MDPI

Article

The Representative Points of Generalized Alpha Skew-*t* Distribution and Applications

Yong-Feng Zhou [1,†], Yu-Xuan Lin [2,3,†] , Kai-Tai Fang [4] and Hong Yin [1,*]

1   School of Mathematics, Renmin University of China, No. 59, Zhongguancun Street, Haidian District, Beijing 100872, China; zyfeng177@ruc.edu.cn
2   Research Center for Frontier Fundamental Studies, Zhejiang Lab, Kechuang Avenue, Zhongtai Sub-District, Yuhang District, Hangzhou 311121, China; yuxuanlin@zhejianglab.com
3   Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, Zhuhai 519087, China
4   Department of Statistics and Data Science, Faculty of Science and Technology, BNU-HKBU United International College, 2000 Jintong Road, Tangjiawan, Zhuhai 519087, China; ktfang@uic.edu.cn
*   Correspondence: yinhong@ruc.edu.cn
†   These authors contributed equally to this work.

Abstract: Assuming the underlying statistical distribution of data is critical in information theory, as it impacts the accuracy and efficiency of communication and the definition of entropy. The real-world data are widely assumed to follow the normal distribution. To better comprehend the skewness of the data, many models more flexible than the normal distribution have been proposed, such as the generalized alpha skew-*t* (GAST) distribution. This paper studies some properties of the GAST distribution, including the calculation of the moments, and the relationship between the number of peaks and the GAST parameters with some proofs. For complex probability distributions, representative points (RPs) are useful due to the convenience of manipulation, computation and analysis. The relative entropy of two probability distributions could have been a good criterion for the purpose of generating RPs of a specific distribution but is not popularly used due to computational complexity. Hence, this paper only provides three ways to obtain RPs of the GAST distribution, Monte Carlo (MC), quasi-Monte Carlo (QMC), and mean square error (MSE). The three types of RPs are utilized in estimating moments and densities of the GAST distribution with known and unknown parameters. The MSE representative points perform the best among all case studies. For unknown parameter cases, a revised maximum likelihood estimation (MLE) method of parameter estimation is compared with the plain MLE method. It indicates that the revised MLE method is suitable for the GAST distribution having a unimodal or unobvious bimodal pattern. This paper includes two real-data applications in which the GAST model appears adaptable to various types of data.

Keywords: entropy; generalized alpha skew-*t* distribution; kernel density estimation; maximum likelihood estimation; moments; quasi-Monte Carlo; representative points

1. Introduction

Statistical distributions play a crucial role in information theory since they describe the probability characteristics of data or signals, and hence directly affect the accuracy and efficiency of the representation, transmission, compression, and reconstruction of information. Entropy, as the most important measure in the field of information theory, depends on the statistical distribution of the random variable. In many applications of information theory, it requires the assumption of the statistical distribution of the data. Although assumed to follow the normal distribution in most statistical analyses due to mathematical convenience and generality, real-world data frequently exhibit skewness, leading to the demand for more flexible models. The geometric Brownian motion (GBM) as a popular model of stochastic processes assumes that its solutions follow the log-normal

distribution. Gupta et al. (2024) [1] indicated that the GBM yields trajectories significantly deviated from the reference distribution when the data do not meet the log-normal assumption. To deal with the limitations in such a scenario, some may consider correcting the model as in [1]. Constructing alternative distributions of the normal distribution has been a common concern.

The skew-normal (SN) distribution is an extension of the normal distribution that allows for skewness, capable of modeling asymmetric data. It was first introduced by Azzalini (1985) [2]. If a random variable Z has a probability density function (pdf) given by

$$\phi(z;s) = 2\phi(z)\Phi(sz), \quad z \in \mathcal{R}, \quad s \in \mathcal{R}, \tag{1}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cumulative distribution function (cdf) of the standard normal distribution, then Z follows the SN distribution, denoted as $Z \sim SN(s)$. The parameter s controls the skewness of the distribution. When $s = 0$, the SN distribution reduces to the standard normal distribution. With $s > 0$, the SN distribution is right-skewed, while $s < 0$ implies left skewness.

The skew-t (ST) distribution is an intriguing example among scale mixtures of SN distributions. It was first formulated by Branco and Dey (2001) [3] and later extensively studied by Azzalini and Capitanio (2003) [4]. An ST random variable, $Y \sim ST(s, \nu)$, can be represented as

$$Y = \frac{Z}{\sqrt{V/\nu}}, \tag{2}$$

where $Z \sim SN(s)$ and $V \sim \chi^2_\nu$, i.e., chi-square distribution with ν degrees of freedom, are independent of each other. The moment of Y exists only when the order is less than ν, which is the same condition required as the Student's t-distribution with ν degrees of freedom, denoted by t_ν. The construction method from the SN distribution to the ST distribution is similar to the approach used to derive the Student's t-distribution from the normal distribution. The pdf of the ST distribution is given by

$$f(y;s,\nu) = 2t(y;\nu)T\left(\sqrt{\frac{1+\nu}{y^2+\nu}}sy;\nu+1\right), \quad y \in \mathcal{R}, \quad s \in \mathcal{R}, \tag{3}$$

where $t(\cdot)$ is the pdf of t_ν, and $T(\cdot)$ is the cdf of $t_{\nu+1}$. The parameter ν controls the tail heaviness. As ν approaches infinity, the ST distribution approaches the SN distribution. Lower values of ν result in heavier tails, providing robustness against outliers. Similar to the SN distribution, the parameter s controls the skewness. When $s = 0$, the ST distribution reduces to the Student's t-distribution. Azzalini and Genton (2008) [5] conducted a quite extensive numerical exploration, demonstrating that the ST distribution can adapt well to various empirical problems. They utilized an autoregressive model of order one, $Y(t) = \beta_0 + \beta_1 Y(t-1) + \epsilon(t)$ with $\beta_0 \in \mathcal{R}$ and $|\beta_1| \leq 1$, to fit the 91 monthly interest rates of an Austrian bank. Their results clearly showed that the error components $\epsilon(t)$ have an ST distribution, where the small degrees of freedom parameter signifies heavy tails in the error distribution, allowing the ST model to better manage outliers than the normal distribution. The ST distribution, which combines the characteristics of the Student's t-distribution and the SN distribution, is particularly suitable for the applications in finance that need to model returns with skewness and excess kurtosis, as well as in environmental studies where the focus is on modeling extreme events. Martínez-Flórez et al. (2020) [6] also mentioned other kinds of skew distributions like skew-Student-t distribution, skew-Cauchy distribution, skew-logistic distribution and skew-Laplace distribution. They summarized those distributions as skew-elliptical distributions since those distributions have a unified expression form of the density function as

$$h_Y(y;s) = 2f(y)F(sy), \quad y, s \in \mathrm{R},$$

where $f(\cdot)$ is a symmetric pdf, and $F(\cdot)$ is the corresponding cdf.

Another type of skew distribution is to add a coefficient function with an α argument to the density function. Elal-Olivero (2010) [7] proposed a distribution called alpha-skew-normal (ASN), with a pdf defined as

$$f(x;\alpha) = \frac{(1-\alpha x)^2 + 1}{2 + \alpha^2}\phi(x), \quad x \in \mathcal{R}, \quad \alpha \in \mathcal{R}. \tag{4}$$

If a random variable X has the pdf as (4), we denote it as $X \sim ASN(\alpha)$. This distribution is more flexible than SN and ST distributions since it can be unimodal or multimodal by adjusting the α parameter. When $\alpha = 0$, the ASN distribution reduces to the standard normal distribution, $X \sim N(0,1)$.

Although the ASN distribution is able to model both skew and bimodal data, it has limitations when data have tails thinner or thicker than the normal distribution. In order to fit stock data more accurately, Altun et al. (2018) [8] introduced a new generalized alpha skew-*t* (GAST) distribution combining the approaches of [4,7]. They combined the GAST distribution with the generalized autoregressive conditional heteroskedasticity (GARCH) model to build a new Value-at-Risk (VaR) prediction model for forecasting daily log returns in three years. They compared the failure rates of the GARCH models under different distribution assumptions including normal, Student's *t*, ST and GAST. The results showed that the GAST distribution performs the best in the backtesting. The definition of GAST distribution and its properties with proof will be elaborated in the next section.

For an unknown continuous statistical distribution, an empirical distribution of a random sample is a traditional way to approximate the target distribution. However, it often leads to low accuracy, and hence the support points for the discrete approximation, also known as representative points (RPs), are explored in order to preserve the information of the target distribution as much as possible. Representative points have a big potential for applications in statistical simulation and inference, see Fang and Pan (2023) [9] for a comprehensive review. Various kinds of representative points of different statistical distributions have been explored in the literature. Especially for complex distributions, the study on the representative points is necessary. The concept of representative points is to simplify complex probability distributions with discrete points easier to manipulate, facilitating efficient computations and analyses. These points serve as a finite set that approximates the distribution of a random variable that can be either discrete or continuous and either univariate or multivariate. In this paper, we focus on the study of the representative points of the GAST distribution and applications. We first introduce the concepts of three kinds of RPs here, while the specific construction procedures are included in Section 4 with their applications on the estimation of moments and densities.

There are many existing criteria for choosing RPs of a distribution, such as Monte Carlo RPs (MC-RPs), quasi-Monte Carlo RPs (QMC-RPs) and mean square error RPs (MSE-RPs) that will be introduced as follows. In fact, the Kullback–Leibler (KL) divergence or relative entropy of two probability distributions is a good criterion for this purpose. The entropy has been utilized as a measure of the experimental design, for example, Lin et al. (2022) [10]. Due to computational complexity, entropy is not popularly used in generating RPs in applications. Therefore, in this article, we study MC-RPs, QMC-RPs, and MSE-RPs of the Generalized alpha skew-*t* distribution only.

1.1. Monte Carlo Representative Points

Let X be the population random variable with the cdf $F(x) = P\{X \leq x\}$, $x \in R$. Various Monte Carlo methods provide ways to generate independent identically distributed (i.i.d.) samples $\{x_1, \ldots, x_n\}$ from the population, and $p(x_i) = \frac{1}{n}$, $i = 1, \ldots, n$. The empirical distribution of the random sample is defined as follows:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{x_i \leq x\}},$$

where I_A is the indicator function of A. The empirical distribution $F_n(x)$ should be close to $F(x)$ in the sense of consistency. Hence, $F_n(x)$ can be regarded as an approximation of $F(x)$. We denote this empirical distribution of random samples generated by the Monte Carlo method as F_{MC}. Traditional statistical inference is based on the empirical distribution. Efron (1979) [11] proposed a resampling technique, *the bootstrap* method, with which we can take a set of random samples from F_{MC} instead of F. Combined with bootstrap, the MC-RPs have proven to be useful in statistical inference, such as parameter estimation, density estimation and hypothesis testing. However, the MC method has many limitations since the convergence rate of $F_n(x) \to F(x)$ in distribution as $n \to \infty$, given by $O(\frac{1}{\sqrt{n}})$, is too slow. The following two kinds of RPs improve the convergence rate nicely.

1.2. Quasi-Monte Carlo Representative Points

For a high-dimensional integration problem:

$$I(f) = \int_0^1 \cdots \int_0^1 f(y_1, \ldots, y_d) dy_1 \cdots dy_d = \int_{\mathcal{C}^d} f(\boldsymbol{y}) d\boldsymbol{y},$$

where f is a continuous function on $\mathcal{C}^d = [0,1]^d$. Suppose that $\mathcal{Y} = \{\boldsymbol{y_1}, \ldots, \boldsymbol{y_n}\}$ is a set of n points uniformly scattered in $\mathcal{C}^d$, we can estimate $I(f)$ by

$$\overline{f(\boldsymbol{y})} = \frac{1}{n} \sum_{i=1}^n f(\boldsymbol{y_i}), \quad \boldsymbol{y_i} \in \mathcal{Y}.$$

If we generate $\mathcal{Y}$ by the MC method, the convergence rate of $\overline{f(\boldsymbol{y})} \to I(f)$ is $O(1/\sqrt{n})$ as $n \to \infty$. The quasi-Monte Carlo (QMC) method provides many ways for the construction of $\mathcal{Y}$ to increase the convergence rate. Through the QMC method, the convergence rate can reach $O(n^{-1}\log^d n)$ according to Fang et al. (1994) [12]. For further theory studies, readers can refer to Hua and Wang (1981) [13] and Niederreiter (1992) [14]. In the study of [12], the F-discrepancy is used to measure the uniformity of $\mathcal{Y}$ in $\mathcal{C}^d$, which is defined by

$$D(F, F_n) = \sup_{\boldsymbol{x} \in \mathbb{R}^d} |F(\boldsymbol{x}) - F_n(\boldsymbol{x})|, \tag{5}$$

where $F(\boldsymbol{x})$ is the cdf of uniform distribution $U(C^d)$ and $F_n(x)$ is the empirical distribution of $\mathcal{Y}$. The $\mathcal{Y}$ that minimizes $D(F, F_n)$ is called QMC-RPs which have equal probability $1/n$.

For the univariate distribution of this paper, the QMC method is designed to sample points that are uniformly distributed on the interval $[0,1]$. If the inverse function of F exists, then the set of n points:

$$\left\{ b_j = F^{-1}\left(\frac{2j-1}{2n}\right), j = 1, \ldots, n \right\}, \tag{6}$$

has been proved to have the minimal F-discrepancy of $1/2n$ from $F(x)$ [12]. Therefore, the set of points $\mathcal{B} = \{b_1, \ldots, b_n\}$ is called the QMC-RPs of $F(x)$. Fang et al. (1994) [12] gave a comprehensive study on QMC methods and their applications in statistical inference, experimental design, geometric probability, and optimization.

1.3. Mean Square Error Representative Points

The concept of MSE-RPs was independently proposed by Cox (1957) [15], Flury (1990) [16] and many others. In the literature, "MSE-RPs" have been called by different names, such as "quantized" and "principal points". Let a random variable $X \sim F(x)$ with finite mean μ and variance σ^2. To provide the best representation of F for a given number

n, we select a set of n representative points having the least mean square error from $F(x)$, and form a discrete distribution $F_{MSE}^{(n)}$. Denote $Y_{MSE}^{(n)} \sim F_{MSE}^{(n)}$ defined as

$$F_{MSE}^{(n)}(y) = \sum_{i=1}^{n} p_i^{(n)} I_{\{b_i^{(n)} \leq y\}},$$

with the probability mass function

$$f(y = b_i^{(n)}) = p_i^{(n)}, \quad i = 1, \cdots, n, \tag{7}$$

where $-\infty < b_1^{(n)} < b_2^{(n)} < \cdots < b_n^{(n)} < \infty$ are MSE-RPs of X and $p_1^{(n)}, \cdots, p_n^{(n)}$ are the corresponding probabilities with respect to

$$MSE(Y_{MSE}) = MSE\left(b_1^{(n)}, \cdots, b_n^{(n)}\right) = \int_{-\infty}^{\infty} \min_{i=1,\cdots,n} \left(x - b_i^{(n)}\right)^2 f(x) dx, \tag{8}$$

and

$$p_1^{(n)} = \int_{-\infty}^{(b_1^{(n)} + b_2^{(n)})/2} f(x) dx,$$

$$p_i^{(n)} = \int_{(b_{i-1}^{(n)} + b_i^{(n)})/2}^{(b_i^{(n)} + b_{i+1}^{(n)})/2} f(x) dx, \quad i = 2, \ldots, n-1,$$

$$p_n^{(n)} = \int_{(b_{n-1}^{(n)} + b_n^{(n)})/2}^{+\infty} f(x) dx.$$

The MSE-RPs have many useful properties. Graf and Luschgy (2007) [17], and Fei (1991) [18] proved that

$$E\left[Y_{MSE}^{(n)}\right] = E[X], \quad \lim_{n \to \infty} E\left[(X - Y_{MSE}^{(n)})^2\right] = \lim_{n \to \infty} \left[\mathrm{Var}(X) - \mathrm{Var}(Y_{MSE}^{(n)})\right] = 0. \tag{9}$$

Hence, $Y_{MSE}^{(n)}$ converges to X in distribution.

In this paper, Section 2 begins by reviewing the definition and properties of the GAST distribution. To explore the relationship between the classification of the GAST distribution and the three parameters α, s, ν, we apply the uniform design (Wang and Fang 1981 [19]) to arrange the values of parameter combinations, and then depict the corresponding density plots. Section 2 also classifies the GAST distribution according to the number of peaks in the density function with some proofs. The first four order moments and stochastic representation of the GAST distribution are shown in this section. Section 3 mainly introduces a maximum likelihood estimation (MLE) method with a distribution-free quantile estimator: QMC-MLE (Li and Fang 2024 [20]). In this QMC-MLE method, the estimated quantiles of the sample are used to replace the original sample, and then the MLE is performed on the estimated quantiles to obtain the parameter estimates. We explore the parameter estimation effectiveness of QMC-MLE for small samples by simulation in this section. In order to cover both unimodal and bimodal cases, we choose the GAST distribution with different parameter settings as the underlying distributions. In this section, we find that the effectiveness of QMC-MLE in parameter estimation is influenced by the number of peaks of sample. Section 4 calculates the three types of RPs, MC-RPs, QMC-RPs, and MSE-RPs, of the GAST distribution for different sample size n. For MSE-RPs, the calculation process requires a parametric k-means algorithm (Stampfer and Stadlober 2002 [21]). We will compare the estimates of four statistics (mean, variance, skewness and kurtosis) by the three types of RPs of the underlying distributions. Another application of RPs is density estimation. Section 4 combines the kernel density method (Rosenblatt 1956 [22]) and the three types of RPs to estimate the density of the underlying

GAST distributions. Section 5 applies the RPs to real data samples to show the outstanding performance of MSE-RPs under the assumption of a GAST model.

2. Generalized Alpha Skew-*t* Distribution

In this section, we give the definition of the density function of the GAST distribution (Altun et al., 2018 [8]) and list some of its commonly used subdistributions. We set the parameter values by the uniform design method (Wang and Fang 1981 [19]) to fully demonstrate the influence of parameters on the shape of the density function. Section 2.2 discusses how the parameters influence the number of peaks of density under four conditions. Sections 2.3 and 2.4 give the moments and stochastic representation of the GAST distribution, respectively.

2.1. Definition of the GAST Distribution

Definition 1. *(GAST distribution). A random variable X is said to follow the GAST distribution, denoted as $X \sim GAST(\alpha, s, \nu)$, if it has the following pdf*

$$f(x; \alpha, s, \nu) = \frac{(1 - \alpha x)^2 + 1}{c(\alpha, s, \nu)} t(x; \nu) T\left(\sqrt{\frac{1 + \nu}{x^2 + \nu}} sx; \nu + 1\right), \qquad \nu > 2, \quad x \in \mathcal{R}, \quad (10)$$

where

$$c(\alpha, s, \nu) = 1 - \alpha \left[\frac{s}{\sqrt{1 + s^2}} \left(\frac{\nu}{\pi}\right)^{1/2} \Gamma\left(\frac{\nu - 1}{2}\right) / \Gamma\left(\frac{\nu}{2}\right)\right] + \frac{\alpha^2}{2}\left(\frac{\nu}{\nu - 2}\right). \quad (11)$$

Proposition 1. *If a random variable $Y \sim ST(s, \nu)$, the $c(\alpha, s, \nu)$ in (11) can be written as*

$$c(\alpha, s, \nu) = 1 - \alpha E[Y] + \frac{\alpha^2}{2} E[Y^2]. \quad (12)$$

Proof. We set a random variable $Z \sim SN(\alpha)$. From the Equation (2), the moments of the ST distribution are given by

$$E[Y^m] = \frac{(\nu/2)^{m/2} \Gamma\left(\frac{\nu - m}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} E[Z^m]. \quad (13)$$

Henze (1986) [23] has given the general expression of the odd moments of Z, which is

$$E\left[Z^{2k+1}\right] = \sqrt{\frac{2}{\pi}} \frac{\delta}{(1 + s^2)^k} \frac{(2k + 1)!}{2^k} \sum_{i=0}^{k} \frac{i!(2s)^{2i}}{(2i + 1)!(k - i)!}, \quad k = 0, 1, \cdots, n, \quad (14)$$

where $\delta = s/\sqrt{1 + s^2}$. The even moments coincide with the standard normal distribution, because $Z^2 \sim \chi_1^2$. Hence, the first two moments of ST distribution are, respectively, given by

$$E[Y] = \frac{\left(\frac{\nu}{\pi}\right)^{1/2} \Gamma\left(\frac{\nu - 1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \delta,$$

$$E\left[Y^2\right] = \frac{\nu}{\nu - 2}.$$

Then the Equation (12) is proved. □

The GAST distribution involves several popular useful distributions:

- If $\alpha = 0$, the GAST distribution reduces to the skew-*t* (ST) distribution.
- If $s = 0$, the GAST distribution reduces to the alpha-skew-*t* (AST) distribution.
- If $\alpha = 0$ and $s = 0$, the GAST distribution reduces to the Student's-*t* distribution.
- If $\nu \to \infty$, the GAST distribution reduces to the alpha-skew-normal (ASN) distribution.
- If $\nu \to \infty$, $\alpha = 0$, the GAST distribution reduces to the skew-normal (SN) distribution.

- If $\nu \to \infty$, $\alpha = 0$ and $s = 0$ the GAST distribution reduces to the normal distribution.

In order to depict the GAST densities, especially the characteristics of unimodal or multimodal with different combinations of parameters, the experimental design is used to arrange the parameter values. The uniform design is a number-theoretic method, proposed by Wang and Fang (1981) [19]. As a robust experimental design method, the uniform design has been widely applied in various fields. A uniform design table provides a scientific arrangement of experiments by tabulating the level combinations of factors of interest. Let $U_n(q^s)$ denote a uniform design with n experimental runs and s factors each having q levels. The uniform design table, $U_{16}(16^3)$, adopted in this paper is derived from the website Uniform-Design-Tables (https://fst.uic.edu.cn/isci/research/Uniform_Design_Tables.htm (accessed on 15 September 2024)). In uniform design tables, the levels of factors are labeled by positive integers. For a unit hypercube experimental region $[0,1]^s$, the levels $\{1, 2, \ldots, q\}$ usually take values $\{\frac{1}{2q}, \frac{3}{2q}, \ldots, \frac{2q-1}{2q}\}$. For any hyperrectangle experimental region $[a,b]^s$, a linear transformation $a + (b-a)\frac{i}{2q}$, $i = 1, 3, \ldots, 2q-1$ is applied. Table 1 lists the arrangement of the uniform design table $U_{16}(16^3)$ for the parametric region, $\alpha \times s \times \nu : [-2.6, 3.8] \times [-3, 3.4] \times [2.5, 18.5]$, indicating the 16 kinds of parameter settings.

Figure 1 shows the density plots corresponding to eight parameter settings in Table 1, which are enough to represent the plot of GAST density. As shown in Figure 1, there are four cases in which the pdfs are bimodal and the No. XII and XIV GAST distributions are AST and ST distributions, respectively.

Table 1. The parameter settings according to a uniform design table $U_{16}(16^3)$.

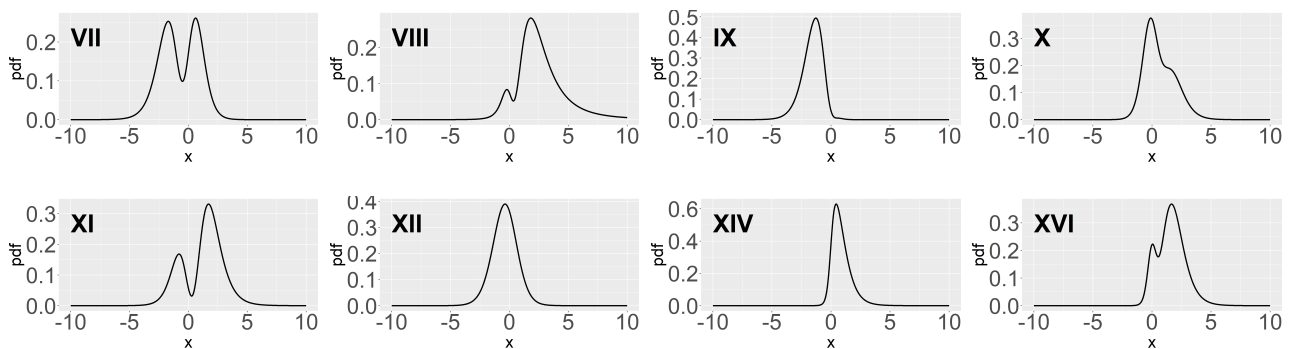| No.	α	s	ν	No.	α	s	ν
I	−2.4	1.6	9	IX	3.2	−2	16
II	−0.8	0.4	3	X	1.2	1.2	15
III	−1.6	−1.6	6	XI	3.6	0.8	11
IV	1.6	−2.4	4	XII	0.4	0	18
V	−1.2	2.4	17	XIII	0.8	−1.2	10
VI	2.4	−0.4	8	XIV	0	3.2	7
VII	−2	−0.8	14	XV	−0.4	−2.8	12
VIII	2.8	2	5	XVI	2	2.8	13

Figure 1. Some plots of GAST densities with parameters in Table 1.

In Section 2.2, we will show how the parameters α, s and ν affect the number of peaks of two special types of the GAST distribution, the AST and ST distributions, leading to the two categories, unimodal and bimodal. The number of peaks in the distribution may affect the parameter estimation. For instance, if the sample size is small and the density function presents a bimodal shape, then the sample is likely to miss the turning points, which will affect the parameter estimation to a certain extent. In addition, the calculation of representative points will also be affected, and the accuracy of derivative density estimation may be reduced.

2.2. Unimodal and Bimodal Properties

Since the density plots of the GAST distribution are varied, we will divide the GAST distribution into two categories: unimodal and bimodal. The number of peaks is determined by the number of zeros of the first derivative of (10). If it has one zero, then the density function is unimodal. If it has three zeros, then the density function is bimodal. To simplify the analysis, we consider the situations under four different parameter combinations of α and s, $(\alpha = 0, s = 0)$, $(\alpha = 0, s \neq 0)$, $(\alpha \neq 0, s = 0)$ and $(\alpha \neq 0, s \neq 0)$. The discussion is as follows:

(1) $\alpha = 0, s = 0$ and $X \sim t(\nu)$

The Student's t-distribution is a well-known unimodal distribution.

(2) $\alpha = 0, s \neq 0$ and $X \sim ST(s, \nu)$

The pdf of $X \sim ST(s, \nu)$, $f(x; s, \nu)$, is given by (3), $\nu > 2$.

Proposition 2. *The skew-t distribution is always unimodal.*

Proof. We derive the first derivative of (3) as follows:

$$f'(x; s, \nu) = 2t'(x; \nu) T\left(sx\sqrt{\frac{\nu+1}{\nu+x^2}}; \nu+1 \right) + 2t(x; \nu) T'\left(sx\sqrt{\frac{\nu+1}{\nu+x^2}}; \nu+1 \right), \quad (15)$$

where

$$t(x; \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} \equiv c\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad (16)$$

$$t'(x; \nu) = -c\left(\frac{2x}{\nu}\right)\left(\frac{\nu+1}{2}\right)\left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}-1} = t(x; \nu)(-x)\left(\frac{\nu+1}{\nu+x^2}\right), \quad (17)$$

$$T'(\cdot) = \left(s\sqrt{\frac{\nu+1}{\nu+x^2}} \right)\left(\frac{\nu}{\nu+x^2} \right) t\left(sx\sqrt{\frac{\nu+1}{\nu+x^2}}; \nu+1 \right). \quad (18)$$

Substituting (17) and (18) in Equation (15), we obtain that

$$f'(x; s, \nu) = 2(-x)t(x; \nu)\left(\frac{\nu+1}{\nu}\right)\left(\frac{\nu}{\nu+x^2}\right)T(\cdot)$$

$$+ 2t(x; \nu)\left(s\sqrt{\frac{\nu+1}{\nu+x^2}} \right)\left(\frac{\nu}{\nu+x^2} \right) t\left(sx\sqrt{\frac{\nu+1}{\nu+x^2}}; \nu+1 \right)$$

$$= 2t(x; \nu)\left(\frac{\nu}{\nu+x^2} \right)\left[(-x)\left(\frac{\nu+1}{\nu}\right)T(\cdot) + \left(s\sqrt{\frac{\nu+1}{\nu+x^2}} \right) t\left(sx\sqrt{\frac{\nu+1}{\nu+x^2}}; \nu+1 \right) \right].$$

As $2t(x; \nu)\left(\frac{\nu}{\nu+x^2}\right) > 0$, the solution to $f'(x; s, \nu) = 0$ can be evaluated by solving the next equation:

$$g(x) = (-x)\left(\frac{\nu+1}{\nu}\right)T(\cdot) + \left(s\sqrt{\frac{\nu+1}{\nu+x^2}} \right) t\left(sx\sqrt{\frac{\nu+1}{\nu+x^2}}; \nu+1 \right) = 0. \quad (19)$$

Since $f(x; s, \nu)$ is symmetric with respect to s:

$$f(x; s, \nu) = f(-x; -s, \nu).$$

The number of peaks is not affected by the sign of s such that we assume $s > 0$. From the expression of (19), we can see that $g(x) > 0$, when $x \in (-\infty, 0]$. When $x \in (0, +\infty)$, we have

$$\left(sx\sqrt{\frac{\nu+1}{\nu+x^2}} \right)' = \left(s\sqrt{\frac{\nu+1}{\nu+x^2}} \right) \left(\frac{\nu}{\nu+x^2} \right) > 0.$$

Hence, $T(\cdot)$ is monotonically increasing, while $t(\cdot)$ is monotonically decreasing. Therefore, we can deduce that $g(x)$ is decreasing when $x \in (0, +\infty)$, i.e., $g(x) \to -\infty$, as $x \to +\infty$. Hence, there is only one solution x_1, s.t. $g(x_1) = 0$. And the ST distribution must be unimodal. $\square$

(3) $\alpha \neq 0, s = 0$ and $X \sim AST(\alpha, \nu)$

The pdf of $X \sim AST(\alpha, \nu)$ is given by

$$f(x; \alpha, \nu) = \frac{(1 - \alpha x)^2 + 1}{2c(\alpha, 0, \nu)} t(x; \nu) \quad, \nu > 2. \tag{20}$$

Proposition 3. *The pdf of the AST distribution, $f(x; \alpha, \nu)$ as (20), is bimodal if and only if $g(x_1) < 0 < g(x_2)$ and $\alpha \notin \left(-\sqrt{\frac{\nu^2-3}{3\nu^2-3\nu}}, \sqrt{\frac{\nu^2-3}{3\nu^2-3\nu}} \right)$, where*

$$g(x) = (1 - \nu)\alpha^2 x^3 + 2\alpha\nu x^2 + (2\alpha^2\nu - 2\nu - 2)x - 2\alpha\nu,$$

$$x_1, x_2 = \frac{-4\alpha\nu \mp \sqrt{\Delta}}{2\alpha^2(3 - 3\nu)}, \quad \Delta = 8\alpha^2 \left[(3\nu^2 - 3\nu)\alpha^2 + (3 - \nu^2) \right]. \tag{21}$$

Otherwise, it is unimodal. It is worth mentioning that a sufficient condition for $f(x; \alpha, \nu)$ to be unimodal is

$$\alpha \in \left(-\sqrt{\frac{\nu^2 - 3}{3\nu^2 - 3\nu}}, \sqrt{\frac{\nu^2 - 3}{3\nu^2 - 3\nu}} \right). \tag{22}$$

Proof. Differentiating (20), we obtain

$$f'(x; \alpha, \nu) = \frac{1}{2c(\alpha, 0, \nu)} \left[(-2\alpha)(1 - \alpha x)t(x; \nu) - \frac{(\nu+1)x}{\nu+x^2} t(x; \nu) \left((1 - \alpha x)^2 + 1 \right) \right].$$

Since $\frac{1}{2c(\alpha, 0, \nu)}$ is a constant, and $t(x; \nu) > 0$, we obtain the equivalent relation expression

$$f'(x; \alpha, \nu) = 0 \iff (-2\alpha)(1 - \alpha x)(\nu + x^2) - (\nu+1)x((1 - \alpha x)^2 + 1) = 0$$

$$\iff g(x) \equiv (1 - \nu)\alpha^2 x^3 + 2\alpha\nu x^2 + (2\alpha^2\nu - 2\nu - 2)x - 2\alpha\nu = 0. \tag{23}$$

Now our problem is transformed into studying the number of zeros of the function $g(x)$. The first derivative of (23) is

$$g'(x) = (3 - 3\nu)\alpha^2 x^2 + 4\alpha\nu x + 2\alpha^2\nu - 2\nu - 2 \quad, \quad \nu > 2.$$

This is a quadratic function with a downward opening. The discriminant Δ of $g'(x)$ is as follows:

$$\Delta = 8\alpha^2 \left[(3\nu^2 - 3\nu)\alpha^2 + (3 - \nu^2) \right].$$

If $\Delta < 0$, namely

$$-\sqrt{\frac{\nu^2 - 3}{3\nu^2 - 3\nu}} < \alpha < \sqrt{\frac{\nu^2 - 3}{3\nu^2 - 3\nu}},$$

then $g(x)$ is monotonically decreasing. Since $\lim_{x\to-\infty} g(x) = +\infty$, $\lim_{x\to+\infty} g(x) = -\infty$, there must be only one root of $g(x) = 0$, i.e., $f(x; \alpha, \nu)$ is unimodal. It is worth mentioning that the parameter setting of No.XII in Table 1 fits this condition.

If $\Delta > 0$, then

$$\exists\, x_1 < x_2, \text{ s.t. } g'(x_1) = g'(x_2) = 0, \text{ where } x_1, x_2 = \frac{-4\alpha\nu \mp \sqrt{\Delta}}{2\alpha^2(3-3\nu)}.$$

We can obtain that

$$\begin{aligned} g'(x) &> 0, \quad x \in (x_1, x_2), \\ g'(x) &< 0, \quad x \in (-\infty, x_1) \cup (x_2, \infty). \end{aligned}$$

Hence, when $x \in (x_1, x_2)$, $g(x)$ is monotonically increasing. When $x \in (-\infty, x_1) \cup (x_2, \infty)$, $g(x)$ is monotonically decreasing. If $g(x_1) < 0 < g(x_2)$, then $g(x)$ has three zeros, and $f(x; \alpha, \nu)$ is bimodal. If condition (22) is met, $f(x; \alpha, \nu)$ is unimodal. To sum up, $f(x; \alpha, \nu)$ is bimodal if and only if $\Delta > 0$, i.e., condition (22) is not satisfied, and $g(x_1) < 0 < g(x_2)$. Otherwise, it is unimodal. $\square$

(4) $\alpha \neq 0, s \neq 0$ and $X \sim GAST(\alpha, s, \nu)$

Differentiating the pdf of the GAST distribution as (10), we obtain

$$f'(x; \alpha, s, \nu) = \frac{t(x; \nu)}{c(\alpha, s, \nu)} \left\{ (-2\alpha)(1-\alpha x)T(\cdot) + \left(\frac{-(\nu+1)x}{\nu+x^2} \right) \left((1-\alpha x)^2 + 1 \right) T(\cdot) \right.$$

$$\left. + \left((1-\alpha x)^2 + 1 \right) \left(s\sqrt{\frac{\nu+1}{\nu+x^2}} \right) \left(\frac{\nu}{\nu+x^2} \right) t\left(sx\sqrt{\frac{\nu+1}{\nu+x^2}}; \nu+1 \right) \right\}.$$

Let

$$\begin{aligned} g(x) =& (-2\alpha)(1-\alpha x)T(\cdot) + \left(\frac{-(\nu+1)x}{\nu+x^2} \right) \left((1-\alpha x)^2 + 1 \right) T(\cdot) + \\ & \left((1-\alpha x)^2 + 1 \right) \left(s\sqrt{\frac{\nu+1}{\nu+x^2}} \right) \left(\frac{\nu}{\nu+x^2} \right) t\left(sx\sqrt{\frac{\nu+1}{\nu+x^2}}; \nu+1 \right). \end{aligned} \tag{24}$$

Then we have $f'(x; \alpha, s, \nu) = \frac{t(x;\nu)}{c(\alpha,s,\nu)} g(x)$, and we obtain

$$f'(x; \alpha, s, \nu) = 0 \iff g(x) = 0.$$

Due to the complexity of the $g(x)$, it is difficult to study its zeros. The discussion of such a situation remains to be studied.

2.3. Moments of the GAST Distribution

From the pdf of the GAST distribution in (10), the kth moment of $X \sim GAST(\alpha, s, \nu)$ is given by

$$E\left[X^k\right] = \frac{E\left[Y^k\right] - \alpha E\left[Y^{k+1}\right] + \frac{\alpha^2}{2} E\left[Y^{k+2}\right]}{c(\alpha, s, \nu)} \equiv \frac{M_k(\alpha, s, \nu)}{c(\alpha, s, \nu)}, \tag{25}$$

where $Y \sim ST(s, \nu)$. Combined with (13) and (14), we have the first four moments of X in which the $M_k(\alpha, s, \nu)$ is given as follows

$$M_1(\alpha, s, \nu) = \delta \left(\frac{\nu}{\pi} \right)^{1/2} \frac{\Gamma \left(\frac{\nu-1}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} - \frac{\alpha \nu}{\nu - 2} + \frac{\alpha^2}{2} \left(\frac{\nu}{2} \right)^{3/2} \frac{\Gamma \left(\frac{\nu-3}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} \sqrt{\frac{2}{\pi}} \left(\frac{3}{1+s^2} \delta + 2\delta^3 \right),$$

$$M_2(\alpha, s, \nu) = \frac{\nu}{\nu - 2} - \alpha \left(\frac{\nu}{2} \right)^{3/2} \frac{\Gamma \left(\frac{\nu-3}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} \sqrt{\frac{2}{\pi}} \left(\frac{3}{1+s^2} \delta + 2\delta^3 \right) + \frac{3\alpha^2 \nu^2}{2(\nu-2)(\nu-4)},$$

$$M_3(\alpha, s, \nu) = \left(\frac{\nu}{2} \right)^{3/2} \frac{\Gamma \left(\frac{\nu-3}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} \sqrt{\frac{2}{\pi}} \left(\frac{3}{1+s^2} \delta + 2\delta^3 \right) - \frac{3\alpha \nu^2}{(\nu-2)(\nu-4)}$$
$$+ \frac{\alpha^2}{2} \left(\frac{\nu}{2} \right)^{5/2} \frac{\Gamma \left(\frac{\nu-5}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} \sqrt{\frac{2}{\pi}} \left(\frac{15}{(1+s^2)^2} \delta + \frac{20}{1+s^2} \delta^3 + 8\delta^5 \right),$$

$$M_4(\alpha, s, \nu) = \frac{3\nu^2}{(\nu-2)(\nu-4)} - \alpha \left(\frac{\nu}{2} \right)^{5/2} \frac{\Gamma \left(\frac{\nu-5}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} \sqrt{\frac{2}{\pi}} \left(\frac{15}{(1+s^2)^2} \delta + \frac{20}{1+s^2} \delta^3 + 8\delta^5 \right)$$
$$+ \frac{15\alpha^2 \nu^3}{2(\nu-2)(\nu-4)(\nu-6)}.$$

2.4. Stochastic Representation the GAST Distribution

Altun (2018) [8] provided a stochastic representation of $X \sim GAST(\alpha, s, \nu)$ as follows.

Theorem 1. *If the random variables $W \sim AST(\alpha, \nu)$ and $Z \sim t(\nu + 1)$ are independent, then we have*

$$W \left| \left\{ \sqrt{\frac{1+\nu}{W^2 + \nu}} sW > Z \right\} \sim GAST(\alpha, s, \nu) \right. \tag{26}$$

According to (26) given by [8], we can generate random samples from the GAST distribution by the following procedure:

Step 1. Generate $W \sim AST(\alpha, \nu)$ and $Z \sim t(\nu + 1)$.
Step 2. If $\sqrt{(1+\nu)/(W^2 + \nu)} sW > Z$, then keep W. Otherwise, go to Step 1.

3. Parameter Estimation

In parameter estimation, the maximum likelihood estimation has been widely utilized because of its transitivity. Let $x = \{x_1, x_2, \ldots, x_n\}$ be a random sample from the $GAST(x; \alpha, s, \nu)$ distribution. The log-likelihood function is given by

$$\ell(\alpha, s, \nu \mid x) = \sum_{i=1}^{n} \log \left[\frac{(1 - \alpha x_i)^2 + 1}{c(\alpha, s, \nu)} \right] + \sum_{i=1}^{n} \log[t(x_i; \nu)] + \\ \sum_{i=1}^{n} \log \left[T \left(\sqrt{\frac{1+\nu}{x_i^2 + \nu}} s x_i; \nu + 1 \right) \right]. \tag{27}$$

By taking the partial derivatives with respect to α, s and ν, we have

$$\frac{\partial \ell}{\partial \alpha} = \sum_{i=1}^{n} \frac{(-2x_i)(1 - \alpha x_i)c(\alpha, s, \nu) - c_\alpha(\alpha, s, \nu)[(1 - \alpha x_i)^2 + 1]}{[(1 - \alpha x_i)^2 + 1]c(\alpha, s, \nu)},$$
$$\frac{\partial \ell}{\partial s} = n \frac{-c_s(\alpha, s, \nu)}{c(\alpha, s, \nu)} + \sum_{i=1}^{n} x_i \sqrt{\frac{\nu+1}{x_i^2 + \nu}} \omega_i^*, \tag{28}$$
$$\frac{\partial \ell}{\partial \nu} = n \frac{-c_\nu(\alpha, s, \nu)}{c(\alpha, s, \nu)} + \sum_{i=1}^{n} \tau_i^* + \sum_{i=1}^{n} \frac{s x_i(x_i^2 - 1)}{2\sqrt{\frac{\nu+1}{x_i^2 + \nu}}(x_i^2 + \nu)^2} \omega_i^*,$$

where

$$
\omega_i^* = \frac{t\left(\sqrt{\frac{1+\nu}{x_i^2+\nu}}sx_i; \nu+1\right)}{T\left(\sqrt{\frac{1+\nu}{x_i^2+\nu}}sx_i; \nu+1\right)}, \qquad \tau_i^* = \frac{t_\nu(x_i;\nu)}{t(x_i;\nu)}.
$$

Remark that $c_\alpha(\alpha,s,\nu)$, $c_s(\alpha,s,\nu)$, $c_\nu(\alpha,s,\nu)$, and $t_\nu(x_i;\nu)$ are the partial derivatives of $c(\alpha,s,\nu)$ and $t(x_i;\nu)$. The solution $(\hat\alpha,\hat s,\hat\nu)$ satisfying $\frac{\partial\ell}{\partial\alpha}=0$, $\frac{\partial\ell}{\partial s}=0$, $\frac{\partial\ell}{\partial\nu}=0$ at the same time is the MLE of (α,s,ν). To solve the system of nonlinear equations in (28), a numerical method is required. In the following subsections, we introduce the algorithm for solving MLE: L-BFGS-B (Byrd et al., 1995 [24]) in Section 3.1. In order to improve estimation accuracy by enhancing sample representativeness, we incorporate a non-parametric quantile estimation method (Harrell and Davis 1982 [25]) introduced in Section 3.2. In Section 3.3, we evaluate the effectiveness of the algorithm and quantile estimation method by simulation. In our study, we use R software version 4.4.1 to conduct simulation.

3.1. L-BFGS-B

L-BFGS-B (Byrd et al., 1995 [24]) is a limited-memory algorithm for solving large nonlinear optimization problems subject to simple bounds on the variables. The essence of the algorithm is a quasi-Newton method. At each iteration, a limited-memory BFGS approximation to the Hessian matrix is updated. This limited-memory matrix is used to define a quadratic model of the objective function, in our study indicating (27). Given a set of samples $\boldsymbol{x}=\{x_i\}_{i=1}^n$, the optimization problem can be formulated as follows:

$$
\max_{\alpha,s,\nu}\ell(\alpha,s,\nu\mid\boldsymbol{x}).
$$

We summarize the procedures of L-BFGS-B as following Algorithm 1.

Algorithm 1 L-BFGS-B for MLE

1: **Input:** Initial guesses for parameters α_0, s_0, ν_0, tolerance ϵ, maximum number of iterations N, bounds $(\alpha_{\min}, s_{\min}, \nu_{\min})$ and $(\alpha_{\max}, s_{\max}, \nu_{\max})$
2: **Output:** Estimated parameters $\hat\alpha, \hat s, \hat\nu$
3: Initialize $k \leftarrow 0$
4: Initialize parameters $\boldsymbol{\theta}^{(0)} \leftarrow (\alpha_0, s_0, \nu_0)$
5: **repeat**
6:     Compute the gradient $\nabla\ell(\boldsymbol{\theta}^{(k)})$
7:     Compute the search direction $\boldsymbol{p}^{(k)}$ using a two-stage approach [24]
8:     Project the search direction $\boldsymbol{p}^{(k)}$ to satisfy the bounds
9:     Line search: find step size $\lambda^{(k)}$ that maximizes $\ell(\boldsymbol{\theta}^{(k)} + \lambda^{(k)}\boldsymbol{p}^{(k)})$
10:    Update parameters: $\boldsymbol{\theta}^{(k+1)} \leftarrow \boldsymbol{\theta}^{(k)} + \lambda^{(k)}\boldsymbol{p}^{(k)}$
11:    $k \leftarrow k+1$
12: **until** $\|\nabla\ell(\boldsymbol{\theta}^{(k)})\| < \epsilon$ or $k \geq N$
13: $(\hat\alpha, \hat s, \hat\nu) \leftarrow \boldsymbol{\theta}^{(k)}$

We chose the L-BFGS-B algorithm because the degree of freedom ν must be greater than 2 for the GAST distribution. If the unconstrained optimization method is used, missing values are likely to appear in the optimization process.

3.2. QMC-MLE

In this subsection, we introduce a method for improving the accuracy of MLE. It is well-known that the accuracy of MLE depends on the sample size to a certain extent. If the sample misses the turning points of the population density, it is less representative, which may lead to lower estimation accuracy. This situation is prone to occur in small sized samples and especially bimodal cases. Fang and Wang (1994) [12] pointed out that

the set of equal quantiles $\{p_i = (2i-1)/2n, i = 1, \ldots, n\}$ has the best representativeness in the sense of *F*-discrepancy. In Section 1, we introduce a QMC method to generate the RPs of a distribution with known parameters. However, for a distribution with unknown parameters, how can we obtain the pth quantile of the distribution *F*? Harrell and Davis (1982) [25] proposed a distribution-free method: the Harrell–Davis (HD) quantile estimator. We use this estimator to calculate the set of equal quantiles of *F*, and then substitute these *n* quantiles into the likelihood function $\ell(\theta \mid x)$ for calculation. Li and Fang (2024) [20] called the MLE method with HD quantile estimator as QMC-MLE, presented below.

Let $x = \{x_1, \ldots, x_n\}$ be a random sample of size *n* from the GAST distribution. Denote $X_{(i)}$ as the ith largest value in x and $F^{-1}(p)$ as the pth population quantile.

Step 1: Generate a set of points uniformly scattered on $(0, 1)$ through

$$p_i = \frac{2i-1}{2n}, \ i = 1, \ldots, n.$$

Step 2: Use the Harrell–Davis quantile estimator to process sample:

$$Q(p_i) = \sum_{i=1}^{n} W_{n,i} X_{(i)},$$

where

$$W_{n,i} = \frac{1}{\beta\{(n+1)p_i, (n+1)(1-p_i)\}} \int_{(i-1)/n}^{i/n} y^{(n+1)p_i - 1} (1-y)^{(n+1)(1-p_i)-1} \, dy$$
$$= I_{i/n}\{p_i(n+1), (1-p_i)(n+1)\} - I_{(i-1)/n}\{p_i(n+1), (1-p_i)(n+1)\},$$

and $I_x\{a, b\}$ denotes the incomplete beta function.

Step 3: Let $z_i = Q(p_i)$, for $i = 1, \ldots, n$. Therefore, the $x = (x_1, \ldots, x_n)$ in the log-likelihood function is replaced by $z = (z_1, \ldots, z_n)$ such that the objective function based on the revised sample is

$$\ell(\theta | z) = \sum_{i=1}^{n} \ln(f(z_i; \alpha, s, \nu)). \tag{29}$$

Step 4: Use the L-BFGS-B algorithm to find the MLE of θ by maximizing (29).

3.3. Simulation

Before the simulation, we introduce four measures of the estimation accuracy: L2.pdf, L2.cdf, absolute bias index (ABI) and Kullback–Leibler (KL) divergence. Denote the true underlying distribution as *F* in cdf or *f* in pdf, and the estimated distribution as $\hat{F}$ or $\hat{f}$. The four measures are defined as follows:

- **L2.pdf** between two densities is defined as

$$L_2\left(f, \hat{f}\right) = \left[\int_{-\infty}^{\infty} \left(f(x) - \hat{f}(x)\right)^2 dx\right]^{1/2};$$

- **L2.cdf** between two cdf's is defined as

$$L_2\left(F, \hat{F}\right) = \left[\int_{-\infty}^{\infty} \left(F(x) - \hat{F}(x)\right)^2 dx\right]^{1/2};$$

- **Absolute bias index** (ABI) is used to evaluate the overall estimation bias in parameters in which $\hat{\mu}$ and $\hat{\sigma}$ denote the estimated expectation and standard deviation of the GAST distribution, defined as

$$ABI = \frac{1}{2}\left(\left|\frac{\mu - \hat{\mu}}{\mu}\right| + \left|\frac{\sigma - \hat{\sigma}}{\sigma}\right|\right);$$

- **Kullback–Leibler (KL)** divergence or the so-called relative entropy is used to measure the difference from one probability distribution to another, defined as follows:

$$D_{KL}(F \| \hat{F}) = \int_{-\infty}^{\infty} f(x) \log\left(\frac{f(x)}{\hat{f}(x)}\right) dx.$$

In the simulation, we generate samples by the inverse transformation method and mainly focus on the small sample case. To study both unimodal and bimodal cases, we choose five parameter settings, No.VII, VIII, IX, X and XI, of the GAST distribution from Figure 1 as the underlying distributions, among which the No.VII, VIII, and XI distributions are bimodal. The sample size n is set to be $25, 50, 100$ and 300. After $N = 100$ times of repetition, the average of $(\hat{\alpha}, \hat{s}, \hat{v})$ is set to be the parameters of the estimated GAST distribution. The precision of the estimates is evaluated by L2.pdf, L2.cdf, ABI and KL, summarized in Table 2, in which "plain" indicates the MLE resulting from the original sample $x = (x_1, \ldots, x_n)$, and "qmc" uses the revised sample $z = (z_1, \ldots, z_n)$.

Table 2. The comparisons between the plain MLE and QMC-MLE in four measures ($n = 25, 50, 100, 300$; $N = 100$).

No.	Method	$n = 25$				$n = 50$				$n = 100$				$n = 300$			
		L2.pdf	L2.cdf	ABI	KL	L2.pdf	L2.cdf	ABI	KL	L2.pdf	L2.cdf	ABI	KL	L2.pdf	L2.cdf	ABI	KL
VII	plain	**0.0326**	0.0469	0.0730	**0.0039**	**0.0133**	**0.0153**	**0.0207**	**0.0012**	**0.0064**	**0.0131**	**0.0260**	**0.0001**	**0.0041**	**0.0056**	0.0098	**0.0001**
VII	qmc	0.0446	**0.0440**	**0.0552**	0.0079	0.0201	0.0218	0.0307	0.0019	0.0163	0.0224	0.0326	0.0008	0.0076	0.0071	**0.0078**	0.0002
VIII	plain	0.0896	0.1161	0.1822	0.0897	0.0575	0.0721	0.1026	0.0256	0.0178	0.0358	0.0809	0.0024	0.0077	0.0186	0.0436	0.0005
VIII	qmc	**0.0670**	**0.0923**	**0.1590**	**0.0830**	**0.0359**	**0.0496**	**0.0749**	**0.0181**	**0.0127**	**0.0286**	**0.0650**	**0.0019**	**0.0058**	**0.0160**	**0.0346**	**0.0004**
IX	plain	**0.0317**	0.0365	0.0321	**0.0073**	0.0185	0.0214	0.0230	0.0034	0.0110	0.0129	0.0172	0.0021	0.0047	0.0049	0.0092	**0.0003**
IX	qmc	0.0302	**0.0363**	**0.0306**	0.0074	**0.0108**	**0.0150**	**0.0160**	**0.0026**	**0.0056**	**0.0082**	**0.0101**	**0.0015**	**0.0029**	**0.0035**	**0.0057**	0.0004
X	plain	**0.0691**	0.0762	0.1443	**0.0097**	0.0453	0.0525	0.0967	0.0056	0.0172	0.0261	0.0525	0.0010	0.0066	0.0125	0.0272	0.0003
X	qmc	0.0570	**0.0663**	**0.1293**	0.0104	**0.0381**	**0.0476**	**0.0886**	**0.0041**	**0.0112**	**0.0207**	**0.0372**	**0.0003**	**0.0058**	**0.0113**	**0.0203**	**0.0002**
XI	plain	0.0689	0.0883	0.1073	**0.0194**	0.0398	0.0487	0.0576	**0.0031**	**0.0134**	**0.0151**	0.0265	**0.0067**	0.0058	**0.0063**	0.0087	**0.0005**
XI	qmc	**0.0563**	**0.0552**	**0.1043**	0.0263	**0.0390**	**0.0342**	**0.0525**	0.0104	0.0279	0.0265	**0.0196**	0.0104	**0.0041**	0.0127	**0.0069**	0.0010

The best performance in the sense of each measure for each pair of distribution type and sample size is highlighted in bold in Table 2. The QMC-MLE method performs better than the plain MLE in most cases, especially for the No.VIII, IX and X distributions. However, for the No.VII and XI distributions, the QMC-MLE has no obvious advantage. The No.IX and X distributions are unimodal, but the No.VIII is bimodal. From the pdf plot of No.VIII distribution, we can see that although it is bimodal, its first peak is not as obvious as the peaks of No.VII and XI distributions. In the pdf plots of No.VII and XI distributions, as x increases, the density function experiences a steep decline after the first peak, while for the No.VIII distribution, the decline lasts only for a short distance before it begins to rise again. Therefore, we have reasons to believe that the QMC-MLE method is more suitable for unimodal functions or bimodal functions of which one peak is not obvious.

In addition, for No.XI GAST distribution, in the sense of KL divergence, the plain MLE is better than the QMC-MLE for all sample sizes. As for the No.XI case under other measures, although the QMC-MLE performs better when $n = 25$ and 50, it becomes less effective for $n = 100$ and 300, which may be caused by the consistency of MLE. According to the discussion above, when we conduct case studies in Section 5, the QMC-MLE will be only used for unimodal samples in parameter estimation, while for bimodal samples, we will use the plain MLE. Nevertheless, this simulation study reveals that the MLE method (both plain and QMC) is appropriate for estimating the GAST parameters due to the small values of four bias measurements.

4. RPs of the GAST Distribution

Recall that in Section 1, we introduced three types of representative points: MC-RPs, QMC-RPs, and MSE-RPs. In this section, we will find these three types of RPs of the GAST distribution for different sample sizes n, and use them to estimate moments and densities in Section 4.1 and Section 4.2, respectively.

4.1. Moment Estimation

For a given n, MC-RPs will be generated by the inverse transformation method. QMC-RPs can be easily obtained by (6) while MSE-RPs are calculated through a parametric k-means algorithm proposed by Stampfer and Stadlober (2002) [21]. We summarize the computation procedure of the k-means algorithm for approximating MSE-RPs of the GAST distribution as follows.

Step 1: For a given pdf $f(x; \alpha, s, \nu)$, the number of RPs: n, and $t = 0$, input a set of initial points $b_1^{(t)} < b_2^{(t)} < \cdots < b_n^{(t)}$. Here we take n QMC-RPs as the initial values. Determine a partition of $\mathcal{R}$ as

$$I_i^{(t)} = (a_{i-1}^{(t)}, a_i^{(t)}], \quad i = 1, \ldots, n-1, \quad I_n^{(t)} = (a_{n-1}^{(t)}, a_n^{(t)}),$$

where

$$a_0^{(t)} = -\infty, \quad a_i^{(t)} = (b_{i-1}^{(t)} + b_i^{(t)})/2, \quad i = 1, \ldots, n-1, \quad a_n^{(t)} = \infty.$$

Step 2: Calculate probabilities

$$p_j^{(t)} = \int_{I_j^{(t)}} f(x; \alpha, s, \nu) dx, \quad j = 1, \ldots, n;$$

and the condition means

$$b_j^{(t+1)} = \frac{\int_{I_j^{(t)}} x f(x; \alpha, s, \nu) dx}{\int_{I_j^{(t)}} f(x; \alpha, s, \nu) dx} = \frac{\int_{I_j^{(t)}} x f(x; \alpha, s, \nu) dx}{p_j^{(t)}}.$$

Step 3: If two sets, $\{b_j^{(t)}\}$ and $\{b_j^{(t+1)}\}$ are identical, the process stops and the outputs $\{b_j^{(t)}\}$ as the MSE-RPs of the distribution with probabilities $\{p_j^{(t)}\}$. Otherwise, let $t := t + 1$ and go back to **Step 1**.

Let Y be a discrete distribution with probability mass function $P(Y = b_j) = p_j$, $j = 1, \ldots, n$, which is an approximate distribution to the GAST distribution. Then, the estimates of mean, variance, skewness and kurtosis can be calculated by

$$E[Y] = \sum_{j=1}^{n} b_j p_j = \mu_Y, \qquad Var[Y] = \sum_{j=1}^{n} (b_j - \mu_Y)^2 p_j = \sigma_Y^2,$$

$$Sk[Y] = \frac{1}{\sigma_Y^3} \sum_{j=1}^{n} (b_j - \mu_b)^3 p_j, \quad Ku[Y] = \frac{1}{\sigma_Y^4} \sum_{j=1}^{n} (b_j - \mu_Y)^4 p_j - 3. \tag{30}$$

We use the No.IX, X and XI as the underlying distributions and consider $n = 10, 20, 30$. It is clear that MC-RPs are random samples of size n. For fair comparisons, we generate N samples of size n and then take the average of the estimated statistics as the results of the MC(N) method. In our study, we choose $N = 10, 100$. The true parameters and four statistics of the three underlying distribution are listed in Table 3. The bias of the estimated results is summarized in Tables 4–6.

Table 3. True parameters and statistics of the underlying distributions.

No.	α	s	v	$E(X)$	**Var**(X)	**Sk**(X)	**Ku**(X)
IX	3.2	−2	16	−1.5997	0.7971	−0.7117	1.4857
X	1.2	1.2	15	0.6272	1.7074	0.6258	0.3101
XI	3.6	0.8	11	1.2301	3.0124	−0.2293	0.0633

Table 4. Estimation bias of four statistics for the No.IX distribution $GAST(3.2, -2, 16)$.

Statistics	Category	10	20	30
Mean	MC(10)	−0.1180	−0.1562	−0.0149
	MC(100)	−0.0291	0.0214	0.0132
	QMC	0.0090	0.0031	0.0014
	MSE	**0.0000**	**0.0000**	**0.0000**
Variance	MC(10)	0.1194	0.3288	0.0769
	MC(100)	−0.0024	−0.0212	−0.0044
	QMC	−0.1455	−0.0889	−0.0668
	MSE	**−0.0231**	**−0.0063**	**−0.0029**
Skewness	MC(10)	0.3486	0.1735	0.1472
	MC(100)	0.3667	0.2676	0.2479
	QMC	0.2740	0.1677	0.1227
	MSE	**0.0259**	**0.0045**	**0.0027**
Kurtosis	MC(10)	−2.4314	−1.3967	−0.7850
	MC(100)	−2.3704	−1.6857	−1.3343
	QMC	−2.0406	−1.6037	−1.3769
	MSE	**−0.5136**	**−0.1855**	**−0.0942**

The best performance within each statistic per sample size is highlighted in bold.

Table 5. Estimation bias of four statistics for the No.X distribution $GAST(1.2, 1.2, 15)$.

Statistics	Category	10	20	30
Mean	MC(10)	−0.0472	0.0170	0.1001
	MC(100)	0.0736	0.0219	−0.0217
	QMC	−0.0102	−0.0051	−0.0035
	MSE	**0.0000**	**0.0000**	**0.0000**
Variance	MC(10)	−0.0684	−0.1739	−0.1644
	MC(100)	0.0757	0.1530	0.0124
	QMC	−0.2179	−0.1213	−0.0861
	MSE	**−0.0394**	**−0.0109**	**−0.0051**
Skewness	MC(10)	−0.6435	−0.1809	−0.0803
	MC(100)	−0.2718	−0.1878	−0.1476
	QMC	−0.1634	−0.1104	−0.0873
	MSE	**−0.0106**	**−0.0034**	**−0.0017**
Kurtosis	MC(10)	−1.4404	−0.5846	−0.5425
	MC(100)	−1.3711	−0.8694	−0.6661
	QMC	−1.0604	−0.7846	−0.6516
	MSE	**−0.2881**	**−0.1005**	**−0.0514**

The best performance within each statistic per sample size is highlighted in bold.

The results indicate that the estimates based on MSE-RPs perform the best for all underlying distributions and sample sizes. The performance of MC-RPs is unstable. Sometimes the average estimates of moments based on MC-RPs are more accurate than those based on QMC-RPs, but in general, appear less effective. In addition, we can observe that with the increase in the number *n*, the overall effect of estimation is better. The estimates of higher-order moments (skewness and kurtosis) are worse than those of lower-order moments (mean and variance).

Table 6. Estimation bias of four statistics for the No.XI distribution $GAST(3.6, 0.8, 11)$.

Statistics	Category	10	20	30
Mean	MC(10)	0.0529	−0.1639	0.0381
	MC(100)	−0.0581	−0.0778	0.0276
	QMC	−0.0189	−0.0082	−0.0032
	MSE	**0.0000**	**0.0000**	**0.0000**
Variance	MC(10)	0.4662	0.1453	−0.3346
	MC(100)	0.2882	−0.0598	−0.0723
	QMC	−0.3569	−0.2171	−0.1635
	MSE	**−0.0705**	**−0.0200**	**−0.0093**
Skewness	MC(10)	−0.2932	−0.1673	−0.0959
	MC(100)	−0.0544	−0.0884	−0.1192
	QMC	−0.1259	−0.0954	−0.0829
	MSE	**−0.0024**	**−0.0051**	**−0.0031**
Kurtosis	MC(10)	−0.9306	−0.7944	−0.6527
	MC(100)	−0.9915	−0.7274	−0.3822
	QMC	−0.9529	−0.7366	−0.6273
	MSE	**−0.3395**	**−0.1294**	**−0.0702**

The best performance within each statistic per sample size is highlighted in bold.

4.2. Kernel Density Estimation

Another application of representative points is density estimation. In the field of signal transmission, the input signal is often converted into discrete data in the transmitter and then reconstructed in the receiver. For a distribution with unknown parameters, how do we use a set of data to construct its overall density function? Here, we introduce a kernel estimation method proposed by Rosenblatt (1956) [22] and Parzen (1962) [26]. Given a fixed number of points $\{x_1, \ldots, x_n\}$ from the original signal, the density estimation of $f(x)$ is given by

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right),$$

where $k(\cdot)$ is the kernel function is the bandwidth and $k_h(y) = \frac{1}{h}k(\frac{y}{h})$. The most popular kernel is the standard normal density function

$$k(x) = \phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}.$$

In our study, we employ the representative points $\{b_1, \ldots, b_n\}$ from the GAST distribution as the samples with their corresponding probabilities $p_i, i = 1, \ldots, n$. The density estimation of $f(x)$ can be extended to

$$\hat{f}_h(x) = \sum_{i=1}^{n} k_h(x - x_i)p_i = \frac{1}{h}\sum_{i=1}^{n} k\left(\frac{x - x_i}{h}\right)p_i.$$

The choice of the bandwidth h is very important. Here, we set a search range $\{0.05, 0.06, \ldots, 1\}$ for h. In the following comparisons, we utilize three types of RPs having sample sizes $n = 10, 20, 30$ for the kernel density estimation of No.IX, X and XI distributions, and evaluate the performances by the minimum L2.pdf between $\hat{f}_h(x; \alpha, s, \nu)$ and $f(x; \alpha, s, \nu)$.

Tables 7–9 show that the kernel density estimation based on MSE-RPs always has the minimum L2.pdf, which decreases as n increases. For the underlying distribution No.IX, we notice that the minimum L2.pdf based on the MSE-RPs with size 10 is only 0.0306, which is even smaller than that based on the QMC-RPs with size 30 (0.0341). Figures 2–4 show the comparing fitting plots of different sets of representative points. It is obvious that the fitting effect increases with n, and the MSE-RPs-based kernel estimation has the best fitting effect, followed by the QMC-RPs-based estimation. It is worth mentioning that for the MC-RPs-based density estimation, due to the randomness of the Monte Carlo method,

the density curve fitted out each time differs greatly, and in many cases, it is not sufficient to reconstruct the original density function.

Table 7. The minimum L2.pdf and the corresponding bandwidth *h* of the kernel density estimation for No.IX distribution.

Method	*n*	*h*	min L2.pdf
MC	10	0.50	0.2108
QMC	10	0.30	0.0583
MSE	10	0.24	**0.0306**
MC	20	0.60	0.1300
QMC	20	0.23	0.0418
MSE	20	0.15	**0.0127**
MC	30	0.48	0.1471
QMC	30	0.21	0.0341
MSE	30	0.13	**0.0098**

The best performance within each sample size is highlighted in bold.
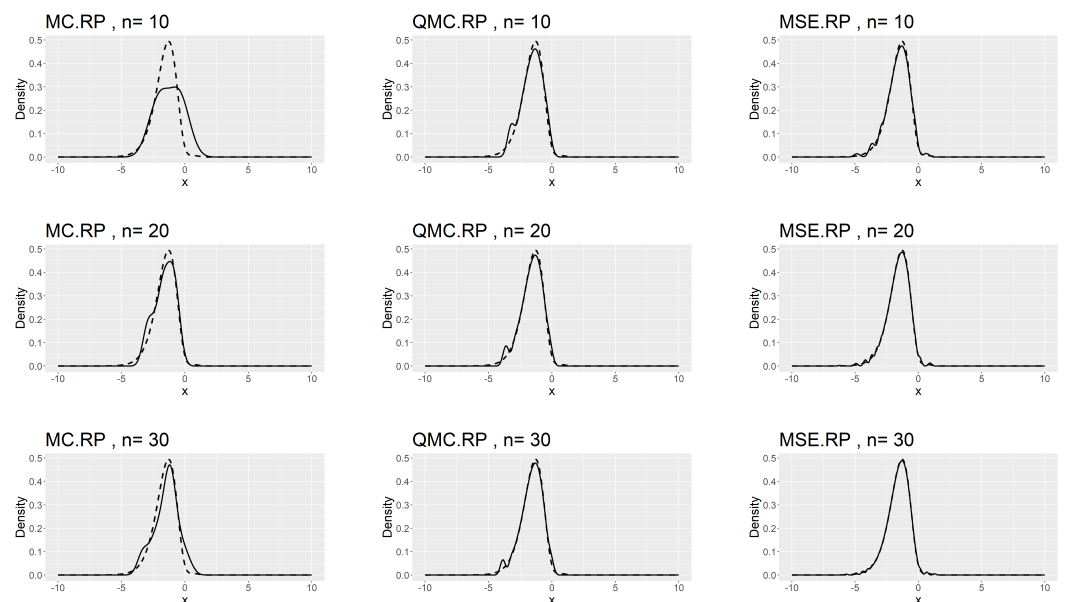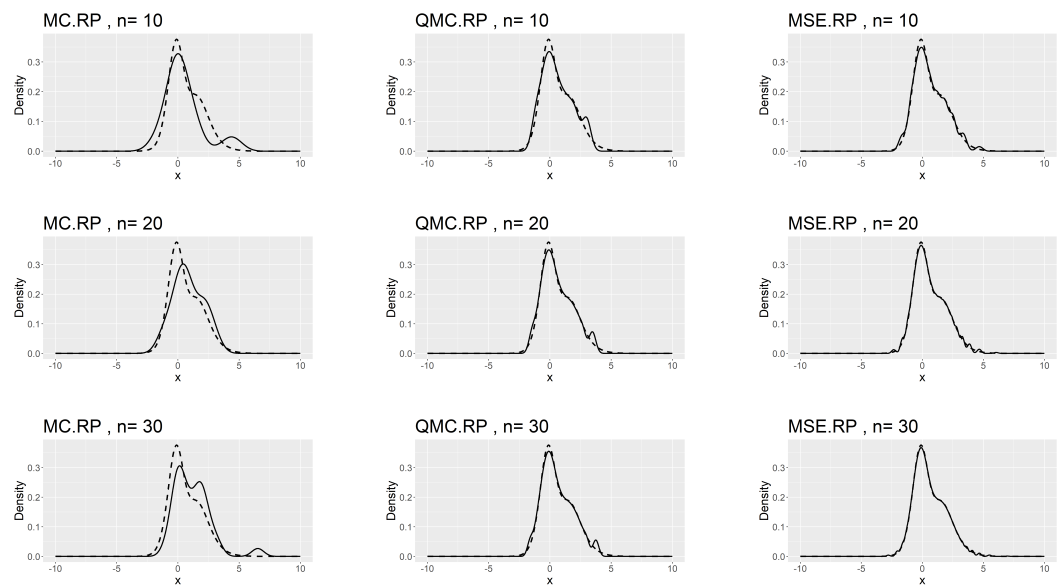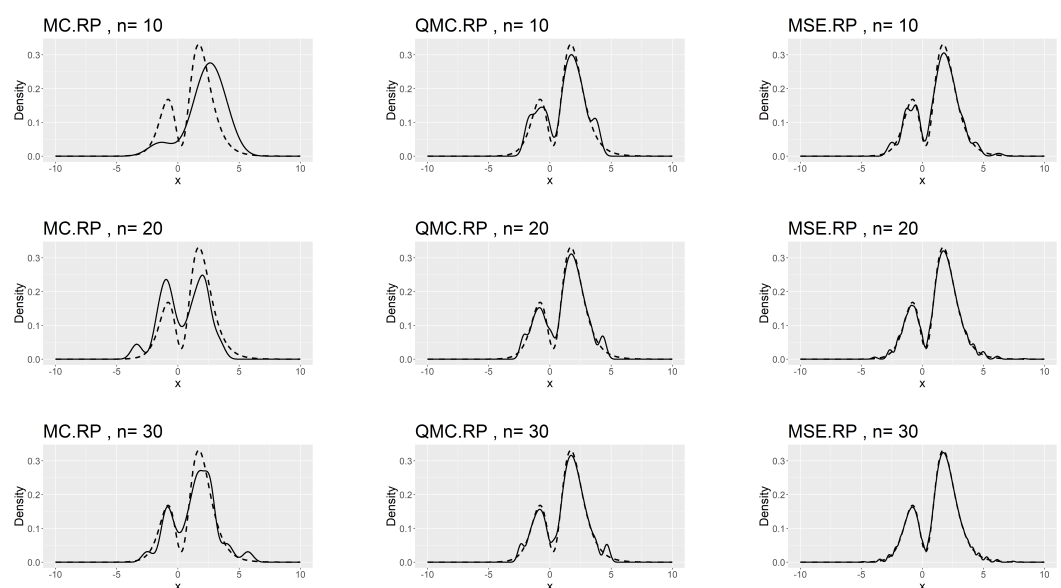

Figure 2. Comparing plots of the fitted densities (in solid lines) by kernel density estimation and the true densities (in dashed lines) for the No.IX distribution.

Table 8. The minimum L2.pdf and the corresponding bandwidth *h* of the kernel density estimation for No.X distribution.

Method	*n*	*h*	min L2.pdf
MC	10	0.90	0.2067
QMC	10	0.36	0.0563
MSE	10	0.29	**0.0305**
MC	20	0.46	0.1504
QMC	20	0.28	0.0379
MSE	20	0.18	**0.0119**
MC	30	0.53	0.1072
QMC	30	0.24	0.0295
MSE	30	0.15	**0.0094**

The best performance within each sample size is highlighted in bold.

Figure 3. Comparing plots of the fitted densities (in solid lines) by kernel density estimation and the true densities (in dashed lines) for the No.X distribution.

Table 9. The minimum L2.pdf and the corresponding bandwidth h of the kernel density estimation for No.XI distribution

Method	n	h	min L2.pdf
MC	10	0.71	0.1607
QMC	10	0.38	0.0628
MSE	10	0.35	**0.0441**
MC	20	0.57	0.1592
QMC	20	0.29	0.0481
MSE	20	0.22	**0.0212**
MC	30	0.76	0.1392
QMC	30	0.25	0.0382
MSE	30	0.17	**0.0137**

The best performance within each sample size is highlighted in bold.

Figure 4. Comparing plots of the fitted densities (in solid lines) by kernel density estimation and the true densities (in dashed lines) for the No.XI distribution.

5. Case Studies

In this section, we will utilize three types of RPs to study real data samples. Before calculating the RPs, we incorporate two additional parameters in the GAST distribution, the location parameter μ and the scale parameter σ, to fit the samples. The pdf is given by

$$f(x; \alpha, s, \nu, \mu, \sigma) = \frac{\left(1 - \alpha\left(\frac{x-\mu}{\sigma}\right)\right)^2 + 1}{\sigma c(\alpha, s, \nu)} t\left(\frac{x - \mu}{\sigma}; \nu\right) T\left(\sqrt{\frac{1 + \nu}{\left(\frac{x-\mu}{\sigma}\right)^2 + \nu}} s\left(\frac{x - \mu}{\sigma}\right); \nu + 1\right),$$

where $c(\alpha, s, \nu)$ is the same as that in Formula (11). For the sample data, we choose both unimodal and bimodal types, which are the O_3 data and the Faithful Geyser data.

5.1. O_3 Data

These data are from the website (https://archive.ics.uci.edu/dataset/360/air+quality (accessed on 15 September 2024)), which contains hourly averaged responses from an Air Quality Chemical Multisensor Device in an Italian city. We selected the "PT08.S5(O_3)" (denoted as "O_3" in this article) data as the study object. After setting the interception time from September 1 to November 30 in 2004, and removing the missing values, we derive 90 observations. We summarize the parameter estimation results of $GAST(\alpha, s, \nu, \mu, \sigma)$ obtained by the QMC-MLE in Table 10, providing the estimated GAST model as follows

$$GAST(-0.1518, -0.2030, 16.9607, 1219.228, 385.0162). \tag{31}$$

Table 10. Parameter estimates of the GAST model based on O_3 data.

Parameters	$\hat{\alpha}$	$\hat{s}$	$\hat{\nu}$	$\hat{\mu}$	$\hat{\sigma}$
QMC-MLE	-0.1518	-0.2030	16.9607	1219.228	385.0162

We present the histogram with the fitted density for O_3 data in Figure 5a. After calculating the $\{p_i = (2i - 1)/2n, i = 1, \ldots, n\}$ quantiles of these data by the HD quantile estimator introduced in Section 3.2, we obtain the associated QQ plot given in Figure 5b. Figure 5 shows the good fitting effect of the GAST model on this unimodal data.
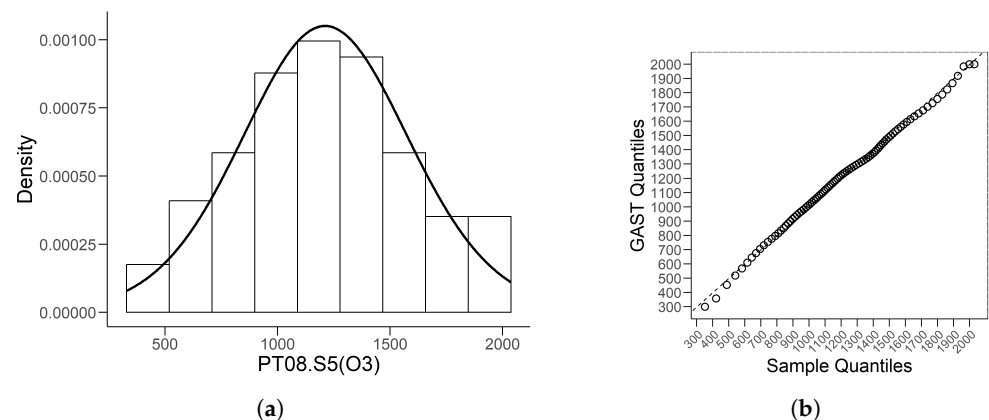

(a)

(b)

Figure 5. (**a**) is the histogram of O_3 data with fitted GAST density line. (**b**) is the associated QQ plot by HD quantile estimator.

The mean, variance, skewness, kurtosis of the distribution (31) are as follows

$$(1209.544, 168359.9, 0.0362, 0.601).$$

We generate MC-RPs, QMC-RPs and MSE-RPs of size 30, from the GAST model (31) using the methods discussed in Section 4.1. Table 11 summarizes the bias of the estimation of the four statistics based on the MC, QMC, and MSE methods.

Table 11. The estimation bias of the four statistics for the fitted GAST model is based on three types of RPs with size 30.

Method	Mean	Variance	Skewness	Kurtosis
MC	39.6251	−46,385.8384	−0.1111	−1.3819
QMC	−0.1250	−10,554.1427	−0.0121	−0.7845
MSE	0.0048	−579.8522	0.0005	−0.0641

Although the bias of the estimated variance in Table 11 is large in value, it is relatively small compared to the true variance of the model, which is 168,359.9. As shown in Table 11, the MSE-RPs estimate the moments of the model more accurately than the other two types of RPs.

The comparisons of the kernel density estimates based on MC, QMC, and MSE RPs are presented in Figure 6.
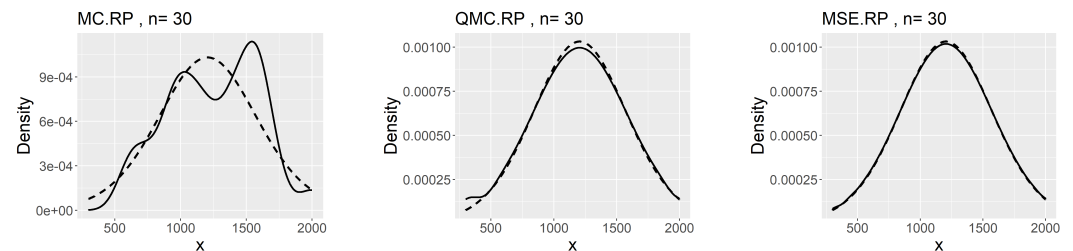
Figure 6. Comparing plots of the fitted densities (in solid lines) by kernel density estimation and the density of the distribution (31).

The corresponding minimum L2.pdf's between the kernel estimates and the density of the model (31) are 0.00237 for MC, 0.00026 for QMC and 0.00025 for MSE. As shown in Figure 6, although the estimated kernel density based on the QMC method is well-fitted, it is not as good as that based on the MSE method at the beginning and at the peak.

5.2. Faithful Geyser Data

The Faithful Geyser Data, a commonly used dataset in R software, is a record of the waiting time between eruptions and the duration time of these eruptions for Old Faithful Geyser in Yellow National Park, Wyoming, USA. In this study, we use the waiting-time samples which include 299 observations.

Since these data are bimodal, we use the plain-MLE to estimate parameters. The results are given in the Table 12, providing the GAST model as

$$GAST(-2.4016, -0.2322, 100, 70.6301, 8.7872).$$ (32)

Table 12. Parameter estimates of the GAST model based on Faithful Geyser data

Parameters	$\hat{\alpha}$	$\hat{s}$	$\hat{v}$	$\hat{\mu}$	$\hat{\sigma}$
Plain-MLE	−2.4016	−0.2322	100	70.6301	8.7872

The histogram with the fitted density for Faithful Geyser data is given in Figure 7a. Denoting $X_{(1)} < \ldots < X_{(n)}$ as the order statistics of this bimodal data, we calculate its quantiles by the traditional estimator:

$$Q_p = (1 - g)X_{(j)} + gX_{(j+1)},$$

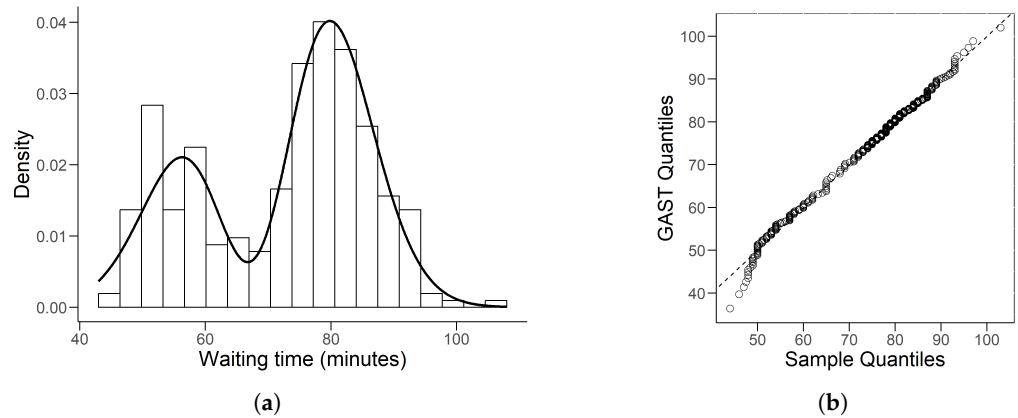where $(n+1)p = j + g$ and j is the integral part of $(n+1)p$. The associated QQ plot is given in Figure 7b.

(a)

(b)

Figure 7. (**a**) is the histogram of Faithful Geyser data with fitted GAST density line. (**b**) is the associated QQ plot by traditional quantile estimator.

Table 12 shows that the estimated ν is 100, which is the upper bound we set. From this point of view, we can assume that $\nu \to \infty$ in this fitting model. As described in Section 2.1, this model is actually a subdistribution of the GAST: ASN, indicating that the GAST model is flexible since it can adapt to different types of data. From the QQ plot in Figure 7, we notice that when data are less than 50, the scatter point deviates far from the line, which can also be observed from Figure 7a. The fitting curve rises slowly at the beginning, so the sample quantiles will be larger than the GAST quantiles. When the data are greater than 50, where more samples are located, this distribution fits the data well. Hence, the estimated GAST model (32) is still acceptable.

The mean, variance, skewness, and kurtosis of the distribution (32) are

$$(72.3205, \ 192.9088, \ -0.4496, \ -0.8031).$$

We generate MC-RPs, QMC-RPs and MSE-RPs of size 50 from the model (32). The estimation biases of the four statistics are summarized in Table 13.

Table 13. The estimation biases of the four statistics for the fitted GAST model based on three types of RPs with size 50.

Method	Mean	Variance	Skewness	Kurtosis
MC	2.8239	−11.0749	0.02917	0.5070
QMC	0.0010	−2.5761	0.0010	−0.0907
MSE	0.0000	−0.1364	−0.0001	−0.0051

We observe that the MSE-RPs have the same mean as the population expectation, which is described in (9). Compared to MC-RPs and QMC-RPs, the MSE method estimates the moments of the model more accurately. The comparison of the kernel density plots is presented in Figure 8.

The corresponding minimum L2.pdf's between the kernel estimates and the density of the model (32) are 0.0041 for MC, 0.0029 for QMC and 0.0027 for MSE. The MSE-RPs still perform the best.
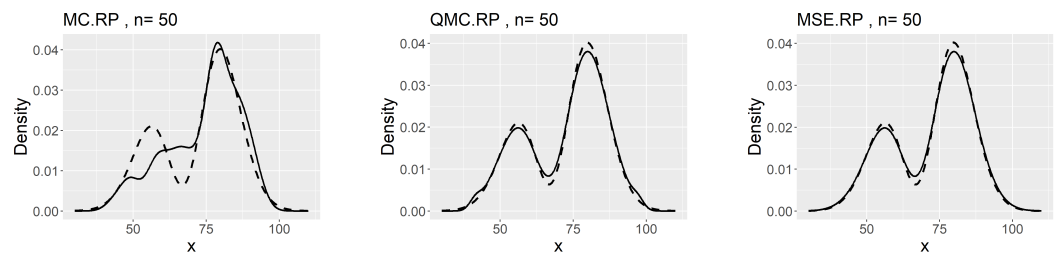
Figure 8. Comparing plots of the fitted densities (in solid lines) by kernel density estimation and the density of the distribution (32).

6. Conclusions

This paper mainly studies different types of representative points of the GAST distribution and the applications of these RPs. The comparative analyses across various sample sizes and both unimodal and bimodal GAST distributions reveal that the RPs obtained by the MSE method consistently outperform the others in the applications of estimating moments and densities. However, the performance on estimating higher-order moments, such as skewness and kurtosis, shows the limitations of RPs on capturing higher-order statistical properties. Therefore, the number of RPs n must adopt a larger value to reduce the bias of higher-order moment estimation. This paper also incorporates QMC-MLE for parameter estimation of the GAST distribution. For unimodal or bimodal data with an unclear peak, the QMC-MLE method improves parameter estimation accuracy. However, in bimodal cases, plain MLE is more effective. Combined with such property, we can model different types of data accordingly.

Author Contributions: Conceptualization, K.-T.F.; Funding acquisition, Y.-X.L.; Methodology, Y.-F.Z. and K.-T.F.; Software, Y.-F.Z.; Writing—original draft, Y.-F.Z.; Writing—review and editing, Y.-X.L., K.-T.F. and H.Y. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: A part of the dataset utilized for case studies in this paper is openly available in UCI Machine Learning Repository at https://archive.ics.uci.edu/dataset/360/air+quality (accessed on 15 September 2024). Another data set for analysis in this paper is obtained from R package *datasets*, named *faithful: Old Faithful Geyser Data*.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Gupta, R.; Drzazga-Szczęśniak, E.; Kais, S.; Szczęśniak, D. The entropy corrected geometric Brownian motion. *arXiv* **2024**, arXiv:2403.06253.
2. Azzalini, A. A class of distributions which includes the normal ones. *Scand. J. Stat.* **1985**, *12*, 171–178.
3. Branco, M.D.; Dey, D.K. A general class of multivariate skew-elliptical distributions. *J. Multivar. Anal.* **2001**, *79*, 99–113. [CrossRef]
4. Azzalini, A.; Capitanio A. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew *t*-distribution. *J. R. Stat. Soc.* **2003**, *65*, 367–389. [CrossRef]
5. Azzalini, A.; Genton, M.G. Robust likelihood methods based on the skew-*t* and related distributions. *Int. Stat. Rev.* **2008**, *76*, 106–129. [CrossRef]
6. Martínez-Flórez, G.; Tovar-Falón, R.; Gómez, H. Bivariate Power-Skew-Elliptical Distribution. *Symmetry* **2020**, *12*, 1327. [CrossRef]
7. Elal-Olivero, D. Alpha-skew-normal distribution. *Proyecciones* **2010**, *29*, 224–240. [CrossRef]

8. Altun, E.; Tatlidil, H.; Ozel, G.; Nadarajah, S. A new generalization of skew-*t* distribution with volatility models. *J. Stat. Comput. Simul.* **2018**, *88*, 1252–1272. [CrossRef]

9. Fang, K.T.; Pan, J. A Review of Representative Points of Statistical Distributions and Their Applications. *Mathematics* **2023**, *11*, 2930. [CrossRef]

10. Lin, Y.X.; Tang, Y.H.; Zhang, J.H.; Fang, K.T. Detecting non-isomorphic orthogonal design. *J. Stat. Plan. Inference* **2022**, *221*, 299–312. [CrossRef]

11. Efron, B. Bootstrap methods: Another look at the jackknife. *Ann. Stat.* **1979**, *7*, 1–26. [CrossRef]

12. Fang, K.T.; Wang, Y.; Bentler, P.M. Some applications of number-theoretic methods in statistics. *Stat. Sci.* **1994**, *9*, 416–428. [CrossRef]

13. Hua, L.K.; Wang, Y. *Applications of Number Theory to Numerical Analysis*; Springer: Berlin/Heidelberg, Germany; Science Press: Beijing, China, 1981.

14. Niederreiter, H. *Random Number Generation and Quasi-Monte Carlo Methods*; Society Industrial and Applied Mathematics (SIAM): Phiadelphia, PA, USA, 1992.

15. Cox, D.R. Note on grouping. *J. Am. Stat. Assoc.* **1957**, *52*, 543–547. [CrossRef]

16. Flury, B.A. Principal points. *Biometrika* **1990**, *77*, 33–41. [CrossRef]

17. Graf, S.; Luschgy, H. *Foundations of Quantization for Probability Distributions*; Springer: Berlin/Heidelberg, Germany, 2007.

18. Fei, R. Statistical relationship between the representative point and the population. *J. Wuxi Inst. Light Ind.* **1991**, *10*, 78–81.

19. Wang, Y.; Fang, K.T. A note on uniform distribution and experimental design. *Kexue Tongbao* **1981**, *6*, 485–489.

20. Li, Y.N.; Fang, K.T. A new approach to parameter estimation of mixture of two normal distributions. *Commun. Stat.-Simul. Comput.* **2024**, *53*, 1161–1187. [CrossRef]

21. Stampfer, E.; Stadlober, E. Methods for estimating principal points. *Commun. Stat.-Simul. Comput.* **2002**, *31*, 261–277. [CrossRef]

22. Rosenblatt, M. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* **1956**, *27*, 832–837. [CrossRef]

23. Henze, N. A probabilistic representation of the 'skew-normal' distribution. *Scand. J. Stat.* **1986**, *13*, 271–275.

24. Byrd, R.H.; Lu, P.; Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208. [CrossRef]

25. Harrell, F.E.; Davis, C.E. A new distribution-free quantile estimator. *Biometrika* **1982**, *69*, 635–640. [CrossRef]

26. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [CrossRef]