三、理赔次数的混合分布

背景：

从保单中随意抽取一份保单，求该保单的理赔次数分布。

同质性：指所有的保单相互独立，且都有相同的风险水平，即各保单的损失额的分布相同，损失次数的分布也相同。

非同质性：保单组合中的每个保单风险水平各不相同。θ 表示其风险水平。

数学模型：

设 Θ 为一个随机变量，当 $\Theta=\theta$ 时，N 的分布为 $p_k(\theta)=P(N=k\,|\,\Theta=\theta)$，令 $v(\theta)=P(\Theta\le\theta)$ 为 Θ 的累积分布，$u(\theta)$ 为 θ 的 pdf，则 N 的分布列为

$$p_k=P(N=k)=\int p_k(\theta)u(\theta)d\theta$$

或者：

$$p_k=P(N=k)=\sum_{\theta_i} p_k(\theta_i)u(\theta_i)$$

N 的分布称为混合分布。$\{p_k(\theta),k=0,1,2\cdots\}$为泊松分布时，N 的分布称为混合泊松分布。

①.混合分布性质

1．母函数

$$P_N(z)=\int P_N(z\,|\,\theta)u(\theta)d\theta$$

或者：

$$P_N(z)=\sum P_N(z\,|\,\theta_i)u(\theta_i)$$

其中$P_N(z\,|\,\theta)$表示在$\Theta=\theta$条件下，N 的母函数。

2．均值和方差

$$E(N)=E(E(N\,|\,\Theta))$$

$$Var(N)=E[Var(N\,|\,\Theta)]+Var[E(N\,|\,\Theta)]$$

定理：任意两个随机变量 X 和 Y，

$$Var(X)=E(Var(X|Y))+Var(E(X|Y))$$

证明： 由于

$$\begin{aligned}Var(X|Y)&=E\{[X-E(X|Y)]^2|Y\}\\&=E(X^2|Y)-[E(X|Y)]^2\end{aligned}$$ 。

因此

$$\begin{aligned}E[Var(X|Y)]&=E\{E(X^2|Y)-[E(X|Y)]^2\}\\&=E(X^2)-E\{[E(X|Y)]^2\}\end{aligned}$$

另一方面

$$\begin{aligned}Var[E(X|Y)]&=E\{[E(X|Y)]^2\}-\{E[E(X|Y)]\}^2\\&=E\{[E(X|Y)]^2\}-[E(X)]^2\end{aligned}$$

从而

$$
\begin{aligned}
&E[Var(X \mid Y)] + Var[E(X \mid Y)] \\
&= E(X^2) - E\{[E(X \mid Y)]^2\} + E\{[E(X \mid Y)]^2\} - [E(X)]^2 \\
&= E(X^2) - [E(X)]^2 \\
&= Var(X)
\end{aligned}
$$

②. 常见的几种混合泊松分布

1、离散型混合

对于规模较小的保单组合，假设保单组合由 n 种不同的风险水平构成，泊松参数 Λ 取值于 $\{\lambda_1,\cdots,\lambda_n\}$，$\lambda_1<\cdots<\lambda_n$，设 $a_k=P(\Lambda=\lambda_k)$，$k=0,1,2,\cdots,n$。当 $\Lambda=\lambda_k$ 时，保单的损失次数服从参数为 λ_k 的泊松分布。则从保单组合中任意抽取一份保单的分布为

$$P(N=k)=\sum_{i=1}^{n}P(N=k\mid\Lambda=\lambda_i)P(\Lambda=\lambda_i)$$

$$=\sum_{i=1}^{n}a_i\frac{\lambda_i^k}{k!}e^{-\lambda_i},\qquad k=0,1,2,\cdots$$

例 2.7：假设投保车险的驾驶员可以分为两类，他们出事的次数服从泊松分布，其中好的一类的泊松参数为 0.11，坏的一类的泊松参数为0.70,好的驾驶员和坏的驾驶员的比例为0.94和 0.06，则任意一个驾驶员出事的次数分布时多少？

解：

$$P(N=k)=p\frac{e^{-\lambda_1}\lambda_1^k}{k!}+(1-p)\frac{e^{-\lambda_2}\lambda_2^k}{k!}$$
$$=0.94\frac{e^{-0.11}0.11^k}{k!}+0.06\frac{e^{-0.70}0.70^k}{k!}$$

2、连续型的混合

对于规模较大的保单组合，可以假设其中的泊松参数Λ

服从连续分布。以$u(\lambda)$表示Λ的密度函数，通常称为结构函数。则从保单组合中随机抽取一份保单的损失次数分布为

$$P(N=k)=\int_0^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}u(\lambda)d\lambda\,,\quad k=0,1,2,\cdots$$

性质（自行阅读）：

1、　母函数的表达式

假设存在一个随机变量Θ和常数λ使得$\Lambda=\lambda\Theta$，则

$$P(z)=P_{\Theta}(\exp(\lambda(z-1)))$$

证明：设$u(\theta)$为θ的密度函数

$$
\begin{aligned}
P(z) &= \int e^{-\lambda\theta(z-1)} u(\theta) d\theta \\
&= \int (e^{-\lambda(z-1)})^{\theta} u(\theta) d\theta \\
&= P_{\Theta}(e^{\lambda(z-1)})
\end{aligned}
$$

2、设 $P_1(z)$ 和 $P_2(z)$ 是两个混合泊松分布的母函数，分别表示为

$$P_1(z) = \int e^{-\lambda\theta(z-1)} u(\theta) d\theta$$

$$P_2(z) = \int e^{-\lambda\theta(z-1)} v(\theta) d\theta$$

若 $P_1(z) = P_2(z)$，则有 $u(\theta) = v(\theta)$。

证明略

3*. 设 $P(z)$ 是混合泊松分布的母函数，如果 $P(z)$ 满足无穷可分

性，即

$$P(z)=(P_n(z))^n$$

其中$P_n(z)$也是一个母函数，则$P(z)$也是一个复合泊松分布的母函数，即

$$P(z)=\exp[\lambda(P_M(z)-1)]$$

特别的，若$P_M(0)=0$，则$P_M(z)$是唯一的。

无穷可分的含义是指，设Y的特征函数是$\varphi(x)$，则对任意的n都存在n个的独立同分布的随机变量$X_1,X_2\cdots X_n$，使得

$$Y=X_1+X_2+\cdots+X_n$$

此时，

$$\varphi(x)=\left(\varphi_n(x)\right)^n$$

其中 $\varphi_n(x)$ 是 $X_1,X_2\cdots X_n$ 的特征函数。

你能举几个无穷可分分布的例子吗？

例 2.8： 设 Θ 的母函数为

$$P_\Theta(z)=(\frac{\beta}{\beta-\log z})^\alpha$$

求 N 的分布。

解：利用母函数公式

$$P_N(z) = P_\Theta(\exp(\lambda(z-1)))$$

$$= \left\{ \frac{\beta}{\beta - \log[\exp(\lambda(z-1))]} \right\}^{\alpha}$$

$$= \left\{ \frac{\beta}{\beta - \lambda(z-1)} \right\}^{\alpha}$$

$$= \left[1 - \frac{\lambda}{\beta}(z-1) \right]^{-\alpha}$$

负二项分布可以看成是泊松分布和 gamma 分布的混合。

定理： 设 $N|_{\Lambda=\lambda}$ 服从参数为λ的泊松分布， Λ 是一个随机变量，服从 gamma 分布， $f_{\Lambda}(\lambda)=\dfrac{\lambda^{\alpha-1}e^{-\frac{\lambda}{\theta}}}{\theta^{\alpha}\Gamma(\alpha)}$， 则 N 服从负二项分布，参数为 $\beta=\theta$， $r=\alpha$ 。

证明：

$$P(N=k)=\int_0^\infty P(N=k\mid\Lambda=\lambda)\,f_\Lambda(\lambda)d\lambda$$

$$=\int_0^\infty \frac{\lambda^{\alpha-1}e^{-\frac{\lambda}{\theta}}}{\theta^\alpha\Gamma(\alpha)}e^{-\lambda}\frac{\lambda^k}{k!}d\lambda$$

$$=\frac{1}{\theta^\alpha\Gamma(\alpha)k!}\int_0^\infty \lambda^{\alpha+k-1}e^{-\lambda(1+\frac{1}{\theta})}d\lambda$$

?

$$=\frac{1}{\theta^\alpha\Gamma(\alpha)k!}\int_0^\infty (\frac{\theta}{\theta+1})^{\alpha+k}$$

$$=\frac{\Gamma(\alpha+k)}{\Gamma(\alpha)k!}\frac{\theta^k}{(\theta+1)^{\alpha+k}}$$

$$=\binom{k+\alpha-1}{k}(\frac{\theta}{1+\theta})^k(\frac{1}{1+\theta})^\alpha$$

例2.9：

Actuaries have modeled auto windshield claim frequencies. They have concluded that the number of windshield claims filed per year per driver follows the Poisson distribution with parameter λ， where λ follows the gamma distribution with mean 3 and variance 3.

Calculate the probability that a driver selected at random will file no more than 1 windshield claim next year.

解：设 gamma 分布参数为α和θ。由 gamma 分布的均值和方差公式有

$$\alpha\theta = 3$$

$$\alpha\theta^2 = 3$$

解得$\alpha = 3$，$\theta = 1$。

由前面的定理知，N 服从负二项分布，参数$\beta = \theta = 1$，$r = \alpha = 3$

于是计算得到

$$P(N \le 1) = p_0 + p_1 = 0.125 + 0.1875 = 0.3125$$

四、理赔次数的复合分布

问题：一次损失事故的发生可能会导致多份保单同时发

生索赔，如何求索赔次数的分布。

例 2.10： 设从城市 A 到城市 B 的某航线每个月有 70 个航班，假设每个航班有 2% 的可能性取消，假设每次飞行有 0.00001的概率出事。进一步假设每趟飞机有 200 个座位，每次飞行有 90% 的就座率和 6 个机组人员，假设出事飞机上的每个人都死亡，并且都买了保险。

求每个月此航线的索赔次数的分布、期望和方差。.

解：令 S 表示下个月此航线的总索赔次数

N 表示下个月出行的航班数

$$N \sim B(n_1, p), n_1 = 70, p = 0.98$$

P 表示飞机上的人员数， M 表示乘客数

$M \sim B(n_2, p), n_2 = 200, p = 0.9 \qquad P = 6 + M$

D 表示发生事故的死亡人数，

$$D = \begin{cases} 0 & 0.99999 \\ P & 0.00001 \end{cases}$$

则 $S = D_1 + D_2 + \cdots + D_N$。

定义：设M和N分别为两个计数随机变量，$M_1, M_2, \cdots M_N \quad i.i.d$ 与M的分别相同，则$S = M_1 + M_2 + \cdots + M_N$的分布称为$M$的复合分布，$N$的分布称为第一分布，$M$称为第二分布。

背景：N表示单位时间内损失事故的发生数，

M_i表示第i个损失事故产生的索赔次数，

$S = M_1 + M_2 + \cdots + M_N$表示单位时间内索赔的总次数。

S 的性质

1、母函数

设N的分布列为 $f_N(n)$，母函数为$P_N(z)$，

M的分布列为$f_M(n)$，母函数为$P_M(z)$

$$P_S(z)=P_N(P_M(z))$$

证明：

$$P_S(z)=\sum_{k=0}^{\infty}\sum_{n=0}^{\infty}P(N=n)P(S=k\mid N=n)z^k$$

$$=\sum_{n=0}^{\infty}P(N=n)\sum_{k=0}^{\infty}z^k P(M_1+M_2+\cdots+M_N=k\mid N=n)$$

$$=\sum_{n=0}^{\infty}P(N=n)P_M(z)^n=P_N(P_M(z))$$

例 2.11：M 服从泊松分布，N 服从泊松分布，$P_N(z)=e^{\lambda_1(z-1)}$，$P_M(z)=e^{\lambda_2(z-1)}$，

$$P_S(z)=P_N(P_M(z))=\exp(\lambda_1(e^{\lambda_2(z-1)}-1))$$

例 2.12：求例 2.10 中 S 的母函数：

$$P_N(z)=(1+p(z-1))^n=(1+0.98(z-1))^{70}$$

$$f_D(0)=0.99999\text{，}\quad f_D(6+j)=0.00001P(M=j)$$

$$P_D(z)=\sum_{k=0}^{\infty}z^k f_D(k)=0.99999+0.00001\times[z^6(1+0.9(z-1))^{200}]$$

$$\begin{aligned}P_S(z)&=P_N(P_D(z))\\&=[1+0.98((0.99999+0.00001z^6(1+0.9(z-1))^{200})-1)]^{70}\end{aligned}$$

2、均值和方差

$$E(S)=E(M)E(N)$$

$$Var(S)=E(N)Var(M)+E(M)^2Var(N)$$

证明：令$S_n=M_1+M_2+\cdots+M_n$，

$$\begin{aligned}E(S) &= \sum_{n=0}^{\infty} E(S \mid N=n)P(N=n) \\ &= \sum_{n=0}^{\infty} nE(M)P(N=n) \\ &= E(M)E(N)\end{aligned}$$

这个证明可以简写成

$$E(S)=E(E(S \mid N))=E(NE(M))=E(M)E(N)$$

利用方差分解公式

$$\begin{aligned}Var(S) &= E(Var(S \mid N))+Var(E(S \mid N)) \\ &= E(NVar(M))+Var(NE(M)) \\ &= E(N)Var(M)+Var(N)E(M)^2\end{aligned}$$

例 2.13： 求例 2.10 中 S 的期望和方差

$$E(N) = n_1 p = 68.6$$

$$Var(N) = 70 \times 0.98 \times 0.02 = 1.372$$

$$E(P) = 6 + 200 \times 0.9 = 186$$

$$Var(P) = 200 \times 0.9 \times 0.1 = 18$$

$$E(P^2) = Var(P) + E(P)^2 = 34614$$

$$E(D) = 0.99999(0) + 0.00001(186) = 0.00186$$

$$E(D^2) = 0.99999(0) + 0.00001(34614) = 0.34614$$

因此

$$E(S) = E(N)E(D)$$

$$Var(S) = E(N)Var(D) + E(D)^2 Var(N)$$

$$E(S) = (68.60)(0.00186) = 0.1276$$

$$Var(S) = (68.60)(0.34137) + (0.00186^2)(1.372) = 23.7450$$

3．S 的分布

我们将在第三章中详细介绍 S 分布的计算。

五、免赔额对理赔次数的分布的影响

注意：当免赔额存在时，理赔次数不等于损失次数。

1、免赔额存在时

以 X 表示损失，N^L 表示损失次数，d 表示免赔额，$v=P(X>d)$，N^P 表示理赔次数。令

$$I=\begin{cases}1 & \text{理赔发生}\\ 0 & \text{理赔不发生}\end{cases}$$

则 $P(I=1)=v$， $N^P=I_1+\cdots+I_{N^L}$，

$$P_I(z)=P(I=0)z^0+P(I=1)z^1=1+v(z-1)$$

由复合分布的母函数有

$$P_{N^P}(z)=P_{N^L}(P_I(z))=P_{N^L}(1+v(z-1))$$

例 2.14： 设某损失事件的损失额有几种可能 25，50，75，10，发生的概率分别为 $0.2, 0.3, 0.2, 0.15, 0.1, 0.05$，假设损失事件的次数服从 $\beta = 0.3, r = 10$ 的负二项分布，免赔额为 50，求赔偿事件的次数的分布。

解

$$v = P(X > 50) = 0.2 + 0.15 + 0.1 + 0.05 = 0.5$$

$$\begin{aligned} P_{N^P}(z) &= (1 - 0.3(1 + 0.5(z-1) - 1))^{-10} \\ &= (1 - 0.15(z-1))^{-10} \end{aligned}$$

所以N^{P}服从负二项分布， $\beta^{*}=0.15, r=10$ 。

命题 1：假设 N^L 的母函数 $P_N(z;\theta)=B(\theta(z-1))$，其中 $B(\cdot)$ 是与参数 θ 无关的函数，则 N^L 和 N^P 的分布类型没有变化。

证明：

$$\begin{aligned}
P_{N^P}(z) &= P_{N^L}(1+v(z-1)) \\
&= B(\theta(1+v(z-1)-1)) \\
&= B(v\theta(z-1)) \\
&= P_{N^L}(z;v\theta)
\end{aligned}$$

注：所有的$(a,b,0)$分布都保持原来的类型。

损失次数 N^L 的分布	理赔次数 N^P 的参数	期望理赔次数数
泊松分布	$\lambda^* = \lambda v$	λv
二项式分布	$q^* = vq$，　$n^* = n$	nqv
负二项式分布	$\beta^* = v\beta, r^* = r$	$r\beta v$

二项式分布的证明：设 N 服从二项式分布 $B(n,q)$，

$$P_N(z) = (1+q(z-1))^n$$

$$
\begin{aligned}
P_{N^*}(z) &= P_N(1+v(z-1)) \\
&= \{1+q[(1+v(z-1)-1]\}^n \\
&= \{1+qv(z-1)\}^n
\end{aligned}
$$

对于（a,b,1）分布，也有类似的结果

命题：假设 N^L 的分布母函数具有如下形式

$$P_{N^L}(z) = P(z;\theta,\alpha) = \alpha + (1-\alpha)\frac{B[\theta(z-1)] - B(-\theta)}{1-B(-\theta)}$$

其中 $\alpha = P_{N^L}(0) = P(N^L = 0)$。则 N^P 的母函数为

$$P_{N^P}(z) = P_{N^L}(z;v\theta,\alpha^*)$$

其中 $\alpha^* = P(N^P = 0) = P_{N^P}(0) = P_{N^L}(1-v;\theta,\alpha)$。

例 2.15：在上例中假设 N^L 服从 ZM-NB 分布，r=2，$\beta=3$，$p_0^M=0.4$，求 N^P 的分布。

解：N^L 的分布母函数为

$$p_0^M+(1-p_0^M)\frac{[1-\beta(z-1)]^{-r}-(1+\beta)^{-r}}{1-(1+\beta)^{-r}}$$

因此 $\alpha=p_0^M, B(z)=(1-z)^{-r}$。根据命题，$N^P$ 服从 ZMNB 分布 r*=2，β*=3*0.5=1.5,。

$$\alpha^*=p_0^{M*}=p_0^M+(1-p_0^M)\frac{[1-v\beta]^{-r}-(1+\beta)^{-r}}{1-(1+\beta)^{-r}}$$

$$=0.4+0.6\times\frac{(1+1.5)^{-2}-4^{-2}}{1-4^{-2}}=0.4595$$

2、免赔额发生变化时

设原来的免赔额为d，现在免赔额调整为d^*，请问调整后新理赔次数发生了什么变化。

记N^d表示免赔额为d的理赔次数，N^{d*}表示理赔额为d^*的理赔次数，设v'表示在免赔额提高后，以前的索赔事件能够继续获得赔偿的比例，则

$$v' = \frac{1 - F_X(d^*)}{1 - F_X(d)}$$

令$I=1$表示继续获得赔偿，$I=0$表示不能继续获得赔偿

$$N^{d*} = I_1 + I_2 + \cdots + I_{N^d}$$

当$d^* > d$时，$v' < 1$，当N^d为（a, b, 0）分布时，N^{d*}的分布类型与N^d相同，只是参数发生变化。

当$d^* < d$时，则$v' = \dfrac{1-F_X(d^*)}{1-F_X(d)} > 1$，此时$N^*$的参数可能超出原频率分布的参数范围，因此我们不能考虑这种情形。

六、保单数目与保单组合总理赔次数

假设当前的保单组合中含有n份保单，令N_j表示第j个保

单产生理赔的次数，则 $N = N_1 + N_2 + \cdots + N_n$ 表示保单组合产生理赔的总次数。若假设 N_j $i.i.d$，则

$$P_N(z) = (P_{N_1}(z))^n$$

问题：假设现在保单组合中的保单数 n' 个，问新保单组合的总理赔次数分布是什么？

回答：令$N' = N_1 + N_2 + \cdots + N_{n'}$，则

$$P_{N'}(z) = (P_{N_1}(z))^{n'} = \left(\sqrt[n]{P_N(z)}\right)^{n'} = (P_N(z))^{n'/n}$$

注：若N是无穷可分的，则N'的分布类型与N相同，只是修改了某些参数。

满足无穷可分性的分布有

泊松、负二项、复合泊松、复合负二项分布等

例 2.16：某家有 300 个员工的公司的劳工补偿险总理赔次数在单位时间内服从 $\beta=0.3, r=10$ 的负二项分布，问如果这家公司扩展到 500 个员工，则理赔次数的分布是什么？

解：设一个员工的理赔次数为 N_1，N＝300，N^*＝500，

$$P_{N^*}(z)=(P_{N_1}(z))^{n^*}=(P_N(z))^{500/300}$$

$$=(1-0.3(z-1)^{-10})^{500/300}$$

$$=(1-0.3(z-1))^{-16.67}$$

N^*服从$\beta=0.3, r=16.67$的负二项分布。