ТЕОРИЯ РИСКА

лекции

Поспелова Ирина Игоревна

Модели совокупных убытков

Страховые компании существует благодаря объединению рисков. При страховании большого количества людей индивидуальные риски объединяются в совокупный риск.

Пример 1.

Страховой случай имеет вероятность 10%, и когда он происходит, то приводит к убытку 5 000. Исследование рынка показало, что потребители готовы заплатить не более 550 за покупку полиса, покрывающего такой ущерб. Сколько полисов должна продать страховая компания, чтобы вероятностью 95% получить доход (издержки не учитываются)?

Пусть n - число проданных полисов. Разумно считать, что число исков C имеет биномиальное распределение с $m = n$ и $q = 0.1$, а суммарный платеж равен 5 000C. Условие получения дохода с вероятностью 95%:

$$0.95 \leq P\{5000C < 550n\} = P\{C < 0.11n\} = P\left\{Z < \frac{0.11n - 0.1n}{\sqrt{0.1 \cdot 0.9n}}\right\}.$$

Используя для приближения центральную предельную теорему, получаем, что

$$\frac{0.11n - 0.1n}{\sqrt{0.1 \cdot 0.9n}} = 1.645,$$

Откуда $n = 2\,435.42$.

Модель коллективного риска

Существуют два варианта построения модели суммы платежа по всем искам, произошедшим за фиксированный период времени и касающихся определенного набора договоров страхования. Первый вариант – учесть все сделанные платежи и затем их сложить. В таком случае получаем совокупные убытки в виде суммы S случайного числа слагаемых N отдельных платежей $X_1, X_2, \dots, X_N$:

$$S = X_1 + X_2 + \dots + X_N, \qquad N = 0, 1, 2, \dots ,$$

где $S = 0$ при $N = 0$.

Определение 1. Модель коллективного риска имеет вид S при условии, что X_j независимы и одинаково распределенные случайные величины, т.е.

1. При $N = n$ случайные величины $X_1, X_2, \dots, X_n$ одинаково распределенные и независимые случайные величины.

2. При $N = n$ общее распределение случайных величин $X_1, X_2, \dots, X_n$ не зависит от n.

3. Распределение N не зависит от $X_1, X_2, \dots$

Модель индивидуального риска

Второй вариант - приписать случайную величину каждому договору страхования.

Определение 2. Модель индивидуального риска представляет суммарный ущерб как сумму

$$X_1 + \cdots + X_n$$

фиксированного числа n договоров страхования. Убытки для n договоров равны $(X_1, X_2, \ldots, X_n)$, где X_j предполагаются независимыми, но необязательно одинаково распределенными. Распределение X_j обычно имеет вероятность в нуле, соответствующую вероятности отсутствия убытка или платежа по договору.

Модель индивидуального риска используется с целью сложения убытков или платежей от фиксированного числа договоров страхования или множества страховых убытков группового страхования жизни или здоровья для группы n работников. Сотрудники могут иметь различное покрытие (выплаты по договору страхования кратны окладу) и различные уровни вероятности убытков (разный возраст и состояние здоровья).

В частном случае, когда X_j одинаково распределены, модель индивидуального риска становится частным случаем модели коллективного риска с вырожденным распределением N, т.е. $P\{N = n\} = 1$

Моделирование суммарного убытка

Распределение суммы S получается из распределения N и общего распределения X_j. При таком подходе частота и убытки моделируются отдельно. Информация об этих распределениях используется для того, чтобы получить информацию об S. Альтернативой к этому подходу является просто сбор информации о S (например, суммарные убытки за каждый месяц в течение нескольких месяцев) и использование какой-либо модели распределений. Моделирование распределения N и распределений X_j отдельно имеет ряд преимуществ:

1. Ожидаемое число исков меняется по мере изменения числа полисов. Необходимо учитывать рост объема бизнеса при прогнозировании будущего числа исков на основе данных за прошлые годы.

2. Последствия инфляции (общей и/или отдельных выплат) отражаются на убытках, понесенных застрахованными лицами, и выплатах страховых компаний по искам. Такие эффекты часто маскируются, когда в страховых полисах предусмотрены франшизы и используются лимиты полисов, не завиящие от инфляции.

Моделирование суммарного убытка

3. Влияние изменения индивидуальных франшиз и лимитов легче учитывать, изменяя распределения убытков и частоту предъявления претензий по отдельности.

4. Данные, являющиеся неоднородными точки зрения франшиз и лимитов, могут быть объединены для получения гипотетического распределения размера убытков. Такой подход полезен, когда объединяются данные за несколько лет, в течение которых менялись условия страхования.

5. Модели убытков страхователей, расходов по претензиям страховщиков и расходов по претензиям перестраховщиков могут быть взаимосогласованными. Это полезно при определении последствий передачи риска перестраховщикам.

6. Форма распределения S зависит от формы распределений N и X. Понимание относительных форм важно для моделирования. Например, если распределение убытков имеет гораздо более тяжелый хвост, чем распределение частоты, форма хвоста распределения совокупных убытков будет определяться распределением индивидуальных убытков и будет нечувствительна к выбору распределения частоты.

Моделирование суммарного убытка

Раздельный анализ убытков и частоты позволят построить более точную и гибкую. Если N представляет фактическое число убытков для застрахованного, то X_j может представлять

- убытки для застрахованного,
- выплаты страховщика,
- выплаты перестраховщика,
- франшизы, выплаченные застрахованным.

В каждом случае S имеет разную интерпретацию, и распределение убытков должно быть с ним согласовано.

Далее:

N - случайная величина числа исков (число исков, иски), частота;

X_j - случайные величины индивидуальных (отдельных) потерь (просто потери, убытки).

Как правило, X_j соответствуют платежам, однако там, где нет необходимости различать платежи и убытки, используется термин потери или убытки.

S - случайная величина суммарных (совокупных) убытков.

Пример 2

Пример 2.

Опишем, как модель коллективного риска может быть использована для описания суммарных платежей в год по полисам КАСКО с франшизой 250.

1. Пусть N^L число аварий, в том числе тех, по которым убыток не превосходит франшизу. Тогда случайная величина индивидуальных потерь есть Y^L .
2. Пусть N^P - число платежей. Тогда индивидуальные убытки задаются Y^P .

Выбор модели

Нередко при подборе распределений по данным для частоты или убытков может быть несколько подходящих вариантов. Однако есть определенные предпочтения.

Желательно, чтобы распределение убытков было масштабируемым, чтобы иметь возможность перейти в другую валюту и учитывать эффект инфляции, в том числе при построении прогнозов на будущее.
Аналогичные соображения касаются распределений частоты. При росте портфеля страховой компании число исков также будет расти при сохранении формы распределения.

Модели с производящей функцией вероятности вида

$$P_N(z,\alpha) = Q(z)^{\alpha}$$

имеют ожидаемое число убытков, пропорциональное α. Увеличение бизнеса на 100%r приводит к ожидаемым искам пропорциональным $\alpha^* = (1 + r)\alpha$.

Выбор модели

Выбранная модель не должна быть зависима от продолжительности периода времени, используемом при определении частоты исков. В частности, ожидаемая частота должна быть пропорциональна длине интервала после корректировки, связанной с ростом бизнеса.
В этом случае исследование, проведенное на периоде в 10 лет, может быть использовано для получения распределения частоты на периоде 1 месяц, 1 год или любом другом.

Форма распределения для однолетнего периода та же, что и для одномесячного с точностью до значения параметра. Параметр α соответствует длине периода времени. Например, если $\alpha = 1.7$ для одного месяца, что аналогичная модель для одного года имеет $\alpha = 12 \cdot 1.7 = 20.4$.

Модифицированные в нуле распределения не имеют такой формы. Однако может оказаться желательным использовать ZM распределение по смыслу ситуации. Например, если определенная доля полисов никогда не порождает претензий из-за дублирования охвата или по другой причине, может оказаться целесообразным использовать эту же долю в будущих периодах для полиса, выбранного случайным образом.

Составное (обобщенное) распределение

Пусть S – суммарный убыток, связанный с множеством из N наблюдаемых исков $X_1, X_2, \dots, X_N$, которые являются независимыми и одинаково распределенными. Чтобы построить модель суммарного убытка, необходимо

1. На основании данных смоделировать распределение N.
2. Найти подходящее общее распределение для X_j.
3. Используя две эти модели, получить распределение S.

Пусть первые две задачи решены.

Случайная сумма случайного числа слагаемых $S = X_1 + X_2 + \dots + X_N$ имеет функцию распределения (составное распределение)

$$F_S(x) = P\{S \le x\} = \sum_{n=0}^{\infty} p_n \, P\{S \le x | N = n\} = \sum_{n=0}^{\infty} p_n F_X^{*n}(x),$$

где $F_X(x) = P\{X \le x\}$ функция распределения X_j, а $p_n = P\{N = n\}$, $F_X^{*n}(x)$ – свертка степени n функций распределения X.

Свертка распределений

Свертка степени n функций распределения X

$F_X^{*n}(x)$:

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

$$F_X^{*k}(x) = \int_{-\infty}^{\infty} F_X^{*(k-1)}(x - y)\, dF_X(y), k = 1, 2, \ldots$$

Составное распределение

Хвост распределения (функция выживания) для $x \geq 0$

$$1 - F_S(x) = \sum_{n=1}^{\infty} p_n\left(1 - F_X^{*n}(x)\right)$$

Если X непрерывная случайная величина с нулевой вероятностью неположительных значений, то

$$F_X^{*k}(x) = \int_0^x F_X^{*(k-1)}(x - y) f_X(y)\, dy, \qquad k = 2, 3, \ldots$$

$F_X^{*1}(x) = F_X(x)$. Дифференцируя, получаем функцию плотности

$$f_X^{*k}(x) = \int_0^x f_X^{*(k-1)}(x - y)\, f_X(y)\, dy, \qquad k = 2, 3, \ldots$$

Составное распределение

Если X непрерывна, то S имеет функцию плотности для $x > 0$

$$f_S(x) = \sum_{n=1}^{\infty} p_n f_X^{*n}(x)$$

и дискретную точку массы $P\{S = 0\} = p_0$ в $x = 0$. Заметим, что $P\{S = 0\} \neq f_S(0) = \lim_{x \to 0+} f_S(x)$.

Если X имеет счетное распределение,

$$F_X^{*k}(x) = \sum_{y=0}^{x} F_X^{*(k-1)}(x - y) f_X(y), \qquad x = 0, 1, \dots; \; k = 2, 3, \dots$$

Соответствующая функция вероятности

$$f_X^{*k}(x) = \sum_{y=0}^{x} f_X^{*(k-1)}(x - y) f_X(y), \qquad x = 0, 1, \dots; \; k = 2, 3, \dots$$

Составное распределение

Будем полагать, что $f_X^{*0}(0) = 1$ и $f_X^{*0}(x) = 0$ при $x \neq 0$. Тогда S имеет дискретное распределение с функцией вероятности

$$f_S(x) = P\{S = x\} = \sum_{n=0}^{\infty} p_n f_X^{*n}(x), \qquad x = 0, 1, \dots$$

Производящая функция вероятности S в силу независимости $X_1, X_2, \dots, X_n$ для фиксированного n

$$P_S(z) = E\left[z^S\right] = E\left[z^0\right] P\{N = 0\} + \sum_{n=1}^{\infty} E\left[z^{X_1+X_2+\dots+X_n} \middle| N = n\right] P\{N = n\}$$

$$= P\{N = 0\} + \sum_{n=1}^{\infty} E\left[\prod_{j=1}^{n} z^{X_j}\right] P\{N = n\} = \sum_{n=0}^{\infty} P\{N = n\} \left(P_X(z)\right)^n = E\left[P_X(z)^N\right] = P_N[P_X(z)]$$

Для производящей функции моментов аналогично $M_S(z) = P_N(M_X(z))$

Моменты составного распределения

Если $P_N(z) = P_1(P_2(z))$, т.е. N само по себе является составным распределением, $P_S(z) = P_1(P_2(P_X(z)))$.

Моменты S легко выражаются через моменты N и X_j:

$$E[S] = \mu'_{S1} = \mu'_{(N1)}\mu'_{X1} = E[N]E[X];$$

$$Var[S] = \mu_{S2} = \mu'_{N1}\mu_{X2} + \mu_{N2}(\mu'_{X1})^2 = E[N]Var[X] + Var[N]E^2[X];$$

$$E[(S - E[S])^3] = \mu_{S3} = \mu'_{N1}\mu_{X3} + 3\mu_{N2}\mu'_{X1}\mu_{X2} + \mu_{N3}(\mu'_{X1})^3$$

Пример 3

Наблюдаемые среднее и стандартное отклонение для числа исков и индивидуальных потерь равны, соответственно: 6.7 и 2.3; 179 247 и 52 141.

Определим среднее и стандартное отклонение суммарных убытков:

$$E[S] = 6.7 \cdot 179\,247 \;=\; 1\,200\,955$$

$$Var[S] = 6.7 \cdot (52\,141)^2 + (2.3)^2 \cdot 179\,247^2 \;=\; 1.88180 \cdot 10^{11}$$

Таким образом, среднее и стандартное отклонение суммарных убытков равны, соответственно, 1200955 и 433797

Пример 4 (продолжение Примера 3)

Используя в качестве аппроксимаций суммарных убытков нормальное и логнормальное распределения, вычислим вероятность того, что убытки превысят 140% ожидаемого значения, т.е.

$$P\{S > 1.40 \cdot 1\,200\,955\} = P\{S > 1\,681\,337\}.$$

Для нормального распределения

$$P\{S > 1\,681\,337\} = P\left\{\frac{S - E[S]}{\sqrt{Var[S]}} > \frac{1\,681\,337 - 1\,200\,955}{433797}\right\} = P\{Z > 1.107\} = 1 - \Phi(1.107) = 0.134$$

Пример 4 (продолжение Примера 3)

Для логнормального распределения

$$E[S] = \exp\left(\mu + \frac{1}{2}\sigma^2\right), \qquad E\left[S^2\right] = \exp\left(2\mu + 2\sigma^2\right)$$

Приравнивая эти величины $1.200955 \cdot 10^6$ и $1.88180 \cdot 10^{11} + \left(1.200955 \cdot 10^6\right)^2 = 1.63047 \cdot 10^{12}$ и логарифмируя, получим два уравнения, из которых $\mu = 13.93731$ и $\sigma^2 = 0.1226361$.Тогда

$$P\{S > 1\,681\,337\} = 1 - \Phi\left(\frac{\ln 1\,681\,337 - 13.93731}{(0.1226361)^{0.5}}\right) = 1 - \Phi(1.135913) = 0.128.$$

Нормальное распределение обеспечивает хорошую аппроксимацию, когда $E[N]$ большое. В частности, если N имеет Пуассоновское, биномиальное или отрицательное биномиальное распределения, то согласно ЦПТ при стремлении λ, m и r, соответственно, к бесконечности, распределение S будет стремиться к нормальному.

В рассматриваемом примере $E[N]$ мало, поэтому распределение S, скорее всего, будет асимметрично. В таком случае логнормальное распределение дает лучшую аппроксимацию.

Страхование стоп-лосс

Частно к суммарным убыткам применяется франшиза. Когда убытки возникают у страхователя, это называется страховым покрытием, а когда убытки возникают у страховой компании, это называется перестраховочным покрытием. Последний вариант является распространенным методом для страховой компании защитить себя от неблагоприятного года (в отличие от защиты от единичного, очень крупного иска).

Определение 3. Страхование суммарных убытков с применением франшизы называется страхованием стоп-лосс. Ожидаемая стоимость такого страхования называется нетто-премией стоп-лосс и она равна $E[(S-d)_+]$, где d - франшиза.

Для совокупного распределения

$$E[(S-d)_+] = \int_d^\infty (1 - F_S(x))\,dx$$

Если распределение является непрерывным при $x > d$, то

$$E[(S-d)_+] = \int_d^\infty (x-d) f_S(x)\,dx$$

Аналогично для дискретного случая

$$E[(S-d)_+] = \sum_{x>d}^{\infty} (x-d) f_S(x)\,dx$$

Линейная интерполяция нетто-премии стоп-лосс

Если есть интервал, в котором значения суммарного убытка отсутствуют, то можно воспользоваться следующей формулой.

Теорема 1. Предположим, что $P\{a < S < b\} = 0$. Тогда для $a \leq d \leq b$

$$E[(S - d)_+] = \frac{b - d}{b - a} E[(S - a)_+] + \frac{d - a}{b - a} E[(S - b)_+].$$

Таким образом, нетто-премия стоп лосс может быть вычислена с помощью линейной интерполяции.

Доказательство. По предположению $F_S(x) = F_S(a), a \leq x < b$. Тогда

$$E[(S - d)_+] = \int_d^\infty \big(1 - F_S(x)\big)\, dx = \int_a^\infty \big(1 - F_S(x)\big)\, dx - \int_a^d (1 - F_S(x))\, dx$$

$$= E[(S - a)_+] - \int_a^d \big(1 - F_S(a)\big)\, dx = E[(S - a)_+] - (d - a)(1 - F_S(a)).$$

Полагая $d = b$,

$$E[(S - b)_+] = E[(S - a)_+] - (b - a)\big(1 - F_S(a)\big) \Rightarrow 1 - F_S(a) = \frac{(E[(S - a)_+] - E[(S - b)_+])}{(b - a)}.$$

Подставляя в предыдущее равенство, получаем требуемое равенство.

Дискретный случай

Теорема 2. Пусть $P\{S = kh\} = f_k \geq 0$ для фиксированного $h > 0,\ k = 0, 1, \dots,$ и $P\{S = x\} = 0$ для всех остальных x. Тогда при условии $d = jh, j \in \mathbb{N} \cup \{0\}$,

$$E[(S - d)_+] = h \sum_{m=0}^{\infty} (1 - F_S(m + j)h).$$

Доказательство.

$$E[(S - d)_+] = \sum_{x>d} (x - d) f_S(x) = \sum_{k=j}^{\infty} (kh - jh) f_k = h \sum_{k=j}^{\infty} \sum_{m=0}^{k-j-1} f_k = h \sum_{m=0}^{\infty} \sum_{k=m+j+1}^{\infty} f_k$$

$$= h \sum_{m=0}^{\infty} (1 - F_S((m + j)h)).$$

В дискретном случае имеет место следующая рекурсия (при равноотстоящий дискретных значениях).
Следствие. В условиях Теоремы 2 $E[(S - (j + 1)h)_+] = E[(S - jh)_+] - h(1 - F_S(jh))$
Этим свойством легко воспользоваться, так как при $d = 0$ $E[(S - 0)_+] = E[S] = E[N]E[X]$

Распределение Твиди (Tweedie)

Распределение Твиди соединяет две концепции.

1. Для определенных значений параметров оно является сложным пуассоновским распределением с гамма-распределением убытков. Таким образом, эта модель может подходить для описания совокупных убытков.
2. Оно является представителем семейства линейных экспоненциальных распределений. Поэтому распределение оказывается полезным при построении обобщенных линейных моделей, связывающих иски с характеристиками держателей полисов.

Рассмотрим сложное пуассоновское распределение. Пусть N имеет распределение Пуассона со средним λ, а X имеет гамма распределение с параметрами α и γ (этот параметр используется вместо θ). Тогда для $S = X_1 + X_2 + \cdots + X_N$

$$P\{S = 0\} = P\{N = 0\} = e^{-\lambda},$$

Так как свертка степени n гамма распределения также является гамма-распределением с параметром масштаба (α), умноженным на n, то

$$f_S(s) = \sum_{n=1}^{\infty} \frac{e^{-\lambda}\,\lambda^n}{n!} \frac{s^{n\alpha-1} e^{-s/\gamma}}{\Gamma(n\alpha)\gamma^{n\alpha}} .$$

$$E[S] = E[N]E[X] = \lambda\alpha\gamma;\ Var[S] = E[N]Var[X] + Var[N]E^2[X] = \lambda\alpha\gamma^2 + \lambda(\alpha\gamma)^2 = \lambda\gamma^2\alpha(\alpha + 1)$$

Распределение Твиди

Распределение Твиди получается в результате следующей параметризации:

$$\lambda = \frac{\mu^{2-p}}{\phi(2 - p)}, \alpha = \frac{2 - p}{p - 1}, \quad \gamma = \phi(p - 1)\mu^{p-1}$$

где $1 < p < 2, \mu > 0, \phi > 0$. Замена приводит мат. ожидание и дисперсию к виду

$$E[S] = \mu, Var[S] = \phi\mu^p.$$

Распределение Твиди, являясь представителем линейного экспоненциального семейства, имеет форму вида

$$f(x;\ \theta) = \frac{p(x)e^{r(\theta)x}}{q(\theta)},$$

где $p(x)$ может иметь параметры, отличные от θ.

Распределение Твиди

Общая форма линейного экспоненциального распределения может быть записана в виде

$$f(x;\ \theta) = p(x)\exp\left(\frac{\theta y\ -\ a(\theta)}{\phi}\right).$$

К такому виду приводится сложное Пуассон-гамма распределение с учетом указанной замены переменных. Новый параметр ϕ называется параметром дисперсии. Он добавляет определенную гибкость в связь между дисперсией и мат. ожиданием.

Распределение Твиди существует для всех неотрицательных p, порождая семейство распределений:

нормальное (p = 0), Пуассона (p = 1), гамма (p = 2), обратное гауссовское (p = 3).

Заметим, что при p = 1 Var$[X]$= $\phi\mu$. Поэтому, чтобы получить распределение Пуассона, нужно положить ϕ = 1. При ϕ>1 получаем распределение, которое называется распределением Пуассона с избыточной дисперсией. Это распределение часто используются при построении обобщенных линейных моделей.

Аналитические результаты

В большинстве случаев выбора распределений N и X_j сложное распределение получается только численно. Однако для некоторых комбинаций получены простые аналитические результаты, заметно уменьшающие объем вычислений.

Пример 5.

Отрицательное биномиальное-экспоненциальное. Определим распределение S при условии отрицательного биномиального распределения с целым r для частоты страховых случаев и экспоненциального распределения для убытков.

Производящая функция моментов для S

$$M_S(z) = P_N(M_X(z)) = P_N((1 - \theta z)^{-1}) = \left(1 - \beta\left((1 - \theta z)^{-1} - 1\right)\right)^{-r}$$

Перепишем в виде

$$M_S(z) = \left(1 + \frac{\beta}{1 + \beta}\left((1 - \theta(1 + \beta)\, z)^{-1} - 1\right)\right)^{r}$$

Таким образом, $M_S(z) = P_N^*(M_X^*(z))$,

где $P_N^*(z) = \left(1 + \frac{\beta}{1 + \beta}(z - 1)\right)^{r}$ - производящая функция вероятности биномиального распределения с параметрами r и $\frac{\beta}{1+\beta}$ и $M_X^*(z)$ – производящая функция моментов экспоненциального распределения со средним $\theta(1 + \beta)$.

Пример 5

Это преобразование приводит вычисление функции распределения к конечной сумме

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

При $r = 1$ S имеет сложное геометрическое распределение

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

При этом $P\{S = 0\} = F_S(0) = (1 + \beta)^{-1}$, и так как $F_S(x)$ дифференцируема, существует функция плотности

$f_S(x) = F_S'(x)$, $x > 0$:

$$f_S(x) = \frac{\beta}{\theta(1 + \beta)^2} \exp\left(-\frac{x}{\theta(1 + \beta)}\right)$$

Таким образом, при $r = 1$ S имеет точку массы вероятности $(1 + \beta)^{-1}$ в нуле и экспоненциально убывающую плотность

Пример 6

Определим функцию распределения S для произвольного сложного распределения с экспоненциальным распределением ущерба. Производящая функция моментов суммы n независимых экспоненциально распределенных случайных величин со средним θ

$$M_{X_1+X_2+\cdots+X_n}(z) = (1 - \theta z)^{-n},$$

которая является производящей функцией моментов гамма распределения с функцией $F_X^{*n}(x) = \Gamma\left(n; \frac{x}{\theta}\right)$.

$$F_S(x) = p_0 + \sum_{n=1}^{\infty} p_n \Gamma\left(n; \frac{x}{\theta}\right)$$

Функция плотности

$$f_S(x) = \sum_{n=1}^{\infty} p_n \frac{x^{n-1} e^{-x/\theta}}{\theta^n \Gamma(n)}$$

Так как для целых n $\Gamma(n;\ x) = 1 - \sum_{j=0}^{n-1} \frac{x^j e^{-x}}{j!}, n = 1, 2, 3 \ldots$, то

$$F_S(x) = 1 - \sum_{n=1}^{\infty} p_n \sum_{j=0}^{n-1} \frac{\left(\frac{x}{\theta}\right)^j e^{-x/\theta}}{j!}, x \geq 0$$

Пример 6

Меняя порядок суммирования, получаем

$$F_S(x) = 1 - e^{-x/\theta} \sum_{j=0}^{\infty} \frac{(x/\theta)^j}{j!} \sum_{n=j+1}^{\infty} p_n = 1 - e^{-x/\theta} \sum_{j=0}^{\infty} \bar{P}_j \frac{(x/\theta)^j}{j!}, \qquad x \geq 0$$

где $\bar{P}_j = \sum_{n=j+1}^{\infty} p_n, j = 0, 1, \ldots$

Для распределений, которые приписывают положительную вероятность всем неотрицательным значениям, сумма может быть оценена с заданной точностью при достаточном числе слагаемых. Распределения, для которых $P\{N > n^*\} = 0$, сумма становится конечной.

Например, при биномиальном распределении частоты

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

Обобщенное распределение Пуассона

В договоре группового страхования у каждого участника есть составная модель Пуассона для совокупных убытков, и необходимо найти распределение общих совокупных убытков. Аналогичным образом можно оценивать совокупные убытки от нескольких независимых направлений бизнеса. В таких случаях совокупное распределение можно получить без вычисления каждого члена группы.

Теорема 3. Пусть S_j имеет сложное распределение Пуассона с параметром λ_j, и распределение ущерба имеет функцию распределения $F_j(x)$, $j = 1, 2, \dots, n$. Также предположим, что $S_1, S_2, \dots, S_n$ независимы. Тогда $S = S_1 + \cdots + S_n$ имеет сложное распределение Пуассона с параметром $\lambda = \lambda_1 + \cdots + \lambda_n$ и распределением ущерба

$$F(x) = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} F_j(x)$$

Доказательство Теоремы 3

Доказательство. Пусть $M_j(t)$ – производящая функция моментов $F_j(x)$ для $j = 1, 2, \dots, n$. Тогда S_j имеет производящую функцию моментов

$$M_{S_j}(t) = E[e^{tS_j}] = \exp(\lambda_j (M_j(t) - 1))$$

и в силу независимости S_j , производящая функция моментов S равна

$$M_S(t) = \prod_{j=1}^{n} M_{S_j}(t) = \prod_{j=1}^{n} \exp\left(\lambda_j \left(M_j(t) - 1\right)\right) = \exp\left(\sum_{j=1}^{n} \lambda_j M_j(t) - \lambda\right) = \exp\left(\lambda\left(\sum_{j=1}^{n} \frac{\lambda_j}{\lambda} M_j(t) - 1\right)\right)$$

Так как

$$\sum_{j=1}^{n} \frac{\lambda_j}{\lambda} M_j(t) -$$

производящая функция

$$F(x) = \sum_{j=1}^{n} \frac{\lambda_j}{\lambda} F_j(x),$$

$M_S(t)$ - производящая функция моментов составного распределения Пуассона.

Пример 7

Полис **A** имеет сложное распределение Пуассона с λ = 2 и вероятности 0.6 для платежа 1 и 0.4 для платежа 2.
Полис **B** имеет сложное распределение Пуассона с λ = 1 и вероятности 0.7 платежа 1 и 0.3 платежа 3. Определим вероятность того, что суммарный платеж от обоих полисов равен 2. (Эта задача достаточно проста и не требует применения Теоремы 3.)
Используя Теорему 3, получим, что суммарный платеж имеет сложное распределение Пуассона *с* λ = 2 + 1 = 3 и распределение убытка

$$f_X(1) = \frac{2}{3}\cdot 0.6 + \frac{1}{3}\cdot 0.7 = 0.63333, \qquad f_X(2) = \frac{2}{3}\cdot 0.4 + \frac{1}{3}\cdot 0.0 = 0.26667, \qquad f_X(3) = \frac{2}{3}\cdot 0.0 + \frac{1}{3}\cdot 0.3 = 0.1$$

Тогда $P\{S = 2\} = P\{N = 1, X_1 = 2\} + P\{N = 2, X_1 = 1, X_2 = 1\} = e^{-3}\cdot\frac{3}{1}\cdot 0.26667 + e^{-3}\cdot\frac{3^2}{2}\cdot 0.63333\cdot 0.63333$

$$= 2.605e^{-3} = 0.12970.$$

Рассматривая каждый полис отдельно, получаем. Для полиса A

$$P\{S_A = 0\} = e^{-2};\ P\{S_A = 1\} = e^{-2}\cdot\frac{2}{1}\cdot 0.6 = 1.2e^{-2};\ P\{S_A = 2\} = e^{-2}\cdot\frac{2}{1}\cdot 0.4 + e^{-2}\cdot\frac{2^2}{2}\cdot 0.6^2 = 1.52e^{-2}.$$

Для полиса B $\quad P\{S_B = 0\} = e^{-1};\ P\{S_B = 1\} = e^{-1}\cdot 0.7, P\{S_B = 2\} = e^{-1}\cdot\frac{1}{2}\cdot 0.7^2 = 0.245e^{-1}.$

Тогда $P\{S_A + S_B = 2\} = e^{-2}\cdot 0.245e^{-1} + 1.2e^{-2}\cdot 0.7e^{-1} + 1.52e^{-2}\cdot e^{-1} = 2.605e^{-3} = 0.12970.$

Численное нахождение функции распределения совокупного ущерба

Вычисление функции распределения $F_S(x) = \sum_{n=0}^{\infty} p_n F_X^{*n}(x)$ или соответствующей функции плотности в общем случае сложная задача.

1. Можно использовать аппроксимирующее распределение для того, чтобы избежать непосредственного вычисления суммы сверток с весами-вероятностями.

Этот метод был использован в примере 4, где для оценки параметров распределений использовался метод моментов. Метод прост, однако имеет значительные недостатки: не всегда есть возможность определить качество аппроксимации, выбор различных аппроксимирующих распределений может привести к существенно различающимся результатам, особенно в правом хвосте распределения; чем больше моментов используется, тем точнее будет аппроксимация, однако после четвертого момента не останется подходящих распределений (так как большего числа параметров нет); приближенное распределение может не отражать особенности истинного распределения.

Например, когда распределение потерь непрерывно и существует максимально возможная претензия (при наличии лимита полиса), распределение ущерба может иметь массу в максимальной точке. Истинное распределение совокупных убытков является смешанным с пиками, кратными максимуму, соответствующими 1, 2, 3 и т.д. максимальными искам. Когда эти пики являются большими, они оказывают значительное влияние на вероятности в окрестности таких кратных значений. Подобные скачки в функции распределения совокупных убытков не могут быть заменены гладким аппроксимирующим распределением.

Численное вычисление интеграла свертки

2. Второй метод состоит в прямом вычислении соответствующей функции плотности. Наиболее сложной (вычислительно) частью является оценка свертки степени n распределений ущерба для n = 2, 3, 4,…. Свертки численно оцениваются с использованием

$$F_X^{*k}(x) = \int_{0-}^{x} F_X^{*(k-1)}(x - y)\, dF\_X(y).$$

(в общем случае интеграл берется по всей прямой).

Этот интеграл записан в форме Лебега-Стилтьеса, так как возможны скачки функции распределения $F_X(x)$ в нуле и других точках. (Достаточно интерпретировать $\int g(y)\, dF_X(y)$ как интеграл $g(y) f_X(y)$ по y в тех точках, где X имеет непрерывное распределение, а затем добавить $g(y_i)\, P\{X = y_i\}$ в тех точках, где $P\{X = y_i\} > 0$.)

Для вычисления интеграла обычно требуются методы численного интегрирования. Из-за того, что интегрируемую функцию распределения необходимо вычислить для всех возможных значений x, такой подход быстро становится вычислительно затратным. Когда распределение ущерба дискретно, вычисления сводятся к многочисленным умножениям и сложениям. Для непрерывных распределений ущерба простым способом избежать технических проблем является замена распределения ущерба дискретным распределением, определяемым в точках, кратных удобной денежной единицы (например,1000).

Дискретизация

Денежную единицу можно сделать достаточно маленькой, чтобы учесть скачки при максимальных страховых суммах. Пик должен быть кратен денежной единице, чтобы он был расположен в точке дискретизации. При уменьшении единицы измерения дискретная функция распределения должна приближаться к истинной функции распределения. Самый простой подход - округлить все суммы до ближайшего значения, кратного денежной единице (например, до ближайшей 1000).

Когда распределение ущерба определено для неотрицательных целых чисел $0, 1, 2, \dots$, вычисление $f_X^{*k}(x)$ для целого x требует $x + 1$ умножений. Чтобы получить распределение от x = 0 до x = n требуется порядка $O(n^3)$ умножений. Когда максимальное значение n, для которого рассчитывается совокупное распределение утверждений, велико, количество вычислений быстро становится большим даже для быстрых компьютеров. В реальных приложениях n может легко достигать 1000, т.е. требуется примерно 10^9 умножений.

Далее, если $P\{X = 0\} > 0$ и распределение частот неограниченно, требуется бесконечное число вычислений для получения одной вероятности. Причина этого в том, что $F_X^{*n}(x) > 0$ для всех n и всех x, и поэтому сумма, определяющая функцию распределения, содержит бесконечное число членов. Когда $P\{X = 0\} = 0$, мы имеем $F_X^{*n}(x) = 0$ для $n > x$, и поэтому сумма будет иметь не более $x + 1$ положительных слагаемых.

Рекурсивный метод

Предположим, что распределение ущерба $f_X(x)$ определено в x = 0, 1, 2,…,m , представляя значения, кратные некоторой денежной единице. Число m - наибольший возможный платеж; может быть бесконечным. Предположим, что распределение частоты p_k является представителем класса $(a, b, 1)$ и поэтому

$$p_k = \left(a + \frac{b}{k}\right) p_{k-1}, k = 2, 3, 4, \dots$$

Теорема 4. Для класса $(a, b, 1)$

$$f_S(x) = \frac{(p_1 - (a + b)p_0) f_X(x) + \sum_{y=1}^{\min(x,m)} \left(a + \frac{by}{x}\right) f_X(y) f_S(x - y)}{\left(1 - a f_X(0)\right)}$$

Следствие. Для класса $(a, b, 0)$

$$f_S(x) = \frac{\sum_{y=1}^{\min(x,m)} \left(a + \frac{by}{x}\right) f_X(y) f_S(x - y)}{\left(1 - a f_X(0)\right)}$$

Когда распределение ущерба имеет максимальное возможное значение m, сумма в рекурсии ограничена не более чем m ненулевыми компонентами. В этом случае вычислительная сложность имеет порядок $O(x)$.

Рекурсия для распределения Пуассона

Заметим, что когда распределение ущерба не имеет вероятности в нуле, знаменатель равен 1.

Для распределения Пуассона

$$f_S(x) = \frac{\lambda}{x} \sum_{y=1}^{\min(x,m)} y f_X(y) f_S(x - y)$$

Начальное значение для выполнения рекурсивной схемы $f_S(0) = P_N(f_X(0))$.

В случае распределения Пуассона

$$f_S(0) = e^{-\lambda}(1 - f_X(0)).$$

Пример 8 (продолжение примера 7)

Вычислим вероятности в 0, 1, 2, 3 и 4, используя рекурсивную формулу.

Чтобы ей воспользоваться, определим $f_X(0) = 0, f_X(1) = 0.63333, f_X(2) = 0.26667, f_X(3) = 0.1.$

Тогда $f_S(0) = e^{(-3)(1-0)} = e^{-3}$. Рекурсивная формула при $\lambda = 3$

$$f_S(x) = \frac{3}{x}\sum_{y=1}^{3} y f_X(y) f_S(x - y) = \frac{3}{x}(1 \cdot 0.63333\, f_S(x-1) + 2 \cdot 0.26667\, f_S(x-2) + 3 \cdot 0.1 f_S(x-3))$$

(когда аргумент $f_S(x)$ отрицательный, это значение равно нулю). Тогда

$$f_S(1) = 3 \cdot 0.63333e^{-3} = 1.9e^{-3};$$

$$f_S(2) = \frac{3}{2}(0.63333 \cdot 1.9e^{-3} + 0.53333e^{-3}) = 2.605e^{-3};$$

$$f_S(3) = \frac{3}{3}(0.63333 \cdot 2.605e^{-3} + 0.53333 \cdot 1.9e^{-3} + 0.3e^{-3}) = 2.96315e^{-3};$$

$$f_S(4) = \frac{3}{4}\left(0.63333 \cdot 2.96315e^{-3} + 0.53333 \cdot 2.605e^{-3} + 0.3 \cdot 1.9e^{-3}\right) = 3.83598e^{-3}.$$

Пример 9

Сложное распределение имеет ZM биномиальное распределение с m = 3, q = 0.3 и $p_0^M = 0.4$. Индивидуальные платежи равны 0, 50 и 150 с вероятностями 0.3, 0.5 и 0.2, соответственно. Используя рекурсивную формулу, определим распределение S.

Так как платежи должны быть равноотстоящими, подходящая единица равна 50. Тогда $f_X(0) = 0.3, f_X(1) = 0.5, f_X(2) = 0, f_X(3) = 0.2$. Начальное значение равно

$$f_X(0) = 0.4 + 0.6\frac{\left(\left(1 + 0.3(-0.7)\right)^3 - 0.7^3\right)}{(1 - 0.7^3)} = 0.53702.$$

Основные величины

$$a = -\frac{0.3}{0.7} = -\frac{3}{7}, \qquad b = (3 + 1)\cdot\frac{0.3}{0.7} = \frac{12}{7}, \qquad p_1 = \frac{3\cdot 0.3\cdot(0.7)^2\cdot 0.6}{1 - 0.7^3} = 0.40274$$

Рекурсивная формула

$$f_S(x) = \frac{(p_1 - (a + b)p_0)f_X(x)\sum_{y=1}^{3}\left(-\frac{3}{7} + \frac{12y}{7x}\right)f_X(y)f_S(x - y)}{1 - \left(-\frac{3}{7}\right)\cdot 0.3}$$

$$= \frac{70}{79}\left[\left(0.40274 - \frac{3.6}{7}\right)f_X(x) + \left(-\frac{3}{7} + \frac{12}{7x}\right)0.5f_S(x - 1) + \left(-\frac{3}{7} + \frac{36}{7x}\right)0.2f_S(x - 3)\right]$$

Пример 9

Первые несколько значений равны

$$f_S(1) = \frac{70}{79}\left[-0.11155 \cdot 0.5 \ + \left(-\frac{3}{7} + \frac{12}{7}\right) \cdot 0.5 \cdot \ 0.53702\right] = \ 0.25648;$$

$$f_S(2) = \frac{70}{79}\left(-\frac{3}{7} + \frac{12}{14}\right) \cdot \ 0.5 \cdot 0.25648 \ = \ 0.04870;$$

$$f_S(3) = \frac{70}{79}\left[-0.11155 \cdot \ 0.2 + \left(-\frac{3}{7} + \frac{12}{21}\right) \cdot \ 0.5 \cdot \ 0.04870 + \left(-\frac{3}{7} + \frac{36}{21}\right) \cdot \ 0.2 \cdot 0.53702\right] \ = \ 0.105.7;$$

$$f_S(4) = \frac{70}{79}\left[\left(-\frac{3}{7} + \frac{12}{28}\right) \cdot \ 0.5 \cdot \ 0.10567 + \left(-\frac{3}{7} + \frac{36}{28}\right) \cdot \ 0.2 \cdot \ 0.25648\right] = \ 0.03896.$$

Остальные значения получаются аналогично.

Применение сложных моделей частот

Когда распределение частоты представляется в виде составного распределения (например, распределение Неймана типа A, т.е. Пуассон-Пуассон, Пуассон-обратный гауссиан), включающего только распределения классов $(a, b, 0)$ или $(a, b, 1)$, рекурсивная формула может быть использована несколько раз для получения распределения суммарного убытка.

Если распределение частоты может быть записано в виде

$$P_N(z) = P_1(P_2(z));$$

тогда распределение суммарного убытка имеет производящую функцию вероятности

$$P_S(z) = P_N(P_X(z)) = P_1(P_2(P_X(z))),$$

которая может быть переписана

$$P_S(z) = P_1(P_{S_1}(z)),$$

где $P_{S_1}(z) = P_2(P_X(z))$.

Последнее выражение имеет ту же форму, что и распределение суммарного ущерба. Таким образом, если $P_2(z)$ принадлежит классу $(a, b, 0)$ или $(a, b, 1)$, распределение S_1 может быть вычислено с помощью рекурсивной формулы. Результирующее распределение является распределением «ущерба». Таким образом, второе применение рекурсии к сложному распределению приводит к распределению S.

Пример 10

Число убытков имеет Пуассон-ETNB распределение, параметр распределения Пуассона λ = 2, параметры ETNB распределения β = 3 и r = 0.2. Размер убытка имеет распределение 0.3, 0.5 и 0.2 в точках 0, 10 и 20, соответственно. Определим распределение суммарных убытков рекурсивно.

N имеет производящую функцию вероятности $P_N(z) = P_1(P_2(z))$, где $P_1(z)$ и $P_2(z)$ – производящие функции распределений Пуассон и ETNB, соответственно. Тогда суммарный убыток имеет производящую функцию вероятностей $P_S(z) = P_1(P_{S_1}(z))$, где $P_{S_1}(z) = P_2(P_X(z))$ – производящая функция вероятности составного ETNB распределения.

Вычислим распределение S_1. Считая единицу измерения ущерба равной 10, имеем

$f_X(0) = 0.3;\ f_X(1) = 0.5, f_X(2) = 0.2.$ Для того, чтобы воспользоваться рекурсией, начнем с

$$f_{S_1}(0) = P_2(f_X(0)) = \frac{\left(1 + \beta\left(1 - f_X(0)\right)\right)^{-r} - (1 + \beta)^{-r}}{1 - (1 + \beta)^{-r}} = \frac{\left(1 + 3(1 - 0.3)\right)^{-0.2} - (1 + 3)^{-0.2}}{1 - (1 + 3)^{-0.2}}$$
$$= 0.16369$$

Оставшиеся значения $f_{S_1}(x)$ могут быть получены из рекурсивной формулы заменой S на S_1

В этом случае $a = \frac{3}{1+3} = 0.75;\ b = (0.2 - 1)a = -0.6;\ p_0 = 0, p_1 = \frac{0.2\cdot 3}{(1+3)^{0.2+1} - (1+3)} = 0.46947.$

Пример 10

Тогда рекурсивная формула принимает вид

$$f_{S_1}(x) = \frac{(0.46947 - (0.75 - 0.6) \cdot 0) f_X(x) + \sum_{y=1}^{x} \left(0.75 - \frac{0.6y}{x}\right) f_X(y) f_{S_1}(x - y)}{1 - 0.75 \cdot 0.3}$$

$$= 0.60577 f_X(x) + 1.29032 \sum_{y=1}^{x} \left(0.75 - 0.6\frac{y}{x}\right) f_X(y) f_{S_1}(x - y)$$

Первые несколько вероятностей равны

$$f_{S_1}(1) = 0.60577 \cdot 0.5 + 1.29032\left(0.75 - 0.6\left(\frac{1}{1}\right)\right) 0.5 \cdot 0.16369 = 0.31873;$$

$$f_{S_1}(2) = 0.60577 \cdot 0.2 + 1.2903\left(\left(0.75 - 0.6 \cdot \frac{1}{2}\right) 0.5 \cdot 0.31873 + \left(0.75 - 0.6 \cdot \frac{2}{2}\right) 0.5 \cdot 0.16369\right) = 0.22002;$$

Пример 10

$$f_{S_1}(3) = 1.29032\left(\left(0.75 - 0.6 \cdot \frac{1}{3}\right)0.5 \cdot 0.22002 + \left(0.75 - 0.6 \cdot \frac{2}{3}\right)0.2 \cdot 0.31873\right) = 0.10686;$$

$$f_{S_1}(4) = 1.29032\left(\left(0.75 - 0.6 \cdot \frac{1}{4}\right)0.5 \cdot 0.10686 + \left(0.75 - 0.6 \cdot \frac{2}{4}\right)0.2 \cdot 0.22002\right) = 0.06692.$$

Теперь перейдем к оценке распределения S с помощью производящей функции вероятности составного распределения Пуассона

$$P_S(z) = P_1(P_{S_1}(z)) = e^{\lambda(P_{S_1}(z)-1)}.$$

Таким образом, распределение $f_{S_1}(x), x = 0, 1, 2, \ldots$ является «вторичным» или распределением «размера убытков» при применении рекурсивной формулы для составного распределения Пуассона. Поэтому

$$f_S(0) = P_S(0) = e^{\lambda(P_{S_1}(0)-1)} = e^{\lambda(f_{S_1}(0)-1)} = e^{2(0.16369-1)} = 0.18775.$$

Пример 10

Первые несколько вероятностей равны

$$f_S(1) = 2 \cdot \frac{1}{1} \cdot 0.31873 \cdot 0.18775 = 0.11968;$$

$$f_S(2) = 2 \cdot \frac{1}{2} \cdot 0.31873 \cdot 0.11968 + 2 \cdot \frac{2}{2} \cdot 0.22002 \cdot 0.18775 = 0.12076;$$

$$f_S(3) = 2 \cdot \frac{1}{3} \cdot 0.31873 \cdot 0.12076 + 2 \cdot \frac{2}{3} \cdot 0.22002 \cdot 0.11968 + 2 \cdot \frac{3}{3} \cdot 0.10686 \cdot 0.18775 = 0.10090;$$

$$f_S(4) = 2 \cdot \frac{1}{4} \cdot 0.31873 \cdot 0.10090 + 2 \cdot \frac{2}{4} \cdot 0.22002 \cdot 0.12076 + 2 \cdot \frac{3}{4} \cdot 0.10686 \cdot 0.11968 + 2 \cdot \frac{4}{4} \cdot 0.06692 \cdot 0.18775 = 0.08696.$$

Проблемы переполнения и недостаточной точности

Рекурсия начинается с вычисления $P\{S = 0\} = P_N(f_X(0))$.

Для больших страховых портфелей эта вероятность очень мала, так что не может быть представлена в машинной памяти.

В таком случае она становится равной нулю, и рекурсивная формула перестает быть применимой. Есть несколько путей решения этой проблемы.

Один из самых простых способов – начинать с произвольных значений $f_S(0), f_S(1), \dots, f_S(k)$, где k находится достаточно далеко слева в распределении, так что истинное значение $F_S(k)$ все еще незначительно.

Обычно достаточно установить k в точку, лежащую на шесть стандартных отклонений левее среднего значения. Рекурсия используется для генерации значений распределения с этим набором начальных значений до тех пор, пока найденные значения не будут стабильно меньше $f_S(k)$. Затем «вероятности» суммируются и делятся на сумму так, чтобы «истинные» вероятности давали в сумме 1.

Насколько маленьким должно быть значение k для конкретной задачи, определяется методом проб и ошибок.

Проблемы переполнения и недостаточной точности

Другим методом получения вероятностей, когда начальное значение слишком мало, является выполнение вычислений для субпортфеля. Например, для распределения Пуассона со средним λ найдем значений

$\lambda^* = \frac{\lambda}{2^n}$ такое, что вероятность в нуле не равна машинному нулю, когда λ^* используется в качестве среднего распределения Пуассона.

Рекурсивная формула теперь используется для совокупного распределения с λ^*. Если $P_*(z)$ является производящей функцией вероятности совокупных убытков со средним распределения Пуассона λ^*, тогда $P_S(z) = \left(P_*(z)\right)^{2^n}$. Следовательно, можно последовательно получить распределения с производящими функциями вероятностей $\left(P_*(z)\right)^2, \left(P_*(z)\right)^4, \left(P_*(z)\right)^8, \ldots, \left(P_*(z)\right)^{2^n}$ путем свертки результата на каждом этапе с самим собой. Этот подход требует n дополнительных сверток при выполнении вычислений, но не предполагает никаких приближений. Это может быть выполнено для любых частотных распределений, которые замкнуты относительно свертки. Для отрицательного биномиального распределения аналогичная процедура начинается с $r^* = \frac{r}{2^n}$. Для биномиального распределения параметр m должен быть целым. В этом случае $m^* = \left\lfloor \frac{m}{2^n} \right\rfloor$.

Когда n сверток выполнено, необходимо выполнить рекурсию для параметра $m - m^* 2^n$. Этот результат затем сворачивается с результатом n сверток. Для составных распределений частоты только первичное распределение должно быть замкнуто относительно свертки.

Численная устойчивость

Любая рекурсивная формула требует точного вычисления значений, поскольку каждое такое значение будет использоваться при вычислении последующих значений.

Рекурсивные схемы подвержены риску распространения ошибок по всем последующим значениям и риску потенциального сбоя.

В рекурсивной формуле ошибки вносятся путем округления на каждом этапе, поскольку компьютеры представляют числа с конечным числом значащих цифр.

Вопрос устойчивости заключается в следующем: насколько быстро растут ошибки в вычислениях по мере подстановки вычисленных значений в последовательные вычисления?

Можно сделать некоторые общие выводы.

Численная устойчивость

Ошибки вносятся в последующие значения при суммировании

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

В правом хвосте распределения S эта сумма положительна (или, по крайней мере, неотрицательна), и последующие значения суммы будут уменьшаться.

Сумма останется положительной, даже с учетом ошибок округления, когда каждый из трех коэффициентов в каждом члене суммы будет положительным. В этом случае рекурсивная формула устойчива, так как приводит к относительным ошибкам, которые растут небыстро.

Для распределений, основанных на распределении Пуассона и отрицательных биномиальных, множители в каждом члене всегда положительны. Однако для биномиального распределения сумма может иметь отрицательные члены, поскольку a отрицательно, b положительно, а y/x - положительная функция, не превышающая 1. В этом случае отрицательные члены могут привести к тому, что последовательные значения будут знакочередующимися. Когда это происходит, бессмысленные результаты сразу становятся очевидными. Хотя на практике это случается нечасто, необходимо знать о такой возможности в моделях, основанных на биномиальном распределении.

Непрерывное распределение ущерба

Рекурсивный метод требует дискретного распределения ущерба, в то время как обычно используют непрерывное распределение. В этом случае аналогом рекурсии будет интегральное уравнение, решением которого является распределение совокупных убытков.

Теорема 5. Для распределения частоты из класса $(a, b, 1)$ и любого непрерывного распределения ущерба с положительными действительными значениями имеет место следующее интегральное уравнение:

$$f_S(x) = p_1 f_X(x) + \int_0^x \left(a + \frac{by}{x}\right) f_X(y) f_S(x - y)\, dy.$$

Заметим, что начальное значение равно $p_1 f_X(x)$, а не $(p_1 - (a + b)p_0) f_X(x)$ как в дискретной рекурсивной формуле. Выражение в теореме 5 выполнено и для класса $(a, b, 0)$.

Интегральное уравнение – это интеграл Вольтерра второго типа. Численное решение можно найти в литературе.

Рассмотрим альтернативный подход к непрерывному распределению ущерба, основанный на дискретизации.

Построение арифметического распределения

Простейший способ получить дискретное распределение из непрерывного – заменить дискретными значениями вероятностей в точках, кратных подходящей единице измерения h (шаг, размах, интервал).

Такое распределение называется *арифметическим*, так как оно определяется в неотрицательных целых точках. Для того, чтобы «арифметизировать» распределение, важно сохранить свойства исходного распределения – как локальные, в рамках диапазона, так и глобально – для всего распределения.

Должна быть сохранена общая форма распределения и количественные характеристики (моменты).

Метод округления

Пусть f_j - вероятность значения jh, j = 0, 1, 2,… . Тогда положим

$$f_0 = P\left\{X < \frac{h}{2}\right\} = F_X\left(\frac{h}{2} - 0\right), f_j = P\left\{jh - \frac{h}{2} \leq X < jh + \frac{h}{2}\right\}$$
$$= F_X\left(jh + \frac{h}{2} - 0\right) - F_X\left(jh - \frac{h}{2} - 0\right), j = 1, 2, \dots$$

Этот метод собирает всю вероятность половины промежутка с каждой стороны jh и помещает ее в точку jh. Есть исключение для вероятности нулевого значения.

По сути, все количество округляется до ближайшей подходящей денежной единицы, h, шага дискретизации.

Когда непрерывное распределение ущерба неограниченно, имеет смысл остановить процесс дискретизации в некоторой точке как только большая часть вероятности будет учтена. Если m – индекс последней точки, то $f_m = 1 - F_X((m - 0.5)h - 0)$. При таком методе вероятности всегда будут неотрицательными и в сумме будут равны 1.

Метод локального согласования моментов

В этом методе строится арифметическое распределение, у которого p моментов соответствуют истинному распределению ущерба. Рассмотрим произвольный интервал $[x_k, x_k + ph)$ длины ph. Поместим массы вероятностей $m_k^0, m_k^1, \ldots, m_k^p$ в точках $x_k, x_k + h, \ldots, x_k + ph$ так, что первые p моментов сохраняют свои значения. Получим систему из p + 1 уравнений:

$$\sum_{j=0}^{p} (x_k + jh)^r m_j^k = \int_{x_k-0}^{x_k+ph-0} x^r dF_X(x), \qquad r = 0, 1, 2, \ldots p.$$

(дискретная вероятность в x_k включается, а дискретная вероятность в $x_k + ph$ исключается). Упорядочим интервалы так, чтобы $x_{k+1} = x_k + ph$ и чтобы концевые точки совпадали. Тогда значения в концевых точках складываются. При $x_0 = 0$ результирующее дискретное распределение имеет последовательные вероятности:

$$f_0 = m_0^0, \qquad f_1 = m_1^0, \; f_2 = m_2^0, \ldots, f_p = m_p^0 + m_0^1, f_{p+1} = m_1^1, \qquad f_{p+2} = m_2^1, \ldots$$

Суммируя равенства по всем положительным k с учетом $x_0 = 0$, видим, что первые p моментов сохраняются для всего распределения, и сумма вероятностей в точности равна 1. Остается решить систему уравнений.

Метод локального согласования моментов

Теорема 6. Решение системы имеет вид

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

Доказательство. Интерполяционный многочлен Лагранжа для полинома $f(y)$ в точках $y_0, y_1, \dots, y_n$ равен

$$f(y) = \sum_{j=0}^{n} f(y_j) \prod_{i\neq j} \frac{y - y_i}{y_j - y_i}.$$

Применяя эту формулу к полиному $f(y) = y^r$ в точках $x_k, x_k + h, \dots, x_k + ph$, получаем

$$x^r = \sum_{j=0}^{p} (x_k + jh)^r \prod_{i\neq j} \frac{x - x_k - ih}{(j - i)h}, \qquad r = 0, 1, \dots, p.$$

Интегрируя по интервалу $[x_k;\ x_k + ph)$ по функции распределения ущерба, получаем

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

где m_j^k определены выше. Таким образом, решение системы сохраняет первые p моментов.

Пример 11

Предположим, что X имеет экспоненциальное распределение с плотностью распределения $f(x) = 0.1e^{-0.1x}$. Используя шаг сетки h = 2, дискретизируем это распределения с помощью метода округления и с помощью согласования первого момента.

Для метода округления

$$f_0 = F(1) = 1 - e^{-0.1\cdot 1} = 0.09516;\ f_j = F(2j+1) - F(2j-1) = e^{-0.1(2j-1)} - e^{-0.1(2j+1)}.$$

Для согласования первого момента $p = 1$ и $x_k = 2k$. Основные уравнения принимают вид

$$m_0^k = \int_{2k}^{2k+2} \frac{x - 2k - 2}{-2}\, 0.1e^{-0.1x}\, dx = 5e^{-0.1(2k+2)} - 4e^{-0.1\cdot 2k};$$

$$m_1^k = \int_{2k}^{2k+2} \frac{x - 2k}{2}\, 0.1e^{-0.1x}\, dx = -6e^{-0.1(2k+2)} + 5e^{-0.1\cdot 2k};$$

тогда

$$f_0 = m_0^0 = 5e^{-0.2} - 4 = 0.09365,$$

$$f_j = m_1^{j-1} + m_0^j = 5e^{-0.1\,(2j-2)} - 10e^{-0.1\cdot 2j} + 5e^{-0.1(2j+2)}$$

Сравнение дискретизации двумя методами

j	f_j округление	f_j моменты
0	0.09516	0.09365
1	0.16402	0.16429
2	0.13429	0.13451
3	0.10995	0.11013
4	0.09002	0.09017
5	0.07370	0.07382
6	0.06034	0.06044
7	0.04940	0.04948
8	0.04045	0.04051
9	0.03311	0.03317
10	0.02711	0.02716

Свойство метода моментов

Метод локального согласования моментов предложен еще в 1976 году (Gerber, Jones) и изучался для различных эмпирических и аналитических распределений ущерба.

При оценке влияния ошибки на совокупную нетто-премию стоп-лосс оказалось, что двух моментов обычно достаточно, и добавление третьего момента лишь незначительно влияет на точность. Более того, метод округления и согласования первого момент дают одинаковые одинаковые ошибки, тогда как второй момент позволяет получить существенное увеличение точности.

Причина выбора согласования нулевого и первого момента заключается в том, что результирующие вероятности всегда будут неотрицательными. При совпадении двух или более моментов это нельзя гарантировать.

Описанные методы качественно аналогичны численным методам, используемым для решения интегральных уравнений Вольтерры, разработанным в рамках численного анализа.

Влияние модификаций индивидуальных условий страхования на совокупные платежи

Рассмотрим влияние ограничений по полису на совокупные убытки. Индивидуальное сострахование и индивидуальные лимиты влияют на индивидуальные убытки, но не на их частоту, поэтому основное внимание уделим франшизам. По-прежнему считаем, что наличие модификаций в полисах не влияет на оценку риска и распределение индивидуальных потерь. Исходное распределение индивидуальных потерь X предполагается неизменным, меняются только сами платежи. Независимо от того, является ли франшиза условной или безусловной, будем считать, что индивидуальный убыток приводит к выплате с вероятностью v.

К индивидуальной случайной величине потерь X сначала применяются модификации полиса (включая франшизы), после чего производится выплата. Затем отдельные платежи рассматриваются в расчете на каждый убыток, где сумма такого платежа, обозначенная знаком Y^L, будет равна нулю, если убыток не приводит к платежу.

Таким образом, сумма выплаты определяется в расчете на каждый убыток. В качестве альтернативы можно рассматривать отдельные платежи. В этом случае сумма платежа обозначается знаком Y^P, и на этой основе суммы платежей определяются только в отношении убытков, которые фактически приводят к ненулевому платежу.

Влияние модификаций индивидуальных условий страхования на совокупные платежи

По определению $P\{Y^P = 0\} = 0$, а распределение Y^P является условным: $Y^P = Y^L | Y^L > 0$. Следовательно, функции распределения связаны между собой

$$F_{Y^L}(y) = (1 - v) + vF_{Y^P}(y), y \geq 0;$$

так как $1 - v = P(Y^L = 0) = F_{Y^L}(0)$

(вспомним, что Y^L имеет точку дискретной вероятности 1 -v в нуле, даже если X, а значит Y^P и Y^L, имеют непрерывную функцию плотности при $y > 0$). Производящие функции моментов Y^L и Y^P связаны соотношением

$$M_{Y^L}(z) = (1 - v) + vM_{Y^P}(z)$$

или через мат. ожидание

$$E\left[e^{zY^L}\right] = E\left[e^{zY^L} \middle| Y^L = 0\right] P(Y^L = 0) + E\left[e^{zY^L} \middle| Y^L > 0\right] P(Y^L > 0).$$

Влияние модификаций индивидуальных условий страхования на совокупные платежи

Число убытков N^L и число платежей N^P связаны через свои производящие функции вероятности

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

где $P_{N^P}(z) = E[z^{N^P}]$ и $P_{N^L}(z) = E[z^{N^L}]$.

Теперь обратимся к совокупным платежам. При расчете на убыток

$$S = Y_1^L + Y_2^L + \cdots + Y_{N^L}^L,$$

причем $S = 0$, если $N^L = 0$. Здесь Y_j^L – платеж по j-му убытку.

Игнорируя убытки, по которым не делают платежей, совокупный платеж при расчете на платеж равен

$$S = Y_1^P + Y_2^P + \cdots + Y_{N^P}^P,$$

S = 0, если $N^P = 0$, Y_j^P – платеж по j-му убытку, приводящему к ненулевому платежу.

S можно представить двумя способами.

Влияние модификаций индивидуальных условий страхования на совокупные платежи

Производящая функция моментов S при расчете на убыток равна

$$M_S(z) = E[e^{zS}] = P_{N^L}(M_{Y^L}(z))$$

при расчете на платеж

$$M_S(z) = E[e^{zS}] = P_{N^P}(M_{Y^P}(z))$$

Эти представления эквиваленты

$$P_{N^L}(M_{Y^L}(z)) = P_{N^L}(1 - v + vM_{Y^P}(z)) = P_{N^P}(M_{Y^P}(z)).$$

Следовательно, любой анализ совокупных платежей S может быть выполнен либо на основе каждого убытка или на основе каждого платежа. Выбранная основа должна определяться тем, что более подходит для конкретной ситуации.

Если (приблизительное) распределение S представляет больший интерес, чем моменты, то обычно предпочтение отдается подсчету платежей. Причина такого выбора заключается в том, что в расчете на потери могут возникнуть проблемы с недостаточным расходом, если $E[N^L]$ велико, и могут возникнуть проблемы с хранением данных из-за наличия большого числа нулевых вероятностей в распределении Y^L, особенно если используется условная франшиза. Кроме того, для удобства обычно сначала применяются модификации полиса к индивидуальному распределению убытков, а затем производится дискретизация (при необходимости), вместо дискретизации и последующего применения модификаций полиса. Однако эта проблема актуальна только в том случае, если размер франшизы и лимит не являются целыми числами, кратными интервалу дискретизации.

Пример 12

Исходное распределение числа убытков распределено по закону Пуассона со средним значением $\lambda = 3$. Индивидуальные убытки имеют распределение Парето с параметрами $\alpha = 4$ и $\theta = 10$. Индивидуальная безусловная франшиза составляет 6, применяется сострахование на 75% и индивидуальные убытки имеют лимит 24 (до применения франшизы и сострахования). Определим среднее, дисперсию и распределение совокупных платежей

Сначала найдем среднее и дисперсию на основе расчета на убыток. Среднее число убытков равно $E[N^L] = 3$, средний индивидуальный платеж при расчете на убыток (теорема 7, Лекция 4 при $r = 0$ и распределении Парето)

$$E\left[Y^L\right] = 0.75\,(E[X \wedge 24] - E[X \wedge 6]) = 0.75 \cdot 3.2485 - 2.5195 = 0.54675.$$

Средний совокупный платеж равен

$$E[S] = E\left[N^L\right]E\left[Y^L\right] = 3 \cdot 0.54675 = 1.64.$$

Второй момент (теорема 8, Лекция 4, $r = 0$, распределение Парето)

$$E\left[\left(Y^L\right)^2\right] = 0.75^2\left(E\left[(X \wedge 24)^2\right] - E\left[(X \wedge 6)^2\right] - 2 \cdot 6 \cdot E[X \wedge 24] + 2 \cdot 6 \cdot E[X \wedge 6]\right)$$
$$= 0.75^2\,(26.3790 - 10.5469 - 12 \cdot 3.2485 + 12 \cdot 2.5195) = 3.98481.$$

Пример 12

Чтобы вычислить дисперсию совокупных платежей, нет необходимости явно находить $\mathrm{Var}[\mathrm{Y}^{\mathrm{L}}]$, так как S имеет составное распределение Пуассона:

$$Var[S] = \lambda E\left[\left(Y^L\right)^2\right] = 3 \cdot 3.98481 = 11.9544 = (3.46)^2.$$

Для получения (приближенного) распределения S, используем расчет на платеж. Заметим, что

$v = P\{X > 6\} = \left(\frac{10}{10+6}\right)^4 = 0.15259$, и число платежей N^P имеет распределение Пуассона с $E\left[N^P\right] = \lambda v = 3 \cdot 0.15259 = 0.45776.$

Пусть $Z = X - 6|X > 6$, т.е. Z - индивидуальный платеж с франшизой 6. Тогда $P\{Z > z\} = \frac{P\{X > z+6\}}{P\{X > 6\}}$.

При сострахoвании 75%, $Y^P = 0.75Z$ имеет функцию распределения

$$F_{Y^P}(y) = 1 - P\{0.75Z > y\} = 1 - \frac{P\left\{X > 6 + \frac{y}{0.75}\right\}}{P\{X > 6\}}.$$

Для y, меньших чем максимальный платеж, равный $0.75(24 - 6) = 13.5$,

$$F_{Y^P}(y) = \frac{P\{X > 6\} - P\left\{X > 6 + \frac{y}{0.75}\right\}}{P\{X > 6\}}, y < 13.5; \qquad F_{Y^P}(y) = 1, \qquad y \geq 13.5$$

Пример 12

Теперь дискретизируем распределение Y^P (т.е. сначала учитываем модификацию полиса, а потом дискретизируем) с шагом 2.25 и используя метод округления. Этот подход приводит $f_0 = F_{Y^P} \cdot 1.125 = 0.30124$; $f_1 = F_{Y^P} \cdot 3.375 - F_{Y^P} \cdot 1.125 = 0.32768$ и т.д.

Здесь необходимо внимательно выполнять оценку f_6:

$f_6 = F_{Y^P}(14.625) - F_{Y^P}(12.375) = 1 – 0.94126 = 0.05874$. Тогда $f_n = 1 - 1 = 0$ при $n = 7, 8, \ldots$. Приближенное распределение S можно получить с помощью рекурсивной формулы для сложного распределения Пуассона,

$f_S(0) = e^{-0.45776 \cdot (1 - 0.30124)} = 0.7262$ и

$$f_S(x) = \frac{0.45776}{x} \sum_{y=1}^{x \wedge 6} y f_y f_S(x - y), x = 1, 2, 3, ..$$

Например, $f_S(1) = 0.45776 \cdot 1 \cdot 0.32768 \cdot 0.72625 = 0.10894$.

Модель индивидуального риска

Модель.

Модель индивидуального риска представляет совокупный убыток как сумму фиксированного числа независимых (но не обязательно одинаково распределенных) случайных величин:

$$S = X_1 + X_2 + \cdots + X_n.$$

Формула часто рассматривается как сумма убытков от n страховых контрактов, например, n человек, участвующих в полисе группового страхования.

Модель индивидуального риска изначально была разработана для страхования жизни, в условиях которого есть вероятность смерти в течение года q_j, и в случае смерти j-го участника платится фиксированная сумма b_j . В таком случае распределение убытков по j-му полису есть

$$f_{X_j}(x) = \begin{cases} 1 - q_j \ , x = 0; \\ q_j \ , x = b_j \end{cases}$$

Среднее и дисперсия совокупных убытков равны

$$E[S] = \sum_{j=1}^{n} b_j q_j \ , Var[S] = \sum_{j=1}^{n} b_j^2 q_j (1 - q_j)$$

так как X_j предполагаются независимыми.

Модель индивидуального риска

Тогда производящая функция вероятности совокупных убытков равна

$$P_S(z) = \prod_{j=1}^{n}\left(1 - q_j + q_j z^{b_j}\right).$$

Когда все риски одинаковы, $q_j = q$ и $b_j = 1$, получаем $P_S(z) = \left(1 + q(z - 1)\right)^n$, т.е. S имеет биномиальное распределение.

Модель индивидуального риска можно обобщить следующим образом. Пусть $X_j = I_j B_j$, где $I_1, \dots, I_n, B_1, \dots, B_n$ независимы. Случайные величины I_j являются индикаторными, они принимают значение 1 с вероятностью q_j и 0 с вероятностью $1 - q_j$.

Эти переменные показывают, производится ли платеж по j-му полису. Случайная величина B_j может иметь любое распределение и представляет величину платежа по j-му полису. В страховании жизни B_j вырождена, со всей вероятностью в значении b_j .

Производящая функция моментов

$$M_S(z) = \prod_{j=1}^{n}\left(1 - q_j + q_j M_{B_j}(z)\right).$$

Если положить $\mu_j = E[B_j], \sigma_j^2 = Var[B_j]$, тогда $E[S] = \sum_{j=1}^{n} q_j \mu_j$, $Var[S] = \sum_{j=1}^{n}(q_j \sigma_j^2 + q_j(1 - q_j)\mu_j^2)$

Пример

Рассмотрим групповое страхование жизни с выплатой в случае смерти от несчастного случая. Пусть для всех членов вероятность смерти в течение следующего года 0.01 и 30% смертей являются следствие несчастного случая. Для 50 сотрудников выплата в случае обычной смерти составляет 50,000 и при несчастном случае 100,000. Для оставшихся 25 сотрудников выплаты равны 75,000 и 150,000, соответственно. Предложим модель индивидуального риска и определим среднее и дисперсию.

Для всех 75 сотрудников $q_j = 0.01$. Для 50 человек B_j принимает значение 50,000 с вероятностью 0.7 и 100,000 с вероятностью 0.3. Для них $\mu_j = 65\,000$ и $\sigma_j^2 = 525\,000\,000$. Для оставшихся 25 сотрудников, B_j принимает значение 75 000 с вероятностью 0.7 и 150 000 вероятностью 0.3. Для них $\mu_j = 97\,500$ и $\sigma_j^2 = 1\,181\,250\,000$. Тогда

$$E[S] = 50 \cdot 0.01 \cdot 65\,000 + 25 \cdot 0.01 \cdot 97500 = 56875$$

$$Var[S] = 50 \cdot 0.01 \cdot 525\,000\,000 + 50 \cdot 0.01 \cdot 0.99 \cdot 65\,000^2 + 25 \cdot 0.01 \cdot 1\,181\,250\,000 + 25 \cdot 0.01 \cdot 0.99 \cdot 97\,500^2$$
$$= 5\,001\,984\,375.$$

При расчете вероятностей существует по крайней мере три варианта. Один из них заключается в выполнении точного вычисления, которое включает в себя многочисленные свертки и почти всегда требует более длительного вычислительного времени. Были разработаны рекурсивные формулы, но они громоздки. Одной из альтернатив является параметрическая аппроксимация. Другой альтернативой является замена модели индивидуального риска аналогичной моделью коллективного риска.

Параметрическая аппроксимация

Для аппроксимации распределения можно использовать нормальное, гамма-, логнормальное или любое другое распределение, обычно это делается путем сопоставления первых нескольких моментов. Поскольку нормальное, гамма-и логнормальное распределения имеют по два параметра, среднего значения и дисперсии достаточно.

ПРИМЕР 13. (групповое страхование жизни) Небольшое производственное предприятие заключило договор группового страхования жизни своих 14 постоянных сотрудников. Актуарий страховщика выбрал таблицу смертности, отражающую смертность в группе. Каждый сотрудник застрахован на сумму своей заработной платы, округленной до ближайшей 1000. Данные группы приведены в таблице. Каковы шансы, что с заданном году страховщик потерпит убытки, если он добавит 45%-ю относительную нагрузку к чистой (нетто) премии? Рассмотрим нормальную и логнормальную аппроксимации.

Сотрудник, j	возраст	пол	выплата	уровень смертности
1	20	M	15,000	0.00149
2	23	M	16,000	0.00142
3	27	M	20,000	0.00128
4	30	M	28,000	0.00122
5	31	M	31,000	0.00123
6	46	M	18,000	0.00353
7	47	M	26,000	0.00394
8	49	M	24,000	0.00484
9	64	M	60,000	0.02182
10	17	F	14,000	0.00050
11	22	F	17,000	0.00050
12	26	F	19,000	0.00054
13	37	F	30,000	0.00103
14	55	F	55,000	0.00479
всего			373,000	

Пример 13

Среднее и дисперсия совокупных убытков для группы

$$E[S] = \sum_{j=1}^{14} b_j q_j = 2054.41, \qquad Var[S] = \sum_{j=1}^{14} b_j^2 q_j \left(1 - q_j\right) = 1.02534 \cdot 10^8.$$

Премия составляет $1.45 \cdot 2\,054.41 = 2978.89$. Для аппроксимации нормальным распределением (в единицах 1000), среднее равно 2.05441 и дисперсия 102.534. Тогда вероятность убытка

$P\{S > 2.97889\} = P\left\{Z > \frac{2.97889 - 2.05441}{102.534^{1/2}}\right\} = P\{Z > 0.0913\} = 0.46$, т.е. 46%.

Для аппроксимации логнормальным распределением

$$\mu + \frac{1}{2}\sigma^2 = \ln 2.05441 = 0.719989$$

и

$$2\mu + 2\sigma^2 = \ln\left(102.534 + 2.05441^2\right) = 4.670533.$$

Отсюда $\mu = -0.895289$ и $\sigma^2 = 3.230555$. Тогда

$P\{S > 2.97889\} = 1 - \Phi\left(\frac{\ln 2.97889 + 0.895289}{3.230555^{1/2}}\right) = 1 - \Phi(1.105) = 0.13$, т.е. 13%.

Аппроксимация сложным распределением Пуассона

В силу большой вычислительной сложности расчета распределения совокупных убытков для портфеля из n рисков при использовании индивидуальной модели, популярной является аппроксимация распределения сложным распределением Пуассона. Использование обобщенного пуассоновского распределения позволяет вычислять распределением совокупных убытков с помощью простой рекурсивной процедуры.

Заметим, что индикаторная случайная величина I_j имеет производящую функцию вероятности $P_{I_j}(z) = 1 - q_j + q_j z$. Поэтому

$$M_S(z) = \prod_{j=1}^{n} P_{I_j}(M_{B_j}(z)).$$

I_j имеет биномиальное распределение с параметрами m = 1 и $q = q_j$. Чтобы получить аппроксимацию сложным распределением Пуассона, предположим, что I_j имеют распределением Пуассона со средним λ_j. Если $\lambda_j = q_j$, то среднее распределения Пуассона совпадает со средним биномиального распределения, что должно обеспечить хорошую аппроксимацию при q_j в окрестности нуля. Альтернативой приравниванию среднего значения является приравнивание вероятности отсутствия потерь. Для биномиального распределения эта вероятность равна $1 - q_j$, а для распределения Пуассона это $\exp(-\lambda_j)$. Приравнивание этих двух вероятностей дает альтернативное приближение $\lambda_j = -\ln(1 - q_j) > q_j$. Это второе приближение уместно в контексте договора группового страхования жизни, когда жизнь «заменяется» после смерти, оставляя интенсивность Пуассона неизменной. Естественно, ожидаемое количество потерь в этом случае окажется больше, чем $\sum_{j=1}^{n} q_j$.

Аппроксимация сложным распределением Пуассона

Еще один вариант предложен Kornya. Он использует $\lambda_j = \frac{q_j}{1 - q_j}$ и приводит к ожидаемому числу убытков, превосходящему то, которое получатся при приравнивании вероятности отсутствия убытков.
Независимо от того, какая используется аппроксимация, из Теоремы 3 следует, что

$$P_{I_j}(x) = \exp\left(\lambda_j (z - 1)\right), \qquad M_S(z) = \prod_{j=1}^{n} \exp(\lambda_j (M_{B_j}(z) - 1)) = \exp(\lambda(M_X(z) - 1)),$$

где

$$\lambda = \sum_{j=1}^{n} \lambda_j, M_X(z) = \lambda^{-1} \sum_{j=1}^{n} \lambda_j M_{B_j}(z);$$

Аппроксимация сложным распределением Пуассона

так что X имеет функцию плотности вероятности

$$f_X(x) = \lambda^{-1} \sum_{j=1}^{n} \lambda_j f_{B_j}(x)$$

которая является взвешенным средним n индивидуальных плотностей убытков.
Если $P\{B_j = b_j\} = 1$, как в страховании жизни, получаем

$$f_X(x) = P\{X = x\} = \lambda^{-1} \sum_{\{j:b_j=x\}} \lambda_j$$

Числитель суммирует все вероятности, связанные с величиной b_j .

Пример 13 (продолжение)

Рассмотрим случай группового страхования жизни. Выведем аппроксимацию на основе сложного распределения Пуассона с согласованием мат. ожиданий.

Используя аппроксимацию на основе сложного распределения Пуассона с параметром

$\lambda = \sum q_j = 0.04813$, получаем функцию распределения в таблице. Когда эти значения сравниваются с точным решением, оказывается, что максимальная ошибка равна 0.0002708 в точке $x = 0$.

X	F_s M	X	F_s M	X	F_s M	X	F_s M
0	0.9530099	20	0.9618348	40	0.9735771	60	0.9990974
1	0.9530099	21	0.9618348	41	0.9735850	61	0.9990986
2	0.9530099	22	0.9618348	42	0.9736072	62	0.9990994
3	0.9530099	23	0.9618348	43	0.9736133	63	0.9990995
4	0.9530099	24	0.9664473	44	0.9736346	64	0.9990995
5	0.9530099	25	0.9664473	45	0.9736393	65	0.9990996
6	0.9530099	26	0.9702022	46	0.9736513	66	0.9990997
7	0.9530099	27	0.9702022	47	0.9736541	67	0.9990997
8	0.9530099	28	0.9713650	48	0.9736708	68	0.9990998
9	0.9530099	29	0.9713657	49	0.9736755	69	0.9991022
10	0.9530099	30	0.9723490	50	0.9736956	70	0.9991091
11	0.9530099	31	0.9735235	51	0.9736971	71	0.9991156
12	0.9530099	32	0.9735268	52	0.9737101	72	0.9991179
13	0.9530099	33	0.9735328	53	0.9737102	73	0.9991341
14	0.9534864	34	0.9735391	54	0.9737195	74	0.9991470
15	0.9549064	35	0.9735433	55	0.9782901	75	0.9991839
16	0.9562597	36	0.9735512	56	0.9782947	76	0.9992135
17	0.9567362	37	0.9735536	57	0.9782994	77	0.9992239
18	0.9601003	38	0.9735604	58	0.9783006	78	0.9992973
19	0.9606149	39	0.9735679	59	0.9783021	79	0.9993307

Пример 13 (продолжение)

Предложить аппроксимации на основе сложного распределения Пуассона, используя все три метода. Вычислить среднее и дисперсию для всех трех и сравнить.

Используя метод соответствия среднего, находим $\lambda = 50 \cdot 0.01 + 25 \cdot 0.01 = 0.75$.

Распределение ущерба

$$f_X(50\,000) = \frac{50 \cdot 0.01 \cdot 0.7}{0.75} = 0.4667; \quad f_X(75\,000) = \frac{25 \cdot 0.01 \cdot 0.7}{0.75} = 0.2333;$$

$$f_X(100\,000) = \frac{50 \cdot 0.01 \cdot 0.3}{0.75} = 0.2000; \quad f_X(150\,000) = \frac{25 \cdot 0.01 \cdot 0.3}{0.75} = 0.1000.$$

Среднее равно $\lambda E[X] = 0.75 \cdot 75\,833.33 = 56\,875$, что соответствует точному значению, дисперсия равна $\lambda E\left[X^2\right] = 0.75 \cdot 6\,729\,166\,667 = 5\,046\,875\,000$, что превосходит точное значение.

Для метода, сохраняющего вероятность отсутствия убытков, $\lambda = -75 \ln 0.99 = 0.753775$. Для этого метода распределение ущерба оказывается в точности таким же, как и раньше (так как все индивидуумы имеют одинаковое значение q_j). Таким образом, среднее равно 57 161 и дисперсия равна 5 072 278 876, оба значения превосходят предыдущую аппроксимацию.

Используя метод Корня, $\lambda = \frac{75 \cdot 0.01}{0.99} = 0.757576$ и снова распределение убытков не меняется. Среднее равно 57 449 и дисперсия 5 097 853 535, они наибольшие из всех.