

基于污染 Gamma 分布的聚合风险模型 及其在风险分类中的应用*

卢志义

(天津财经大学统计学院, 天津 300222)

刘乐平

(天津财经大学统计学院, 天津 300222; 中国人民大学应用统计科学研究中心, 北京 100872)

孟生旺

(中国人民大学应用统计科学研究中心, 北京 100872)

摘要 分析了污染 Gamma 分布及其性质, 讨论了基于污染 Gamma 分布的聚合风险模型. 对模型的概率特性和参数估计进行了分析, 并对该模型在风险分类中的应用进行了讨论. 为克服索赔总量的分布函数在计算上的困难, 利用同单调性理论得到了随机凸序意义下索赔总量随机变量 S 的随机上界 S^c , 对 S^c 的分布函数及限额损失保费进行了讨论. 通过一个例子对所述结论的有效性进行验证.

关键词 污染 Gamma 分布, 聚合风险模型, 风险分类, 同单调性, 随机凸界.

MR(2000) 主题分类号 62P05

1 引言

Huber^[1] 在研究统计方法的稳健性时, 提出了污染分布. 假定随机变量 X 服从分布 $F(x)$, 且 $F(x) = (1 - \varepsilon)F_1(x) + \varepsilon F_2(x)$, $\varepsilon \in (0, 1)$, $F_1(x)$ 与 $F_2(x)$ 均为分布函数. 其背景是, 观察到的数据本应服从分布 $F_1(x)$, 但受到来自分布 $F_2(x)$ 的数据的影响. 或者是, 对每一观察到的数据, 有 $1 - \varepsilon$ 的可能来自 $F_1(x)$, 有 ε 的可能来自 $F_2(x)$. Huber^[1] 给出了一个被污染的正态分布的实例, 此例中, 随机变量 X 的分布函数为

$$F(x) = (1 - \varepsilon)\Phi\left(\frac{x - \mu}{\sigma}\right) + \varepsilon\Phi\left(\frac{x - \mu}{3\sigma}\right),$$

即考察的数据来自两个正态分布 $N(\mu, \sigma^2)$ 和 $N(\mu, 9\sigma^2)$, 它们的期望相同, 方差不同, 后者的方差大于前者.

* 天津市社科研究规划项目 (TJ05-TJ001), 教育部重大项目 (05JJD91015) 和中国人民大学应用统计科学研究中心重大项目 (05JJD910152) 资助.

收稿日期: 2006-12-15.

Landsman^[2] 讨论了以指数型分布为背景分布的污染分布. 设随机变量 X 的密度函数为

$$f_\varepsilon(x|\theta, \phi, \beta) = (1 - \varepsilon)\exp\left\{\frac{\theta x - b(\theta)}{\phi} + c(x, \phi)\right\} + \varepsilon\exp\left\{\frac{\theta x - b(\theta)}{\frac{\phi}{\beta}} + c(x, \phi)\right\}, \quad (1)$$

$\theta \in \Theta$ 为自然参数, $\phi > 0$ 为离散参数, $S \subset R$ 为支撑集. 此污染分布的特点是, $f_1(x)$ 与 $f_2(x)$ 为期望相同, 方差不同的指数型分布, $f_2(x)$ 的方差是 $f_1(x)$ 方差的 $\frac{1}{\beta}$ 倍, 称此模型为污染指数分布. 易知, 若 $\beta = 1$, X 的分布为标准的指数型分布, 因此, 污染指数分布是指数型分布的推广. Landsman^[2] 得到该污染分布具有如下性质: 取 $\beta \in (0, 1]$, 此污染分布的偏度和峰度可达到任意大, 且偏度和峰度的增大只取决于 ε 和 β 的值. 因而, 无论数据集中在分布的中心还是偏移到分布的尾部, 都可用该分布进行很好的拟合.

保险精算中常需计算若干个随机变量和的分布函数, 而和的分布函数的计算常常是非常复杂的. 同单调性 (Comonotonicity) 理论的出现, 为克服上述困难提供了很好的工具. Dhaene^[3] 和 Dhaene^[4] 对该理论及其应用进行了系统的介绍. 本文讨论污染指数分布下保险理赔的聚合风险模型. 文中以 Gamma 分布作为污染指数分布的背景分布, 对应的分布称作污染 Gamma 分布. 文中分析了基于污染 Gamma 分布的复合 Poisson 分布及其性质, 对参数的估计进行了分析, 并对此模型在风险分类中的应用进行了讨论. 为克服索赔总量的分布函数在计算上的困难, 利用同单调性理论得到了随机凸序意义下索赔总量随机变量 S 的随机上界 S^c , 并对 S^c 的分布函数及限额损失保费进行了讨论.

2 污染 Gamma 分布及其性质

Gamma 分布密度函数的指数型标准形式为

$$\exp\{\nu[x\theta - \kappa(\theta)] + c(x, \nu)\}, \quad (2)$$

其中

$$\theta = -\frac{1}{\mu}, \quad \kappa(\theta) = -\log(-\theta), \quad c(x, \nu) = \nu \log(\nu x) - \log x - \log(\Gamma(\nu)),$$

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx, \quad a > 0.$$

Gamma 分布的方差函数为 $V(\mu) = \mu^2$.

设随机变量 Y 服从污染 Gamma 分布, 由上述, Y 的密度函数为

$$f_\varepsilon(x) = (1 - \varepsilon)\exp\{\nu[y\theta + \log(-\theta)] + c(y, \nu)\} + \varepsilon\exp\{\beta\nu[y\theta + \log(-\theta)] + c(y, \beta\nu)\}, \quad (3)$$

其中 $\theta, c(y, \nu)$ 含义同上, $c(y, \beta\nu) = \beta\nu \log(\beta\nu y) - \log y - \log(\Gamma(\beta\nu))$. 记此分布为 CGD(μ, ν, y) (Contaminated Gamma Distribution). Y 的矩母函数为

$$M_Y(t) = (1 - \varepsilon)\left(1 - \frac{t}{\mu}\right)^{-\nu} + \varepsilon\left(1 - \frac{t}{\frac{\beta\nu}{\mu}}\right)^{-\beta\nu}, \quad t < \min\left\{\frac{\nu}{\mu}, \frac{\beta\nu}{\mu}\right\}. \quad (4)$$

由 Landsman^[2] 的结论, 污染 Gamma 分布的偏度为

$$\gamma_1(\nu, \varepsilon, \beta) = \frac{2}{\sqrt{\nu}} \frac{1 - \varepsilon + \frac{\varepsilon}{\beta^2}}{(1 - \varepsilon + \frac{\varepsilon}{\beta})^{\frac{3}{2}}} = \gamma_1(\nu)\varphi(\beta), \quad (5)$$

其中, $\gamma_1(\nu) = \frac{2}{\sqrt{\nu}}$ 为 Gamma 分布 (2) 的偏度, $\varphi(\beta) = \frac{1-\varepsilon+\frac{\varepsilon}{\beta^2}}{(1-\varepsilon+\frac{\varepsilon}{\beta})^{\frac{3}{2}}}$, $\varphi(\beta)$ 在区间 $(0, 1]$ 是单调减函数, 且 $\varphi(0) = \infty$, $\varphi(1) = 1$. 若取定 ε 的值, 污染 Gamma 分布的偏度随 β 的不同取值而变化, 且可以达到任意大.

污染 Gamma 分布的峰度为

$$\gamma_2(\nu, \varepsilon, \beta) = \left(\frac{6}{\nu} + 3\right) \frac{1 - \varepsilon + \frac{\varepsilon}{\beta^2}}{(1 - \varepsilon + \frac{\varepsilon}{\beta})^2} = \gamma_2(\nu)\psi(\beta), \quad (6)$$

其中, $\gamma_2(\nu) = \frac{6}{\nu} + 3$ 为 Gamma 分布 (2) 的峰度. $\psi(\beta)$ 在区间 $(0, 1]$ 是单调减函数, 且 $\psi(0) = \infty$, $\psi(1) = 1$. 若取定 ε 的值, 污染 Gamma 分布的偏度随 β 的不同取值而变化, 且可以达到任意大.

3 聚合风险模型与风险分类

设 Y_i 为保险公司某类保单的第 i 次赔付额随机变量, N 是单位时间内 (如一个会计年或保单年) 所有该类保单的发生理赔总次数的随机变量, S 为该时间段内赔付总额随机变量, 则 $S = \sum_{i=1}^n Y_i$. 且设

- 1) 随机变量 $N, Y_i (i = 1, 2, \dots, n)$ 相互独立;
- 2) $Y_i \stackrel{i.i.d.}{\sim} \text{CGD}(\mu, \nu, \beta), (i = 1, 2, \dots, n)$;
- 3) $N \sim \text{Poi}(\lambda\omega)$, Poi 表示 Poisson 分布, ω 为风险暴露单位.

则随机变量 S 的矩母函数为

$$M_S(t) = \exp\left\{(1-\varepsilon)\omega\lambda\left(\left(1-\frac{t}{\frac{\nu}{\mu}}\right)^{-\nu}-1\right)\right\} \cdot \exp\left\{\varepsilon\omega\lambda\left(\left(1-\frac{t}{\frac{\beta\nu}{\mu}}\right)^{-\beta\nu}-1\right)\right\}, \quad (7)$$

其中, $t < \min\{\frac{\nu}{\mu}, \frac{\beta\nu}{\mu}\}$. 由矩母函数的形式可知, S 可以看作两个独立随机变量 S_1 与 S_2 的和, 即 $S \stackrel{d}{=} S_1 + S_2$. 其中

1) S_1 服从理赔次数为 N_1 , 每次理赔额大小为 Y_{1i} 的复合 Poisson 分布, 且 $N_1 \sim \text{Poi}((1-\varepsilon)\omega\lambda)$, $Y_{1i} \sim G(\nu, \frac{\nu}{\mu})$, G 表示 Gamma 分布, S_1 的分布密度为

$$f_{S_1}(s) = \left\{ \sum_r \frac{\left(\frac{\nu}{\mu}\right)^{r\nu} ((1-\varepsilon)\omega\lambda)^r s^{r\nu-1}}{\Gamma(r\nu)r!} \right\} \exp\left\{-\frac{\nu}{\mu}s - (1-\varepsilon)\omega\lambda\right\}. \quad (8)$$

经参数变换, 可变形为

$$f_{S_1}(s) = \left\{ \sum_r \frac{s^{r\nu-1}}{((p-1)^\nu(2-p)\phi^{\nu+1})^r \Gamma(r\nu)r!} \right\} \exp\left\{\frac{\tau^{1-p}}{1-p}s - \frac{\tau^{2-p}}{2-p}\right\}, \quad (9)$$

其中

$$p = \frac{\nu+2}{\nu+1} \in (1, 2), \quad \tau = (1-\varepsilon)\omega\lambda\mu, \quad \phi = \frac{(1-\varepsilon)^{1-p}\omega^{1-p}\lambda^{1-p}\mu^{2-p}}{2-p}.$$

进一步，记

$$\theta = \frac{\tau^{1-p}}{1-p}, \quad c(s; \phi, p) = \sum_r \frac{s^{r\nu-1}}{((p-1)^\nu(2-p)\phi^{\nu+1})^r \Gamma(r\nu)r!},$$

$$\kappa_p(\theta) = \frac{1}{2-p} ((1-p)\theta)^{\frac{2-p}{1-p}},$$

则有

$$f_{S_1}(s) = c(s; \phi, p) \exp\left\{\frac{s\theta - \kappa_p(\theta)}{\phi}\right\}, \tag{10}$$

因而， S_1 的分布属于 Tweedie 类分布.

2) S_2 服从理赔次数为 N_2 , 每次理赔额大小为 Y_{2i} 的复合 Poisson 分布, 且 $N_2 \sim \text{Poi}(\varepsilon\omega\lambda)$, $Y_{2i} \sim G(\beta\nu, \frac{\beta\nu}{\mu})$, S_2 的分布密度为

$$f_{S_2}(s) = \left\{ \sum_r \frac{(\frac{\beta\nu}{\mu})^{r\nu} (\varepsilon\omega\lambda)^r s^{r\beta\nu-1}}{\Gamma(r\beta\nu)r!} \right\} \exp\left\{-\frac{\beta\nu}{\mu}s - \varepsilon\omega\lambda\right\}, \tag{11}$$

类似于 1), 经参数变换, 上式可变形为

$$f_{S_2}(s) = \left\{ \sum_r \frac{s^{r\beta\nu-1}}{((p-1)^{\beta\nu}(2-p)\phi^{\beta\nu+1})^r \Gamma(r\beta\nu)r!} \right\} \exp\left\{\frac{\tau^{1-p}}{1-p}s - \frac{\tau^{2-p}}{2-p}\right\}, \tag{12}$$

其中, $p = \frac{\beta\nu+2}{\beta\nu+1} \in (1, 2)$, $\tau = \varepsilon\omega\lambda\mu$, $\phi = \frac{\varepsilon^{1-p}\omega^{1-p}\lambda^{1-p}\mu^{2-p}}{2-p}$. 显然, S_2 的分布也属于 Tweedie 类分布.

于是, 可得到赔付总量 S 的密度函数为 $f_S = f_{S_1} * f_{S_2}$, 符号 “*” 表示卷积.

由以上分析知, 若设每次赔付额服从污染 Gamma 分布, 每一保单的索赔次数服从参数为 λ 的 Poisson 分布, 则单位时间内总数为 ω 的保单的索赔总额可看作由两部分组成, 一部分为, 保单的每次索赔额服从 $G(\nu, \frac{\nu}{\mu})$ 分布, 这部分保单的数目为 $(1-\varepsilon)\omega$; 另一部分数目为 $\varepsilon\omega$ 的保单, 其每次索赔额服从 $G(\beta\nu, \frac{\beta\nu}{\mu})$ 分布. 两类保单的每次赔付额服从期望相同, 方差不同的 Gamma 分布, 后者的方差是前者的 $\frac{1}{\beta}$ 倍.

上述方法可作进一步推广, 设赔付总额服从分布密度为

$$f_S(s) = \varepsilon_1 \exp\{\beta_1\nu(s\theta + \log(-\theta)) + c(s, \beta_1, \nu)\} + \dots + \varepsilon_n \exp\{\beta_n\nu(s\theta + \log(-\theta)) + c(s, \beta_n, \nu)\} \tag{13}$$

的复合 Poisson 分布, 这里 $\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n = 1$, 类似于上面的分析, 我们可将此保单组合分为 n 类, 第 i 类保单的数目为 $\varepsilon_i\omega$, 各类保单的每次索赔额服从期望相同, 方差不同的 Gamma 分布.

基于污染 Gamma 分布的复合 Poisson 模型在风险管理及保险精算中具有重要的应用价值. 若设给定赔付总量数据服从每次赔付额为污染 Gamma 分布的复合 Poisson 分布, 用此分布对损失数据进行拟合, 并根据参数的估计值将该组保单进行分类, 每一保单索赔次数服从相同的分布, 索赔额服从期望相同, 方差不同的 Gamma 分布. 通过分类, 为保险人提供详细的承保风险信息, 为保单费率的厘定提供依据. 反之, 若某保单集合存在异质性, 可采

用每次赔付额分布为污染 Gamma 分布 (或其它分布) 的复合 Poisson 分布进行拟合, 即可将相关的异质风险合并为一类, 为承保人提供合并后保单组信息.

4 参数的估计

赔付总量 S 的分布密度函数的计算非常复杂, 因而考虑用矩估计法求参数的估计值. 由随机变量 S 的矩母函数, 可得其期望和各阶中心矩 (用 v_k 表示)

$$\begin{aligned} E[S] &= \omega\lambda\mu, \\ v_2 &= \omega\lambda\mu^2 \left((1-\varepsilon)\frac{\nu+1}{\nu} + \varepsilon\frac{\beta\nu+1}{\beta\nu} \right), \\ v_3 &= \omega\lambda\mu^3 \left((1-\varepsilon)\frac{(\nu+1)(\nu+2)}{\nu^2} + \varepsilon\frac{(\beta\nu+1)(\beta\nu+2)}{(\beta\nu)^2} \right), \\ v_4 &= \omega\lambda\mu^4 \left((1-\varepsilon)\frac{(\nu+1)(\nu+2)(\nu+3)}{\nu^3} + \varepsilon\frac{(\beta\nu+1)(\beta\nu+2)(\beta\nu+3)}{(\beta\nu)^3} \right) + 3v_2^2. \end{aligned}$$

设 λ 已知, 参数的矩估计可由以下方程组得到

$$\begin{cases} \omega\lambda\mu = \bar{S}, \\ \omega\lambda\mu^2\psi_1(\varepsilon, \nu, \beta) = \tilde{v}_2, \\ \omega\lambda\mu^3\psi_2(\varepsilon, \nu, \beta) = \tilde{v}_3, \\ \omega\lambda\mu^4\psi_3(\varepsilon, \nu, \beta) = \tilde{v}_4 - 3\tilde{v}_2^2, \end{cases} \quad (14)$$

其中

$$\begin{aligned} \psi_1(\varepsilon, \nu, \beta) &= (1-\varepsilon)\frac{\nu+1}{\nu} + \varepsilon\frac{\beta\nu+1}{\beta\nu}, \\ \psi_2(\varepsilon, \nu, \beta) &= (1-\varepsilon)\frac{(\nu+1)(\nu+2)}{\nu^2} + \varepsilon\frac{(\beta\nu+1)(\beta\nu+2)}{(\beta\nu)^2}, \\ \psi_3(\varepsilon, \nu, \beta) &= (1-\varepsilon)\frac{(\nu+1)(\nu+2)(\nu+3)}{\nu^3} + \varepsilon\frac{(\beta\nu+1)(\beta\nu+2)(\beta\nu+3)}{(\beta\nu)^3}, \end{aligned}$$

\bar{S} 为样本均值, \tilde{v}_k 表示样本中心矩, λ 可利用赔付次数数据进行估计, 也可由经验或历史数据得到.

5 赔付总额 S 的随机凸界

求索赔总额 S 分布函数的计算涉及到复杂的卷积运算. 在 (13) 中, 随着 n 的增加, 积分的重数不断增加, 因而 S 的分布函数是很难用解析的方法得到的. 一种良好的途径是, 借助同单调性理论的相关方法得到 S 的随机凸界, 作为 S 分布的一种近似, 通过对随机凸界概率特性的讨论, 得到有关索赔总额的信息. 下面就最简单的情形 ((13) 中 $n=2$) 为例说明该方法的实现, $n>2$ 的情形可类似得到.

设随机变量 X 和 Y 的分布函数分别为 $F_X(\cdot)$ 和 $F_Y(\cdot)$. 用 $R(G, H)$ 表示边际分布为 G 和 H 的二维随机向量的集合.

定义 1 对概率空间 (Ω, \mathcal{F}, P) 上的两随机变量 X 和 Y , 若有

$$[X(\omega_1) - X(\omega_2)][Y(\omega_1) - Y(\omega_2)] \geq 0, \quad \text{a.s.} \quad (15)$$

则称 X 与 Y 同单调.

定义 2 称随机变量 X 在凸序意义下优于 Y . 若对任意使得 $E[u(X)]$ 与 $E[u(Y)]$ 存在的实的凸函数 u , 有 $E[u(X)] \leq E[u(Y)]$. 记作 $X \leq_{cx} Y$. 也称 Y 是 X 的随机上界.

对随机凸序, 有如下结论

1) 若 $X \leq_{cx} Y$, 则 $E[X] = E[Y]$, 且 $\text{Var}[X] \leq \text{Var}[Y]$; 若 $X \leq_{cx} Y$ 且 $\text{Var}[X] = \text{Var}[Y]$, 则 X 与 Y 必有相同的分布.

2) 若 $X \leq_{cx} Y$, 则 $E[(X - d)_+] \leq E[(Y - d)_+]$.

定义随机变量 X 的三种形式的逆分布函数

$$F_X^{-1}(p) = \inf\{x \in R | F_X(x) \geq p\}, \quad F_X^{-1+}(p) = \sup\{x \in R | F_X(x) \leq p\},$$

$$F_X^{-1(\alpha)}(p) = \alpha F_X^{-1}(p) + (1 - \alpha) F_X^{-1+}(p), \quad p \in [0, 1].$$

下面给出两个引理, 其证明参见文 [3].

引理 1 设随机变量 X_1 与 X_2 的分布函数分别为 $F_{X_1}(\cdot)$ 和 $F_{X_2}(\cdot)$, $S = X_1 + X_2$, U 服从 $(0, 1)$ 上的均匀分布, 则有

$$S \leq_{cx} S^c \stackrel{\text{def}}{=} F_{X_1}^{-1}(U) + F_{X_2}^{-1}(U). \quad (16)$$

事实上, 随机变量 X 与 $F_{X_1}^{-1}(U)$, X_2 与 $F_{X_2}^{-1}(U)$ 的分布分别相同, 因而 S 与 S^c 的期望相同, 但由于 $F_{X_1}^{-1}(U)$ 与 $F_{X_2}^{-1}(U)$ 具有很强的相依关系 (同单调性), 从而使 S^c 的方差较大.

汪荣明 [5] 证明了, 如果随机变量 X 和 Y 的相关系数达到最大, 则 X 和 Y 同单调. 从效用论的角度, $S \leq_{cx} S^c$ 意味着对所有的风险回避者来讲更偏好风险 S 而讨厌 S^c , 若用风险 S^c 代替 S , 决策会更谨慎, 这是风险管理中的一个重要原则.

引理 2 设随机向量 X_1 与 X_2 的分布函数分别为 $F_{X_1}(\cdot)$ 和 $F_{X_2}(\cdot)$, U 服从 $(0, 1)$ 上的均匀分布, $S^c \stackrel{\text{def}}{=} F_{X_1}^{-1}(U) + F_{X_2}^{-1}(U)$, 则有

$$F_{S^c}^{-1}(p) = F_{X_1}^{-1}(p) + F_{X_2}^{-1}(p). \quad (17)$$

由引理 2, 可得到 S^c 的分布函数

$$\begin{aligned} F_{S^c}(s) &= \sup\{p \in (0, 1) | F_{S^c} \geq p\} \\ &= \sup\{p \in (0, 1) | F_{S^c}^{-1}(p) \leq s\} \\ &= \sup\{p \in (0, 1) | F_{X_1}^{-1}(p) + F_{X_2}^{-1}(p) \leq s\}. \end{aligned} \quad (18)$$

定理 设随机向量 X_1 与 X_2 的分布函数为 $F_{X_1}(\cdot)$ 和 $F_{X_2}(\cdot)$, 分布密度函数为 $f_{X_1}(\cdot)$ 和 $f_{X_2}(\cdot)$, U 服从 $(0, 1)$ 上的均匀分布, $S = X_1 + X_2$, $S^c \stackrel{\text{def}}{=} F_{X_1}^{-1}(U) + F_{X_2}^{-1}(U)$, 则对 $d \in (F_{S^c}^{-1+}(0), F_{S^c}^{-1}(1))$ 有

$$E[(S^c - d)_+] = E[S] - d - \int_0^{d_1} (x - d_1) f_{X_1}(x) dx - \int_0^{d_2} (x - d_2) f_{X_2}(x) dx, \quad (19)$$

其中 $d_i = F_{X_i}^{-1(\alpha_d)}(F_{S^c}(d))$ ($i = 1, 2$), 且 $d_1 + d_2 = d$, $\alpha_d \in [0, 1]$ 由下式确定

$$F_{S^c}^{-1(\alpha_d)}(F_{S^c}(d)) = d.$$

证 由文 [3] 定理 7, 有

$$\begin{aligned} & E[(S^c - d)_+] \\ &= E[(X_1 - d_1)_+] + E[(X_2 - d_2)_+] \\ &= \int_{d_1}^{\infty} (x - d_1)f_{X_1}(x)dx + \int_{d_2}^{\infty} (x - d_2)f_{X_2}(x)dx \\ &= \int_0^{\infty} (x - d_1)f_{X_1}(x)dx - \int_0^{d_1} (x - d_1)f_{X_1}(x)dx + \int_0^{\infty} (x - d_2)f_{X_2}(x)dx \\ &\quad - \int_0^{d_2} (x - d_2)f_{X_2}(x)dx \\ &= E[X_1] - d_1 + E[X_2] - d_2 - \int_0^{d_1} (x - d_1)f_{X_1}(x)dx - \int_0^{d_2} (x - d_2)f_{X_2}(x)dx \\ &= E[S] - d - \int_0^{d_1} (x - d_1)f_{X_1}(x)dx - \int_0^{d_2} (x - d_2)f_{X_2}(x)dx, \end{aligned}$$

其中, $d_i = F_{X_i}^{-1(\alpha_d)}(F_{S^c}(d))$ ($i = 1, 2$), 由文 [3] 定理 6, 有

$$d_1 + d_2 = F_{X_1}^{-1(\alpha_d)}(F_{S^c}(d)) + F_{X_2}^{-1(\alpha_d)}(F_{S^c}(d)) = F_{S^c}^{-1(\alpha_d)}(F_{S^c}(d)) = d.$$

由引理 1 可知, 对赔付总额随机变量 S , 我们可以找到它的一个随机上界

$$S^c \stackrel{\text{def}}{=} F_{S_1}^{-1}(U) + F_{S_2}^{-1}(U).$$

将随机界作为 S 的一个近似, 并通过随机上界变量概率特性的分析, 得到关于赔付总额 S 的信息, 这既解决了上文遇到的求赔付总额分布函数的困难, 也符合保险精算的谨慎性原则. 由引理 2, S^c 的逆分布函数 (即分位数) 是很容易求得的, 而随机变量分布的分位数正是风险管理和保险精算人员最关心的指标. 另外, 对索赔总量随机变量, 我们更多的是关心其右尾的性质, 如限额损失保费 (Stop-loss premiums 简称 SLP) 的估计, 本文定理给出了计算索赔总额上界的限额损失保费的一个简单而非常有用的结论. 关于上述方法的具体应用, 将在下例中讨论.

6 数值实验

设某类保单索赔总量数据服从基于污染 Gamma 分布的复合 Poisson 分布, 保单数为 50, 精算师和理赔人员根据历史数据和经验判断确定每一保单的索赔次数服从均值为 0.2 的 Poisson 分布, 观察值的样本均值和各阶样本中心矩分别为: $\bar{S} = 20$, $\tilde{v}_2 = 52.8$, $\tilde{v}_3 = 176.64$, $\tilde{v}_4 = 11889.408$. 由矩估计方程组 (14) 解得 $\mu = 2$, $\varepsilon = 0.4$, $\nu = 5$, $\beta = 0.4$. 由以上计算结果可以得出, 该组保单可分为两类, 两类保单的每次索赔额服从期望都为 2 方差分别为 0.8 和 2 的 Gamma 分布.

下面讨论索赔总量 S 的随机上界. 显然, $F_{S_1}(\cdot)$ 和 $F_{S_2}(\cdot)$ 都为连续函数, 因而 $F_{S_1}^{-1}(\cdot)$ 和 $F_{S_2}^{-1}(\cdot)$ 都为严格递增函数. 由引理 2 知

$$F_{S^c}^{-1}(p) = F_{S_1}^{-1}(p) + F_{S_2}^{-1}(p),$$

但 $F_{S_1}^{-1}(\cdot)$ 和 $F_{S_2}^{-1}(\cdot)$ 的显式不易求得, 我们可采用线性插值法得到其近似值. 以 $F_{S_1}^{-1}(\cdot)$ 为例, 将 S_1 的支撑集分割为区间 $(0, \delta], (\delta, 2\delta] \cdots$ 的并, 若 $p \in (F_{S_1}((k-1)\delta), F_{S_1}(k\delta)]$, 则用线性插值法可得到

$$F_{S_1}^{-1}(p) \approx k\delta + \frac{\delta p - \delta F_{S_1}(k\delta)}{F_{S_1}(k\delta) - F_{S_1}((k-1)\delta)}, \quad p \in (0, 1). \quad (20)$$

表 1 F_{S^c} 的分位数

p	$F_{S^c}^{-1}(p)$	p	$F_{S^c}^{-1}(p)$	p	$F_{S^c}^{-1}(p)$	p	$F_{S^c}^{-1}(p)$	p	$F_{S^c}^{-1}(p)$
0.02	3.7125	0.22	11.9262	0.42	17.0810	0.62	22.3657	0.82	29.6314
0.04	5.1862	0.24	12.4784	0.44	17.5819	0.64	22.9557	0.84	30.6576
0.06	6.3005	0.26	13.0169	0.46	18.0855	0.66	23.5647	0.86	31.7979
0.08	7.2363	0.28	13.5442	0.48	18.5928	0.68	24.1957	0.88	33.0901
0.10	8.0604	0.30	14.0625	0.50	19.1054	0.70	24.8521	0.90	34.5943
0.12	8.8079	0.32	14.5739	0.52	19.6242	0.72	25.5379	0.92	36.4172
0.14	9.4999	0.34	15.0800	0.54	20.1509	0.74	26.2581	0.94	38.7776
0.16	10.1497	0.36	15.5825	0.56	20.6866	0.76	27.0185	0.96	42.2767
0.18	10.7664	0.38	16.0825	0.58	21.2333	0.78	27.8272	0.98	51.9069
0.20	11.3568	0.40	16.5817	0.60	21.7923	0.80	28.6938		

显然

$$\lim_{\delta \rightarrow 0} k\delta + \frac{\delta p - \delta F_{S_1}(k\delta)}{F_{S_1}(k\delta) - F_{S_1}((k-1)\delta)} = F_{S_1}^{-1}(p),$$

另外

$$F_{S_1}^{-1}(0) = 0, \quad F_{S_1}^{-1}(\infty) = 1.$$

表 1 给出了 $p = 0.02i$ ($i = 1, 2, \dots, 49$) 时 $F_{S^c}^{-1}(p)$ 的近似值. 在坐标系中将上面得到的各点描出并用光滑曲线连结可得到随机变量的分布函数的近似图 (图 1).

由本文定理知

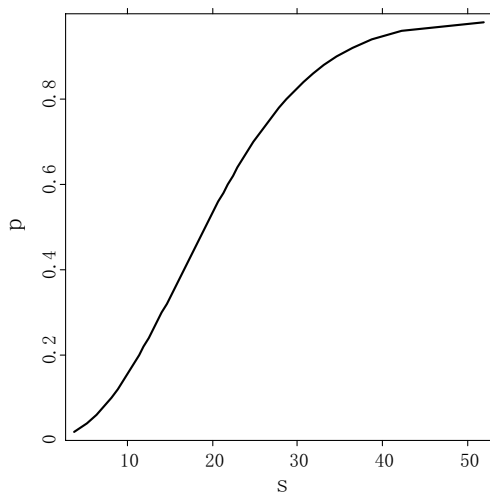
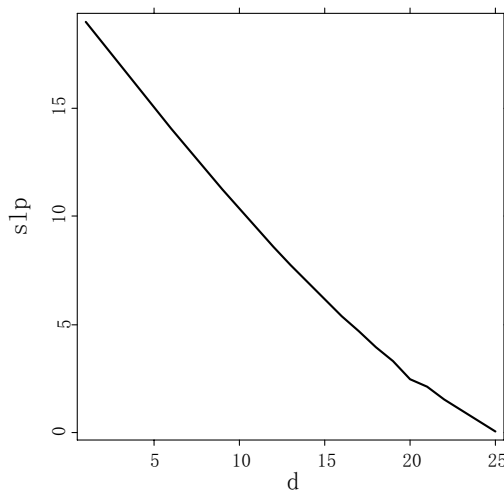
$$E[(S^c - d)_+] = E[S] - d - \int_0^{d_1} (s - d_1)f_{S_1}(s)ds - \int_0^{d_2} (s - d_2)f_{S_2}(s)ds,$$

其中 $d_i = F_{S_i}^{-1}(F_{S^c}(d))$ ($i = 1, 2$), d_1 和 d_2 可仿上述插值法得到. 表 2 给出了 d 取 1-25 的整数时限额损失保费 (SLP) $E[(S^c - d)_+]$ 的近似值. 将各点在坐标系中描出并用光滑曲线连结得到图 2. 事实上, $E[(S^c - d)_+]$ 的一阶导数为 $F_{S^c}(d) - 1 \leq 0$, 二阶导数为 $f_{S^c}(d) > 0$ 因而 $E[(S^c - d)_+]$ 是下凸的严格递减函数, 且有

$$\lim_{d \rightarrow \infty} E[(S^c - d)_+] = 0.$$

表 2 S^c 的限额损失保费

d	$E[(S^c - d)_+]$	d	$E[(S^c - d)_+]$	d	$E[(S^c - d)_+]$	d	$E[(S^c - d)_+]$	d	$E[(S^c - d)_+]$
1	19.0002	6	14.0708	11	9.4494	16	5.3836	21	2.0929
2	18.0022	7	13.1137	12	8.5806	17	4.6586	22	1.5321
3	17.0121	8	12.1699	13	7.7426	18	3.9643	23	1.0099
4	16.0204	9	11.2445	14	6.9297	19	3.3146	24	0.5258
5	15.0414	10	10.3357	15	6.1406	20	2.4384	25	0.0414

图 1 S^c 分布函数的近似图形图 2 $E[(S^c - d)_+]$ 的近似图形

参 考 文 献

- [1] Huber H J. Robust Statistics. New York: Wiley, 1981.
- [2] Landsman Z M and Makov U E. Contaminated exponential dispersion loss models. *North American Actuarial Journal*, 2003, **7**(2): 116–127.
- [3] Dhaene J, Denuit M, Goovaerts M J, Kaas R and Vyncke D. The concept of comonotonicity in actuarial science and finance: Theory. *Insurance: Mathematics and Economics*, 2002, **31**(1): 3–33.
- [4] Dhaene J, Denuit M, Goovaerts M J, Kaas R and Vyncke D. The concept of comonotonicity in actuarial science and finance: Applications. *Insurance: Mathematics and Economics*, 2002, **31**(2): 133–161.
- [5] 汪荣明, 杨亚松. 同单调相依结构下两重生命模型的概率分布. *应用数学学报*, 2006, **29**(1): 131–138.
- [6] Kaas R, Dhaene J, Goovaerts M J. Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, 2000, **27**: 151–168.
- [7] Wüthrich M V. Claims reserving using Tweedie's compound Poisson model. *Astin Bulletin*, 2003, **33**(2): 331–346.

COLLECTIVE RISK MODEL BASED ON THE CONTAMINATED
GAMMA DISTRIBUTION AND ITS APPLICATION
IN RISK CLASSIFICATION

LU Zhiyi

(School of Statistics, Tianjin University of Finance and Economics, Tianjin 300222)

LIU Leping

*(School of Statistics, Tianjin University of Finance and Economics, Tianjin 300222;
The Reserch Center for Applied Statistics of Renmin University of China, Beijing 100872)*

MENG Shengwang

(The Reserch Center for Applied Statistics of Renmin University of China, Beijing 100872)

Abstract The contaminated Gamma distribution and its properties are analyzed. The collective risk model based on contaminated Gamma distribution is put forward and its probability character is considered. Then the application of the model in risk classification is discussed. To overcome the difficulty in calculating the distribution function of claim amount S , a stochastic upper bound S^c of S in the sense of stochastic convex order is obtained by using the comonotonicity theory. The distribution function of S^c and stop-loss premium are discussed. A numerical examples is given to illustrate the validity of the proposed method.

Key words Contaminated Gamma distribution, collective risk model, risk classification, comonotonicity, stochastic convex bounds.