第二节 集体风险模型

假定单位时间内保单组合理赔的次数是一个随机变量，我们记为N， $X_1, X_2, \cdots, X_N$表示按次序到来的的理赔，设 S 表示单位时间内的总理赔额，N表示单位时间内的理赔次数，集体风险模型可以描述为

$$S = X_1 + X_2 + \cdots + X_N \qquad (**)$$

假定

（1） $X_1, X_2, \cdots, X_n, \cdots$ 是独立同分布的随机变量；

（2） N 与$\{X_i\}$独立。

思考：个体风险模型与集体风险模型两者之间的关系。

一、S 的数字特征

$$E(S)=E(E(S\mid N))=E(NE(X))=E(X)E(N) \qquad (1)$$

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

例 3.9：设理赔次数 N 服从负二项分布，

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

已知参数 $\beta=2$，$Var(N)=24$，个别理赔额 $X\sim\begin{pmatrix}2 & 3 & 4\\ 0.1 & 0.4 & 0.5\end{pmatrix}$，求总理赔额 S 的均值与方差之和。

解：由负二项分布的性质，$Var(N)=r\beta(1+\beta)$，$E(N)=r\beta$，所以

$$Var(N)=E(N)(1+\beta)=24$$

$$E(N)=24\times\frac{1}{3}=8$$

又
$$E(X)=2\times0.1+3\times0.4+4\times0.5=3.4$$

$$Var(X)=2^2\times0.1+3^2\times0.4+4^2\times0.5-3.4^2=0.44$$

因此

$$\begin{aligned}E(S)+Var(S)&=E(X)E(N)+Var(X)E(N)+E(X)^2Var(N)\\&=8\times3.4+3.4^2\times24+8\times0.44\\&=308.16\end{aligned}$$

85. Computer maintenance costs for a department are modeled as follows:

(i) The distribution of the number of maintenance calls each machine will need in a year is Poisson with mean 3.

(ii) The cost for a maintenance call has mean 80 and standard deviation 200.

(iii) The number of maintenance calls and the costs of the maintenance calls are all mutually independent.

The department must buy a maintenance contract to cover repairs if there is at least a 10% probability that aggregate maintenance costs in a given year will exceed 120% of the expected costs.

Using the normal approximation for the distribution of the aggregate maintenance costs, calculate the minimum number of computers needed to avoid purchasing a maintenance contract.

Let N = number of computers in department
Let X = cost of a maintenance call
Let S = aggregate cost

$$\text{Var}(X) = [\text{Standard Deviation}(X)]^2 = 200^2 = 40,000$$

$$E(X^2) = \text{Var}(X) + [E(X)]^2$$

$$= 40,000 + 80^2 = 46,400$$

$$E(S) = N\lambda E(X) = N(3)(80) = 240N$$

$$\text{Var}(S) = N\lambda \times E(X^2) = N(3)(46,400) = 139,200N$$

We want

$$0.1 \geq \Pr\left(S > 1.2E(S)\right)$$

$$\geq \Pr\left(\frac{S - E(S)}{\sqrt{139,200N}} > \frac{0.2E(S)}{\sqrt{139,200N}}\right) \Rightarrow \frac{0.2(240)N}{373.1\sqrt{N}} \geq 1.282 = \Phi(0.9)$$

$$N \geq \left(\frac{1.282(373.1)}{48}\right)^2 = 99.3$$

下面的例子是 N 与 X 不独立的例子

53. You are given:

Number of Claims	Probability	Claim Size	Probability
0	1/5		
1	3/5	25	1/3
		150	2/3
2	1/5	50	2/3
		200	1/3

Claim sizes are independent.

Calculate the variance of the aggregate loss.

First obtain the distribution of aggregate losses:

Value	Probability
0	1/5
25	(3/5)(1/3) = 1/5
100	(1/5)(2/3)(2/3) = 4/45
150	(3/5)(2/3) = 2/5
250	(1/5)(2)(2/3)(1/3) = 4/45
400	(1/5)(1/3)(1/3) = 1/45

$$\mu = (1/5)(0) + (1/5)(25) + (4/45)(100) + (2/5)(150) + (4/45)(250) + (1/45)(400) = 105$$

$$\sigma^2 = (1/5)(0^2) + (1/5)(25) + (4/45)(100^2) + (2/5)(150^2) + (4/45)(250^2)$$
$$+(1/45)(400^2) - 105^2 = 8100$$

二、S 的分布

方法一：直接收集 S 的信息，如收集一个会计年度内每月的总赔付额，然后使用建模方法去拟合 S 的分布。

方法二：首先分开研究个体理赔额 X 和理赔次数 N 的分布的信息，再利用复合分布公式来研究 S 的分布。本章主要讨论方法二，常用的方法有：

（一）卷积法

设 X 的分布函数为 $F_X(x)$，N 的分布列为 $\{p_n\}$。从公式(**)可以看出，S 的分布是一个复合分布。由第三章的复合分布公式有

$$F_S(x)=\sum_{n=0}^{\infty}P(X_1+X_2+\cdots+X_N\mid N=n)P(N=n)$$
$$=\sum_{n=0}^{\infty}F^{*n}(x)p_n \qquad （3）$$

将分布函数(3)两边求导得到分布密度为

$$f(x)=\sum_{n=0}^{\infty}f^{*n}(x)P(N=n)=\sum_{n=0}^{\infty}f^{*n}(x)p_n \qquad （4）$$

其中 $F^{*n}(x)$， $f^{*n}(x)$ 为 F 的 n 重卷积，它的定义为

$$F_X^0(x)=\begin{cases}0 & x<0\\ 1 & x\geq 0\end{cases}$$

和

$$F_X^{*n}(x)=\int_{-\infty}^{\infty}F_X^{*(n-1)}(x-y)dF_X(y) \qquad （5）$$

由于 X 是非负随机变量，因此

$$F_X^{*n}(x)=\int_0^x F_X^{*(n-1)}(x-y)dF_X(y)$$

当理赔额的分布是离散型分布时，n 重卷积变为

$$F^{*n}(x)=\sum_{y=0}^{x}F_X{}^{*(n-1)}(x-y)f_X(y)$$

$$f^{*n}(x)=\sum_{y=0}^{x}f_X{}^{*(n-1)}(x-y)f_X(y)$$

例 3.10：假设有一组保单组合，在单位时间内可能发生的理赔次数为 0，1，2 和 3，相应的概率为 $0.1, 0.3, 0.4$ 和 0.2，可能产生的理赔额为 1，2，3，相应的概率为 $0.5, 0.4$ 和 0.1，试计算理赔总量 S 的概率分布。

解： 设 N 表示理赔次数，C 表示理赔额，则

$$N \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ 0.1 & 0.3 & 0.4 & 0.2 \end{pmatrix} \qquad C \sim \begin{pmatrix} 1 & 2 & 3 \\ 0.5 & 0.4 & 0.1 \end{pmatrix}$$

因此，由公式（3）

$$f_S(x)=\sum_{n=0}^{3}P(N=n)f^{*n}(x)$$
$$=0.1\times f^{*0}(x)+0.3\times f^{*1}(x)+0.4\times f^{*2}(x)+0.2\times f^{*3}(x)$$

$$f^{*n}(x)=\sum_{y=0}^{x}f^{*(n-1)}(x-y)f(y)$$

其中 S 的取值范围是 $X=0,1,2\cdots 9$。具体计算结果见下表。

x	$f^{*0}(x)$	$f^{*1}(x)$	$f^{*2}(x)$	$f^{*3}(x)$	$f_S(x)$	$F_S(x)$
0	1.0	—	—		0.1	0.1
1		0.5	—		0.15	0.25
2		0.4	0.25		?	0.47

3		0.1	0.40	?	0.215	0.685
4			?	?	0.164	0.849
5			0.08	0.315	0.095	0.944
6			0.01	0.184	0.0408	0.9848
7				0.063	0.0126	0.9974
8				0.012	0.0024	0.9998
9				0.001	0.0002	1

例如：

$f^{*2}(2) = P(X_1 = 1, X_2 = 1) = 0.5 * 0.5 = 0.25$，

$$f^{*3}(5)=P(X_1+X_2=2,X_3=3)+P(X_1+X_2=3,X_3=2)$$
$$+P(X_1+X_2=4,X_3=1)$$
$$=0.4\times0.4+0.1\times0.25+0.26\times0.5=0.315$$

$$f_S(1)=0.1\times f^{*0}(1)+0.3\times f^{*1}(1)+0.4\times f^{*2}(1)+0.2\times f^{*3}(1)$$
$$=0+0.3\times0.5+0+0=0.15$$

练习：计算 $f^{*2}(4)$， $f^{*3}(3)$， $f^{*3}(4)$， $f_S(2)$。

答案 $f^{*2}(4)=0.26$， $f^{*3}(3)=0.125$， $f^{*3}(4)=0.3$， $f_S(2)=0.22$。

练习 Exam C Sample questions 2015 第 95 题

假设某保单组合在一年内的理赔次数N服从均值为4的几何分布，每次理赔额X的分布为$P(X=x)=0.25 \quad x=1,2,3,4$。已知理赔次数和理赔额相互独立，设$S$表示理赔总量，求$F_S(3)$。

解：由几何分布的均值公式$E(N)=r\beta$，**可得参数**β=4，试验成功概率0.2，$p_k=\dfrac{4^k}{5^{k+1}}$。***N*的分布列为**

N	P（$N=n$）
0	0.2
1	0.16

2	0.128
3	0.1024

由公式（3）计算X分布的3重卷积为

x	$f^{*1}(x)$	$f^{*2}(x)$	$f^{*3}(x)$
0	0	0	0
1	0.25	0	0
2	0.25	0.0625	0
3	0.25	0.125	0.0156

由于X的最小可能取值为1，因此 $f^{*k}(x)=0,\quad k>x$，因此

$$f_S(x)=\sum_{k=0}^{\infty}P(N=k)f^{*k}(x)=\sum_{k=0}^{x}P(N=k)f^{*k}(x)$$

计算得

$$f_S(0)=0.2,\ f_S(1)=0.04,\ f_S(2)=0.048,\ f_S(3)=0.0576$$

因此 $F_S(3)=0.346$。

```
#加载 actuar 包
library(actuar)
#索赔额 X 的分布
fx = c(0, 0.25, 0.25, 0.25, 0.25)
#索赔次数 N 的分布
pn = dgeom(0:3, 0.2)
#卷积法计算总索赔额复合分布
```

```
Fs1 = aggregateDist("convolution", model.freq = pn, model.sev = fx)
#计算卷积方法下总索赔额不大于 3 的概率
Fs1(3)
```

例 3.11：已知总理赔额 S 的分布密度函数为

$$f_S(x) = \sum_{n=0}^{\infty} f_X^{*n}(x) \binom{n+2}{n} (0.6)^3 (0.4)^n$$

其中理赔额 X 的分布为 $f_X(1) = 0.5$， $f_X(2) = 0.3$， $f_X(3) = 0.2$，则 $E(S)$ 等于（ ）。

A.1.7　　B.2.4　　C. 3.4　　D. 4.5　　E. 5.6

【解】由总理赔额的密度函数知

$$P\left(N=n\right)=\binom{n+2}{n}^{3}\left(0.6\right)^{3}\left(0.4\right)^{n}$$

故 N 服从参数 $r=3,\beta=\dfrac{2}{3}$ 的负二项分布，故

$$E\left(N\right)=r\beta=2$$

$$Var\left(N\right)=r\beta\left(1+\beta\right)=\frac{10}{3}$$

由理赔额的分布知

$$E\left(X\right)=1\times0.5+2\times0.3+3\times0.2=1.7$$

$$E(S)=E(N)E(X)=2\times1.7=3.4$$

例 3.12：设个别理赔额分布 X 服从指数分布，均值为θ，理赔次数 N 服从二项式分布，求 S 的分布。

解：固定$N=n$，由 gamma 分布的可加性知，$X_1+X_2+\cdots+X_n$ 的分布为 gamma 分布。因此，X 的 n 重卷积为

$$F_X^{*n}=\Gamma(n,\frac{x}{\theta})$$

由于任何 gamma 函数$\Gamma(n,x)$都可以写成

$$\Gamma(n,x)=1-\sum_{j=0}^{n-1}\frac{x^j e^{-x}}{j!}$$

（证明参见随机过程教材（张波、张景肖著）或者 SOA 的 Table）

因此，

$$\begin{aligned} F_S(x) &= p_0 + \sum_{n=1}^{\infty} p_n \Gamma(n, \frac{x}{\theta}) \\ &= p_0 + \sum_{n=1}^{\infty} p_n (1 - \sum_{j=0}^{n-1} \frac{(x/\theta)^j e^{-x/\theta}}{j!}) \\ &= 1 - e^{-x/\theta} \sum_{j=0}^{\infty} \frac{(x/\theta)^j}{j!} \sum_{n=j+1}^{\infty} p_n \\ &= 1 - e^{-x/\theta} \sum_{j=0}^{\infty} \overline{p}_j \frac{(x/\theta)^j}{j!} \qquad\qquad x \geq 0 \end{aligned}$$

其中，$\overline{p}_j = \sum_{n=j+1}^{\infty} p_n$ ，$j = 0,1,2\cdots$。

由于 $N \sim B(m, p)$， $P(N \geq m) = 0$，因此

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

应用卷积法注意的问题

1、连续分布转化为离散的分布

2、计算量巨大，约为 $O(n^3)$

其他例子

289. A compound Poisson distribution has $\lambda = 5$ and claim amount distribution as follows:

x	$p(x)$
100	0.80
500	0.16
1000	0.04

Calculate the probability that aggregate claims will be exactly 600.

Question #289
Key: D

600 can be obtained only 2 ways, from $500+100$ or from 6(100).

Since $\lambda = 5$ and $p(100) = 0.8$, $p(500) = 0.16$.

$$\Pr(6 \text{ claims for } 100) = \frac{e^{-5}5^6}{6!}(0.8)^6 = 0.03833 \text{ or } 3.83\%$$

$$\Pr(\overset{\text{2 claims}}{500+100}) = \frac{e^{-5}5^2}{2!}\left[(0.8)^1(0.16)^1(2)\right] = 0.02156 = 2.16\%$$

The factor of 2 inside the bracket is because you could get a 500 then 100 or you could get a 100 then 500. Total is 3.83% + 2.16% = 5.99%.

（二）矩母函数法

设 $P_S(z)$ 表示 S 的母函数， $M_S(z)$ 表示 S 的矩母函数

$$P_S(z) = P_N(P_X(z)) \qquad （6）$$

$$M_S(z) = E(e^{zS}) = P_N(P_X(e^z)) = P_N(M_X(z)) \qquad （7）$$

例 3.13: 设理赔额 $X_1, X_2, \cdots, X_n$ 服从指数分布，均值为 θ 。设理赔次数 N 服从几何分布，参数为 β 。求 S 的分布。

解： X_i 的矩母函数为

$$M_X(z) = \int_0^\infty e^{zx} \frac{1}{\theta} e^{-\frac{x}{\theta}} dx = \frac{1}{\theta} \int_0^\infty e^{zx} e^{(z-\frac{1}{\theta})x} dx$$
$$= (1-\theta z)^{-1}$$

N 的母函数为

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

由公式（7），S 的矩母函数是

$$\begin{aligned} M_S(z) &= P_N(M_X(z)) \\ &= \{1-\beta[(1-\theta z)^{-1}-1]\}^{-1} \\ &= \frac{\beta}{1+\beta}(1-(1+\beta)\theta z)^{-1} + \frac{1}{1+\beta} \end{aligned}$$

第二个等式是因为

$$
\begin{aligned}
\{1-\beta[(1-\theta z)^{-1}-1]\}^{-1} &= (1-\frac{\beta\theta z}{1-\theta z})^{-1} \\
&= \frac{1-\theta z}{1-(1+\beta)\theta z} \\
&= \frac{(1+\beta)(1-\theta z)}{(1+\beta)(1-(1+\beta)\theta z)} \\
&= \frac{1-(1+\beta)\theta z+\beta}{(1+\beta)(1-(1+\beta)\theta z)} \\
&= \frac{\beta}{1+\beta}(1-(1+\beta)\theta z)^{-1}+\frac{1}{1+\beta}
\end{aligned}
$$

这是一个由指数分布和贝努利（0－1）分布组成的混合分布。它的分布密度和分布函数为

$$f_s(x)=\begin{cases}\dfrac{1}{1+\beta} & x=0\\[2ex] \dfrac{\beta}{\theta(1+\beta)^2}\exp\left(-\dfrac{x}{\theta(1+\beta)}\right) & x>0\end{cases}$$

$$F_S(x)=1-\frac{\beta}{1+\beta}\exp\left(-\frac{x}{\theta(1+\beta)}\right) \qquad x\geq 0$$

例 3.14：设频率 N 服从负二项分布， r 是整数，X 服从指数分布 $e(\theta)$，求 S 的分布。

解：负二项分布的母函数为

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

利用与上例类似的方法我们有

$$
\begin{aligned}
M_s(z) &= P_N(M_X(z)) \\
&= P_N((1-\theta z)^{-1}) \\
&= \{1-\beta[(1-\theta z)^{-1}-1]\}^{-r} \\
&= \left\{\frac{\beta}{1+\beta}(1-(1+\beta)\theta z)^{-1}+\frac{1}{1+\beta}\right\}^{r} \\
&= \left\{1+\frac{\beta}{1+\beta}[(1-(1+\beta)\theta z)^{-1}-1]\right\}^{r}
\end{aligned}
$$

这是一个复合分布，因此

$$
M_s(z) = P_{N*}(M_{X*}(z))
$$

其中，$P_{N*}(z) = (1+\dfrac{\beta}{1+\beta}(z-1))^{r}$，这是二项分布。

$M_X(z)=(1-(1+\beta)\theta z)^{-1}$，这是指数分布，参数为$(1+\beta)\theta$。

因此，$F_S(x)=1-\sum_{n=1}^{r}\binom{r}{n}(\frac{\beta}{1+\beta})^n(\frac{1}{1+\beta})^{r-n}\sum_{j=0}^{n-1}\frac{(x\theta^{-1}(1+\beta)^{-1})^j e^{-x\theta^{-1}(1+\beta)^{-1}}}{j!}$

（三）递推法

设个体保单损失额 X 取值$0,1,2\cdots r$，这r个值表示货币单位的整数倍，r表示最大的赔付额，r可以取值无穷。假设损失次数N属于（a,b,0）分布族，即分布列满足下面关系式

$$p_k = P(N=k) = (a+\frac{b}{k})p_{k-1} \qquad k=1,2,3,4\cdots$$

定理 如果N的分布属于（a,b,0）分布族，损失额 X 取有限个整数值$0,1,2,\cdots$，r，则总损失额的分布为

$$f_S(x) = \frac{\sum_{y=1}^{x\wedge r}(a+\frac{by}{x})f_X(y)f_S(x-y)}{1-af_X(0)} \tag{8}$$

$$f_S(0)=P_N(f_X(0))$$

假设 N 服从泊松分布，

$$p_k=\lambda^k\frac{e^{-\lambda}}{k!}=0+\frac{\lambda}{k}p_{k-1}$$

因此 a=0, b=λ，这时 S 的分布为

$$f_S(x)=\sum_{y=1}^{x\wedge r}\frac{\lambda}{x}yf_X(y)f_S(x-y) \qquad (9)$$

$$f_S(0)=P_N(f_X(0))=e^{\lambda(f_X(0)-1)} \qquad (10)$$

有兴趣的同学阅读下面的证明：

＃＃＃＃＃＃＃

证明：由（$a, b, 0$）分布族的性质，我们可知

$$np_n = a(n-1)p_{n-1} + (a+b)p_{n-1} \tag{4.2.11}$$

设 $P_X(z)$ 为 X 的母函数，在公式（4.2.11）两边乘以 $(P_X(z))^{n-1}P_X'(z)$，并对 n 求和，得到

$$\sum_{n=1}^{\infty} np_n (P_X(z))^{n-1} P_X'(z) = a\sum_{n=1}^{\infty}(n-1)p_{n-1}(P_X(z))^{n-1}P_X'(z)$$
$$+(a+b)\sum_{n=1}^{\infty} p_{n-1}(P_X(z))^{n-1}P_X'(z)$$

由于

$$P_S(z)=\Sigma_{n=0}^{\infty}p_n(P_X(z))^n,$$

由前面的公式得到

$$P_S^{'}(z)=a\sum_{n=0}^{\infty}np_n(P_X(z))^nP_X{}'(z)+(a+b)\sum_{n=0}^{\infty}p_n(P_X(z))^nP_X{}'(z)$$

因此，

$$P_S^{'}(z)=aP_S^{'}(z)P_X(z)+(a+b)P_S(z)P_X^{'}(z)$$

两边对 z 用级数展开，比较公式两边 z^{x-1} 的系数，得到

$$xf_S(x)=a\sum_{y=0}^{x\wedge r}(x-y)f_X(y)f_S(x-y)+(a+b)\sum_{y=0}^{x\wedge r}yf_X(y)f_S(x-y)$$

$$=axf_X(0)f_S(x)+a\sum_{y=1}^{x\wedge r}(x-y)f_X(y)f_S(x-y)+(a+b)\sum_{y=1}^{x\wedge r}yf_X(y)f_S(x-y)$$

$$=axf_X(0)f_S(x)+ax\sum_{y=1}^{x\wedge r}f_X(y)f_S(x-y)+b\sum_{y=1}^{x\wedge r}yf_X(y)f_S(x-y)$$

因此，

$$f_S(x)=af_X(0)f_S(x)+\sum_{y=1}^{x\wedge r}(a+\frac{by}{x})f_X(y)f_S(x-y),$$

整理后即得（4.2.9）。

由全概率公式我们有

$$
\begin{aligned}
f_S(0) &= \sum_{n=0}^{\infty} P(X_1 + \cdots + X_n = 0) P(N = n) \\
&= \sum_{n=0}^{\infty} f_X(0)^n P(N = n) \\
&= P_N(f_X(0))
\end{aligned}
$$

证毕。

＃＃＃＃＃＃＃＃＃

例 3.15 设某险种的总理赔额服从复合泊松分布，平均理赔次数为 0.2 次，在任何一次理赔中，有 80％的概率会损失 5000 元，有 20％的概率会损失 10000 元。试计算保险人所面临的总理赔额的分布。

解：设 X 为个别理赔额，则 X 取值为 1，2 两个数，货币单位为 5000 元，$\lambda = 0.2$，由公式（9）

$$f_S(0) = e^{-\lambda} = e^{-0.2} = 0.8187$$

利用 $f_S(x) = \sum_{y=1}^{x \wedge r} \frac{\lambda}{x} y f_X(y) f_S(x-y)$

$$f_S(1) = \lambda f_X(1) f_S(0) = 0.2 \times 0.8 \times \exp(-0.2) = 0.1309$$

$$f_S(2) = \frac{\lambda}{2}[f_X(1) f_S(1) + 2 f_X(2) f_S(0)] = 0.043229$$

$$f_S(3) = \frac{\lambda}{3}[f_X(1) f_S(2) + 2 f_X(2) f_S(1)] = 0.005796$$

如此递推下去，结果列入下表：

x	$f_S(x)$	$F_S(x)$
0	0.818731	0.818731
1	0.130997	0.949728
2	0.043229	0.992957
3	0.005796	0.998756
4	0.001097	0.999853
5	0.000128	0.999981
6	0.000018	0.999999

例 3.16：设总理赔额 S 的分布列为

$$f_S(x)=\frac{1}{x}[0.2f_S(x-1)+kf_S(x-2)+0.6f_S(x-3))],x=1,2,3\ldots$$

已知 $E(S)=1.5$，求个体理赔额的期望 E(X)。

解：从 S 的分布知，S 服从由复合泊松分布的递归公式

$$f_S(x)=\sum_{y=1}^{x\wedge r}\frac{\lambda}{x}yf_X(y)f_S(x-y)$$

知 X 只取 1、2、3，

$$f_S(x)=\frac{1}{x}\left[\lambda f_X(1)f_S(x-1)+2\lambda f_X(2)f_S(x-2)+3\lambda f_X(3)f_S(x-3)\right]$$

故有方程组

$$\lambda f_X(1)=0.2 \qquad (1)$$

$$2\lambda f_X(2)=k$$

$$3\lambda f_X(3)=0.6$$

$$r=3$$

由于

$$
\begin{aligned}
E(S) = E(N)E(X) &= \lambda[1 f_X(1) + 2 f_X(2) + 3 f_X(3)] \\
&= 0.2 + k + 0.6 \\
&= 1.5
\end{aligned}
$$

解得 k＝0.7, 代入方程组（1）有

$$\lambda[f_X(1) + f_X(2) + f_X(3)] = 0.2 + \frac{0.7}{2} + \frac{0.6}{3} = 0.75$$

因此 $\lambda = 0.75$，即 $E(N) = 0.75$。故

$$E(X) = \frac{E(S)}{E(N)} = 2$$

若 N 服从(a,b,1)分布族，也有类似的结果：

$$f_S(x)=\frac{[p_1-(a+b)p_0]f_X(x)+\sum_{y=1}^{x\wedge m}(a+by/x)f_X(y)f_S(x-y)}{1-af_X(0)}$$

特别的，若 N 为 ZT-Poisson 或 ZM-poission 分布，则

$$f_S(x)=p_1^T f_X(x)+\sum_{y=1}^{x\wedge m}(\lambda y/x)f_X(y)f_S(x-y)$$

$$f_S(x)=[p_1^M-\lambda p_0^M]f_X(x)+\sum_{y=1}^{x\wedge m}(\lambda y/x)f_X(y)f_S(x-y)$$

R中实现的例子：

假设某保单组合在一年内的理赔次数N服从均值为4的几何分布，每次理赔额X的分布为$P(X=x)=0.25 \quad x=1,2,3,4$。已知理赔次数和理赔额相

互独立，设S表示理赔总量，用递推法计算$F_S(3)$。

```
#加载 actuar 包
library(actuar)
#索赔额 X 的分布
fx = c(0, 0.25, 0.25, 0.25, 0.25)
#索赔次数 N 的分布
pn = dgeom(0:3, 0.2)
#卷积法计算总索赔额复合分布
Fs1 = aggregateDist("convolution", model.freq = pn, model.sev = fx)
#计算卷积方法下总索赔额不大于 3 的概率
#递推法计算总索赔额复合分布
Fs2 = aggregateDist("recursive", model.freq = "geometric", model.sev = fx, prob = 0.2)
#计算卷积方法下总索赔额不大于 3 的概率
Fs1(3)
```

[1] 0.3456
#计算递推方法下总索赔额不大于 3 的概率
Fs2(3)
[1] 0.345

（四）数值计算方法（自主性学习）

1、卷积法的数值计算

在卷积法中要计算 $F_S(x)$，那么就需要计算每个 F 的 k 重卷积，即

$$F_X^{*k}(x)=\int_0^x F_X^{*k-1}(x-y)dF_X(y)$$

这是一个递推关系式。为此需要区间[0, x]上所有点的 k－1 重卷积，这所需要的计算工作量是非常巨大的。

为了避免这种技术难题，通常的方法是先将连续分布转化为离散分布，即将 $[0,\infty)$ 分成无穷个小区间 $[0,h/2),[h/2,3h/2),\cdots,[jh-h/2,jh+h/2),\cdots$，$h$ 为正数，令 f_j 为 X 在区间 $[jh-h/2,jh+h/2)$ 上的概率值，即

$$f_0 = P(X < h/2) = F_X(h/2),$$

$$\begin{aligned} f_j &= P(jh-h/2 \le X < jh+h/2) \\ &= F_X(jh+h/2-0) - F_X(jh-h/2-0) \\ &= \int_{jh-h/2}^{jh+h/2} f_X(x)dx \end{aligned}$$

$$\vdots$$

对于 $x = jh$，卷积公式变为

$$F_j^{*k} = \sum_{i=0}^{j} F_{j-i}^{*(k-1)} f_i \qquad (1)$$

$$f_j^{*k} = \sum_{i=0}^{j} f_{j-i}^{*(k-1)} f_i$$

当 h 趋于 0 时， f_j 将趋于 $f_X(x)$。然后利用公式（1）来计算 S 的分布。这种离散化的方法在实践中是可行的，因为任何损失和赔付都是某个货币单位的整数倍。但是，即使离散化了，卷积法的计算量仍然很大，约为 $O(n^3)$

傅立叶法变换利用了分布的特征函数与分布密度函数一一对应的性质：对

于任何一个分布都存在唯一的特征函数，对于任意一个特征函数都对应着唯一的概率分布。

对于一个连续函数 f，它的傅立叶变换定义为

$$\tilde{f}(z)=\int_{-\infty}^{\infty} f(x)e^{izx}dx \tag{2}$$

傅立叶变换是一个双射变换，因此任何一个原函数 f 可以通过它的逆傅立叶变换得到：

$$f(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{f}(z)e^{-izx}dz \tag{2}$$

若 $f_X(x)$ 是一个概率密度函数，则可以证明 $\tilde{f}(z)$ 是它的特征函数 $\varphi_X(x)$。

对于向量 $(f_0, f_1,\cdots, f_{n-1})$，离散傅立叶变换定义为

$$\tilde{f}_k = \sum_{j=0}^{n-1} f_j \exp(\frac{2\pi i}{n} kj), \qquad k = \ldots -1, 0, 1, \ldots \quad (3)$$

映射（3）是双射，$(f_0, f_1, \cdots, f_n)$ 也可以通过它的逆傅立叶变换得到：

$$f_j = \frac{1}{n} \sum_{k=0}^{n-1} \tilde{f}_k \exp(-\frac{2\pi i}{n} kj) \qquad \mathrm{j} = \ldots -1, 0, 1, \ldots \quad (4)$$

由公式（4）知，计算$\{\tilde{f}_j, j = 0, \cdots, n-1\}$的值的计算复杂度是$O(n^2)$。快速傅立叶变换（FFT）是建立在离散傅立叶变换的基础上的快速算法，它的计算复杂度为$O(n \log_2 n)$。它的基本思想是将离散傅立叶变换中的求和项分为两部分：奇数项求和以及偶数项求和。用公式可以表示为

$$\begin{aligned}\tilde{f}_k &= \sum_{j=0}^{n-1} f_j \exp(\frac{2\pi i}{n} kj) \\ &= \sum_{j=0}^{n/2-1} f_{2j} \exp(\frac{2\pi i}{n} 2jk) + \sum_{j=0}^{n/2-1} f_{2j+1} \exp(\frac{2\pi i}{n}(2j+1)k) \\ &= \sum_{j=0}^{m-1} f_{2j} \exp(\frac{2\pi i}{m} jk) + \exp(\frac{2\pi i}{n} k) \sum_{j=0}^{m-1} f_{2j+1} \exp(\frac{2\pi i}{m} jk)\end{aligned}$$

其中，$m = n/2$。令

$$\tilde{f}_k^a = \sum_{j=0}^{m-1} f_{2j} \exp(\frac{2\pi i}{m} jk),$$

$$\tilde{f}_k^b = \sum_{j=0}^{m-1} f_{2j+1} \exp(\frac{2\pi i}{m} jk)$$

从而可以得到，

$$\tilde{f}_k = \tilde{f}_k^a + \exp(\frac{2\pi i}{n}k)\tilde{f}_k^b$$

这样， $\tilde{f}_k$ 就可以写为两个求和项的和。$\{\tilde{f}_k^a\}$， $\{\tilde{f}_k^b\}$ 分别是 m 维向量 $(f_0, f_2, \cdots, f_{2(m-1)})$ 和 $(f_1, f_3, \cdots, f_{2m-1})$ 的傅立叶变换。因此，类似地，同样可以分解成奇数项求和以及偶数项求和两部分。这个过程可以继续下去。由于 $n/2, m/2, \ldots.$ 都是整数，因此快速傅立叶变换要求初始向量 $(f_0, f_1, \cdots, f_{n-1})$ 中的 n 是 2 的某个整数幂，即 $n = 2^r$，r 为某个正整数。

注意快速傅立叶算法计算 S 的分布时需要将个别理赔额的分布进行离散化。离散化后的个别理赔额的分布列的维数应满足 $n = 2^r$。利用快速傅立叶变换求解

S 的分布步骤为，

（1）将 X 的连续分布离散化为向量 $f_X(0), f_X(1), \cdots, f_X(n-1)$，其中 $n=2^r$。

（2）使用 FFT 算法求解 X 的特征函数，得到一个 $n=2^r$ 维的复向量；

（3）利用复合函数的特征函数公式 $\varphi_S(z)=P_N(\varphi_X(z))$，求得 S 的特征函数

（4）使用快速逆傅立叶变换求 $f_S(x)$，得到 S 的分布列 $f_S(0), f_S(1), \cdots, f_S(n-1)$。

例 3.17 设个体理赔额 X 取值为 1，2，3，相应的分布列为 0.5，0.4，0.1。假设理赔次数 N 服从泊松分布，$\lambda=3$。当 $n=8$、$n=16$、$n=32$、$n=4096$ 时，

使用 FFT 算法[1]得到 S 的分布如下表，

表 4.5.1

S	$n=8$ $f_S(s)$	$n=16$ $f_S(s)$	$n=32$ $f_S(s)$	$n=4096$ $f_S(s)$
0	0.11227	0.050932	0.049787	0.04979
1	0.11821	0.075283	0.074681	0.07468
2	0.1447	0.11606	0.11575	0.11575
3	0.151	0.13271	0.13256	0.13256
4	0.14727	0.13604	0.13597	0.13597
5	0.13194	0.12529	0.12525	0.12525
6	0.10941	0.1056	0.10558	0.10558

[1] 许多计算软件都有快速傅立叶变换程序，可以直接使用。

7	0.08518	0.083058	0.08305	0.08305

从表可以看出，随着 n 的增加，$f_S(s)$的数值虽然在变化，但变化的幅度越来越小，说明 n 是控制精度的有效参数。

R 程序实现的例子：

（1）损失金额为离散的情况：

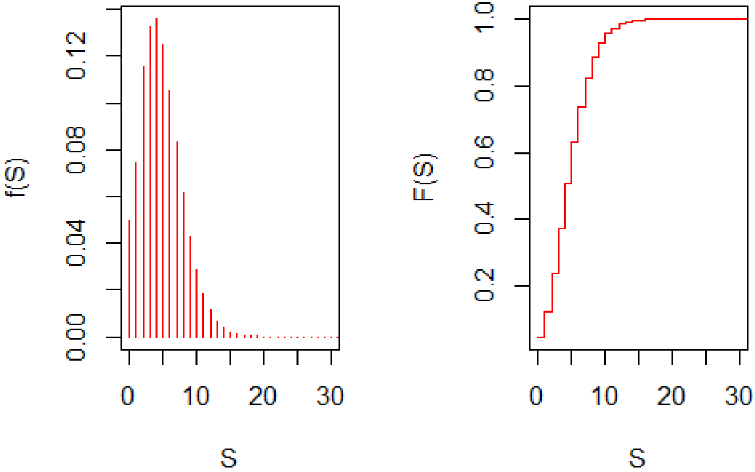
损失金额 X 等于 1，2 和 3 的概率：P（X=1）=0.5，P（X=2）=0.4，P（X=3）=0.1 。损失次数 N 服从 Poisson 分布，$\lambda=3$ 。应用 FFT 计算 S 的分布

解：

R 程序：

```
x = c(0, 0.5, 0.4, 0.1 , rep(0, 40))    #损失金额分布
z = fft(x)    #损失金额的特征函数
y = exp(3*( z - 1));    #累积损失特征函数
```

```
s = fft(y, inverse = TRUE)/length(x)    #FFT 逆变换
s = Re(s) #  累积损失的概率函数
S = cumsum(s) #分布函数
par(mfrow = c(1, 2))
plot(0:(length(S)-1), s, type = 'h', xlim = c(0, 30), col= 2, xlab='S',ylab='f(S)')
plot(0:(length(S)-1), S, type = 's', xlim = c(0, 30), col = 2, xlab='S',ylab='F(S)')
```



f(S)
0.00
0.04
0.08
0.12
0
10
20
30
S
F(S)
0.2
0.4
0.6
0.8
1.0
0
10
20
30
S

（2）损失金额为连续的情况，首先对其离散化：

损失参数服从参数为 lam=2 的泊松分布，损失金额服从参数为(shape=2, scale=500)的伽马分布，用 FFT 方法求累积索赔金额的分布。

解：

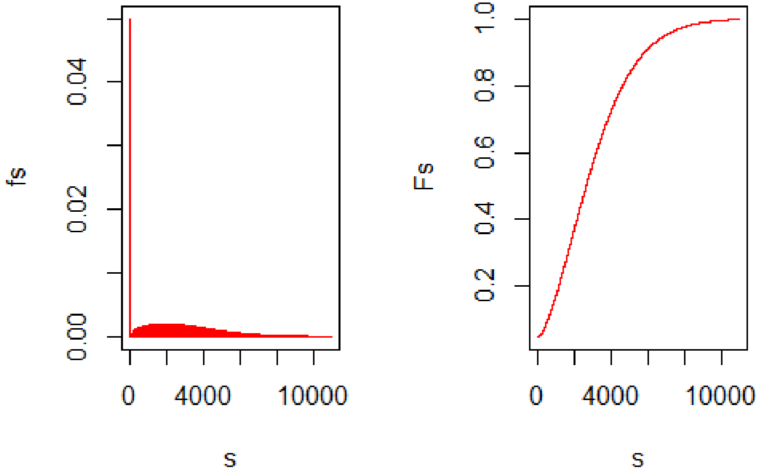
R 程序

```
x <- seq(0, 10000, by = 10)    #等间隔离散化，步长为 10
F_gam <- pgamma(x, shape = 2, scale = 500) #产生 gamma 的分布函数
f_gam <- diff(F_gam)    #离散化以后伽马分布的概率
f_gam <- c(f_gam, 1-pgamma(10000, shape = 2, scale = 500))    #补足概率
sum(f_gam)
[1]  1

plot(x, f_gam, type = 's', col = 2)
f_gam <- c(f_gam, rep(0, 100))    #右尾补充足够的零
phi_x=fft(f_gam)    #损失金额的特征函数
```

```
phi_s=exp(lam*(fft(f_gam)-1))    #累计损失的特征函数
fs <- Re(fft(phi_s, inverse = TRUE)/length(f_gam) ) #累积索赔金额的分布
sum(fs)

Fs <- cumsum(fs) #求出 s 的分布
par(mfrow=c(1,2))
plot(10*(0:(length(fs)-1)), fs, type = 'h', col=2, xlab='s')
plot(10*(0:(length(Fs)-1)), Fs, type = 's', col=2, xlab='s')
```


fs
0.00
0.02
0.04
0
4000
10000
s
Fs
0.2
0.4
0.6
0.8
1.0
0
4000
10000
s

（五）、随机模拟的方法

本小节内容来自教材《精算模型》第九章

1、用反变换法产生一般分布的随机数

设随机变量 X 的分布函数为 $F(x)$，定义

$$F^{-1}(u)=\sup\{x:F(x)\le u\}，\quad 0<u<1 \qquad (1)$$

假设随机变量 U 服从（0,1）上的均匀分布，则随机变量 $F^{-1}(U)$ 的分布函数是 $F(x)$，即在分布相等的意义上有 $X=F^{-1}(U)$，

要证明 x 为所要求的随机数，只要证明所构造的随机变量 $F^{-1}(U)$ 具有分布函数 $F(x)$ 即可，事实上，$X=F^{-1}(U)$ 的分布函数为：

$$P(X \le x) = P(F^{-1}(U) \le x) = P(F(F^{-1}(U)) \le F(x)) = P(U \le F(x)) = F(x)$$ 。

反函数模拟的具体步骤为

（1）由 $U(0,1)$抽取 u，

（2）计算 $x = F^{-1}(u)$，其中 F^{-1} 如(1)中定义。

以上步骤的含义是，若已知分布函数 $F(x)$，只要产生一个均匀分布随机数 u，对应的取分布 $F(x)$的 u 分位点即可。形象的说就是，由均匀分布随机数 u 确定了纵坐标，然后根据分布 F，找对应的原象（横坐标）。

分几种情况讨论

（1）如果 $F(x)$是连续且严格单调的，则 $F^{-1}(u)$为 F 的随机数。

例 3.18 假设随机变量 X 的密度函数为 $f\left(x\right)=\begin{cases}24x^2 & 0\le x\le 0.5\\ 0 & 其他\end{cases}$，定义 v 是(0,1)均匀分布的随机数，已知 v=0.125。请使用反变换法确定 X 的随机观测值。

解: 当 $0\le x\le 0.5$， $F(x)=\int_{-\infty}^{x} f\left(t\right)dt=\int_{0}^{x}24t^2dt=8x^3$，

当 $x\ge 0.5$， $F(x)=1$。

所以，由反变换法知，X 的模拟数为方程 $v=F(x)$ 的解，即 $0.125=8x^3$，

$x=0.25$。

例 3.19 已知分布函数为$F(0)=0$，$F(1)=0.4$，$F(2)=1.0$，$F(x)$在[0,1]和[1,2]上是线性函数，使用下列来自（0,1）均匀分布的随机数：0.2，0.4，0.7，用反变换法生成上面分布的三个模拟数，并计算三个模拟值的均值。

解 由题意知分布函数为

$$F(x)=\begin{cases}0, & x<0\\ 0.4x, & 0\le x\le 1\\ 0.6x-0.2, & 1<x\le 2\\ 1, & x>2\end{cases}$$

如果$0<u\le 0.4$，$F(x)=0.4x$，如果$0.4<u\le 1, F(x)=0.6x-0.2$。

因为模拟数为方程$u=F(x)$的解，所以

$$0.2=u_1=F\left(x_1\right)=0.4x_1，得x_1=0.5;$$

$$0.4=u_2=F\left(x_2\right)=0.4x_2，得\ x_2=1;$$

$$0.7=u_3=F\left(x_3\right)=0.6x_3-0.2，得x_3=1.5.$$

所以模拟值的样本均值为$\dfrac{0.5+1+1.5}{3}=1$。

（2）如果纵坐标 u 有多个原象（多个横坐标的函数值相等 $u = Fx(x)$），由(2.1)式，则取数值最大的原象 x_0（横坐标）为对应的随机数，见图 1。

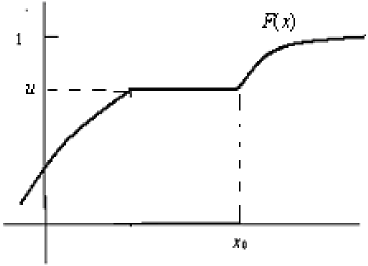


图 1 一个特殊情况时分位点的取法

例 3.20

Suppose

$$F_X(x) = \begin{cases} 0.5x, & 0 \le x < 1, \\ 0.5, & 1 \le x < 2, \\ 0.5x - 0.5 & 2 \le x < 3. \end{cases}$$

Determine the simulated values of x resulting from the uniform numbers 0.3, 0.5, and 0.9.

解：当 u=0.3 时，由 $0.3 = 0.5x$ 得到第一个随机数为 0.6，

当 u=0.5 时，取最大值 x=2 作为第二个随机数。

当 u=0.9 时，由 $0.9 = 0.5x - 0.5$，得到第三个随机数为 2.8。

（3）如果纵坐标 u 没有对应的原象（u 不在分布函数的值域中），

则取值域中大于 u 且最接近 u 的纵坐标 v 的原象 x_1，即跳跃发生时的横坐标为对应的随机数，见图 2。

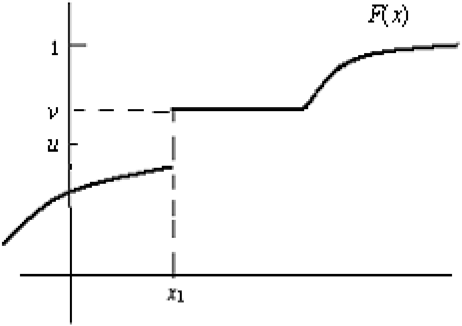


图 2 有跳跃时分位点的取法

例 3.21：

Suppose

$$F_X(x) = \begin{cases} 0.5x, & 0 \le x < 1, \\ 0.5 + 0.25x & 1 \le x \le 2. \end{cases}$$

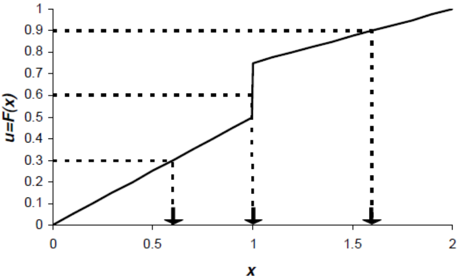


图 3

Determine the simulated values of x resulting from the uniform numbers 0.3, 0.6, and 0.9.

解：由图 3 可知 0.3, 0.6, and 0.9 对应的随机数为 0.6、1 和 1.6。

例 3.22:

Simulate values from a binomial distribution with $m = 4$ and $q = 0.5$ using th
uniform numbers 0.3, 0.6875, and 0.95.

The distribution function is

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 0.0625, & 0 \leq x < 1, \\ 0.3125, & 1 \leq x < 2, \\ 0.6875, & 2 \leq x < 3, \\ 0.9375, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$

0.3→1,0.6875→3, 0.95→4

Table 20.2 Simulation lookup for Example 20.4.

For u in this range,	the simulated value is
$0 \le u < 0.0625$,	0
$0.0625 \le u < 0.3125$,	1
$0.3125 \le u < 0.6875$,	2
$0.6875 \le u < 0.9375$,	3
$0.9375 \le u < 1$,	4

$$\sum_{i=0}^{j-1} p_i \leq u < \sum_{i=0}^{j} p_i \rightarrow x_{\mathrm{j}}$$

总结：离散随机变量可用反变换法生成。因为离散型随机变量的分布函数是阶梯函数，针对这种特殊情形，把反变换法具体改述如下：

设随机变量 X 有概率分布为$P(X=x_j)=p_j$， $j=0,1,2,\cdots$，其中 $p_0+p_1+p_2+\cdots=1$。对均匀分布随机数 u，只需找整数 j，满足

$\sum_{i=0}^{j-1} p_i \le u < \sum_{i=0}^{j} p_i$ ，则 X 的模拟值取 x_j 。

如果 $u<p_0$，则 X 的模拟值取 x_0 。

如果 $p_0 \le u < p_0 + p_1$，则 X 的模拟值取 x_1 。。。。。。。以此类推。

练习假设随机变量 X 的分布函数为：

(1) 当 $x<0$ 时，$F(x)=0$，

(2) $F(0)=1/2$，

(3) 当 $0<x<1$ 时，$F'(x)=x$ 。

若有(0,1)均匀分布的三个观测值：0.25,0.625 和 0.52。请按上述 X 的分布模拟三个样本 X_1, X_2, X_3，并确定 X 的样本均值。

解

因为 $P\left[X=0\right]=0.5$，$F\left(x\right)=\frac{1}{2}x^2+c, F\left(0\right)=\frac{1}{2}$，所以 $c=\frac{1}{2}$。

由反变换法知，

当 $0<u\leq\frac{1}{2}$ 时，X 的模拟数为 0，所以当 $u_1=0.25$ 时，$x_1=0$。

当 $\frac{1}{2}<u<1$， $u=F\left(x\right)=\frac{1}{2}x^2+\frac{1}{2}$，模拟数 x_1、 x_2 分别满足

$u_2=0.625=\frac{1}{2}x_2^2+\frac{1}{2}$， $u_3=0.52=\frac{1}{2}x_3^2+\frac{1}{2}$,

即 $x_2=0.5$ $x_3=0.2$。

例 3.23:指数分布的模拟

解 假设指数分布的分布函数为 $F(x)=1-\exp(-\lambda x), x>0, \lambda>0$，则其反函数为

$$F^{-1}(y)=-\frac{1}{\lambda}\ln(1-y).$$

所以，只要将（0,1）均匀分布的随机数 u 代入 $F^{-1}(y)$ 中即可得到指数

分布的随机数 $x=-\frac{1}{\lambda}\ln(1-u)$.

若 U 服从（0,1）均匀分布，则 1-U 仍然是服从（0,1）均匀分布，所以 1-u 仍然是（0,1）均匀分布的随机数。又由于分布函数是单调递增的，小的数值对应的分位点也要小，所以对于较小的均匀分布的随机数 u，如果取 u 对应的分位点，则意味着小的随机数产生小的模拟数。如果取 1-u 对应的分位点，则意味着小的随机数产生大的模拟数，如在上例中可得到另一个指数分布的随机数 $x'=-\frac{1}{\lambda}\ln(1-(1-u))=-\frac{1}{\lambda}\ln(u)$。

（4） 有些分布的反函数没有显式，如正态分布，这时只能数值计算或采用其他方法。

2、随机模拟方法计算总索赔额的分布例子

例 3.24 假设一个医疗保险合约由三个独立的被保险人组成，他们一年中的总看病次数服从二项式分布 n=3，p=0.9。设每次看病的医疗费用为 X，X 服从以下分布

x	100	10,000
$P(X=x)$	0.9	0.1

每年，三人的总医疗费用超过 5000 元的部分由保险公司支付。

假设 0.01，0.20 为(0,1)均匀分布的随机数，用这两个数来分别产生第一年和第二年的总看病次数。利用下列(0,1)均匀分布的随机数来产生每次看病的费用：0.80，0.95，0.70，0.96，0.54，0.01。请计算这两年内保

险公司给付的总数的一个模拟结果。

【解】因为

$$p_0=P\left[N=0\right]=0.1^3=0.001,\ p_1=P\left[N=1\right]=0.3\times0.9\times0.1^2=0.027,$$

$$p_2=P\left[N=2\right]=0.3\times0.9^2\times0.1=0.243,\ p_3=P\left[N=3\right]=0.9^3=0.729,$$

$$p_0+p_1=0.028,\ p_0+p_1+p_2=0.271$$

由于0.01满足 $p_0=0.001\le0.01<0.028=p_0+p_1$，所以，第一年总看病次数的模拟数是 1，而 0.2满足$p_0+p_1=0.028\le0.2<0.271=p_0+p_1+p_2$，所以，第二年总看病次数的模拟数是 2。

又因为每次看病的医疗费用的分布函数为 $F_X(100)=0.9, F_X(10000)=1$，根据离散随机变量的反变换法，由 $0.8<F_X(100)$，知(0,1)均匀分布的随机数 0.8 对应的医疗费用的模拟数是 100。如果用 0.95 来生成费用的随机数，则由 $F_X(100)=0.9\le 0.95<1=F_X(10000)$ 知模拟数取 10000。同理 $0.7\le 0.9=F_X(100)$，所以 0.7 对应的模拟费用为 100。

由于每年总看病次数的模拟要求，第一年只需要一个费用模拟数，第二年只需要两个费用模拟数，所以可以认为在第一年，总看病次数为 1，费用为 100，在第二年，总看病次数为 2，费用为 10100。

所以保险公司第一年给付为 0，第二年给付为 10100-5000=5100。

例 3.25 保险公司对于一个剧院的供电故障造成的损失提供四个月的保险赔偿。

(1) 每月有 1000 元的绝对免赔额。

(2) 保险公司假设这个剧院每月的损失相互独立的服从均值为 1500，标准差为 200 的正态分布。

(3) 为了模拟这四个月的保险赔偿费用，保险公司运用反变换方法（小的随机数对应低的赔偿费用）.

(4) 从（0, 1）中随机抽取的 4 个数分别为: 0.5438, 0.1131, 0.0013, 0.7910.

计算保险公司模拟得到的保险费用。

解：

根据反变换方法，由(0,1)均匀分布的随机数 u，通过查表可以得到

模拟的标准正态分布 Z 的值 z，这里 $P(Z < z) = u$。所以从已知的 4 个随机数，可以得到 4 个标准正态分布的模拟随机数，即：

U	0.5438	0.1131	0.0013	0.7910.
Z	0.11	-1.21	-3.01	0.81

损失 X 服从均值为 1500，标准差为 200 的正态分布，所以若令 $X = 200Z + 1500$，可得模拟的损失随机数：

1522, 1258 898 1662

扣除绝对免赔额，保险公司的最终赔付为：

522 258 0 662

于是，这四个月的总赔付为 1442。

例 3.26 对复合总索赔额的分布 S 进行模拟，首先进行索赔次数的模拟，然后进行索赔额的模拟。反变换法被用于索赔次数和索赔额的模拟（小的模拟值对应少的索赔次数和少的索赔额）。索赔次数服从 m=5，p=0.5 的二项分布。索赔额服从均值为 2000，方差为 20000000 的帕雷托分布。均匀随机数 0.3 被用于索赔次数的模拟，使用 0.1，0.7，0.5，0.3 中尽可能多的数进行索赔额的模拟。求 S 的模拟值。

解 由于索赔次数服从 n＝5，p＝0.5 的二项分布，即

K	0	1	2	3	4
$f(k)$	0.03125	0.15625	0.31250	0.31250	0.15625
$F(k)$	0.03125	0.18750	0.50000	0.81250	0.96875

若用随机数 0.3 进行模拟，因为 $F(1)=0.1875\le 0.3<F(2)=0.500$，根

据离散分布的模拟数产生原理可得索赔次数的模拟值是 2，所以要模拟两个索赔额。

参数为 α 和 θ 的帕雷托分布的均值为 $E[X]=\frac{\theta}{\alpha-1}$，二阶矩为

$$E\left[X^2\right]=\frac{2\theta^2}{(\alpha-1)(\alpha-2)}$$

所以方差为

$$Var[X]=E\left[X^2\right]-\left(E[X]\right)^2=\frac{2\theta^2}{(\alpha-1)(\alpha-2)}-\left(\frac{\theta}{\alpha-1}\right)^2=\frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)}$$

根据题意，有 $E\left[X\right]=\frac{\theta}{\alpha-1}=2000$，$Var\left[X\right]=\frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)}=20000000$

解得 $\alpha=2.5,\theta=3000$。

因此帕雷托分布的分布函数为 $F\left(x\right)=1-\left(\frac{\theta}{x+\theta}\right)^{2.5}=1-\left(\frac{3000}{x+3000}\right)^{2.5}$

对于给定的均匀随机数 u，索赔额的模拟值为 x，并且有 $u=F\left(x\right)$；

对于 u=0.1 有 $0.1=1-\left(\frac{3000}{x_1+3000}\right)^{2.5}$，得 $x_1=129$；

对于 u=0.7 有 $0.7=1-\left(\frac{3000}{x_2+3000}\right)^{2.5}$，得 $x_2=1856$；

所以总索赔额 S 的模拟值为 129＋1856＝1985。

R 程序实现：

```
#固定随机数种子
set.seed(123)
#从(0,1)均匀分布中 20 个随机数
s1 = runif(5, min = 0, max = 1)
s1
## [1] 0.2875775 0.7883051 0.4089769 0.8830174 0.9404673
#反变换法将 s1 转化为保险事故发生的时间间隔 t，间隔服从均值为 0.5 的指数分布
t = qexp(s1, rate = 2)
```

```
t
## [1] 0.1695421 0.7763047 0.2629501 1.0728651 1.4106146
#从(0,1)均匀分布中 5 个随机数
s2 = runif(5, min = 0, max = 1)
s2
## [1] 0.0455565 0.5281055 0.8924190 0.5514350 0.4566147
#反变换法将 s2 转化为赔偿额 x
x = qpareto(s2, shape = 2, scale = 1000)
x
## [1] 23.5873 455.7190 2048.8236 493.0946 356.5821
#计算 1 时刻时一共发生了多少次事故 n
for(i in 1:4){t[i+1] = t[i+1] + t[i]}
for(i in 1:4){if(t[i]>=1) break}
n = i-1
n
```

```
##[1] 2
#计算公司 1 时刻的盈余 k
k = 1000 + 2400 - x[1] - x[2]
k
##[1] 2920.694
```

自主性学习作业 2

（1）分别用卷积法、递推法和快速傅里叶变换法求讲义例 3.26 的总索赔额 S 分布，绘制分布函数曲线图，计算 99%的分位数

（2）尝试直接用 R 生成 pareto 分布随机数和泊松随机数，求讲义例 3.26 的总索赔额 S 分布，绘制分布函数曲线图，计算 99%的分位数。不会的同学可参考 https://openacttexts.github.io/LDARcode/#61_collective_risk_model:_without_coverage_modifications

（六）S 的近似分布

1、 正态近似

定理 设个别理赔额分布函数为 $f(x)$，$u_1=E(X)$，$u_2=E(X^2)$。

（1）如果S是复合泊松分布，参数为λ，则当$\lambda \to \infty$时$Z=\dfrac{S-\lambda u_1}{\sqrt{\lambda u_2}}$的分布趋于标准正态分布。

（2）如果S是复合负二项式分布，参数为r,β，

$$P(N=k)=\binom{k+r-1}{r}\left(\frac{\beta}{1+\beta}\right)^k\left(\frac{1}{1+\beta}\right)^r,k=0,1,2,\cdots$$

个别理赔额分布函数为$f(x)$，则

$$Z=\frac{S-r\beta u_1}{\sqrt{r\beta u_2+r\beta^2u_1^2}}$$

的分布在$r \to \infty$时趋于标准正态分布。

证明：我们将利用$\lim\limits_{\lambda \to \infty} M_Z(t) = e^{\frac{t^2}{2}}$来证明（1）和（2）。对于泊松分布情形，由$Z = \dfrac{S - E(S)}{\sqrt{Var(S)}} = \dfrac{S - \lambda u_1}{\sqrt{\lambda u_2}}$得到

$$M_Z(t) = E(\exp(t\frac{S-\lambda u_1}{\sqrt{\lambda u_2}})) = M_S(\frac{t}{\sqrt{\lambda u_2}})\exp\{-\frac{\lambda u_1 t}{\sqrt{\lambda u_2}}\}$$

由公式（5）知$M_S(t) = \exp[\lambda(M_X(t) - 1)]$，因此

$$M_Z(t) = \exp\{\lambda[M_X(\frac{t}{\sqrt{\lambda u_2}}) - 1] - \frac{\lambda u_1 t}{\sqrt{\lambda u_2}}\}$$

由矩母函数的级数展开式

$$M_X(t)=E(e^{tX})=1+E(X)t+\frac{E(X^2)}{2}t^2+\cdots+\frac{E(X^n)}{n!}t^n+\cdots,$$

我们可以得到，

$$M_Z(t)=\exp(\frac{1}{2}t^2+\frac{1}{6}\frac{1}{\sqrt{\lambda}}\frac{u_3^{'}}{(u_2)^{3/2}}t^3+\cdots)$$

当$\lambda\to\infty$， $M_Z(t)\to e^{\frac{t^2}{2}}$，即$\lim\limits_{\lambda\to\infty}M_Z(t)=e^{\frac{t^2}{2}}$。从而，Z 分布在$r\to\infty$时趋于标准正态分布。

对于负二项分布，令

$$Z=\frac{S-E(S)}{\sqrt{Var(S)}}=\frac{S-E(N)E(X)}{\sqrt{E(N)Var(X)+Var(N)E(X)^2}}$$

$$=\frac{S-r\beta u_1}{\sqrt{r\beta u_2+r\beta^2u_1^2}}$$

再用类似的方法证明 Z 分布在$r\to\infty$时趋于标准正态分布。此处不再叙述。

例 3.27: (SOA 2001-11 30) The claims department of an insurance company receives envelopes with claims for insurance coverage at a Poisson rate of $\lambda = 50$ envelopes per week. For any period of time, the number of envelopes and the numbers of claims in the envelopes are independent. The numbers of claims in the envelopes have the following distribution:

Number of Claims	Probability
1	0.20
2	0.25
3	0.40
4	0.15

Using the normal approximation, calculate the 90th percentile of the number of claims received in 13 weeks.

解：设 Y_i 表示第 i 个信封中的索赔数。设 $X(13)$ 表示 13 周内收到的总索赔数。

$$E(Y_i)=1\times0.2+2\times0.25+3\times0.4+4\times0.15=2.5$$

$$E(Y_i^2)=1\times0.2+4\times0.25+9\times0.4+16\times0.15=7.2$$

$$E(X(13))=50\times13\times2.5=1625$$

$$Var(X(13))=50\times13\times7.2=4680$$

由 $P(X(13)\le Z)=0.9=\Phi(1.282)$

$$=> P(\frac{X(13)-1625}{\sqrt{4680}} \le 1.282)$$

因此， $X(13) \le 1712.7$

2、平移 gamma 近似

设 $G(x;\alpha,\beta)=\int_0^x \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t} dt$ 为 gamma 分布，对任意一点 x_0 ，

定义一个新的分布函数 $H(x,\alpha,\beta,x_0)=G(x-x_0;\alpha,\beta)$ 。若设 $h(x)$ 和 $g(x)$ 分布为 $H(x,\alpha,\beta,x_0)$ 和 $G(x,x_0,\alpha,\beta)$ 的分布函数密度，则从图形上 H 和 G 只差了一个平移变换。

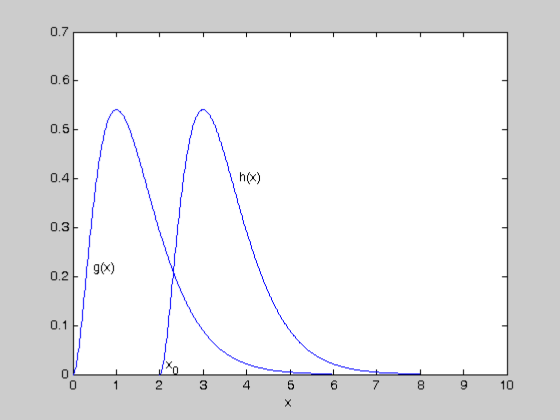


图 5.1 H 和 G 的分布密度图

下面我们用 $h(x)$ 来描述总理赔额 S 的分布密度。因为 $h(x)$ 有三个参数，所以只需根据 S 的均值、方差和三阶中心矩来定出 $h(x)$ 的形状和位置。又因为 H 的均值，方差和三阶中心矩分别为 $x_0+\frac{\alpha}{\beta}$， $\frac{\alpha}{\beta^2}$， $\frac{2\alpha}{\beta^3}$，所以用 $h(x)$ 来描述 S 的分布时，下面三个等式近似成立。

$$\begin{cases} E(S) = x_0 + \dfrac{\alpha}{\beta} \\ Var(S) = \dfrac{\alpha}{\beta^2} \\ E\left[(S - E(S))^3\right] = \dfrac{2\alpha}{\beta^3} \end{cases} \tag{7}$$

这是一方程组，解出x_0, α, β有

$$\begin{cases} x_0 = E(S) - \dfrac{2Var(S)^2}{E((S-E(S))^3)} \\ \alpha = \dfrac{4Var(S)^3}{\left[E((S-E(S))^3)\right]^2} \\ \beta = \dfrac{2Var(S)}{E((S-E(S))^3)} \end{cases}$$

这样就可以得到一个平移 gamma 分布。当S为复合泊松分布时，（7）可简化为

$$x_0 = \lambda u_1 - 2\lambda\frac{u_2^2}{u_3};\quad \alpha = \frac{4\lambda u_2^3}{u_3^2};\quad \beta = \frac{2u_2}{u_3}$$

其中$u_3 = E(X^3)$，这是因为$E((S-E(S))^3) = \lambda E(X^3)$。

若 $x_0+\dfrac{\alpha}{\beta}=u \quad \dfrac{\alpha}{\beta^2}=\sigma^2$，当 $x\to-\infty,\alpha\to\infty,\beta\to\infty$ 时，可以证明 $H(x,\alpha,\beta,x_0)$ 趋于正态分布 $N(u,\sigma^2)$。因此，正态分布可以看作是这种三参数分布的一种极限情况。从这个意义上来说，平移 gamma 分布近似是正态近似的推广。

例 3.28：设 S 为复合泊松分布，$\lambda=12$，当个体理赔额分布为 $[0,1]$ 上的均匀分布时，试分别用（1）正态近似 （2）平移 gamma 近似计算 $P(S<10)$。

解：（1）由条件易知$u_1 = E(X) = \frac{1}{2}$，$u_2 = E(X^2) = \frac{1}{3}$，于是得到

$$E(S) = \lambda u_1 = 6，\ Var(S) = \lambda u_2 = 4$$

所以，$P(S < 10) = P(\frac{S-6}{2} < 2) = \Phi(2) = 0.9772$

（2）令$u_3 = E(X^3) = \int_0^1 x^3 dx = \frac{1}{4}$，则$E(S - E(S))^3 = \lambda u_3 = 3$，解方程组

$$\begin{cases} 6 = x_0 + \alpha / \beta \\ 4 = \alpha / \beta^2 \\ 3 = 2\alpha / \beta^3 \end{cases}$$

得 $x_0 = -4.67, \alpha = 28.44, \beta = 2.67$ 。因此 S 的分布函数为

$G(x+4.67;28.44,2.67)$,

$$P(S<10)=G(14.67,28.44,2.67)=0.9683$$

3、对数正态近似

此外，实际中还使用对数正态分布 $\ln(u,\sigma^2)$ 来近似 S 的分布，即考虑方程组

$$\begin{aligned} E(S) &= \exp(u+\sigma^2/2) \\ E(S^2) &= \exp(2u+2\sigma^2) \end{aligned} \qquad (3.41)$$

解出 u,σ^2，然后用$\ln(u,\sigma^2)$来描述S的分布。

当$E(N)$的值较大时，正态分布近似的效果不错，特别是当 N 为泊松分布，二项式分布和负二项式分布时，由中心极限定理知，当λ,m,r趋于无穷时，S 的分布将趋于正态分布。而当$E(N)$的值较小时，S 的分布是有偏斜的，这时使用平移 gamma 和对数正态近似可能更为恰当。

例 3.29：在过去的 10 个月里每月某险种的理赔次数和个体损失额的样本均值（和标准差）分别为6.7(2.3)，179747(52,141)。

(1) 计算每个月总损失额的样本均值和方差。

（2）分别用正态近似和对数正态近似计算损失额超过期望的140% 的概

率 $P(S > 1.40E(S))$。

解：（1）由公式（1）和（2）得到 S 的样本均值和样本方差为

$$E(S) = E(X)E(N) = 6.7 \times 179747 = 1200955$$

$$\begin{aligned} Var(S) &= E(Var(S \mid N)) + Var(E(S \mid N)) \\ &= 6.7 \times 52{,}141^2 + 2.3^2 \times 179747^2 \\ &= 188180 \times 10^{11} \end{aligned}$$

若使用正态近似计算，则

$$
\begin{aligned}
P(S>1.40E(S)) &= P(S>1681337) \\
&= P\left(\frac{S-E(S)}{\sqrt{Var(S)}} > \frac{1681337-1200955}{433791}\right) \\
&= P(Z>1.107) \\
&= 1-\Phi(1.107) \\
&= 0.134
\end{aligned}
$$

若使用对数正态近似，则解下列方程

$$u+\sigma^2/2 = \ln(E(S)) = \ln(1.200955\times 10^6) = 13.99863$$

$$2u+2\sigma^2 = \ln(E(S^2)) = \ln(1.63047\times 10^{12}) = 28.11989$$

得到　$u=13.93731, \sigma^2 = 0.1226361$。因此

$$P(S > 1681337) = 1 - \Phi\left(\frac{\log 1681337 - 13.93731}{0.1226361^{0.5}}\right)$$

$$= 1 - \Phi(1.135913)$$

$$= 0.128$$