第二章 理赔次数的分布

主要内容：

1、母函数与矩母函数

2、一张保单的理赔次数分布

3、理赔次数的混合分布

4、理赔次数的复合分布

5、免赔额对理赔次数分布的影响

6、保单数目与保单组合总理赔次数

一、母函数与矩母函数

设 N 是一个离散随机变量，取值于 $0,1,2\cdots$

分布列：

$$p_k = P(N = k) \qquad k = 0\,,1\cdots$$

母函数：

$$P_N(z) = E(z^N) = \sum_{k=0}^{\infty} p_k z^k$$

矩母函数

$$M_N(z) = E(e^{zN}) = \sum_{k=0}^{\infty} e^{zk} p_k\,, \quad M_N(z) = P_N(e^z)$$

母(矩母)函数性质

1、若 N 的母(矩母)函数存在，那么母(矩母)函数与分布函数是相互唯一决定的。

2、由母(矩母)函数可以导出矩的计算：

$$P'(1)=\sum_{k=1}^{\infty}kp_k=E(N)$$

$$P''(1)=\sum_{k=2}^{\infty}k(k-1)p_k=E(N(N-1))$$
$$=E(N^2)-E(N)$$

$$Var(N)=E(N^2)-E(N)^2$$
$$=P''(1)+P'(1)-(P'(1))^2$$

请问：

$$M_N'(0)=?$$

$$M_N''(0)=?$$

3、设 $N=N_1+N_2+\cdots+N_n$， $N_1,\cdots,N_n$ 相互独立，则

$$P_N(z)=\prod_{j=1}^{n}P_{N_j}(z)$$

$$M_N(z)=\prod_{j=1}^{n}M_{N_j}(z)$$

二、一张保单的理赔次数分布

1、泊松分布(Poisson)

对于保险公司而言，客户因发生损失而提出理赔的人数类似于等待服务现象，因此对大多数险种来说，个别保单的理赔次数可用泊松分布来表示，即在单位时间内个别保单发生理赔次数 N 的分布列为：

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

令 $t=1$，则在单位时间内理赔次数 N 的分布列为：

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

泊松分布的性质：

（1）、均值和方差

$$E(N)=Var(N)=\lambda$$

（2）、母函数

$$P_N(z)=\sum_{k=o}^{\infty} z^k e^{-\lambda}\frac{\lambda^k}{k!}=\sum_{k=o}^{\infty} e^{-\lambda}\frac{(\lambda z)^k}{k!}=\exp(\lambda(z-1))$$

（3）、矩母函数

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

（4）、可加性

定理： 设 $N_1, N_2, \cdots, N_n$ 是相互独立的泊松随机变量，参数分别为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，则 $N = N_1 + N_2 + \cdots + N_n$ 服从泊松分布，参数为 $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$。

证明：

$$P_N(z) = \prod_{i=1}^{n} P_{N_i}(z) = \prod_{i=1}^{n} \exp(\lambda_i (z-1)) = \exp(\sum_{i=1}^{n} \lambda_i (z-1))$$

故 N 服从泊松分布，参数 $\lambda = \lambda_1 + \lambda_2 + \cdots + \lambda_n$。

（5）、可分解性

假设损失事故可以分为m个不同类型$C_1, C_2, \cdots C_m$，

E_i表示第i类事故发生。

$p_i = P(E_i)$表示第i类事故发生的概率，$i = 1, 2 \cdots m$

N_i表示第i类事故发生的次数，$i = 1, 2 \cdots m$

$N = N_1 + N_2 + \cdots + N_m$表示所有事故发生的次数。

定理：若 N 服从参数为 λ 的泊松分布，则 $N_1, N_2, \cdots N_m$ 都是相互独立的，且服从泊松分布，参数分别是 λp_i，$i=1,2\cdots,m$。

证明：给定 $N=n$，$N_i \mid N=n$ 服从二项分布 $B(n, p_i)$，$(N_1, N_2, \cdots N_m)$ 服从多项分布 $B(n, p_1, \cdots p_m)$。

因此，

$$
\begin{aligned}
& P(N_1=n_1, \cdots, N_m=n_m) \\
& = P(N_1=n_1, \cdots, N_m=n_m \mid N=n) P(N=n) \\
& = \frac{n!}{n_1! n_2! \cdots n_m!} p_1^{n_1} \cdots p_m^{n_m} \frac{e^{-\lambda} \lambda^n}{n!} \\
& = \prod_{j=1}^{m} e^{-\lambda p_j} \frac{(\lambda p_j)^{n_j}}{n_j!}
\end{aligned}
$$

其中， $n = n_1 + n_2 + \cdots + n_m$

$$P(N_j = n_j) = \sum_{n=n_j}^{\infty} P(N_j = n_j \mid N = n) P(N = n)$$

$$= \sum_{n=n_j}^{\infty} C_n^{n_j} p_j^{n_j} (1 - p_j)^{n-n_j} \frac{e^{-\lambda} \lambda^n}{n!}$$

$$= e^{-\lambda} \frac{(\lambda p_j)^{n_j}}{n_j!} e^{\lambda(1-p_j)}$$

$$= e^{-\lambda p_j} \frac{(\lambda p_j)^{n_j}}{n_j!}$$

因此， $N = (N_1, N_2, \cdots, N_m)$ 的联合分布等于 $N_1, N_2, \cdots, N_m$ 分布的乘积， $N_1, N_2, \cdots, N_m$ 是相互独立的随机变量。

例 2.1：设 N 表示损失事故发生的次数，X 表示损失额，N 服从泊松分布，$\lambda = 10$，$X \sim U[0,20]$。问损失额超过 5 的事故发生次数的概率分布。

解：令 E 表示事件"损失额超过 5"

$$P(E) = \int_{5}^{20} \frac{1}{20} dx = 0.75$$

所以损失额超过 5 的次数服从参数为 $10 \times 0.75 = 7.5$ 的泊松分布。

例 2.2：假设某险种的个体保单损失 X 的分布为

$$f_X(1)=0.40, f_X(2)=0.35, f_X(3)=0.25$$

又假设个体保单在一年内发生的损失事件的次数 N 服从泊松分布，$\lambda=200$。N_i 表示损失额为 i 的损失事件的次数。

（1）求 N_1, N_2, N_3 的分布。

解

（1）由于$N=N_1+N_2+N_3$，且 N 服从泊松分布，由定理知，

N_1, N_2, N_3相互独立且服从泊松分布。

参数λ_i等于

$$\lambda_i=\lambda P(X=i)=200P(X=i)$$

计算得到$\lambda_1=80; \lambda_2=70; \lambda_3=50$

练习：假设免赔额为 1，求个体保单在一年内发生的理赔事件次数的分布。

（2）当免赔额为 1 时，赔偿事件为损失额等于 2 或 3 的损失事件。发生的赔偿事件的次数等于 N_1+N_2，服从参数为 $\lambda_2+\lambda_3=120$ 的分布。

2、其他常见的理赔次数分布

（1）负二项分布（negative binomial distribution）

$$p_k = P(N=k) = \binom{k+r-1}{k}\left(\frac{\beta}{1+\beta}\right)^k \left(\frac{1}{1+\beta}\right)^r，\quad r>0, \beta>0$$

其中$\binom{x}{k} = \dfrac{x(x-1)\cdots(x-k+1)}{k!}$

注：

$r=1 \quad p_k = \left(\frac{\beta}{1+\beta}\right)^k \left(\frac{1}{1+\beta}\right)$，为几何分布 (Geometric)

令 $p=\dfrac{1}{1+\beta}, q=\dfrac{\beta}{1+\beta}$，负二项分布也可以写为：

$$p_k = P(N=k) = \binom{k+r-1}{k} p^r q^k$$

背景：贝努利试验系列中第r次成功正好出现在第$r+k$次试验上的概率。k为第r次成功前失败的次数。p为成功概率

母函数：

$$
\begin{aligned}
P_N(z) &= \sum_{k=0}^{\infty}\binom{k+r-1}{k}q^k(1-q)^r z^k \\
&= \sum_{k=0}^{\infty}\binom{k+r-1}{k}(qz)^k(1-qz)^r\left(\frac{1-q}{1-qz}\right)^r \\
&= \left(\frac{1-q}{1-qz}\right)^r
\end{aligned}
$$

将$\left(\frac{1-q}{1-qz}\right)^r$化简得到，

$$P_N(z) = (1-\beta(z-1))^{-r}$$

均值和方差

$$E(N) = r\beta \quad \text{或者} \quad E(N) = \frac{rq}{p}$$

$$Var(N) = r\beta(1+\beta) \quad \text{或者} \quad Var(N) = \frac{rq}{p^2}$$

$$E(N) < Var(N)$$

注意：我们这里的负二项是广义的负二项分布，r 可以为非整数。

（2）二项式分布

$$p_k = P(N=k) = \binom{m}{k}(q)^k(1-q)^{m-k} \qquad 0<q<1 \qquad k=0,1,\cdots,m$$

背景：m 次贝努利试验中成功的次数。

m 个死亡率相同的投保人，q 表示死亡概率

母函数与矩母函数

$$P_N(z)=\sum_{k=0}^{\infty} z^k p_k=\sum_{k=0}^{\infty}\binom{m}{k}(qz)^k(1-q)^{m-k}$$
$$=(qz+(1-q))^m$$
$$=(1+q(z-1))^m$$

$$M_N(z)=(1+q(e^z-1))^m$$

均值与方差

$$E(N)=P_N{}'(z)\,|_{z=1}=mq(1+q(z-1))^{m-1}\,|_{z=1}=mq$$

$$Var(N)=mq(1-q)$$

$$E(N)<Var(N)$$

例 2.3：设有 100 个 40 岁的投保人投保生命险，q 表示一个投保人明年死亡的概率，问明年死亡人数的分布是什么？

3、(a, b, 0)分布族

上述 3 种分布都可以用$(a, b, 0)$分布来表示

定义： 设随机变量 N 的分布列$\{p_k\}$满足

（1） $p_k \geq 0$ 且 $p_0 \neq 0$， $\sum_{k=0}^{\infty} p_k = 1$,

（2） $\dfrac{p_k}{p_{k-1}} = a + \dfrac{b}{k}$ $\quad k = 1，2\cdots$

则称分布族$\{p_k, k = 1, 2\cdots\}$为$(a, b, 0)$分布族。

注：泊松分布，二项分布，负二项分布是$(a, b, 0)$分布族。

（1）、泊松分布：

$$\frac{p_k}{p_{k-1}}=\frac{e^{-\lambda}\dfrac{(\lambda)^k}{k!}}{e^{-\lambda}\dfrac{(\lambda)^{k-1}}{(k-1)!}}=\frac{\lambda}{k}$$

因此，泊松分布属于(a, b, 0)分布族，

$$a=0,\ \ b=\lambda \quad p_0=e^{-\lambda}$$

泊松分布是唯一使得$a=0$的分布。

（2）、负二项分布：

$$\frac{p_k}{p_{k-1}} = \frac{\binom{k+r-1}{k}(\frac{\beta}{1+\beta})^k(\frac{1}{1+\beta})^r}{\binom{k+r-2}{k-1}(\frac{\beta}{1+\beta})^{k-1}(\frac{1}{1+\beta})^r}$$

因为

$$\frac{\binom{k+r-1}{k}}{\binom{k+r-2}{k-1}} = \frac{(k+r-1)(k+r-2)\cdots r}{k!} \cdot \frac{(k-1)!}{(k+r-2)\cdots r}$$

$$= \frac{k+r-1}{k}$$

$$\frac{p_k}{p_{k-1}} = \frac{\beta}{1+\beta} \times \frac{k+r-1}{k}$$

$$= \frac{\beta}{1+\beta} + \frac{\beta(r-1)}{1+\beta}\frac{1}{k}$$

因此，负二项式分布属于(a, b, 0)分布族

$$a = \frac{\beta}{1+\beta} > 0 \qquad b = \frac{\beta(r-1)}{1+\beta} \qquad p_0 = (\frac{1}{1+\beta})^r$$

当$r=1$时，

$$\frac{p_k}{p_{k-1}} = \frac{\beta}{1+\beta}$$

$$a=\frac{\beta}{1+\beta} \qquad b=0 \qquad p_0=\frac{1}{1+\beta}=(1-q)$$

几何分布是唯一使得$b=0$的分布。

（3）、二项分布：

$$\frac{p_k}{p_{k-1}}=\frac{\binom{n}{k}(p)^k(q)^{n-k}}{\binom{n}{k-1}(p)^{k-1}(q)^{n-k+1}}$$

$$=\frac{n(n-1)(n-2)\cdots(n-k+1)}{k!}\cdot\frac{(k-1)!}{n(n-1)(n-2)\cdots(n-k)}\cdot\frac{p}{q}$$

$$=\frac{n-k+1}{k}\frac{p}{q}$$

$$=-\frac{p}{q}+\frac{(n+1)p}{q}\frac{1}{k}$$

因此，二项分布属于(a, b, 0)分布族，

$$a = -\frac{p}{q} \qquad b = \frac{(n+1)p}{q}, \quad p_0 = q^n$$

问题：如何简单的区别泊松、负二项和二项分布？

例 2.4：设 N 是一随机变量，令 $p_k = P(N = k)$，如果

$$\frac{p_k}{p_{k-1}} = -\frac{1}{3} + \frac{4}{k}, \quad k=1,2,3,\ldots$$

问 N 的分布是什么？

解：$-\frac{1}{3}<0$，N 服从二项式分布，$\frac{p}{q}=\frac{1}{3}$，$\frac{(n+1)p}{q}=4$，解出 $n=11, p=\frac{1}{4}, q=\frac{3}{4}$。

练习：设 X 的分布属于 $(a,b,0)$ class，已知

$$P(X=0)=P(X=1)=0.25$$

$$P(X=2)=0.1875$$

求 $P(X=3)$？

解： $\dfrac{p_1}{p_0}=a+\dfrac{b}{1}=1$

$$\frac{p_2}{p_1}=a+\frac{b}{2}=\frac{0.1875}{0.25}=0.75$$

解得 $a=0.5,\ \ b=0.5$

因此

$$p_3=(0.5+\frac{0.5}{3})p_2=0.125$$

25. The distribution of accidents for 84 randomly selected policies is as follows:

Number of Accidents	Number of Policies
0	32
1	26
2	12
3	7
4	4
5	2
6	1
Total	84

Which of the following models best represents these data?

(A) Negative binomial

(B) Discrete uniform

(C) Poisson

(D) Binomial

(E) Either Poisson or Binomial

Question #25
Key: A

k	kn_k / n_{k-1}
0	
1	0.81
2	0.92
3	1.75
4	2.29
5	2.50
6	3.00

Positive slope implies that the negative binomial distribution is a good choice. Alternatively, the sample mean and variance are 1.2262 and 1.9131 respectively. With the variance substantially exceeding the mean, the negative binomial model is again supported.

总结：如何简单的区别泊松、负二项和二项分布？

研究性自主学习内容

- 如何产生上述三种分布的随机数

例子

Excle：BINOM.INV(n,p,RAND())

R：rbinom(N,n,p)

```
> no_of_simul = 10
> no_of_risks = 5
> p = 0.3 # probability of having a claim
> rbinom(no_of_simul,no_of_risks,p) # no of claims in each simulation
```

4、$(a,b,1)$ 分布族

定义：随机变量 N 的分布列满足 ：

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \qquad k = 2,3,\cdots$$

$(a,b,1)$ class 可以分为两类，

（1） $p_0 = 0$， 则称为截断的 ZT （Z—truncated）

（2） $p_0 > 0$，则称为 ZM (Z—modified)

记号：

$(a,b,0)$ $class$: $\quad p_k = P(N = k)$， $\quad P(z) = \sum_{k=0}^{\infty} p_k z^k$

$ZT-(a,b,0)$： $p_k^T = P(N=k)$ ， $P^T(z)=\sum_{k=0}^{\infty} p_k^T z^k$

$ZM-(a,b,0)$： $p_k^M = P(N=k)$， $P^M(z)=\sum_{k=0}^{\infty} p_k^M z^k$

三类分布的关系：

1、 $ZM-(a,b,0)$ 与 $(a,b,0)$ 分布的关系：

$$p_k^M = cp_k = \frac{1-p_0^M}{1-p_0} p_k, k=1,2,\cdots$$

$$P^M(z) = (1-\frac{1-p_0^M}{1-p_0}) + \frac{1-p_0^M}{1-p_0} P(z)$$

证明：由于 $p_k^M = cp_k, k = 1, 2, 3, \cdots$

$$P^M(z) = \sum_{k=1}^{\infty} p_k^M z^k + p_0^M$$

$$= p_0^M + c\sum_{k=1}^{\infty} p_k z^k$$

$$= p_0^M + c(P(z) - p_0)$$

由于 $P^M(1) = P(1) = 1$，

$$1 = p_0^M + c(1 - p_0) \Rightarrow c = \frac{1 - p_0^M}{1 - p_0}$$

$$p_k^M = cp_k = \frac{1 - p_0^M}{1 - p_0} p_k, k = 1, 2, \cdots$$

$$P^M(z) = p_0^M + \frac{1-p_0^M}{1-p_0}(P(z)-p_0)$$

$$= (1-\frac{1-p_0^M}{1-p_0}) + \frac{1-p_0^M}{1-p_0}P(z)$$

ZM 分布可以看作一个$(a,b,0)$族分布和一个退化分布的混合。

$$E^M(N) = \frac{1-p_0^M}{1-p_0}E(N)$$

2、$ZT-(a,b,0)$分布与$(a,b,0)$族分布的关系。

$$p_k^T = \frac{1}{1-p_0}p_k$$

$$P^T(z) = \frac{1}{1-p_0}(P(z)-p_0)$$

证明：

令 $p_0^M = 0$，即得到

$$P^T(z) = (1 - \frac{1}{1-p_0}) + \frac{1}{1-p_0} P(z)$$

$$= \frac{1}{1-p_0}(P(z) - p_0)$$

$$p_k^T = \frac{1}{1-p_0} p_k$$

$$E^T(N) = \frac{1}{1-p_0} E(N)$$

3、ZT 和 ZM 的关系

$$p_k^M = cp_k = \frac{1-p_0^M}{1-p_0} p_k, k=1,2,\cdots;$$

$$p_k^T = \frac{1}{1-p_0} p_k$$

$$p_k^M = (1-p_0^M) p_k^T$$

$$P_k^M(z) = p_0^M + (1-p_0^M)P^T(z)$$ (请同学们推导一遍)

例 2.5：设 N 服从负二项分布，$\beta = 0.5, r = 2.5$，求 $p_k, p_k^T, p_k^M, k = 1,2,3,$，其中 $p_0^M = 0.6$。

解：由于负二项式分布属于(a, b, 0)分布，其中

$$p_0=(\frac{1}{1+\beta})^{\gamma}=(1+0.5)^{-2.5}=0.362887$$

$$a=\frac{\beta}{1+\beta}=\frac{0.5}{1.5}=\frac{1}{3}$$

$$b=\frac{\beta(\gamma-1)}{1+\beta}=\frac{(2.5-1)(0.5)}{(1.5)}=0.5$$

于是计算得到

$$p_1=p_0(a+\frac{b}{1})=0.302406$$

$$p_2=p_1(a+\frac{b}{2})=0.176404$$

$$p_3=0.176404(\frac{1}{3}+\frac{1}{2}\frac{1}{2})=0.088202$$

$$p_1^T = \frac{p_2}{1-p_0} = \frac{0.302406}{1-0.362887} = 0.474651$$

$$p_2^T = 0.474651(a+\frac{b}{2}) = 0.276680$$

$$p_3^T = 0.276680(a+\frac{b}{3}) = 0.138440$$

由 $p_0^M = 0.6$， $p_k^M = \dfrac{1-p_0^M}{1-p_0} p_k$ 知道，

$$p_1^M = \frac{1-0.6}{1-0.362887} 0.302406 = 0.189860$$

$$p_2^M = \frac{1-0.6}{1-0.362887} p_2 = p_1^M (a+\frac{b}{2}) = 0.110752$$

例2.6：Given

(i) p_k denotes the probability that the number of claims equals k for $k = 0,1,2, \ldots$

(ii) $$\frac{p_n}{p_m} = \frac{m!}{n!}$$

Using the corresponding zero-modified claim count distribution with $p_0^M = 0.1$， calculate p_1^M

解：由 $\frac{p_n}{p_m}=\frac{m!}{n!}$ 知，$\frac{p_n}{p_{n-1}}=\frac{(n-1)!}{n!}=0+\frac{1}{n}$，

所以$\{p_k\}$是一个(a, b, 0)分布，a＝0，　b＝1，

这是一个 ZM 泊松分布 $\lambda=1$，$p_0=e^{-1}=0.368=p_1$

由公式计算得到

$$p_1^M=\frac{1-p_0^M}{1-p_0}p_1=\frac{0.9}{0.632}0.368=0.52$$

练习,样题 166

关于（a, b, 0）分布的性质，请同学们参看 loss model 中第 4.6.5 和 4.6.6 节。