\hat{\psi}_{LB}$. We will investigate if this is the

case with problems that come from various fields in subsequent chapters. If we see that many applications do not satisfy the bilinear assumption, it motivates us to extend the copula-based criterion so that it includes higher order polynomials.

We can define a search criterion for the Korobov lattice by using quasi-regression. This criterion searches for the generator that minimizes (3.9), where we take $\tilde{f}_{LB}$ for $\tilde{f}$. We refer to this criterion as the bilinear regression based criterion or more shortly the BR criterion. If $f$ is well approximated by $\tilde{f}_{LB}$, we expect that the optimal generator with respect to the BR criterion is also a good generator for $f$. The precise expression for the BR criterion will be given in (4.14).

The BR criterion differs from other criteria such as the copula-based criterion and spectral test [11] in that it is adapted to a specific problem. The generators based on the BR criterion are designed to work well with specific problems. Consequently, the generators may work well for one problem, but may perform poorly for other problems. On the other hand, other criteria such as spectral test are designed to produce generators that work well in general settings. We can think of the BR criterion as the modification of the copula-based criterion specialized to a particular problem.

# Chapter 4

# Bilinear Function

The copula-based criterion (3.10) is specifically concerned with bilinear functions. It is natural to start our analysis by looking at bilinear functions. We would like to know if the Korobov lattice with generators based on the copula-based criterion actually integrate bilinear functions with a small error. We will assess this by comparing the Korobov lattice to Monte Carlo and the Korobov lattice with generators based on other search criteria. The comparisons are made by comparing theoretical variance of MC and Korobov lattice estimator of $I(f)$.

The rest of the chapter is organized as follows. In Section 4.1, we describe the three search criteria, spectral test, $\mathcal{P}_2$ and $\mathcal{R}_2$, which we have chosen to compare against the copula-based criterion. We also list generators based on those criteria. In Section 4.2, we derive the expression for the theoretical variance for MC and Korobov lattice estimators of $I(f)$, where $f$ is a bilinear function. Also, the precise definition of the bilinear regression-based is given in that section.

We also investigate the robustness of the copula-based criterion $\gamma_M$ to the different choices of $M$.

## 4.1 Selection of Search Criteria for Comparison

We first introduce some terminology. Henceforth, we refer to the generators obtained from optimizing with respect to the copula-based criterion as Copula generators. The point sets constructed based on the Copula generators are referred to as Copula Korobov or Copula Korobov point set. The estimator obtained from the Copula Korobov point set is referred

to as Copula estimator. Also the variance of the Copula estimator is referred to as Copula variance. We refer to generators, point sets, estimators and variances for other criteria in a similar way.

We would like to examine the performance of the copula-based criterion in subsequent sections. The approach we take is to compare Copula generators to generators based on other search criteria through numerical experiments. The criteria we chose for comparison are spectral tests, $\mathcal{P}_2$ Test and $\mathcal{R}_2$ Test. We first briefly describe these criteria.

### 4.1.1 Spectral Test

We follow [11] for the description of the spectral test.

The Fourier expansion of $f$, where $f : [0,1)^t \to \mathbb{R}$, is

$$f(\mathbf{u}) = \sum_{\mathbf{h} \in \mathbf{Z}^s} \hat{f}(\mathbf{h}) \exp(2\pi\sqrt{-1}\mathbf{h}\mathbf{u}),$$

with *Fourier coefficients*

$$\hat{f}(\mathbf{h}) = \int_{[0,1)^t} f(\mathbf{u}) \exp(-2\pi\sqrt{-1}\mathbf{h}\mathbf{u}) d\mathbf{u}.$$

Sloan and Osborn [21] have shown that if $P_n$ is a lattice node set, i.e., $P_n = L_s \cap [0,1)^s$, the integration error is the sum of the Fourier coefficients over the nonzero vectors of the dual lattice $L_s^*$:

$$E_n = \sum_{\mathbf{0} \neq \mathbf{h} \in L_s^*} \hat{f}(\mathbf{h}). \tag{4.1}$$

The error expression (4.1) immediately suggests a quality criterion of the form

$$D'(P_n) = \sup_{\mathbf{0} \neq \mathbf{h} \in L_s^*} \omega(\mathbf{h}) \tag{4.2}$$

for lattice rule, where $\omega(\mathbf{h})$ are arbitrary nonnegative weights that decrease with the size of $\mathbf{h}$.

For spectral test, we take $\omega(\mathbf{h}) = ||\mathbf{h}||_2^{-1}$, where $||\mathbf{h}||_2 = \sqrt{h_1^2 + \cdots + h_t^2}$, the $\mathcal{L}_2$ norm of $\mathbf{h}$. We use $l_t$ to denote the length of the shortest vector in $L_s^*$. Since small vectors $\mathbf{h}$ are considered the most damaging, that is, lead to large Fourier coefficient $|\hat{f}(\mathbf{h})|$, maximizing

$l_t$ makes sense. Let $l_s = l_{\{1,\ldots,s\}}$ for $s \leq t$ and let $l^*_{|I|}(n)$ denote an upper bound on the length.

For arbitrary positive integers $t_1 \geq \cdots \geq t_d \geq d$, the following criterion is defined in [11]:

$$M_{\{t_1,\cdots,t_d\}} = \min\left[\min_{2\leq s\leq t_1} l_t/l^*_t(n), \min_{2\leq s\leq d}\min_{I\in S(s,t_s)} l_I/l^*_{|I|}(n)\right],$$

where $S(s,t_s) = \{I = \{i_1,\ldots i_s\} : 1 = i_1 < \cdots < i_s \leq t_s\}$.

For this thesis, we take generators from [11] listed as the best generators with respect to $M_{32,24,12,8}$.

### 4.1.2 $\mathcal{P}_2$ Test

We follow [5] to present $\mathcal{P}_2$, which is a particular case of $\mathcal{P}_\alpha$. For a point set of size $n$ in dimension $s$ generated by $a$, the weighted $\mathcal{P}_\alpha$ is defined as

$$\mathcal{P}_{\alpha,n,s}(a) = \sum_{\mathbf{0}\neq\mathbf{h}:\mathbf{h}\cdot\mathbf{a}\equiv 0 \bmod n} \gamma_{I_\mathbf{h}}||\mathbf{h}||^{-\alpha}, \tag{4.3}$$

where $\mathbf{h}\cdot\mathbf{a} = h_1 + h_2 a + \ldots + h_s a^{s-1}, I_\mathbf{h} = \{j : h_j \neq 0, 1 \leq j \leq s\}, \{\gamma_I, \emptyset \neq I \subseteq 1,\ldots\}$ is a set of weights, $||\mathbf{h}|| = \prod_{i=1}^{s} \bar{h}_i,$ and $\bar{h} = \max(1,|h|)$. The criterion searches for the generators that minimize (4.3).

As in [5], we have that

$$\mathcal{P}_{\alpha,n,s}(a) = \sum_{I\subseteq\{1,\ldots,s\}} \gamma_I \mathcal{P}_{\alpha,n,s,I}(a), \tag{4.4}$$

where $\mathcal{P}_{\alpha,n,s,I}(a)$ is the value of the measure $\mathcal{P}_{\alpha,n,|I|}(a)$ for the projection of the point set over $I$ when all weights are set to 1.

We use the weights of the form $\gamma_I = \prod_{j\in I} r_j$, where $r_j = 0.1$ for $j = 1,\ldots s$. We run *Lattice Builder*[10](software), a software that implements several search criteria, to obtain the best generators.

### 4.1.3 $\mathcal{R}_2$ Test

We follow the notation used to describe $\mathcal{P}_2$ in the previous section.

For a Korobov point set of size $n$ in dimension $s$ generated by $a$, the weighted $\mathcal{R}_\alpha$ is defined as the weighted sum of the unweighted components

$$\mathcal{R}_{\alpha,n,s}(a) = \sum_{I \subseteq \{1,\ldots,s\}} \gamma_I \mathcal{R}_{\alpha,n,s,I}(a). \tag{4.5}$$

As in [9] we have,

$$\mathcal{R}_{\alpha,n,s,I}(a) = \frac{1}{n} \sum_{i=0}^{n-1} \prod_{j \in u} r_{\alpha,n}(i(a^{j-1} \bmod n)/n \bmod 1),$$

where

$$r_{\alpha,n} = \sum_{h=-\lfloor (n-1)/2 \rfloor}^{\lfloor n/2 \rfloor} |\max(1,h)|^{-\alpha} e^{2\pi i h x} - 1.$$

For given weights, the best generator minimizes (4.5). We again use the weights of the form $\gamma_I = \prod_{j \in I} r_j$, where $r_j = 0.1$ for $j = 1, \ldots s$. We run *Lattice Builder*[10](software) to obtain the best generators with respect to $\mathcal{R}_2$.

### 4.1.4 Comparison of the generators

Having introduced the search criteria against which we compare the copula-based criterion, we present the generators we have chosen. Table 4.1 lists the generators which we use for numerical experiments.

Table 4.2 lists the value $\gamma_8$ for each generator. This table shows that the variation of $\gamma_8$ across the four generators becomes smaller as $n$ gets larger. The first three digits are identical for all generators when $n = 131071$ while the second digit are all different when $n = 1021$. This does not mean $\mathcal{P}_2$, $\mathcal{R}_2$ and Spectral generators happen to be good generators with respect to Copula Criteria, as shown in Table 4.3.

Table 4.3 lists the ranking for each $\gamma_8$. The rank $k$ here would mean that the generator produced the $k^{th}$ smallest value of $\gamma_8$ among all generators. This table presents $\gamma_8$ in relative sense. We see that generators from $\mathcal{P}_2$ and $\mathcal{R}_2$ are similar with respect to the

Table 4.1: Korobov Generators

| $n$ | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|
| 1021 | 396 | 76 | 401 | 401 |
| 2039 | 778 | 1487 | 496 | 713 |
| 4093 | 1902 | 1516 | 1428 | 450 |
| 8191 | 2950 | 5130 | 1527 | 1527 |
| 16381 | 13716 | 4026 | 7097 | 5715 |
| 32749 | 6934 | 14251 | 14844 | 14844 |
| 65521 | 23529 | 8950 | 11070 | 30902 |
| 131071 | 15674 | 28823 | 42645 | 45599 |

Table 4.2: Maximum Spearman rho for $M = 8$

| $n$ | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|
| 1021 | -3.439295E-04 | -3.170025E-04 | -3.308221E-04 | -3.308221E-04 |
| 2039 | -1.728790E-04 | -1.724252E-04 | -1.725575E-04 | -1.725575E-04 |
| 4093 | -8.625873E-05 | -8.287868E-05 | -8.539463E-05 | -8.539463E-05 |
| 8191 | -4.311918E-05 | -4.301152E-05 | -4.291234E-05 | -4.291234E-05 |
| 16381 | -2.157169E-05 | -2.152508E-05 | -2.152955E-05 | -2.152955E-05 |
| 32749 | -1.079241E-05 | -1.077341E-05 | -1.024041E-05 | -1.024041E-05 |
| 65521 | -5.395000E-06 | -5.372306E-06 | -5.393171E-06 | -5.393171E-06 |
| 131071 | -2.697086E-06 | -2.690664E-06 | -2.695285E-06 | -2.695285E-06 |

copula-based criterion, that is, they are close in ranking for all $n$. The rankings reveal that generators from $\mathcal{P}_2$, $\mathcal{R}_2$ and Spectral test rarely make it to the top 10% in ranking.

If the problem at hand can be well approximated by a function in a class of functions for which the copula-based criterion has been defined, the Copula generators should perform better than generators from the other three criteria. Conversely, if Copula generators and Spectral generators, for instance, show close performance, it suggests that the class of functions does not represent the problem well.

Note that the quality measure $\gamma_8$ aggregates seven Spearman rhos, that is, $\rho_{1,m}$ for $m = 2, \ldots 8$. All those are required in order to compute the variance of the corresponding estimator theoretically in the next section. Table 4.4 lists $\rho_{1,m}$ for generators based on the four criteria when $n = 1021$ and $n = 16381$.

Table 4.3: Ranking of maximum Spearman rho for $M = 8$

| $n$ | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|
| 1021 | 1 | 384 | 296 | 296 |
| 2039 | 1 | 45 | 25 | 29 |
| 4093 | 1 | 2156 | 1066 | 1056 |
| 8191 | 1 | 719 | 1488 | 1488 |
| 16381 | 1 | 2548 | 2304 | 2300 |
| 32749 | 1 | 6134 | 26258 | 26258 |
| 65521 | 1 | 30581 | 2855 | 2852 |
| 131071 | 1 | 63542 | 28548 | 28557 |

Table 4.4: $\rho_{1,m}$ for $n = 1021$ and $n = 16381$

| $n = 1021$ | $\rho_{1,2}$ | $\rho_{1,3}$ | $\rho_{1,4}$ | $\rho_{1,5}$ | $\rho_{1,6}$ | $\rho_{1,7}$ | $\rho_{1,8}$ |
|---|---|---|---|---|---|---|---|
| Copula | -3.453252E-04 | -3.448801E-04 | -3.442355E-04 | -3.440783E-04 | -3.453538E-04 | -3.443018E-04 | -3.439295E-04 |
| Spectra Test | -3.439036E-04 | -3.425688E-04 | -3.412977E-04 | -3.170025E-04 | -3.409652E-04 | -3.422423E-04 | -3.427998E-04 |
| $\mathcal{P}_2$ | -3.441316E-04 | -3.308221E-04 | -3.412977E-04 | -3.450824E-04 | -3.350692E-04 | -3.422423E-04 | -3.451316E-04 |
| $\mathcal{R}_2$ | -3.441316E-04 | -3.308221E-04 | -3.412977E-04 | -3.450824E-04 | -3.350692E-04 | -3.422423E-04 | -3.451316E-04 |
| $n = 16381$ | $\rho_{1,2}$ | $\rho_{1,3}$ | $\rho_{1,4}$ | $\rho_{1,5}$ | $\rho_{1,6}$ | $\rho_{1,7}$ | $\rho_{1,8}$ |
| Copula | -2.157207E-05 | -2.157636E-05 | -2.157548E-05 | -2.157542E-05 | -2.157169E-05 | -2.157408E-05 | -2.157233E-05 |
| Spectra Test | -2.156863E-05 | -2.157127E-05 | -2.153361E-05 | -2.152508E-05 | -2.155613E-05 | -2.154522E-05 | -2.157349E-05 |
| $\mathcal{P}_2$ | -2.157105E-05 | -2.156966E-05 | -2.157380E-05 | -2.152955E-05 | -2.156606E-05 | -2.156902E-05 | -2.157628E-05 |
| $\mathcal{R}_2$ | -2.157105E-05 | -2.156966E-05 | -2.157380E-05 | -2.152955E-05 | -2.156606E-05 | -2.156902E-05 | -2.157628E-05 |

## 4.2 Comparison of Monte Carlo and Korobov Lattice for Bilinear Function

For the rest of the chapter, assume that $f$ is a sum of linear and bilinear functions written as

$$f(\mathbf{u}) = \sum_{i=1}^{s} c_i u_i + \sum_{i<j}^{s} d_{i,j} u_i u_j, \tag{4.6}$$

for $c_i, d_{ij} \in \mathbb{R}$.

In this section we first derive formulas of $\mathrm{Var}(\hat{\mu}_{mc,n})$ and $\mathrm{Var}(\hat{\mu}_{latt,n})$ for a bilinear function, where $\hat{\mu}_{latt,n}$ denotes the estimator based on a rank-1 lattice. Then we make certain assumptions on coefficients and compare $\mathrm{Var}(\hat{\mu}_{latt,n})$ from one generator to another.

### 4.2.1 A close look at the variance expression

**Proposition 4.2.1** *Suppose $f$ is of the form* (4.6)*. Let $\hat{\mu}_{mc,n}$ and $\hat{\mu}_{latt,n}$ be MC and rank-1 lattice estimators, respectively, of $I(f)$. Then,*

$$\mathrm{Var}(\hat{\mu}_{mc,n}) = \frac{1}{12n}\sum_{i=1}^{s} {c_i}^2 + \frac{1}{48n}\sum_{k=1}^{s}\sum_{\substack{k\in G,H\\ G\neq H}} d_G d_H + \frac{1}{12n}\sum_{i=1}^{s} c_i \sum_{i\in G} d_G + \frac{7}{144n}\sum_{i<j}^{s} {d_{i,j}}^2 \tag{4.7}$$

*and*

$$\mathrm{Var}(\hat{\mu}_{latt,n}) = \frac{1}{12n^2}\sum_{i=1}^{s} {c_i}^2 + \frac{1}{48n^2}\sum_{k=1}^{s}\sum_{\substack{k\in G,H\\ G\neq H}} d_G d_H + \frac{1}{12n^2}\sum_{i=1}^{s} c_i \sum_{i\in G} d_G + \sum_{i<j}^{s}\left(\frac{7}{144n} + \frac{11}{80}\rho_{lr,\{i,j\}}\right){d_{i,j}}^2, \tag{4.8}$$

*where $\rho_{lr,\{i,j\}}$ is defined as in* (3.1.13)*.*

The proof of Proposition 4.2.1 is found in the Appendix.

**Remark 4.2.2** If $\rho_{lr,\{k,l\}} = -\frac{35}{99n}$ for all $\{k,l\}$, (4.8) collapses to

$$\mathrm{Var}(\hat{\mu}_{latt,n}) = \frac{1}{n}\mathrm{Var}(\hat{\mu}_{mc,n}). \tag{4.9}$$

Intuitively, if $\rho_{lr,\{k,l\}}$ are all close to $-\frac{35}{99n}$, the Korobov variance is much smaller than the MC variance.

As in [13], we note that $\rho_{lr,\{k,l\}} = \rho_{lr,\{k+c,l+c\}}$ since a Korobov point set is dimension-stationary. Hence in that case we can further simplify (4.8) to

$$\mathrm{Var}(\hat{\mu}_{latt,n}) = \frac{1}{12n^2}\sum_{i=1}^{s} {c_i}^2 + \frac{1}{48n^2}\sum_{k=1}^{s}\sum_{\substack{k\in G,H\\ G\neq H}} d_G d_H + \frac{1}{12n^2}\sum_{i=1}^{s} c_i \sum_{i\in G} d_G$$
$$+\frac{7}{144n}\sum_{i<j}^{s} {d_{i,j}}^2 + \frac{11}{80}\sum_{m=2}^{s}\rho_{lr,\{1,m\}}\sum_{k=1}^{s-m+1} {d_{k,k+m-1}}^2. \tag{4.10}$$

For (4.10), only the $\frac{11}{80}\sum_{m=2}^{s}\rho_{lr,\{1,m\}}\sum_{k=1}^{s-m+1} {d_{k,k+m-1}}^2$ term depends on the choice of the generator through $\rho_{lr,\{1,m\}}$. All the other terms are constant with respect to the choice of the generator.

The copula criterion $\gamma_M$ effectively assumes that $d_{i,j} = 0\ \forall\ (i,j)$ such that $M \leq j-i$ for a given $M$. Let $\mathcal{I} := \{\{i,j\} : j-i+1 \leq M, j \leq s\}$. With this assumption, (4.10) becomes

$$\mathrm{Var}(\hat{\mu}_{latt,n}) = \frac{1}{12n^2}\sum_{i=1}^{s} {c_i}^2 + \frac{1}{48n^2}\sum_{k=1}^{s}\sum_{\substack{k\in G,H\in\mathcal{I}\\ G\neq H}} d_G d_H + \frac{1}{12n^2}\sum_{i=1}^{s} c_i \sum_{i\in G\in\mathcal{I}} d_G$$
$$+\frac{7}{144n}\sum_{i<j,(i,j)\in\mathcal{I}}^{s} {d_{i,j}}^2 + \frac{11}{80}\sum_{m=2}^{M}\rho_{lr,\{1,m\}}\sum_{k=1}^{s-m} {d_{k,k+m}}^2. \tag{4.11}$$

### 4.2.2 The bilinear regression based (BR) criterion

With (4.10), (4.11), and (3.21), we can provide the precise definition of the bilinear regression based (BR) criterion discussed in Section 3.2.4. Let $\tilde{f}_{LB}$ be the bilinear regression of $f$ as defined in (3.21). Then $\tilde{f}_{LB}$ is of the form

$$\tilde{f}_{LB} = \hat{\mu} + \sum_{i=1}^{s}\hat{c}_i(2u_i - 1) + \sum_{i<j}^{s}\hat{d}_{i,j}(2u_i-1)(2u_i-1), \tag{4.12}$$

where:
$\hat{\mu}$ is the estimate of $\beta_{\mathbf{r}}$ with all the elements of $\mathbf{r}$ are zero,

$\hat{c}_i$ is the estimate of $\beta_{\mathbf{r}}$ with only the $i$th element of $\mathbf{r}$ is 1 and all the other elements are zero,
$\hat{d}_{i,j}$ is the estimate of $\beta_{\mathbf{r}}$ with only the $i$th an $j$th element of $\mathbf{r}$ are 1 and all the other elements are zero.

The BR criterion searches for the generator that minimizes $\mathrm{Var}(\hat{\mu}_{latt,n})$, assuming the underlying function is $\tilde{f}_{LB}$ which comes from quasi-regression. Note that $\tilde{f}_{LB}$ is a bilinear function but is not of the form (4.6), which we have considered to derive the variance expression. However, with simple algebras, it is trivial to see that

$$\mathrm{Var}(\hat{\mu}_{latt,n}) = \alpha + \frac{11}{20}\sum_{i<j}^{s} \rho_{lr,\{i,j\}}\hat{d}_{i,j}^2, \tag{4.13}$$

where $\alpha$ is a constant independent of the choice of the generator. So, when searching for the optimal generator, minimizing $\mathrm{Var}(\hat{\mu}_{latt,n})$ is equivalent to minimizing $\sum_{i<j}^{s} \rho_{lr,\{i,j\}}\hat{d}_{i,j}^2$. Thus, using the property that Korobov point set is dimension-stationary, the BR criterion seeks for the generator $a$ that minimizes the quantity

$$\sum_{m=2}^{s} \rho_{lr,\{1,m\}} \sum_{k=1}^{s-m} \hat{d}_{k,k+m}^2. \tag{4.14}$$

We will obtain the BR generators for problems such as Asian option pricing in Chapter 5.

### 4.2.3 Comparison of MC and Korobov Variance

We are often interested in how good QMC is compared to MC. We would like to know if a Korobov lattice outperforms MC for any generators, or there exist some bad generators for which a Korobov lattice does worse than MC.

First, consider the linear case. Proposition 3.7 in [13] says for $f(u) = u$, we have that

$$\mathrm{Var}(\hat{\mu}_{latt,n}) = \frac{1}{n}\mathrm{Var}(\hat{\mu}_{mc,n})$$

This also holds for $s$-dimensional linear functions with arbitrary coefficients. To see this, assume $d_{i,j} = 0\ \forall(i,j)$ in (4.6), which means

$$f(\mathbf{u}) = \sum_{i=1}^{s} c_i u_i.$$

Then we have $\text{Var}(\hat{\mu}_{mc,n}) = \frac{1}{12n}\sum_{i=1}^{s} {c_i}^2$ and $\text{Var}(\hat{\mu}_{latt,n}) = \frac{1}{12n^2}\sum_{i=1}^{s} {c_i}^2$. So, $\text{Var}(\hat{\mu}_{latt,n}) = \frac{1}{n}\text{Var}(\hat{\mu}_{mc,n})$.

Hence the use of a Korobov lattice over MC gives more accurate estimators than MC for $s$-dimensional linear functions.

Note that there is no Spearman rho involved in $\text{Var}(\hat{\mu}_{mc,n})$ for an s-dimensional linear function. It means that $\text{Var}(\hat{\mu}_{latt,n})$ is independent of the choice of generator $a$. This holds more generally with an s-dimensional function which is a sum of univariate functions.

Now consider the bilinear case. As bilinear functions are more complex than linear ones, we expect $\text{Var}(\hat{\mu}_{latt,n}) > \frac{1}{n}\text{Var}(\hat{\mu}_{mc,n})$. In Remark (4.2.2), we saw that if $\rho_{lr,\{k,l\}} = -\frac{35}{99n}$, we have $\text{Var}(\hat{\mu}_{latt,n}) = \frac{1}{n}\text{Var}(\hat{\mu}_{mc,n})$, which is identical to the linear case. This suggests that $\rho_{lr,\{k,l\}} > -\frac{35}{99n}$. Empirically, we found that $\rho_{lr,\{k,l\}} > -\frac{35}{99n}$ for $n = 1021$, 2039, 4093, 8191, 16381, 32749, 65521, and 131071.

The inequality $\text{Var}(\hat{\mu}_{latt,n}) > \frac{1}{n}\text{Var}(\hat{\mu}_{mc,n})$ is a lower bound of Korobov variances, so it does not guarantee that Korobov lattices always outperform MC.

Pick $a = 1$, which is considered the worst generator. If Korobov lattice outperforms MC even when $a = 1$, it is safe to say Korobov lattice is always better than MC for bilinear functions.

But, $\rho_{lr,\{k,l\}} = \frac{n^3-34n^2+6n+6}{99n^3} \to \frac{1}{99}$ asymptotically, for any $k$ and $l$ when $a = 1$. Using (4.11), we see that $\text{Var}(\hat{\mu}_{latt,n}) \to p$, for some $p$ independent of $n$. Since $\text{Var}(\hat{\mu}_{mc,n}) \to 0$, there exists $n$ such that $\text{Var}(\hat{\mu}_{mc,n}) < \text{Var}(\hat{\mu}_{latt,n})$ unless the $b_{i,j}$ are all zero.

So, depending on the choice of generator, Korobov lattice could do worse than MC. A sufficient condition to have $\text{Var}(\hat{\mu}_{latt,n}) < \text{Var}(\hat{\mu}_{mc,n})$ is $\rho_{lr,\{1,m\}} < 0$ for all $2 \le m \le M$. The optimal generators with respect to the four criteria satisfy this condition as shown in Table 4.4. As Spectral generators are often not close to optimal with respect to the copula-based criterion yet they satisfy $\rho_{lr,\{1,m\}} < 0$ for all $m \le 8$, it seems that the condition of having a negative Spearman rho is not too restrictive.

As Table 4.4 shows, when $n = 1021$, $\rho_{lr,\{1,m\}}$ is at maximum 10% greater than the lower bound $-\frac{35}{99n}$ which is $-3.462638 \cdot 10^{-4}$. We thus expect that the corresponding Korobov estimator performs better than MC estimator by an order of magnitude.

## 4.3 Numerical experiments for bilinear functions

We make certain assumptions on the coefficients of $f$ and compute variances to see if Copula variances are smaller than the variances from other Korobov lattices. This comparison will determine if the copula-based criterion is valid. Also, we vary assumed and actual $M$ in the definition of the copula-based criterion (3.10) to assess the robustness of the copula-based criterion. We have done numerical experiments for 30-dimensional bilinear functions. Thus, we assume $s = 30$ for the rest of this section.

### 4.3.1 Constant Coefficient

Here, we assume that $c_{\{i\}} = c$ for all $i = 1, \ldots, s$ and $d_{\{i,j\}} = \alpha c$ for some $\alpha \in \mathbb{R}$ if $\{i, j\} \in \mathcal{I} = \{\{i, j\} : j - i + 1 \leq M, j \leq s\}$ and 0 otherwise. The implication is that all univariate terms are equally important and all bivariate terms are also equally important. We start with all coefficients being equal to 1, that is, $c = 1$ and $\alpha = 1$. Table 4.5 lists variances under this assumption. The L_Bound column represents the lower bound of Korobov variance, which is $\frac{1}{n}$ of MC variance. In Table 4.6, each variance is divided by the entry in L_Bound that has the same value of $n$.

We see that the Copula Korobov gives the smallest variance for all $n$ as we expected. The Copula variance is often half the size of variance from Spectral, $\mathcal{P}_2$, and $\mathcal{R}_2$ Korobov lattices. This means that the copula-based criterion is valid. The Copula variances are often 1.5 times as large as lower bounds which is $\frac{1}{n}$ of MC variance. So, there is a great advantage to use Korobov lattice over MC since Korobov variance with a good generator produces a much smaller variance.

Table 4.5: Variances for $M = 8$, $c = 1$, and $\alpha = 1$

| $n$ | MC | L_Bound | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|
| 1021 | 1.27E-01 | 1.24E-04 | 1.57E-04 | 3.05E-04 | 2.62E-04 | 2.62E-04 |
| 2039 | 6.36E-02 | 3.12E-05 | 3.88E-05 | 4.61E-05 | 4.28E-05 | 4.28E-05 |
| 4093 | 3.17E-02 | 7.75E-06 | 9.44E-06 | 3.94E-05 | 1.37E-05 | 1.37E-05 |
| 8191 | 1.58E-02 | 1.93E-06 | 2.52E-06 | 3.18E-06 | 4.31E-06 | 4.31E-06 |
| 16381 | 7.92E-03 | 4.84E-07 | 6.53E-07 | 1.17E-06 | 8.76E-07 | 8.76E-07 |
| 32749 | 3.96E-03 | 1.21E-07 | 1.65E-07 | 3.78E-07 | 2.05E-06 | 2.05E-06 |
| 65521 | 1.98E-03 | 3.02E-08 | 4.38E-08 | 1.37E-07 | 6.20E-08 | 6.20E-08 |
| 131071 | 9.90E-04 | 7.55E-09 | 1.12E-08 | 7.46E-08 | 1.97E-08 | 1.97E-08 |

Table 4.6: Relative Variances for M=8, c=1, and $\beta = 1$

| $n$ | MC | L_Bound | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|
| 1021 | 1021 | 1 | 1.26 | 2.45 | 2.10 | 2.10 |
| 2039 | 2039 | 1 | 1.24 | 1.48 | 1.37 | 1.37 |
| 4093 | 4093 | 1 | 1.22 | 5.09 | 1.77 | 1.77 |
| 8191 | 8191 | 1 | 1.30 | 1.65 | 2.23 | 2.23 |
| 16381 | 16381 | 1 | 1.35 | 2.41 | 1.81 | 1.81 |
| 32749 | 32749 | 1 | 1.36 | 3.12 | 16.92 | 16.92 |
| 65521 | 65521 | 1 | 1.45 | 4.54 | 2.05 | 2.05 |
| 131071 | 131071 | 1 | 1.48 | 9.87 | 2.61 | 2.61 |

Table 4.7: Variances for $M = 8$, $c = 1$, $n = 1021$.

| $\alpha$ | MC | L_Bound | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|
| 0 | 2.45E-03 | 2.40E-06 | 2.40E-06 | 2.40E-06 | 2.40E-06 | 2.40E-06 |
| 0.01 | 2.76E-03 | 2.70E-06 | 2.70E-06 | 2.72E-06 | 2.71E-06 | 2.71E-06 |
| 0.1 | 6.37E-03 | 6.24E-06 | 6.57E-06 | 8.04E-06 | 7.61E-06 | 7.61E-06 |
| 0.5 | 4.10E-02 | 4.02E-05 | 4.84E-05 | 8.54E-05 | 7.46E-05 | 7.46E-05 |
| 0.75 | 7.81E-02 | 7.65E-05 | 9.50E-05 | 1.78E-04 | 1.54E-04 | 1.54E-04 |
| 1 | 1.27E-01 | 1.24E-04 | 1.57E-04 | 3.05E-04 | 2.62E-04 | 2.62E-04 |
| 2 | 4.42E-01 | 4.33E-04 | 5.64E-04 | 1.16E-03 | 9.82E-04 | 9.82E-04 |
| 10 | 9.79E+00 | 9.59E-03 | 1.29E-02 | 2.77E-02 | 2.33E-02 | 2.33E-02 |
| 100 | 9.52E+02 | 9.33E-01 | 1.26E+00 | 2.74E+00 | 2.31E+00 | 2.31E+00 |

We vary $\alpha$ to examine the relationship between the value of $\alpha$ and the relative performance of the four Korobov lattices. Table 4.7 shows the results. This time we only look at the case $n = 1021$, as the relative performance does not vary much for different choices of $n$, except for a few cases.
When $\alpha = 0$, the function is linear, so all four Korobov lattices give the same variance. As $\alpha$ increases, the relative performance of Copula Korobov improves, until $\alpha$ hits 0.75. After that, the Copula variance remains about half the size of other Korobov variances, no matter how large $\alpha$ gets.

We now would like to see if the copula-based criterion is robust to a mismatch of $M$. So far, we have used Copula generators obtained by assuming $M = 8$. We denote the assumed $M$ in the criterion and actual $M$ of a bilinear function as $M_{assumed}$ and $M_{actual}$, respectively. The ideal case is when $M_{assumed} = M_{actual}$. Since we usually do not know

$M_{actual}$, we should examine if Copula Korobov outperforms other Korobov methods when $M_{assumed} \neq M_{actual}$. Table 4.8 shows the Copula variances when $M_{assumed}$ and $M_{actual}$ vary between 4 to 24. The last row contains Spectral variances for comparison.

Table 4.8: Variances for various $M_{assumed}$ and $M_{actual}$ with $c = 1$, $\alpha = 0.5$, $n = 1021$

| $M_{assumed} \backslash M_{actual}$ | 4 | 8 | 12 | 16 | 20 | 24 | 28 |
|---|---|---|---|---|---|---|---|
| 4 | 1.644E-05 | 6.175E-04 | 6.846E-04 | 7.662E-04 | 8.301E-04 | 1.238E-03 | 1.265E-03 |
| 8 | 1.724E-05 | 4.842E-05 | 9.430E-05 | 2.008E-04 | 2.545E-04 | 2.771E-04 | 3.566E-04 |
| 12 | 2.000E-05 | 5.119E-05 | 8.649E-05 | 1.269E-04 | 1.569E-04 | 2.903E-04 | 3.166E-04 |
| 16 | 2.157E-05 | 5.599E-05 | 9.756E-05 | 1.342E-04 | 1.632E-04 | 1.905E-04 | 2.133E-04 |
| 20 | 2.157E-05 | 5.599E-05 | 9.756E-05 | 1.342E-04 | 1.632E-04 | 1.905E-04 | 2.133E-04 |
| 24 | 1.765E-05 | 5.138E-05 | 9.138E-05 | 1.291E-04 | 1.671E-04 | 1.900E-04 | 4.340E-04 |
| 28 | 2.361E-05 | 8.535E-05 | 1.210E-04 | 1.696E-04 | 1.994E-04 | 2.220E-04 | 2.352E-04 |
| Spectral | 2.324901E-005 | 8.409192E-005 | 1.183574E-004 | 1.651131E-004 | 1.937807E-004 | 3.130848E-004 | 3.216814E-004 |

As expected, variances for estimators based on the Copula Korobov are smaller than those based on Spectral counterpart under the ideal case. This table reveals that most of the time, mismatches between $M_{assumed}$ and $M_{actual}$ result in a greater variance. In some cases, the variance is ten times as large if the optimal generator is not chosen because of this mismatch. Spectral Korobov often does better than copula Korobov if $M_{assumed} \neq M_{actual}$

We also see that the Copula criterion is more sensitive to underestimation than overestimation of $M_{actual}$. When we overestimate $M_{actual}$, Copula Korobov does worse than in the ideal case but still does better than Spectral Korobov. On the other hand when we underestimate $M_{actual}$, Copula Korobov often does worse than Spectral Korobov. This makes sense because when $M_{assumed} > M_{actual}$ , the copula-based criterion makes sure $\rho_{lr,\{1,m\}}, m = 2, \dots M_{actual}$ are all small. On the other hand when $M_{assumed} < M_{actual}$, the criterion ensures only $\rho_{lr,\{1,m\}}, m = 2, \dots M_{assumed}$ are small, but $\rho_{lr,\{1,k\}}$, for $k = M_{assumed} + 1, \dots M_{actual}$ could possibly be very large. We have to be careful about the selection of $M$ in the criterion.

### 4.3.2 Geometric Coefficients

In this section we assume $f$ has coefficients that vary according to a geometric progression. In particular, we are interested in the case where coefficients geometrically decay. The implication is that univariate terms with low indices are more important than those with high indices. Also, bivariate terms with small sum of indices are more important than those with large sum of indices.

We would like to examine whether the superior performance of Copula Korobov shown in the previous sections still holds under geometric constant assumptions.

Given $c, \alpha, r \in \mathbb{R}$ and $|r| < 1$, assume $c_{\{i\}} = cr^{i-1} \; \forall \; i = 1, \ldots, s$ and $d_{\{i,j\}} = \alpha c r^{i+j-3}$ if $\{i, j\} \in \mathcal{I} = \{\{i, j\} : j - i + 1 \leq M, j \leq s\}$ and 0 otherwise.

Table 4.9 lists variances assuming $c = 1$, $\alpha = 1$, and $r = 0.95$. Table 4.10 lists relative variances.

Table 4.9: Variances for $M = 8$, $c = 1$, $\alpha = 1$ and $r = 0.95$

| $n$ | MC | L_Bound | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|
| 1021 | 3.44E-02 | 3.36E-05 | 4.24E-05 | 8.25E-05 | 7.27E-05 | 7.27E-05 |
| 2039 | 1.72E-02 | 8.44E-06 | 1.06E-05 | 1.26E-05 | 1.17E-05 | 1.17E-05 |
| 4093 | 8.57E-03 | 2.09E-06 | 2.56E-06 | 1.13E-05 | 3.62E-06 | 3.62E-06 |
| 8191 | 4.28E-03 | 5.23E-07 | 6.90E-07 | 8.95E-07 | 1.19E-06 | 1.19E-06 |
| 16381 | 2.14E-03 | 1.31E-07 | 1.77E-07 | 3.16E-07 | 2.39E-07 | 2.39E-07 |
| 32749 | 1.07E-03 | 3.27E-08 | 4.48E-08 | 1.08E-07 | 4.65E-07 | 4.65E-07 |
| 65521 | 5.35E-04 | 8.17E-09 | 1.19E-08 | 4.07E-08 | 1.67E-08 | 1.67E-08 |
| 131071 | 2.68E-04 | 2.04E-09 | 3.04E-09 | 2.24E-08 | 5.20E-09 | 5.20E-09 |

Table 4.10: Relative Variances for $M = 8$, $c = 1$, $\alpha = 1$ and $r = 0.95$

| $n$ | MC | L_Bound | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|
| 1021 | 1021 | 1 | 1.26 | 1.95 | 2.16 | 2.16 |
| 2039 | 2039 | 1 | 1.25 | 1.19 | 1.39 | 1.39 |
| 4093 | 4093 | 1 | 1.22 | 4.40 | 1.73 | 1.73 |
| 8191 | 8191 | 1 | 1.32 | 1.30 | 2.27 | 2.27 |
| 16381 | 16381 | 1 | 1.35 | 1.79 | 1.82 | 1.82 |
| 32749 | 32749 | 1 | 1.37 | 2.41 | 14.22 | 14.22 |
| 65521 | 65521 | 1 | 1.46 | 3.41 | 2.04 | 2.04 |
| 131071 | 131071 | 1 | 1.49 | 7.37 | 2.55 | 2.55 |

We see that Copula Korobov still gives the smallest theoretical variance for all $n$ as we expected. The relative performance of Copula Korobov compared to Spectral, $\mathcal{P}_2$ and $\mathcal{R}_2$ Korobov does not differ much from the constant coefficient case.

We vary $\alpha$ to examine the relationship between the value of $\alpha$ and the relative performance of the four Korobov lattices. Table 4.11 shows the results.

The same analysis as for the constant coefficients case applies here. The relative performance of Copula Korobov improves at diminishing rate as $\alpha$ increases.

Table 4.11: Variances for $M = 8$, $c = 1$, $r = 0.95$, $n = 1021$.

| $\alpha$ | MC | L_Bound | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|
| 0 | 7.99E-04 | 7.82E-07 | 7.82E-07 | 7.82E-07 | 7.82E-07 | 7.82E-07 |
| 0.01 | 8.87E-04 | 8.69E-07 | 8.70E-07 | 8.74E-07 | 8.73E-07 | 8.73E-07 |
| 0.1 | 1.91E-03 | 1.87E-06 | 1.96E-06 | 2.36E-06 | 2.26E-06 | 2.26E-06 |
| 0.5 | 1.13E-02 | 1.11E-05 | 1.33E-05 | 2.33E-05 | 2.09E-05 | 2.09E-05 |
| 0.75 | 2.13E-02 | 2.08E-05 | 2.58E-05 | 4.84E-05 | 4.28E-05 | 4.28E-05 |
| 1 | 3.44E-02 | 3.36E-05 | 4.24E-05 | 8.25E-05 | 7.27E-05 | 7.27E-05 |
| 2 | 1.18E-01 | 1.15E-04 | 1.51E-04 | 3.11E-04 | 2.72E-04 | 2.72E-04 |

We vary $r$ to examine the relationship between the value of $r$ and the relative performance of the four Korobov lattices. Table 4.12 shows the results.

Table 4.12: Variances for M=8, c=1, $\alpha = 0.5$, n=1021.

| $r$ | MC | L_Bound | Copula | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|
| 0.1 | 1.47E-04 | 1.44E-07 | 1.65E-07 | 2.15E-07 | 2.11E-07 | 2.11E-07 |
| 0.25 | 1.79E-04 | 1.75E-07 | 2.00E-07 | 2.58E-07 | 2.76E-07 | 2.76E-07 |
| 0.5 | 3.16E-04 | 3.09E-07 | 3.55E-07 | 4.76E-07 | 5.81E-07 | 5.81E-07 |
| 0.75 | 1.04E-03 | 1.02E-06 | 1.20E-06 | 1.91E-06 | 2.06E-06 | 2.06E-06 |
| 0.9 | 4.69E-03 | 4.59E-06 | 5.47E-06 | 9.54E-06 | 8.86E-06 | 8.86E-06 |
| 0.95 | 1.13E-02 | 1.11E-05 | 1.33E-05 | 2.33E-05 | 2.09E-05 | 2.09E-05 |
| 0.975 | 2.03E-02 | 1.99E-05 | 2.39E-05 | 4.21E-05 | 3.71E-05 | 3.71E-05 |

The relative performance of Copula Korobov seems to improve at diminishing rate as $r$ increases, similar to an increase in $\alpha$. This makes sense because more of the variation of $f$ comes from the bilinear part as $r$ increases.

Now we investigate the robustness of the copula-based criterion under the geometric coefficients assumptions. We look at the case where $r = 0.5$ and $r = 0.85$ to see if the robustness is influenced by $r$.

Table 4.13 shows variances when $r = 0.85$. This table is quite different from the one for the constant coefficient case. For fixed $M_{assumed}$, the variances increase with $M_{actual}$ at diminishing rate. This makes sense because additional terms associated with increased $M_{actual}$ are small because of geometric terms.

We can connect this observation with the variance expression (4.11). The variance is

Table 4.13: Variances for various $M_{assumed}$ and $M_{actual}$. $c = 1$. $\alpha = 0.5$. $r = 0.85$. $n = 1021$.

| $M_{assumed} \backslash M_{actual}$ | 4 | 8 | 12 | 16 | 20 | 24 | 28 |
|---|---|---|---|---|---|---|---|
| 4 | 1.58E-06 | 3.25E-05 | 3.37E-05 | 3.41E-05 | 3.43E-05 | 3.47E-05 | 3.47E-05 |
| 8 | 1.62E-06 | 2.90E-06 | 3.62E-06 | 4.13E-06 | 4.31E-06 | 4.38E-06 | 4.42E-06 |
| 12 | 1.90E-06 | 3.19E-06 | 3.82E-06 | 4.15E-06 | 4.30E-06 | 4.42E-06 | 4.46E-06 |
| 16 | 2.10E-06 | 3.49E-06 | 4.20E-06 | 4.52E-06 | 4.67E-06 | 4.74E-06 | 4.77E-06 |
| 20 | 2.10E-06 | 3.49E-06 | 4.20E-06 | 4.52E-06 | 4.67E-06 | 4.74E-06 | 4.77E-06 |
| 24 | 1.70E-06 | 3.02E-06 | 3.71E-06 | 4.03E-06 | 4.19E-06 | 4.26E-06 | 4.36E-06 |
| 28 | 1.89E-06 | 3.18E-06 | 3.97E-06 | 4.30E-06 | 4.45E-06 | 4.52E-06 | 4.55E-06 |
| Spectral | 2.20E-06 | 4.95E-06 | 5.59E-06 | 5.96E-06 | 6.11E-06 | 6.18E-06 | 6.20E-06 |

small if $\sum_{m=2}^{M} \rho_{lr,\{1,m\}} \sum_{k=1}^{s-m} {d_{k,k+m}}^2$ is small. With geometric coefficient assumption, we can rewrite it as $\alpha c \sum_{m=2}^{M} \rho_{lr,\{1,m\}} r^{2m} \sum_{k=1}^{s-m} r^{4k}$, where $0 < r < 1$.
We can also write it as $\alpha c \sum_{m=2}^{M} t_m \rho_{lr,\{1,m\}}$, where $t_m = r^{2m} \sum_{k=1}^{s-m} r^{4k}$.

If $r$ is not close to 1, say less than 0.9, $t_m$ decays sufficiently fast that $t_m$ becomes negligible for large $m$. So, a large value of $\rho_{lr,\{1,m\}}$ will be offset by a small value of $t_m$ if $m$ is sufficiently large. Therefore, in order to have a small variance, we only need $\rho_{lr,\{1,m\}}$ to be small for $1 < m < M'$ for some $M' < M$. Hence, when selecting $M_{assumed}$ for the copula-based criterion, we should choose $M_{assumed}$ such that $M_{assumed} \leq M_{actual}$. This suggests that $M_{opt}$ defined in Section 3.2.4 is smaller than $M_{actual}$.

As expected, the generators based on the ideal case never give the smallest variances, unlike in the constant coefficient case. For the constant coefficient case, underestimation of $M_{actual}$ led to greater variances, but this time it generally leads to smaller variances. The Copula Korobov never does worse than Spectral Korobov except when $M_{assumed} = 4$. It seems that the Copula generator with $M_{assumed} = 4$ is good only when $M_{actual}$ is also equal to 4.

This time we have $r = 0.95$ instead of $r = 0.85$. Unlike the previous case, coefficients do not decay as fast, so an increase in $M_{actual}$ does increase the variance. It seems that the robustness analysis for the constant coefficient case applies here. The generators under the ideal case give smaller variances and underestimation of $M_{actual}$ generally results in greater variances. Overall, the robustness of the Copula criteria depends on the value of $r$.

Table 4.14: Variances for various $M_{assumed}$ and $M_{actual}$. $c = 1$, $\alpha = 0.5$, $r = 0.95$ and $n = 1021$.

| $M_{assumed} \backslash M_{actual}$ | 4 | 8 | 12 | 16 | 20 | 24 | 28 |
|---|---|---|---|---|---|---|---|
| 4 | 5.14E-06 | 1.67E-04 | 1.80E-04 | 1.93E-04 | 2.01E-04 | 2.40E-04 | 2.42E-04 |
| 8 | 5.37E-06 | 1.33E-05 | 2.24E-05 | 3.78E-05 | 4.47E-05 | 4.75E-05 | 5.31E-05 |
| 12 | 6.24E-06 | 1.42E-05 | 2.15E-05 | 2.83E-05 | 3.27E-05 | 4.47E-05 | 4.71E-05 |
| 16 | 6.78E-06 | 1.55E-05 | 2.40E-05 | 3.03E-05 | 3.46E-05 | 3.78E-05 | 3.99E-05 |
| 20 | 6.78E-06 | 1.55E-05 | 2.40E-05 | 3.03E-05 | 3.46E-05 | 3.78E-05 | 3.99E-05 |
| 24 | 5.52E-06 | 1.40E-05 | 2.22E-05 | 2.86E-05 | 3.39E-05 | 3.67E-05 | 5.53E-05 |
| 28 | 6.35E-06 | 1.45E-05 | 2.38E-05 | 3.05E-05 | 3.54E-05 | 3.83E-05 | 3.97E-05 |
| Spectral | 7.33E-06 | 2.33E-05 | 3.07E-05 | 3.87E-05 | 4.31E-05 | 4.59E-05 | 4.73E-05 |

### 4.3.3 Conclusion for this chapter

In this chapter we saw that the Copula Korobov performs much better than the Spectral, $\mathcal{P}_2$ and $\mathcal{R}_2$ Korobov, if we select the right value of $M$ for the Copula criterion $\gamma_M$. We also saw that the robustness of the Copula criterion to the mismatch of $M_{supposed}$ and $M_{actual}$ depends on how fast the coefficients decay with the geometric progression, which is controlled by the parameter $r$ in this chapter. The functions that arise from practical applications do not have an apparent bilinear form. Our interest is whether Copula Korobov will work well with more practical problems, which we will investigate in the next chapter. We will also look at generators from the BR criterion defined in (4.14) in the numerical experiments.

# Chapter 5

# Numerical Experiments

In this chapter, we compare the Copula generators and the BR generators to the generators based on the spectral test, the $\mathcal{P}_2$ test and the $\mathcal{R}_2$ test through numerical experiments. We also compare the Korobov lattice to Monte Carlo to assess whether we get smaller error bounds with Korobov lattice. In this chapter, we consider Asian option pricing, single-queue problem, valuation of mortgage-backed securities, and test functions. We will also use the quasi-regression techniques to gain insight on the problems at hand. This will help us assess if the assumptions made by the copula-based criterion are valid.

For each problem we consider, we will compute the RQMC estimate of $\psi_{LB}$ defined in Section 3.2.4 and the half width (HW) of the approximate 95% confidence interval of the estimate based on MC and the randomly shifted Korobov lattice. We will also compute $\psi_L$, $\psi_{B,M}$, $\psi_{LB,M}$ for different values of $M$. Unless otherwise stated, we will build a quasi-regression of $f$ using the Korobov lattice with the parameters $n = 32749, Q = 30, R = 15$ defined in Section 3.2.3 and the generator $a = 14251$ which comes from the spectral test.

## 5.1 Asian Option Pricing

The first problem we examine is the problem of Asian option pricing. The lower order ($<= 2$) quasi-regression has been shown to work well with this problem in [15], depending on the choice of parameters of the option.

### 5.1.1 Description of the problem

As defined in [8], an option gives the holder of the option the right to do something, but the holder does not have to exercise his right. An Asian call option gives the holder the right to receive a payoff at maturity of the option. The payoff depends on the arithmetic average of the price of the underlying stock during the life span of the option. More precisely, the payoff of an Asian call option at maturity is $\max(0, S_{avg} - K)$, where $S_{avg}$ denotes the (continuous) arithmetic average of the underlying stock and $K$ is the strike price of the option. The goal is to calculate the fair price of an Asian call option.

We consider the pricing of Asian call options under the Black-Scholes framework. We follow [13] for the presentation of the model.

Let $S(t)$ denotes the price of a stock at time $t$. Suppose that a stock price follows the risk-neutral stochastic differential equation

$$dS(t) = rS(t)dt + \sigma S(t)dW(t), \tag{5.1}$$

where $r$ is the risk-neural interest rate, $\sigma$ is the volatility of the stock price, and $\{W(t), t \geq 0\}$ is a standard Brownian motion.

It is generally impossible to simulate continuous models. For this Asian option problem, we cannot continuously observe the price of the underlying stock and compute the average. Instead, we observe the price of the underlying stock $s$ times in an equally spaced manner in time and take the average of the discretely observed prices. As the randomness in the problem only comes from the behaviour of $S_t$, the dimension of this problem is $s$, if we try to solve this problem by simulation.

The application of the Euler-Maruyama method [17] to (5.1) yields the following recurrence formula

$$S(t_i) = S(t_{i-1}) \exp\left\{(r - \sigma^2/2)\Delta + \sigma\sqrt{\Delta}X_i\right\}, \tag{5.2}$$

where $t_i = i\Delta$, $\Delta = T/s$, $X_i \sim N(0,1)$, $T$ is the maturity time of the option. The option's price is given by the expectation

$$\mathrm{E}\left(e^{-rT} \max\left(0, \frac{1}{s}\sum_{j=1}^{s} S(t_i) - K\right)\right), \tag{5.3}$$

where $K$ is the strike price of the option.

As the randomness of this model comes from $X_i's$ which are normally distributed, we need to obtain normally distributed random variables from uniformly distributed random

variables. Let $\psi(x)$ be the distribution function of a standard normal distribution and $\psi^{-1}(u)$ be the inverse of $\psi(x)$. Then, if $U \sim U[0,1)$, $\psi^{-1}(U) \sim N(0,1)$.

Based on this model and simulation method, we can write the corresponding function as

$$f(\mathbf{u}) = e^{(-rt)} \left( 0, \frac{1}{s} \sum_{j=1}^{s} \exp\{(r - \sigma^2/2)j\Delta + \sigma\sqrt{j\Delta}(\psi^{-1}(u_1) + \cdots + \psi^{-1}(u_j)\} - K \right).$$

The problem is completely specified by the set of parameters $(S_0, K, r, \sigma, T, s)$. It is argued in [15] that the larger $S_0$ is compared to $K$, the larger the proportion of the variance of $f$ explained by lower order quasi-regression terms. Thus, we expect that $\tilde{f}_{LB}$ will well approximate $f$ if $S_0$ is large. We consider the following three sets of parameters with only $S_0$ different: $(S_0, K, r, \sigma, T, s) = (120, 100, 0.05, 0.2, 1, 30)$, $(100, 100, 0.05, 0.2, 1, 30)$, and $(80, 100, 0.05, 0.2, 1, 30)$.

### 5.1.2 Simulation Study

Table 5.1 shows some of the estimated coefficients for bilinear components of $\tilde{f}_{LB}$ for the $S_0 = 100$ case. The estimates are normalized so that $\hat{d}_{1,2} = 1$.

Table 5.1: Estimated coefficients with $S_0 = 100$

| Coeff $\setminus j$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $d_{1,1+j}$ | 1.00E+00 | 9.76E-01 | 9.32E-01 | 9.06E-01 | 9.50E-01 |
| $d_{2,2+j}$ | 9.60E-01 | 8.84E-01 | 8.71E-01 | 8.27E-01 | 7.61E-01 |
| $d_{3,3+j}$ | 9.08E-01 | 8.28E-01 | 8.31E-01 | 7.50E-01 | 8.32E-01 |
| $d_{4,4+j}$ | 8.00E-01 | 8.54E-01 | 7.85E-01 | 7.83E-01 | 7.19E-01 |

We see that the geometrically decaying coefficients assumption which we studied in Section 4.3.2 is more applicable to this problem than the constant coefficients assumption. The implication of this observation is that the generators from the Copula criterion $\gamma_M$ with $M = 8$ will work better than the generators from $\gamma_M$ with a large value of $M$ like 28. This was not the case under the constant coefficients assumption studied in Section 4.3.1

Table 5.2 shows estimates of various $\psi$, the fraction of variance of $f$ captured by the quasi-regression, defined in Section 3.2.3 and Section 3.2.4. The HW of the approximate confidence intervals of $\psi_{LB}$ for the problem with $S_0 = 120$, 100, and 80 are 1.471e-04,

Table 5.2: Various estimates of $\psi$ for Asian option pricing

| $S_0$ | $\hat{\psi}_L$ | $\hat{\psi}_{LB,4}$ | $\hat{\psi}_{LB,8}$ | $\hat{\psi}_{LB,16}$ | $\hat{\psi}_{LB}$ |
|---|---|---|---|---|---|
| 120 | 9.31e-01 | 9.38e-01 | 9.44e-01 | 9.49e-01 | 9.50e-01 |
| 100 | 7.23e-01 | 7.85e-01 | 8.44e-01 | 9.02e-01 | 9.17e-01 |
| 80 | 1.35e-01 | 2.20e-01 | 3.03e-01 | 3.89e-01 | 4.11e-01 |

4.120e-04, and 1.425e-03, respectively. It is safe to say that the estimates of $\psi$ listed in the table are accurate enough to base our analysis on.

Table 5.2 shows that $\psi_L$ and $\psi_{LB}$ are positively correlated with $S_0$ while $\psi_B$ is negatively correlated with $S_0$. When $S_0 = 120$, about 93% of the variance of $f$ is explained by the linear approximation and only 2% of the variance is explained by the bilinear part. On the other hand, when $S_0 = 80$, only 13.5% of the variance is explained by the linear approximation of $f$ but 27.5% of the variance is captured by the bilinear components.

It is expected that the Copula Korobov would work well in the $S_0 = 100$ case because the assumption of the copula-based criterion holds and a moderate proportion of variance (12%) is captured by $\tilde{f}_{B,8}$. It is not obvious whether the Copula Korobov works better for the case with $S_0 = 120$ or the case with $S_0 = 80$. It may work better in the $S_0 = 120$ case because 95% of the variance of $f$ is captured by $\tilde{f}_{LB,8}$. Certainly, the assumption for the Copula criterion holds with this case. However, recall that when integrating a linear function with Korobov lattice, the choice of generator has no effect on the linear parts of $f$. In the $S_0 = 120$ case, the linear part, which accounts for 93% of the entire variance, is integrated in the same way regardless the choice of the generator. The copula-based criterion $\gamma_8$ focuses on the $\tilde{f}_{B,8}$ part of the approximation which accounts for only 1.7% of the variance when $S_0 = 120$. On the other hand, when $S_0 = 80$, a greater proportion (17%) of the variance is explained by $\tilde{f}_{B,8}$. So, the Copula Korobov may work better in this case despite the fact that bilinear regression explains only 40% of the variance. We will investigate this through numerical experiments.

Beside working with Copula, Spectral, $\mathcal{P}_2$ and $\mathcal{R}_2$ Korobov, we can also include BR Korobov in our numerical experiments. The BR Korobov is the Korobov lattice obtained from the optimal generators with respect to the BR criterion defined in (4.14). For different value of $S_0$, we have different $f$. Thus, $\tilde{f}_{LB}$ are different in each of the $S_0 = 120$, 100 and 80 case. However, the BR criterion gave exactly the same generators for all the three cases. Table 5.3 lists the generators.

Table 5.4, lists the half width (HW) of the confidence intervals of the estimators based

Table 5.3: BR generators for Asian option pricing

| $n$ | 1021 | 2039 | 4093 | 8191 | 16381 | 32749 | 65521 | 131071 |
|---|---|---|---|---|---|---|---|---|
| $a$ | 263 | 1831 | 1853 | 6975 | 1513 | 8602 | 9819 | 24768 |

on MC and the Korobov lattice for all the three cases. Henceforth, when we just say HW, it means the half width of the approximate 95% confidence interval for an estimate. Unless otherwise stated, we obtain the HW based on 30 replications in this thesis.

In all the three cases, the Korobov lattice based on the five criteria gives smaller HW than MC does. As expected, the Copula Korobov and the BR Korobov seem to work best in the $S_0 = 100$ case. When $S_0 = 100$, Copula and BR generally outperform the Spectral, $\mathcal{P}_2$ and $\mathcal{R}_2$ Korobov. For the other two cases, the Copula and BR Korobov are still competitive to the Spectral, $\mathcal{P}_2$ and $\mathcal{R}_2$ Korobov, but their advantages are not apparent. It seems that the Copula Korobov and the BR Korobov slightly perform better in the $S_0 = 120$ case than in the $S_0 = 80$ case. So, in this Asian option pricing problem, having only $\hat{\psi}_B$ moderately high does not mean the Copula and BR Korobov work well.

Now we fix $S_0 = 100$ and $n = 1021$, 4093, 16381, and vary the $s$. Table 5.5 lists HW for $n = 1021$, 4093, and 16381.

For all the dimensions considered, the Korobov lattice gives smaller HW than MC does. On the other hand, the size of HW based on MC is not affected by, or even decreases when the dimension of the problem increases. Nevertheless, the Korobov lattice still gives smaller HW when $s = 1000$, so it is safe to say that for this Asian option pricing problem, there is an advantage to use the Korobov lattice over MC.

We see that the Copula Korobov and the BR Korobov show superior performance even when the dimension of the problem is large. It is surprising that the BR Korobov works well for the high dimensional problems like when $s = 500$ despite the fact that generators are based on $f$ for $s = 30$. Overall, we can say that the Copula criterion and the BR criterion give generators that work well for Asian option pricing.

Table 5.4: HW for $S_0 = 120, 100$, and 80

| $S_0$ | $n$ | MC | Copula | BR | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|---|
| 120 | 1021 | 1.52E-01 | 1.50E-02 | 1.84E-02 | 1.68E-02 | 2.67E-02 | 2.67E-02 |
| | 2039 | 1.08E-01 | 1.78E-02 | 9.57E-03 | 1.11E-02 | 8.92E-03 | 8.80E-03 |
| | 4093 | 7.63E-02 | 8.49E-03 | 5.71E-03 | 8.30E-03 | 6.29E-03 | 8.59E-03 |
| | 8191 | 5.41E-02 | 2.90E-03 | 3.91E-03 | 3.94E-03 | 5.27E-03 | 5.27E-03 |
| | 16381 | 3.82E-02 | 1.60E-03 | 1.85E-03 | 2.30E-03 | 2.16E-03 | 2.18E-03 |
| | 32749 | 2.71E-02 | 2.25E-03 | 1.64E-03 | 1.33E-03 | 1.48E-03 | 1.48E-03 |
| | 65521 | 1.91E-02 | 1.03E-03 | 7.68E-04 | 1.17E-03 | 8.40E-04 | 1.24E-03 |
| | 131071 | 1.35E-02 | 6.19E-04 | 4.61E-04 | 5.63E-04 | 8.21E-04 | 6.10E-04 |
| 100 | 1021 | 9.08E-02 | 2.63E-02 | 2.02E-02 | 1.78E-02 | 5.10E-02 | 5.10E-02 |
| | 2039 | 6.46E-02 | 3.33E-02 | 1.32E-02 | 1.46E-02 | 1.26E-02 | 1.26E-02 |
| | 4093 | 4.57E-02 | 8.98E-03 | 7.44E-03 | 1.24E-02 | 1.22E-02 | 1.23E-02 |
| | 8191 | 3.24E-02 | 6.03E-03 | 4.99E-03 | 8.02E-03 | 8.37E-03 | 8.37E-03 |
| | 16381 | 2.29E-02 | 3.14E-03 | 3.21E-03 | 4.03E-03 | 4.25E-03 | 3.52E-03 |
| | 32749 | 1.62E-02 | 3.16E-03 | 1.62E-03 | 3.01E-03 | 2.96E-03 | 2.96E-03 |
| | 65521 | 1.15E-02 | 1.45E-03 | 1.63E-03 | 1.82E-03 | 2.20E-03 | 1.98E-03 |
| | 131071 | 8.10E-03 | 1.01E-03 | 8.63E-04 | 9.64E-04 | 1.36E-03 | 1.12E-03 |
| 80 | 1021 | 1.52E-02 | 1.16E-02 | 1.26E-02 | 1.19E-02 | 1.21E-02 | 1.21E-02 |
| | 2039 | 1.11E-02 | 1.36E-02 | 6.81E-03 | 9.49E-03 | 7.53E-03 | 7.13E-03 |
| | 4093 | 7.84E-03 | 9.86E-03 | 4.64E-03 | 6.23E-03 | 6.23E-03 | 6.28E-03 |
| | 8191 | 5.72E-03 | 3.38E-03 | 5.08E-03 | 4.91E-03 | 3.30E-03 | 3.30E-03 |
| | 16381 | 4.02E-03 | 2.64E-03 | 2.06E-03 | 3.04E-03 | 2.18E-03 | 2.43E-03 |
| | 32749 | 2.85E-03 | 1.58E-03 | 2.30E-03 | 1.79E-03 | 1.46E-03 | 1.46E-03 |
| | 65521 | 2.02E-03 | 1.61E-03 | 9.82E-04 | 1.01E-03 | 9.61E-04 | 1.18E-03 |
| | 131071 | 1.43E-03 | 7.63E-04 | 9.55E-04 | 6.41E-04 | 8.83E-04 | 1.00E-03 |

Table 5.5: HW for varying $s$

| $n$ | $s$ | MC | Copula | BR | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|---|
| 1021 | 10 | 9.55E-02 | 1.81E-02 | 2.42E-02 | 2.19E-02 | 2.62E-02 | 2.62E-02 |
| | 30 | 9.08E-02 | 2.63E-02 | 2.02E-02 | 1.78E-02 | 5.10E-02 | 5.10E-02 |
| | 50 | 9.05E-02 | 3.43E-02 | 1.68E-02 | 3.36E-02 | 4.99E-02 | 4.99E-02 |
| | 100 | 9.01E-02 | 3.18E-02 | 3.25E-02 | 3.98E-02 | 5.94E-02 | 5.94E-02 |
| | 200 | 8.93E-02 | 3.54E-02 | 3.09E-02 | 3.91E-02 | 4.71E-02 | 4.71E-02 |
| | 500 | 8.95E-02 | 3.62E-02 | 3.21E-02 | 3.95E-02 | 6.06E-02 | 6.06E-02 |
| | 1000 | 8.85E-02 | 4.84E-02 | 3.23E-02 | 3.95E-02 | 5.38E-02 | 5.38E-02 |
| 4093 | 10 | 4.81E-02 | 2.32E-02 | 5.99E-03 | 1.46E-02 | 7.49E-03 | 7.49E-03 |
| | 30 | 4.57E-02 | 8.98E-03 | 7.44E-03 | 1.24E-02 | 3.19E-02 | 3.19E-02 |
| | 50 | 4.54E-02 | 1.05E-02 | 8.64E-03 | 1.09E-02 | 3.37E-02 | 3.37E-02 |
| | 100 | 4.51E-02 | 1.20E-02 | 1.31E-02 | 9.29E-03 | 3.52E-02 | 3.52E-02 |
| | 200 | 4.46E-02 | 1.61E-02 | 1.42E-02 | 1.24E-02 | 2.58E-02 | 2.58E-02 |
| | 500 | 4.46E-02 | 1.72E-02 | 1.49E-02 | 1.94E-02 | 2.65E-02 | 2.65E-02 |
| | 1000 | 4.45E-02 | 1.62E-02 | 1.81E-02 | 1.53E-02 | 2.48E-02 | 2.48E-02 |
| 16381 | 10 | 2.41E-02 | 2.29E-03 | 1.85E-03 | 3.42E-03 | 2.22E-03 | 2.11E-03 |
| | 30 | 2.29E-02 | 3.14E-03 | 3.21E-03 | 4.03E-03 | 4.25E-03 | 3.52E-03 |
| | 50 | 2.27E-02 | 3.54E-03 | 3.82E-03 | 4.84E-03 | 4.83E-03 | 4.63E-03 |
| | 100 | 2.25E-02 | 4.12E-03 | 5.03E-03 | 6.78E-03 | 4.42E-03 | 4.34E-03 |
| | 200 | 2.24E-02 | 6.09E-03 | 5.49E-03 | 7.20E-03 | 5.17E-03 | 4.06E-03 |
| | 500 | 2.23E-02 | 7.94E-03 | 6.96E-03 | 1.08E-02 | 9.31E-03 | 1.06E-02 |
| | 1000 | 2.23E-02 | 7.97E-03 | 5.44E-03 | 9.98E-03 | 8.30E-03 | 8.38E-03 |

## 5.2 Single Queue Problem

In the Asian option problem, we constructed the quasi-regression only for the problem with $s = 30$. In this section, we construct two quasi-regressions: one for a small value of $s$ and another one for a large value of $s$. It allows us to study if the linearity of a problem drastically changes for different values of $s$.

### 5.2.1 Description of the problem

The problem we consider here is a modification of two related examples in [14]: Example 1.2 and Example 4.6. The following description of the problem largely comes from the description for Example 1.2 and Example 4.6 in [14].

We consider a bank that operates all day. We assume that there is only one teller, that the clients arrive according to Poisson process at a rate of 1 per minute, and that each client stays with the teller for a random length of time that is exponentially distributed with mean $\mu_s$. The mean $\mu_s$ is randomly determined at the beginning of the day. More precisely, $\mu_s = 35$ seconds with probability 0.2, $\mu_s = 50$ seconds with probability 0.7 , and $\mu_s = 55$ seconds with probability 0.1. We assume these service times and all inter arrival times are independent from each other. The goal is to estimate the expected length of time that the first $d$ customers will wait for a teller at the bank during a given day of operation, where $d$ is a constant specified beforehand.

Given a sample point, we can evaluate the realization of the average waiting time by simulation, instead of formulating the closed form of the corresponding $f$. The simulation method below is described in [14]. In order to simulate the waiting time for the $j$th customer, we use Lindley's equation [16], which provides us a recurrence relation for the waiting time $W_j$:

$$W_j = \max(0, W_{j-1} + S_{j-1} - A_j), \;\; j \geq 2, \tag{5.4}$$

where

$W_j$ = waiting time in the queue of the $j$th customer,
$A_j$ = interarrival time between $(j-1)$th and $j$th customers,
$S_j$ = service time of the $j$th customer,
and $W_1 = 0$.

Let $C_d :=$ the expected wait time for the first $d$ customers in the queue. Then,

$$C_d = \mathrm{E}\left(\frac{1}{d}\sum_{j=1}^{d} W_j\right).$$

The dimension of this problem is $2d-1$. We need one random number to determine the mean speed of the server, $d-1$ random numbers to simulate $A_2, \ldots, A_d$, and another $d-1$ random numbers to simulate $S_1, \ldots, S_{d-1}$. Notice that $A_1$ is irrelevant for this simulation. The recurrence equation (5.4) allows us to approximate $C_d$ by simulation.

### 5.2.2 Simulation Study

We consider $d = 30$ and $d = 100$ cases to construct quasi-regression. Table 5.6 lists some of the estimates of the coefficients for the bilinear parts of $\tilde{f}_{LB}$ for the two cases.

Table 5.6: Estimated coefficients for the bank problem

| $d$ | Coeff $\setminus j$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| 30 | $d_{1,1+j}$ | 3.62E-01 | -1.72E-01 | 4.76E-01 | -2.58E-01 | 5.17E-01 |
| | $d_{2,2+j}$ | -7.06E-01 | 3.54E-01 | -4.35E-01 | 2.65E-01 | -3.09E-01 |
| | $d_{3,3+j}$ | -2.16E-01 | 2.34E-01 | -1.82E-01 | 1.73E-01 | -6.29E-02 |
| | $d_{4,4+j}$ | -7.28E-01 | 4.07E-01 | -4.88E-01 | 3.08E-01 | -3.98E-01 |
| 100 | $d_{1,1+j}$ | 2.02E-01 | -8.78E-02 | 2.18E-01 | -1.49E-01 | 3.44E-01 |
| | $d_{2,2+j}$ | -2.93E-01 | 1.30E-01 | -1.82E-01 | 1.10E-01 | -1.79E-01 |
| | $d_{3,3+j}$ | -9.35E-02 | 7.27E-02 | -1.00E-01 | 1.05E-01 | -2.46E-02 |
| | $d_{4,4+j}$ | -3.16E-01 | 1.80E-01 | -1.91E-01 | 1.35E-01 | -1.58E-01 |

We see that all the coefficients for the $d = 100$ case are smaller in absolute value. This makes sense because the variance of $f$ is spread over a larger number of terms in the $d = 100$ case. There does not seem to be a clear pattern in the coefficients of the bilinear components, unlike the Asian option pricing problem.

Table 5.7 lists the estimates of various $\psi$. The HW for the estimate of $\psi_{BL}$ is 4.559e-04 for the $d = 30$ case and 8.211e-04 for the $d = 100$ case. We believe that the figures in the table is accurate enough to base our analysis on.

Table 5.7: Various estimates of $\psi$ for the bank problem

| $d$ | $\hat{\psi}_L$ | $\hat{\psi}_{LB,4}$ | $\hat{\psi}_{LB,8}$ | $\hat{\psi}_{LB,16}$ | $\hat{\psi}_{LB}$ |
|---|---|---|---|---|---|
| 30 | 4.86e-01 | 5.39e-01 | 5.67e-01 | 5.89e-01 | 6.12e-01 |
| 100 | 4.58e-01 | 4.84e-01 | 5.04e-01 | 5.28e-01 | 5.95e-01 |

About 60% of the variance of $f$ is explained by the bilinear regression of $f$. So the Copula Korobov and BR Korobov may not perform well for this bank problem. The underlying function $f$ is sightly more linear but less bilinear in the $d = 30$ case. However, the changes are marginal. It suggests that the structure of the problem does not alter much when the dimensionality of the problem gets larger. For fixed $M$, we see that $\hat{\psi}_{LB,M}$ for the $d = 30$ case is greater than the one for the $d = 100$ case. This makes sense because for fixed $M$, the bilinear regression $\tilde{f}_{LB,M}$ include relatively less bilinear terms as the dimensionality of the problem increases. When $M = 8$ and $s = 30$, for instance, $\tilde{f}_{LB,M}$ include about 43% of all bilinear terms. When $M = 8$ and $s = 100$, it is just 17%. Since $\tilde{f}_{LB,M}$ exclude more bilinear terms that explain variance of $f$, $\hat{\psi}_{LB,M}$ gets smaller as $s$ increases. We can make the exactly same observation for the Asian option pricing problem.

We obtain BR generators for the both cases. Table 5.8 lists the BR generators for this bank problem.

Table 5.8: BR generators for the bank problem

| $d\backslash n$ | 1021 | 2039 | 4093 | 8191 | 16381 | 32749 | 65521 | 131071 |
|---|---|---|---|---|---|---|---|---|
| 30 | 559 | 349 | 1183 | 3159 | 10611 | 17991 | 27552 | 68361 |
| 100 | 263 | 976 | 2702 | 1337 | 638 | 20375 | 16182 | 32408 |

Intuitively, the generators based on the $d = 30$ case would perform better for lower dimensional problems and the generators based on the $d = 100$ case perform better for the higher dimensional problems. First we fix $d = 30$ and $d = 100$ and vary the value of $n$. The column for BR(30) and BR(100) represent HW for BR Korobov with generator based on the $d = 30$ and $d = 100$ cases, respectively.

We first notice that the Korobov lattice in overall performs much better than MC does for all values of $n$ considered. Copula Korobov seems to give relatively large HW when $d = 30$. This is not the case when $d = 100$. Despite the fact that the bilinear regression explains only 60% of the variance of $f$, Copula Korobov and BR Korobov perform competitive compared to other Korobov lattices when $d = 100$. We had a similar observation with the Asian option pricing in the $S_0 = 80$ case. This suggests that the function does not need to be highly bilinear for the Copula Korobov to become competitive to the Korobov lattice based on other search criteria. Contrary to our intuition, the BR(30) Korobov and BR(100) Korobov do not perform significantly different from each other. This suggests that we may be able to get away with building the bilinear regression for the low dimensional version of the problem. In order to assess whether or not BR(100) Korobov performs better than

Table 5.9: HW for $d = 30$ and 100

| $d$ | $n$ | MC | Copula | BR(30) | BR(100) | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|---|---|
| 30 | 1021 | 1.73E-02 | 6.93E-03 | 5.99E-03 | 6.75E-03 | 9.87E-03 | 1.08E-02 | 1.08E-02 |
| | 2039 | 1.23E-02 | 8.07E-03 | 3.62E-03 | 4.42E-03 | 3.82E-03 | 3.97E-03 | 3.89E-03 |
| | 4093 | 8.68E-03 | 5.14E-03 | 2.20E-03 | 2.56E-03 | 2.70E-03 | 2.26E-03 | 2.70E-03 |
| | 8191 | 6.13E-03 | 1.57E-03 | 1.58E-03 | 1.46E-03 | 2.86E-03 | 1.69E-03 | 1.69E-03 |
| | 16381 | 4.34E-03 | 1.22E-03 | 9.93E-04 | 9.00E-04 | 8.13E-04 | 1.15E-03 | 1.13E-03 |
| | 32749 | 3.07E-03 | 7.92E-04 | 1.95E-03 | 1.31E-03 | 5.35E-04 | 6.66E-04 | 6.66E-04 |
| | 65521 | 2.17E-03 | 8.77E-04 | 5.31E-04 | 4.44E-04 | 3.73E-04 | 4.51E-04 | 4.47E-04 |
| | 131071 | 1.54E-03 | 3.94E-04 | 3.39E-04 | 9.34E-04 | 3.85E-04 | 3.47E-04 | 3.51E-04 |
| 100 | 1021 | 2.58E-02 | 1.32E-02 | 1.16E-02 | 1.04E-02 | 1.47E-02 | 1.61E-02 | 1.61E-02 |
| | 2039 | 1.84E-02 | 8.63E-03 | 7.88E-03 | 6.85E-03 | 9.73E-03 | 7.79E-03 | 7.50E-03 |
| | 4093 | 1.30E-02 | 6.97E-03 | 4.31E-03 | 5.48E-03 | 5.37E-03 | 5.86E-03 | 4.41E-03 |
| | 8191 | 9.16E-03 | 2.94E-03 | 3.57E-03 | 2.99E-03 | 4.80E-03 | 4.42E-03 | 4.42E-03 |
| | 16381 | 6.45E-03 | 2.88E-03 | 2.44E-03 | 1.60E-03 | 2.76E-03 | 2.26E-03 | 3.02E-03 |
| | 32749 | 4.56E-03 | 1.30E-03 | 2.10E-03 | 2.23E-03 | 1.46E-03 | 1.32E-03 | 1.32E-03 |
| | 65521 | 3.22E-03 | 1.52E-03 | 1.12E-03 | 8.11E-04 | 1.28E-03 | 1.07E-03 | 1.01E-03 |
| | 131071 | 2.28E-03 | 6.95E-04 | 5.27E-04 | 1.03E-03 | 5.83E-04 | 7.99E-04 | 7.44E-04 |

BR(30) Korobov, we vary the value of $d$ while fixing $n$. Table 5.10 lists the HW for varying $d$ while $n$ is fixed to 1021,4093,16381.

We see that BR(100) performs somewhat better than BR(30) in high dimensional cases and vice versa. However, BR(30) is still competitive to the Korobov lattice used for comparison in high dimensional cases. We also see that the Copula Korobov gives competitive results to other Korobov lattices.

Overall we see that Copula and BR Korobov are competitive to Spectral, $\mathcal{P}_2$, and $\mathcal{R}_2$ Korobov even though the problem does not satisfy the assumption made by the copula-based and BR criterion.

Table 5.10: HW for varying $d$

| $n$ | $d$ | MC | Copula | BR(30) | BR(100) | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|---|---|
| 1021 | 10 | 1.06E-02 | 2.69E-03 | 2.77E-03 | 3.23E-03 | 2.64E-03 | 3.63E-03 | 2.64E-03 |
| | 20 | 1.47E-02 | 6.85E-03 | 4.05E-03 | 5.65E-03 | 6.88E-03 | 6.37E-03 | 6.88E-03 |
| | 30 | 1.73E-02 | 6.93E-03 | 5.99E-03 | 6.75E-03 | 9.87E-03 | 1.08E-02 | 9.87E-03 |
| | 50 | 2.09E-02 | 9.47E-03 | 6.83E-03 | 1.04E-02 | 1.05E-02 | 1.33E-02 | 1.05E-02 |
| | 100 | 2.58E-02 | 1.32E-02 | 1.16E-02 | 1.04E-02 | 1.47E-02 | 1.61E-02 | 1.47E-02 |
| | 500 | 3.23E-02 | 2.30E-02 | 1.90E-02 | 1.56E-02 | 2.10E-02 | 2.62E-02 | 2.10E-02 |
| | 1000 | 3.26E-02 | 1.62E-02 | 1.92E-02 | 1.90E-02 | 1.98E-02 | 2.88E-02 | 1.98E-02 |
| 4093 | 10 | 5.41E-03 | 4.57E-03 | 8.76E-04 | 1.18E-03 | 8.07E-04 | 1.77E-03 | 8.07E-04 |
| | 20 | 7.36E-03 | 5.96E-03 | 1.56E-03 | 2.17E-03 | 1.92E-03 | 2.56E-03 | 1.92E-03 |
| | 30 | 8.68E-03 | 5.14E-03 | 2.20E-03 | 2.56E-03 | 2.17E-03 | 2.70E-03 | 2.17E-03 |
| | 50 | 1.05E-02 | 5.38E-03 | 3.40E-03 | 3.63E-03 | 3.40E-03 | 3.99E-03 | 3.40E-03 |
| | 100 | 1.30E-02 | 6.97E-03 | 4.31E-03 | 5.48E-03 | 5.06E-03 | 5.37E-03 | 5.06E-03 |
| | 500 | 1.61E-02 | 7.44E-03 | 9.71E-03 | 7.56E-03 | 8.48E-03 | 8.62E-03 | 8.48E-03 |
| | 1000 | 1.62E-02 | 9.24E-03 | 8.65E-03 | 6.84E-03 | 9.35E-03 | 8.36E-03 | 9.35E-03 |
| 16381 | 10 | 2.70E-03 | 4.14E-04 | 3.52E-04 | 3.12E-04 | 3.85E-04 | 4.84E-04 | 3.85E-04 |
| | 20 | 3.68E-03 | 1.07E-03 | 5.59E-04 | 7.53E-04 | 6.53E-04 | 9.32E-04 | 6.53E-04 |
| | 30 | 4.34E-03 | 1.22E-03 | 9.93E-04 | 9.00E-04 | 8.47E-04 | 8.13E-04 | 8.47E-04 |
| | 50 | 5.23E-03 | 1.45E-03 | 1.64E-03 | 1.12E-03 | 1.66E-03 | 2.13E-03 | 1.66E-03 |
| | 100 | 6.45E-03 | 2.88E-03 | 2.44E-03 | 1.60E-03 | 1.67E-03 | 2.76E-03 | 1.67E-03 |
| | 500 | 8.04E-03 | 3.54E-03 | 3.86E-03 | 3.73E-03 | 3.30E-03 | 4.58E-03 | 3.30E-03 |
| | 1000 | 8.08E-03 | 3.70E-03 | 3.69E-03 | 3.68E-03 | 3.63E-03 | 3.84E-03 | 3.63E-03 |

## 5.3 Mortgage-Backed Securities

We consider the valuation of mortgage-backed securities (MBS). The holder of a mortgage-backed security receives the cash flows which come from the payments by a mortgage holder. The goal is to find the present value of such securities. We follow the presentation of the problem in [14], which in turn follows [2] and [20].

The problem here is to compute an expectation of the form

$$M_0 = \mathrm{E}\left(\sum_{l=0}^{M} v_l c_l\right),$$

which represents the present value of this security. Here $v_l$ is the discount factor for month $l$ and $c_l$ is the cash flow for month $l$. Both of these quantities depend on the interest rate process in the following way. Let $i_l$ be the interest rate for month $l$. As in [2], we use the interest rate model

$$i_l = K_0 e^{\xi_l} i_{l-1},\ l \geq 1,$$

where $\xi_l \sim N(0, \sigma^2)$. Then,

$$v_l = \prod_{k=0}^{l-1} (1 + i_k)^{-1}$$

and

$$c_l = c r_l((1 - w_l) + w_l f_l,$$

where

$$\begin{aligned}
c &= \text{ monthly mortgage payment,} \\
w_l &= \text{fraction of remaining mortgages prepaying in month } l, \\
&= K_1 + K_2 \arctan(K_3 \times i_l + K_4), \\
r_l &= \text{ fraction of remaining mortgage at month } l, \\
&= \prod_{k=0}^{l=1} (1 - w_k), \\
f_l &= \text{ (remaining annuity at month } l)/c \\
&= \sum_{k=0}^{M-1} (1 + i_0)^{-k}.
\end{aligned}$$

As in [2], we choose $K_0 = \exp(-\sigma^2/2)$ so that $\mathrm{E}(i_k) = i_0$. Hence, this problem is completely characterized by $(K_1, K_2, K_3, K_4, \sigma, i_0)$. In [2], two sets of parameters are considered. The first set is given by

$$(K_1, K_2, K_3, K_4, \sigma, i_0) = (0.01, -0.05, 10, 0.5, 0.02, 0.007)$$

and such that the corresponding function $f(\cdot)$ satisfying

$$M_0 = \int_{[0,1)^s} f(\mathbf{u}) d\mathbf{u}$$

is almost linear in its inputs $u_1, \ldots, u_s$. We refer this case as the linear case. The second set

$$(K_1, K_2, K_3, K_4, \sigma, i_0) = (0.04, 0.0222, -1500, 7, 0.02, 0.007)$$

does not have such a strong linear component. We refer this case as the non-linear case as done in [2].

Sine the randomness of this problem comes from the behaviour of the interest rate that fluctuates monthly, the dimensionality of the problem is equal to the duration of the life of the MBS in months. As the length of a mortgage is typically 30 years (360 months) or longer, the valuation of MBS is considered a large dimensional problem.

## 5.4 Simulation Study

We are interested in the valuation of 30 years (360 dimensional) MBS for both the linear and non-linear case. As this problem is very high dimensional problem and it is very computationally expensive to build the full bilinear regression $\tilde{f}_{LB}$, we build $\tilde{f}_{LB,30}$ instead.

Table 5.11 lists the selected coefficients for the bilinear components of the bilinear regression of $f$. There does not seem to be an obvious pattern in the estimated coefficients.

Table 5.12 lists various estimates of $\psi$. The table has $\hat{\psi}_{LB,30}$ instead of $\hat{\psi}_{LB,M}$ as we did not build a full bilinear regression. The HW for $\hat{\psi}_{LB,30}$ for the linear and the non-linear case are 4.531e-04 and 7.904e-04, respectively. So those figures on the table are accurate.

Over 95% of the variance of $f$ is captured by the linear components in the linear case. In the non-linear case, the linear components still explain close to 90% of the variance of $f$ in the non-linear case. So, the function corresponding for the non-linear case is actually very linear. This observation suggests that the Korobov lattice will perform much better

Table 5.11: Estimated coefficients for the MBS problem

| case | Coeff \ $j$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Linear | $d_{1,1+j}$ | 1.52E-01 | -1.81E-01 | 2.57E-01 | 1.06E+00 | 4.64E-01 |
| | $d_{2,2+j}$ | 3.05E-01 | -2.80E-02 | -3.49E-02 | 3.20E-01 | 3.28E-01 |
| | $d_{3,3+j}$ | 1.47E-01 | 1.60E-01 | 1.21E-01 | -2.23E-01 | 1.41E-01 |
| | $d_{4,4+j}$ | 8.36E-03 | -1.00E-01 | 2.53E-01 | 1.13E-01 | 1.21E-01 |
| Non-linear | $d_{1,1+j}$ | -1.47E-01 | -6.00E-01 | -2.45E-01 | 6.42E-01 | -4.81E-02 |
| | $d_{2,2+j}$ | -1.74E-01 | -4.53E-01 | -4.24E-01 | -7.32E-02 | -1.09E-01 |
| | $d_{3,3+j}$ | -1.53E-01 | -2.24E-01 | -2.69E-01 | -5.52E-01 | -2.71E-01 |
| | $d_{4,4+j}$ | -4.04E-01 | -5.84E-01 | -1.18E-01 | -2.47E-01 | -4.61E-01 |

Table 5.12: Various estimates of $\psi_\cdot$ for the MBS problem

| case | $\hat{\psi}_L$ | $\hat{\psi}_{LB,4}$ | $\hat{\psi}_{LB,8}$ | $\hat{\psi}_{LB,12}$ | $\hat{\psi}_{LB,30}$ |
|---|---|---|---|---|---|
| Linear | 9.55e-01 | 9.55e-01 | 9.55e-01 | 9.55e-01 | 9.55e-01 |
| Non-linear | 8.96e-01 | 9.02e-01 | 9.08e-01 | 9.19e-01 | 9.32e-01 |

than MC for this problem. In either case, a significant proportion of the variance of $f$ is captured by the bilinear function of $f$. Thus, we expect that the Copula generators performs well with this problem.

Instead of obtaining BR generators from this MBS problem, we test if the BR generators from the bank problem for the $d = 100$ case perform well for this problem. Table 5.13 lists HW for linear and non-linear cases with different values $n$.

There is a clear advantage to use the Korobov lattice over MC. The HW we obtain from MC with $n = 131071$ is greater than the HW the Korobov lattice gives with $n = 1021$. Copula Korobov performs well in the linear case, but not so well in the non-linear case. The BR Korobov performs well in either case. This is a surprising result. The generators designed to work well for the bank problem actually work well for the MBS problem. This suggests that the generators based on the BR criterion tailored to the bank problem may work well in other problems, instead of only the bank problem.

Table 5.13: HW for $s = 360$

| case | $n$ | MC | Copula | BR | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|---|
| Linear | 1021 | 7.23E-02 | 1.52E-04 | 1.62E-04 | 1.50E-04 | 1.79E-04 | 1.79E-04 |
| | 2039 | 5.11E-02 | 8.41E-05 | 1.00E-04 | 6.69E-05 | 9.63E-05 | 7.71E-05 |
| | 4093 | 3.61E-02 | 1.38E-05 | 1.34E-05 | 1.87E-05 | 1.53E-05 | 1.28E-05 |
| | 8191 | 2.56E-02 | 5.87E-06 | 4.62E-06 | 6.00E-06 | 3.93E-06 | 3.93E-06 |
| | 16381 | 1.81E-02 | 1.37E-06 | 1.63E-06 | 1.88E-06 | 1.87E-06 | 1.74E-06 |
| | 32749 | 1.28E-02 | 5.10E-07 | 4.86E-07 | 5.51E-07 | 6.26E-07 | 6.26E-07 |
| | 65521 | 9.04E-03 | 1.92E-07 | 1.28E-07 | 1.68E-07 | 1.27E-07 | 1.38E-07 |
| | 131071 | 6.39E-03 | 4.93E-08 | 4.50E-08 | 3.52E-08 | 2.89E-08 | 3.86E-08 |
| Non-linear | 1021 | 4.81E-02 | 5.98E-04 | 5.99E-04 | 1.18E-03 | 2.38E-03 | 2.38E-03 |
| | 2039 | 3.40E-02 | 7.53E-04 | 1.94E-04 | 3.41E-04 | 1.05E-04 | 2.95E-04 |
| | 4093 | 2.41E-02 | 1.18E-04 | 6.75E-05 | 1.31E-04 | 6.75E-05 | 6.03E-05 |
| | 8191 | 1.71E-02 | 3.12E-05 | 2.58E-05 | 1.84E-04 | 5.98E-05 | 5.98E-05 |
| | 16381 | 1.21E-02 | 1.09E-05 | 9.54E-06 | 2.42E-05 | 1.21E-05 | 5.96E-06 |
| | 32749 | 8.52E-03 | 1.20E-05 | 5.42E-06 | 1.04E-05 | 7.37E-06 | 7.37E-06 |
| | 65521 | 6.03E-03 | 7.52E-06 | 7.95E-07 | 1.49E-06 | 7.44E-07 | 1.52E-06 |
| | 131071 | 4.26E-03 | 1.26E-06 | 6.71E-07 | 6.67E-07 | 6.71E-07 | 1.10E-06 |

## 5.5 Test Function

Lastly, we look at a function with known properties. The test function we look at is perfectly non-linear and non-bilinear. That is, the bilinear regression of $f$ does not capture any variance of $f$. The test function we consider is experimented in [23].

$$f(\mathbf{u}) = \prod_{j=1}^{s} \frac{|4u_j - 2| + a_j}{1 + a_j}. \tag{5.5}$$

We consider the following three set of parameters:
(i) $a_1 = 1$ for $j = 1, \dots s$
(ii) $a_j = j^2$ for $j = 1, \dots s$
(iii) $a_j = (s - j)^2$ for $j = 1, \dots s$.

The influence of $u_k$ to the variance of $f$ has an inverse relationship with the relative size of $a_k$, as discussed in [23]. In particular, if $a_1 < a_2 < \cdots < a_s$, the order of the importance of the variables is $a_k < a_{k-1} < \cdots a_1$. So, all the variables have the same significance in (i), the variables with lower indices contribute more in (ii), and the variables with higher indices contribute more in (iii).

As in [23], this function has the property that $I(f) = 1$ for any $a_1, \dots, a_s > 0$. To see this, we have

$$\begin{aligned}
\int_{[0,1)} f(\mathbf{u}) &= \int_{[0,1)^s} \prod_{j=1}^{s} \frac{|4u_j - 2| + a_j}{1 + a_j} d\mathbf{u} \\
&= \prod_{j=1}^{s} \int_{[0,1)} \frac{|4u_j - 2| + a_j}{1 + a_j} du_j \\
&= \prod_{j=1}^{s} 1 = 1.
\end{aligned}$$

This function is perfectly non-linear since for any $k \in \{1, \dots s\}$ since,

$$\begin{aligned}\int_{[0,1)} f(\mathbf{u})\sqrt{3}(2u_k - 1) &= \int_{[0,1)^s} \prod_{j=1}^{s} \frac{|4u_j - 2| + a_j}{1 + a_j}\sqrt{3}(2u_k - 1)d\mathbf{u} \\ &= \prod_{j\neq k}^{s} \int_{[0,1)} \frac{|4u_j - 2| + a_j}{1 + a_j} du_j \int_{[0,1)} \frac{|4u_j - 2| + a_j}{1 + a_j}\sqrt{3}(2u_k - 1)du_k \\ &= 1 \cdot 0 = 0.\end{aligned}$$

Similarly, we can show that for any $k < l \in \{1, \dots s\}$

$$\int_{[0,1)} f(\mathbf{u})\sqrt{3}(2u_k - 1)\sqrt{3}(2u_l - 1)d\mathbf{u} = 0.$$

So, the coefficients corresponding to the linear and bilinear components of $f$ are all zero. Thus, the bilinearity assumption made by the copula and BR criteria are violated for this function. We expect the Copula and BR Korobov to perform poorly with this function.

Since this function is perfectly non-bilinear, we cannot build a bilinear regression of $f$. Hence, we cannot obtain BR generators for this function. So, we take the BR generators we obtained for Asian option pricing in Section 5.1.2. We are interested in whether or not the BR generators designed to work well with Asian option pricing can give good results for completely unrelated problems.

## 5.6 Simulation study

For each of the three cases, we fix $s$ to 30, 100, 500 and vary $n$. Table 5.14, 5.15 and 5.16 lists HW for the case (i), (ii) and (iii), respectively.

In the case (i), all MC and the Korobov lattice struggle to approximate the mean of $f$, except when $s = 30$. When $s = 500$, we have small HW, but the estimated integrals are far from the true value. Thus, we exclude the $s = 500$ case from analysis.

Again, the Korobov lattice outperforms MC. We see that $\mathcal{R}_2$ often gives smallest HW among all MC and the Korobov lattice. Even if $f$ violates the assumption made by the copula-based criterion, Copula Korobov is competitive to other Korobov lattices. Also, the BR generators based on the Asian option pricing problem perform well for this test

function. This reinforces our guess that the generators based on the BR criterion may work well in a general setting.

Overall, the Copula Korobov is competitive with other Korobov lattice even when the bilinear assumption does not hold. We also saw that BR generators based on one problem could work well for other problems.

Table 5.14: HW for the case (i)

| $s$ | $n$ | MC | Copula | BR | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|---|
| 30 | 1021 | 3.53E-02 | 2.23E-02 | 2.17E-02 | 2.70E-02 | 1.63E-02 | 1.63E-02 |
| | 2039 | 2.43E-02 | 3.49E-02 | 1.09E-02 | 1.82E-02 | 1.96E-02 | 1.76E-02 |
| | 4093 | 1.73E-02 | 3.73E-02 | 6.71E-03 | 1.52E-02 | 5.47E-03 | 7.88E-03 |
| | 8191 | 1.21E-02 | 3.47E-03 | 6.40E-03 | 5.71E-03 | 7.46E-03 | 7.46E-03 |
| | 16381 | 8.63E-03 | 5.34E-03 | 2.39E-03 | 1.27E-02 | 3.69E-03 | 2.96E-03 |
| | 32749 | 6.33E-03 | 3.00E-03 | 4.33E-03 | 4.01E-03 | 3.30E-03 | 3.30E-03 |
| | 65521 | 4.40E-03 | 1.80E-03 | 2.05E-03 | 2.93E-03 | 3.12E-03 | 4.01E-03 |
| | 131071 | 3.09E-03 | 7.65E-04 | 6.75E-04 | 8.89E-04 | 7.55E-04 | 8.42E-04 |
| 100 | 1021 | 1.67E-01 | 2.38E-01 | 9.73E-02 | 1.64E-01 | 3.81E-01 | 3.81E-01 |
| | 2039 | 1.43E-01 | 3.07E-01 | 1.35E-01 | 1.99E-01 | 1.33E-01 | 2.18E-01 |
| | 4093 | 1.10E-01 | 1.23E-01 | 8.54E-02 | 1.19E-01 | 4.38E-02 | 1.15E-01 |
| | 8191 | 8.96E-02 | 5.45E-02 | 1.02E-01 | 9.47E-02 | 1.76E-01 | 1.76E-01 |
| | 16381 | 2.12E-01 | 6.20E-02 | 7.35E-02 | 9.68E-02 | 8.37E-02 | 4.14E-02 |
| | 32749 | 1.28E-01 | 2.16E-01 | 8.70E-02 | 7.18E-02 | 7.72E-02 | 7.72E-02 |
| | 65521 | 6.77E-02 | 8.39E-02 | 4.22E-02 | 3.52E-02 | 7.03E-02 | 6.82E-02 |
| | 131071 | 4.32E-02 | 4.51E-02 | 2.89E-02 | 2.48E-02 | 6.30E-02 | 4.58E-02 |
| 500 | 1021 | 3.53E-02 | 6.17E-03 | 4.62E-03 | 5.22E-01 | 1.52E-01 | 1.52E-01 |
| | 2039 | 2.97E-01 | 3.80E-03 | 4.30E-03 | 2.45E-01 | 7.06E-01 | 4.83E-02 |
| | 4093 | 1.48E-01 | 9.27E-03 | 6.37E-03 | 4.03E-02 | 2.72E-01 | 2.14E-01 |
| | 8191 | 7.62E-02 | 1.85E-01 | 1.26E-01 | 1.62E+00 | 3.51E-01 | 3.51E-01 |
| | 16381 | 3.88E-02 | 5.51E-02 | 8.68E-03 | 7.72E-03 | 5.15E-02 | 3.10E-01 |
| | 32749 | 6.75E-02 | 5.51E-02 | 1.08E+00 | 4.63E-01 | 8.38E-02 | 8.38E-02 |
| | 65521 | 3.64E-02 | 5.14E-01 | 5.41E+00 | 1.48E-01 | 2.75E-02 | 7.07E-02 |
| | 131071 | 3.18E-02 | 2.40E-01 | 1.01E-01 | 9.81E-02 | 3.51E-02 | 6.89E-02 |

Table 5.15: HW for the case (ii)

| $s$ | $n$ | MC | Copula | BR | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|---|
| 30 | 1021 | 7.68E-03 | 1.68E-05 | 3.73E-05 | 8.13E-06 | 1.14E-04 | 1.14E-04 |
| | 2039 | 5.43E-03 | 3.84E-05 | 1.07E-05 | 5.33E-06 | 7.94E-06 | 1.23E-05 |
| | 4093 | 3.83E-03 | 1.76E-03 | 2.43E-06 | 3.88E-06 | 2.48E-06 | 9.60E-07 |
| | 8191 | 2.71E-03 | 5.85E-07 | 6.03E-06 | 6.83E-07 | 1.65E-06 | 1.65E-06 |
| | 16381 | 1.92E-03 | 1.76E-05 | 1.65E-06 | 6.63E-07 | 2.31E-07 | 1.16E-06 |
| | 32749 | 1.36E-03 | 1.82E-07 | 2.69E-07 | 5.49E-07 | 1.91E-07 | 1.91E-07 |
| | 65521 | 9.60E-04 | 8.93E-07 | 1.49E-07 | 5.34E-07 | 3.89E-08 | 1.02E-07 |
| | 131071 | 6.79E-04 | 8.70E-07 | 2.92E-08 | 4.00E-08 | 3.07E-08 | 1.25E-08 |
| 100 | 1021 | 7.68E-03 | 1.69E-05 | 2.64E-05 | 8.05E-06 | 9.62E-05 | 9.62E-05 |
| | 2039 | 5.45E-03 | 3.19E-05 | 8.90E-06 | 6.50E-06 | 7.79E-06 | 1.14E-05 |
| | 4093 | 3.85E-03 | 1.74E-03 | 1.76E-06 | 3.80E-06 | 2.69E-06 | 8.60E-07 |
| | 8191 | 2.72E-03 | 3.89E-07 | 6.98E-06 | 6.12E-07 | 8.30E-07 | 8.30E-07 |
| | 16381 | 1.92E-03 | 2.49E-07 | 9.32E-08 | 6.25E-07 | 1.98E-07 | 1.08E-06 |
| | 32749 | 1.36E-03 | 1.89E-07 | 2.23E-07 | 5.87E-07 | 1.15E-07 | 1.15E-07 |
| | 65521 | 9.60E-04 | 9.52E-07 | 6.78E-08 | 4.73E-07 | 3.83E-08 | 1.10E-07 |
| | 131071 | 6.79E-04 | 7.87E-07 | 2.26E-08 | 2.92E-08 | 2.73E-08 | 1.19E-08 |
| 500 | 1021 | 7.65E-03 | 1.59E-05 | 3.66E-05 | 7.88E-06 | 1.03E-04 | 1.03E-04 |
| | 2039 | 5.43E-03 | 3.59E-05 | 8.79E-06 | 5.82E-06 | 7.01E-06 | 1.55E-05 |
| | 4093 | 3.84E-03 | 1.75E-03 | 1.98E-06 | 4.33E-06 | 3.22E-06 | 1.11E-06 |
| | 8191 | 2.71E-03 | 6.00E-07 | 5.78E-06 | 5.52E-07 | 1.37E-06 | 1.37E-06 |
| | 16381 | 1.92E-03 | 1.85E-05 | 1.95E-06 | 1.30E-06 | 2.52E-07 | 1.07E-06 |
| | 32749 | 1.36E-03 | 2.03E-07 | 2.50E-07 | 5.66E-07 | 2.57E-07 | 2.57E-07 |
| | 65521 | 9.60E-04 | 8.53E-07 | 1.53E-07 | 4.80E-07 | 6.65E-08 | 1.36E-07 |
| | 131071 | 6.79E-04 | 8.93E-07 | 3.44E-08 | 3.64E-08 | 3.00E-08 | 1.16E-08 |

Table 5.16: HW for the case (iii)

| $s$ | $n$ | MC | Copula | BR | Spectral | $\mathcal{P}_2$ | $\mathcal{R}_2$ |
|---|---|---|---|---|---|---|---|
| 30 | 1021 | 8.21E-03 | 1.30E-04 | 1.33E-03 | 2.96E-04 | 7.74E-04 | 7.74E-04 |
| | 2039 | 5.83E-03 | 1.83E-03 | 1.25E-04 | 2.75E-04 | 3.03E-04 | 2.85E-04 |
| | 4093 | 4.11E-03 | 2.76E-03 | 6.08E-05 | 9.46E-05 | 2.30E-05 | 4.44E-05 |
| | 8191 | 2.90E-03 | 8.63E-06 | 6.70E-05 | 1.83E-05 | 7.37E-05 | 7.37E-05 |
| | 16381 | 2.05E-03 | 1.20E-05 | 4.89E-06 | 2.84E-04 | 6.73E-05 | 1.21E-05 |
| | 32749 | 1.45E-03 | 3.83E-06 | 1.82E-05 | 1.77E-05 | 7.33E-06 | 7.33E-06 |
| | 65521 | 1.03E-03 | 5.96E-06 | 5.55E-06 | 5.62E-06 | 7.89E-06 | 1.44E-05 |
| | 131071 | 7.25E-04 | 1.09E-06 | 3.18E-06 | 4.88E-07 | 1.21E-06 | 3.08E-06 |
| 100 | 1021 | 8.32E-03 | 4.32E-04 | 1.56E-03 | 2.97E-04 | 6.35E-04 | 6.35E-04 |
| | 2039 | 5.89E-03 | 1.88E-03 | 1.55E-04 | 3.53E-04 | 3.20E-04 | 3.37E-04 |
| | 4093 | 4.15E-03 | 3.27E-03 | 8.44E-05 | 2.82E-04 | 7.11E-05 | 6.26E-05 |
| | 8191 | 2.93E-03 | 3.25E-05 | 6.51E-05 | 4.44E-05 | 7.52E-05 | 7.52E-05 |
| | 16381 | 2.07E-03 | 4.03E-05 | 2.88E-05 | 3.26E-04 | 7.93E-05 | 1.72E-05 |
| | 32749 | 1.47E-03 | 9.92E-06 | 6.76E-05 | 1.74E-05 | 1.27E-05 | 1.27E-05 |
| | 65521 | 1.04E-03 | 1.45E-05 | 8.37E-06 | 8.15E-06 | 1.22E-05 | 1.42E-05 |
| | 131071 | 7.33E-04 | 6.58E-06 | 3.77E-06 | 7.89E-06 | 3.19E-06 | 4.63E-06 |
| 500 | 1021 | 8.37E-03 | 5.58E-04 | 1.56E-03 | 4.42E-04 | 7.52E-04 | 7.52E-04 |
| | 2039 | 5.90E-03 | 1.99E-03 | 2.12E-04 | 3.32E-04 | 3.58E-04 | 3.10E-04 |
| | 4093 | 4.16E-03 | 2.91E-03 | 1.17E-04 | 3.01E-04 | 1.16E-04 | 7.23E-05 |
| | 8191 | 2.94E-03 | 7.00E-05 | 6.52E-05 | 5.80E-05 | 9.30E-05 | 9.30E-05 |
| | 16381 | 2.08E-03 | 4.13E-05 | 4.48E-05 | 2.78E-04 | 7.40E-05 | 3.40E-05 |
| | 32749 | 1.47E-03 | 1.04E-05 | 6.28E-05 | 2.43E-05 | 2.49E-05 | 2.49E-05 |
| | 65521 | 1.04E-03 | 1.93E-05 | 1.18E-05 | 2.72E-05 | 1.29E-05 | 1.98E-05 |
| | 131071 | 7.36E-04 | 7.17E-06 | 1.36E-05 | 9.76E-06 | 3.51E-06 | 7.32E-06 |

# Chapter 6

# Conclusion

We have analyzed bilinearity of functions using quasi-regression. We found that some functions are not very bilinear, but still the generators based on the copula-based criterion work well with those functions. Also, we found that the quasi-regression based criterion gives good generators. For further research, we would like to include higher order terms in the quasi-regression criterion. This will allow us to search for the generators that work well with functions that have significant quadratic components as well as linear and bilinear parts. Also, we would like to investigate the linearity of more complex finance models such as the Heston's model.

# Appendix

Proof of Proposition 4.2.1.

We first derive $\mathrm{Var}(\hat{\mu}_{mc,n})$ and $\sigma_{uv,f}$. Then we use the fact that

$$\mathrm{Var}(\hat{\mu}_{grid,n}) = \mathrm{Var}(\hat{\mu}_{mc,n}) + \frac{n-1}{n}\sigma_{uv,f} \tag{1}$$

as in [13]. We start by finding the expression for $\mathrm{Var}(f(\mathbf{u}))$.

$$\mathrm{Var}(f(\mathbf{u})) = \mathrm{Var}(\sum_{i=1}^{s} c_i u_i) + \mathrm{Var}(\sum_{i<j}^{s} d_{i,j} u_i u_j) + 2\mathrm{Cov}(\sum_{i=1}^{s} c_i u_i, \sum_{i<j}^{s} d_{i,j} u_i u_j) \tag{2}$$

We simplify the right hand side of (2)

$$\mathrm{Var}(\sum_{i=1}^{s} c_i u_i) = \sum_{i=1}^{s} {c_i}^2 \mathrm{Var}(u_i) = \frac{1}{12}\sum_{i=1}^{s} {c_i}^2 \tag{3}$$

$$\begin{aligned}\mathrm{Var}(\sum_{i<j}^{s} d_{i,j} u_i u_j) &= \sum_{i<j}^{s} \mathrm{Var}(d_{i,j} u_i u_j) + \sum_{(i,j)\neq(l,k)}^{s} \mathrm{Cov}(d_{i,j} u_i u_j, d_{k,l} u_k u_l) \\ &= (\frac{1}{9} - \frac{1}{16})\sum_{i<j}^{s} {d_{i,j}}^2 + \frac{1}{48}\sum_{k=1}^{s}\sum_{\substack{k\in G,H \\ G\neq H}} d_G d_H \end{aligned} \tag{4}$$

$$\begin{aligned}\mathrm{Cov}(\sum_{i=1}^{s} c_i u_i, \sum_{i<j}^{s} d_{i,j} u_i u_j) &= \sum_{i=1}^{s} c_i \mathrm{Cov}(u_i, \sum_{k<l}^{s} d_{k,l} u_k u_l) \\ &= \sum_{i=1}^{s} c_i \mathrm{Cov}(u_i, \sum_{i\in G} d_G u_{G_1} u_{G_2}) \\ &= \frac{1}{24}\sum_{i=1}^{s} c_i \sum_{i\in G} d_G \end{aligned} \tag{5}$$

Substituting (3), (4) and (5) into (2), we get

$$\mathrm{Var}(f(\mathbf{u})) = \frac{1}{12}\sum_{i=1}^{s} {c_i}^2 \frac{7}{144}\sum_{i<j}^{s} {d_{i,j}}^2 + \frac{1}{48}\sum_{k=1}^{s}\sum_{\substack{k\in G,H \\ G\neq H}} d_G d_H + \frac{1}{12}\sum_{i=1}^{s} c_i \sum_{i\in G} d_G. \tag{6}$$

Since $\mathrm{Var}(\hat{\mu}_{mc,n}) = \frac{1}{n}\mathrm{Var}(f(\mathbf{u}))$,

$$\mathrm{Var}(\hat{\mu}_{mc,n}) = \frac{1}{12n}\sum_{i=1}^{s} {c_i}^2 + \frac{7}{144n}\sum_{i<j}^{s} {d_{i,j}}^2 + \frac{1}{48n}\sum_{k=1}^{s}\sum_{\substack{k\in G,H \\ G\neq H}} d_G d_H + \frac{1}{12n}\sum_{i=1}^{s} c_i \sum_{i\in G} d_G \tag{7}$$

Next, we calculate $\sigma_{uv,f}$. From Proposition 5.1 in [13], we have

$$\begin{aligned}\sigma_{uv,f} &= \sum_{\substack{K\in I \\ |K|\leq 2}} r_K \rho_{lr,K} = \sum_{\substack{K\in I \\ |K|=1}} r_K \rho_{lr,K} + \sum_{\substack{K\in I \\ |K|=2}} r_K \rho_{lr,K} \\ &= \sum_{\substack{K\in I \\ |K|=1}} \frac{1}{12} t_K \rho_{lr,K} + \sum_{\substack{K\in I \\ |K|=2}} \frac{11}{80} t_K \rho_{lr,K} = -\frac{1}{12n}\sum_{k=1}^{s} t_{\{k\}} + \frac{11}{80}\sum_{k<l} t_{\{k,l\}}\rho_{lr,\{k,l\}},\end{aligned}$$

where $t_K = \sum_{I,J\subseteq S: I\cap J = K} (-1)^{|I|+|J|} 2^{-l} \sum_{I\in G} d_G \sum_{J\in H} d_H$ and $l = |I| + |J| - 2|K|$.
The third equality follows since $\rho_{lr,K} = -\frac{1}{n}$ for $|K| = 1$.

Start with looking at the first sum, that is, when $|K| = 1$.
For each $K = \{k\}$ , $1 \leq k \leq s$, we consider following three cases of $I$ and $J$ to derive $t_K$
For each case, we take sum over all combination of $I$ and $J$ and combine the three at the end.

Case 1: $I = \{k\}$, $J = \{k\}$.

Inside the sum is
$(-1)^{1+1}2^0(c_k + \sum_{k\in G} d_G)^2 = (c_k + \sum_{k\in G} d_G)^2$.

Summing over k, we get

$$\sum_{k=1}^{s}(c_k+\sum_{k\in G}d_G)^2=\sum_{k=1}^{s}(c_k+\sum_{k\in G}d_G)(c_k+\sum_{k\in G}d_G)$$
$$=\sum_{k=1}^{s}(c_k+\sum_{k\in G}d_G)(c_k+\sum_{k\in G}d_G) \tag{8}$$
$$=\sum_{k=1}^{s}(c_k+\sum_{k\in G}d_G)c_k+\sum_{k=1}^{s}(c_k+\sum_{k\in G}d_G)\sum_{k\in G}d_G. \tag{9}$$

Case 2: I = $\{k\}$, J = $\{k,l\}$ or I = $\{k,l\}$, J = $\{k\}$.

Inside the sum is $(-1)^{1+2}2^{-1}d_{k,l}(c_k+\sum_{k\in G}d_G)=-\frac{1}{2}d_{k,l}d_{k,l}(c_k+\sum_{k\in G}d_G)$.
Summing over all combination of I = $\{k\}$, J = $\{k,l\}$ and multiply by 2 to account for $I=\{k,l\}$, $J=\{k\}$ case, we get

$$2\sum_{k=1}^{s}\sum_{k\neq l}(-\frac{1}{2})d_{k,l}(c_k+\sum_{k\in G}d_G)=-\sum_{k=1}^{s}(c_k+\sum_{k\in G}d_G)\sum_{k\in G}d_G. \tag{10}$$

Case 3: $I=\{k,l\}$, $J=\{k,q\}$.

Inside the sum is $(-1)^2 2^{-2}d_{k,l}d_{kq}=\frac{1}{4}d_{k,l}d_{kq}$.
Summing over all $I=\{k,l\}$, $J=\{k,q\}$ , we get

$$\frac{1}{4}\sum_{k=1}^{s}\sum_{\substack{k\in G,H\\ G\neq H}}d_G d_H. \tag{11}$$

Summing (9),(10) and (11), we get

$$\sum_{k=1}^{s}t_{\{k\}}=\sum_{k=1}^{s}(c_k^2+c_k\sum_{k\in G}d_G)+\frac{1}{4}\sum_{k=1}^{s}\sum_{\substack{k\in G,H\\ G\neq H}}d_G d_H. \tag{12}$$

With (8) and (12),

$$\sum_{\substack{K\in I\\ |K|=1}}r_K\rho_{lr,K}=\frac{1}{12n}\sum_{k=1}^{s}(c_k^2+c_k\sum_{k\in G}d_G)+\frac{1}{48n}\sum_{k=1}^{s}\sum_{\substack{k\in G,H\\ G\neq H}}d_G d_H. \tag{13}$$

Next, look at the second sum, that is, $|K| = 2$.
Suppose $K = \{k, l\}$ for some $1 \leq k < l \leq s$.
This time we have only one case of $I$ and $J$ to consider when deriving $t_{k,l}$, namely $I = \{k, l\}$, $J = \{k, l\}$.
So, $t_{k,l} = (-1)^{2+2} d_{k,l}^2 = d_{k,l}^2$.
With (8),

$$\begin{aligned} \sum_{\substack{K \in I \\ |K|=2}} r_K \rho_{lr,K} &= \frac{11}{80} \sum_{k<l} d_{\{k,l\}}^2 \rho_{lr,\{k,l\}} \\ &= \frac{11}{80} \sum_{k=1}^{s} \sum_{k<l} d_{\{k,l\}}^2 \rho_{lr,\{k,l\}}. \end{aligned} \tag{14}$$

Combining (8),(13), and (14), we get

$$\sigma_{uv,f} = \frac{1}{12n} \sum_{k=1}^{s} c_k^2 + \frac{1}{12n} \sum_{k=1}^{s} c_k \sum_{k \in G} d_G + \frac{1}{48n} \sum_{k=1}^{s} \sum_{\substack{k \in G, H \\ G \neq H}} d_G d_H + \frac{11}{80} \sum_{k=1}^{s} \sum_{k<l} d_{\{k,l\}}^2 \rho_{lr,\{k,l\}}. \tag{15}$$

Combining (1), (7) and (15), we get

$$\mathrm{Var}(\hat{\mu}_{grid,n}) = \frac{1}{12n^2} \sum_{i=1}^{s} {c_i}^2 + \frac{1}{48n^2} \sum_{k=1}^{s} \sum_{\substack{k \in G, H \\ G \neq H}} d_G d_H + \frac{1}{12n^2} \sum_{i=1}^{s} c_i \sum_{i \in G} d_G + \sum_{i<j}^{s} \left( \frac{7}{144n} + \frac{11}{80} \rho_{lr,\{i,j\}} \right) {d_{i,j}}^2. \tag{16}$$

# References


[1] Jian An and Art Owen. Quasi-regression. *Journal of Complexity*, 17(4):588–607, 2001.

[2] Russel E Caflisch, William J Morokoff, and Art B Owen. *Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension*. Department of Mathematics, University of California, Los Angeles, 1997.

[3] RR Coveyou and Robert D MacPherson. Fourier analysis of uniform random number generators. *Journal of the ACM (JACM)*, 14(1):100–119, 1967.

[4] Bradley Efron and Charles Stein. The jackknife estimate of variance. *The Annals of Statistics*, pages 586–596, 1981.

[5] Hardeep S Gill and Christiane Lemieux. Searching for extensible korobov rules. *Journal of Complexity*, 23(4):603–613, 2007.

[6] JM Hammersley and KW Morton. A new monte carlo technique: antithetic variates. In *Mathematical proceedings of the Cambridge philosophical society*, volume 52, pages 449–475. Cambridge Univ Press, 1956.

[7] Edmund Hlawka. Funktionen von beschränkter variation in der theorie der gleichverteilung. *Annali di Matematica Pura ed Applicata*, 54(1):325–333, 1961.

[8] John Hull. *Options, Futures, and Other Derivatives*. Pearson Education India, 7 edition, 2010.

[9] P L'Ecuyer and D Munger. A general software tool for constructing rank-1 lattice rules, 2012.

[10] P L'Ecuyer and D Munger. Latticebuilder: A general software tool for constructing rank-1 lattice rules, 2012.

[11] Pierre L'Ecuyer and Christiane Lemieux. Variance reduction via lattice rules. *Management Science*, 46(9):1214–1235, 2000.

[12] Erich Leo Lehmann. Some concepts of dependence. *The Annals of Mathematical Statistics*, pages 1137–1153, 1966.

[13] Christiane Lemieux. Dependence concepts for randomized quasi-monte carlo methods. preprint, `https://math.uwaterloo.ca/statistics-and-actuarial-science/sites/ca.statistics-and-actuarial-science/files/uploads/files/Lemieux-2013-04.pdf`.

[14] Christiane Lemieux. *Monte Carlo and Quasi-Monte-Carlo Sampling*. Springer, 2009.

[15] Christiane Lemieux and Art B Owen. Quasi-regression and the relative importance of the anova components of a function. In *Monte Carlo and Quasi-Monte Carlo Methods 2000*, pages 331–344. Springer, 2002.

[16] David V Lindley. The theory of queues with a single server. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 48, pages 277–289. Cambridge University Press, 1952.

[17] Xuerong Mao. *Stochastic differential equations and their applications*. Horwood publishing, 1997.

[18] Michael D McKay, Richard J Beckman, and William J Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[19] Roger B Nelsen. *An introduction to copulas*. Springer, 2006.

[20] Brian D Ripley. The lattice structure of pseudo-random number generators. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 389(1796):197–204, 1983.

[21] IH Sloan and TR Osborn. Multiple integration over bounded and unbounded regions. *Journal of computational and applied mathematics*, 17(1):181–196, 1987.

[22] Arthur H Stroud. *Approximate calculation of multiple integrals*, volume 431. Prentice-Hall Englewood Cliffs, NJ, 1971.

[23] Xiaoqun Wang and Fred J Hickernell. Randomized halton sequences. *Mathematical and Computer Modelling*, 32(7):887–899, 2000.

[24] SK Zaremba. Some applications of multidimensional integration by parts. *Ann. Polon. Math*, 21:85–96, 1968.