ARTICLE IN PRESS

Journal of Multivariate Analysis xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Three kinds of discrete approximations of statistical multivariate distributions and their applications

Jun Yang [a], Ping He [a], Kai-Tai Fang [a,b,*]

[a] *Division of Science and Technology, BNU-HKBU United International College, Zhuhai, China*

[b] *The Key Lab of Random Complex Structures and Data Analysis, The Chinese Academy of Sciences, Beijing, China*

A R T I C L E I N F O

Article history:
Received 13 August 2021
Received in revised form 28 August 2021
Accepted 28 August 2021
Available online xxxx

AMS 2020 subject classifications:
primary 62H12
secondary 62F12

Keywords:
Bootstrap
k-means
Quasi-Monte Carlo method
Representative points
Resampling
Statistical simulation

A B S T R A C T

Discrete approximations of statistical continuous distributions have been widely requested in various fields. Using random samples generated by Monte Carlo (MC) method to infer the population has been dominant in statistics. The empirical distribution of a random sample can be regarded as a discrete approximation of the population distribution in a certain statistical sense. However, MC has a poor performance in many problems. This paper concerns some alternative methods, such as Quasi-Monte Carlo (QMC) F-numbers and Mean Square Error Representative Points (MSE-RPs), and constructs approximation distributions for elliptically contoured distributions and skew-normal distributions. Numerical comparisons are given for two geometric probability problems and for estimation accuracy by resampling from the discrete approximation distributions obtained by MC, QMC and MSE-RPs. Our simulation results indicate that QMC and MSE-RPs have better performance in most comparisons. These results show that QMC and MSE-RPs have high potential in statistical inference. In addition, we also discuss the relationship between principal component analysis and MSE-RPs for elliptically contoured distributions, as well as its potential applications.

1. Introduction

Discrete approximations of statistical continuous distribution have been widely used in various situations. For example, we often concern various characteristics of the continuous distribution being studied, however, sometimes integration techniques may fail to provide analytic solutions due to the complicated distribution function, especially for multivariate continuous distributions. Let $\mathbf{x}$ be a d-variate random vector following a continuous distribution $F(\mathbf{x})$. Monte Carlo (MC) method provides ways to generate a random sample $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ from the population, and its empirical distribution is given by

$$F_k(\mathbf{x}) = \frac{1}{k}\sum_{j=1}^{k} I\{\mathbf{x}_j \le \mathbf{x}\}, \tag{1}$$

where $I\{A\}$ is the indicator function of event A. Under some regular assumption, the weak law of large numbers implies that $F_k(\mathbf{x})$ converges to $F(\mathbf{x})$ in probability as $k \to \infty$. Hence, $F_k(\mathbf{x})$ can be considered as an estimate of the population distribution $F(\mathbf{x})$, and the characteristics of $F(\mathbf{x})$ can be estimated by the random sample. The MC method has a lot of

* Corresponding author at: Division of Science and Technology, BNU-HKBU United International College, Zhuhai, China.
E-mail address: ktfang@uic.edu.cn (K.-T. Fang).

https://doi.org/10.1016/j.jmva.2021.104829
0047-259X/

Please cite this article as: J. Yang, P. He and K.-T. Fang, Three kinds of discrete approximations of statistical multivariate distributions and their applications, Journal of Multivariate Analysis (2021) 104829, https://doi.org/10.1016/j.jmva.2021.104829.

advantages in estimation and performing statistical tests. However, it sometimes has poor performance. For example, the convergence rate for a numerical integration by the MC method is $O(k^{-1/2})$ independent of dimension d, which is slow. Therefore, some alternative approaches including number-theoretic methods or Quasi-Monte Carlo (QMC) methods [32] are proposed. The convergence rate for a numerical integration by the QMC methods can be reached to $O((\log k)^d/k)$, which is faster than that of the MC method if d is not large, and a set of points of size k generated by the QMC methods is called quasi-random F-numbers (or QMC F-numbers) with respect to $F(\mathbf{x})$ (see Definition 2 of the paper). Another way to generate a set of points which can represent $F(\mathbf{x})$ well is the Mean Square Error (MSE) methods, which provides a set of MSE representative points of $F(\mathbf{x})$ by minimizing the MSE (see Definition 3 of the paper).

The bootstrap method proposed by [8] is one of the resampling techniques. The idea of bootstrap is sampling from the empirical distribution based on a random sample, which can be regarded as an approximation of the distribution $F(\mathbf{x})$. Hence, a natural question is whether sampling from a better discrete approximation of $F(\mathbf{x})$ can provide more accurate estimations? Fang et al. [14] first employed resampling techniques to the approximations of the univariate normal distribution constructed by Quasi-Monte Carlo methods (QMC) and MSE representative points, and the results indicated a significant improvement in estimation of the mean, variance, skewness and kurtosis. In this paper, one of the purposes is to compare the performance of resampling techniques based on discrete approximations in statistical inference for the elliptically contoured distribution and the skew-normal distribution, where the discrete approximations are constructed by MC, QMC and MSE methods, respectively.

Discrete approximation of a probability distribution typically is determined by a set of representative points and associated probabilities. Suppose a discrete approximation of $F(\mathbf{x})$ with k points is denoted as $\hat{F}_k(\mathbf{x})$. There are various ways to measure the closeness between $\hat{F}_k(\mathbf{x})$ and $F(\mathbf{x})$, such as consistency, L_p-distance and the mean square error (MSE) which will be introduced in (18), etc. Here, L_p-distance between $\hat{F}_k(\mathbf{x})$ and $F(\mathbf{x})$ is defined as

$$D_p(\hat{F}_k, F) = \left[\int_{-\infty}^{\infty} |\hat{F}_k(\mathbf{x}) - F(\mathbf{x})|^p d\mathbf{x}\right]^{1/p}, \tag{2}$$

where $p \geq 1$. The L_∞-distance also known as F-discrepancy, L_1-distance and L_2-distance are commonly used to measure the closeness between $\hat{F}_k(\mathbf{x})$ and $F(\mathbf{x})$. Therefore, finding an optimal discrete approximation can be considered as a problem of selecting a given number of points that retain as much information of the population as possible under some appropriate criterion, and such a set of points are also called representative points (RPs). In other words, the set of RPs of $\mathbf{x}$ is chosen by minimizing the bias between $\hat{F}_k(\mathbf{x})$ and $F(\mathbf{x})$. In the following discussion, we mainly focus on the RPs obtained by minimizing F-discrepancy and MSE.

Number-theoretic methods or Quasi-Monte Carlo (QMC) methods [32] are proposed as the alternative approaches to the MC method. A set of points generated by the QMC methods can be regarded as RPs of the continuous distribution $F(\mathbf{x})$ (QMC-RPs for short) which has greater uniformity than a set of random points generated by the MC method (MC-RPs for short). It is known that the QMC-RPs are constructed under the criterion F-discrepancy which is often used to measure the closeness of the set of points, so QMC-RPs have better performance than MC-RPs. Using the QMC methods, the discrete approximation of the continuous distribution $F(\mathbf{x})$ is constructed by a set of k QMC-RPs as the support points and the corresponding probability at each point $p_j = 1/k$. The QMC methods have been widely applied in various areas of statistics [12,13,33], including tests for multi-normality, multivariate goodness-of-fit, Markov chain Monte Carlo (MCMC), permutation tests, and robust regression model fitting, etc., and they are also utilized to solve financial problems [18,22,36,48] and partial differential equations [20,21].

Another important criterion for representativeness is MSE. Cox [5] discussed the problem of classifying the observations of a univariate continuous distribution into k groups and assigning the individuals in the jth group with an associated value $\xi_j, j \in \{1, \ldots, k\}$, where the MSE is proposed to measure the "representativeness" of the selected points to the continuous distribution. In engineering, the problem of finding RPs is described as a problem of optimal quantization. The optimal quantization is originally applied to optimize signal transmission by appropriate discretization procedures [34]. Max [31] discussed the problem of minimizing the distortion of a signal by a quantizer with a fixed number of output levels. The distortion is defined as the expected value of squared error between the input and output of the quantizer, so the quantizer with k output levels having a minimized distortion of a signal is exactly the same as a set of RPs that minimizing the MSE (MSE-RPs for short). For more applications of MSE-RPs (or the optimal vector quantization) in the field of signal and image processing and information theory, one can see in [17,19]. In spite of this, Fang and He [10] proposed a computational procedure for selecting RPs with minimized MSE to develop clothing standards. Flury [15,16] studied the MSE-RPs of elliptically contoured distribution with the motivation of determining optimal sizes and shapes of protection masks for the Swiss Army. Flury and his co-authors used the term "principal points" for MSE-RPs due to the relationship between MSE-RPs and principal components (see [15,42]). In addition, the MSE-RPs can be utilized in problems for numerical computation of conditional expectations, stochastic differential equations and stochastic partial differential equations, which are mostly motivated by problems arising in finance (e.g., option pricing) [4,9,34,35], and they are also applied for classification to distinguish drug responders from placebo responders in psychiatric studies [44,45]. Fang et al. [14] suggested to apply the MSE-RPs in statistical inference by resampling technique. Recently, Lemaire et al. [27] proposed a new variance reduction technique for Monte Carlo estimators with one dimensional MSE-RPs. In the following discussion, the discrete approximation distributions of $F(\mathbf{x})$ constructed with MC-RPs, QMC-RPs and MSE-RPs are denoted as $\hat{F}_{\mathrm{MC}}(\mathbf{x})$, $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$, and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$, respectively.

Although the three types of discrete approximations, $\hat{F}_{\mathrm{MC}}(\mathbf{x})$, $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ have been widely applied to various fields, to our knowledge, there are few studies comparing their performance in statistical inference, especially for multivariate distributions. Therefore, it is meaningful to compare the performance of these discrete approximation methods in statistical inference for multivariate distributions, and the research results can provide guidance for the choice of discrete methods in practical problems. In addition, Fang et al. [14] proposed to resample from these three kinds of discrete approximations of univariate normal distribution using the idea of bootstrap, and results showed that $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ has superior performance in the estimation accuracy among the three kinds of approximations. In this paper, we will focus on the comparison of the estimation accuracy of these three discrete approximations for some multivariate distributions, including the elliptically contoured distributions and the skew-normal distribution.

Algorithms which can generate the QMC-RPs and MSE-RPs accurately and effectively are very important in this study. Fang and Wang [12] provided an efficient algorithm to compute the QMC-RPs of spherical distributions. For generating a set of MSE-RPs, k-means algorithm is important and widely used. For a random sample of $F(\mathbf{x})$ and a given k, Pollard [37] showed that the set of optimal sample cluster means converge to the optimal population cluster means almost surely under some regularity conditions when the sample size goes to infinity. Hence, the cluster means of k-means algorithm are self-consistent points for the empirical distribution of $F(\mathbf{x})$ [40], which gives the nonparametric estimators of MSE-RPs of the underlying distribution based on the self-consistency of MSE-RPs [41]. However, the training set generated from the underlying distribution $F(\mathbf{x})$ and the set of initial points selected are also crucial for the k-means algorithm. From this point of view, the estimation of MSE-RPs can be improved with a better training set and initial points. Fang and Wang [12] provided an effective algorithm by means of the QMC methods, which is the NTLBG algorithm, and they showed the details for manipulation of the method. In this paper, besides the NTLBG algorithm, we also make an attempt to improve the accuracy of MSE-RPs by employing the k-means++ seeding method for selecting initial points [2] instead of the QMC methods, as it can be more competitive than the QMC methods in some cases with a QMC training set.

Another topic of interest is that MSE-RPs relate to principal components. Several studies have attempted to determine the geometric pattern of MSE-RPs for the elliptically contoured distributions (see [15,16,24,39]). The simplest pattern formed by MSE-RPs of the elliptically contoured distributions is a line, and this line is spanned by the first principal component based on the principal subspace theorem. Tarpey [39] provided a lower bound $\sigma_0(k)$ of the value σ in a covariance matrix of the form $\mathbf{\Sigma}_2(\sigma) = \mathrm{diag}(\sigma^2, 1)$ such that the MSE-RPs of a bivariate normal distribution with this covariance matrix form a line if $\sigma > \sigma_0(k)$, where $\sigma_0(k)$ is related to the number of MSE-RPs k. Inspired by this numerical result, we consider the covariance matrix with another special structure $\mathbf{\Sigma}_2(\rho) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. For a bivariate normal distribution with $\mathbf{\Sigma} = \mathbf{\Sigma}_2(\rho)$ there also exists a lower bound $\rho_0(k)$ for ρ such that all the MSE-RPs locate on the direction of the first principal component if $\rho \geq \rho_0(k)$. For the other subclasses of the elliptically contoured distribution of dimension $d = 2$ such as the Kotz type distribution, we have the similar findings with the bivariate normal distribution, and the numerical results for the Kotz type distribution are also presented. Due to the relationship between MSE-RPs of the elliptically contoured distribution and principal components, the contribution of the first principal component C_1 will also be discussed. For given k, we observe that the covariance matrices $\mathbf{\Sigma}_2(\sigma_0(k))$ and $\mathbf{\Sigma}_2(\rho_0(k))$ have the same value of C_1, and the k MSE-RPs of the elliptically contoured distribution with these two covariance matrices are located in the direction of the first principal component. This may provide a potential application of MSE-RPs for principal component analysis, and the discussion is in Section 5. In addition, we will also discuss the relationship between the magnitude of MSE and contribution of the first principal component.

The rest of this paper is organized as follows. In Section 2, the elliptically contoured distributions and the algorithms for generating different types of RPs will be reviewed. The applications of discrete approximations $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ in two geometric probability problems will be considered in Section 3. Estimation of the mean vector and covariance matrix for the elliptically contoured distribution by resampling from $\hat{F}_{\mathrm{MC}}(\mathbf{x})$, $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ will be investigated and numerical results will be shown in Section 4. We will discuss the relationship between the MSE-RPs and principal component analysis in Section 5. Finally, the conclusions and further study will be considered in Section 6.

2. Preliminary

In this section, we will briefly introduce the elliptically contoured distributions (ECDs) [11], skew-normal distribution [1,3] and the algorithm for generating QMC-RPs and the MSE-RPs [12,13].

2.1. Elliptically contoured distributions

Let $\mathbf{x} \in \mathbb{R}^d$ be a d-variate random vector of elliptically contoured distribution (ECD), denoted as $\mathbf{x} \sim \mathrm{ECD}_d(\boldsymbol{\mu}, \mathbf{\Psi}; g)$, where g is the density generator. The elliptically contoured distribution is originally defined with the density $p(\mathbf{x})$ as

$$p(\mathbf{x}) = |\mathbf{\Psi}|^{-1/2} g\left((\mathbf{x}-\boldsymbol{\mu})^\top \mathbf{\Psi}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \tag{3}$$

where $\mathbf{\Psi}$ is a positive definite matrix of order d, $\boldsymbol{\mu} = \mathrm{E}(\mathbf{x})$ is the mean vector, and the covariance matrix can be expressed as $\mathbf{\Sigma} = c\mathbf{\Psi}$ with a constant $c > 0$. We may see that the density of $\mathbf{x}$ is related to the corresponding function g for each subclass. However, some elliptically contoured distributions may not have a density, such as the uniform distribution on the surface of a sphere. In general, the elliptically contoured distribution is defined by the stochastic representation

$$\mathbf{x} \stackrel{d}{=} \boldsymbol{\mu} + R\boldsymbol{\Psi}^{1/2}\mathbf{u}^{(d)}, \tag{4}$$

where $\stackrel{d}{=}$ means that the two sides have identical distributions, and $\boldsymbol{\Psi}^{1/2}$ is the positive definite square root of $\boldsymbol{\Psi}$. In the stochastic representation (4), $\mathbf{u}^{(d)}$ is uniformly distributed on the unit sphere surface of $\mathbb{R}^d$, and the random variable $R > 0$ is referred to as the generating variate of $\mathbf{x}$, independent of $\mathbf{u}^{(d)}$. When $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\Psi} = \mathbf{I}_d$, then

$$\mathbf{y} \stackrel{d}{=} R\mathbf{u}^{(d)} \tag{5}$$

has a spherical distribution, denoted as $\mathbf{y} \sim \mathrm{S}_d(g)$. A d-variate random vector $\mathbf{y}$ is said to have a spherical distribution if $\mathbf{y} \stackrel{d}{=} \mathbf{P}\mathbf{y}$ for every orthogonal matrix $\mathbf{P}$, and (5) is the stochastic representation of $\mathbf{y}$. The relationship between $\mathbf{x}$ and $\mathbf{y}$ is

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Psi}^{1/2}\mathbf{y}. \tag{6}$$

It has been proved that $\mathbf{y}$ has a density generator g if and only if R has a density. Let $f_R(r)$ be the density of the random variable R, and the relationship between g and f_R is

$$f_R(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} r^{d-1} g(r^2). \tag{7}$$

Therefore, the stochastic representation (6) of $\mathbf{x}$ is determined by the distribution of the random variable R for different subclasses of ECDs. The following three subclasses are useful for our discussion in the next two sections:

1. uniform distribution on a unit sphere: when a random vector $\mathbf{y}$ is uniformly distributed on a unit sphere $\mathbf{S}^d$ in $\mathbb{R}^d$, the random variable R in (5) has a beta distribution beta$(d, 1)$;
2. multivariate normal distribution: when a random vector $\mathbf{y} \sim \mathrm{N}_d(\mathbf{0}, \mathbf{I}_d)$, the random variable R in (5) has a χ distribution with d degrees of freedom, then $\mathbf{x} \sim \mathrm{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ can be obtained by (4) with $R \sim \chi_d$ and $\boldsymbol{\Psi} = \boldsymbol{\Sigma}$;
3. Kotz type distribution: when a random vector $\mathbf{y} \sim \mathrm{Kotz}_d(N, b, s)$, a Kotz type distribution in $\mathbb{R}^d$ with parameters $N, b > 0$, $s > 0$ and $2N + d > 2$, the random variable $R' = R^{2s}$ has a gamma distribution with parameters $(2N + d - 2)/(2s)$ and $1/b$, i.e., $R' \sim \mathrm{gamma}((2N + d - 2)/(2s), b)$, and R can be obtained by $R = (R')^{1/(2s)}$.

For more information about some other subclasses of d-dimensional ECDs, it can be found in Chapter 3 of [11].

2.2. Skew-normal distribution

The skew-normal distribution can be applied to approximate the distribution of real data in many situations. Suppose $\mathbf{x}$ be a d-variate random vector of skew-normal distribution with location vector $\mathbf{a} \in \mathbb{R}^d$, dispersion matrix $\mathbf{G} \in \mathbb{R}^{d\times d}$ and skewness vector $\boldsymbol{\eta} \in \mathbb{R}^d$ where $\mathbf{G}$ is positive definite, denoted as $\mathbf{x} \sim \mathrm{SN}_d(\mathbf{a}, \mathbf{G}, \boldsymbol{\eta})$. The density of $\mathbf{x}$ is given by

$$p(\mathbf{x}) = 2\phi_d(\mathbf{x}|\mathbf{a}, \mathbf{G})\Phi_1(\boldsymbol{\eta}^\top \mathbf{G}^{-1/2}(\mathbf{x} - \mathbf{a})|\mathbf{a}, \mathbf{G}), \tag{8}$$

where $\phi_d(\mathbf{x}|\mathbf{a}, \mathbf{G})$ and $\Phi_d(\mathbf{x}|\mathbf{a}, \mathbf{G})$ are the density function and distribution function of $\mathrm{N}_d(\mathbf{a}, \mathbf{G})$, respectively. If $\mathbf{a} = \mathbf{0}$ and $\mathbf{G} = \mathbf{I}_d$, then $\mathrm{SN}_d(\mathbf{a}, \mathbf{G}, \boldsymbol{\eta})$ is denoted as $\mathrm{SN}_d(\boldsymbol{\eta})$. Let $\mathbf{w} \sim \mathrm{SN}_d(\boldsymbol{\eta})$, the stochastic representation of $\mathbf{w}$ is given by

$$\mathbf{w} \stackrel{d}{=} \boldsymbol{\delta}|z_0| + (\mathbf{I}_d - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}\mathbf{z}_1, \tag{9}$$

where $\boldsymbol{\delta} = \boldsymbol{\eta}/\sqrt{1 + \boldsymbol{\eta}^\top\boldsymbol{\eta}}$, and $z_0 \sim \mathrm{N}(0, 1)$ independent of $\mathbf{z}_1 \sim \mathrm{N}_d(\mathbf{0}, \mathbf{I}_d)$. For $\mathbf{x} \sim \mathrm{SN}_d(\mathbf{a}, \mathbf{G}, \boldsymbol{\eta})$, the relation between $\mathbf{w} \sim \mathrm{SN}_d(\boldsymbol{\eta})$ and $\mathbf{x}$ is

$$\mathbf{x} \stackrel{d}{=} \mathbf{a} + \mathbf{G}^{1/2}\mathbf{w}, \tag{10}$$

and the mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{x}$ are

$$\boldsymbol{\mu} = \mathbf{a} + \sqrt{\frac{2}{\pi}}\mathbf{G}^{1/2}\boldsymbol{\eta}, \qquad \boldsymbol{\Sigma} = \mathbf{G} - \frac{2}{\pi}\mathbf{G}^{1/2}\boldsymbol{\eta}\boldsymbol{\eta}^\top\mathbf{G}^{1/2}, \tag{11}$$

respectively.

2.3. Quasi-Monte Carlo methods and QMC-RPs

As we mentioned before in Section 1, QMC-RPs are a set of points with low discrepancy, where the discrepancy is a measurement of uniformity of the set of points. The F-discrepancy proposed by Wang and Fang [47] is defined as follows.

Definition 1. Let $\mathcal{P} = \{\mathbf{x}_j \in \mathbb{R}^d, j \in \{1, \ldots, k\}\}$ be a set of points, then

$$D_F(k, \mathcal{P}) = \sup_{\mathbf{x}\in\mathbb{R}^d} |F_k(\mathbf{x}) - F(\mathbf{x})| \tag{12}$$

is the F-discrepancy of $\mathcal{P}$ with respect to $F(\mathbf{x})$.

For a set of points $\mathcal{P}^* = \{\mathbf{x}_j^*, j \in \{1, \ldots, k\}\}$, if

$$D_F(k, \mathcal{P}^*) = \inf_{\mathcal{P}_k} D_F(k, \mathcal{P}),$$

where the set $\mathcal{P}_k$ consists of all sets of k points $\mathcal{P}$ in $\mathbb{R}^d$, then $\mathcal{P}^*$ is said to be the set of QMC-RPs (or quasi-random F-numbers in [12]) with respect to $F(\mathbf{x})$. For univariate case, let $F^{-1}(x)$ be the inverse function of the continuous distribution function $F(x)$, then the set of k optimal QMC-RPs are given by $\{F^{-1}((2j-1)/(2k)), j \in \{1, \ldots, k\}\}$ with F-discrepancy $1/(2k)$. However, for multivariate cases ($d \geq 2$), it is difficult to find the optimal set of QMC-RPs with the lowest F-discrepancy, so a set of points with F-discrepancy that is near to the lowest one can be regarded as a set of QMC-RPs with respect to $F(\mathbf{x})$.

Definition 2. Let $\mathcal{P}_k$ be a sequence of sets of points under a certain structure in $\mathbb{R}^d$ and $F(\mathbf{x})$ be a distribution function. The points of $\mathcal{P}_k$ are called quasi-random F-numbers if $D_F(k, \mathcal{P}_k) = O((\log k)^d/k)$ as $k \to \infty$, and $\mathcal{P}_k$ is called quasi-random F-numbers sequence.

The points of $\mathcal{P}_k$ are referred to as quasi-random numbers when $F(\mathbf{x})$ is the uniform distribution on a d-dimensional unit cube $\mathbf{C}^d$, denoted by $U(\mathbf{C}^d)$. A set of quasi-random numbers can be produced by several methods, such as the good point (*gp*) method, the good lattice point (*glp*) method, the Halton method and (t, s)-sequence [12]. However, there is no common method for generating quasi-random F-numbers, so we will apply the following method which has good properties (see [47]) and can be employed for most multivariate distributions. Let $\mathbf{c} \sim \mathrm{U}(\mathbf{C}^t)$, $t \leq d$ with its set of QMC-RPs $\{\mathbf{c}_j, j \in \{1, \ldots, k\}\}$, and $\mathbf{h}$ be a continuous function on $\mathbf{C}^t$. Suppose $\mathbf{x} \in \mathbb{R}^d$ is a random vector of a continuous distribution $F(\mathbf{x})$ with stochastic representation

$$\mathbf{x} = \mathbf{h}(\mathbf{c}), \tag{13}$$

then, the F-discrepancy of $\{\mathbf{x}_j = \mathbf{h}(\mathbf{c}_j), j \in \{1, \ldots, k\}\}$ is the same as the discrepancy of $\{\mathbf{c}_j\}$, which implies that $\{\mathbf{x}_j\}$ is a set of QMC-RPs with respect to $F(\mathbf{x})$. Tashiro [46] proposed an efficient algorithm for generating random points of the uniform distribution on the surface of unit sphere. This algorithm can be applied to generate quasi-random F-numbers of the uniform distribution on the surface of a unit sphere (see TFWW algorithm in [12] for details). Combining the TFWW algorithm and stochastic representation (5), an NTSR algorithm produces the QMC-RPs of $\mathbf{s} \sim \mathrm{S}_d(g)$ with the following steps [12].

Step 1. Generate a set of quasi-random numbers $\{\mathbf{c}_j = (c_{j1}, \ldots, c_{jd}), j \in \{1, \ldots, k\}\}$ on a d-dimensional unit cube $\mathbf{C}^d$.
Step 2. Let $F_R(r)$ be the distribution function of R and F_R^{-1} be its inverse function. Compute $r_j = F_R^{-1}(c_{jd})$ for $j \in \{1, \ldots, k\}$.
Step 3. Generate a set of quasi-random F-numbers $\{\mathbf{u}_j, j \in \{1, \ldots, k\}\}$ of the uniform distribution on the surface of unit sphere by TFWW algorithm with the first $(d-1)$-components of $\mathbf{c}_j$'s, $j \in \{1, \ldots, k\}$.
Step 4. A set of quasi-random F-numbers of the given spherical distribution $\mathrm{S}_d(g)$ is $\{\mathbf{s}_j = r_j\mathbf{u}_j, j \in \{1, \ldots, k\}\}$.

With the set of quasi-random F-numbers of $\mathbf{s} \sim \mathrm{S}_d(g)$ provided by the NTSR algorithm, we can obtain the QMC-RPs of $\mathbf{x} \sim \mathrm{EC}_d(\boldsymbol{\mu}, \boldsymbol{\Psi}; g)$ by relation (6), so the QMC-RPs of $\mathbf{x}$ form a set

$$\mathcal{B} = \{\mathbf{b}_j = \boldsymbol{\mu} + \boldsymbol{\Psi}^{1/2}\mathbf{s}_j, j \in \{1, \ldots, k\}\}. \tag{14}$$

In Step 1 of the NTSR algorithm, the quasi-random numbers are generated by the *glp* method in our study.

The NTSR algorithm for a skew-normal distribution is analogous to that for the elliptically contoured distributions. Due to the stochastic representation of skew-normal distribution and $(z_0, \mathbf{z}_1^\top)^\top \sim \mathrm{N}_{d+1}(\mathbf{0}, \mathbf{I}_{d+1})$ [3], where $z_0 \sim \mathrm{N}(0, 1)$ and $\mathbf{z}_1 \sim \mathrm{N}_d(\mathbf{0}, \mathbf{I}_d)$, a set of quasi-random F-numbers of $(z_0, \mathbf{z}_1^\top)^\top$ can be generated by the NTSR algorithm. Then, the set of quasi-random F-numbers of $\mathrm{SN}_d(\boldsymbol{\eta})$ is $\{\mathbf{w}_j = \boldsymbol{\delta}|(z_0)_j| + (\mathbf{I}_d - \boldsymbol{\delta}\boldsymbol{\delta}^\top)^{1/2}(\mathbf{z}_1)_j, j \in \{1, \ldots, k\}\}$, where $\boldsymbol{\delta} = \boldsymbol{\eta}/\sqrt{1 + \boldsymbol{\eta}^\top\boldsymbol{\eta}}$. Therefore, a set of QMC-RPs of $\mathbf{x} \sim \mathrm{SN}_d(\mathbf{a}, \mathbf{G}, \boldsymbol{\eta})$ can be obtained by (10), i.e., $\mathcal{B} = \{\mathbf{b}_j = \mathbf{a} + \mathbf{G}^{1/2}\mathbf{w}, j \in \{1, \ldots, k\}\}$.

2.4. Algorithms for generating MSE-RPs

Let $\mathbf{x} \sim F(\mathbf{x})$ with the density function $p(\mathbf{x})$ in $\mathbb{R}^d$. The set of MSE-RPs of $\mathbf{x}$ are obtained by minimizing MSE. In engineering, the MSE-RPs is also known as the optimal quantizer, but these two types of definitions are exactly the same. We may begin with the following definition of MSE-RPs.

Definition 3. The vector $\boldsymbol{\xi}_j \in \mathbb{R}^d$ for $j \in \{1, \ldots, k\}$ is called the MSE representative points (MSE-RPs) of $\mathbf{x} \sim F(\mathbf{x})$ if

$$\mathrm{E}\left[d^2(\mathbf{x}|\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k)\right] = \min_{\substack{\mathbf{y}_j\in\mathbb{R}^d \\ j\in\{1,\ldots,k\}}} \mathrm{E}\left[d^2(\mathbf{x}|\mathbf{y}_1, \ldots, \mathbf{y}_k)\right], \tag{15}$$

where

$$d^2(\mathbf{x}|\mathbf{y}_1,\ldots,\mathbf{y}_k)=\min_{h\in\{1,\ldots,k\}}\|\mathbf{x}-\mathbf{y}_h\|^2 \tag{16}$$

is the minimum distance between $\mathbf{x}$ and $\mathbf{y}_j, j\in\{1,\ldots,k\}$, and $\|\cdot\|$ is the ℓ_2-norm.

Let $\Xi=\{\boldsymbol{\xi}_j, j\in\{1,\ldots,k\}\}$ be a set of MSE-RPs, then a partition $\mathcal{S}=\{S_j, j\in\{1,\ldots,k\}\}$ of $\mathbf{x}$ is formed by

$$S_j=\{\mathbf{x}:\|\mathbf{x}-\boldsymbol{\xi}_j\|<\|\mathbf{x}-\boldsymbol{\xi}_h\|, j\neq h, 1\leq h,j\leq k\}, \tag{17}$$

and the partition $\mathcal{S}$ is also known as Voronoi partition [19] or domains of attraction. The following definition of optimal quantizer is equivalent to Definition 3.

Definition 4. An k-level quantizer $\mathcal{Q}=\{\mathcal{Y},\mathcal{S}\}$ of $\mathbf{x}$ consists of

(i) a set of output vectors $\mathcal{Y}=\{\mathbf{y}_j, j\in\{1,\ldots,k\}\}$;
(ii) a partition $\mathcal{S}=\{S_j, j\in\{1,\ldots,k\}\}$ of $\mathbb{R}^d$ with k disjoint and exhaustive regions;
(iii) a mapping $Q:\mathbb{R}^d\to\mathcal{S}$ defined by $Q(\mathbf{x})=\mathbf{y}_j$ if $\mathbf{x}\in S_j$.

Then, the optimal k-level quantizer minimizes

$$\mathrm{MSE}(\mathcal{Q})=\mathrm{E}(\|\mathbf{x}-Q(\mathbf{x})\|^2). \tag{18}$$

By Definition 4, a set of MSE-RPs $\Xi=\{\boldsymbol{\xi}_j, j\in\{1,\ldots,k\}\}$ satisfies $\mathrm{MSE}(\Xi)=\min\mathrm{MSE}(\mathcal{Q})$. There are many algorithms proposed for estimating MSE-RPs of a given distribution. For multivariate distributions, most of the methods are based on the k-means algorithm [23]. Suppose $\mathbf{x}$ is a d-dimensional random vector with a given distribution $F(\mathbf{x})$, Linde et al. [28] proposed a so-called LBG algorithm to compute the approximated MSE-RPs with respect to $F(\mathbf{x})$. In the LBG algorithm, the steps of computation are exactly the same as of the k-means algorithm, but the training set of size N and the k initial centroids are randomly generated from $F(\mathbf{x})$. As the set of initial centroids is randomly generated, the k-means algorithm does not guarantee that a set of MSE-RPs can be exactly obtained, as it strongly depends on the initial points and may converge to a local optimum. The same problem appears when using the LBG algorithm. To improve the LBG algorithm for generating MSE-RPs of the elliptically contoured distributions, the QMC methods are applied to generate the training set and the initial centroids from $F(\mathbf{x})$, i.e., a set of N QMC-RPs of $F(\mathbf{x})$ is regarded as the training set, and another set of k QMC-RPs of $F(\mathbf{x})$ as the initial centroids. This method combines the QMC methods and the LBG algorithm, and it is called the NTLBG algorithm [12]. The following steps are included in the NTLBG algorithm.

Step 1. For a given $F(\mathbf{x})$, generate a set of representative points $\mathcal{Y}=\{\mathbf{y}_1,\ldots,\mathbf{y}_N\}$ as a training set by the NTSR algorithm and the relation (6).

Step 2. Set $t=0$. For a given k, generate a set of representative points $\mathcal{W}_t=\{\mathbf{w}_{t1},\ldots,\mathbf{w}_{tk}\}$ of $F(\mathbf{x})$ as a set of initial centroids by means of NTSR algorithm and the relation (6).

Step 3. Find the associated partition $\mathcal{S}_t=\{S_{t1},\ldots,S_{tk}\}$ of $\mathcal{Y}$ by (17) and assign each $\mathbf{y}_i$ to the nearest cell S_{tj} for $i\in\{1,\ldots,N\}, j\in\{1,\ldots,k\}$.

Step 4. Compute the conditional mean $\mathbf{w}_{t+1,j}$ of $\mathbf{y}_i\in S_{tj}$ by

$$\mathbf{w}_{t+1,j}=\frac{1}{N_{tj}}\sum_{\mathbf{y}_i\in S_{tj}}\mathbf{y}_i, \tag{19}$$

where $N_{tj}=\sum_{i=1}^N I\{\mathbf{y}_i\in S_{tj}\}$ is the number of $\mathbf{y}_i$ in S_{tj}, $i\in\{1,\ldots,N\}, j\in\{1,\ldots,k\}$.

Step 5. Let $\mathcal{W}_{t+1}=\{\mathbf{w}_{t+1,1},\ldots,\mathbf{w}_{t+1,k}\}$. If $\mathcal{W}_{t+1}=\mathcal{W}_t$, then $\Xi=\{\boldsymbol{\xi}_j=\mathbf{w}_{t+1,j}, j\in\{1,\ldots,k\}\}$ is the set of MSE-RPs of $F(\mathbf{x})$ and terminate the algorithm. Otherwise, let $t=t+1$ and go to Step 2.

It is proved by Flury [16] that the MSE-RPs are self-consistent, and the definition of self-consistent points is given as follows.

Definition 5. Let $\mathbf{x}\in\mathbb{R}^d$ be a random vector and $\mathcal{Y}=\{\mathbf{y}_j, j\in\{1,\ldots,k\}\}$ be a set of points with partition $\mathcal{S}=\{S_j\}$ defined in (17). Then, the set of points $\mathcal{Y}$ is said to be self-consistent for $\mathbf{x}$ if $\mathrm{E}(\mathbf{x}|\mathbf{x}\in S_j)=\mathbf{y}_j, j\in\{1,\ldots,k\}$.

Hence, a set of point which are self-consistent can be regarded as MSE-RPs if the MSE is minimized, however, we need to note that self-consistent points are not necessarily MSE-RPs. As stated in [40], the k-means algorithm is a special case of self-consistency algorithms. So the k-means based algorithms including the LBG and NTLBG algorithms are also self-consistency algorithms. In the Step 4 in NTLBG algorithm, the conditional mean (19) is a sample version of self-consistent points when putting the same probability mass at each point $\mathbf{y}_i$ for $i\in\{1,\ldots,N\}$, i.e., for each t(19) can be written as

$$\mathbf{w}_{t+1,j}=\mathrm{E}(\mathbf{x}|\mathbf{x}\in S_{tj}), j\in\{1,\ldots,k\}.$$

This step is to compute the new centroids, and it provides a set of self-consistent points, but not necessarily a set of MSE-RPs. Compared with using random initial points, when the initial points are selected by the QMC method, a set of

Table 1
Notation of the three types of RPs with the probability of each point to be generated and the corresponding generating method.

	Generating method	Notation	Probability of the jth point
MC-RPs	MC method	$\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$	$p_j = 1/k, j \in \{1, \ldots, k\}$
QMC-RPs	NTSR and (6)	$\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$	$p_j = 1/k, j \in \{1, \ldots, k\}$
MSE-RPs	NTLBG and k-means++	$\Xi = \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_k\}$	$\hat{p}_j = N_j/N, j \in \{1, \ldots, k\}$

MSE-RPs with lower MSE can be obtained in most cases (see [12] for details), so the NTLBG algorithm can provide more accurate MSE-RPs than the LBG algorithm.

Another algorithm proposed by Arthur and Vassilvitskii [2], k-means++, is also the one to improve the k-means method by selecting the initial centroids with a probability-based method. The steps of the k-means++ seeding method are as follows.

Step (a) Set $j = 1$. Choose an initial centroid $\mathbf{w}_{0j}$ randomly from the training set $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$, i.e., $\mathbf{w}_{0j} = \mathbf{y}_i$ for some i. In this step, each $\mathbf{y}_i$ has a same probability $p_{ij} = 1/N$.

Step (b) Denote by $D(\mathbf{y})$ the shortest distance from $\mathbf{y}$ to the closest centroid that has been selected. Then, the next centroid is selected by $\mathbf{w}_{0,j+1} = \mathbf{y}_i$ with corresponding probability computed by

$$p_{i,j+1} = \frac{D(\mathbf{y}_i)^2}{\sum_{i=1}^N D(\mathbf{y}_i)^2}.$$

Step (c) If $j = k$, then all the k initial centroids have been selected, and stop the procedure. Otherwise, let $j = j + 1$ and return to Step (b).

As the probability p_{ij} of each $\mathbf{y}_i$ in the training set becomes higher with a larger $D(\mathbf{y})^2$, so the k-means++ seeding method is trying to choose each initial centroid as far away from the selected centroids as possible.

When applying the k-means++ algorithm to generate MSE-RPs of a given $F(\mathbf{x})$, we only need to replace the Step 2 in the NTLBG algorithm by the k-means++ seeding method, and other steps remain the same, that is, the training set is still generated by QMC method instead of MC method because the samples generated by the QMC method have better representativeness. In our numerical examples, we will apply both the NTSR algorithm and k-means++ seeding method to generate the set of initial points and choose the set of representative points with lower MSE as the estimation of MSE-RPs.

In the following two sections, we will apply representative points to geometric probability problems and to statistical inference for elliptically contoured distributions. For convenience, we list the notations for the MC-RPs, QMC-RPs, MSE-RPs and their corresponding probability in Table 1. For the MC-RPs and QMC-RPs, the probabilities at each point are the same. However, the probability of each point of MSE-RPs may not be the same and should be calculated individually. Commonly, the probability of $\boldsymbol{\xi}_j$ is estimated by $\hat{p}_j = N_j/N$, where N_j is the number of points of the training set located in S_j, and N is the total number of points in the training set.

3. Geometric probability

Statistical simulation has been widely used for problems in statistics with no analytic solution. For example, let D be a domain in $\mathbb{R}^d$ and E be the output of a stochastic process working on D. One needs to find the distribution of the area or volume of E. Statistical simulations need a set of points that are uniformly scattered on D and on any subarea of D including E. A set of random samples fails to provide such a set of points. Wang and Fang [47] suggested to apply QMC methods to construct such a net on D when D is not a rectangle, which is the NT-net on D. The following two case studies discussed in [12] show that both QMC-RPs and MSE-RPs can be used to generate a net on D.

3.1. *Area of intersection between a fixed circle and several random circles*

Let K be a unit circle with center at the origin $\mathbf{o} = (0, 0)^\top$. Suppose that $K_1, \ldots, K_m$ are the m random circles centered at $\mathbf{o}_1, \ldots, \mathbf{o}_m$ with given radii $r_1, \ldots, r_m$, respectively. The centers $\mathbf{o}_1, \ldots, \mathbf{o}_m$ are independent and $\mathbf{o}_j \sim N_2(\mathbf{0}, \sigma_j^2 \mathbf{I}_2)$, where $\sigma_j > 0$ is known. Let $S_m = K \cap (K_1 \cup \cdots \cup K_m)$, the overlapping region of K and the union of the m random circles $K_1, \ldots, K_m$, and denote $A(S_m)$ the area of S_m. The goal of the problem is to find the distribution of $A(S_m)$. It is not difficult to find the distribution of the overlapping area with a single random circle, however, there is no analytic solution for the cases when $m \geq 2$. Besides, Zhou and Fang [50] studied the problem in details and provided the analytic formulas to calculate $A(S_m)$ for $m \leq 2$, but the analytic solution to the overlapping area with $m > 2$ is complicated. Hence, an accurate estimation to the area $A(S_m)$ is crucial to solve the problem. Fang and Wang [12] suggested to apply the NT-net to estimate the overlapping area $A(S_m)$ so that we can obtain a random sample $A_1(S_m), \ldots, A_k(S_m)$ of $A(S_m)$, and then the empirical distribution of the sample can be regarded as an approximation of the distribution of $A(S_m)$. In the following discussion, we will focus on the estimation of $A(S_m)$ with a net generated by QMC-RPs and MSE-RPs.

Table 2
Selected values of $\{r_1, r_2, \sigma_1, \sigma_2\}$, center $\mathbf{o}_j$ of the jth random circle K_j randomly generated from $N_2(\mathbf{0}, \sigma_j^2 \mathbf{I}_2)$, $j \in \{1, 2\}$, and the corresponding analytical solution of $A(S_2)$.

	r_1	r_2	σ_1	σ_2	$\mathbf{o}_1$	$\mathbf{o}_2$	$A(S_2)$
Case 1	0.7500	0.7500	0.1000	0.9000	$(0.1093, 0.1109)^\top$	$(-0.7773, 0.0696)^\top$	2.3068
Case 2	0.8500	0.9500	0.5000	0.1000	$(-0.4565, 0.0605)^\top$	$(-0.1206, 0.1656)^\top$	2.6321
Case 3	0.5500	0.8500	0.7000	0.7000	$(-1.0248, -0.4244)^\top$	$(0.3796, 0.9042)^\top$	1.2633
Case 4	0.9500	0.6500	0.9000	0.5000	$(1.0942, -0.2511)^\top$	$(0.3349, -0.0812)^\top$	1.4690
Case 5	0.6500	0.5500	0.3000	0.3000	$(0.2604, 0.1034)^\top$	$(0.2515, -0.6134)^\top$	1.7881

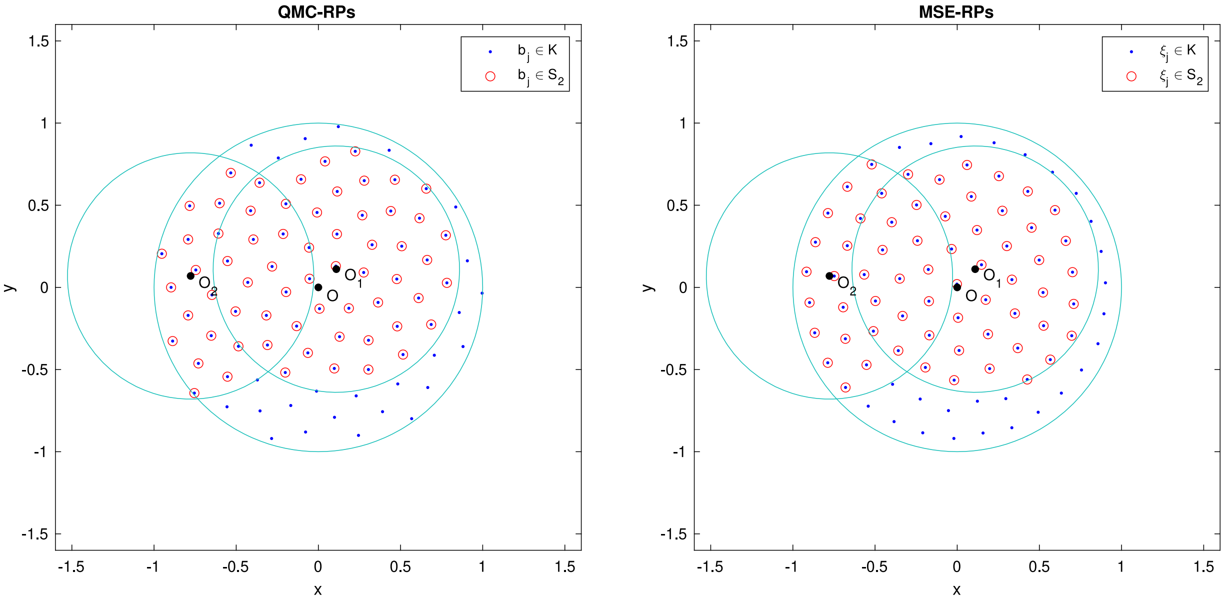

Fig. 1. The nets on D constructed by QMC-RPs (left figure) and MSE-RPs (right figure) of size $k = 89$. The dots in blue are the points in K and the circles in red are those in S_2 with $r_1 = 0.7500, r_2 = 0.7500, \sigma_1 = 0.1000, \sigma_2 = 0.9000$, $\mathbf{o}_1 = (0.1093, 0.1109)^\top$, $\mathbf{o}_2 = (-0.7773, 0.0696)^\top$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To estimate the area of S_m we need to generate a sample uniformly distributed on the unit circle. As a multi-uniform distribution is a subclass of elliptically contoured distribution, the QMC-RPs can be generated by the NTSR algorithm with the random variable $R \sim \text{beta}(d, 1)$, and the MSE-RPs are computed by the two k-means based methods (the NTLBG and the k-means++ algorithms). Let $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be a sample from the uniform distribution on K, then the area of S_m is estimated by

$$\hat{A}(S_m) = \frac{\pi}{k} \sum_{j=1}^{k} I\{\mathbf{u}_j \in S_m\}. \tag{20}$$

Let $d_{S_m} = |A(S_m) - \hat{A}(S_m)|$ be the difference between the analytical solution and the estimate of $A(S_m)$, and the performance of the two types of representative points can be evaluated by d_{S_m}. Consider the cases of two random circles K_1 and K_2, the parameters are $\{r_1, r_2, \sigma_1, \sigma_2\}$. Assume that $0.5 < r_j < 1$ and $0 < \sigma_j < 1$ for $j \in \{1, 2\}$. The values of r_j and σ_j are selected by a uniform design table $U_5(5^4)$ for the numerical experiments as shown in Table 2, and the corresponding analytical solution $A(S_2)$ of each combination of the parameters is also presented. Fig. 1 shows the two nets constructed by QMC-RPs and MSE-RPs, respectively, as well as the two random circles K_1, K_2 with $r_1 = 0.7500, r_2 = 0.7500, \sigma_1 = 0.1000, \sigma_2 = 0.9000$, $\mathbf{o}_1 = (0.1093, 0.1109)^\top$, $\mathbf{o}_2 = (-0.7773, 0.0696)^\top$.

Table 3 shows the estimation bias of $A(S_2)$ using the two types of representative points for the 5 cases listed in Table 2. Table 4 counts the number of smallest estimation bias obtained by each type of representative points in the experiments. Based on the numerical results, both the QMC-RPs and MSE-RPs provide a good estimation of $A(S_2)$. Compared with MSE-RPs, the QMC-RPs seems to have a slightly better performance in this problem, though the values of estimation biases of QMC-RPs and MSE-RPs are very close.

Table 3
Estimation bias d_{S_m} with QMC-RPs and MSE-RPs on unit circle K for the 5 cases in Table 2. Smallest estimation bias is in bold.

Case 1	k	QMC-RPs	MSE-RPs	k	QMC-RPs	MSE-RPs
	34	**0.00327604**	0.09567583	610	0.00569964	**0.00460066**
	89	**0.04759575**	0.11819334	987	0.00543032	**0.05317489**
	144	**0.03779053**	**0.03779053**	1597	**0.00708060**	0.00904779
	233	**0.01239705**	0.02805264	2584	**0.00280289**	0.01374497
	377	**0.00156040**	0.01510588	4181	**0.00067984**	0.01195081
Case 2	k	QMC-RPs	MSE-RPs	k	QMC-RPs	MSE-RPs
	34	**0.04487344**	**0.04487344**	610	**0.00481038**	0.00548992
	89	**0.01534215**	**0.01534215**	987	**0.00611608**	0.00929905
	144	0.03589012	**0.01407351**	1597	**0.00002464**	0.00194254
	233	0.01632071	**0.01064575**	2584	**0.00132646**	0.01326298
	377	**0.00712923**	0.03212864	4181	**0.00067359**	0.00082921
Case 3	k	QMC-RPs	MSE-RPs	k	QMC-RPs	MSE-RPs
	34	0.12262137	**0.03022159**	610	**0.01901243**	0.02931274
	89	0.04267998	**0.00738119**	987	0.01246768	**0.00610173**
	144	0.04164491	**0.00198832**	1597	**0.00044334**	0.00241052
	233	**0.02291820**	0.03101472	2584	**0.00104271**	0.01597831
	377	**0.00326144**	0.02826085	4181	**0.00102771**	0.00498347
Case 4	k	QMC-RPs	MSE-RPs	k	QMC-RPs	MSE-RPs
	34	0.10177038	**0.00937059**	610	**0.00123266**	0.01153296
	89	0.05707419	**0.02177540**	987	**0.00804213**	0.02077402
	144	**0.00731270**	**0.00731270**	1597	**0.00046041**	**0.00046041**
	233	**0.05457907**	**0.05457907**	2584	**0.00329166**	0.01980828
	377	**0.00239383**	0.00593930	4181	0.00079537	**0.00004397**
Case 5	k	QMC-RPs	MSE-RPs	k	QMC-RPs	MSE-RPs
	34	**0.03256731**	**0.03256731**	610	**0.00408966**	0.02984042
	89	**0.01207529**	0.02322351	987	**0.00888226**	**0.00888226**
	144	0.04283395	**0.00079928**	1597	**0.00197413**	0.00392742
	233	**0.00837681**	0.07579296	2584	**0.00633787**	0.03378309
	377	0.02846067	**0.02012753**	4181	**0.00134011**	0.00542246

Table 4
The number of smallest estimation bias of $A(S_2)$ by QMC-RPs and MSE-RPs in Table 3.

	Case 1	Case 2	Case 3	Case 4	Case 5	Total
QMC-RPs	8	8	6	7	8	37
MSE-RPs	3	4	4	6	4	21

3.2. Area of random belts with a fixed width covering a unit sphere

Let $\mathbf{S}^3$ be a unit sphere in $\mathbb{R}^3$ centering at the origin $(0,0,0)^\top$, and it can be covered by random belts with a fixed width. Each belt is symmetric about a unit circle in $\mathbf{S}^3$ centered at the origin $(0,0)^\top$. Let $\mathbf{n} \in \mathbb{R}^3$ be a normal direction, then the belt on the surface of $\mathbf{S}^3$ with thickness $2h$ can be written as

$$G_h(\mathbf{n}) = \{\mathbf{w} : |\mathbf{w}^\top \mathbf{n}| \le h\}, \tag{21}$$

and the area of the belt can be computed by $A(G_h(\mathbf{n})) = 4\pi h$. The original problem arises in steel rolling and aims to find the distribution of a roller's lifespan and find some ways to increase the lifespan (see [12] for details). Similarly to the problem of finding distribution of overlapping area in Section 3.1, a net on D is also necessary for estimating the area of each belts with low bias. We will only focus on one random belt and estimate its area with QMC-RPs and MSE-RPs and compare their performance. Suppose $\mathcal{U}^{(3)} = \{\mathbf{u}_1^{(3)}, \dots, \mathbf{u}_k^{(3)}\}$ is a set of points uniformly distributed on the surface of $\mathbf{S}^3$, the area of $G_h(\mathbf{n})$ can be estimated by

$$\hat{A}(G_h(\mathbf{n})) = \frac{4\pi}{k} \sum_{j=1}^{k} I\{|(\mathbf{u}^{(3)})^\top \mathbf{n}| \le h\} \tag{22}$$

with the estimation bias evaluated by $d_{G_h} = |A(G_h(\mathbf{n})) - \hat{A}(G_h(\mathbf{n}))|$. In fact, the area of a random belt on the surface of $\mathbf{S}^3$ with a given h remains the same with any direction vector , we will assume that $\mathbf{n} = (0,0,1)^\top$ for convenience and denote $G_h = G_h((0,0,1)^T)$ as the belt with thickness $2h$ and $A(G_h)$ as its area.

Similar to the estimation of overlapping area $A(S_m)$, we can also generate different sets of representative points of the uniform distribution on the surface of $\mathbf{S}^d$, including QMC-RPs and MSE-RPs. Suppose $\mathbf{s} \sim S_d(g)$, and its stochastic

Table 5
Estimation bias d_{G_h} by QMC-RPs and MSE-RPs on the surface of $\mathbf{S}^3$ for $h \in \{0.05, 0.10, 0.15\}$. Smallest estimation bias is in bold.

$h = 0.05$, $A(G_{0.05}) = 0.6283$					
k	QMC-RPs	MSE-RPs	k	QMC-RPs	MSE-RPs
185	**0.05094475**	0.08490791	1626	**0.00231852**	0.01313827
266	0.06141459	**0.03306940**	2440	**0.01030030**	0.01545046
418	0.08718295	**0.02705678**	3237	**0.01999284**	0.02271031
597	**0.00315738**	0.03894101	4044	0.00372889	**0.00248593**
828	0.02428284	**0.00910607**	5037	**0.00212059**	0.00461540
1220	0.02060061	**0.01030030**	6066	0.00890791	**0.00683630**
$h = 0.10$, $A(G_{0.10}) = 1.2566$					
k	QMC-RPs	MSE-RPs	k	QMC-RPs	MSE-RPs
185	**0.03396316**	**0.03396316**	1626	0.06491852	**0.03400494**
266	0.12282919	**0.01889680**	2440	0.04635137	**0.02060061**
418	0.05411356	**0.00601262**	3237	0.00271747	**0.00116463**
597	**0.00631476**	0.07788203	4044	**0.00807927**	0.01118668
828	0.02731820	**0.00303536**	5037	0.00573807	**0.00324326**
1220	0.05150152	**0.02060061**	6066	**0.00538618**	0.01781582
$h = 0.15$, $A(G_{0.15}) = 1.8850$					
k	QMC-RPs	MSE-RPs	k	QMC-RPs	MSE-RPs
185	0.22076056	**0.05094475**	1626	**0.00695556**	0.03014074
266	0.08975979	**0.00472420**	2440	0.01030030	**0.00515015**
418	0.12927128	**0.02104416**	3237	**0.00601726**	0.02115747
597	**0.01157706**	**0.01157706**	4044	**0.00186445**	**0.00186445**
828	**0.01821213**	0.06374246	5037	**0.00137215**	0.04852410
1220	0.05150152	**0.01030030**	6066	**0.01263681**	0.02258052

representation is $\mathbf{s} \stackrel{d}{=} R\mathbf{u}^{(d)}$, where $\|\mathbf{s}\| \stackrel{d}{=} R$, $\mathbf{s}/\|\mathbf{s}\| \stackrel{d}{=} \mathbf{u}^{(d)}$, and $\|\mathbf{s}\|$ is independent of $\mathbf{s}/\|\mathbf{s}\|$ (see Theorem 2.3 of [11]). Hence, the set of QMC-RPs of uniform distributions on the surface of the unit sphere is generated by the following two steps. We can firstly generate a set of QMC-RPs $\{\mathbf{z}_j, j \in \{1, \ldots, k\}\}$ of $\mathrm{N}_d(\mathbf{0}, \mathbf{I}_d)$, which is also a spherical distribution, and then the set of QMC-RPs of uniform distributions on the surface of $\mathbf{S}^d$ can be computed by $\mathbf{b}_j = \mathbf{z}_j/\|\mathbf{z}_j\|$.

When the NTLBG algorithm is used to generate MSE-RPs, the Euclidean distance is commonly used to calculate the distance from a point to the centroids. However, the points are on the surface of sphere, so a so-called cosine dissimilarity seems more reasonable to evaluate the difference between two points, i.e.,

$$d(\mathbf{a}, \mathbf{b}) = 1 - \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}, \tag{23}$$

where $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$. Moreover, for the training set $\mathcal{Y} = \{\mathbf{y}_i, i \in \{1, \ldots, N\}\}$, we can also have the partition $\mathcal{S} = \{S_{tj}\}$ associated with $\mathbf{w}_{tj}, j \in \{1, \ldots, k\}$, but the new centroids are then computed by

$$\mathbf{w}_{t+1,j} = \frac{\sum_{\mathbf{y}_i \in S_{tj}} \mathbf{y}_i}{\|\sum_{\mathbf{y}_i \in S_{tj}} \mathbf{y}_i\|}, \tag{24}$$

that is, each centroid is the mean of $\mathbf{y}_i \in S_{tj}$ after normalizing these points to unit by Euclidean length. The cosine dissimilarity is originally suggested by Dhillon and Modha [7] for the problems of processing and analyzing textual data to reduce the effect of document length, and the k-means algorithm with cosine dissimilarity (23) and new centroids computed by (24) is the so-called spherical k-means algorithm [6]. This algorithm seems more reasonable than the standard k-means algorithm here because the points are on the surface of a sphere instead of a plane, and the new centroids will be inside the sphere if they are not normalized as shown in (24), whereas a set of new centroids on the surface of the sphere is required. For these reasons, we will replace (19) by (24) for the calculation of MSE-RPs on the surface of a sphere.

In the following experiments, take $h \in \{0.05, 0.1, 0.15\}$, for example, and the corresponding analytical solutions of the belt area are $A(G_{0.05}) = 0.6283$, $A(G_{0.1}) = 1.2566$ and $A(G_{0.15}) = 1.8850$, respectively, and we will estimate $A(G_h)$ with different values of k by the sets of representative points generated by the QMC methods (TFWW algorithm in this problem), and the two k-means based methods. The numerical results are shown in Tables 5 and 6. In this problem, both the two types of RPs have good performance when estimating $A(G_{0.05})$, $A(G_{0.15})$, but the MSE-RPs are more competitive than QMC-RPs for the estimation of $A(G_{0.1})$.

The numerical results of the first problem show that the QMC-RPs provide a better estimation of the overlapping area $A(S_m)$ in most cases. Hence, $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ may have a better approximation to the uniform distribution on a unit circle, but it does not imply that MSE-RPs do not perform well in this problem. For the second problem, the MSE-RPs is slightly better than the QMC-RPs, and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ may have a better representation of uniform distribution on the surface of a unit sphere.

Table 6
The number of smallest estimation bias of $A(G_h)$ by QMC-RPs and MSE-RPs in Table 5.

	$A(G_{0.05})$	$A(G_{0.10})$	$A(G_{0.15})$	Total
QMC-RPs	6	4	7	17
MSE-RPs	6	9	7	22

Therefore, we may recommend both QMC-RPs and MSE-RPs of the uniform distribution on a region D for the geometric probability problems.

4. Statistical inference by resampling

For a given continuous distribution $F(\mathbf{x})$ with finite mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, a discrete approximation $\hat{F}_k(\mathbf{x})$ with support points $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ and related probabilities $p_j = \Pr(\mathbf{x} = \mathbf{x}_j)$, $j \in \{1, \ldots, k\}$ can be constructed by the MC-RPs, QMC-RPs and MSE-RPs. The corresponding mean vector and covariance matrix are calculated by

$$\hat{\boldsymbol{\mu}} = \sum_{j=1}^{k} \mathbf{x}_j p_j,\ \hat{\boldsymbol{\Sigma}} = \sum_{j=1}^{k} (\mathbf{x}_j - \hat{\boldsymbol{\mu}})(\mathbf{x}_j - \hat{\boldsymbol{\mu}})^\top p_j, \tag{25}$$

which are the estimates of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, respectively. A ℓ_2-norm is applied to evaluate the estimation bias of the mean vector $\boldsymbol{\mu}$, and a Frobenius norm $\|\boldsymbol{\Sigma} - \hat{\boldsymbol{\Sigma}}\|_F$ for the covariance matrix $\boldsymbol{\Sigma}$, where the Frobenius norm of a matrix $\mathbf{A}$ can be computed by $\|\mathbf{A}\|_F = \sqrt{\mathrm{trace}(\mathbf{A}^\top \mathbf{A})}$.

In the following numerical experiments, we try to estimate the mean vector and covariance matrix of elliptically contoured distributions and the skew-normal distribution of dimension $d = 3$ by the three types of representative points, and this is a preliminary comparison among the MC-RPs, QMC-RPs and MSE-RPs. Assume that the mean vector $\boldsymbol{\mu} = \mathbf{0}$, and the three covariance matrices are randomly chosen for fair comparisons in the following way. Firstly, we choose three matrices $\mathbf{U}_i$ of size 3×3, $i \in \{1, 2, 3\}$, where all elements in $\mathbf{U}_i$ are i.i.d. random numbers from the uniform distribution $U(0, 1)$, which are

$$\mathbf{U}_1 = \begin{bmatrix} 0.8147 & 0.9134 & 0.2785 \\ 0.9058 & 0.6324 & 0.5469 \\ 0.1270 & 0.0975 & 0.9575 \end{bmatrix},\quad \mathbf{U}_2 = \begin{bmatrix} 0.4393 & 0.4706 & 0.6357 \\ 0.5778 & 0.4499 & 0.7960 \\ 0.8024 & 0.4614 & 0.8750 \end{bmatrix},$$
$$\mathbf{U}_3 = \begin{bmatrix} 0.6849 & 0.8635 & 0.9878 \\ 0.9878 & 0.1841 & 0.3504 \\ 0.9100 & 0.7210 & 0.0060 \end{bmatrix}.$$

Then, we construct the covariance matrices by $\boldsymbol{\Sigma}_i = \mathbf{U}_i^\top \mathbf{U}_i$. We have carried out the numerical experiments for normal and Kotz type distributions. Assume that the normal distribution $\mathrm{N}_3(\mathbf{0}, \boldsymbol{\Sigma})$ and the Kotz type distribution $\mathrm{Kotz}_3(3, 3, 1)$ share the same mean vector $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_i$ in the experiments. Since the numerical results of the normal distribution is quite similar to that of the Kotz type distribution, we omit the tables of results for the normal distribution to save space. The results of estimation bias for $\mathrm{Kotz}_3(3, 3, 1)$ are shown in Table 7 with the counts of smallest estimation bias for each set of representative points in Table 8. For the skew-normal distribution $\mathrm{SN}_3(\mathbf{a}, \mathbf{G}, \boldsymbol{\eta})$, suppose that the location vector $\mathbf{a} = \mathbf{0}$, the dispersion matrix $\mathbf{G} = \boldsymbol{\Sigma}_i$ and the skewness vector $\boldsymbol{\eta} = (2, 3, 1)^\top$, then the mean vector and the covariance matrix of the skew-normal distribution with these parameters can be calculated by (11). Tables 9, 10 are the results for $\mathrm{SN}_3(\mathbf{0}, \mathbf{G}, (2, 3, 1)^\top)$. The numbers of representative points k are chosen from the tables of the *glp* set in Appendix A of [12] for convenience, as the generating vectors of some specific number N are required when using the QMC methods to generate a set of QMC-RPs. From the numerical results, we can see that estimating the mean vector $\boldsymbol{\mu}$ with MSE-RPs provides smaller bias than QMC-RPs in all the experiments, and the results of estimation bias are all equal regardless of the value of k. This is reasonable due to the self-consistency of MSE-RPs, i.e.,

$$\mathrm{E}(\mathbf{x}) = \sum_{j=1}^{k} \mathrm{E}(\mathbf{x} | \mathbf{x} \in S_j) \Pr(\mathbf{x} \in S_j) = \sum_{j=1}^{k} \boldsymbol{\xi}_j p_j,$$

where $\mathrm{E}(\mathbf{x}|\mathbf{x} \in S_j) = \boldsymbol{\xi}_j$, $j \in \{1, \ldots, k\}$ by self-consistency and $p_j = \Pr(\mathbf{x} \in S_j)$ is the corresponding probability of $\boldsymbol{\xi}_j$, so the mean vector of $\mathbf{x}$ is exactly the same as that of its estimate. However, $\boldsymbol{\xi}_j$ and its probability are estimated for $j \in \{1, \ldots, k\}$ by

$$\boldsymbol{\xi}_j = \frac{1}{N_j} \sum_{\mathbf{y}_i \in S_j} \mathbf{y}_i,\ \hat{p}_j = \frac{N_j}{N}$$

Table 7
Estimation bias of the estimated mean vector and covariance matrices $\mathbf{\Sigma}_i = \mathbf{U}_i^\top \mathbf{U}_i$, $i \in \{1, 2, 3\}$, for Kotz$_3$(3, 3, 1). $\hat{F}_{\mathrm{MC}}(\mathbf{x})$, $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ are used for discrete approximations to Kotz$_3$(3, 3, 1). Smallest estimation bias is in bold.

$\mathbf{\Sigma} = \mathbf{\Sigma}_1$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\mathbf{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.332666	0.014961	**0.000040**	2.132877	0.451716	**0.107800**
25	0.091474	0.080595	**0.000040**	1.508648	0.257488	**0.053842**
35	0.242490	0.021176	**0.000040**	0.516929	0.564609	**0.039294**
101	0.029241	0.006214	**0.000040**	0.592528	0.392158	**0.013585**
185	0.187304	0.020161	**0.000040**	0.137735	0.130086	**0.007840**
418	0.079177	0.002024	**0.000040**	0.418887	0.008194	**0.003777**
828	0.115085	0.002889	**0.000040**	0.416516	0.012089	**0.002137**
1459	0.004132	0.005112	**0.000040**	0.174390	0.014945	**0.001427**
2440	0.009230	0.001968	**0.000040**	0.225844	0.026886	**0.001063**
3237	0.022368	0.001951	**0.000040**	0.159070	0.004403	**0.000919**
$\mathbf{\Sigma} = \mathbf{\Sigma}_2$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\mathbf{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.717651	0.007931	**0.000033**	1.863166	0.276473	**0.039740**
25	0.217945	0.033538	**0.000033**	0.211429	0.085495	**0.024462**
35	0.198866	0.014331	**0.000033**	0.272105	**0.165903**	0.017020
101	0.232616	0.003511	**0.000033**	0.769038	0.104015	**0.006570**
185	0.120270	0.011898	**0.000033**	0.079402	0.029729	**0.004190**
418	0.173440	0.001293	**0.000033**	0.049242	**0.001864**	0.002387
828	0.062791	0.001910	**0.000033**	0.156528	0.002817	**0.001504**
1459	0.031152	0.003180	**0.000033**	0.141750	0.004449	**0.001050**
2440	0.006956	0.000998	**0.000033**	0.057752	0.006836	**0.000764**
3237	0.071384	0.000823	**0.000033**	0.024114	0.001395	**0.000642**
$\mathbf{\Sigma} = \mathbf{\Sigma}_3$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\mathbf{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.599516	0.026194	**0.000088**	0.527672	0.350963	**0.252247**
25	0.304867	0.067033	**0.000088**	0.724271	**0.125155**	0.179515
35	0.359905	0.019325	**0.000088**	1.105650	0.262044	**0.143330**
101	0.299775	0.005155	**0.000088**	0.413505	0.118756	**0.074556**
185	0.268338	0.014523	**0.000088**	0.137162	**0.042243**	0.051571
418	0.084572	0.002398	**0.000088**	0.247787	**0.007914**	0.030914
828	0.052073	0.001752	**0.000088**	0.051831	**0.005881**	0.020004
1459	0.038571	0.004452	**0.000088**	0.217465	**0.005527**	0.013885
2440	0.092121	0.001066	**0.000088**	0.068231	**0.005999**	0.009895
3237	0.022408	0.001222	**0.000088**	0.055010	**0.004821**	0.008193

Table 8
The number of smallest estimation bias of estimated mean vector and covariance matrix in Table 7 for each type of RPs.

$\mathbf{\Sigma} = \mathbf{\Sigma}_i$	Estimation of $\boldsymbol{\mu}$			Estimation of $\mathbf{\Sigma}$		
	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
$\mathbf{\Sigma}_1$	0	0	10	0	0	10
$\mathbf{\Sigma}_2$	0	0	10	0	2	8
$\mathbf{\Sigma}_3$	0	0	10	0	7	3
Total	0	0	30	0	9	21

based on the training set $\mathcal{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$, which gives

$$\hat{\boldsymbol{\mu}} = \sum_{j=1}^{k} \left(\frac{1}{N_j} \sum_{\mathbf{y}_i \in S_j} \mathbf{y}_i \right) \frac{N_j}{N} = \frac{1}{N} \sum_{j=1}^{k} \sum_{\mathbf{y}_i \in S_j} \mathbf{y}_i = \frac{1}{N} \sum_{i=1}^{N} \mathbf{y}_i.$$

Since the training set for generating MSE-RPs by k-means based methods is generated by the QMC methods, the mean vector $\boldsymbol{\mu}$ is essentially estimated by a set of N QMC-RPs in $\mathcal{Y}$ with respect to $F(\mathbf{x})$. For the estimation of $\mathbf{\Sigma}$ with MSE-RPs,

$$\hat{\mathbf{\Sigma}} = \sum_{j=1}^{k} \boldsymbol{\xi}_j \boldsymbol{\xi}_j^\top \hat{p}_j - \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top = \sum_{j=1}^{k} \left(\frac{1}{N_j} \sum_{\mathbf{y}_i \in S_j} \mathbf{y}_i \right) \left(\frac{1}{N_j} \sum_{\mathbf{y}_i \in S_j} \mathbf{y}_i \right)^\top \frac{N_j}{N} - \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top = \frac{1}{N} \sum_{j=1}^{k} N_j \bar{\mathbf{y}}_j \bar{\mathbf{y}}_j^\top - \hat{\boldsymbol{\mu}}\hat{\boldsymbol{\mu}}^\top,$$

where $\hat{\boldsymbol{\mu}}$ is estimated by all points in the training set $\mathcal{Y}$ generated by QMC methods while $\bar{\mathbf{y}}_j = (\sum_{\mathbf{y}_i \in S_j} \mathbf{y}_j)/N_j$ is related to the MSE-RPs. Compared with MC-RPs, both QMC-RPs and MSE-RPs can improve the accuracy of the estimation. In general, the MSE-RPs have a better performance, but the results obtained by QMC-RPs are more accurate in some cases. Hence, the performance of QMC-RPs and MSE-RPs may depend on the underlying distribution and their parameters.

The bootstrap method proposed by Efron [8] is one of the resampling techniques, and it has been widely used in statistical inference. Let $\mathbf{v}$ be a random vector of the discrete distribution $\hat{F}_k(\mathbf{x})$ that approximate $F(\mathbf{x})$ with support $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ with corresponding probabilities $p_1, \ldots, p_k$. Then, the bootstrap method for estimating the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ involves the following steps.

Step 1. Set $i = 1$. Generate a sample $\mathcal{V}_i^* = \{\mathbf{v}_{1i}, \ldots, \mathbf{v}_{ki}\}$ from $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ with probability $p_1, \ldots, p_k$.
Step 2. Compute

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{k}\sum_{j=1}^{k} \mathbf{v}_{ji}^*, \; \hat{\boldsymbol{\Sigma}}_i = \frac{1}{k}\sum_{j=1}^{k} (\mathbf{v}_{ji}^* - \hat{\boldsymbol{\mu}}_i)(\mathbf{v}_{ji}^* - \hat{\boldsymbol{\mu}}_i)^\top.$$

Step 3. Repeat the first two steps B times. Then, the mean vector and covariance matrix of $\mathbf{x}$ are estimated by

$$\hat{\boldsymbol{\mu}} = \frac{1}{B}\sum_{i=1}^{B} \hat{\boldsymbol{\mu}}_i, \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{B}\sum_{i=1}^{B} \hat{\boldsymbol{\Sigma}}_i.$$

The sets of MC-RPs, QMC-RPs and MSE-RPs are also used for estimating the mean vector and the covariance matrix of normal, Kotz type and skew-normal distributions in the above experiments by the bootstrap method. In the following experiments, let $B = 2000$, and the numerical results are shown in the same form as in the preliminary comparison in Tables 7–10. The results of $N_3(\mathbf{0}, \boldsymbol{\Sigma})$ by resampling are also omitted, as they are similar to $\mathrm{Kotz}_3(3, 3, 1)$ with the same mean vector and covariance matrix. Based on the results of $\mathrm{Kotz}_3(3, 3, 1)$ (Tables 11, 12), the QMC-RPs perform slightly better than the MSE-RPs for the estimation of the covariance matrix. However, Tables 13 and 14 for $SN_3(\mathbf{0}, \mathbf{G}, (2, 3, 1)^\top)$ show that a more accurate estimation can be obtained by MSE-RPs, which is quite different from the conclusion drawn from the results of the other two distributions. Hence, both QMC-RPs and MSE-RPs are useful for estimating the covariance matrix by resampling techniques, and their performance may depend on the underlying distribution. Besides, MSE-RPs provide an estimation of the mean vector by the bootstrap method with higher accuracy than the QMC-RPs in most experiments.

In the traditional bootstrap method, a set of MC-RPs is used for the support of the discrete approximation. As we may see in Tables 11–14, there is a significant improvement in most cases if the QMC-RPs and MSE-RPs are regarded as the support of $\mathbf{v}$ in the bootstrap approach, so these types of representative points have a great potential for statistical inference by performing resampling techniques. Estimating the covariance matrix by resampling from QMC-RPs appears to provide slightly higher estimation accuracy for elliptically contoured distributions, while MSE-RPs may be more competitive for the skew-normal distribution.

5. Relation between MSE-RPs and principal component analysis

The MSE-RPs generated under the criterion MSE are related to principal components. We may begin with the criterion MSE. Let $\mathbf{x} \in \mathbb{R}^d$ be a random vector of a continuous distribution $F(\mathbf{x})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and let the density function of $\mathbf{x}$ be $p(\mathbf{x})$. The set of MSE-RPs with respect to $F(\mathbf{x})$ can be regarded as the support of $\boldsymbol{\xi}$, a random vector of a discrete approximation of $F(\mathbf{x})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$. With the partition $\mathcal{S} = \{S_j, j \in \{1, \ldots, k\}\}$ of the support of $\mathbf{x}$, the MSE of $\Xi = \{\boldsymbol{\xi}_j, j \in \{1, \ldots, k\}\}$ can be written as

$$\begin{aligned}
\mathrm{MSE}(\Xi) &= \int \min_{j=1,\ldots,k} \|\mathbf{x} - \boldsymbol{\xi}_j\|^2 p(\mathbf{x}) d\mathbf{x} = \sum_{j=1}^{k} \int_{S_j} \|\mathbf{x} - \boldsymbol{\xi}_j\|^2 p(\mathbf{x}) d\mathbf{x} = \sum_{j=1}^{k} \mathrm{E}(\|\mathbf{x} - \boldsymbol{\xi}_j\|^2 | \mathbf{x} \in S_j) \Pr(\mathbf{x} \in S_j) \\
&= \sum_{j=1}^{k} \mathrm{E}[(\mathbf{x} - \boldsymbol{\xi}_j)^\top (\mathbf{x} - \boldsymbol{\xi}_j) | \mathbf{x} \in S_j] \Pr(\mathbf{x} \in S_j) = \sum_{j=1}^{k} \mathrm{trace}(\mathrm{Cov}(\mathbf{x} | \mathbf{x} \in S_j)) p_j \qquad (26) \\
&= \mathrm{E}[(\mathbf{x} - \boldsymbol{\mu})^\top (\mathbf{x} - \boldsymbol{\mu})] - \sum_{j=1}^{k} (\boldsymbol{\xi}_j - \boldsymbol{\mu})^\top (\boldsymbol{\xi}_j - \boldsymbol{\mu}) \Pr(\mathbf{x} \in S_j) = \mathrm{trace}(\boldsymbol{\Sigma}) - \sum_{j=1}^{k} \|\boldsymbol{\xi}_j - \boldsymbol{\mu}\|^2 p_j \qquad (27) \\
&= \mathrm{trace}(\boldsymbol{\Sigma}) - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}}), \qquad (28)
\end{aligned}$$

where $\mathrm{E}(\mathbf{x} | \mathbf{x} \in S_j) = \boldsymbol{\xi}_j$ by use of the self-consistency of MSE-RPs and $p_j = \Pr(\mathbf{x} \in S_j)$. Eq. (26) leads to one more definition of MSE-RPs given in [16], which implies that the MSE-RPs minimize the sum of the trace of the covariance matrix of $\mathbf{x} \in S_j$. The expression of MSE (27) provided in [38] is more convenient to compute the MSE, and the expression (28) has been given in [41].

Table 9
Estimation bias of the estimated mean vector and covariance matrices for $SN_3(\mathbf{0}, \mathbf{G}, (2, 3, 1)^\top)$, where $\mathbf{G} = \boldsymbol{\Sigma}_i = \mathbf{U}_i^\top \mathbf{U}_i$, $i \in \{1, 2, 3\}$. $\hat{F}_{MC}(\mathbf{x})$, $\hat{F}_{QMC}(\mathbf{x})$ and $\hat{F}_{MSE}(\mathbf{x})$ are used for discrete approximation to $SN_3(\mathbf{0}, \mathbf{G}, (2, 3, 1)^\top)$. Smallest estimation bias is in bold.

$\mathbf{G} = \boldsymbol{\Sigma}_1$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\boldsymbol{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.167373	0.066652	**0.000073**	0.466210	0.450941	**0.159836**
25	0.255815	0.022767	**0.000073**	0.651583	0.424710	**0.103057**
31	0.387120	0.057031	**0.000073**	0.367563	0.285593	**0.086323**
307	0.031687	0.019381	**0.000073**	0.077712	0.031647	**0.018398**
562	0.070129	0.009711	**0.000073**	0.057653	0.027562	**0.012376**
701	0.014221	0.007442	**0.000073**	0.080039	0.077150	**0.010650**
1019	0.021079	0.005285	**0.000073**	0.090031	0.016345	**0.008202**
2129	0.023713	0.001252	**0.000073**	0.023101	0.008862	**0.004812**
3001	0.014740	0.001113	**0.000073**	0.024302	0.006504	**0.003716**
4001	0.005111	0.001911	**0.000073**	0.025445	0.004668	**0.002957**
$\mathbf{G} = \boldsymbol{\Sigma}_2$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\boldsymbol{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.442373	0.086722	**0.000036**	1.128469	0.581315	**0.033676**
25	0.468967	0.019518	**0.000036**	0.631236	0.322768	**0.021285**
31	0.203203	0.024511	**0.000036**	0.632891	0.419159	**0.017286**
307	0.076553	0.016586	**0.000036**	0.189267	0.036246	**0.002495**
562	0.017958	0.010257	**0.000036**	0.029340	0.009044	**0.001648**
701	0.009657	0.007103	**0.000036**	0.124827	0.021135	**0.001423**
1019	0.058045	0.006286	**0.000036**	0.026544	0.021020	**0.001106**
2129	0.048222	0.001549	**0.000036**	0.045962	0.007927	**0.000660**
3001	0.002164	0.001219	**0.000036**	0.052943	0.006965	**0.000513**
4001	0.005792	0.002459	**0.000036**	0.030282	0.004677	**0.000411**
$\mathbf{G} = \boldsymbol{\Sigma}_3$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\boldsymbol{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.471961	0.120231	**0.000101**	0.601581	0.351456	**0.217226**
25	0.129859	0.068746	**0.000101**	0.443513	0.471129	**0.158461**
31	0.139151	0.072319	**0.000101**	0.960364	0.283401	**0.139317**
307	0.159581	0.024045	**0.000101**	0.105484	0.040397	**0.033554**
562	0.078494	0.012173	**0.000101**	0.109267	0.027626	**0.022504**
701	0.048212	0.008130	**0.000101**	0.108085	0.082647	**0.019329**
1019	0.057387	0.007629	**0.000101**	0.065771	0.024161	**0.015044**
2129	0.021726	0.001427	**0.000101**	0.038704	0.012330	**0.008847**
3001	0.031081	0.002069	**0.000101**	0.011811	**0.005175**	0.006812
4001	0.022209	0.002352	**0.000101**	0.069842	**0.003683**	0.005447

Table 10
The number of smallest estimation bias of estimated mean vector and covariance matrix in Table 9 for each type of RPs.

$\mathbf{G} = \boldsymbol{\Sigma}_i$	Estimation of $\boldsymbol{\mu}$			Estimation of $\boldsymbol{\Sigma}$		
	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
$\boldsymbol{\Sigma}_1$	0	0	10	0	0	10
$\boldsymbol{\Sigma}_2$	0	0	10	0	0	10
$\boldsymbol{\Sigma}_3$	0	0	10	0	2	8
Total	0	0	30	0	2	28

The following two properties provided by Tarpey et al. [42] allow us to assume the mean vector $\boldsymbol{\mu}$ of $\mathbf{x}$ to be $\mathbf{0}$ without loss of generality. Let $\mathbf{x}_0 \in \mathbb{R}^d$ be a random vector and $\mathbf{x} = \boldsymbol{\mu} + b\mathbf{P}\mathbf{x}_0$ for some $\boldsymbol{\mu} \in \mathbb{R}^d$, $b \in \mathbb{R}$ and an orthogonal matrix $\mathbf{P} \in \mathbb{R}^{d\times d}$. Then, (i) if $\{\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_k\}$ is a set of self-consistent points of $\mathbf{x}_0$, the set of self-consistent points of $\mathbf{x}$ is given by $\{\boldsymbol{\mu} + b\mathbf{P}\boldsymbol{\gamma}_j, j \in \{1, \dots, k\}\}$, and (ii) if $\Xi_0 = \{\boldsymbol{\xi}_j'\}$ is a set of MSE-RPs of $\mathbf{x}_0$ for $j \in \{1, \dots, k\}$, then $\Xi = \{\boldsymbol{\xi}_j = \boldsymbol{\mu} + b\mathbf{P}\boldsymbol{\xi}_j'\}$ is the set of MSE-RPs of $\mathbf{x}$ with $\mathrm{MSE}(\Xi) = b^2\mathrm{MSE}(\Xi_0)$.

The principal subspace theorem is one of the most important results for the study on MSE-RPs, which originally appeared in [42] for the elliptically contoured distributions.

Proposition 1 *([42], Theorem 4.1).* *Suppose $\mathbf{x} \in \mathbb{R}^d$ is a random vector of elliptically contoured distribution with $E(\mathbf{x}) = \mathbf{0}$ and $Cov(\mathbf{x}) = \boldsymbol{\Sigma}$. Let $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_d$ be the eigenvectors of $\boldsymbol{\Sigma}$ with associated eigenvalues $\lambda_1 \geq \cdots \geq \lambda_d$. If a set of MSE-RPs $\{\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_k\}$ of $\mathbf{x}$ spans a subspace $\mathcal{V}$ of dimension q, then $\mathcal{V}$ is spanned by the first q eigenvectors of $\boldsymbol{\Sigma}$.*

Another useful result for self-consistent points of the elliptically contoured distribution is also presented in Tarpey et al. [42].

Table 11
Estimation bias of the estimated mean vector and covariance matrices $\mathbf{\Sigma}_i = \mathbf{U}_i^\top \mathbf{U}_i$, $i \in \{1, 2, 3\}$, for Kotz$_3$(3, 3, 1) by the bootstrap method. $\hat{F}_{\mathrm{MC}}(\mathbf{x})$, $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ are used for discrete approximation to Kotz$_3$(3, 3, 1). Smallest estimation bias is in bold.

$\mathbf{\Sigma} = \mathbf{\Sigma}_1$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\mathbf{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.339382	**0.011185**	0.017939	2.705093	**0.391274**	0.919508
25	0.073918	0.087890	**0.008259**	0.938121	**0.163319**	0.582056
35	0.251636	0.012450	**0.010119**	0.296953	**0.170563**	0.351828
101	0.030892	0.015529	**0.007249**	0.508193	0.278315	**0.093341**
185	0.174243	0.024130	**0.001357**	0.087749	0.238956	**0.034090**
418	0.081861	0.004841	**0.002727**	0.448038	**0.047386**	0.048232
828	0.115587	0.003883	**0.000307**	0.397930	0.014841	**0.007239**
1459	0.003603	0.005111	**0.001660**	0.182671	0.009133	**0.006990**
2440	0.007819	0.000823	**0.000213**	0.225136	0.035529	**0.005571**
3237	0.022146	0.003785	**0.000068**	0.158965	0.004412	**0.003260**
$\mathbf{\Sigma} = \mathbf{\Sigma}_2$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\mathbf{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.721514	0.008834	**0.001261**	1.986047	**0.090411**	0.258858
25	0.226749	0.028999	**0.014650**	0.378748	**0.029228**	0.174074
35	0.201690	0.025219	**0.006142**	0.341458	**0.043095**	0.111629
101	0.236874	0.001123	**0.003710**	0.794988	0.066310	**0.052264**
185	0.115714	0.006730	**0.000727**	0.076345	0.039866	**0.023155**
418	0.175352	**0.000405**	0.000967	0.064786	**0.003232**	0.009732
828	0.064103	0.002092	**0.000087**	0.166602	0.005807	**0.003575**
1459	0.030388	0.003551	**0.001631**	0.140709	0.002446	**0.001854**
2440	0.007688	0.001285	**0.000575**	0.056606	0.008328	**0.001932**
3237	0.071739	0.002000	**0.000409**	0.022251	0.002360	**0.000556**
$\mathbf{\Sigma} = \mathbf{\Sigma}_3$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\mathbf{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.590542	0.026633	**0.004299**	0.757688	**0.362763**	0.463363
25	0.306403	0.068660	**0.004390**	0.561337	**0.107533**	0.316966
35	0.355764	0.027596	**0.002901**	1.175356	**0.126148**	0.253164
101	0.301302	0.004995	**0.003495**	0.440069	**0.082796**	0.099135
185	0.269042	0.015239	**0.000977**	0.146629	0.063476	**0.062544**
418	0.082881	0.003072	**0.001442**	0.254755	**0.007814**	0.037470
828	0.053578	0.001871	**0.000539**	0.047664	**0.004239**	0.022205
1459	0.039575	0.004632	**0.000961**	0.215437	**0.004571**	0.015924
2440	0.090849	0.001651	**0.000810**	0.069431	**0.008963**	0.010297
3237	0.022392	0.000784	**0.000190**	0.053341	**0.004153**	0.008442

Table 12
The number of smallest estimation bias of estimated mean vector and covariance matrix by the bootstrap method in Table 11 for each type of RPs.

$\mathbf{\Sigma} = \mathbf{\Sigma}_i$	Estimation of $\boldsymbol{\mu}$			Estimation of $\mathbf{\Sigma}$		
	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
$\mathbf{\Sigma}_1$	0	1	9	0	4	6
$\mathbf{\Sigma}_2$	0	1	9	0	4	6
$\mathbf{\Sigma}_3$	0	0	10	0	9	1
Total	0	2	28	0	17	13

Proposition 2 (42). *Let $\mathbf{x} \in \mathbb{R}^d$ be a random vector of elliptically contoured distribution with $\boldsymbol{\mu} = E(\mathbf{x})$ and $\mathbf{\Sigma} = Cov(\mathbf{x})$. The diagonal decomposition of $\mathbf{\Sigma}$ is given by $\mathbf{\Sigma} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}^\top$, where $\mathbf{B} \in \mathbb{R}^{d\times d}$ is an orthogonal matrix which consists of the d eigenvectors of $\mathbf{\Sigma}$ and the corresponding d eigenvalues are in the diagonal matrix $\mathbf{\Lambda}$. Suppose the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ span the same subspace as the q eigenvectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_q$. Let $\mathbf{B}_1 \in \mathbb{R}^{d\times q}$ consist of the q eigenvectors, and $\mathbf{x}' = \mathbf{B}_1^\top \mathbf{x}$. If $\mathbf{v}_j' = \mathbf{B}_1^\top \mathbf{v}_j$ are self-consistent points of $\mathbf{x}'$ for $j \in \{1, \ldots, k\}$, then $\mathbf{v}_j$ are self-consistent points of $\mathbf{x}$.*

The principal subspace theorem implies that a set of MSE-RPs of $\mathbf{x}$ having a elliptically contoured distribution is in the linear subspace that is spanned by some eigenvectors of Cov($\mathbf{x}$). For a d-dimensional random vector $\mathbf{x}$ of a distribution $F(\mathbf{x})$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, the single MSE-RP of $\mathbf{x}$ is $\boldsymbol{\xi} = \boldsymbol{\mu}$ with MSE($\boldsymbol{\xi}$) = trace($\mathbf{\Sigma}$). When the number of MSE-RPs is $k = 2$, Flury [15] proved the analytical results of MSE-RPs for elliptically contoured distribution as follows.

Proposition 3 (15). *Let $\mathbf{x} \in \mathbb{R}^d$ be a random vector following an elliptically contoured distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$, and let $\boldsymbol{\beta}_1$ be the associated normalized eigenvector of the largest eigenvalue of $\mathbf{\Sigma}$. The two MSE*

Table 13
Estimation bias of the estimated mean vector and covariance matrices for $SN_3(\mathbf{0}, \mathbf{G}, (2,3,1)^\top)$ by the bootstrap method, where $\mathbf{G} = \boldsymbol{\Sigma}_i = \mathbf{U}_i^\top \mathbf{U}_i$, $i \in \{1,2,3\}$. $\hat{F}_{MC}(\mathbf{x})$, $\hat{F}_{QMC}(\mathbf{x})$ and $\hat{F}_{MSE}(\mathbf{x})$ are used for discrete approximation to $SN_3(\mathbf{0}, \mathbf{G}, (2,3,1)^\top)$. Smallest estimation bias is in bold.

$\mathbf{G} = \boldsymbol{\Sigma}_1$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\boldsymbol{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.169509	0.070159	**0.001665**	0.440627	0.523081	**0.250810**
25	0.254025	0.023229	**0.005527**	0.582180	0.466042	**0.156491**
31	0.382753	0.055064	**0.007285**	0.368922	0.308939	**0.124137**
307	0.030967	0.020740	**0.001416**	0.071353	0.025172	**0.021591**
562	0.069603	0.009183	**0.000667**	0.055223	0.026068	**0.016563**
701	0.013907	0.006590	**0.000480**	0.076101	0.078252	**0.013565**
1019	0.020837	0.004922	**0.000636**	0.091190	0.019389	**0.009631**
2129	0.024097	0.001941	**0.001034**	0.021186	0.009948	**0.006082**
3001	0.014637	0.000993	**0.000588**	0.024313	0.006227	**0.004081**
4001	0.005070	0.001997	**0.000785**	0.025391	0.004994	**0.003832**
$\mathbf{G} = \boldsymbol{\Sigma}_2$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\boldsymbol{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.449962	0.085720	**0.008419**	1.201827	0.680859	**0.161385**
25	0.464787	0.022741	**0.000510**	0.511655	0.377068	**0.081532**
31	0.216299	0.024945	**0.000724**	0.537853	0.472456	**0.058401**
307	0.072198	0.017018	**0.003739**	0.198651	0.026776	**0.011563**
562	0.019458	0.009303	**0.002093**	0.031125	**0.006181**	0.010149
701	0.010601	0.006691	**0.001769**	0.130132	0.019875	**0.005393**
1019	0.057499	0.005709	**0.001052**	0.029475	0.024653	**0.001789**
2129	0.047615	0.001996	**0.000658**	0.043335	0.010142	**0.001982**
3001	0.001911	0.000743	**0.000562**	0.052557	0.007514	**0.001208**
4001	0.006454	0.003519	**0.000374**	0.031507	0.004526	**0.001172**
$\mathbf{G} = \boldsymbol{\Sigma}_3$	Estimation bias of mean vector $\boldsymbol{\mu}$			Estimation bias of covariance matrix $\boldsymbol{\Sigma}$		
k	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
15	0.467046	0.123618	**0.002434**	0.640243	0.428801	**0.306574**
25	0.134477	0.066911	**0.007946**	0.396748	0.511097	**0.233344**
31	0.142672	0.073466	**0.010198**	0.879896	0.320486	**0.203242**
307	0.157332	0.022044	**0.001468**	0.104098	0.038326	**0.036394**
562	0.079122	0.012673	**0.001882**	0.115173	0.027061	**0.025532**
701	0.049189	0.009260	**0.000397**	0.109363	0.088129	**0.019277**
1019	0.057299	0.006725	**0.000304**	0.064895	0.027700	**0.018006**
2129	0.021702	0.001779	**0.001540**	0.038961	0.012144	**0.011136**
3001	0.031076	0.001767	**0.000468**	0.011745	**0.006485**	0.008143
4001	0.021988	0.002057	**0.000382**	0.070026	**0.004224**	0.005844

Table 14
The number of smallest estimation bias of estimated mean vector and covariance matrix by the bootstrap method in Table 13 for each type of RPs.

$\mathbf{G} = \boldsymbol{\Sigma}_i$	Estimation of $\boldsymbol{\mu}$			Estimation of $\boldsymbol{\Sigma}$		
	MC-RPs	QMC-RPs	MSE-RPs	MC-RPs	QMC-RPs	MSE-RPs
$\boldsymbol{\Sigma}_1$	0	0	10	0	0	10
$\boldsymbol{\Sigma}_2$	0	0	10	0	1	9
$\boldsymbol{\Sigma}_3$	0	0	10	0	2	8
Total	0	0	30	0	3	27

representative points of $\mathbf{x}$ *are* $\boldsymbol{\xi}_1 = \boldsymbol{\mu} - \gamma\boldsymbol{\beta}_1$, $\boldsymbol{\xi}_2 = \boldsymbol{\mu} + \gamma\boldsymbol{\beta}_1$, *where* $-\gamma$ *and* γ *are the two MSE-RPs of the univariate random variable* $\boldsymbol{\beta}_1^\top(\mathbf{x} - \boldsymbol{\mu})$.

The two MSE-RPs of an elliptically contoured distribution locate on the eigenvector $\boldsymbol{\beta}_1$ associated to the largest eigenvalue λ_1 of the covariance matrix, which is also the direction of the first principal component, and $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are symmetric about $\boldsymbol{\mu}$. The MSE-RPs $\{-\gamma, \gamma\}$ of $\boldsymbol{\beta}_1^\top(\mathbf{x} - \boldsymbol{\mu})$ can be computed by $\gamma = \sqrt{\lambda_1}\mathrm{E}(|\boldsymbol{\beta}_1^\top \boldsymbol{\Sigma}^{-1/2}\mathbf{x}|)$, which follows from the Theorem 1 in [15] for computing two MSE-RPs of univariate symmetric distributions. By (27), the MSE of the two MSE-RPs is $\mathrm{MSE}(\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2\}) = \mathrm{trace}(\boldsymbol{\Sigma}) - \gamma^2$ due to the symmetry of elliptically contoured distribution and $\|\boldsymbol{\beta}_1\| = 1$.

For a more general version of the principal subspace theorem, it can be referred to Theorems 4.1 and 4.2 of [41]. Moreover, the principal subspace theorem has been extended to more distributions, including multivariate location mixtures of spherically symmetric distributions (see [25,26,29,49]) and a mixture of elliptically contoured distribution that form an allometric extension model (see [30]). Tarpey and Loperfido [43] provided a generalized principal subspace theorem for multivariate location mixtures of spherically symmetric distributions based on the notion of self-consistency,

Table 15
Lower bounds σ_0 and ρ_0 for a line pattern of MSE-RPs for bivariate Normal and Kotz type distributions $\mathrm{Kotz}_2(3,6,4)$ with $\boldsymbol{\mu}=0$ and covariance matrices $\boldsymbol{\Sigma}_2(\sigma)$, $\boldsymbol{\Sigma}_2(\rho)$. $C_1(k,\sigma_0)$ and $C_1(k,\rho_0)$ are the contribution rates of $\boldsymbol{\Sigma}_2(\sigma_0(k))$ and $\boldsymbol{\Sigma}_2(\rho_0(k))$, respectively, and $R_1(k,\sigma_0(k))$, $R_1(k,\rho_0(k))$ are the corresponding rates $R_1(k)$ in Eq. (29).

Distribution	k	$\sigma_0(k)$	$C_1(k,\sigma_0(k))$	$R_1(k,\sigma_0(k))$	$\rho_0(k)$	$C_1(k,\rho_0(k))$	$R_1(k,\rho_0(k))$
$\mathrm{N}_2(\mathbf{0},\boldsymbol{\Sigma})$	3	1.5416	0.7038	0.5700	0.4078	0.7039	0.5700
	4	2.1509	0.8223	0.7257	0.6445	0.8223	0.7257
	5	2.6147	0.8724	0.8027	0.7448	0.8724	0.8027
$\mathrm{Kotz}_2(3,6,4)$	3	2.0746	0.8115	0.7230	0.6234	0.8117	0.7232
	4	3.4250	0.9214	0.8668	0.8414	0.9207	0.8652
	5	4.2630	0.9478	0.9120	0.8957	0.9430	0.9093

including the following result for skew-normal distribution, where the skew-normal distribution can be regarded as an infinite location mixture of multivariate standard normal distributions.

Proposition 4. *Let $\mathbf{x}\in\mathbb{R}^d$ be a random vector with skew-normal distribution with location parameter $\mathbf{a}=\mathbf{0}$, scale matrix $\mathbf{I}_d+\boldsymbol{\eta}\boldsymbol{\eta}^\top$ and shape parameter $\mathbf{G}^{-1/2}\boldsymbol{\eta}/\sqrt{1+\boldsymbol{\eta}^\top\mathbf{G}^{-1}\boldsymbol{\eta}}$, where the dispersion matrix $\mathbf{G}=\mathbf{I}_d$ with skewness vector $\boldsymbol{\eta}\in\mathbb{R}^d$. If a set of k MSE-RPs of $\mathbf{x}$ lie on a line, then the k MSE-RPs lie on the line spanned by $\boldsymbol{\eta}$.*

We may see that the analytical solutions of MSE-RPs of size $k=1$ and $k=2$ follow the principal component theorem, as the MSE-RPs are in the linear subspace spanned by the eigenvector $\boldsymbol{\beta}_1$. However, so far there is no analytical solution to the MSE-RPs for $k\geq 3$, and the principal component theorem is useful for constructing the approximation of MSE-RPs. Flury [15,16] and Tarpey et al. [42] consider the patterns of MSE-RPs of the elliptically contoured distribution for $k\in\{3,4,5\}$. A more detailed discussion for the patterns of self-consistent points for a symmetric multivariate distribution is given by Tarpey [39], especially for the bivariate normal distribution. As we mentioned in Section 1, the covariance matrix considered in the numerical results of [39] is $\boldsymbol{\Sigma}_2(\sigma)$. When the value σ in $\boldsymbol{\Sigma}_2(\sigma)$ satisfies certain condition, the MSE-RPs form different patterns, including line pattern, triangle pattern, product pattern, pentagon pattern, etc. It is more convenient to estimate the MSE-RPs with a particular pattern, so the principal subspace theorem is a powerful tool for generating MSE-RPs when k is small. However, it is difficult to estimate the MSE-RPs when k becomes larger.

For an elliptically contoured distribution, a set of MSE-RPs forming a line pattern are on the direction of the first principal component. Now, we will discuss the relation between the MSE and the contribution of the first principal component. Let $\mathbf{x}\in\mathbb{R}^d$ be a random vector of an elliptically contoured distribution with the mean vector $\boldsymbol{\mu}=\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, and let $\lambda_1\geq\cdots\geq\lambda_d$ be the eigenvalues of $\boldsymbol{\Sigma}$ with the associated normalized eigenvectors $\boldsymbol{\beta}_1,\ldots,\boldsymbol{\beta}_d$. By the principal subspace theorem, the MSE-RPs of $\mathbf{x}$ forming a line pattern are in the subspace spanned by $\boldsymbol{\beta}_1$ of the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{x}$. Suppose $\Xi=\{\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_k\}$ is a set of MSE-RPs of $\mathbf{x}$, and $\Xi^1=\{\xi_1^1,\ldots,\xi_k^1\}$ is a set of MSE-RPs of $\boldsymbol{\beta}_1^\top\boldsymbol{\Sigma}^{-1/2}\mathbf{x}$. When the MSE-RPs of $\mathbf{x}$ form a line, they locate on the direction of $\boldsymbol{\beta}_1$ and can be written as $\boldsymbol{\xi}_j=\sqrt{\lambda_1}\xi_j^1\boldsymbol{\beta}_1$ for $j\in\{1,\ldots,k\}$. Then, the MSE of Ξ can be rewritten as

$$\mathrm{MSE}(\Xi)=\mathrm{trace}(\boldsymbol{\Sigma})-\sum_{j=1}^k\|\boldsymbol{\xi}_j\|^2p_j=\mathrm{trace}(\boldsymbol{\Sigma})-\sum_{j=1}^k\|\sqrt{\lambda_1}\xi_j^1\boldsymbol{\beta}_1\|^2p_j=\mathrm{trace}(\boldsymbol{\Sigma})-\lambda_1\sum_{j=1}^k(\xi_j^1)^2p_j.$$

For the jth principal component, the contribution is $C_j=\lambda_j/\mathrm{trace}(\boldsymbol{\Sigma})$. If we define a ratio $R_1(k)=\mathrm{trace}(\boldsymbol{\Sigma}_\xi)/\mathrm{trace}(\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}_\xi$ is the covariance matrix of the MSE-RPs of $\mathbf{x}$, then we can find a relationship between C_1 and $R_1(k)$, i.e.,

$$R_1(k)=\frac{\mathrm{trace}(\boldsymbol{\Sigma}_\xi)}{\mathrm{trace}(\boldsymbol{\Sigma})}=\left(\sum_{j=1}^k(\xi_j^1)^2p_j\right)\frac{\lambda_1}{\mathrm{trace}(\boldsymbol{\Sigma})}=e_kC_1,$$

where $e_k=1-\mathrm{MSE}(\Xi^1)$. As $\mathrm{trace}(\boldsymbol{\Sigma}_\xi)=\mathrm{trace}(\boldsymbol{\Sigma})-\mathrm{MSE}(\Xi)$ from (28), the ratio $R_1(k)$ can also be expressed as

$$R_1(k)=1-\frac{\mathrm{MSE}(\Xi)}{\mathrm{trace}(\boldsymbol{\Sigma})}=e_kC_1. \tag{29}$$

Then, $R_1(2)=[\lambda_1(\mathrm{E}(|\boldsymbol{\beta}_1^\top\boldsymbol{\Sigma}^{-1/2}\mathbf{x}|))^2]/\mathrm{trace}(\boldsymbol{\Sigma})$ gives $C_1=1$ for $k=2$. This implies that the two MSE-RPs of $\mathbf{x}$ are always in a linear subspace spanned by $\boldsymbol{\beta}_1$. For a given $k>2$ and a covariance matrix $\boldsymbol{\Sigma}$, the corresponding C_1 can be determined by the known $\mathrm{MSE}(\Xi)$ and e_k. If there is a covariance matrix $\boldsymbol{\Sigma}$ such that a line pattern of k MSE-RPs of an elliptically contoured distribution is optimal, then $\mathrm{MSE}(\Xi)$ is lower than the MSE of self-consistent points forming other patterns. Hence, let $C_1(k)$ be the lower bound of C_1 such that all the k MSE-RPs locate on a line spanned by $\boldsymbol{\beta}_1$ for given k and $\boldsymbol{\Sigma}$, then we have the value of $C_1(k)$ obtained by $R_1(k)/e_k$ from (29).

Consider a covariance matrix with the special structure $\boldsymbol{\Sigma}_2(\sigma)=\mathrm{diag}(\sigma^2,1)$. Tarpey [39] provided the lower bound $\sigma_0(k)$ of σ such that the k MSE-RPs form a line pattern for the bivariate normal distribution with the mean vector $\boldsymbol{\mu}=\mathbf{0}$ and the covariance matrix $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_2(\sigma)$ for $k\in\{3,4,5\}$, as shown in Table 15. For each k, the value of $C_1(k)$

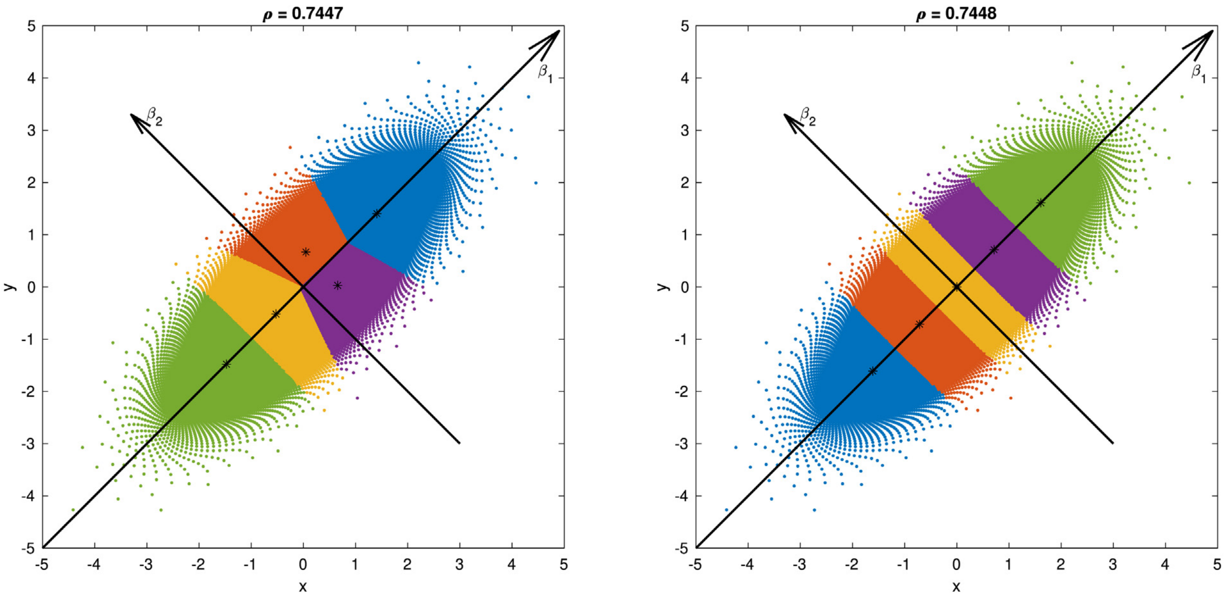

Fig. 2. Optimal pattern of MSE-RPs for the bivariate normal distribution $\mathrm{N}_2(\mathbf{0}, \mathbf{\Sigma}_2(\rho))$ with the number of MSE-RPs $k = 5$. A line pattern is formed when $\rho = 0.7448$ as shown in the right figure.

for $\mathbf{\Sigma}_2(\sigma_0(k))$ is given as well, which is denoted by $C_1(k, \sigma_0(k))$. We also present the values of $\sigma_0(k)$ for $\mathrm{Kotz}_2(3, 6, 4)$ with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}_2(\sigma)$ for more numerical results. Furthermore, we consider the covariance matrix with another structure, $\mathbf{\Sigma}_2(\rho) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. For this type of matrix, there also exists a lower bound $\rho_0(k)$ such that the k MSE-RPs form a line pattern if $\rho \geq \rho_0(k)$. Fig. 2 shows the pattern of MSE-RPs of $\mathrm{N}_2(\mathbf{0}, \mathbf{\Sigma}_2(\rho))$ with $k = 5$ when $\rho = 0.7447$ and $\rho = \rho_0(5) = 0.7448$. The results for $\mathrm{N}_2(\mathbf{0}, \mathbf{\Sigma}_2(\rho))$ and $\mathrm{Kotz}_2(3, 6, 4)$ are also presented in Table 15 with the corresponding value of $C_1(k)$, denoted by $C_1(k, \rho_0(k))$. We can find that, for a same subclass of elliptically contoured distribution, $C_1(k, \sigma_0(k)) \approx C_1(k, \rho_0(k))$. Based on this observation, a conjecture is that the lower bound $C_1(k)$ has a same value for a subclass of elliptically contoured distribution with covariance matrix of any structure, and the set of k MSE-RPs is given by $\{\boldsymbol{\xi}_j = \sqrt{\lambda_1}\xi_j^1\boldsymbol{\beta}_1, j \in \{1, \ldots, k\}\}$ if $C_1 \geq C_1(k)$. Due to the relationship in (29), we can also observe that $R_1(k, \sigma) \approx R_1(k, \rho)$. The results in Table 15 also provide some instruction for the principal component analysis. When using principal component analysis, we always assume that if most (e.g., 80% or 90%) of the population variance can be attributed to the first q principal components, then these components can replace the original variables without losing too much information. However, the choice of the contribution rate is always based on experience, without clear guidance or explanation. From aspect of MSE-RPs, we can try to determine the number q of the principal components used to replace the variables of the original data based on the dimension of the subspace spanned by the MSE-RPs. For example, let $C_1(k) = C_1(k, \sigma_0(k)) = C_1(k, \rho_0(k))$ for convenience, then the values $C_1(k)$ in Table 15 can be regarded as the reference values for the contribution rate to determine whether we can only use the first principal component of the underlying distribution.

6. Conclusions and further study

In this paper, we have discussed some applications of representative points of multivariate distributions in statistical inference. We focus on the elliptically contoured distributions and skew-normal distribution in our study. We consider three types of representative points (MC-RPs, QMC-RPs and MSE-RPs) for a given continuous distribution $F(\mathbf{x})$ and each of them is generated by a different method. The MSE-RPs are computed by k-means based methods including NTLBG and k-means++ algorithm. These three types of representative points can be applied to construct the distribution approximations of $F(\mathbf{x})$, which are $\hat{F}_{\mathrm{MC}}(\mathbf{x})$, $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$. In the two case studies of geometric probability problems, QMC-RPs and MSE-RPs are used to construct a net which consists of k points uniformly distributed on a closed and bounded domain D with good uniformity, and the corresponding discrete distribution $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ are regarded as approximations of the uniform distribution on D. The QMC-RPs provided a better estimation of overlapping area of a fixed circle and the union of two random circles, while both QMC-RPs and MSE-RPs have a good performance in the problem of estimating area of a random belt on the surface of a unit sphere.

When the three kinds of discrete approximations are employed for estimating the mean vector and the covariance matrix of $\mathbf{x}$ by the bootstrap method, both of $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ estimates with higher accuracy, compared with $\hat{F}_{\mathrm{MC}}(\mathbf{x})$.

A lower estimation bias for the mean vector is provided by $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ in most cases, whereas the performance of $\hat{F}_{\mathrm{QMC}}(\mathbf{x})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x})$ for estimating the covariance matrix depends on the underlying distribution. In addition, the relationship between MSE-RPs and principal component analysis is also discussed.

There is a big potential of applications of QMC-RPs and MSE-RPs in statistics. When the parameter $\boldsymbol{\theta}$ of the underlying distribution $F(\mathbf{x};\boldsymbol{\theta})$ is unknown, we might use some kind of estimate $\hat{\boldsymbol{\theta}}$ and construct approximate distributions $\hat{F}_{\mathrm{QMC}}(\mathbf{x};\hat{\boldsymbol{\theta}})$ and $\hat{F}_{\mathrm{MSE}}(\mathbf{x};\hat{\boldsymbol{\theta}})$. Resampling techniques can be applied to these two approximate distributions.

CRediT authorship contribution statement

Jun Yang: Methodology, Software, Writing – original draft, Visualization, Investigation. **Ping He:** Methodology, Supervision, Validation, Writing – reviewing and editing, Funding acquisition. **Kai-Tai Fang:** Conceptualization, Methodology, Writing – reviewing & editing.

Acknowledgments

The authors would thank Dr. Dietrich von Rosen and Dr Tõnu Kollo for their valuable comments. This work was partially supported by the Beijing Normal University-Hong Kong Baptist University United International College (UIC) Internal Research Grant (No. R201912 and No. R202010).

References

[1] R.B. Arellano-Valle, H. Bolfarine, V.H. Lachos, Skew-normal linear mixed models, J. Data Sci. 3 (2005) 415–438.
[2] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, in: SODA'07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007, pp. 1027–1035.
[3] A. Azzalini, A. Dalla Valle, The multivariate skew-normal distribution, Biometrika 83 (1996) 715–726.
[4] G. Callegaro, L. Fiorin, M. Grasselli, Pricing via recursive quantization in stochastic volatility models, Quant. Finance 17 (2017) 855–872.
[5] D.R. Cox, Note on grouping, J. Amer. Statist. Assoc. 52 (1957) 543–547.
[6] I.S. Dhillon, Y. Guan, J. Kogan, Iterative clustering of high dimensional text data augmented by local search, in: 2002 IEEE International Conference on Data Mining, Proceedings, 2002, pp. 131–138.
[7] I.S. Dhillon, D.S. Modha, Concept decompositions for large sparse text data using clustering, Mach. Learn. 42 (2001) 143–175.
[8] B. Efron, Bootstrap methods: Another look at the Jacknife, Ann. Statist. 7 (1979) 1–26.
[9] M.R. El Amri, C. Helbert, O. Lepreux, M.M. Zuniga, C. Prieur, D. Sinoquet, Data-driven stochastic inversion via functional quantization, Stat. Comput. 30 (2020) 525–541.
[10] K.T. Fang, S.D. He, The Problem of Selecting a Given Number of Representative Points in a Normal Population and a Generalized Mills' Ratio, Technical Report No 5, Department of Statistics, Stanford University, 1982.
[11] K.T. Fang, S. Kotz, K.W. Ng, Symmetric Multivariate and Related Distributions, Chapman and Hall, London/New York, 1990.
[12] K.T. Fang, Y. Wang, Number-Theoretic Methods in Statistics, Chapman and Hall, London, 1994.
[13] K.T. Fang, Y. Wang, P.M. Bentler, Some applications of number-theoretic methods in statistics, Statist. Sci. 9 (1994) 416–428.
[14] K.T. Fang, M. Zhou, W. Wang, Applications of the representative points in statistical simulations, Sci. China Math. 57 (2014) 2609–2620.
[15] B.A. Flury, Principal points, Biometrika 77 (1990) 33–41.
[16] B.D. Flury, Estimation of principal points, J. R. Stat. Soc. Ser. C. Appl. Stat. 42 (1993) 139–151.
[17] A. Gersho, R.M. Gary, Vector Quantization and Signal Compression, Kluwer Academic Publishers, Boston, 1992.
[18] M.B. Giles, F.Y. Kuo, I.H. Sloan, B.J. Waterhouse, Quasi-monte carlo for finance applications, in: Proceedings of the 14th Biennial Computational Techniques and Applications Conference, CTAC-2008, volume 50 of ANZIAM Journal, 2008, pp. C308–C323.
[19] S. Graf, H. Luschgy, Foundations of Quantization for Probability Distributions, Springer, New York, 2000.
[20] I.G. Graham, F.Y. Kuo, J.A. Nichols, R. Scheichl, C. Schwab, I.H. Sloan, Quasi-Monte Carlo finite element methods for elliptic PDEs with lognormal random coefficients, Numer. Math. 131 (2015) 329–368.
[21] I.G. Graham, F.Y. Kuo, D. Nuyens, R. Scheichl, I.H. Sloan, Quasi-Monte Carlo methods for elliptic PDEs with random coefficients and applications, J. Comput. Phys. 230 (2011) 3668–3694.
[22] M. Griebel, F. Kuo, I. Sloan, The smoothing effect of integration in $\mathbb{R}^d$ and the anova decomposition, Math. Comp. 82 (2013) 383–400.
[23] J.A. Hartigan, Clustering Algorithms, Wiley, New York, 1975.
[24] S. Iyengar, H. Solomon, Selecting representative points in normal populations, in: Recent Advances in Statistics: Papers in Honor of Herman Chernoff on His 60th Birthday, Elsevier, 1983, pp. 579–591.
[25] H. Kurata, On principal points for location mixtures of spherically symmetric distributions, J. Statist. Plann. Inference 138 (2008) 3405–3418.
[26] H. Kurata, D. Qiu, Linear subspace spanned by principal points of a mixture of spherically symmetric distributions, Comm. Statist. Theory Methods 40 (2011) 2737–2750.
[27] V. Lemaire, T. Montes, G. Pagès, New weak error bounds and expansions for optimal quantization, J. Comput. Appl. Math. 371 (2020).
[28] Y. Linde, A. Buzo, R. Gray, An algorithm for vector quantizer design, IEEE Trans. Commun. 28 (1980) 84–95.
[29] S. Matsuura, H. Kurata, A principal subspace theorem for 2-principal points of general location mixtures of spherically symmetric distributions, Statist. Probab. Lett. 80 (2010) 1863–1869.
[30] S. Matsuura, H. Kurata, Principal points for an allometric extension model, Statist. Papers 55 (2014) 853–870.
[31] J. Max, Quantizing for minimum distortion, IRE Trans. Inform. Theo. 6 (1960) 7–12.
[32] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, SIAM, Philadelphia, 1992.
[33] A.B. Owen, Monte Carlo And quasi-Monte Carlo for statistics, in: Monte Carlo and Quasi-Monte Carlo Methods 2008, Springer, Berlin, Heidelberg, 2009, pp. 3–18.
[34] G. Pagès, Introduction to vector quantization and its applications for numerics, ESAIM: Proc. Surv. 48 (2015) 29–79.
[35] G. Pagès, H. Pham, J. Printems, Optimal quantization methods and applications to numerical problems in finance, in: Handbook of Computational and Numerical Methods in Finance, Springer, 2004, pp. 253–297.
[36] S. Paskov, J.F. Traub, Faster valuation of financial derivatives, J. Portfolio Manag. 22 (1995) 113–123.
[37] D. Pollard, Strong consistency of k-means clustering, Ann. Statist. 9 (1981) 135–140.

[38] T. Tarpey, Principal points and self-consistent points of symmetrical multivariate distributions, J. Multivariate Anal. 53 (1995) 39–51.
[39] T. Tarpey, Self-consistent patterns for symmetric multivariate distributions, J. Classification 15 (1998) 57–79.
[40] T. Tarpey, Self-consistency algorithms, J. Comput. Graph. Statist. 8 (1999) 889–905.
[41] T. Tarpey, B. Flury, Self-consistency: A fundamental concept in statistics, Statist. Sci. 11 (1996) 229–243.
[42] T. Tarpey, L. Li, B.D. Flury, Principal points and self-consistent points of elliptical distributions, Ann. Statist. 23 (1995) 103–112.
[43] T. Tarpey, N. Loperfido, Self-consistency and a generalized principal subspace theorem, J. Multivariate Anal. 133 (2015) 27–37.
[44] T. Tarpey, E. Petkova, Principal point classification: Applications to differentiating drug and placebo responses in longitudinal studies, J. Statist. Plann. Inference 140 (2010) 539–550.
[45] T. Tarpey, E. Petkova, R.T. Ogden, Profiling placebo responders by self-consistent partitioning of functional data, J. Amer. Statist. Assoc. 98 (2003) 850–858.
[46] Y. Tashiro, On methods for generating uniform random points on the surface of a sphere, Ann. Inst. Statist. Math. 29 (1977) 295–300.
[47] Y. Wang, K.T. Fang, Number-theoretic method in applied statistics, Chinese Ann. Math. Ser. B 11 (1990a) 31–55.
[48] X. Wang, I.H. Sloan, Quasi-Monte Carlo methods in financial engineering: An equivalence principle and dimension reduction, Oper. Res. 59 (2011) 80–95.
[49] W. Yamamoto, N. Shinozaki, Two principal points for multivariate location mixtures of spherically symmetric distributions, J. Japan Statist. Soc. 30 (2000) 53–63.
[50] Y.D. Zhou, K.T. Fang, A note on statistics simulation for geometric probability problems (in chinese), Sci. Sinica Math. 41 (2011) 253–264.