The general mixture of Type I gamma distribution with mixture variable following generalized inverse Gaussian family and its applications

Yuanfan ZHAO[a,†], **Yuefan WU**[a,†], **Hua ZHOU**[b,c], **Xun-Jian LI**[b,*] and **Guo-Liang TIAN**[a,*]

[a]*Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen 518055, Guangdong Province, P. R. China*

[b]*Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA 90095, USA*

[c]*Department of Computational Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA*

[†]Equally contributed

[*]Corresponding author's Email: xunjianli@ucla.edu tiangl@sustech.edu.cn

14 FEB 2026 ZYF; 14 JAN 2026 TGL

Abstract: Positive continuous data arise frequently in a wide range of scientific fields, including biomedicine, engineering, finance, and economics, and are often characterized by right-skewness, over-dispersion, and latent heterogeneity. Existing distributional models such as the gamma, inverse Gaussian, and Weibull distributions provide fundamental tools for modeling such data, yet their structural assumptions–particularly the restrictive relationship between the mean and variance may limit flexibility and lead to inadequate fit when substantial heterogeneity is present. In this paper, we propose a general mixture of Type I gamma distribution constructed by introducing a random mixture variable τ into the Type I gamma distribution. This framework preserves a clearly interpretable structure while allowing for flexible variance inflation through latent heterogeneity. The proposed construction enables a natural decomposition of structural variability and random heterogeneity at the variance level. Within this unified framework, several new distributional models are obtained by specifying τ to specific distributions of the generalized inverse Gaussian family. Statistical inference for the proposed models is developed using the normalized-expectation maximization (N-EM) algorithm, which facilitates efficient parameter estimation and extension to mean regression models. Asymptotic properties of the MLEs are established, and practical inference procedures are discussed. Extensive simulation studies and real data analysis demonstrate that the proposed models have the better performance compared with existing alternatives, particularly for over-dispersed and highly skewed data.

Keywords: Ge-Ga distributions, Mean regression model, N-EM algorithm, Over-dispersion, Right skewness.

1. Introduction

Positive continuous data widely appear in biomedical research, engineering, insurance, finance and economics. Typical examples include drug response intensity, cell growth rates, and gene expression levels; lifetime and failure-time; insurance claim amounts; transaction volumes, and risk exposure measures. Such data often exhibit right-skewness, substantial individual variability or latent heterogeneity, which pose nontrivial challenges for statistical modeling and inference. Consequently, the development of flexible and robust probabilistic models for positive continuous data has long been an important topic in statistical methodology and applied research.

To model quantitative measurements with positive support $\mathbb{R}_+ \triangleq (0, \infty)$, a variety of classical distributions have been proposed and extensively studied in the statistical literature. Among them, the gamma distribution (Bownan & Shenton 1982) with density

$$\text{Gamma}(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \mathrm{e}^{-\beta y}, \quad y > 0,\ \alpha > 0,\ \beta > 0 \tag{1.1}$$

is particularly popular due to its simple functional form and clear parameter interpretation, especially within its mean regression frameworks, and has been widely adopted in biostatistics and generalized linear models. The inverse Gaussian distribution (Chhikara & Folks 1989) with density

$$\text{IGaussian}(y|\mu, \lambda) = \sqrt{\frac{\lambda}{2\pi}} y^{-\frac{3}{2}} \exp\left[-\frac{\lambda(y-\mu)^2}{2\mu^2 y}\right], \quad y > 0,\ \mu > 0,\ \lambda > 0 \tag{1.2}$$

provides greater flexibility in capturing strongly right-skewed and heavy-tailed data, and has found important applications in engineering reliability and survival analysis. In addition, the Weibull distribution (Weibull 1951), owing to its diverse shape characteristics, also plays a central role in lifetime analysis and reliability modeling, whose density function can be expressed as

$$\text{Weibull}(y|k, \lambda) = \frac{k}{\lambda} \left(\frac{y}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{y}{\lambda}\right)^k\right], \quad y > 0,\ k > 0,\ \lambda > 0.$$

These distributions constitute fundamental tools for the analysis of positive continuous data and have been successfully applied in numerous empirical studies.

Despite their widespread applicability, the structural assumptions underlying these classical distributions inevitably impose limitations on model flexibility. For instance, in the gamma distribution, the mean and variance are linked through a fixed functional relationship, which often makes it difficult to simultaneously achieve adequate goodness of fit and meaningful parameter interpretation when modeling over-dispersed data or data exhibiting substantial heterogeneity. Similar limitations arise for the inverse Gaussian and Weibull distributions when more complex variance structures or latent individual–level variability are present. In many real–world applications, the variability of observations may arise not only from sampling–level randomness, but also from unobserved factors or latent mechanisms, and such heterogeneity cannot be adequately captured by a two-parameter distribution. Therefore, introducing more flexible variance structures while preserving interpretability remains a key challenge in modeling positive continuous data.

To address these issues, this paper proposes a new family of distributions based on mixture of Type I gamma distribution. By introducing a random mixture variable τ into the Type I gamma distribution, we construct a *general mixture of Type I gamma* (Ge-Ga) distribution that retains a clearly interpretable while naturally accommodating additional sources of heterogeneity. This construction enables an effective decomposition of structural variability and random heterogeneity at the variance level. Within this unified framework, we further investigate several representative special cases and develop corresponding methods for parameter estimation and regression modeling. Compared with existing distributional models, the proposed approach exhibits greater flexibility and robustness in fitting over-dispersed and right-skewed data.

The remainder of this paper is organized as follows. In Section 2, we systematically introduce the general mixture of Type I gamma distribution and basic distributional properties, discuss parameter estimation, computing the MLE and *confidence interval* (CI), and extensions to mean regression models. In Section 3, we assume that the mixture variable τ follows several special cases of the generalized inverse Gaussian distribution family, leading to three specific statistical distributions, for which statistical inferences are developed. Section 4 presents numerical simulation studies to systematically assess the finite-sample properties and fitting performance of the proposed models. The fitting performance of the our proposed distributions on real data is demonstrated in Section 5. Finally, the last section concludes the paper and discusses potential directions for future research.

2. General mixture of Type I gamma distribution

2.1 Definition and basic properties of Ge-Ga distribution family

Let a positive *random variable* (r.v.) τ follow a general continuous distribution with *probability density function* (pdf) $f_\tau(\tau|\boldsymbol{\lambda})$, where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m)^\top$ is an unknown parameter vector. If a positive r.v. X has the following *mixture representation* (MR):

$$\tau \sim f_\tau(\tau|\boldsymbol{\lambda}) \quad \text{and} \quad X|\tau \sim \text{Ga}(\alpha, \mu\tau), \tag{2.1}$$

where $\text{Ga}(\cdot, \cdot)$ represents the *Type I gamma* (Ga) distribution in Appendix B.1, then X is said to follow the *general mixture of Type I gamma* (Ge-Ga) distributions, denoted by $X \sim \text{Ge-Ga}(\alpha, \mu, \boldsymbol{\lambda})$ or $\text{Ge-Ga}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = \{\alpha, \mu, \boldsymbol{\lambda}\}$ is the parameter vector, α and μ are two unknown positive parameters. Its pdf is

$$\text{Ge-Ga}(x|\boldsymbol{\theta}) = \frac{\alpha^\alpha x^{\alpha-1}}{\mu^\alpha \Gamma(\alpha)} \int_0^\infty \frac{f_\tau(\tau|\boldsymbol{\lambda})}{\tau^\alpha} \exp\left(-\frac{\alpha x}{\mu\tau}\right) \mathrm{d}\tau, \quad x > 0, \tag{2.2}$$

and the conditional distribution of $\tau|X$ is

$$f_{\tau|X}(\tau|x) \propto \frac{f_\tau(\tau|\boldsymbol{\lambda})}{\tau^\alpha} \exp\left(-\frac{\alpha x}{\mu\tau}\right),$$

which can be employed to calculate the conditional expectation of $\tau|(X, \boldsymbol{\theta})$ in the *expectation–maximization* (EM) and *data augmentation* (DA) algorithms. From the MR (2.1), we easily obtain the expectation, variance of Ge-Ga distribution as

$$E(X) = \mu E(\tau) \quad \text{and} \quad \text{Var}(X) = \mu^2 \left\{ \left(1 + \frac{1}{\alpha}\right) \text{Var}(\tau) + \frac{[E(\tau)]^2}{\alpha} \right\}.$$

Remark 2.1 (Variance decomposition model of gamma distribution). In the practice of specific distribution, in order to ensure the identifiability of parameters and facilitate the construction of mean regression model, we usually set the expectation $E(\tau) = 1$. Therefore, the mixture variable τ does not affect the mean value of Ge-Ga distribution, but only increases its variance, which divides the variance into two parts:

(1) The first one part is a constant/parameter

$$\text{Original Variance} = \frac{\mu^2}{\alpha},$$

which is common across individuals and reflects the structural or population–level intensity;

(2) The second part is related to mixture variable τ, which is given by

$$\text{Incremental Variance} = \mu^2\left(1+\frac{1}{\alpha}\right)\text{Var}(\tau),$$

accounting for random variation or individual-level heterogeneity.

Therefore, Ge-Ga distribution actually decompose the variance of original Type I gamma distribution by introducing mixture variable τ to MR (2.1), which consider the homogeneity in population and also the heterogeneity caused by various individuals, and are suitable for the over-dispersion data defined on $\mathbb{R}_+$. ‖

2.2 MLEs of parameters via the N-EM algorithm

Let $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} \text{Ge-Ga}(\alpha, \mu, \boldsymbol{\lambda})$ and $Y_{\text{obs}} = \{x_i\}_{i=1}^n$ denote the observed data, where x_i is the realization of X_i, then the log-likelihood function of $\boldsymbol{\theta} = \{\alpha, \mu, \boldsymbol{\lambda}\}$ is

$$\begin{aligned}\ell_*(\boldsymbol{\theta}|Y_{\text{obs}}) \overset{(2.2)}{=} & \; n[\alpha\log\alpha - \log\Gamma(\alpha)] - n\alpha\log\mu + (\alpha-1)\sum_{i=1}^n \log x_i \\ & + \sum_{i=1}^n \log\left[\int_0^\infty h_*(\tau|x_i,\boldsymbol{\theta})\,\mathrm{d}\tau\right],\end{aligned} \tag{2.3}$$

where the function $h_*(\cdot|x_i,\boldsymbol{\theta})$ is defined as

$$h_*(\tau|x_i,\boldsymbol{\theta}) = \frac{f_\tau(\tau|\boldsymbol{\lambda})}{\tau^\alpha}\exp\left(-\frac{\alpha x_i}{\mu\tau}\right), \quad i = 1,\ldots,n.$$

We notice that there are n complicated integrals in (2.3), therefore, some traditional optimization methods like Newton–Raphson algorithm, Fisher scoring algorithm and *minorization–maximization* (MM) algorithm are not appropriate to such situations. To address these challenges, we adopt the N-EM algorithm (Tian & Liu 2022) with the following three steps:

N-step: By normalizing $h_*(\cdot|x_i, \boldsymbol{\theta})$, we construct a *normalizing density function* (ndf)

$$g_*(\tau|x_i, \boldsymbol{\theta}) \triangleq \frac{h_*(\tau|x_i, \boldsymbol{\theta})}{\int_0^\infty h_*(s|x_i, \boldsymbol{\theta})\,\mathrm{d}s}, \quad \tau > 0,$$

so that $g_*(\tau|x_i, \boldsymbol{\theta}^{(t)})$ is also a density function defined on $\mathbb{R}_+$, where $\boldsymbol{\theta}^{(t)}$ denotes the t-th approximation of the MLE $\hat{\boldsymbol{\theta}}$.

E-step: To construct a minoring function satisfying $Q_*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) \leqslant \ell_*(\boldsymbol{\theta}|Y_{\text{obs}})$ and $Q_*(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t)}) = \ell_*(\boldsymbol{\theta}^{(t)}|Y_{\text{obs}})$, by using the integral version of Jensen's inequality, we obtain

$$\begin{aligned} Q_*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) &= c_*^{(t)} + n[\alpha\log\alpha - \log\Gamma(\alpha)] - n\alpha(\log\mu + \bar{d}_*^{(t)}) + \alpha\sum_{i=1}^n \log x_i \\ &\quad - \frac{\alpha}{\mu}\sum_{i=1}^n b_{i*}^{(t)} x_i + \sum_{i=1}^n \int_0^\infty \log[f_\tau(\tau|\boldsymbol{\lambda})] \cdot g_*(\tau|x_i, \boldsymbol{\theta}^{(t)})\,\mathrm{d}\tau, \end{aligned} \tag{2.4}$$

which minorizes $\ell_*(\boldsymbol{\theta}|Y_{\text{obs}})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$, where $c_*^{(t)}$ is a constant free from $\boldsymbol{\theta}$ and

$$b_{i*}^{(t)} = \int_0^\infty \frac{1}{\tau} \cdot g_*(\tau|x_i, \boldsymbol{\theta}^{(t)})\,\mathrm{d}\tau, \quad i = 1, \ldots, n,$$

$$d_{i*}^{(t)} = \int_0^\infty \log\tau \cdot g_*(\tau|x_i, \boldsymbol{\theta}^{(t)})\,\mathrm{d}\tau \quad \text{and} \quad \bar{d}_*^{(t)} \triangleq \frac{1}{n}\sum_{i=1}^n d_{i*}^{(t)}.$$

M-step: Given $\boldsymbol{\theta}^{(t)}$, by maximizing $Q_*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ with respect to $\boldsymbol{\theta}$, we update $\boldsymbol{\theta}^{(t)}$ by

$$\boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Q_*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}).$$

Given a pre-specified small positive number δ_0, the N-, E- and M-steps are alternately repeated until

$$\left|\frac{\ell_*(\boldsymbol{\theta}^{(t+1)}|Y_{\text{obs}}) - \ell_*(\boldsymbol{\theta}^{(t)}|Y_{\text{obs}})}{\ell_*(\boldsymbol{\theta}^{(t)}|Y_{\text{obs}})}\right| \leqslant \delta_0,$$

then we obtain the MLE $\hat{\boldsymbol{\theta}}$ of parameters $\boldsymbol{\theta}$.

We compute the first partial derivatives of $Q_*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ as

$$\frac{\partial Q_*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial\alpha} = n[C_*^{(t)} + \log\alpha - \psi(\alpha)],$$

$$\frac{\partial Q_*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial\mu} = -\frac{n\alpha}{\mu} + \frac{\alpha}{\mu^2}\sum_{i=1}^n b_{i*}^{(t)} x_i,$$

$$\frac{\partial Q_*(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})}{\partial\lambda_k} = \sum_{i=1}^n \int_0^\infty \frac{\partial\log f_\tau(\tau|\boldsymbol{\lambda})}{\partial\lambda_k} \cdot g_*(\tau|x_i, \boldsymbol{\theta}^{(t)})\,\mathrm{d}\tau, \quad k = 1, \ldots, m,$$

where the constant $C_*^{(t)}$ is defined as

$$C_*^{(t)} \triangleq 1 - \log\mu + \frac{1}{n}\sum_{i=1}^{n}\log x_i - \bar{d}_*^{(t)} - \frac{1}{n\mu}\sum_{i=1}^{n} b_{i*}^{(t)} x_i,$$

and $\psi(\alpha) = \mathrm{d}\log\Gamma(\alpha)/\,\mathrm{d}\alpha$ is the digamma function. We obtain the $(t+1)$-th approximation of $\hat{\boldsymbol{\theta}}$ as

$$\begin{cases} \alpha^{(t+1)} = \mathrm{sol}\Big\{C_*^{(t)} + \log\alpha - \psi(\alpha) = 0, \quad \alpha > 0\Big\}, \\[2ex] \mu^{(t+1)} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} b_{i*}^{(t)} x_i, \\[2ex] \lambda_k^{(t+1)} = \mathrm{sol}\left\{\displaystyle\sum_{i=1}^{n}\int_0^{\infty}\frac{\partial\log f_\tau(\tau|\boldsymbol{\lambda})}{\partial\lambda_k}\cdot g_*(\tau|x_i,\boldsymbol{\theta}^{(t)})\,\mathrm{d}\tau = 0\right\}, \quad k = 1,\dots,m, \end{cases} \tag{2.5}$$

where $\alpha^{(t+1)}$ can be solved by the US algorithm (Li & Tian 2022), see Appendix A.1.

2.3 Asymptotic normality of MLEs and CIs of parameters

One of the key questions in large–sample statistical inferences is to verify the asymptotic normality of MLEs, i.e.,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{\mathrm{true}}) \xrightarrow{\mathrm{D}} N_{m+2}(\mathbf{0}, \boldsymbol{\Sigma}) \quad \text{as} \quad n \to \infty,$$

where $\hat{\boldsymbol{\theta}}_n$ denotes the MLE of $\boldsymbol{\theta}$ based on the sample size n, $\boldsymbol{\theta}_{\mathrm{true}}$ is the true value of $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ is the variance–covariance matrix of the asymptotic normal distribution. The asymptotic normality depends on a series of assumptions known as regularization or Cramér–Rao conditions (Lehmann & Casella 1998, Lehmann 1999), which are presented in Theorem 1 of Appendix A.2. By verifying the Cramér–Rao conditions for Ge-Ga distributions (see Appendix A.3 for more details), we obtain the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ as

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_{\mathrm{true}}) \xrightarrow{\mathrm{D}} N_{m+2}\left(\mathbf{0}, \mathbf{J}^{-1}(\boldsymbol{\theta}_{\mathrm{true}})\right) \quad \text{as} \quad n \to \infty,$$

where $\mathbf{J}(\boldsymbol{\theta}_{\mathrm{true}})$ is the Fisher information matrix. In practice, we adopt the sandwich covariance matrix estimation to replace $\mathbf{J}^{-1}(\boldsymbol{\theta}_{\mathrm{true}})$, which can be computed as

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{H}}^{-1}\widehat{\mathbf{S}}\widehat{\mathbf{H}}^{-1} \triangleq (\hat{\sigma}_{kk'}), \quad k, k' = 1,\dots,m+2,$$

where $\widehat{\mathbf{H}} = -\nabla^2 \log[\ell_*(\boldsymbol{\theta}|Y_{\rm obs})]\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}$ is the observed information and $\widehat{\mathbf{S}} = (1/n)\sum_{i=1}^n \boldsymbol{s}_i\boldsymbol{s}_i^\top$ is the outer product of scores with $\boldsymbol{s}_i = \nabla\log[\text{Ge-Ga}(x_i|\boldsymbol{\theta})]\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}$. Finally, the $100(1-\alpha)\%$ asymptotic CI of θ_k can be constructed as $\hat{\theta}_{k,n} \pm z_{\alpha/2}\cdot\widehat{\text{se}}_k$, where $\widehat{\text{se}}_k = \sqrt{\hat{\sigma}_{kk'}}$.

2.4 Bootstrap CIs of $\boldsymbol{\theta}$

The large–sample method for calculating CIs of parameters has two limitations: (1) The asymptotic CI is not reliable for small to moderate sample sizes; and (2) even though for a large sample size, the range of CI may be beyond its bounds when the true value is close to the bounds. To overcome the two difficulties, this subsection presents the bootstrap method (Efron 1992) to construct the *bootstrap confidence interval* (BCI) (Hall 1988, Diciccio & Romano 1988) of $\vartheta \triangleq h(\boldsymbol{\theta})$, where $h(\cdot)$ is an arbitrary function of the parameter vector $\boldsymbol{\theta}$ in the Ge-Ga($\boldsymbol{\theta}$) distributions.

Based on the observations $Y_{\rm obs} = \{x_i\}_{i=1}^n$, we can first compute the MLEs $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ through the N-EM algorithm (2.5). Then, we generate a bootstrap sample $\{X_i^* = x_i^*\}_{i=1}^n \stackrel{\text{iid}}{\sim} \text{Ge-Ga}(\hat{\boldsymbol{\theta}})$ and compute the bootstrap replication $\hat{\vartheta}^* = h(\hat{\boldsymbol{\theta}}^*)$. Independently repeating this process G times, we obtain G bootstrap replications $\{\hat{\vartheta}^*(g)\}_{g=1}^G$. Hence, a $100(1-\alpha)\%$ BCI of ϑ is $[\hat{\vartheta}^*_{\rm L},\ \hat{\vartheta}^*_{\rm U}]$, where $\hat{\vartheta}^*_{\rm L}$ and $\hat{\vartheta}^*_{\rm U}$ are the $(\alpha/2)G$-th and the $(1-\alpha/2)G$-th order statistics of $\{\hat{\vartheta}^*(g)\}_{g=1}^G$.

2.5 The Ge-Ga mean regression model

Let $\{X_i\}_{i=1}^n \stackrel{\text{ind}}{\sim} \text{Ge-Ga}(\alpha, \mu_i, \boldsymbol{\lambda})$ and $Y_{\rm obs} = \{x_i, \boldsymbol{z}_{(i)}\}_{i=1}^n$ denote the observed data, we consider the following Ge-Ga mean regression model (McCullagh 2019)

$$\log E(X_i) = \boldsymbol{z}_{(i)}^\top\boldsymbol{\beta} \quad \text{or} \quad \mu_i = \frac{\exp(\boldsymbol{z}_{(i)}^\top\boldsymbol{\beta})}{E(\tau)}, \quad i = 1, \ldots, n,$$

where $\boldsymbol{z}_{(i)} = (z_{i1}, \ldots, z_{ip})^\top$ is a p-dimensional vector of covariates for the i-th subject and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is the regression coefficients with $p+1+m \leqslant n$. The log-likelihood function of $\boldsymbol{\gamma} = \{\alpha, \boldsymbol{\beta}, \boldsymbol{\lambda}\}$ is

$$\begin{aligned}\ell(\boldsymbol{\gamma}|Y_{\rm obs}) &= n[\alpha \log \alpha - \log \Gamma(\alpha)] - \alpha \sum_{i=1}^{n} \log \mu_i + (\alpha - 1) \sum_{i=1}^{n} \log x_i \\ &\quad + \sum_{i=1}^{n} \log \left[\int_0^\infty h(\tau|x_i, \boldsymbol{z}_{(i)}, \boldsymbol{\gamma}) \, \mathrm{d}\tau \right],\end{aligned}$$

where $h(\tau|x_i, \boldsymbol{z}_{(i)}, \boldsymbol{\gamma}) \triangleq f_\tau(\tau|\boldsymbol{\lambda}) \exp[-(\alpha x_i)/(\mu_i \tau)]/\tau^\alpha$. Similar to §2.2, we again adopt the N-EM algorithm to calculate the MLEs of $\boldsymbol{\gamma}$. The N-step is to compute the ndf

$$g(\tau|x_i, \boldsymbol{z}_{(i)}, \boldsymbol{\gamma}) = \frac{h(\tau|x_i, \boldsymbol{z}_{(i)}, \boldsymbol{\gamma})}{\int_0^\infty h(s|x_i, \boldsymbol{z}_{(i)}, \boldsymbol{\gamma}) \, \mathrm{d}s}, \quad \tau > 0,$$

and the E-step is to construct the surrogate function

$$\begin{aligned}Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)}) &= c^{(t)} + n[\alpha \log \alpha - \log \Gamma(\alpha)] - n\bar{d}^{(t)}\alpha + \alpha \sum_{i=1}^{n} (\log x_i - \log \mu_i) \\ &\quad - \alpha \sum_{i=1}^{n} \frac{b_i^{(t)} x_i}{\mu_i} + \sum_{i=1}^{n} \int_0^\infty \log[f_\tau(\tau|\boldsymbol{\lambda})] \cdot g(\tau|x_i, \boldsymbol{z}_{(i)}, \boldsymbol{\gamma}^{(t)}) \, \mathrm{d}\tau,\end{aligned}$$

where $c^{(t)}$ is a constant free from $\boldsymbol{\gamma}$, and for $i = 1, \ldots, n$, we define

$$\begin{aligned}b_i^{(t)} &\triangleq \int_0^\infty \frac{1}{\tau} \cdot g(\tau|x_i, \boldsymbol{z}_{(i)}, \boldsymbol{\gamma}^{(t)}) \, \mathrm{d}\tau, \\ d_i^{(t)} &\triangleq \int_0^\infty \log \tau \cdot g(\tau|x_i, \boldsymbol{z}_{(i)}, \boldsymbol{\gamma}^{(t)}) \, \mathrm{d}\tau \quad \text{and} \quad \bar{d}^{(t)} \triangleq \frac{1}{n} \sum_{i=1}^{n} d_i^{(t)}.\end{aligned}$$

Finally, the M-step is to solve the following system of equations

$$\begin{aligned}\frac{\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)})}{\partial \alpha} &= n[C^{(t)} + \log \alpha - \psi(\alpha)] = 0, \\ \frac{\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)})}{\partial \boldsymbol{\beta}} &= \alpha \sum_{i=1}^{n} \boldsymbol{z}_{(i)} \left(\frac{b_i^{(t)} x_i}{\mu_i} - 1 \right) = \mathbf{0}_p, \\ \frac{\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)})}{\partial \lambda_k} &= \frac{\partial E(\tau)}{\lambda_k} \cdot \alpha \left[\frac{n}{E(\tau)} - \sum_{i=1}^{n} \frac{b_i^{(t)} x_i}{\exp(\boldsymbol{z}_{(i)}^\top \boldsymbol{\beta})} \right] \\ &\quad + \sum_{i=1}^{n} \int_0^\infty \frac{\partial \log f_\tau(\tau|\boldsymbol{\lambda})}{\partial \lambda_k} \cdot g(\tau|x_i, \boldsymbol{z}_{(i)}, \boldsymbol{\gamma}^{(t)}) \, \mathrm{d}\tau = 0,\end{aligned}$$

where the constant $C^{(t)}$ is defined as

$$C^{(t)} \triangleq 1 + \frac{1}{n}\sum_{i=1}^{n}(\log x_i - \log \mu_i) - \bar{d}^{(t)} - \frac{1}{n}\sum_{i=1}^{n}\frac{b_i^{(t)} x_i}{\mu_i},$$

and we obtain the $(t+1)$-th approximation of $\hat{\boldsymbol{\gamma}}$ as

$$\begin{cases} \alpha^{(t+1)} = \mathrm{sol}\Big\{C^{(t)} + \log\alpha - \psi(\alpha) = 0, \quad \alpha > 0\Big\}, \\ \boldsymbol{\beta}^{(t+1)} = (\mathbf{Z}^\top \mathbf{W}^{(t)} \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{W}^{(t)} \boldsymbol{r}^{(t)}, \\ \lambda_k^{(t+1)} = \mathrm{sol}\left\{\dfrac{\partial Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)})}{\partial \lambda_k} = 0\right\}, \quad k = 1, \ldots, m, \end{cases} \tag{2.6}$$

where $\mathbf{Z} = (\boldsymbol{z}_{(1)}, \ldots, \boldsymbol{z}_{(n)})^\top$, $\mathbf{W}^{(t)} = \mathrm{diag}(b_1^{(t)}, \ldots, b_n^{(t)})$, and $\boldsymbol{r}^{(t)} = (r_1^{(t)}, \ldots, r_n^{(t)})^\top$ with

$$r_i^{(t)} = \boldsymbol{z}_{(i)}^\top \boldsymbol{\beta}^{(t)} + \frac{x_i}{\mu_i^{(t)}} - \frac{1}{b_i^{(t)}},$$

and the iteration of the regression coefficient $\boldsymbol{\beta}$ utilizes the IRLS of gamma regression, see more details in Appendix B.2.

2.6 Three specific cases of Ge-Ga distribution family

In the previous discussion, we assumed that mixture r.v. τ follows an arbitrary distribution defined on $\mathbb{R}_+$. In this section, we will fix the specific distribution of τ and discuss the special case of Ga-Ga distribution family. We study the following three distributions: inverse gamma, inverse Gaussian and reciprocal inverse Gaussian distributions. We define the so-called IGa-Ga, IGau-Ga and RIGau-Ga distributions and study their basic properties, statistical inference together with the corresponding mean regression models (see supplementary materials for details). This not only expands our alternative distribution for fitting right-skewness positive continuous data, but also brings the existing inverse beta distribution as a special case (i.e. IGa-Ga distribution) into our Ge-Ga family framework, and gives a unified statistical inference method. Under different parameter combinations, the density function images of three new distributions proposed by us are shown in Figure 1.

[Insert Figure 1 here.]

3. Simulation studies

3.1 Accuracy of MLEs and reliability of asymptotic normal CIs

In §2.6, we propose three specific distributions of the Ge-Ga distribution family and compute the MLEs via the N-EM algorithm. In addition, CIs of parameters can also be obtained by the sandwich method or the bootstrap method. To assess the performance of the proposed algorithms and methods, we conduct some numerical experiments by employing the *mean square error* (MSE) as the criterion to evaluate the accuracy of the MLE for each parameter and apply the *coverage rate* (CR) to confirm the reliability of CIs.

Consider different sample sizes $n = 200(300)800$ and repetition numbers $k = 1, \ldots, K$ with $K = 10{,}000$ in our simulations. We conduct some numerical experiments with following three cases:

Case (A_1): Independently generating $\{X_i^{(k)}\}_{i=1}^n \overset{\text{iid}}{\sim} \text{IGa-Ga}(\alpha, \mu, \lambda)$ with the true values being set as $\alpha = 3$, $\mu = 2$ and $\text{Var}(\tau) = 0.5$;

Case (A_2): Independently generating $\{X_i^{(k)}\}_{i=1}^n \overset{\text{iid}}{\sim} \text{IGau-Ga}(\alpha, \mu, \lambda)$ with the true values being set as $\alpha = 2$, $\mu = 3$ and $\text{Var}(\tau) = 0.2$;

Case (A_3): Independently generating $\{X_i^{(k)}\}_{i=1}^n \overset{\text{iid}}{\sim} \text{RIGau-Ga}(\alpha, \mu, \lambda)$ with the true values being set as $\alpha = 1$, $\mu = 2$ and $\text{Var}(\tau) = 0.96$;

Given a combination of $(n, k, \alpha, \mu, \text{Var}(\tau))$ for Case (A_1), let $Y_{\text{obs}}^{(k)} = \{x_i^{(k)}\}_{i=1}^n$ denote the k-th generated observations, where $x_i^{(k)}$ is a realization of $X_i^{(k)}$. We use the N-EM algorithm to compute the MLEs that are denoted by $\{\hat{\alpha}^{(k)}, \hat{\mu}^{(k)}, \hat{\lambda}^{(k)}\}$. Next, we calculate the *average MLE* (A-MLE) for each parameter and the corresponding MSE based on the K repetitions. The CIs of parameters are obtained by the sandwich method, the CR of CIs is presented in Table 1 together with A-MLE and MSE. Similarly, for Case (A_2) and (A_3), the corresponding results are also reported in Table 1.

[Insert Table 1 here]

We found that with the increase of sample size n, the MLEs of parameters becomes more and more accurate and stable. Specifically, A-MLE is getting closer to the set true value, and MSE is getting smaller and smaller. For interval estimation, a small sample size (such as $n = 200$) may lead to dissatisfaction with asymptotic normality, and CR is slightly lower than 95%. At this time, we recommend adopting bootstrap method to build BCIs. When the sample size increases, the asymptotic normality of the MLE is better, and the CR is stable at about 95%, so a more reliable interval estimation is obtained.

3.2 Model comparisons

To assess the performance of the goodness-of-fit tests among the three specific distributions (i.e., IGa-Ga, IGau-Ga and RIGau-Ga) of the Ge-Ga distribution family and three commonly-employed distributions (i.e., gamma, inverse Gaussian and Weibull), we adopt two evaluation criteria: The *Akaike information criterion* (AIC, Akaike 1974), and the *Bayesian information criterion* (BIC, Schwarz 1978).

In our numerical studies, the sample size, the number of sampling replications are respectively set to be $n = 800$ and $K = 1{,}000$. We set the parameters $\alpha = 2$, $\mu = 3$ and λ in each of the three specific distributions (denoted by $\lambda_{\text{IGa-Ga}}$, $\lambda_{\text{IGau-Ga}}$, $\lambda_{\text{RIGau-Ga}}$, respectively) can be determined by assuming that $\text{Var}(\tau) = 0.5, 1$.

For a given $\{n, k, \alpha, \mu, \text{Var}(\tau)\}$ with $k = 1, \ldots, K$, firstly, based on the theoretical mean and variance of the Ge-Ga distributions, we use the method of moment estimation to obtain the parameters corresponding to the gamma, inverse Gaussian and Weibull distributions, denoted as $\{\alpha_{\text{Ga}}, \mu_{\text{Ga}}, \mu_{\text{IGau}}, \lambda_{\text{IGau}}, k_{\text{Wei}}, \lambda_{\text{Wei}}\}$. Secondly, we independently generate the samples $\{X_i^{(k)} = x_i^{(k)}\}_{i=1}^n$ from the mixed distribution:

$$
\begin{aligned}
& p_1 \cdot \text{IGa-Ga}(\alpha, \mu, \lambda_{\text{IGa-Ga}}) + p_2 \cdot \text{IGau-Ga}(\alpha, \mu, \lambda_{\text{IGau-Ga}}) + p_3 \cdot \text{RIGau-Ga}(\alpha, \mu, \lambda_{\text{RIGau-Ga}}) \\
& + p_4 \cdot \text{Ga}(\alpha_{\text{Ga}}, \mu_{\text{Ga}}) + p_5 \cdot \text{IGaussian}(\mu_{\text{IGau}}, \lambda_{\text{IGau}}) + p_6 \cdot \text{Weibull}(k_{\text{Wei}}, \lambda_{\text{Wei}}).
\end{aligned}
$$

where $p_s \geqslant 0$ for $s = 1, \ldots, 6$ with $\sum_{s=1}^6 p_s = 1$. We consider the following situations for selecting $\boldsymbol{p} = (p_1, \ldots, p_6)^\top$:

Case (B_1): Single–component distribution: $p_s = 1$ for $s = 1, \ldots, 3$ and $p_{s'} = 0$ with $s' \neq s$, indicating that the sample is generated from each of the proposed three specific distributions;

Case (B_2): Dominant one–component mixed distribution: One specific case of Ge-Ga distribution has a larger weight p_s and the other two have smaller weights $p_{s'}$;

Case (B_3): Uniform six–component mixed distribution: $\boldsymbol{p} = 1/6 \times \mathbf{1}_6$;

Case (B_4): Mixed distribution dominated by traditional distribution: Three specific cases of Ge-Ga distribution with small weight $p_s = 1/9$ (or even $p_s = 0$), and three traditional distributions with large weight $p_{s'} = 2/9$ (or $p_{s'} = 1/3$).

Based on the generated sample $\{x_i^{(k)}\}_{i=1}^n$, we can compute the MLEs $\{\hat{\alpha}^{(k)}, \hat{\mu}^{(k)}, \hat{\lambda}^{(k)}\}$ for the three specific distributions of the Ge-Ga distribution family via the N-EM algorithm, calculate three AIC values $\{\mathsf{AIC}_1^{(k)}, \mathsf{AIC}_2^{(k)}, \mathsf{AIC}_3^{(k)}\}$ and fit the gamma, inverse Gaussian, Weibull distributions with $\mathsf{AIC}_4^{(k)}, \mathsf{AIC}_5^{(k)}, \mathsf{AIC}_6^{(k)}$ by using Newton-Raphson algorithm. Then we sort AICs to obtain the corresponding six ranks $\{\mathsf{Rank}_1^{(k)}, \ldots, \mathsf{Rank}_6^{(k)}\}$, where $\mathsf{Rank}_s^{(k)}$ represents the rank of $\mathsf{AIC}_s^{(k)}$ in $\{\mathsf{AIC}_1^{(1)}, \ldots, \mathsf{AIC}_6^{(k)}\}$. Finally, the average AICs (denoted by AIC) and the average ranks (denoted by Rank) are calculated by

$$\mathsf{AIC}_s = \frac{1}{K} \sum_{k=1}^{K} \mathsf{AIC}_s^{(k)} \quad \text{and} \quad \mathsf{Rank}_s = \frac{1}{K} \sum_{k=1}^{K} \mathsf{Rank}_s^{(k)}, \quad s = 1, \ldots, 6.$$

Table 2 and Table 3 presents the AICs and Ranks of six fitted distributions. From the result in Table 2 and Table 3, we found that the generated Ge-Ga distribution provides the best fitting effect, which outstanding performance is significantly better than other distributions. And the IGau-Ga distribution has good fitting effect in various mixed cases.

[Insert Table 2 and Table 3 here.]

4. Real data analysis

In this section, we apply the proposed models and estimation methods to analyze a real data set and compare them with the traditional models to illustrate that the Ge-Ga distributions have a more flexible and robust fitting performance.

4.1 Data description

The real data set considered in this study is a personal medical cost data set, which is publicly available from Kaggle (https://www.kaggle.com/datasets/mirichoi0218/insurance/data). This data set has been widely used in the literature as a benchmark example for modeling positive and right-skewed medical expenditure data (Jecong & Valdez 2020).

The data consist of observations on 1,338 individuals, each corresponding to a policy-holder covered by a health insurance plan. The primary response variable of interest is the individual medical charges (denoted as charges with unit k \$), representing the total medical expenditure incurred by each individual. This variable takes strictly positive values and exhibits pronounced right-skewness, with a small proportion of individuals incurring substantially higher costs than the majority of the population. Such characteristics make the data particularly suitable for evaluating statistical models designed for positive, over-dispersed continuous outcomes.

In addition to the response variable, the data set contains several demographic and health-related covariates. Specifically, the available explanatory variables include:

age: the age of the individual (in years);

sex: a binary indicator of gender (female or male);

BMI: body mass index, a continuous measure commonly used to assess body fat;

children: the number of dependents covered by the insurance plan;

smoker: a binary indicator of smoking status (yes or no);

region: residential region in the United States (northeast, northwest, southeast, southwest).

These covariates are routinely considered important determinants of medical expenditure and have clear interpretations in health economics and biostatistical studies. In particular, variables such as age, body mass index, and smoking status are known to be strongly associated with both the level and variability of medical costs, suggesting the presence of substantial heterogeneity across individuals.

In §4.2, we apply the three specific distributions of the Ge-Ga distribution family to fit individual medical expenditure, with the aim of systematically characterizing the overall distributional features of the data, particularly their skewness and dispersion. The proposed distributions are further compared with the gamma, inverse Gaussian and Weibull distributions to assess their relative performance in capturing right-skewed and over-dispersed data. Then in §4.3, we further construct mean regression models to investigate the relationship between individual medical expenditures and the associated covariates (De & Heller 2008, Goldburd *et al.* 2016, Gagnon & Wang 2024, Hayawi *et al.* 2025). The proposed regression models are compared with existing mean regression models, like gamma regression (Manning *et al.* 2002) and inverse Gaussian regression, in terms of goodness-of-fit, stability of parameter estimation, and predictive accuracy, thereby providing a comprehensive evaluation of their practical performance in real data analysis.

4.2 Distribution fitting of medical expenses

In this section, we use three cases of Ge-Ga distribution compared with gamma, inverse Gaussian and Weibull distributions to fit the variable charges, so as to describe the distribution characteristics and morphology of individual medical expenditure. We adopt the log-likelihood value, AIC and BIC as evaluation criteria to evaluate the performance of fitting.

[Insert Table 4 here.]

The comparison of the fitting performance for six distributions is presented in Table 4. We found that majority of the distributions fit the mean accurately, but the gamma and Weibull distribution are seriously inadequate in variance fitting, which is limited by the coupling between the variance and the mean for these two distributions. In contrast, the heavy tailed inverse Gaussian and the variance decomposition Ge-Ga distribution family proposed in this paper have better fitting effect.

[Insert Figure 2 here.]

At the same time, IGau-Ga, RIGau-Ga and inverse Gaussian distributions have larger log-likelihood and smaller AIC, BIC, indicating that these three distributions have the best fitting effect. We provide the MLEs and CIs of parameters for these three best fitting distributions in Table 5 as follows.

[Insert Table 5 here.]

4.3 Construction and prediction of mean regression model

In this section, we divide all 1338 sample points into training set and test set according to the ratio $1000 + 388$, and assume that individual medical expenditure for each sample with

$$\mathsf{charges}_i \overset{\text{ind}}{\sim} \text{IGa-Ga/IGau-Ga/RIGau-Ga}(\alpha, \mu_i, \lambda) \;\; \text{or} \;\; \text{Ga}(\alpha, \mu_i)/\text{IGaussian}(\mu_i, \lambda)$$

Then we construct the corresponding mean regression models as

$$\begin{aligned} \log \mu_i \;\; &= \;\; \beta_0 + \beta_1 \times \mathsf{age}_i + \beta_2 \times \mathsf{BMI}_i + \beta_3 \times \mathsf{children}_i + \beta_4 \times I(\mathsf{Sex}_i = \text{male}) \\ & \quad + \beta_5 \times I(\mathsf{smoker}_i = \text{yes}), \quad i = 1, \ldots, 1000. \end{aligned}$$

The comparison of the fitting effects of the five mean regression models on the training set is shown in Table 6. We found that the Ge-Ga mean regression models have better fitting performance than the existing gamma and inverse Gaussian mean regression models as a whole. In three Ge-Ga models, the IGa-Ga and IGau-Ga models have the best fitting effect.

[Insert Table 6 here.]

The estimated values of the regression coefficients of the five mean regression models are shown in Table 7. The impact of covariates depicted by each regression model on medical expenses is basically the same. As people grow older, the increase of BMI and the number of children raised will increase the expenditure of medical expenses. At the same time, the average medical expenditure of men is $\exp(\hat{\beta}_4)$ times that of women. Most importantly, we found that smoking can significantly affect medical costs, which also confirms that smoking is indeed harmful to human health.

[Insert Table 7 here.]

Finally, we present the fitting effect and prediction accuracy of each regression model on the test set in Table 8. On the test set, the log-likelihood, AIC and BIC of IGa-Ga and IGau-Ga models are still excellent, which also means that these two models have the most outstanding ability of data fitting and modeling.

[Insert Table 8 here.]

But at the same time, gamma regression model has the smallest MAE, the most robust prediction performance for the mean μ. This is because gamma regression implies hypothesis $\mathrm{Var}(\mathsf{charges}|\mu) \propto \mu^2$, while inverse Gaussian and Ge-Ga regression models with

$$\mathrm{Var}(\mathsf{charges}|\mu) \propto \mu^3 \quad \text{and} \quad \mathrm{Var}(\mathsf{charges}|\mu) \propto \mu^2 + \text{Incremental Variance}.$$

The empirical results suggest that while the gamma regression remains competitive for mean prediction, it fails to adequately capture the tail behavior of the response distribution. The proposed Gamma mixture models substantially improve distributional fit, as evidenced by higher predictive likelihoods, at the cost of a modest increase in MAE.

5. Discussions

In this paper, we proposed a new family of gamma mixture distributions (Ge-Ga distribution family) to solve the common problems of right-skewness, over-dispersion and potential individual heterogeneity in positive continuous data. By introducing random mixture variables τ into the Type I gamma distribution $\mathrm{Ga}(\alpha, \mu)$, the model not only maintains the simplicity of the model structure and the interpretability of the parameters, but also significantly enhances the ability to describe the complex variance structure, providing a unified and flexible probability framework for the modeling of positive continuous data.

From the theoretical point of view, this paper systematically constructed the mixture representation of Ge-Ga distribution family, and derived its probability density function, moment property and mean variance structure. By normalizing the mixture variables, the model can effectively decompose the sources of variance, thus providing a clear statistical perspective for explaining the phenomenon of over-dispersion. In terms of inference methods, this paper estimated the parameters of Ge-Ga distributions uniformly based on the normalized expectation maximization (N-EM) algorithm, and establishes the asymptotic normality of the MLE under Cramér–Rao conditions, which provides a theoretical basis for statistical inference and uncertainty evaluation. In terms of model construction and application, this paper selected a variety of generalized inverse Gaussian distributions as mixture variables, and constructed several representative Ge-Ga special cases, which are further extended to the regression modeling framework.

Numerical simulation results show that this kind of model has good estimation accuracy and stability. In addition, in the model comparison experiment, the Ge-Ga distribution family shows better fitting performance under a variety of mixed generation situations, especially in the case of obvious heterogeneity or over-dispersion. In the real data analysis, the Ge-Ga distribution and their regression models were applied to the individual medical expenditure data set. The results show that the Ge-Ga models is more robust and consistently better than the existing gamma distribution and some commonly used alternative models in terms of information criteria, goodness of fit and parameter interpretation. In particular, the additional flexibility introduced by the mixture variable enables the model to more effectively depict the potential heterogeneity in the data, thus improving the overall depiction ability of the model on the mean relationship and variation structure.

In the future research, there are still a number of potentially valuable issues worthy of further discussion. Firstly, it is necessary to further study the extension of the representation framework in this paper to the multivariate case

$$\tau \sim f_\tau(\tau|\lambda) \quad \text{and} \quad X_j|\tau \sim \mathrm{Ga}(\alpha, \mu_j\tau),$$

for $j = 1, \ldots, d$, which can characterize the multivariate continuous positive data with positive correlation, right skewness, over-dispersion and individual heterogeneity, so as to expand the application scope of Ge-Ga distribution family in complex data structures. Secondly, the model proposed in this paper mainly characterizes the mixture variable τ based on the parametric distribution. A natural extension direction is to introduce non-parametric or semi-parametric methods to model the mixture variable, so as to obtain a more flexible and more universal representation of it. Finally, in the context of multivariate modeling, how to construct a reasonable mechanism to depict the positive and negative correlation structure at the same time is still a challenging but significant research problem, which needs to be further explored in future work.

Supplementary materials

In the supplement, we provide three special cases of Ge-Ga distribution family in detail, specifically introduce their definitions, basic properties, statistical inference and corresponding mean regression models. At the same time, we added some US algorithms we constructed and the related distributions we need to use. Interested readers can see the supplement in https://github.com/YuanfanZhao/GeGa/blob/main/Thesis/Supplementary_Materials.pdf.

Acknowledgement

Reference

Akaike H (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, **19**(6), 716-723.

Barndorff-Nielsen O & Halgreen C (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **38**(4), 309-311.

Bownan KO & Shenton LR (1982). Properties of estimators for the gamma distribution, *Communications in Statistics–Simulation and Computation*, **11**(4), 377-519. DOI: 10.1080/03610918208812270

Chhikara RS & Folks JL (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications.* Marcel Dekker, New York.

Cordeiro GM & Lemonte AJ (2012). The McDonald inverted beta distribution. *Journal of the Franklin Institute*, **349**(3), 1174-1197.

Dubey SD (1970). Compound gamma, beta and F distributions. *Metrika*, **16**(1), 27-31.

Diciccio TJ & Romano JP (1988). A review of bootstrap confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **50**(3), 338-354.

De Jong P & Heller GZ (2008). *Generalized linear models for insurance data.* Cambridge University Press, Cambridge.

Efron B (1992). *Bootstrap methods: another look at the jackknife. In Breakthroughs in statistics: Methodology and distribution (pp. 569-593).* Springer, New York.

Goldburd M, Khare A, Tevet D & Guller D (2016). Generalized linear models for insurance rating. *Casualty Actuarial Society, CAS Monographs Series*, **5**, 77.

Gagnon P & Wang Y (2024). Robust heavy-tailed versions of generalized linear models with applications in actuarial science. *Computational Statistics & Data Analysis*, **194**, 107920.

Hall P (1988). On symmetric bootstrap confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **50**(1), 35-45.

Huber PJ and Ronchetti EM (2011). *Robust statistics.* John Wiley & Sons, New York.

Hayawi H, Ali TH & Omer AW (2025). Gamma regression model for insurance claims analysis: A comparative study on link functions and their influence on cost predictions. Preprints, 202508.0841. v1.

Jorgensen, B. (2012). *Statistical properties of the generalized inverse Gaussian distribution (Vol. 9).* Springer Science & Business Media, Berlin.

Jeong H & Valdez EA (2020). Predictive compound risk models with dependence. *Insurance: Mathematics and Economics*, **94**, 182-195.

Lehmann EL & Casella G (1998). *Theory of Point Estimation.* Springe, New York.

Lehmann EL (1999). *Elements of Large Sample Theory.* Springer, New York.

Li X and Tian GL (2022). The upper–crossing/solution (US) algorithm for root–finding with strongly stable convergence. arXiv preprint arXiv:2212.00797

Manning WG, Basu A, Mullahy J & Manning W (2002). *Modeling costs with generalized gamma regression.* The University of Chicago, Chicago.

McCullagh, P. (2019). *Generalized linear models.* Routledge, Oxford.

Norman L, Johnson, Samuel Kotz & N. Balakrishnan (1994). *Continuous univariate distributions.* John Wiley, New York.

Schwarz G (1978). Estimating the dimension of a model. *The annals of statistics*, 461-464.

Tian GL and Liu XY (2022). The normalized expectation–maximization (N-EM) algorithm. *Technical Report at Department of Statistics and Data Science of SUSTech.*

Weibull W (1951). A statistical distribution function of wide application. *Journal of applied mechanics*, **18**, 287-293.

A. Some technical derivations

A.1 The root of $C + \log s - \psi(s) = 0$ with $s > 0$ via a US algorithm

The aim of this appendix is to solve the root of the equation $g(s) = C + \log s - \psi(s) = 0$ with $s > 0$ by employing a US algorithm. We know the digamma function $\psi(\cdot)$ is defined as

$$\psi(s) \triangleq \frac{\Gamma'(s)}{\Gamma(s)} = -\gamma_0 + \sum_{m=0}^{\infty} \left(\frac{1}{m+1} - \frac{1}{m+s} \right),$$

where $\gamma_0 = \lim\limits_{n\to\infty} \left(\sum\limits_{m=1}^{n} m^{-1} - \log n \right) \approx 0.5772$ is the Euler–Mascheroni constant. Since

$$\begin{aligned} g'(s) &= \frac{1}{s} - \psi'(s) > -\psi'(s) = -\sum_{m=0}^{\infty} \frac{1}{(m+s)^2} = -\frac{1}{s^2} - \sum_{m=1}^{\infty} \frac{1}{(m+s)^2} \\ &> -\frac{1}{s^2} - \sum_{m=1}^{\infty} \frac{1}{m^2} = -\frac{1}{s^2} - \frac{\pi^2}{6} \triangleq b(s), \end{aligned}$$

the US iteration for calculating $s^{(t+1)}$ is

$$\begin{aligned} s^{(t+1)} &= \text{sol}\left\{ g(s^{(t)}) + \int_{s^{(t)}}^{s} b(z)\,\mathrm{d}z = 0,\ \forall\ s, s^{(t)} > 0 \right\} \\ &= \text{sol}\left\{ g(s^{(t)}) + \frac{1}{s} - \frac{\pi^2}{6}s - \frac{1}{s^{(t)}} + \frac{\pi^2}{6}s^{(t)} = 0,\ \forall\ s, s^{(t)} > 0 \right\} \\ &= \text{sol}\left\{ \frac{\pi^2}{6}s^2 - q^{(t)}s - 1 = 0,\ \forall\ s, s^{(t)} > 0 \right\} \\ &= \frac{q^{(t)} + \sqrt{q^{(t)2} + 2\pi^2/3}}{\pi^2/3}, \end{aligned}$$

where

$$q^{(t)} = C + \log s^{(t)} - \psi(s^{(t)}) + \frac{\pi^2 s^{(t)}}{6} - \frac{1}{s^{(t)}}.$$

A.2 Asymptotic normality theorem for MLE

Suppose that $\{X_i\}_{i=1}^n \stackrel{\text{iid}}{\sim} F(x|\boldsymbol{\theta})$, the likelihood function be expressed as $\mathcal{L}_n(\boldsymbol{\theta})$. Then we calculate the MLE $\hat{\boldsymbol{\theta}}_n$. To study the consistency and asymptotic normality of $\hat{\boldsymbol{\theta}}_n$, we need a laundry list of assumptions, which has come to be known as the Cramér–Rao conditions.

Theorem 1 (Cramér–Rao conditions)

Let $X_1, \ldots, X_n$ be i.i.d. r.v. with density $f(x|\boldsymbol{\theta})$ and *cumulative distribution function* (cdf) $F(x|\boldsymbol{\theta})$. In order to study the consistency and asymptotic normality of MLE $\hat{\boldsymbol{\theta}}_n$ of parameter $\boldsymbol{\theta}$, we assume the following assumptions:

(A1) Identifiability, that is, $F(x|\boldsymbol{\theta}_1) = F(x|\boldsymbol{\theta}_2) \Longleftrightarrow \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$;

(A2) $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ is an open field in $\mathbb{R}^d$;

(A3) The observations are $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$, where the X_i are i.i.d. with probability density $f(x|\boldsymbol{\theta})$ which will be assumed to be continuous in x;

(A4) The support $\mathbb{S} = \{x : \ f(x|\boldsymbol{\theta}) > 0\}$ is free from $\boldsymbol{\theta}$;

(A5) $\forall\ x \in \mathbb{S}$, $f(x|\boldsymbol{\theta})$ is differentiable with respect to $\boldsymbol{\theta}$, that is, the likelihood function is smooth as a function of the parameter.

Let $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta}$ be the true value of $\boldsymbol{\theta}$. Then there exists a sequence $\hat{\boldsymbol{\theta}}_n$ such that $\hat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ in probability. Under the condition of consistency, Assumption (A6) and (A7) provide sufficient conditions for the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$.

(A6) For all $x \in \mathbb{S}$ (specified in (A4)), the density $f(x|\boldsymbol{\theta})$ is three times differentiable with respect to $\boldsymbol{\theta}$ and the third derivative is continuous. The corresponding integral of partial derivative

$$\int \frac{\partial^3 f(x|\boldsymbol{\theta})}{\partial\theta_r \partial\theta_s \partial\theta_t}\, \mathrm{d}x$$

can be obtained by differentiating under the integral sign;

(A7) There exists a positive number $c(\boldsymbol{\theta}_0)$ and a function $M_{\boldsymbol{\theta}_0}(x)$ such that

$$\left|\frac{\partial^3}{\partial\theta_r \partial\theta_s \partial\theta_t} \log f(x|\boldsymbol{\theta})\right| \leqslant M_{\boldsymbol{\theta}_0}(x) \quad \text{for all} \quad x \in \mathbb{S},\ ||\boldsymbol{\theta} - \boldsymbol{\theta}_0|| < c(\boldsymbol{\theta}_0),$$

and $E_{\boldsymbol{\theta}_0}(M_{\boldsymbol{\theta}_0}(X)) < \infty$.

Then any consistent sequence $\hat{\boldsymbol{\theta}}_n$ satisfies

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{\mathrm{D}} N\left(\mathbf{0}, \mathbf{J}^{-1}(\boldsymbol{\theta}_0)\right),$$

where $\mathbf{J}(\boldsymbol{\theta}) = [J_{ij}(\boldsymbol{\theta})]$ with

$$J_{ij}(\boldsymbol{\theta}) = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log\mathcal{L}(\boldsymbol{\theta}|\boldsymbol{X})\right]$$

is the Fisher information matrix, here $\boldsymbol{X} = (X_1, \ldots, X_n)^\top$. $\|$

A.3 Validation of Cramér–Rao conditions for Ge-Ga distributions

Suppose that $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} \text{Ge-Ga}(\alpha, \mu, \boldsymbol{\lambda})$ with parameters $\alpha, \mu > 0$ and an arbitrary mixture density $f_\tau(\tau|\boldsymbol{\lambda})$, which satisfies:

(C1) $\boldsymbol{\lambda} \in \boldsymbol{\Lambda}$ is an open field in $\mathbb{R}^m$;

(C2) $f_\tau(\tau|\boldsymbol{\lambda})$ is three times differentiable with respect to $\boldsymbol{\lambda}$ and the third derivative is continuous.

Because the pdf of Ge-Ga, which is given by (2.2), can be expressed as

$$\text{Ge-Ga}(x|\alpha, \mu, \boldsymbol{\lambda}) = \frac{\alpha^\alpha x^{\alpha-1}}{\mu^\alpha\Gamma(\alpha)}\int_0^\infty \frac{f_\tau(\tau|\boldsymbol{\lambda})}{\tau^\alpha}\exp\left(-\frac{\alpha x}{\mu\tau}\right)\,\mathrm{d}\tau, \quad x > 0,$$

This is an elementary function, therefore, the analytic properties of Ge-Ga distributions are inherited from the kernel $f_\tau(\tau|\boldsymbol{\lambda})$ inside the integral. Firstly, in constructing the specific Ge-Ga distribution, we ensure the identifiability of the parameters by setting $E(\tau) = 1$ (see §3 for details), which guarantees the validity of Assumption (A1). Note that the parameter space $\boldsymbol{\Theta} \triangleq \mathbb{R}_+^2 \times \mathbb{R}_+^m$ is an open field and the support of Ge-Ga distributions are $\mathbb{S} = \mathbb{R}_+$, which does not depend on any parameter, hence, the Assumptions (A2) and (A4) hold. Assumption (A3) will automatically hold when we introduce r.v. sequence $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} \text{Ge-Ga}(\alpha, \mu, \boldsymbol{\lambda})$. Regarding Assumption (A5), we know the term

$$\frac{\alpha^\alpha x^{\alpha-1}}{\mu^\alpha\tau^\alpha\Gamma(\alpha)}\exp\left(-\frac{\alpha x}{\mu\tau}\right)$$

is differentiable with respect to α and μ, we just need $f_\tau(\tau|\boldsymbol{\lambda})$ is once differentiable with respect to $\boldsymbol{\lambda}$ under integration, which holds by Condition (C2). Therefore, we obtain the MLEs $\hat{\boldsymbol{\theta}}$ of Ge-Ga distributions converge to the true value $\boldsymbol{\theta}_0$ in probability.

Further verify the asymptotic normality, because $f_\tau(\tau|\boldsymbol{\lambda})$ is three times differentiable with respect to $\boldsymbol{\lambda}$ and the third derivative is continuous, therefore, the density $\text{Ge-Ga}(x|\alpha,\mu,\boldsymbol{\lambda})$ composed of basic elementary functions will also hold these properties, which means the Assumption (A6) holds. Finally, for the smooth function $\text{Ge-Ga}(x|\alpha,\mu,\boldsymbol{\lambda})$, its derivatives are bounded in the neighborhood $c(\boldsymbol{\theta}_0)$ of $\boldsymbol{\theta}_0$. Therefore, there exists a constant C such that

$$\left|\frac{\partial^3}{\partial\theta_r\partial\theta_s\partial\theta_t}\log f(x|\boldsymbol{\theta})\right| \leqslant C \triangleq M_{\boldsymbol{\theta}_0}(x) \ \text{ for all } \ x \in \mathbb{S},\ ||\boldsymbol{\theta}-\boldsymbol{\theta}_0|| < c(\boldsymbol{\theta}_0),$$

and the expectation $E_{\boldsymbol{\theta}_0}(M_{\boldsymbol{\theta}_0}(X)) = C < \infty$. In conclusion, Ge-Ga distributions satisfy the Cramér–Rao conditions. The MLEs of parameters satisfies the consistency and asymptotic normality.

B. Type I gamma distribution and mean regression model

B.1 Type I gamma distribution

The pdf of the r.v. $Y \sim \text{Gamma}(\alpha,\beta)$ is given by (1.1), where $\alpha\,(>0)$ is called the shape parameter, $\beta\,(>0)$ the rate parameter, and $\mu \triangleq \alpha/\beta\,(>0)$ the mean parameter. If we adopt parameters $\{\alpha,\mu\}$ instead of $\{\alpha,\beta\}$, then the reparametrized gamma distribution is called *Type I gamma* (Ga) distribution, denoted by $X \sim \text{Ga}(\alpha,\mu)$ with pdf

$$\text{Ga}(x|\alpha,\mu) = \frac{\alpha^\alpha}{\mu^\alpha\Gamma(\alpha)}x^{\alpha-1}\exp\left(-\frac{\alpha x}{\mu}\right), \quad x>0.$$

The *moment generating function* (mgf) and the k-th moments of X are given by

$$\begin{aligned} M_X(t) &= \left(\frac{\alpha}{\alpha-\mu t}\right)^\alpha, \quad t<\frac{\alpha}{\mu} \quad \text{and} \\ E(X^k) &= \left(\frac{\mu}{\alpha}\right)^k\prod_{j=0}^{k-1}(\alpha+j), \quad k\in\mathbb{N}_+ \triangleq \{1,2,\ldots,\infty\}. \end{aligned}$$

Therefore, we can easily compute the expectation, variance and *coefficient of variation* (CV) of X as

$$E(X)=\mu, \quad \text{Var}(X)=\frac{\mu^2}{\alpha} \quad \text{and} \quad \text{CV}(X) \triangleq \frac{\sqrt{\text{Var}(X)}}{E(X)} = \sqrt{\frac{1}{\alpha}},$$

respectively. Thus, we say μ the mean parameter of Ga distribution and α the shaper (or CV) parameter. In addition, Ga distribution has the following properties:

Theorem 2 (Some properties of Ga distribution)

1° If $Y \sim \mathrm{Ga}(\alpha, \mu)$, then $Y^{-1} \sim \mathrm{IGamma}(\alpha, \alpha/\mu)$, whose pdf is defined as

$$\mathrm{IGamma}(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \mathrm{e}^{-\beta/y}, \quad y > 0;$$

2° $\mathrm{Ga}(1, \mu) = \mathrm{Exponential}(1/\mu)$;

3° $\mathrm{Ga}(\nu/2, \nu) = \chi^2(\nu)$;

4° If $Y \sim \mathrm{Ga}(\alpha, \mu)$, then $cY \sim \mathrm{Ga}(\alpha, c\mu)$ for any constant $c > 0$. ‖

B.2 Type I gamma mean regression model and iterative reweighted least squares approach

Suppose that $\{Y_i\}_{i=1}^n \stackrel{\text{ind}}{\sim} \mathrm{Ga}(\alpha, \mu_i)$ with common shape parameter α and each mean parameter μ_i, we consider the following Type I gamma mean regression model

$$\log \mu_i = \boldsymbol{z}_{(i)}^\top \boldsymbol{\beta} \quad \text{or} \quad \mu_i = \exp\left(\boldsymbol{z}_{(i)}^\top \boldsymbol{\beta}\right),$$

where $\boldsymbol{z}_{(i)}^\top = (z_{i1}, \ldots, z_{ip})^\top$ is a p-dimensional vector of covariates for the i-th subject and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is the regression coefficients with $p+1 \leqslant n$. The log-likelihood function of $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^\top)^\top$ is

$$\ell(\boldsymbol{\theta}|Y_{\text{obs}}) = \text{constant} + n\left[\left(\log\alpha + \frac{1}{n}\sum_{i=1}^n \log y_i\right)\alpha - \log\Gamma(\alpha)\right] - \alpha\sum_{i=1}^n\left(\log\mu_i + \frac{y_i}{\mu_i}\right),$$

where $Y_{\text{obs}} = \{y_i, \boldsymbol{z}_i\}_{i=1}^n$. For the shape parameter α, we easily obtain

$$\frac{\partial \ell(\boldsymbol{\theta}|Y_{\text{obs}})}{\partial\alpha} = n\left[1 + \frac{1}{n}\log y_i - \frac{1}{n}\sum_{i=1}^n\left(\log\mu_i + \frac{y_i}{\mu_i}\right) + \log\alpha - \psi(\alpha)\right] = 0,$$

which can be solved by US algorithm in appendix A.1. And for the regression coefficients $\boldsymbol{\beta}$, by taking the first-order and second-order partial derivatives of the log-likelihood function, we obtain

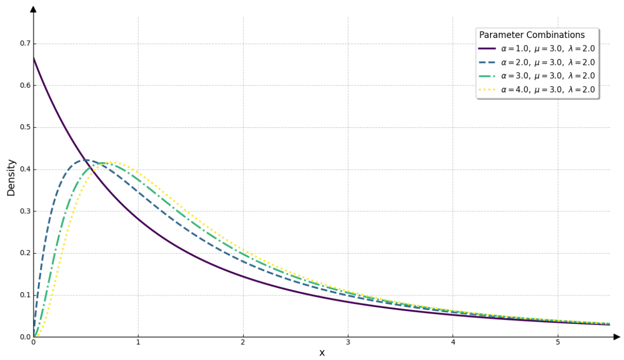
$$
\begin{aligned}
\nabla \ell(\boldsymbol{\beta}) &\triangleq \frac{\partial \ell(\boldsymbol{\theta}|Y_{\rm obs})}{\partial \boldsymbol{\beta}} = \alpha \sum_{i=1}^{n} \left(\frac{y_i - \mu_i}{\mu_i} \right) \boldsymbol{z}_{(i)} = \mathbf{Z}^\top \mathbf{W} \boldsymbol{u}, \\
\mathbf{I}(\boldsymbol{\beta}|Y_{\rm obs}) &\triangleq -\frac{\partial^2 \ell(\boldsymbol{\theta}|Y_{\rm obs})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} = \alpha \sum_{i=1}^{n} \frac{y_i \boldsymbol{z}_{(i)} \boldsymbol{z}_{(i)}^\top}{\mu_i} = \mathbf{Z}^\top \mathbf{W} \mathbf{Y} \mathbf{M}^{-1} \mathbf{Z},
\end{aligned}
$$

where $\boldsymbol{u} = (u_1, \dots, u_p)^\top$ with $u_i = (y_i - \mu_i)/\mu_i$, $\mathbf{Y} = \mathrm{diag}\{y_1, \dots, y_n\}$, $\mathbf{M} = \mathrm{diag}\{\mu_1, \dots, \mu_n\}$, $\mathbf{W} = \alpha \cdot \mathbf{I}_n$ is the weight matrix and $\mathbf{Z} = (\boldsymbol{z}_{(1)}, \dots, \boldsymbol{z}_{(n)})^\top$ is the covariate matrix. Therefore, we obtain the Fisher information matrix of $\boldsymbol{\beta}$ is $\mathbf{J}(\boldsymbol{\beta}) = E[\mathbf{I}(\boldsymbol{\beta}|Y_{\rm obs})] = \mathbf{Z}^\top \mathbf{W} \mathbf{Z}$. Based on the Fisher scoring algorithm, we have

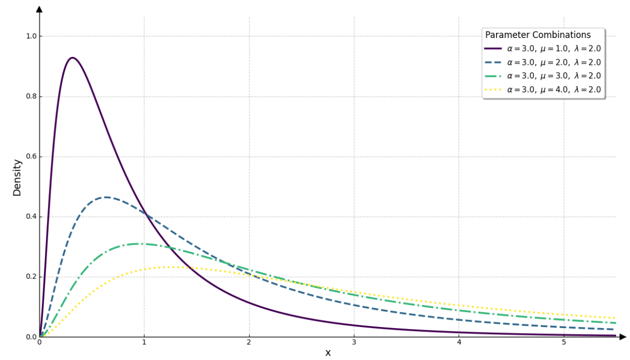
$$
\begin{aligned}
\boldsymbol{\beta}^{(t+1)} &= \boldsymbol{\beta}^{(t)} + \mathbf{J}^{-1}\left(\boldsymbol{\beta}^{(t)}\right) \nabla \ell \left(\boldsymbol{\beta}^{(t)}\right) = \boldsymbol{\beta}^{(t)} + \left(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}\right)^{-1} \mathbf{Z}^\top \mathbf{W} \boldsymbol{u}^{(t)} \\
&= \left(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}\right)^{-1} \mathbf{Z}^\top \mathbf{W} \left(\mathbf{Z} \boldsymbol{\beta}^{(t)} + \boldsymbol{u}^{(t)}\right) \\
&= \left(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}\right)^{-1} \mathbf{Z}^\top \mathbf{W} \boldsymbol{r}^{(t)}
\end{aligned}
$$

where $\boldsymbol{r} = \mathbf{Z} \boldsymbol{\beta}^{(t)} + \boldsymbol{u}^{(t)}$ is the response vector and $\mathbf{W}$ is a diagonal weight matrix, which is the form of *iterative reweighted least squares* (IRLS).

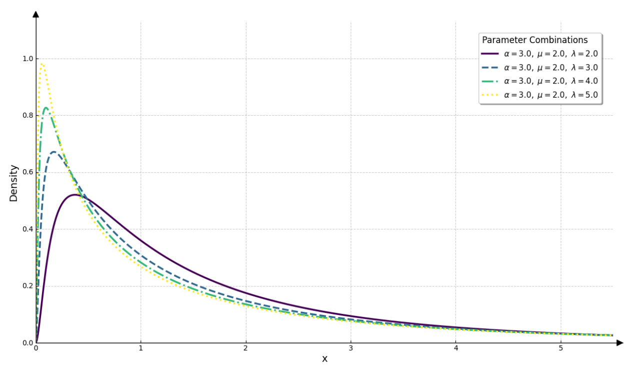
C. Figures and tables



(a) IGa-Ga distribution;



(b) IGau-Ga distribution;



(c) RIGau-Ga distribution.

Figure 1: The density functions of three specific cases for Ge-Ga distribution under various parameter combinations.

Table 1. Comparisons of A-MLE, MSE and CR for three distributions under various sample sizes

Distribution	Sample size	Parameter	A-MLE	MSE	CR
IGa-Ga	200	α	3.2193	0.9627	0.9376
		μ	2.0054	0.0200	0.9353
		$\mathrm{Var}(\tau)$	0.6141	1.0147	0.8895
	500	α	3.0720	0.2174	0.9476
		μ	2.0017	0.0077	0.9484
		$\mathrm{Var}(\tau)$	0.5289	0.0455	0.9200
	800	α	3.0436	0.1232	0.9495
		μ	2.0019	0.0047	0.9526
		$\mathrm{Var}(\tau)$	0.5173	0.0234	0.9327
IGau-Ga	200	α	2.2394	0.5591	0.9734
		μ	3.0062	0.0357	0.9493
		$\mathrm{Var}(\tau)$	0.2407	0.0214	0.9655
	500	α	2.0580	0.0857	0.9576
		μ	3.0019	0.0146	0.9464
		$\mathrm{Var}(\tau)$	0.2089	0.0076	0.9403
	800	α	2.0291	0.0458	0.9548
		μ	3.0005	0.0092	0.9465
		$\mathrm{Var}(\tau)$	0.2035	0.0048	0.9364
RIGau-Ga	200	α	1.1430	0.4432	0.9669
		μ	3.0082	0.0804	0.9426
		$\mathrm{Var}(\tau)$	0.4484	0.0701	0.9103
	500	α	1.0281	0.0455	0.9620
		μ	2.9970	0.0311	0.9493
		$\mathrm{Var}(\tau)$	0.4104	0.0242	0.9382
	800	α	1.0127	0.0098	0.9561
		μ	2.9995	0.0196	0.9499
		$\mathrm{Var}(\tau)$	0.4040	0.0139	0.9428

The convergence accuracy of MLE is 10^{-8}.

Table 2. Model comparisons based on $K = 1{,}000$ replications for Cases (B_1)–(B_4) with $\mathrm{Var}(\tau) = 0.5$

Mixed weight $\boldsymbol{p}$	Crit-erion	Fitted distribution					
		IGa-Ga	IGau-Ga	RIGau-Ga	Ga	IGau	Weibull
$(1,0,0,0,0,0)^\top$	AIC	3269.1	**3270.7**	3271.9	3324.3	3406.5	3344.2
	Rank	5.44	**5.20**	4.37	2.88	1.34	1.79
$(0,1,0,0,0,0)^\top$	AIC	3298.6	3296.9	**3297.2**	3347.7	3439.5	3358.1
	Rank	4.52	5.39	**5.09**	2.92	1.19	1.89
$(0,0,1,0,0,0)^\top$	AIC	3307.2	**3304.7**	3304.5	3349.6	3456.5	3357.2
	Rank	4.28	**5.29**	5.43	2.96	1.08	1.96
$\frac{1}{5}\times(3,1,1,0,0,0)^\top$	AIC	**3287.6**	3287.9	3288.7	3339.1	3428.6	3353.9
	Rank	**5.06**	5.29	4.65	2.92	1.21	1.87
$\frac{1}{5}\times(1,3,1,0,0,0)^\top$	AIC	3293.7	3292.5	**3293.0**	3342.9	3431.5	3354.5
	Rank	4.64	5.34	**5.01**	2.93	1.19	1.88
$\frac{1}{5}\times(1,1,3,0,0,0)^\top$	AIC	3299.9	3298.4	**3298.6**	3345.9	3443.8	3356.1
	Rank	4.52	5.35	**5.13**	2.95	1.12	1.93
$\frac{1}{6}\times\mathbf{1}_6$	AIC	**3339.3**	3339.3	3339.5	3358.8	4265.3	3360.3
	Rank	**4.98**	5.26	4.75	2.76	1.00	2.25
$\frac{1}{9}\times(1,1,1,2,2,2)^\top$	AIC	3348.0	3347.9	**3348.0**	3361.4	4452.1	3361.0
	Rank	4.83	5.20	**4.86**	2.51	1.00	2.61

The convergence accuracy of MLE is 10^{-8}.

Table 3. Model comparisons based on $K = 1{,}000$ replications for Case (B_1)–(B_4) with $\mathrm{Var}(\tau) = 1$

Mixed weight $\boldsymbol{p}$	Criterion	Fitted distribution					
		IGa-Ga	IGau-Ga	RIGau-Ga	Ga	IGau	Weibull
$(1,0,0,0,0,0)^{\top}$	AIC	3251.4	**3255.4**	3259.4	3349.7	3377.8	3356.6
	Rank	5.67	**5.18**	4.15	2.45	1.84	1.71
$(0,1,0,0,0,0)^{\top}$	AIC	3258.1	3253.2	**3256.4**	3352.3	3387.5	3337.9
	Rank	4.47	5.71	**4.82**	1.64	1.65	2.72
$(0,0,1,0,0,0)^{\top}$	AIC	3273.0	**3264.9**	3261.8	3338.3	3451.3	3317.2
	Rank	4.03	**5.17**	5.80	1.94	1.07	2.98
$\frac{1}{10} \times (8,1,1,0,0,0)^{\top}$	AIC	3263.6	**3265.9**	3269.3	3355.1	3413.1	3355.3
	Rank	5.49	**5.27**	4.24	2.37	1.50	2.13
$\frac{1}{10} \times (1,8,1,0,0,0)^{\top}$	AIC	3269.4	3265.1	**3267.8**	3360.2	3407.0	3346.3
	Rank	4.50	5.64	**4.86**	1.70	1.54	2.75
$\frac{1}{10} \times (1,1,8,0,0,0)^{\top}$	AIC	3280.5	**3274.4**	3273.1	3349.4	3459.1	3330.9
	Rank	4.12	**5.30**	5.59	1.94	1.09	2.98
$\frac{1}{6} \times \mathbf{1}_6$	AIC	**3290.7**	3290.8	3291.4	3319.7	—	3304.3
	Rank	**5.11**	5.33	4.53	2.00	1.00	3.03
$\frac{1}{3} \times (0,0,0,1,1,1)^{\top}$	AIC	3220.6	**3220.0**	3220.0	3231.2	—	3221.6
	Rank	4.04	**4.95**	4.98	2.15	1.00	3.88

The convergence accuracy of MLE is 10^{-8}.

Table 4. Fitting performance for individual medical expenditure (with sample mean 13.27 and standard error 12.11) of six distributions, respectively

Distribution	Mean	Standard Error	Log-likelihood	AIC	BIC
IGa-Ga	13.61	16.24	−4728.33	9462.65	9478.25
IGau-Ga	13.43	**14.93**	**−4712.64**	**9431.28**	**9446.88**
RIGau-Ga	**13.28**	12.50	−4700.91	9407.83	9423.43
Gamma	**13.27**	4.33	−4753.51	9511.03	9521.42
IGaussian	13.27	**14.82**	**−4713.32**	**9430.63**	**9441.03**
Weibull	13.34	3.03	−4769.08	9542.16	9552.56

The convergence accuracy of MLE is 10^{-8}.

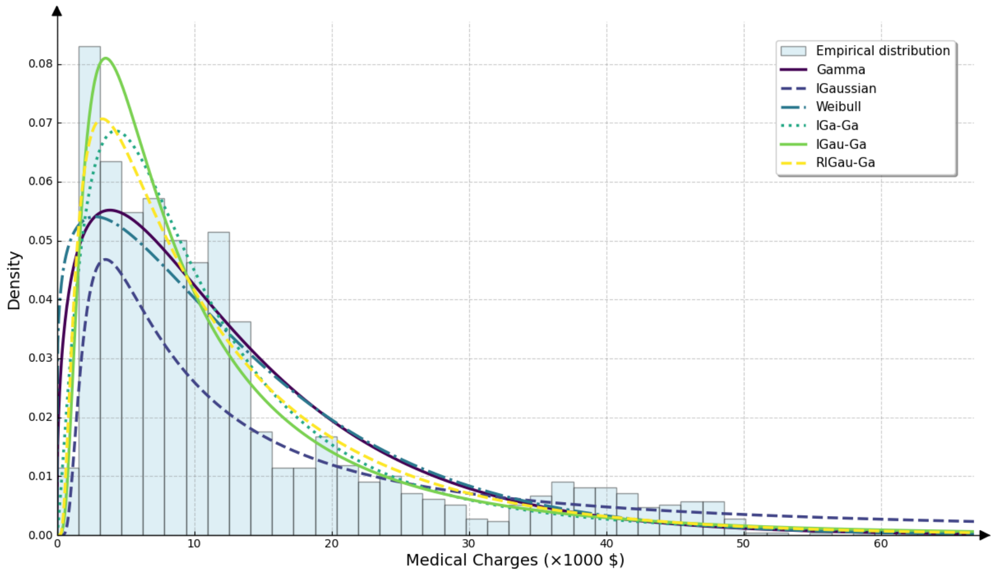


Figure 2: Comparison of fitting curves for medical expenditure with various distributions.

Table 5. Parameter estimations of IGau-Ga, RIGau-Ga and inverse Gaussian distributions for individual medical expenditure fitting

Distribution	Parameter	MLE	Asymptotic normal CI		Bootstrap CI	
			Lower	Upper	Lower	Upper
IGau-Ga	α	8.4449	−3.8280	20.7178	4.9203	22.4373
	μ	13.4324	12.7217	14.1431	12.6508	14.1897
	$\mathrm{Var}(\tau)$	0.9987	0.6006	1.3968	0.7881	1.1943
RIGau-Ga	α	43.7479	38.8045	48.6913	43.6630	43.8152
	μ	13.2789	12.6234	13.9344	12.6338	13.9634
	$\mathrm{Var}(\tau)$	0.8438	0.7998	0.8879	0.7840	0.8982
IGaussian	μ	13.2703	12.4765	14.0641	12.4867	14.0920
	λ	10.6469	9.8401	11.4537	9.9179	11.5105

The convergence accuracy of MLE is 10^{-8}.

Table 6. Fitting performance of five mean regression models for individual medical expenditure on training set

Model	Log-likelihood	AIC	BIC
IGa-Ga	-2727.019	5470.039	5509.301
IGau-Ga	**-2824.760**	**5665.520**	**7604.811**
RIGau-Ga	-2835.595	5687.189	5726.451
Ga	-2948.919	5911.838	5946.193
IGaussian	-3126.475	6266.950	6301.304

The convergence accuracy of MLE is 10^{-4}.

Table 7. Regression coefficient estimations of IGa-Ga, IGau-Ga, RIGau-Ga, gamma and inverse Gaussian regression models

Coefficient \ Model	IGa-Ga	IGau-Ga	RIGau-Ga	Ga	IGaussian
β_0	0.1356	0.2588	0.2434	0.4408	0.4004
β_1	0.0366	0.0334	0.0328	0.0273	0.0323
β_2	0.0108	0.0119	0.0125	0.0141	0.0080
β_3	0.1063	0.0998	0.0966	0.0817	0.1287
β_4	-0.0645	-0.0536	-0.0483	-0.0259	-0.0766
β_5	1.5087	1.5418	1.5438	1.4813	1.6994

The convergence accuracy of MLE is 10^{-4}.

Table 8. Fitting performance of five mean regression models for individual medical expenditure on test set

Model	Log-likelihood	AIC	BIC	MAE
IGa-Ga	-914.4527	1844.905	1875.490	4.8658
IGau-Ga	**-950.3276**	**1916.655**	**1947.240**	4.9382
RIGau-Ga	-955.0877	1926.176	1956.760	**4.6823**
Ga	-1029.6402	2073.280	2100.042	4.2275
IGaussian	-1110.9934	2235.987	2262.748	5.7197

Mean absolute error can be computed as $\mathsf{MAE} = \frac{1}{n'}\sum_{i=1}^{n'}|x_i - \hat{\mu}_i|$.

The convergence accuracy of MLE is 10^{-4}.