应用概率统计　第二十四卷
第一期　2008年2月

Chinese Journal of Applied Probability
and Statistics Vol.24 No.1 Feb. 2008

指数类混合型索赔次数的分布及其应用 *

毛泽春[1,2]　　刘锦萼[2]

([1]湖北大学商学院金融系, 武汉, 430062; [2]山东经济学院概率统计与保险精算研究所, 济南, 250014)

摘　要

基于保险索赔的实际应用背景, 本文引出了一类指数类混合型索赔次数的分布并研究了其散度(dispersion)性质, 这类分布由复合Poisson-Geometric分布与指数类分布混合而得到. 本文给出了拟合这类分布的矩估计方法及我国汽车保险索赔数据的应用实例, 并给出了相应的检验结果.

关键词: 索赔次数, 复合Poisson-Geometric分布, 混合PG分布.

学科分类号: O212.1, F840.6.

《应用概率统计》版权所有

§1. 引　言

1.1 背景

Poisson分布常用来刻画个体保单的索赔次数, 对于集合保单, 若假定集合中各保单是同质的而且其索赔次数相互独立, 仍可以用Poisson分布来描述集合保单的索赔次数. 但是, 当集合中各保单存在不同质性时, 用Poisson分布来描述就不合适了. Poisson分布的一个重要性质是方差等于均值, 但是实际上索赔次数并不完全遵循Poisson分布规律, 方差往往偏离均值, 当方差大于均值时, 这种现象相对于Poisson分布来说称为散度偏大(over-dispersion)[3, 4]. 在保险中引起散度偏大现象的原因很复杂, 其中一个原因就是集合中各保单存在着差异性, 这时通常可以采用混合Poisson分布来刻画这种差异性.

设在集合中随机抽取一份保单, 其索赔次数为N, 用随机变量Λ来表示个体之间的差异性, 其分布函数为$U(\lambda)$, 并设

$$N|\Lambda=\lambda\sim \mathrm{Po}(\lambda).$$

此时N的分布为

$$\mathsf{P}(N=k)=\int_0^{+\infty}\frac{\lambda^k}{k!}e^{-\lambda}\mathrm{d}U(\lambda). \tag{1.1}$$

称(1.1)式所确定的分布为混合Poisson分布. 例如, 当$U(\lambda)$为Gamma分布时, 得到的混合Poisson分布为负二项分布. 对混合Poisson分布, 可以计算

$$\mathsf{E}N=\mathsf{E}\Lambda,\qquad \mathsf{E}(N(N-1))=\mathsf{E}\Lambda^2,\qquad \mathsf{Var}\,N=\mathsf{E}N+\mathsf{Var}\,\Lambda.$$

*国家自然科学基金资助项目(批号10471076), 山东省自然科学基金资助项目(批号Y2004A05, Y2004A07), 湖北省教育厅科学研究计划重点项目(批号D200610002).

本文2002年12月20日收到, 2006年9月30日收到修改稿.

若个体之间存在差异性, 则$\mathsf{Var}\,\Lambda > 0$, 即有$\mathsf{Var}\,N > \mathsf{E}N$, 这表明相对于Poisson分布来说混合Poisson分布具有散度偏大的性质.

另一方面, 对于某确定的个体保单而言, 其索赔次数也会偏离Poisson分布. 事实上, 由于保险公司采用了回避风险机制, 如免赔制度、无赔款优待(NCD)制度(或BMS)等[2, 7], 使得投保人在发生事故时会权衡其利益得失而是否进行索赔, 这样, 与发生事故的次数相比, 索赔次数在0处有更大的概率; 此外, 若保险公司没有采用了回避风险机制, 或管理方面的原因使得投保人的索赔次数与发生事故的次数相比, 索赔次数在0处有更小的概率, 即产生索赔次数虚报夸大现象. 下面引出一类被称为复合Poisson-Geometric分布的分布来刻画这类现象[12].

1.2 PG分布

设随机变量X为某个体保单在n分之一单位时间内发生事故的次数, Y为保单在n分之一单位时间内实际索赔次数. 对于$\mathsf{P}(Y=0)\geq\mathsf{P}(X=0)$的情况:

$$\begin{aligned}&\mathsf{P}(X=0)=1-\alpha, \qquad \mathsf{P}(X\geq1)=\alpha, \qquad \alpha\in[0,1),\\&\mathsf{P}(Y=0)=1-\alpha+\rho\alpha, \qquad \mathsf{P}(Y\geq1)=\alpha(1-\rho), \qquad \rho\in[0,1),\end{aligned}$$

称ρ为偏离参数. 由于

$$\mathsf{P}(Y\geq1)=\alpha(1-\rho)=\sum_{k=1}^{\infty}\alpha(1-\rho)^2\rho^{k-1}.$$

进一步, 设$\mathsf{P}(Y=k)=\alpha(1-\rho)^2\rho^{k-1}$且$\pi=\alpha(1-\rho)$, 则$Y$的分布可写为

$$\mathsf{P}(Y=0)=(1-\pi), \qquad \mathsf{P}(Y=k)=\pi(1-\rho)\rho^{k-1}, \qquad k=1,2,\cdots. \tag{1.2}$$

所以, Y的母函数为

$$G_Y(t)=\sum_{k=0}^{\infty}t^k\mathsf{P}(Y=k)=1-\frac{\pi(1-t)}{1-\rho t}. \tag{1.3}$$

对于$\mathsf{P}(Y=0)\leq\mathsf{P}(X=0)$的情况:

$$\begin{aligned}&\mathsf{P}(Y=0)=1-\alpha-\rho\alpha, \qquad \mathsf{P}(Y\geq1)=\alpha(1+\rho), \qquad \rho\in[0,1),\\&\mathsf{P}(Y\geq1)=\alpha(1+\rho)=\sum_{k=1}^{\infty}\alpha(1-\rho^2)\rho^{k-1}.\end{aligned}$$

设$\mathsf{P}(Y=k)=\alpha(1-\rho^2)\rho^{k-1}$且$\pi=\alpha(1+\rho)$, 则$Y$的母函数同(1.3)式.

记Y_k $(k=1,2,\cdots,n)$为个体保单在第k个n分之一时间段内的索赔次数, 并且它们独立同分布, 其概率分布如(1.2)式所示, 则$S_n\triangleq\sum\limits_{k=1}^{n}Y_k$的母函数为

$$G_{S_n}(t)=\Big(1-\frac{\pi(1-t)}{1-\rho t}\Big)^n.$$

《应用概率统计》版权所有

令$n\to\infty$, $\pi\to 0$, $n\pi\to\lambda$, 则有

$$G(t)=\exp\Big(\frac{\lambda(t-1)}{1-\rho t}\Big). \tag{1.4}$$

定义 1.1　若随机变量ξ的母函数如(1.4)所示, 其中$\rho\in[0,1)$, $\lambda>0$为参数, 则称ξ服从复合Poisson-Geometric分布, 简记PG(ρ,λ).

定理 1.1　设随机变量ξ服从PG(ρ,λ)分布, 则

1) ξ的概率分布为

$$\begin{cases}\mathsf{P}(\xi=0)=e^{-\lambda},\\ \mathsf{P}(\xi=k)=e^{-\lambda}\sum\limits_{j=1}^{k}\dfrac{(k-1)!}{(j-1)!(k-j)!j!}(\lambda(1-\rho))^j\rho^{k-j}, \qquad k=1,2,\cdots;\end{cases} \tag{1.5}$$

2) ξ的数学期望及方差分别为

$$\mathsf{E}\xi=\frac{\lambda}{1-\rho},\qquad \mathsf{Var}\,\xi=\frac{\lambda(1+\rho)}{(1-\rho)^2};$$

3) ξ的散度参数为

$$\phi=\frac{\mathsf{Var}\,\xi}{\mathsf{E}\xi}=\frac{1+\rho}{1-\rho}.$$

证明:　见文献[12].　□

注 1.1　1) 当$\rho=0$时, 由(1.4)知, PG(λ,ρ)退化为参数为λ的Poisson分布, 因此PG(λ,ρ)是Poisson分布一个推广.

2) 当$0<\rho<1$时$\phi>1$, 此时PG分布比Poisson分布有更大的散度系数.

3) 之所以称之为"复合Poisson-Geometric分布", 是因为我们可以用以下方法得到此分布:

设随机变量N服从参数为λ/ρ的Poisson分布, $0<\rho<1$, 其矩母函数为$\exp\{\lambda/\rho(e^s-1)\}$; $\{\xi_i\}$独立且同服从参数为$1-\rho$的Geometric分布, 其矩母函数为$(1-\rho)/(1-\rho e^s)$. 所以, $S=\sum\limits_{i=1}^{N}\xi_i$的矩母函数为$\exp\{\lambda(e^r-1)/(1-\rho e^r)\}$, 可见, S的母函数与(1.4)式一致.

§2.　混合PG分布

2.1　指数类分布

以下简要介绍指数类分布, 详见[1].

定义 2.1　对参数空间$(\Theta\times\Lambda)$及定义在Θ上的二阶连续可微且一阶导函数严格单调的函数$\kappa(\theta)$, 若某一随机变量Z的分布密度或概率函数有如下的形式:

$$p^*(z;\theta,\lambda)=c^*(z;\lambda)\exp\{\theta z-\lambda\kappa(\theta)\}, \tag{2.1}$$

《应用概率统计》版权所有

则称Z服从ED*类分布, 记为$\mathrm{ED}^*(\theta,\lambda;\kappa)$; 当$\kappa$确定或默认的情况下, 简记为$\mathrm{ED}^*(\theta,\lambda)$; 其全体记为$\mathrm{ED}^*(\Theta,\Lambda;\kappa)$. 称$\theta$为典型参数, λ为散度参数, 称$\kappa(\theta)$为累积量函数, 记$\tau(\theta)\,\hat{=}\,\kappa'(\theta)$, 称$\mu=\tau(\theta)$为均值映射; 并称$V(\mu)\,\hat{=}\,\tau'(\tau^{-1}(\mu))$为单位方差函数.

由(2.1)式易计算, 其矩母函数为:

$$M_Z(r)=\exp\{\lambda(\kappa(\theta+r)-\kappa(\theta))\}. \tag{2.2}$$

当λ确定时, 单位方差函数可以唯一地决定指数分布类(详见[1] P.51).

特别地, 当$V(\mu)=\mu^p$, $p\in(-\infty,0]\cup[1,\infty)$, 称$V(\mu)$所决定的ED*为$\mathrm{TW}_p^*$类分布, 对应的分布记为$\mathrm{TW}_p^*(\theta,\lambda)$, $\theta\in\Theta$, $\lambda\in\Lambda$, 累积量函数及均值映射分别为

$$\kappa_p(\theta)=\begin{cases}\dfrac{\alpha-1}{\alpha}\Big(\dfrac{\theta}{\alpha-1}\Big)^{\alpha} & p\neq 1,2;\\ -\log(-\theta) & p=2;\\ e^{\theta} & p=1,\end{cases} \tag{2.3}$$

$$\tau_p(\theta)=\begin{cases}\Big(\dfrac{\theta}{\alpha-1}\Big)^{\alpha-1} & p\neq 1;\\ e^{\theta} & p=1,\end{cases} \tag{2.4}$$

其中: $\alpha=(p-2)/(p-1)$.

对$\mathrm{PG}(\rho,\lambda^*)$, 若$\rho=0$, 则退化为参数为$\lambda^*$的Poisson分布. 以下讨论$0<\rho<1$的情况, 将$\mathrm{PG}(\rho,\lambda^*)$表示为ED*类的形式为

$$a^*(z;\lambda)\exp\{\theta z-\lambda\kappa_a(\theta)\}, \tag{2.5}$$

其中

$$\lambda=\frac{\lambda^*(1-\rho)}{\rho}\qquad 0<\rho<1, \tag{2.6}$$

典型参数$\theta=\log\rho$, 累积量函数为

$$\kappa_a(\theta)=\frac{e^{\theta}}{1-e^{\theta}}=\frac{\rho}{1-\rho}\qquad 0<\rho<1,\ \theta\in\Theta=R_-, \tag{2.7}$$

$$a^*(z,\lambda)=\sum_{j=1}^{z}\frac{(z-1)!}{(j-1)!(z-j)!j!}\lambda^j\qquad 0<\rho<1. \tag{2.8}$$

为了少用记号, 当$0<\rho<1$时, 仍用$\mathrm{PG}(\rho,\lambda)$来表示(2.5)所示的分布, 这里$\lambda>0$表示散度参数.

《应用概率统计》版权所有

2.2 混合PG分布

由§1的讨论, 对集合保单, 用随机变量X表示个体保单之间的差异性, 在保单集合中随机抽取一份保单, 其索赔次数为Z, 若

$$\begin{cases} Z|_{X=x} \sim \mathrm{PG}(\rho, x), \\ X \sim \mathrm{ED}^*(\phi, \lambda; \kappa_X), \end{cases} \tag{2.9}$$

其中$\lambda \in \Lambda$, $\phi \in \Phi$. 称Z的分布为混合PG分布. (当$\rho = 0$时, 混合PG分布为混合Poisson分布.) 可计算Z的概率函数为

$$p_Z(z; \theta, \lambda, \psi) = c(z; \lambda, \psi) \exp\{\theta z - \lambda \kappa_X(\psi + \kappa_a(\theta))\}. \tag{2.10}$$

这里

$$c(z; \lambda, \psi) = \int_{R_+} a^*(z; x) c^*(x; \lambda) \exp\{\psi z\} \mathrm{d}x,$$

ψ满足$\psi + \kappa_a(\theta) \in \Phi$.

因此, Z仍服从ED*类, 其累积量函数为

$$\kappa(\theta) = \kappa_X(\psi + \kappa_a(\theta)).$$

特别地, 对混合PG分布, 当随机变量X服从TW$_p^*$类分布时, 称Z服从混合Tweedie*–PG分布, 记TPG$_p^*$类. 当$\rho = 0$且X服从TW$_p^*$类分布时, 称Z服从混合TW$_p^*$–Po类. 对TPG$_p^*$, 可计算其单位方差函数$V(\cdot)$满足

$$\frac{V(\mu)}{\mu} = \frac{1+\rho}{1-\rho} + \mu_X^{p-1} \frac{\rho}{(1-\rho)^2}, \tag{2.11}$$

其中$\mu_X = \kappa_X'(\phi)$. 由此可见, 相对于PG类分布, 当$\mu_X > 0$时, TPG$_p^*$类具有散度偏大性质.

例 1 当$p = 0$时, $X \sim \mathrm{TW}_0^*(\phi, \lambda)$即正态分布, 称此分布为混合正态–PG分布.

例 2 当$p = 2$时, $X \sim \mathrm{Ga}^*$分布, 可计算Z的方差函数为

$$V(\mu) = \mu\Big(\mu + 1 + \frac{2\rho}{1-\rho}\Big). \tag{2.12}$$

称此单位方差函数所对应的分布为广义负二项分布(简记为GNB). 因为, 当$\rho \to 0$时, $V(\mu) \to \mu(1+\mu)$, 表明广义负二项分布是负二项分布的一个推广.

例 3 当$p = 3$时, $X \sim \mathrm{IG}^*$分布(逆高斯分布). 称Z的分布为混合IG*–PG类分布.

《应用概率统计》版权所有

§3. 用TPG_p^*类分布来拟合索赔次数数据

3.1 TPG_p^*类分布中参数的矩估计方法

设随机变量Z服从TPG_p^*类分布, 当$0<\rho<1$时, 其概率分布表示式如(2.10)所示, 其中: $\kappa_a(\theta)$, $\kappa_p(\theta)$分别如(2.7)、(2.3)所示. 则Z的母函数为:

$$G(t)=\exp\{\lambda(\kappa_p(\psi+\kappa_a(\theta+\log t))-\kappa_p(\psi+\kappa_a(\theta)))\}. \tag{3.1}$$

记$m_i \,\widehat{=}\, \mathsf{E}Z^i$ $(i=1,2,3,4)$, 由母函数与矩之间的关系得:

$$\begin{cases} G'(1)=m_1, \\ G''(1)=m_2-m_1, \\ G'''(1)=m_3-3m_2+2m_1, \\ G^{(4)}(1)=m_4-6m_3+11m_2-6m_1. \end{cases} \tag{3.2}$$

记

$$K(t)=\lambda(\kappa_p(\psi+\kappa_a(\theta+\log t))-\kappa_p(\psi+\kappa_a(\theta))),$$

则有

$$\begin{cases} K'(1)=m_1 \,\widehat{=}\, u_1, \\ K''(1)=m_2-m_1-m_1^2 \,\widehat{=}\, u_2, \\ K'''(1)=m_3-3m_2+2m_1-m_1^3-3m_1u2 \,\widehat{=}\, u_3, \\ K^{(4)}(1)=m_4-6m_3+11m_2-6m_1-m_1^4-6m_1u_2^2-3u_2^2-4m_1u_3 \,\widehat{=}\, u_4. \end{cases} \tag{3.3}$$

用样本观察值代替u_1, u_2, u_3, u_4, 由(3.3)式所解出的$\widehat{\rho}$, $\widehat{\lambda}$, $\widehat{\psi}$, $\widehat{p}$即为TPG_p^*分布中参数的矩估计.

对$p\neq 1,2$及$\rho\neq 0$的情况, 参数的矩估计步骤如下:

1) 由方程$c_4\rho_1^4+c_3\rho_1^3+c_2\rho_1^2++c_1\rho_1+c_0=0$解出$\rho_1$, 其中:

$$\begin{cases} c_0=2a_3^2+a_2^2a_3-2a_2a_4, \\ c_1=-16a_2a_3-6a_2^3+2a_4, \\ c_2=4a_3+6a_2^2, \\ c_3=-96a_2, \\ c_4=48, \end{cases} \qquad \text{其中} \begin{cases} a_1=u_1, \\ a_2=u_2/u_1, \\ a_3=u_3/u_1, \\ a_4=u_4/u_1; \end{cases}$$

《应用概率统计》版权所有

2) 计算ρ, ψ, p, λ

$$\begin{cases}\rho=\rho_1/(1+\rho_1),\\ p=[a_3-6\rho_1^2-6\rho_1(a_2-2\rho_1)]/(a_2-2\rho_1),\\ \psi=-\rho_1(1+\rho_1)/[(p-1)(a_2-2\rho_1)]-\rho_1,\\ \lambda=a_1(\rho_1(1+\rho_1))^{-(p-2)/(p-1)}(a_2-2\rho_1)^{-1/(p-1)};\end{cases}$$

3) 取满足条件$\rho>0$, $\lambda>0$, $\psi<0$, $p\in(-\infty,0]\cup(1,\infty)$且$p\neq 1,2$的解作为估计值. 对$p=1,2$或$\rho=0$的情况, 只需要三阶矩即可得到参数$\rho$, ψ, λ的矩估计.

3.2 实际例子

利用混合PG分布对我国某保险公司1996年的35072辆投保车辆的第三责任保险索赔次数的拟合结果如下表所示(数据来源见文献[7]).

表1 拟合结果

索赔次数	观察值	拟合值				
		Possion	PG	IG*–PG	TW*–Po	TW*–PG
0	27141	25528.60	27330.45	27194.69	27181.29	27184.64
1	5789	8107.99	5352.54	5638.59	5677.48	5667.53
2	1443	1287.56	1673.54	1556.86	1524.47	1533.15
3	457	136.31	506.14	460.38	461.96	461.13
4	155	10.82	149.23	145.61	149.44	148.44
5	56	0.69	43.11	48.79	50.33	49.99
6	27	0.36	12.25	17.12	17.42	17.38
7	2	0.0017	3.43	6.23	6.15	6.18
8	1	0	0.95	2.33	2.2	2.24
9	1	0	0.26	0.89	0.79	0.82
≥10	0	0	0.1	0.58	0.46	0.18
χ^2		37185.87	98.04	24.07	21.14	22.94

表中所有分布均采用矩估计, 用SAS/IML模块[10]编程计算, 结果如下:

1) 对PG分布, 参数$\widehat{\rho}=0.2147$, $\widehat{\lambda}=0.2494$;

2) 对广义负二项分布, 得不到参数的矩估计;

3) 对混合IG*–PG分布, 参数$\widehat{\rho}=0.0776$, $\widehat{\lambda}=1.709$, $\widehat{\psi}=-0.2046$;

4) 对$\rho=0$的情况, 参数$\widehat{p}=2.21899$, $\widehat{\lambda}=0.52146$, $\widehat{\psi}=-2.4999$;

《应用概率统计》版权所有

5) 对$\rho\neq0$的情况, 参数$\widehat{\rho}=0.0226$, $\widehat{p}=2.3578$, $\widehat{\lambda}=1.4181$, $\widehat{\psi}=-0.0579$.

从直观拟合效果来看, 4)及5)的拟合结果比较接近.

§4. 假设检验

由表1可以直观地看到, 用混合PG类布拟合上述索赔次数数据时, $\rho=0$与$\rho\neq0$的拟合效果看不出明显的差异, 它们的χ^2值分别为21.14及22.94. 对后者, ρ的估计值为0.0226, 是否可以接受$\rho=0$? 由于这里采用的是矩估计而不是最大似然估计, 无法由χ^2值作出判断. 因此, 还需要用其它方法进行检验.

设集合中有n份保单, 从中随机抽取一份保单, 由第1、2节的分析, 设其索赔次数Z服从$\mathrm{TPG}_p^*(\rho,\psi,\lambda)$分布, 我们提出假设

$$H_0:\ \rho=0.$$

为了构造检验统计量, 记X_i为第i份个体保单的索赔次数$(i=1,2,\cdots,n)$, 并设

$$\{X_i\}_{i=1}^n \stackrel{\text{i.i.d.}}{\sim} \mathrm{TPG}_p^*(\rho,\psi,\lambda),$$

即$X=(X_1,\cdots,X_n)$为简单子样, 其k阶样本矩记为:

$$\overline{N^k}\triangleq\frac{1}{n}\sum_{i=1}^n X_i^k.$$

记

$$Y_i=\begin{cases}1 & X_i=0\\ 0 & X_i\neq 0\end{cases}\quad(i=1,2,\cdots,n),\qquad T_0\triangleq\frac{1}{n}\sum_{i=1}^n Y_i,$$

由于$\{X_i\}_{i=1}^n$ i.i.d., 所以$\{Y_i\}_{i=1}^n$ i.i.d..

设母体分布的k阶矩为m_k $(k=1,2,3)$, n_0为集合保单中0次索赔的保单数, 记$t_0\triangleq n_0/n$为0次索赔的频率, 它可以看为T_0的一个观察值, 记$\mu_0\triangleq \mathsf{E}T_0=\mathsf{E}Y_i$. 当$H_0$为真即$\rho=0$时, 可以得到矩估计结果:

$$\begin{cases}p=p(m_1,m_2,m_3),\\ \psi=\psi(m_1,m_2,m_3),\\ \lambda=\lambda(m_1,m_2,m_3).\end{cases}$$

记Z在0处的概率为

$$p_0(\psi,p,\lambda)\triangleq\exp\{\lambda[\kappa_p(\psi)-\kappa_p(\psi+1)]\}.\tag{4.1}$$

若拟合实际数据且效果较好, 则$|t_0-p_0|$应该"较小". 基于这一思想, 构造统计量(参见文献[9])

$$T\triangleq T_0-p_0(\widehat{\psi},\widehat{p},\widehat{\lambda}),\tag{4.2}$$

《应用概率统计》版权所有

这里$\widehat{\psi}$, $\widehat{p}$, $\widehat{\lambda}$为样本矩估计. 由中心极限定理有

$$\sqrt{n}\left[\begin{pmatrix}\overline{N^1}\\ \overline{N^2}\\ \overline{N^3}\\ T_0\end{pmatrix}-\begin{pmatrix}m_1\\ m_2\\ m_3\\ \mu_0\end{pmatrix}\right]\xrightarrow{d}N(\overrightarrow{0},\Sigma), \tag{4.3}$$

其中

$$\overrightarrow{0}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix},\qquad \Sigma=\begin{pmatrix}m_2-m_1^2 & m_3-m_1m_2 & m_4-m_1m_3 & -m_1\mu_0\\ m_3-m_1m_2 & m_4-m_2^2 & m_5-m_2m_3 & -m_2\mu_0\\ m_4-m_1m_3 & m_5-m_2m_3 & m_6-m_3^2 & -m_3\mu_0\\ -m_1\mu_0 & -m_2\mu_0 & -m_3\mu_0 & \mu_0(1-\mu_0)\end{pmatrix}. \tag{4.4}$$

将T视为$\overline{N^1}$, $\overline{N^2}$, $\overline{N^3}$, T_0的函数, 记为$T(m_1,m_2,m_3,\mu_0)$, 令

$$\beta\triangleq\Big(\frac{\partial T}{\partial m_1},\frac{\partial T}{\partial m_2},\frac{\partial T}{\partial m_3},\frac{\partial T}{\partial \mu_0}\Big),$$

当假设H_0成立时, 由估计量的相合渐近正态性(参见文献[6] P.181–184)有

$$\sqrt{n}T\xrightarrow{d}N(0,\sigma^2), \tag{4.5}$$

其中$\sigma^2=\beta\Sigma\beta'$.

方差σ^2可以表示为$\sigma^2(m_1,m_2,m_3,\mu_0)$, 由矩估计的相合性(参见文献[6] P.193–195)有

$$\sqrt{n}T\xrightarrow{d}N(0,\widehat{\sigma}^2), \tag{4.6}$$

其中$\widehat{\sigma}^2=\sigma^2(\overline{N^1},\overline{N^2},\overline{N^3},T_0)$.

对给定的检验水平α, 检验假设$H_0:\rho=0$, 由上面的结论可知: 当$|\sqrt{n/\sigma^2}T|>U_{\alpha/2}$时, 拒绝$H_0$. 这里$U_{\alpha/2}$为正态分布的临界值. 具体计算步骤如下:

1) 由矩估计得

$$\begin{cases}p=\dfrac{u_1u_3}{u_2^2},\\ \psi=-\dfrac{u_1u_2}{u_1u_3-u_2^2}-1,\\ \lambda=u_1(u_2/u_1)^{-1/(p-1)};\end{cases} \tag{4.7}$$

2) 由

$$T=\exp\Big\{-\frac{\lambda}{p-2}[(-\psi(p-1))^{(p-2)/(p-1)}-(-(\psi+1)(p-1))^{(p-2)/(p-1)}]\Big\}, \tag{4.8}$$

《应用概率统计》版权所有

计算

$$\frac{\partial T}{\partial(p,\psi,\lambda)} \widehat{=} \Big(\frac{\partial T}{\partial p}, \frac{\partial T}{\partial \psi}, \frac{\partial T}{\partial \lambda}\Big);$$

3) 分别由(4.7)式、(3.3)式计算

$$\frac{\partial(p,\psi,\lambda)}{\partial(u_1,u_2,u_3)} \quad 及 \quad \frac{\partial(u_1,u_2,u_3)}{\partial(m_1,m_2,m_3)};$$

4) 计算

$$\beta = \Big(\frac{\partial T}{\partial(p,\psi,\lambda)} \cdot \frac{\partial(p,\psi,\lambda)}{\partial(u_1,u_2,u_3)} \cdot \frac{\partial(u_1,u_2,u_3)}{\partial(m_1,m_2,m_3)}, 1\Big);$$

5) 计算$\sigma^2 = \beta\Sigma\beta'$及$|\sqrt{n/\sigma^2}T|$.

用Matlab软件[11]编程计算, 结果如下:

$$\begin{cases} \widehat{p} = 2.21899, & \widehat{t_0} = 0.7738652, \\ \widehat{\psi} = -2.499, & \widehat{p_0} = 0.775014, \\ \widehat{\lambda} = 0.52146, & \widehat{\sigma}^2 = 0.00501787, \\ |\sqrt{n/\widehat{\sigma}^2}T| = 3.037245. \end{cases} \tag{4.9}$$

取检验水平为$\alpha = 0.05$, 查表得临界值为$U_{\alpha/2} = 1.96$, 因为$|\sqrt{n/\widehat{\sigma}^2}T| > 1.96$, 所以, 拒绝原假设$H_0: \rho = 0$. 我们接受$\rho \neq 0$, 即认为用混合$\mathrm{TW}_p^*$–Po类分布来描述该集合保单的索赔次数不合适.

值得一提的是, 作者对文献[9]以及文献[5]等国外保险索赔数据用上述方法进行检验, 其结果均不能拒绝原假设$H_0: \rho = 0$, 表明用混合TW_p^*–Po类分布来描述该集合保单的索赔次数较合适, 这与本文例子的结果相反. 因此, 从这一角度上说, 本文中的数据与上述国外保险索赔数据有显著的不同. 因此在制定保费定价政策时, 不能完全套用国外的方法, 这一点值得保险公司及有关部门的注意. 详细的计算结果见表2.

表2 国外保险索赔数据及计算结果(数据(1)来源于文献[9], 数据(2)(3)来源于文献[5]

索赔次数	数据(1)	数据(2)	数据(3)	计算结果	数据(1)	数据(2)	数据(3)
0	103704	370412	7840	$\widehat{p}$	3.132036	4.496455	3.959785
1	14075	46545	1317	$\widehat{\psi}$	-4.0100	-6.5542	-1.9715
2	1766	3935	239	$\widehat{\lambda}$	0.371032	0.307718	0.306272
3	255	317	42	$\widehat{t_0}$	0.8652599	0.879337	0.828665
4	45	28	14	$\widehat{p_0}$	0.8652898	0.879340	0.828617
5	6	3	4	$\widehat{\sigma}^2$	0.000310	0.0000515	0.001703
6	2		4	$\lvert\sqrt{n/\widehat{\sigma}^2}T\rvert$	0.578464	0.294650	0.111656
7			1				

《应用概率统计》版权所有

参 考 文 献

[1] Jørgensen, B., *The Theory of Dispersion Models*, Chapman and Hall, London, 1997.
[2] Lemaire, J., *Bonus-Malus Systems in Automobile Insurance*, Kluwer Academic Publishers, 1995.
[3] McCullagh, P. and Nelder, J.A., *Generalized Linear Models* (Second edition), Chapman and Hall, London, 1989.
[4] Nelder, J.A. and Pregibon, D., An extended quasi-likelihood function, *Biometrika*, **74**(1987), 221–232.
[5] Willmot, G., Sundt and Jewell's family of discrete distribution, *ASTIN Bulltin*, **18**(1988), 17–29.
[6] 陈希孺, 数理统计引论, 科学出版社, 1999.
[7] 孟生旺, 袁卫, 汽车保险的精算模型及应用, 数理统计与管理, **20**(2001), 60–65.
[8] 孟生旺, 袁卫, 实用非寿险精算学, 经济科学出版社, 2000.
[9] 吴大伟, 复合Poisson过程参数的检验, 应用概率统计, **18**(2002), 409–412.
[10] SAS/IML Software: Usage and Reference, Version 6, First Edition, SAS Institute Inc..
[11] 云舟工作室, MATLAB 6数学建模基础教程, 人民邮电出版社, 2001.
[12] 毛泽春, 刘锦萼, 一类索赔次数的回归模型及其在风险分级中的应用, 应用概率统计, **20**(2004), 359–367.

Mixed Exponential Dispersion Distributions and Their Applications in Insurance Claims

MAO ZECHUN[1,2] LIU JINE[2]

([1]*Department of Finance, School of Business, Hubei University, Wuhan, 430062*)

([2]*Institute of Statistics and Actuary, Shandong Economics University, Jinan, 250014*)

The mixed exponential dispersion distributions has been studied based on the background in insurance claims. The distribution can be generated by mixing ED* and compound Poisson-Geometric distribution. The steps of fitting the numbers of claims by using moment estimation and the method about hypothesis test have been given in the paper. As an example, the data in automobile insurance in China has been investigated.

Keywords: Numbers of claims, compound Poisson-Geometric distribution;mixed PG distribution.

AMS Subject Classification: 62E20, 62G09.

《应用概率统计》版权所有