Article

New Approaches on Parameter Estimation of the Gamma Distribution

Xiao Ke [1]**, Sirao Wang** [2,3,4]**, Min Zhou** [2,3] **and Huajun Ye** [2,3,*]

[1] College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China
[2] Faculty of Science and Technology, BNU-HKBU United International College, Zhuhai 519087, China
[3] Guangdong Provincial Key Laboratory of Interdisciplinary Research and Application for Data Science, BNU-HKBU United International College, Zhuhai 519087, China
[4] Department of Mathematics, Hong Kong Baptist University, Hong Kong, China
[*] Correspondence: hjye@uic.edu.cn

Abstract: This paper discusses new approaches to parameter estimation of gamma distribution based on representative points. In the first part, the existence and uniqueness of gamma mean squared error representative points (MSE-RPs) are discussed theoretically. In the second part, by comparing three types of representative points, we show that gamma MSE-RPs perform well in parameter estimation and simulation. The last part proposes a new Harrel–Davis sample standardization technique. Simulation studies reveal that the standardized samples can be used to improve estimation performance or generate MSE-RPs. In addition, a real data analysis illustrates that the proposed technique yields efficient estimates for gamma parameters.

Keywords: parameter estimation; gamma distribution; representative points; mean squared error; quantile estimator

MSC: 62F10

1. Introduction

The term representative points (RPs) indicates a set of supporting points with corresponding probabilities, which can be used as the best approximation of a d-dimensional probability distribution. Representative points can be regarded as a discretization of a continuous distribution, and are expected to retain as much information as possible. In the univariate case, X is considered to be a population random variable with cumulative distribution function (cdf) $F(x)$, a discrete random variable Z is defined to approximate X with probability mass function (pmf) by a set of supporting points $z = \{z_1, z_2, \cdots, z_k\}$ ($-\infty < z_1 < z_2 < \cdots < z_k < \infty$) with probabilities $\{p_1, p_2, \cdots, p_k\}$, where $P(Z = z_i) = p_i$ and $\sum_{i=1}^{k} p_i = 1$. In the literature, there are several approaches to choosing the supporting point set z. For example, a set of random samples from $F(x)$ can be viewed as a representative of the distribution; Fang and Wang [1] suggest generating representative points based on the number theoretic method. In 1957, Cox [2] proposes the idea of using mean squared error (MSE) to measure the loss of information from $F(x)$, where

$$\text{MSE}(z) = \text{MSE}(z_1, z_2, \cdots, z_k) = E(\min_{i=1,\cdots,k}(z_i - X)^2) = \int_{-\infty}^{\infty} \min_{i=1,\cdots,k}(z_i - X)^2 \mathrm{d}F(x). \quad (1)$$

The point set $z^{MSE} = \{z_1^{MSE}, \cdots, z_k^{MSE}\}$ such that $\text{MSE}(z)$ arrives its minimum is called the mean squared error representative points (MSE-RPs) of $F(x)$. MSE-RPs are found to have many good properties and have been applied in study fields such as signal compression (Gersho and Gray [3]), numerical integration computation (Pagès [4,5]), simulating stochastic differential equation (Gobet et al. [6]; El Amri et al. [7]), statistical simulation (Fang et al. [8],

Fang et al. [9]) and clothing standard settings (Fang and He [10]; Flury [11]). To compute MSE-RPs for different distributions, effective numerical methods are proposed. Fang-He algorithm (Fang and He [10]) calculates MSE-RPs by solving a system of non-linear equations; Lloyd I algorithm (Lloyd [12]), LBG algorithm (Linde et al. [13]) and Competitive Learning Vector Quantization algorithm (Pagès [5]) obtain MSE-RPs by iterating a long training sequence of data; Tarpey's self-consistency algorithm (Tarpey [14]) brings the idea of k-means algorithm for generating MSE-RPs; Chakraborty et al. [15] provides an accelerate algorithm using Newton's method. When the number of MSE-RPs (k) is large, obtaining MSE-RPs becomes computationally intensive. Fang and He [10] presents some discussion on the optimum choice of k.

Recently, the use MSE-RPs properties for some distributions have been studied in detail, including normal distribution (Fang et al. [8]), mixed normal distribution (Fang et al. [9] and Li et al. [16]), arcsine distribution (Jiang et al. [17]) and exponential distribution (Xu et al. [18]). A general relationship between MSE-RPs and population distribution can be found in the work of Fei [19] and Fang et al. [9]. The study of the gamma distribution's MSE-RPs (gamma MSE-RPs) can be traced back to Fu [20], which discusses the existence of gamma MSE-RPs and establishes an algorithm for computing these points. The gamma distribution is one of the most important distributions in statistics and probability theory, it is worth taking a closer look at gamma MSE-RPs and discovering their merits. The innovations of this paper are listed as follows:

1. New theoretical results prove the uniqueness of gamma MSE-RPs;
2. Gamma MSE-RPs are found to outperform other types of representative points in parameter estimation;
3. A new standardization technique is proposed to improve the estimation performance of random samples from the gamma distribution.

Our discussion will focus on these three perspectives. Section 2 provides some preliminary knowledge of the gamma distribution and different types of representative points for readers to access our content easily. Section 3 gives some theoretical discussion on the existence and uniqueness of gamma MSE-RPs. An algorithm for generating gamma MSE-RPs is recommended. Section 4 compares the performance of three typical gamma representative points in parameter estimation and simulation. The results demonstrate that gamma MSE-RPs take advantage of other representative points in many scenarios. Section 5 introduces a new Harrel–Davis standardization technique. Simulation studies show that the standardized samples have better performances than random samples in estimation and can be used to generate gamma MSE-RPs. Section 6 provides a real clinical data analysis and illustrates that the standardization technique yields efficient estimates for gamma parameters.

2. Preliminaries

2.1. The Gamma Distribution and Gamma MSE-RPs

A gamma-distributed random variable with shape parameter a and rate parameter b is denoted $X \sim Gamma(a, b) \equiv Ga(a, b)$. The corresponding probability density function (pdf) in the shape-rate parametrization is

$$f(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad \text{for } x > 0, \ a, b > 0, \tag{2}$$

where $\Gamma(\cdot)$ is the gamma function. The mean, variance, skewness and kurtosis of X are

$$\mu = \frac{a}{b}, \quad \sigma^2 = \frac{a}{b^2}, \quad \text{Sk}(X) = \frac{2}{\sqrt{a}} \quad \text{and} \quad \text{Ku}(X) = \frac{6}{a} \tag{3}$$

accordingly. Let $z^{MSE} = \{z_1^{MSE}, z_2^{MSE}, \cdots, z_k^{MSE}\}$ be a set of MSE-RPs for $Ga(a, b)$, derive the following intervals

$$I_1 = \left(0, \frac{z_1^{MSE} + z_2^{MSE}}{2}\right), \ I_2 = \left[\frac{z_1^{MSE} + z_2^{MSE}}{2}, \frac{z_2^{MSE} + z_3^{MSE}}{2}\right), \cdots,$$

$$I_{k-1} = \left[\frac{z_{k-2}^{MSE} + z_{k-1}^{MSE}}{2}, \frac{z_{k-1}^{MSE} + z_k^{MSE}}{2}\right), \ I_k = \left[\frac{z_{k-1}^{MSE} + z_k^{MSE}}{2}, +\infty\right) \tag{4}$$

with the corresponding probabilities in these intervals as

$$p_i = \int_{I_i} f(x; a, b)\, dx, \quad i = 1, \cdots, k. \tag{5}$$

Here $f(x; a, b)$ is the pdf in (2).

2.2. Other Types of Representative Points

In addition to MSE-RPs, two other types of representative points are frequently discussed in the literature: Monte Carlo representative points and number theoretic representative points.

(A) *Monte Carlo representative points (MC-RPs)*

MC-RPs are generated by the Monte Carlo method. Consider a random sample $\{x_1, x_2, \cdots, x_k\}$ from the distribution function $F(x)$; this can be treated as a set of MC-RPs, written as $z^{MC} = \{z_1^{MC}, z_2^{MC}, \cdots, z_k^{MC}\}$, where $z_i^{MC} = x_i$ and $p(z_i^{MC}) = 1/k$, $i = 1, \cdots, k$.

(B) *Number theoretic representative points (NT-RPs)*

NT-RPs are determined from the number theoretic method (Fang and Wang [1]). Given an one dimensional interval $(0, 1)$, it is known that point set $\left\{\frac{2i-1}{2k}\right\}$ ($i = 1, \cdots, k$) is uniformly scattered on this interval. Based on the inverse transformation method, points

$$z_i^{NT} = F^{-1}\left(\frac{2i-1}{2k}\right), \ i = 1, \cdots, k$$

are k NT-RPs of $F(x)$. The supporting point set is $z^{NT} = \{z_1^{NT}, z_2^{NT} \cdots z_k^{NT}\}$ with probability $p(z_i^{NT}) = 1/k$, $i = 1, \cdots, k$.

2.3. Harrel-Davis Quantile Estimator

In Harrel and Davis [21], a distribution-free quantile estimator is proposed, which consists of a linear combination of the order statistics admitting a jackknife variance. Let $X_1, X_2, \cdots, X_n$ denote a random sample of size n from $Ga(a, b)$; the pth quantile estimator is

$$Q_p = \frac{1}{\beta\{(n+1)p, (n+1)(1-p)\}} \int_0^1 F_n^{-1}(y) y^{(n+1)p-1} (1-y)^{(n+1)(1-p)-1} dy, \tag{6}$$

where $F_n(x)$ is the empirical distribution function. That is, $F_n(x) = n^{-1}\sum I(X_i \leq x)$, and $I(\cdot)$ is the indicator function of the set A. This method can be used for sample standardization. More details are discussed in Section 5.

3. The Existence and Uniqueness of Gamma MSE-RPs

Let a random variable $X \sim Ga(a, b)$ with $E(X) = \mu$ and $z = \{z_1, z_2, \cdots z_k\}$ ($0 < z_1 < z_2 < \cdots < z_k < \infty$) is the supporting points set of X, to minimize $\text{MSE}(z)$, by taking partial derivative of (1), we have

$$\begin{cases} \int_0^{\frac{1}{2}(z_1+z_2)} (z_1 - x)f(x)dx = 0 \\ \int_{\frac{1}{2}(z_1+z_2)}^{\frac{1}{2}(z_2+z_3)} (z_2 - x)f(x)dx = 0 \\ \cdots\cdots\cdots\cdots \\ \int_{\frac{1}{2}(z_{k-1}+z_k)}^{\infty} (z_k - x)f(x)dx = 0. \end{cases} \tag{7}$$

where $f(x)$ is the pdf of the gamma distribution (2). When $k = 1$, system of Equation (7) has only one equation

$$\int_0^{\infty} (z_1 - x)f(x)dx = 0.$$

Obviously, it has one solution $z_1 = \frac{a}{b} = \mu$, which is the only representative point. When $k \geq 2$, the existence of MSE-RPs is true if the system of Equation (7) has a solution. After several transformations, (7) becomes

$$\begin{cases} (z_1 - \mu)[F(\frac{z_1+z_2}{2}) - F(0)] = -\frac{1}{2b}(z_1 + z_2)f(\frac{z_1+z_2}{2}) \\ (z_2 - \mu)[F(\frac{z_2+z_3}{2}) - F(\frac{z_1+z_2}{2})] = \frac{1}{2b}(z_1 + z_2)f(\frac{z_1+z_2}{2}) - \frac{1}{2b}(z_2 + z_3)f(\frac{z_2+z_3}{2}) \\ \cdots\cdots\cdots\cdots \\ (z_k - \mu)[1 - F(\frac{z_{k-1}+z_k}{2})] = \frac{1}{2b}(z_{k-1} + z_k)f(\frac{z_{k-1}+z_k}{2}), \end{cases} \tag{8}$$

where $F(x)$ is the cdf. Theorem 1 shows that the system of Equation (8) has a solution:

Theorem 1.

1. *For given $z_1 > 0$, equation*

$$(z_1 - \mu)F\left(\frac{z_1 + z_2}{2}\right) = -\frac{1}{2b}(z_1 + z_2)f\left(\frac{z_1 + z_2}{2}\right) \tag{9}$$

   a solution z_2 exists if and only if $z_1 < \mu$.

2. *For given $z_i > z_{i-1} > 0$, $i = 2, \cdots, k - 1$, Equation*

$$(z_i - \mu)\left[F\left(\frac{z_i + z_{i+1}}{2}\right) - F\left(\frac{z_{i-1} + z_i}{2}\right)\right] = \frac{1}{2b}(z_{i-1} + z_i)f\left(\frac{z_{i-1} + z_i}{2}\right)$$
$$- \frac{1}{2b}(z_i + z_{i+1})f\left(\frac{z_i + z_{i+1}}{2}\right), \tag{10}$$

   exists a solution z_{i+1} when $z_{i-1} < z_{i,i-1}$, where $z_{i,i-1}$ is the $i - 1$th representative point in the set of gamma MSE-RPs, which has $k = i$.

3. *For a given $z_{k-1} > 0$, Equation*

$$(z_k - \mu)\left[1 - F\left(\frac{z_{k-1} + z_k}{2}\right)\right] = \frac{1}{2b}(z_{k-1} + z_k)f\left(\frac{z_{k-1} + z_k}{2}\right) \tag{11}$$

   a solution z_k exists.

Theorem 1 guarantees the existence of gamma MSE-RPs. Its proof is provided in Appendix A. For the special case $k = 2$, the existence can be provided by statements 1 and 3 in Theorem 1. Next, we show the uniqueness of gamma MSE-RPs in Theorem 2.

Theorem 2. *Suppose $X \sim Ga(a, b)$. For any $k \in \mathbb{N}^+$, the set of gamma MSE-RPs is unique if $a \geq 1$.*

The proof of Theorem 2 is provided in Appendix A. As a result, these two theorems guarantee the existence and uniqueness of gamma MSE-RPs. Furthermore, throughout this paper, gamma MSE-RPs are generated based on the self-consistency algorithm [22]. The details of this algorithm are provided in Appendix B.

4. Gamma MSE-RPs in Parameter Estimation and Simulation

This section compares the performances of gamma MSE-RPs with other types of representative points, i.e., NT-RPs and MC-RPs, in terms of parameter estimation and simulation. Recall that random variable $X \sim Ga(a, b)$ and Z is a discrete approximation of X. The mean, variance, skewness and kurtosis of Z are

$$E(Z) = \sum_{i=1}^{k} z_i p_i \equiv \mu_z, \quad Var(Z) = \sum_{i=1}^{k} (z_i - \mu_z)^2 p_i \equiv \sigma_z^2,$$

$$Sk(Z) = \frac{1}{\sigma_z^3} \sum_{i=1}^{k} (z_i - \mu_z)^3 p_i, \quad Ku(Z) = \frac{1}{\sigma_z^4} \sum_{i=1}^{k} (z_i - \sigma_z)^4 p_i - 3.$$

By the method of moments, we have

$$\hat{a}_{m_2} = \frac{\mu_z^2}{\sigma_z^2} \text{ and } \hat{b}_{m_2} = \frac{\mu_z}{\sigma_z^2}, \tag{12}$$

which are the point estimators of a and b in $Ga(a, b)$. As Z is a discrete approximation of X, it is expected that the moments of Z and estimates in (12) are close to the moments of X, a and b accordingly. The following theorem shows some connections between gamma MSE-RPs and the corresponding $Ga(a, b)$.

Theorem 3. *Let $X \sim Ga(a, b)$ with $Var(X) = \sigma^2 < \infty$, $\mathbf{z} = \{z_1, z_2, \cdots, z_k\}$ is a set of gamma MSE-RPs of $Ga(a, b)$ with corresponding probabilities in (4); then,*

$$E(Z) = E(X) \text{ and } \lim_{k \to \infty} Var(Z) = Var(X).$$

The proof of Theorem 3 is provided in Appendix A. Note that Theorem 3 is established not only for the gamma distribution but also for all continuous population distribution. Next, moments and estimates in (12) are calculated from MSE-RPs, NT-RPs, and MC-RPs of different $Ga(a, b)$. Three typical shapes of gamma distributions ($Ga(1, 0.5)$—monotone decreasing; $Ga(2, 0.5)$—right skewed and $Ga(7.5, 1)$—bell-shaped; their pdfs are plotted in Figure 1). These are chosen and the representative points are set to three sizes ($k = 5, 20, 100$). The first part of Tables 1–3 summarizes the results in different scenarios. The last line of each table presents the moments and parameters of $Ga(a, b)$. It is clear that if k is fixed, the moments and estimates of MSE-RPs are closer to the true values than other representative points. Moreover, we can observe that the means of MSE-RPs are almost equal to the means of $Ga(a, b)$ in all scenarios; when k becomes large, the moments and estimates of MSE-RPs converge to the true values much faster than other representative points. These results are consistent with the description in Theorem 3.

Table 1. Summary of results from RPs of $Ga(1, 0.5)$ in parameter estimation.

	k	μ	σ^2	**Skewness**	**Kurtosis**	$\hat{a}_{m_2}$	$\hat{b}_{m_2}$	$\overline{PD}_{\hat{a}_{m_2}}$	$\overline{PD}_{\hat{b}_{m_2}}$	$\overline{PD}_{\hat{a}_{mle}}$	$\overline{PD}_{\hat{b}_{mle}}$
	5	**2.001**	**3.708**	**1.850**	**3.971**	**1.080**	**0.540**	**0.086**	**0.088**	0.380	0.381
MSE-RPs	20	**2.001**	**3.978**	**1.989**	**5.767**	**1.007**	**0.503**	**0.050**	**0.056**	0.088	0.089
	100	**2.001**	**3.998**	**2.003**	**5.996**	**1.002**	**0.501**	**0.050**	**0.056**	0.036	**0.043**
	5	1.866	2.419	0.775	−0.752	1.440	0.772	0.441	0.545	**0.244**	**0.334**
NT-RPs	20	1.967	3.419	1.394	1.470	1.132	0.576	0.134	0.153	**0.060**	**0.078**
	100	1.995	3.839	1.759	3.662	1.036	0.520	0.054	0.059	**0.034**	**0.043**
	5	2.069	3.516	0.576	−0.839	2.720	1.586	1.818	2.312	1.569	1.939
MC-RPs [1]	20	1.945	3.629	1.329	1.571	1.269	0.684	0.348	0.401	0.244	0.327
	100	1.991	3.987	1.751	3.797	1.051	0.536	0.168	0.181	0.108	0.127
$Ga(1, 0.5)$		2	4	2	6	1	0.5	-	-	-	-

[1] MC-RPs are randomly generated 100 times for each $Ga(a, b)$. All values are the average values of the repeats. This practice is also carried out in Tables 3–6, A1, and A2.
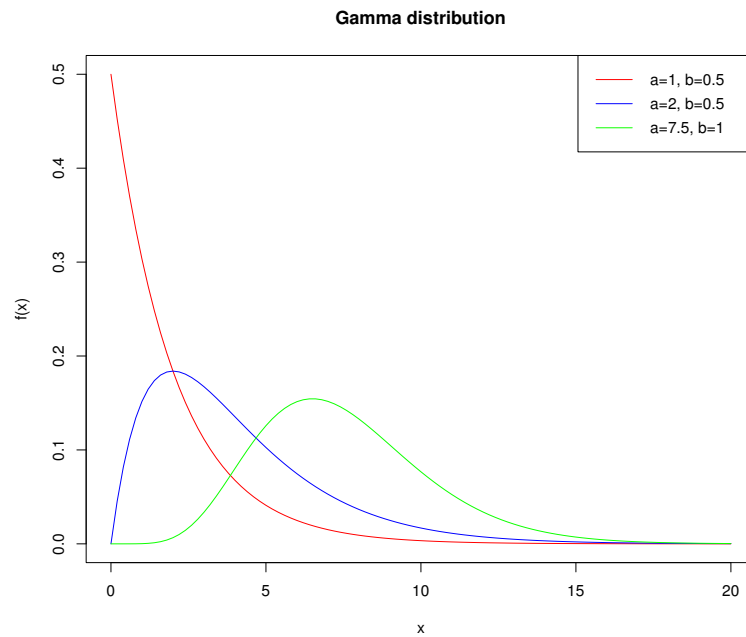
Gamma distribution

Figure 1. Probability density function for $Ga(1, 0.5)$, $Ga(2, 0.5)$ and $Ga(7.5, 1)$.

Table 2. Summary of results from RPs of $Ga(2, 0.5)$ in parameter estimation.

	k	μ	σ^2	Skewness	Kurtosis	$\hat{a}_{m_2}$	$\hat{b}_{m_2}$	$\overline{\text{PD}}_{\hat{a}_{m_2}}$	$\overline{\text{PD}}_{\hat{b}_{m_2}}$	$\overline{\text{PD}}_{\hat{a}_{mle}}$	$\overline{\text{PD}}_{\hat{b}_{mle}}$
	5	**3.999**	**7.396**	**1.304**	**1.685**	**2.163**	**0.541**	**0.085**	**0.086**	**0.222**	**0.223**
MSE-RPs	20	**3.999**	**7.954**	**1.403**	**2.842**	**2.011**	**0.503**	**0.044**	**0.048**	**0.042**	**0.046**
	100	**3.999**	**7.997**	**1.412**	**2.976**	**2.001**	**0.500**	**0.044**	**0.047**	**0.033**	**0.037**
	5	3.855	5.449	0.552	−0.934	2.727	0.707	0.365	0.416	0.269	0.317
NT-RPs	20	3.963	7.149	1.010	0.532	2.196	0.554	0.101	0.112	0.061	0.071
	100	3.993	7.780	1.266	1.825	2.049	0.513	0.044	0.048	0.034	0.038
	5	4.089	6.912	0.416	−0.904	5.261	1.382	1.769	1.925	1.672	1.800
MC-RPs	20	3.917	7.269	0.969	0.755	2.425	0.633	0.310	0.339	0.260	0.304
	100	3.981	7.982	1.266	1.930	2.074	0.525	0.146	0.153	0.112	0.120
$Ga(2, 0.5)$		4	8	1.414	3	2	0.5	-	-	-	-

Table 3. Summary of results from RPs of $Ga(7.5, 1)$ in parameter estimation.

	k	μ	σ^2	Skewness	Kurtosis	$\hat{a}_{m_2}$	$\hat{b}_{m_2}$	$\overline{\text{PD}}_{\hat{a}_{m_2}}$	$\overline{\text{PD}}_{\hat{b}_{m_2}}$	$\overline{\text{PD}}_{\hat{a}_{mle}}$	$\overline{\text{PD}}_{\hat{b}_{mle}}$
	5	**7.499**	**6.911**	**0.672**	**0.024**	**8.139**	**1.085**	**0.088**	**0.088**	**0.121**	**0.121**
MSE-RPs	20	**7.499**	**7.455**	**0.724**	**0.711**	**7.545**	**1.006**	**0.038**	**0.039**	**0.036**	**0.037**
	100	**7.499**	**7.498**	**0.730**	**0.795**	**7.502**	**1.000**	**0.038**	0.040	**0.035**	**0.036**
	5	7.424	5.575	0.284	−1.067	9.885	1.332	0.319	0.333	0.296	0.309
NT-RPs	20	7.480	6.946	0.530	−0.213	8.056	1.077	0.076	0.079	0.067	0.070
	100	7.496	7.374	0.662	0.381	7.620	1.017	0.038	**0.039**	0.036	0.037
	5	7.575	6.394	0.225	−0.950	19.282	2.561	1.480	1.431	1.480	1.422
MC-RPs	20	7.416	6.901	0.494	0.076	8.754	1.187	0.286	0.296	0.277	0.290
	100	7.478	7.490	0.670	0.443	7.714	1.034	0.127	0.127	0.116	0.116
$Ga(7.5, 1)$		7.5	7.5	0.730	0.8	7.5	1	-	-	-	-

Next, the comparison focuses on the estimating performance of samples from representative points. We take samples from different shapes of gamma distributions ($Ga(1, 0.5)$, $Ga(2, 0.5)$ and $Ga(7.5, 1)$), as well as their representative points with different sizes ($k = 5, 20, 100$). Setting sample size $N = 1000$ and repeat sampling $M = 10,000$ times

for each scenario, the method of moment estimates ($\hat{a}_{m_2}$ and $\hat{b}_{m_2}$) and maximum likelihood estimates ($\hat{a}_{mle}$ and $\hat{b}_{mle}$) are calculated. Define

$$\overline{\mathrm{PD}}_{\hat{a}} = \sum_{i=1}^{M} \frac{|\hat{a}_i - a|}{a} / M \text{ and } \overline{\mathrm{PD}}_{\hat{b}} = \sum_{i=1}^{M} \frac{|\hat{b}_i - b|}{b} / M$$

as the average proportional deviation between estimations and parameters. The second part of Tables 1–3 show that MSE-RPs samples have the smallest average proportional deviation in most of the selected scenarios. Tables A1 and A2 in Appendix C give medians and 95% empirical confidence intervals of $\hat{a}_{m_2}$, $\hat{b}_{m_2}$, $\hat{a}_{mle}$ and $\hat{b}_{mle}$. In this simulation study, we observe that the point estimates of a and b from MSE-RPs samples generally have good estimation accuracy with both the moment and maximum likelihood methods. Meanwhile, when k is large, the estimation performances of MSE-RPs samples are similar to those samples from the corresponding $Ga(a, b)$. It is also worth mentioning that when $k = 5$, the proportional deviation $\overline{\mathrm{PD}}_{\hat{a}_{m_2}}$ and $\overline{\mathrm{PD}}_{\hat{b}_{m_2}}$ are much smaller than $\overline{\mathrm{PD}}_{\hat{a}_{mle}}$ and $\overline{\mathrm{PD}}_{\hat{b}_{mle}}$. That is, when the size of gamma MSE-RPs is small, it is better to estimate parameters using the method of moments.

5. Generating MSE-RPs from Harrel–Davis Standardized Samples

This section discusses how to generate MSE-RPs from a gamma-distributed sample. A commonly used approach has two steps as follows:

1. Calculate the maximum likelihood estimates (MLEs) for a and b, namely $\hat{a}$ and $\hat{b}$, based on the sample dataset;
2. Generate MSE-RPs from the gamma distribution with the estimated parameters, i.e., $Ga(\hat{a}, \hat{b})$.

As we know, the representativeness of MSE-RPs depends on the estimate of gamma parameters. More accurate estimates will produce better representativeness. However, if a random sample does not represent the population well, the estimates may show large deviations from the true parameters. Hence, the MSE-RPs that are generated are not good representatives of the population distribution. This usually occurs when the sample size is small or medium. Next, we introduce a new Harrel–Davis (HD) standardization technique that can reduce the effect of randomness from samples. This technique transfers a random sample to a set of HD quantile estimators and then treats these estimators as a new "sample". Recall that a set of quantiles with equal probability is a set of NT-RPs for population; a similar idea is utilized for sample standardization.

Definition 1 (HD standardized sample). *Let $\boldsymbol{x} = \{x_1, x_2, \cdots, x_n\}$ be a set of sample data from a gamma distribution; set $\boldsymbol{x}' = \{Q_{p_1}, Q_{p_2}, \cdots, Q_{p_n}\}$, which is called the HD standardized sample of $\boldsymbol{x}$, where Q_{p_i} is the p_ith HD quantile estimator defined in (6), $p_i = \frac{2i-1}{2n}$ and $P(Q_{p_i}) = 1/n$ ($i = 1, 2, \cdots, n$).*

Note here that $\boldsymbol{x}'$ is not a random sample because $Q_{p_1}, Q_{p_2}, \cdots, Q_{p_n}$ are not independent. However, since quantile estimators are equiprobable ($P(Q_{p_i}) = 1/n$), set $\boldsymbol{x}'$ is treated as an arbitrarily selected sample, which can be used to calculate MLEs for a and b. A new approach to generate MSE-RPs is proposed as follows:

1. Obtain the HD standardized sample;
2. Calculate the MLEs for a and b, namely $\hat{a}$ and $\hat{b}$, based on the HD standardized sample;
3. Generate MSE-RPs from $Ga(\hat{a}, \hat{b})$.

Next, a simulation study is provided to show the good performance of HD standard samples in parameter estimation. Consider three gamma distributions ($Ga(1, 0.5)$, $Ga(2, 0.5)$ and $Ga(7.5, 1)$) and three different sample sizes ($n = 50, 200, 500$), in each scenario, a number of $N = 10,000$ random samples are generated and their HD standardized samples are obtained. The MLEs are calculated for each sample/standardized sample and

summarized in Table 4. This shows that the means of estimates from HD standardized samples are closer to the true value in most scenarios. Moreover, the estimates from HD standardized samples appear to have smaller standard deviations than those from random samples. We conclude that HD standardized samples outperform random samples in terms of estimation accuracy and stability based on these results. Therefore, it is recommended to use the new three-step approach to generate MSE-RPs. Here, a comparison study between the MSE-RPs generated by random samples and HD-samples is provided. The estimates ($\hat{a}$ and $\hat{b}$) in Table 4 are used to generate gamma MSE-RPs. Table 5 summarizes the results when $n = 200$ with the size of MSE-RPs $k = 20$. It shows that the moments of gamma MSE-RPs from HD-samples are close to the moments of the origin $Ga(a, b)$. Meanwhile, the method of moment estimates in (12) are obtained. The estimates from HD samples have a better accuracy than those from random samples. This conclusion is generally valid when $n = 50$ and 500.

Table 4. Mean (Standard deviation) of MLEs from samples and HD standardized samples.

		Ga(1, 0.5)		*Ga*(2, 0.5)		*Ga*(7.5, 1)	
n		**Sample**	**HD-Sample**	**Sample**	**HD-Sample**	**Sample**	**HD-Sample**
50	$\hat{a}$	**1.060**(0.195)	1.075(**0.192**)	2.115(0.415)	**2.104**(**0.405**)	7.964(1.645)	**7.794**(**1.604**)
	$\hat{b}$	0.540(0.127)	**0.535**(**0.126**)	0.534(0.118)	**0.524**(**0.116**)	1.064(0.227)	**1.038**(**0.221**)
200	$\hat{a}$	1.021(**0.090**)	**1.020**(**0.090**)	2.029(0.192)	**2.012**(**0.189**)	7.616(0.759)	**7.502**(**0.746**)
	$\hat{b}$	0.512(**0.058**)	**0.507**(**0.058**)	0.508(0.054)	**0.502**(**0.054**)	1.016(0.104)	**0.999**(**0.103**)
500	$\hat{a}$	1.012(**0.056**)	**1.010**(**0.056**)	2.010(0.120)	**2.000**(**0.119**)	7.543(0.472)	**7.477**(**0.468**)
	$\hat{b}$	0.507(**0.036**)	**0.504**(**0.036**)	0.503(0.034)	**0.499**(**0.033**)	1.006(0.065)	**0.997**(**0.064**)

It is noteworthy that the HD standardization technique can also be applied in resampling. Consider another simulation study with the same settings as Table 4. We resample from each sample/standardized sample using $n_r = n$ and calculate the MLEs. The means and standard deviations of the resampled MLEs are summarized in Table 6. This shows that estimates from standardized samples generally have a better accuracy and smaller standard deviations when resampling.

Table 5. Summary of results for MSE-RPs from the esitmated gamma distributions.

$n = 200$	$k = 20$	μ	σ^2	**Skewness**	**Kurtosis**	$\hat{a}_{m_2}$	$\hat{b}_{m_2}$
	sample	**1.995**	3.873	1.966	5.641	1.028	0.515
$Ga(1, 0.5)$	HD-sample	2.013	**3.946**	**1.967**	**5.647**	**1.027**	**0.510**
	origin	2	4	2	6	1	0.5
	sample	**3.994**	7.818	1.393	2.801	2.041	0.511
$Ga(2, 0.5)$	HD-sample	4.008	**7.938**	**1.399**	**2.825**	**2.023**	**0.505**
	origin	4	8	1.414	3	2	0.5
	sample	**7.496**	7.334	0.719	0.699	7.662	1.022
$Ga(7.5, 1)$	HD-sample	7.509	**7.472**	**0.724**	**0.711**	**7.547**	**1.005**
	origin	7.5	7.5	0.730	0.8	7.5	1

Table 6. Mean (Standard deviation) of resampled MLEs from samples and HD standardized samples.

		Ga(1, 0.5)		*Ga*(2, 0.5)		*Ga*(7.5, 1)	
$n(n_r)$		**Sample**	**HD-Sample**	**Sample**	**HD-Sample**	**Sample**	**HD-Sample**
50	$\hat{a}$	**1.114**(0.295)	1.127(**0.283**)	2.234(0.630)	**2.216**(**0.594**)	8.446(2.504)	**8.245**(**2.340**)
	$\hat{b}$	0.577(0.196)	**0.572**(**0.185**)	0.569(0.181)	**0.558**(**0.170**)	1.131(0.347)	**1.101**(**0.323**)
200	$\hat{a}$	1.034(0.131)	**1.030**(**0.127**)	2.059(0.279)	**2.034**(**0.269**)	7.741(1.102)	**7.594**(**1.061**)
	$\hat{b}$	0.522(0.085)	**0.515**(**0.082**)	0.518(0.079)	**0.509**(**0.076**)	1.034(0.151)	**1.013**(**0.146**)
500	$\hat{a}$	1.017(**0.079**)	**1.016**(**0.079**)	2.021(**0.167**)	**2.012**(0.168)	7.584(**0.660**)	**7.527**(0.662)
	$\hat{b}$	0.510(**0.051**)	**0.508**(**0.051**)	0.506(0.047)	**0.503**(**0.047**)	1.012(**0.091**)	**1.004**(**0.091**)

6. Real Data Illustration

In this section, we consider a real-world dataset and illustrate the HD standardized technique proposed in the previous section. In this clinical study, 97 Swiss females ($n = 97$) aged 70–74 inclusive at the time of diagnosis of dementia (a form of mental disorder) were studied for survival times (in years) by Elandt–Johnson and Johnson [23]. These data were analyzed by Ozonur and Paul [24] using the likelihood ratio test and score test with p-values 0.233 and 0.140, which are greater than 0.05. Both tests suggest that the two-parameter gamma distribution adequately fits the dementia data.

Point estimates (MLE) and the bootstrap interval estimates [25] based on the origin sample data and the corresponding HD sample are calculated. The approximate $(1 - \alpha)$ bootstrap percentile interval is defined as

$$\left[\hat{\boldsymbol{\theta}}_{\text{lower}}, \hat{\boldsymbol{\theta}}_{\text{upper}}\right] = \left[\hat{\boldsymbol{\theta}}_M^{*\left(\frac{\alpha}{2}\right)}, \hat{\boldsymbol{\theta}}_M^{*\left(1-\frac{\alpha}{2}\right)}\right].$$

In practice, we resample the original data $M = 1000$ times to obtain 1000 replications of the parameter estimate $\hat{\boldsymbol{\theta}}^*$ (i.e., $\hat{a}$ and $\hat{b}$ for the gamma distribution) with $\alpha = 0.05$. These estimates are sorted and the 25th value is used as the lower bound; the 975th value is the upper bound. The MLEs based on the HD standardized sample are $\hat{a}_{HD} = 1.4602$ and $\hat{b}_{HD} = 0.2886$ with confidence intervals $(1.3846, 1.8073)$ and $(0.2637, 0.3839)$. The lengths of confidence intervals are shorter than those based on the origin sample data, where $\hat{a}_{origin} = 1.4602$ and $\hat{b}_{origin} = 0.2886$ with confidence intervals $(1.3777, 1.8632)$ and $(0.2659, 0.3914)$.

7. Concluding Remarks

In the first part of this paper, the existence and uniqueness of gamma MSE-RPs are proved using two different approaches. An effective algorithm is recommended for the generation of gamma MSE-RPs. The second part of this paper compares gamma MSE-RPs with other representative points in terms of parameter estimation and simulation. This shows that the moments and estimates based on gamma MSE-RPs are the closest to the true values in different scenarios. In addition, samples from gamma MSE-RPs show a good general estimation accuracy. The last part of this paper introduces the new HD standardization technique. When a gamma-distributed sample is at hand, we recommend first transferring it to the HD standardized sample and then using it to estimate gamma parameters or generate MSE-RPs.

In future work, we would like to study whether the MSE-RPs of other distributions can also perform well in parameter estimation. It would also be interesting to explain how HD standardization technique reduces the randomness from samples through a theoretical demonstration.

Appendix A. Proofs of Theorems

Proof of Theorem 1. The proof of three points in this theorem are provided as follows. Without loss of generality, consider a gamma distribution with $b = 1$ (i.e., $\mu = a$) in all proofs.

Proof of point 1. Let

$$G(z_1, z_2) = (z_1 - a)F(\frac{z_1 + z_2}{2}) + \frac{1}{2}(z_1 + z_2)f(\frac{z_1 + z_2}{2}),$$

there is

$$G(z_1, z_1) = (z_1 - a)F(z_1) + z_1 f(z_1).$$

Because $\lim_{z_1 \to 0} G(z_1, z_1) = 0$ and

$$\begin{aligned}
\frac{dG(z_1, z_1)}{dz_1} \equiv G'_{z_1}(z_1, z_1) &= F(z_1) + (z_1 - a)f(z_1) + f(z_1) + (a - 1)f(z_1) - z_1 f(z_1) \\
&= F(z_1) > 0,
\end{aligned}$$

Hence $G(z_1, z_1) > 0$. In addition,

$$G(z_1, \infty) = \lim_{z_2 \to \infty} G(z_1, z_2) = z_1 - a,$$

we have

$$G(z_1, \infty) < 0 \Leftrightarrow z_1 < a. \tag{A1}$$

Combine (A1) with the condition $G(z_1, z_2)$ is continuous for $z_2 \in [z_1, \infty)$ and $G(z_1, z_1) > 0$; point 1 of Theorem 1 is proved.

Proof of point 3.

Let

$$G(z_{k-1}, z_k) = (z_k - a)[1 - F(\frac{z_{k-1} + z_k}{2})] - \frac{1}{2}(z_{k-1} + z_k)f(\frac{z_{k-1} + z_k}{2}), \tag{A2}$$

therefore

$$G(z_{k-1}, z_{k-1}) = (z_{k-1} - a)(1 - F(z_{k-1})) - z_{k-1}f(z_{k-1}).$$

Since $G(0, 0) = -a < 0$, $\lim_{z_{k-1} \to \infty} G(z_{k-1}, z_{k-1}) = 0$ and

$$\begin{aligned}
G'_{z_{k-1}}(z_{k-1}, z_{k-1}) &= 1 - F(z_{k-1}) - (z_{k-1} - a)f(z_{k-1}) - f(z_{k-1}) \\
&\quad -(a - 1)f(z_{k-1}) + z_{k-1}f(z_{k-1}) \\
&= 1 - F(z_{k-1}) > 0,
\end{aligned}$$

we have

$$G(z_{k-1}, z_{k-1}) < 0 \quad (z_{k-1} > 0). \tag{A3}$$

Next, we show that for $z_k \in [z_{k-1}, \infty)$, $G(z_{k-1}, z_k)$ is firstly monotone-increasing and then monotone-decreasing. Derive $G(z_{k-1}, z_k)$ by z_k to obtain

$$\begin{aligned}
G'_{z_k}(z_{k-1}, z_k) &= (1 - F(\frac{z_{k-1} + z_k}{2})) - \frac{1}{2}(z_k - a)f(\frac{z_{k-1} + z_k}{2}) \\
&\quad -\frac{1}{2}f(\frac{z_{k-1} + z_k}{2}) - \frac{1}{2}(a - 1)f(\frac{z_{k-1} + z_k}{2}) + \frac{1}{4}(z_{k-1} + z_k)f(\frac{z_{k-1} + z_k}{2}) \\
&= 1 - F(\frac{z_{k-1} + z_k}{2}) - \frac{1}{4}(z_k - z_{k-1})f(\frac{z_{k-1} + z_k}{2}).
\end{aligned}$$

Let

$$H(z_{k-1}, z_k) = G'_{z_k}(z_{k-1}, z_k), \quad z = \frac{1}{2}(z_{k-1} + z_k),$$

we have

$$
\begin{aligned}
H'_{z_k}(z_{k-1}, z_k) &= -\frac{1}{2}f(\frac{z_{k-1} + z_k}{2}) - \frac{1}{4}f(\frac{z_{k-1} + z_k}{2}) \\
&\quad -\frac{1}{4}(z_k - z_{k-1})[\frac{2}{z_{k-1} + z_k} \cdot \frac{a-1}{2} - \frac{1}{2}]f(\frac{z_{k-1} + z_k}{2}) \\
&= [-\frac{3}{4} + \frac{1}{8}(z_k - z_{k-1}) - \frac{(a-1)(z_k - z_{k-1})}{4(z_{k-1} + z_k)}]f(\frac{z_{k-1} + z_k}{2}) \\
&= [z^2 - (a + 2 + z_{k-1})z + (a-1)z_{k-1}]f(z)/z. \quad\quad\quad (A4)
\end{aligned}
$$

Note that in Equation (A4), $z^2 > 0$, as long as

$$H(z_{k-1}, z_{k-1}) = 1 - F(z_{k-1}) > 0, \text{ and } \lim_{z_k \to \infty} H(z_{k-1}, z_k) = 0,$$

$C_0 > z_{k-1}$ must exist that satisfies

$$\text{when } z_k < C_0, \ H'_{z_k}(z_{k-1}, z_k) < 0;$$
$$\text{when } z_k > C_0, \ H'_{z_k}(z_{k-1}, z_k) > 0.$$

Therefore, C^* ($z_{k-1} < C^* < C_0$) satisfies

$$\text{when } z_k < C^*, \ G'_{z_k}(z_{k-1}, z_k) = H(z_{k-1}, z_k) < 0;$$
$$\text{when } z_k > C^*, \ G'_{z_k}(z_{k-1}, z_k) = H(z_{k-1}, z_k) > 0,$$

which means that $G(z_{k-1}, z_k)$ is firstly monotone-increasing and then monotone-decreasing. In addition, we have

$$\lim_{z_k \to \infty} G(z_{k-1}, z_k) = 0 \quad\quad\quad (A5)$$

and $G(z_{k-1}, z_{k-1}) < 0$. Thus, the function $G(z_{k-1}, z_k)$ must cross the x-axis and the solution z_k exists. One more step:

$$\frac{\mathrm{d}z_k}{\mathrm{d}z_{k-1}} = -G'_{z_{k-1}}(z_{k-1}, z_k)/G'_{z_k}(z_{k-1}, z_k)$$

and $G'_{z_k}(z_{k-1}, z_k) > 0$ in the neighborhood domain; furthermore,

$$G'_{z_{k-1}}(z_{k-1}, z_i) = -\frac{1}{4}(z_k - z_{k-1})f(\frac{z_{k-1} + z_k}{2}) < 0,$$

We find that z_k is a monotone-increasing function of z_{k-1}.

Proof of point 2. Proving point 2 is complicated. Here, we provide the prove of a special case when $X \sim Ga(1, 1)$. Let

$$
\begin{aligned}
G(z_{i-1}, z_i, z_{i+1}) &= (z_i - a)[F(\frac{z_i + z_{i+1}}{2}) - F(\frac{z_{i-1} + z_i}{2})] \\
&\quad + \frac{1}{2}(z_i + z_{i+1})f(\frac{z_i + z_{i+1}}{2}) - \frac{1}{2}(z_{i-1} + z_i)f(\frac{z_{i-1} + z_i}{2}),
\end{aligned}
$$

thus:

$$G(z_{i-1}, z_i, z_i) = (z_i - a)[F(z_i) - F(\frac{z_{i-1} + z_i}{2})]$$
$$+ z_i f(z_i) - \frac{z_{i-1} + z_i}{2} f(\frac{z_{i-1} + z_i}{2}).$$

Deriving $G(z_{i-1}, z_i, z_i)$ by z_i, we have

$$G'_{z_i}(z_{i-1}, z_i, z_i) = F(z_i) - F(\frac{z_{i-1} + z_i}{2}) - \frac{z_i - z_{i-1}}{4} f(\frac{z_{i-1} + z_i}{2}). \tag{A6}$$

Let

$$H(z_{i-1}, z_i) = G'_{z_i}(z_{i-1}, z_i, z_i),$$

we have

$$H'_{z_i}(z_{i-1}, z_i) = f(z_i) + [-\frac{3}{4} + \frac{1}{8}(z_i - z_{i-1}) - (a-1)\frac{z_i - z_{i-1}}{4(z_{i-1} + z_i)}] f(\frac{z_{i-1} + z_i}{2}). \tag{A7}$$

For $X \sim Ga(1, 1)$, (A7) can be simplified to

$$H'_{z_i}(z_{i-1}, z_i) = f(z_i) + [-\frac{3}{4} + \frac{1}{8}(z_i - z_{i-1})] f(\frac{z_{i-1} + z_i}{2}), \tag{A8}$$

set $x = \frac{z_i - z_{i-1}}{2}$, we have

$$f(z_i) + [-\frac{3}{4} + \frac{1}{8}(z_i - z_{i-1})] f(\frac{z_{i-1}+z_i}{2}) = 0$$
$$\Rightarrow \qquad \frac{e^{-z_i}}{e^{-\frac{z_{i-1}+z_i}{2}}} = \frac{3}{4} - \frac{1}{8}(z_i - z_{i-1})$$
$$\Rightarrow \qquad e^{-\frac{z_{i-1}-z_i}{2}} = \frac{3}{4} - \frac{1}{4}\frac{z_{i-1}-z_i}{2}$$
$$\Rightarrow \qquad e^{-x} = \frac{3}{4} - \frac{1}{4}x.$$

Therefore, $H'_{z_i}(z_{i-1}, z_i)$ crosses the x-axis twice for $z_i > z_{i-1}$.

As $H'_{z_i}(z_{i-1}, z_{i-1}) = \frac{1}{4} f(z_{i-1}) > 0$ and $\lim_{z_i \to \infty} \frac{\partial H(z_{i-1}, z_i)}{\partial z_i} = 0$, combined with the facts $H(z_{i-1}, z_{i-1}) = 0$ and $\lim_{z_i \to \infty}(H(z_{i-1}, z_i)) = 0$, we know that, for $z_i \in [z_{i-1}, \infty)$, $G(z_{i-1}, z_i, z_i)$ is first monotone-increasing and then monotone-decreasing. In addition, $G(z_{i-1}, z_{i-1}, z_{i-1}) = 0$ and $\lim_{z_i \to \infty} G(z_{i-1}, z_i, z_i) = 0$; we conclude $G(z_{i-1}, z_i, z_i) > 0$ when $z_{i-1} > 0$. Next, consider

$$\begin{aligned} G'_{z_{i+1}}(z_{i-1}, z_i, z_{i+1}) &= \frac{z_i - a}{2} f(\frac{z_i + z_{i+1}}{2}) + \frac{1}{2} f(\frac{z_i + z_{i+1}}{2}) \\ &\quad + \frac{a-1}{2} f(\frac{z_i + z_{i+1}}{2}) - \frac{1}{4}(z_i + z_{i+1}) f(\frac{z_i + z_{i+1}}{2}) \\ &= \frac{z_i - z_{i+1}}{4} f(\frac{z_i + z_{i+1}}{2}) < 0, \end{aligned}$$

therefore, the solution z_{i+1} exists if

$$G(z_{i-1}, z_i, \infty) < 0.$$

We find that

$$G(z_{i-1}, z_i, \infty) = (z_i - a)[1 - F(\frac{z_{i-1} + z_i}{2})] - \frac{z_{i-1} + z_i}{2} f(\frac{z_{i-1} + z_i}{2})$$

is exactly (A2). From the analysis in the proof of point 3, we conclude that the solution z_{i+1} exists when $z_{i-1} < z_{i,i-1}$. □

Proof of Theorem 2. Let a random variable $X \sim Ga(a, b)$ with the pdf

$$g(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad \text{for } x > 0, \text{ and } a, b > 0. \tag{A9}$$

If X has a log-concave density function, there exists a unique set of MSE-RPs (Trushkin [26]). Function $g(x; a, b)$ is log-concave if

$$(-\ln g(x; a, b))'' \geq 0. \tag{A10}$$

Based on (A9), we have

$$\ln g(x; a, b) = (a - 1) \ln x + a \ln b - bx - \ln \Gamma(a),$$
$$(-\ln g(x; a, b))' = (1 - a)x^{-1} + b,$$
$$(-\ln g(x; a, b))'' = (a - 1)x^{-2}.$$

The inequality (A10) holds when $a \geq 1$. □

Proof of Theorem 3. By taking partial diffrenciation of (7), we have

$$\begin{cases} z_1 \int_0^{\frac{1}{2}(z_1+z_2)} f(x)\mathrm{d}x = \int_0^{\frac{1}{2}(z_1+z_2)} x f(x)\mathrm{d}x \\ z_2 \int_{\frac{1}{2}(z_1+z_2)}^{\frac{1}{2}(z_2+z_3)} f(x)\mathrm{d}x = \int_{\frac{1}{2}(z_1+z_2)}^{\frac{1}{2}(z_2+z_3)} x f(x)\mathrm{d}x \\ \quad\cdots\cdots\cdots\cdots \\ z_k \int_{\frac{1}{2}(z_{k-1}+z_k)}^{\infty} f(x)\mathrm{d}x = \int_{\frac{1}{2}(z_{k-1}+z_k)}^{\infty} x f(x)\mathrm{d}x. \end{cases} \tag{A11}$$

Summing up the LHS and RHS of (A11), $E(X) = E(Z)$ is obtained. The first part of this theorem is proved. Next, from Theorem 3 in Fei [19], we have

$$\lim_{k \to \infty} \mathrm{MSE}(z_1, z_2, \cdots, z_k) = 0. \tag{A12}$$

Theorem 5 of Fei [19] shows that

$$Var(Z) = (1 - \mathrm{MSE}(z_1, z_2, \cdots, z_k)) Var(X). \tag{A13}$$

Combining (A12) and (A13), the second part of this theorem is proved. □

Appendix B. Self-Consistency Algorithm for Generating Gamma MSE-RPs

The self-consistency algorithm [22] has the following steps:
1. Let the $\mathbf{z}_0 = \{z_1^{NT}, z_2^{NT} \cdots z_k^{NT}\}$ be the initial set.
2. Compute the conditional expectation $\mathbf{z}_1 = \mathcal{E}[X \mid \mathbf{z}_0]$ using the system of equation,

$$z_i = \frac{\int_{I_i} x dF(x)}{\int_{I_i} dF(x)}, \quad i = 1, 2, \cdots, k$$

and compare the distance between $\mathbf{z}_0$ and $\mathbf{z}_1$ for each z_i. If the minimum distance is not smaller than the pre-defined error, e.g., $\epsilon = 10^{-10}$, proceed to check the next step.
3. Repeat steps 1 and 2, obtaining corresponding $\mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4, \cdots$, until convergence is reached.

Appendix C. Median Estimates and Confidence Intervals of *a* and *b*

Table A1. Median estimates and confidence intervals of *a* and *b* (method of moments).

		$Ga(1, 0.5)$		$Ga(2, 0.5)$		$Ga(7.5, 1)$	
k	RP	$\hat{a}_{m_2}$	$\hat{b}_{m_2}$	$\hat{a}_{m_2}$	$\hat{b}_{m_2}$	$\hat{a}_{m_2}$	$\hat{b}_{m_2}$
	MSE	**1.081**(0.979, 1.198)	**0.540**(0.484, 0.608)	**2.166**(1.985, 2.371)	**0.541**(0.493, 0.598)	**8.143**(7.548, 8.820)	**1.086**(1.003, 1.180)
5	NT	1.440(1.336, 1.555)	0.772(0.735, 0.813)	2.728(2.552, 2.922)	0.707(0.670, 0.748)	9.890(9.316, 10.516)	1.332(1.261, 1.413)
	MC	2.719(2.533, 2.930)	1.585(1.501, 1.681)	5.260(4.934, 5.632)	1.382(1.307, 1.467)	17.210(16.200, 18.380)	2.240(2.114, 2.387)
	MSE	**1.011**(0.894, 1.139)	**0.505**(0.440, 0.577)	**2.015**(1.815, 2.238)	**0.504**(0.450, 0.565)	**7.550**(6.896, 8.281)	**1.007**(0.916, 1.108)
20	NT	1.133(1.045, 1.229)	0.576(0.532, 0.626)	2.197(2.034, 2.377)	0.554(0.513, 0.601)	8.050(7.477, 8.719)	1.076(1.000, 1.167)
	MC	1.180(1.090, 1.282)	0.603(0.563, 0.650)	2.253(2.086, 2.440)	0.568(0.529, 0.614)	8.185(7.586, 8.868)	1.093(1.014, 1.184)
	MSE	**1.007**(0.888, 1.135)	**0.503**(0.436, 0.576)	**2.004**(1.799, 2.223)	**0.501**(0.446, 0.561)	**7.516**(6.844, 8.242)	**1.002**(0.911, 1.101)
100	NT	1.039(0.941, 1.151)	0.521(0.467, 0.584)	2.050(1.870, 2.251)	0.513(0.466, 0.569)	7.626(7.013, 8.322)	1.017(0.934, 1.111)
	MC	1.038(0.942, 1.149)	0.518(0.468, 0.579)	2.043(1.866, 2.246)	0.510(0.464, 0.564)	7.591(6.966, 8.300)	1.011(0.926, 1.108)
$Ga(a, b)$		1.004(0.885, 1.131)	0.501(0.436, 0.574)	2.004(1.797, 2.230)	0.501(0.446, 0.563)	7.509(6.838, 8.257)	1.001(0.909, 1.104)

Table A2. Median estimates and confidence intervals of *a* and *b* (MLEs).

		$Ga(1, 0.5)$		$Ga(2, 0.5)$		$Ga(7.5, 1)$	
k	RP	$\hat{a}_{mle}$	$\hat{b}_{mle}$	$\hat{a}_{mle}$	$\hat{b}_{mle}$	$\hat{a}_{mle}$	$\hat{b}_{mle}$
	MSE	1.379(1.305, 1.459)	0.689(0.627, 0.759)	**2.442**(2.298, 2.599)	**0.610**(0.563, 0.662)	**8.394**(7.844, 9.022)	**1.120**(1.039, 1.209)
5	NT	**1.243**(1.175, 1.320)	**0.667**(0.625, 0.713)	2.535(2.395, 2.689)	0.658(0.619, 0.701)	9.709(9.179, 10.297)	1.308(1.236, 1.389)
	MC	2.383(2.258, 2.528)	1.354(1.277, 1.443)	4.947(4.683, 5.253)	1.289(1.217, 1.372)	16.959(16.009, 18.067)	2.203(2.078, 2.349)
	MSE	1.087(1.018, 1.161)	0.543(0.494, 0.596)	**2.066**(1.929, 2.241)	**0.516**(0.474, 0.566)	**7.589**(6.982, 8.261)	**1.012**(0.929, 1.106)
20	NT	1.057(0.988, 1.134)	**0.538**(0.493, 0.586)	2.116(1.973, 2.278)	0.534(0.494, 0.580)	7.977(7.425, 8.618)	1.067(0.990, 1.155)
	MC	**1.083**(1.011, 1.164)	0.554(0.508, 0.607)	2.168(2.017, 2.339)	0.548(0.505, 0.597)	8.166(7.571, 8.846)	1.091(1.008, 1.186)
	MSE	1.022(0.950, 1.101)	**0.510**(0.463, 0.564)	**2.009**(1.857, 2.177)	**0.501**(0.460, 0.550)	**7.521**(6.901, 8.192)	**1.002**(0.919, 1.096)
100	NT	**1.017**(0.945, 1.098)	**0.510**(0.465, 0.561)	2.025(1.876, 2.190)	0.507(0.466, 0.555)	7.605(7.025, 8.267)	1.015(0.934, 1.105)
	MC	1.020(0.948, 1.101)	0.509(0.463, 0.562)	2.028(1.876, 2.199)	0.506(0.463, 0.555)	7.604(7.006, 8.280)	1.013(0.930, 1.106)
$Ga(a, b)$		1.002(0.928, 1.083)	0.501(0.453, 0.553)	2.003(1.849, 2.176)	0.501(0.456, 0.549)	7.512(6.908, 8.208)	1.001(0.918, 1.097)

References

1. Fang, K.T.; Wang, Y. *Number-Theoretic Methods in Statistics*; Chapman and Hall: London, UK, 1994.
2. Cox, D.R. Note on grouping. *J. Am. Stat. Assoc.* **1957**, *52*, 543–547. [CrossRef]
3. Gersho, A.; Gray, R.M. *Vector Quantization and Signal Compression*; Springer Science & Business Media: New York, NY, USA, 2012; Volume 159.
4. Pagès, G. A space quantization method for numerical integration. *J. Comput. Appl. Math.* **1998**, *89*, 1–38. [CrossRef]
5. Pagès, G. Introduction to vector quantization and its applications for numerics. *ESAIM Proc. Surv.* **2015**, *48*, 29–79. [CrossRef]
6. Gobet, E.; Pagès, G.; Pham, H.; Printems, J. Discretization and simulation of the Zakai equation. *SIAM J. Numer. Anal.* **2006**, *44*, 2505–2538. [CrossRef]
7. El Amri, M.R.; Helbert, C.; Lepreux, O.; Zuniga, M.M.; Prieur, C.; Sinoquet, D. Data-driven stochastic inversion via functional quantization. *Stat. Comput.* **2020**, *30*, 525–541. [CrossRef]
8. Fang, K.T.; Zhou, M.; Wang, W.J. Applications of the representative points in statistical simulations. *Sci. China Math.* **2014**, *57*, 2609–2620. [CrossRef]
9. Fang, K.T.; He, P.; Yang, J. Set of representative points of statistical distributions and their applications. *Sci. Sin. Math.* **2020**, 50, 1–20.
10. Fang, K.T.; He, S.D. The problem of selecting a specified number of representative points from a normal population. *Acta Math. Appl. Sin.* **1984**, *7*, 293–306.
11. Flury, B.A. Principal points. *Biometrika* **1990**, *77*, 33–41. [CrossRef]
12. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [CrossRef]

13. Linde, Y.; Buzo, A.; Gray, R. An algorithm for vector quantizer design. *IEEE Trans. Commun.* **1980**, *28*, 84–95. [CrossRef]
14. Tarpey, T. Self-consistency algorithms. *J. Comput. Graph. Stat.* **1999**, *8*, 889–905.
15. Chakraborty, S.; Roychowdhury, M.K.; Sifuentes, J. High precision numerical computation of principal points for univariate distributions. *Sankhya B* **2021**, *83*, 558–584. [CrossRef]
16. Li, Y.N.; Fang, K.T.; He, P.; Peng, H. Representative points from a mixture of two normal distributions. *Mathematics* **2022**, *10*, 3952. [CrossRef]
17. Jiang, J.J.; He, P.; Fang, K.T. An interesting property of the arcsine distribution and its applications. *Stat. Probab. Lett.* **2015**, *105*, 88–95. [CrossRef]
18. Xu, L.H.; Fang, K.T.; He, P. Properties and generation of representative points of the exponential distribution. *Stat. Pap.* **2022**, *63*, 197–223. [CrossRef]
19. Fei, R.C. Statistical relationship between the representative points and the population. *J. Wuxi Inst. Light Ind.* **1991**, *10*, 78–83.
20. Fu, H.H. The problem of selecting a specified number of representative points from a gamma population. *J. Min. Sci. Technol.* **1985**, 107–116.
21. Harrell, F.E.; Davis, C.E. A new distribution-free quantile Estimator. *Biometrika* **1982**, *69*, 635–640. [CrossRef]
22. Stampfer, E.; Stadlober, E. Methods for estimating principal points. *Commun. Stat.-Simul. Comput.* **2002**, *31*, 261–277. [CrossRef]
23. Elandt-Johnson, R.; Johnson, N. *Survival Models and Data Analysis*; Wiley Series in Probability and Statistics: New York, NY, USA, 1999.
24. Ozonur, D.; Paul, S. Goodness of fit tests of the two-parameter gamma distribution against the three-parameter generalized gamma distribution. *Commun. Stat.-Simul. Comput.* **2020**, *51*, 687–697. [CrossRef]
25. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap*; Chapman and Hall/CRC: New York, NY, USA, 1994.
26. Trushkin, A. Sufficient conditions for uniqueness of a locally optimal quantizer for a class of convex error weighting functions. *IEEE Trans. Inf. Theory* **1982**, *28*, 187–198. [CrossRef]