第3讲 其他聚合风险模型

2019年9月17日

本讲主要内容

- 索赔数分布——计数变量
  - 一类特殊的计数分布
  - 负二项分布
- 索赔额分布——非负连续型随机变量（右截断）
  - 指数、对数正态
  - Gamma、Beta
  - 混合分布

一类计数分布：

$$(a, b, 0)$$

$$p_n = (a + \frac{b}{n}) p_{n-1}, n = 1, 2, \ldots$$

常见分布

	p_n	a	b	p_0
Poisson	$\frac{\lambda^n}{n!}e^{-\lambda}$	0	λ	$e^{-\lambda}$
二项	$\binom{m}{n}p^{m-n}q^n$	$-q/p$	$(m+1)q/p$	p^m
负二项 $q=\beta/(1+\beta)$	$\frac{\Gamma(n+r)}{\Gamma(n+1)\Gamma(r)}p^r q^n$	q	$(r-1)q$	p^r
几何 $q=\beta/(1+\beta)$	pq^n	q	0	p

引理1.1 满足$(a,b,0)$条件的非负整数分布只有：Poisson、二项、几何和负二项分布。

证明：

在(a,b)的平面上进行划分：

$1)\ a+b\le 0 \to impossible$

$2)\ a+b>0, a\ge 1 \to impossible$

$3)\ a+b>0, a<0 \xrightarrow{?} Binomial(b=-a(m+1))$

$4)\ a+b>0, 0<a<1$

$\to (Generalized)NB(b=a(r-1))$

引理1.1 证明（续）

两个边界：

$$a = 0, b > 0 \rightarrow Poisson$$

$$b = 0, 0 < a < 1 \rightarrow Geometric$$

引理1.1的应用-实际数据建模

由定义，有：

$$np_n / p_{n-1} = an + b, \quad n = 1, 2, \ldots$$

进而，对于频数观测n_k，

可以考虑以下的参数估计：

$$kn_k / n_{k-1} = \hat{a}k + \hat{b}, \quad k = 1, 2, \ldots$$

定理1.4：（离散）总损失的递推计算

$$S=\begin{cases} X_1+X_2+\cdots+X_N, & N>0 \\ 0, & N=0 \end{cases}$$

$N \sim (a,b,0)$ X的取值域为 $\{1,2,\ldots\}$

则总损失S的概率函数可以通过以下递推公式计算

$$f_S(0)=p_0,$$

$$f_S(x)=\sum_{y=1}^{x}\left(a+b\frac{y}{x}\right)f_X(y)\,f_S(x-y),\ x=1,2,\ldots$$

定理1.5：（连续）总损失的迭代计算

- 个体损失X为取值于正实数的连续随机变量，则总损失S的密度函数可通过以下迭代计算：

$$f_S(0) = p_0 = \Pr\left(N = 0\right)$$

$$f_S(x) = E\left(X\right) f_X(x) + \int_0^x (a + b\frac{y}{x}) f_X(y) f_S(y - x) dy, x > 0$$

计数变量为负二项分布

- 除Poisson分布外使用较广的计数分布
- 有一些特别的优良性质
  - 混合Poisson分布（Mixture）
  - 复合Poisson分布（compound）

备忘

以下几个分布的概念：**条件分布 无条件分布 混合分布**

- 定义：
  - 条件分布（conditional distribution）
  - 无条件分布（unconditional distribution ）
  - 混合分布（mixture distribution ）
- 例子：

例1 负二项分布：Poisson-Gamma

- 负二项分布是条件Poisson分布与先验Gamma分布的无条件分布。
- 或称之为Poisson-Gamma的混合分布。

$$N \sim NB(r, p)$$

$$\Leftrightarrow \begin{cases} N \mid \Lambda = \lambda \sim Poisson(\lambda) \\ \Lambda \sim Gamma(r, \beta) \end{cases}$$

$$p = \frac{\beta}{1+\beta}$$

混合分布的实际背景

- 假设某组合中的风险个体由一些子类构成，风险特征如下：
  - 在同一个风险子类中“索赔次数”（风险发生频率）服从Poisson分布，平均索赔次数为Poisson参数；
  - 在组合层面，各个风险子类的“平均索赔次数”（随机的）服从Gamma分布。
- 则：风险组合中每个个体的“索赔次数”（不分子类）服从负二项分布。

混合Poisson分布的方差均大于其均值

若N为混合Poisson分布，条件变量为Λ，则有：

$$\begin{cases} E(N)=E(\Lambda) \\ Var(N)=Var(\Lambda)+E(\Lambda) \end{cases}$$

进而有：

$$\frac{Var(N)}{E(N)}=1+\frac{Var(\Lambda)}{E(\Lambda)}\geq 1$$

例2 负二项分布可以表示为CP分布

$$N \sim NB(r,p) \Leftrightarrow N \sim CP(\lambda, f_X(x))$$

$$N = \begin{cases} M_1 + \cdots M_{N_1}, & N_1 > 0 \\ 0, & N_1 = 0 \end{cases}$$

$$N_1 \sim Poisson(\lambda), \ \lambda = r\log(1+\beta), \ \beta = \frac{1-p}{p},$$

$$M_i \sim f_X(x) = \frac{1}{x\log(1+\beta)}\left(\frac{\beta}{1+\beta}\right)^x, \ x = 1, 2, \ldots$$

实际背景：同一个意外事故（accident）产生多次索赔（claim）

- 例如：机动车辆险
- 索赔起因为交通事故，特别是高速公路上的交通事故，事故发生的频率服从Poisson分布；
- 每次事故可能会造成多个索赔案件：
  - 每辆车的多个索赔：车损、第三者责任、意外伤害；
  - 多个车辆的索赔
- 索赔次数随机变量则为混合分布

例如：信用风险

- 数据量有限时，将会把不同质的个体作为同一个观测样本；
- 可以设想：这些个体之间存在一种潜在的分类，每个类的内部是一致的。
- 例如：违约概率－PD的背景是频率，如果Bernoulli变量的参数是随机的，将会导致什么混合分布？

例1-14

索赔次数		0	1	2	3	4	5	总和
观测频数		3719	232	38	7	3	1	4000
统计估计	Poisson	3668.54	317.33	13.72	0.40	0.01	0.00	4000
	负二项	3719.22	229.90	39.91	8.42	1.93	0.46	4000

例（续）

- 第一步统计建模（MLE估计）：
  - 样本基本统计计算：
    - 均值0.0865（8.65%）——平均索赔次数
    - 方差0.122518
  - 用Poisson分布进行拟合：
    - 显然方差大于均值，Poisson分布不适合，从拟合结果看也不理想（请同学对表中的Poisson拟合进行拟合优度检验）。

例（续）

- 第二步：负二项模型拟合
  - MLE估计：$r = 0.216600$ 和 $p = 0.2854$
  - 将负二项分布看做复合Poisson分布：
    - 平均事故次数为0. 27（复合模型的Poisson参数）
    - 每次事故中的索赔次数分布（对数分布）：
      - 84.942%为1次索赔
      - 12.114%为2次索赔
      - 2.304%为3次索赔。

基于前面的分析，复合负二项模型可以表示为

- （需要证明）

$$S \sim CP\left(\lambda, f_Y\left(x\right)\right)$$

$$\lambda = r\log\left(1+\beta\right)$$

$$Y = X_1 + \cdots + X_M$$

$$M \sim \log\left(1+\beta\right)$$

特殊的索赔额分布下的复合分布计算

- 索赔额为指数分布：

$$F_S(x)=1-\sum_{n=1}^{\infty}p_n\sum_{j=0}^{n-1}\frac{(\beta x)^j}{j!}e^{-\beta x}$$

- 特别的，索赔数为二项情形（有限和）

$$F_S(x)=1-\sum_{n=1}^{m}\binom{m}{n}q^n(1-q)^{m-n}\sum_{j=0}^{n-1}\frac{(\beta x)^j}{j!}e^{-\beta x}$$

- 特别的，索赔数为负二项情形：
  - 总索赔为二项分布（参数变换）与指数的复合分布：

$$F_S(x)=1-\sum_{n=1}^{r}\binom{n}{r}\left(\frac{p}{1+p}\right)^{n}\left(\frac{1}{1+p}\right)^{r-n}\sum_{j=0}^{n-1}\frac{\left(\frac{p}{1+p}x\right)^{j}}{j!}e^{-\beta x}$$

$$r=1$$

$$F_S(x)=p+q\left[1-e^{-\beta px}\right]$$

分布的数值化近似

- 将索赔额用算术分布近似
- 算术分布：

$$k_j = \Pr(X = jh) > 0, \;\; j = 0, 1, 2, \ldots,$$

$$\sum_{j=0}^{\infty} k_j = 1$$

对一般的分布转换为算术分布

- 在整数点平滑：考虑三个算术分布：

$$K_h^A(x)=\begin{cases} k_0^A=F_X(h-0), & x=0 \\ k_j^A=F_X(jh+h-0)-F_X(jh-0), & x=jh,\ j=1,2,\ldots \end{cases}$$

$$K_h^C(x)=\begin{cases} k_0^C=0, & x=0 \\ k_j^C=F_X(jh+0)-F_X(jh-h-0), & x=jh,\ j=1,2,\ldots \end{cases}$$

$$K_h^B(x)=\begin{cases} k_0^B=F_X\left(\frac{h}{2}-0\right), & x=0 \\ k_j^B=F_X\left(jh+\frac{h}{2}+0\right)-F_X\left(jh-\frac{h}{2}-0\right), & x=jh,\ j=1,2,\ldots \end{cases}$$

- 显然有：

$$K_h^A(x) \geq F_X(x) \geq K_h^C(x), x \geq 0$$

- 和：

$$K_h^A(x) \geq K_h^B(x) \geq K_h^C(x), x \geq 0$$

- 可以证明，对复合分布也有下面的近似：

$$\sum_{n=0}^{\infty} p_n (K_h^A(x))^{*n} \geq F_S(x) = \sum_{n=0}^{\infty} p_n (F_X(x))^{*n} \geq \sum_{n=0}^{\infty} p_n (K_h^C(x))^{*n}, x \geq 0$$

矩方法：算术分布与原分布的原点矩相同

$$\sum_{j=0}^{p}(x_k+jh)^r m_j^k=\int_{x_k}^{x_k+ph}x^r dF_X(x),\ r=0,1,\ldots,p$$

- 进而有：

$$m_j^k=\int_{x_k}^{x_k+ph}\prod_{i\neq j}\frac{x-x_k-ih}{(j-i)h}dF_X(x),\ j=0,1,\ldots,p$$

作业

证明**定理1-4**