第3章　损失次数分布模型

损失次数是指在一个保险期间内，保单所发生的事故次数或索赔次数。损失次数的取值为非负整数，可以用非负的离散型随机变量进行描述。最常见的损失次数分布模型包括泊松分布、二项分布、负二项分布和几何分布，这些分布的概率计算满足一个相同形式的递推公式，所以称为$(a,b,0)$分布类。

对于某些保险合同，$(a,b,0)$分布类可能无法准确描述它们的损失次数在零点的概率，此时，可以对$(a,b,0)$分布类在零点的概率进行调整，从而得到零截断分布或零调整分布，这些分布的概率计算也满足一个相同形式的递推公式，所以统称为$(a,b,1)$分布类。对零点的概率进行调整的另一种方式是对$(a,b,0)$分布类与零点的退化分布进行混合，从而形成所谓的零膨胀分布。零膨胀分布在零点的概率总是大于$(a,b,0)$分布类在零点的概率。

如果损失次数数据存在较长的右尾，常用的分布类型也将无法很好地拟合数据。为了刻画右尾较长的损失次数数据，可以对常用的分布进行复合或混合，从而生成新的分布，如复合泊松分布和混合泊松分布。

在财产与责任保险合同中，免赔额条款的使用十分普遍。如果保险合同规定了免赔额，则索赔次数将不同于损失次数，即对于小于免赔额的损失，保险公司是不予赔偿的。在这种情况下，需要对损失次数分布模型的参数进行适当调整，从而得到索赔次数的分布模型。

本章首先介绍最基本的损失次数分布模型，即$(a,b,0)$分布类，然后讨论对零点的概率进行调整的$(a,b,1)$分布类和零膨胀分布，再介绍对尾部概率进行调整的复合泊松分布和混合泊松分布，最后讨论免赔额对损失次数模型的影响。

3.1　$(a, b, 0)$分布类

$(a,b,0)$分布类包括泊松分布、二项分布、负二项分布和几何分布，其中几何分布是负二项分布的特例。

3.1.1　泊松分布

泊松分布的概率函数为

$$p_k = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0,1,2,\cdots \tag{3.1}$$

式中，p_k 表示发生 k 次损失的概率，λ 表示泊松分布的参数。

泊松分布的母函数 $P(z)$为

$$P(z) = E(z^N) = \sum_{k=0}^{\infty} p_k z^k = e^{\lambda(z-1)} \tag{3.2}$$

泊松分布的矩母函数 $M(t)$为

$$M(t) = E(e^{tN}) = e^{\lambda(e^t-1)} \tag{3.3}$$

泊松分布的均值、方差和偏度分别为

$$E(N)=\lambda$$
$$\mathrm{Var}(N)=\lambda$$
$$\kappa=\lambda^{-1/2}$$

泊松分布的参数 λ 越大，泊松分布的偏度越小，越接近对称分布，用正态分布近似的效果也越好，如图 3-1 所示，其中竖条表示泊松分布的概率。绘制该图的 R 程序代码如下。

```
x = 0:20
par(mfrow = c(2, 2))
for (lam in c(1, 2, 5, 10)) {
    barplot(dpois(x, lam), names.arg = x, main = paste("lambda = ", lam, sep
 = ""))
}
```

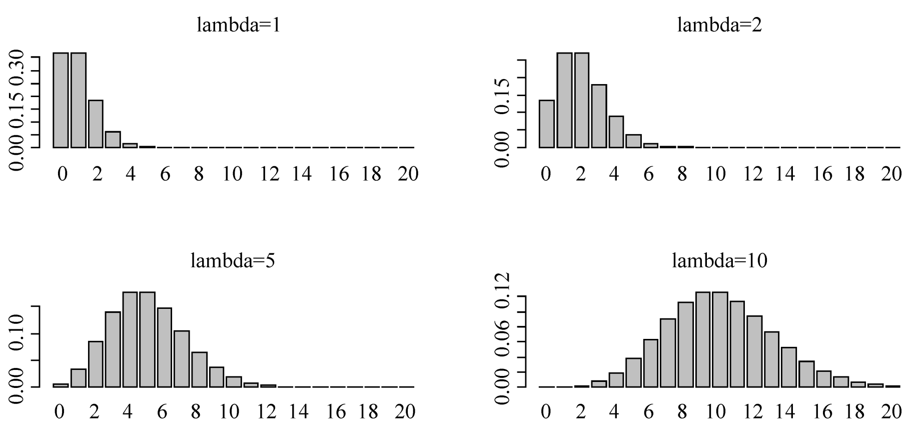

图 3-1　泊松分布的概率函数

【例 3-1】 假设索赔次数 N 服从参数为 $\lambda=2$ 的泊松分布，请计算：

(1) 索赔次数等于 3 的概率 $\Pr(N=3)$；

(2) 索赔次数小于等于 4 的概率 $\Pr(N\leqslant 4)$；

(3) 索赔次数大于等于 3 小于等于 5 的概率 $\Pr(3\leqslant N\leqslant 5)$；

(4) 模拟 20 个索赔次数的观察值。

【解】 求解本例的 R 程序代码如下。

```
#索赔次数等于 3 的概率
dpois(3, lambda = 2)
##[1] 0.180447
```

```
＃索赔次数小于等于 4 的概率为
ppois(4, lambda = 2)
＃＃[1] 0.947347

＃索赔次数大于等于 3 小于等于 5 的概率
ppois(5, 2) - ppois(2, 2)
＃＃[1] 0.30676

＃模拟 20 个索赔次数观察值
set.seed(111)
sim = rpois(20, 2)
sim
＃＃ [1] 2 3 1 2 1 2 0 2 2 0 2 2 0 0 1 2 1 5 1 2

＃对索赔次数的模拟值进行汇总
table(sim)
＃＃sim
＃＃0 1 2 3 5
＃＃4 5 9 1 1
```

【例 3-2】 假设 N_1 和 N_2 相互独立，分别是参数为 λ_1 和 λ_2 的泊松随机变量，证明 $N=N_1+N_2$ 是参数为$(\lambda_1+\lambda_2)$的泊松随机变量。

【解】 泊松分布的母函数为

$$P(z)=\mathrm{e}^{\lambda(z-1)}$$

根据独立性假设，N 的母函数可以表示为

$$\begin{aligned}P_N(z)&=P_{N_1}(z)P_{N_2}(z)\\&=\mathrm{e}^{\lambda_1(z-1)}\mathrm{e}^{\lambda_1(z-1)}\\&=\mathrm{e}^{(\lambda_1+\lambda_2)(z-1)}\\&=\mathrm{e}^{\lambda(z-1)}\end{aligned}$$

上式就是参数为 $\lambda=(\lambda_1+\lambda_2)$的泊松分布的母函数。

【例 3-3】 假设两种保险事故相互独立，它们的总索赔次数 N 服从参数为 λ 的泊松分布，这两种保险事故发生的概率分别为 p_1 和 p_2。证明这两种事故的索赔次数 N_1 和 N_2 相互独立且均服从泊松分布，泊松参数分别为 λp_1 和 λp_2。

【解】 首先证明 N_1 和 N_2 服从参数为 λp_1 和 λp_2 泊松分布。

$$\begin{aligned}\Pr(N_1=n_1)&=\sum_{n=n_1}^{\infty}\Pr(N_1=n_1\mid N=n)\Pr(N=n)\\&=\sum_{n=n_1}^{\infty}\binom{n}{n_1}p_1^{n_1}(1-p_1)^{n-n_1}\frac{\mathrm{e}^{-\lambda}\lambda^n}{n!}\end{aligned}$$

$$= \frac{p_1^{n_1} e^{-\lambda}}{n_1!} \sum_{n=n_1}^{\infty} \frac{(1-p_1)^{n-n_1} \lambda^{n_1+(n-n_1)}}{(n-n_1)!}$$

$$= e^{-\lambda} \frac{(\lambda p_1)^{n_1}}{n_1!} \sum_{n-n_1=0}^{\infty} \frac{[\lambda(1-p_1)]^{n-n_1}}{(n-n_1)!}$$

$$= e^{-\lambda} \frac{(\lambda p_1)^{n_1}}{n_1!} e^{\lambda(1-p_1)}$$

$$= e^{-\lambda p_1} \frac{(\lambda p_1)^{n_1}}{n_1!}$$

上式表明，N_1 服从参数为 λp_1 的泊松分布，同理可以证明 N_2 服从参数为 λp_2 泊松分布。

下面再证明 N_1 与 N_2 相互独立。

$$\begin{aligned}
\Pr(N_1 = n_1, N_2 = n_2) &= \Pr(N_1 = n_1, N_2 = n_2 \mid N = n)\Pr(N = n) \\
&= \frac{n!}{n_1! n_2!} p_1^{n_1} p_2^{n_2} \frac{e^{-\lambda}\lambda^n}{n!} \\
&= \frac{n!}{n_1! n_2!} p_1^{n_1} p_2^{n_2} \frac{e^{-\lambda(p_1+p_2)}\lambda^{(n_1+n_2)}}{n!} \\
&= \frac{e^{-\lambda p_1}(\lambda p_1)^{n_1}}{n_1!} \frac{e^{-\lambda p_2}(\lambda p_2)^{n_2}}{n_2!} \\
&= \Pr(N_1 = n_1)\Pr(N_2 = n_2)
\end{aligned}$$

上式表明，N_1 与 N_2 相互独立。

本例的结论在保险实践中具有重要的应用价值。譬如，假设某险种的索赔次数服从泊松分布，在引入免赔额 d 后，剩余的索赔次数仍将服从泊松分布，只是泊松参数不同而已，即从 λ 变为 $[1-F(d)]\lambda$。又如，假设某险种的索赔次数服从参数为 2 的泊松分布，如果将保险责任减少一项（假设此项责任的索赔次数占总索赔次数的 10%），那么剩余责任的索赔次数仍将服从泊松分布，泊松参数成为 $0.9\times2=1.8$。

3.1.2 二项分布

二项分布的概率表示在 n 次试验中，成功了 k 次的概率，其概率函数为

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0,1,2,\cdots,n \tag{3.4}$$

二项分布的概率母函数为

$$P(z) = (1-p+pz)^n \tag{3.5}$$

二项分布的矩母函数为

$$M(t) = (1-p+pe^t)^n \tag{3.6}$$

二项分布的均值、方差和偏度分别为

$$E(N) = np$$

$$\mathrm{Var}(N) = np(1-p)$$

$$\kappa = \frac{1-2p}{\sqrt{np(1-p)}}$$

二项分布的均值越大，其分布形态越接近对称，如图 3-2 所示。绘制该图的 R 程序代码如下。

```
x = 0:15
par(mfrow = c(2, 2))
for (n in c(1, 5, 10, 20)) {
    barplot(dbinom(x, size = n, prob = 0.3), names.arg = x, xlab = "", ylab =
"", main = paste("n = ", n, ", p = 0.3", sep = ""))
}
```

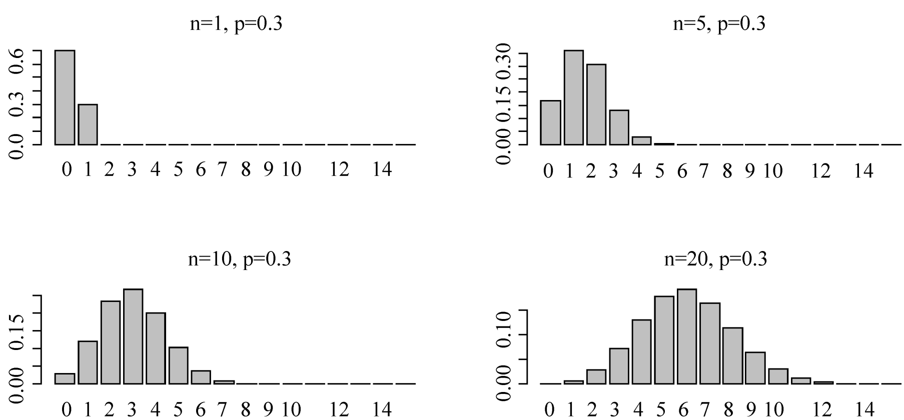

图 3-2　参数 p 给定参数 n 变化时二项分布的概率函数

【例 3-4】　假设损失次数随机变量 N 服从参数为(10,0.2)的二项分布，请计算：

(1) 损失次数等于 3 的概率；

(2) 损失次数大于 3 的概率；

(3) 损失次数小于等于 5 的概率；

(4) 模拟 20 个损失次数观察值。

【解】　求解本例的 R 程序代码如下。

```
#损失次数等于 3 的概率
dbinom(3, 10, 0.2)
```

```
##[1] 0.2013266

#损失次数大于 3 的概率
1 - pbinom(3, 10, 0.2)
##[1] 0.1208739

#损失次数小于等于 5 的概率
pbinom(5, 10, 0.2)
##[1] 0.9936306

#模拟 20 个损失次数观察值
set.seed(111)
sim = rbinom(20, 10, 0.2)
sim
## [1] 2 3 1 2 2 2 0 2 2 0 2 2 0 0 1 2 1 4 1 2

#对损失次数的模拟值进行汇总
table(sim)
##
##0 1 2 3 4
##4 4 10 1 1
```

3.1.3 负二项分布

负二项分布的概率函数为

$$\begin{aligned} p_k &= \binom{k+r-1}{k} p^r (1-p)^k \\ &= \frac{(r+k-1)!}{k!(r-1)!} p^r (1-p)^k, \quad k = 0,1,2,\cdots \end{aligned} \tag{3.7}$$

式中，$r>0$，$0 \leqslant p \leqslant 1$。

负二项分布的均值、方差和偏度分别为

$$E(N) = \frac{r(1-p)}{p}$$

$$\mathrm{Var}(N) = \frac{r(1-p)}{p^2}$$

$$\kappa = \frac{2-p}{\sqrt{r(1-p)}}$$

负二项分布的另一种参数形式是令 $p=1/(1+\beta)$，此时，负二项分布的概率函数可以表示为

$$p_k = \frac{(r+k-1)!}{k!(r-1)!}\left(\frac{1}{1+\beta}\right)^r\left(\frac{\beta}{1+\beta}\right)^k \tag{3.8}$$

负二项分布的母函数可以表示为

$$P(z) = (1+\beta-z\beta)^{-r} \tag{3.9}$$

即

$$\begin{aligned} P(z) &= \sum_{k=0}^{\infty} p_k z^k = \sum_{k=0}^{\infty}\left[\frac{\Gamma(k+r)}{\Gamma(r)\Gamma(k+1)}\left(\frac{1}{1+\beta}\right)^r\left(\frac{\beta}{1+\beta}\right)^k\right]z^k \\ &= \sum_{k=0}^{\infty}\left[\frac{\Gamma(k+r)}{\Gamma(r)\Gamma(k+1)}\left(\frac{1+\beta-z\beta}{1+\beta}\right)^r\left(\frac{z\beta}{1+\beta}\right)^k\right](1+\beta-z\beta)^{-r} \\ &= (1+\beta-z\beta)^{-r} \end{aligned}$$

相应地，负二项分布的均值和方差分别为

$$E(N) = r\beta$$

$$\mathrm{Var}(N) = r\beta(1+\beta)$$

负二项分布的形状受两个参数的共同影响。在给定参数 p 的条件下，参数 r 的取值越大，负二项分布越接近对称，如图 3-3 所示，绘图的 R 程序代码如下。

```
#负二项分布的概率函数随着参数 r 的变化
x = 0:20
par(mfrow = c(2, 2))
for (r in c(1, 2, 3, 4)) {
    barplot(dnbinom(x, r, 0.3), names.arg = x, xlab = "", ylab = "", main = paste
("r = ", r, ", p = 0.3", sep = ""))
}
```

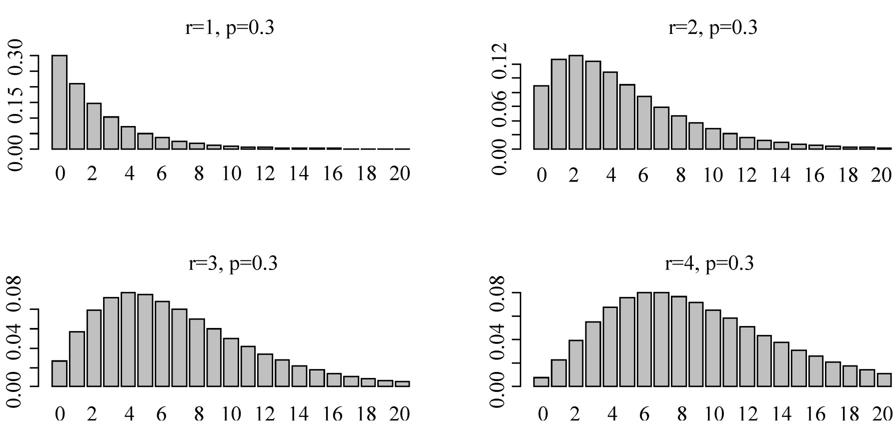

图 3-3　参数 p 给定条件下负二项分布的概率函数

在给定参数 r 的条件下，参数 p 的取值越大，负二项分布的尾部越短，如图 3-4 所示，绘图的 R 程序代码如下。

```
#负二项分布的概率函数随着参数 p 的变化
x = 0:30
par(mfrow = c(2, 2))
for (p in c(2, 3, 4, 5)/10) {
    barplot(dnbinom(x, 3, p), names.arg = x, ylim = c(0, 0.2), xlab = "", ylab =
"", main = paste("r = 3, p = ", p, sep = ""))
}
```

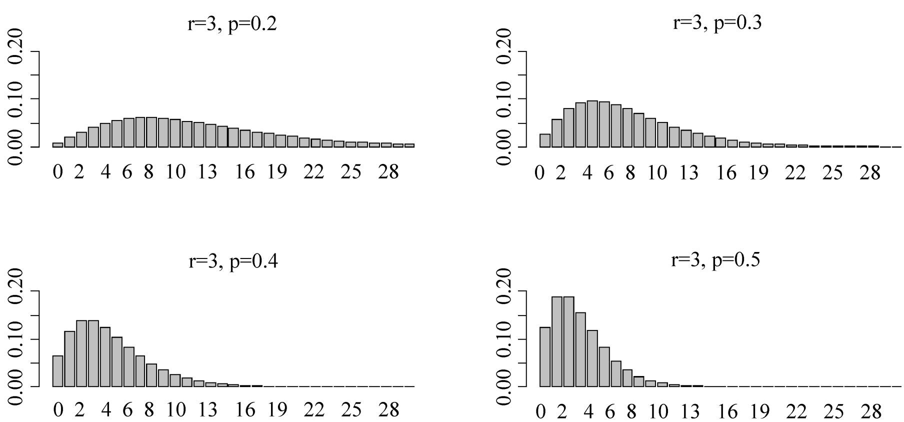

图 3-4　参数 r 给定条件下负二项分布的概率函数

【例 3-5】　假设损失次数随机变量 N 服从参数为($r=5$，$p=0.6$)的负二项分布，请计算：

(1) 损失次数等于 3 的概率；

(2) 损失次数小于 3 的概率；

(3) 损失次数小于等于 5 的概率；

(4) 模拟 20 个损失次数观察值。

【解】　求解本例的 R 程序代码如下。

```
#损失次数等于 3 的概率
dnbinom(3, 5, 0.6)
##[1] 0.1741824

#损失次数大于 3 的概率
1 - pnbinom(3, 5, 0.6)
```

```
##[1] 0.4059136

#损失次数小于等于5的概率
pnbinom(5, 5, 0.6)
##[1] 0.8337614

#模拟20个损失次数的观察值
sim = rnbinom(20, 5, 0.6)
sim
## [1] 1 2 4 3 2 7 3 4 6 2 2 3 4 2 2 1 2 1 4 3
```

3.1.4 几何分布

如果负二项分布的参数 $r=1$，则负二项分布就退化为几何分布，相应地，概率函数就简化为

$$p_k = (1-p)^k p, \quad k = 0,1,2,\cdots \tag{3.10}$$

几何分布的概率母函数为

$$P(z) = \frac{p}{1-(1-p)z} \tag{3.11}$$

几何分布的矩母函数为

$$M(t) = \frac{p}{1-(1-p)\mathrm{e}^t} \tag{3.12}$$

几何分布的均值、方差和偏度分别为

$$E(N) = \frac{(1-p)}{p}$$

$$\mathrm{Var}(N) = \frac{(1-p)}{p^2}$$

$$\kappa = \frac{2-p}{\sqrt{(1-p)}}$$

几何分布只有一个参数 p，该参数在(0，1)区间取值，参数值越大，几何分布的尾部越短，如图 3-5 所示，绘制该图的 R 程序代码如下。

```
x = 0:20
par(mfrow = c(2, 2))
for (p in c(1, 2, 3, 4)/10) {
    barplot(dgeom(x, p), names.arg = x, xlab = "", ylab = "", main = paste("p = ",
p, sep = ""))
}
```