

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М. В. ЛОМОНОСОВА»

МЕХАНИКО-МАТЕМАТИЧЕСКИЙ ФАКУЛЬТЕТ

КАФЕДРА теории вероятностей

КУРСОВАЯ РАБОТА

на тему

«МОДЕЛЬ ИНДИВИДУАЛЬНОГО РИСКА»

Выполнил студент 309 группы:

Искорнев Матвей Евгеньевич

подпись студента

Научный руководитель:

профессор, д.ф.-м.н.

Фалин Геннадий Иванович

подпись научного руководителя

Москва

2020

Оглавление

1. Введение	3
2. Резюме.....	4
3. Свёртка	4
4. Аппроксимация	5
4.1. Гамма-приближение.....	6
4.2. Нормальное степенное приближение.....	10
4.3. Приближение ортогональными многочленами	12
4.4. Приближение Гамма-функцией Бауэrsa	14
4.5. Аппроксимация Грама – Шарлье	17
4.6. Разложение Эджуорта	21
4.7. Аппроксимация Эшера	23
5. Связь между моделями индивидуального и коллективного риска	27
5.1. Описание модели индивидуального иска	27
5.2. Описание модели коллективного риска	30
5.3. Связь между моделями	31
5.4. Примеры	33
6. Ошибки и недостатки	39
7. Заключение	40
8. Список литературы	41

1. Введение

Модель индивидуального риска имеет большое значение в страховой математике. Несмотря на относительную простоту своего устройства она находит практическое применение во многих областях страхования. Данная модель позволяет получить определенную информацию о важнейшей составляющей теории риска – функции распределения величины суммарного иска. Исходя из полученной информации представляется возможным вычислять вероятность разорения страховой компании или прогнозировать вероятность её успешного и безоблачного будущего, основываясь на начальном капитале. Цель данной работы – подробно и детально изучить этот современный раздел теории риска, описать основные методы и инструменты для его анализа и сделать определенные выводы. Наше изложение главным образом будет базироваться на следующих статьях: Dhaene, J., Vyncke, D., The individual risk model; Gerber, H. U., An Introduction to Mathematical Risk Theory, а также на разделах 5.2 и 5.5 главы 5 книги Panjer, H. H., Willmot, G. E., Insurance Risk Models*. Хотелось бы заметить, что многие доказательства и рассуждения в вышеупомянутых монографиях содержат пробелы или вовсе приводятся без каких-либо обоснований. Кроме того, большое число важных деталей и логических переходов попросту опущено. Однако в настоящей работе мы восполним все эти пробелы. Утверждения, требующие обоснований, будут сопровождены полными и исчерпывающими доказательствами, содержащими при необходимости точные ссылки на различные результаты из области теории вероятностей или других математических разделов. К тем рассуждениям, которые, по нашему мнению, нуждаются в более детальном и подробном рассмотрении, будут добавлены различные графики, рисунки и таблицы (см. рис. **4.1.1, 4.1.3, 4.1.5, 5.1.1, 5.4.1, 5.4.5**, таблицы **5.1.1, 5.4.1 – 5.4.4**). Отметим, что все графики, иллюстрирующие теорию, реализованы посредством языка программирования Python, все алгоритмы для их построения были написаны и составлены лично нами (их можно найти на рис. **4.1.2, 4.1.4, 4.1.6, 5.4.2 – 5.4.4, 5.4.6**). Естественно, одной лишь теории в столь важной и практически значимой теме, конечно, недостаточно. Именно поэтому мы приведем большое количество примеров, некоторые из которых используют реальные данные. Благодаря им сравнение изучаемых методов станет более объективным, полным и позволит сделать правильные и разумные выводы. Вся работа, которая была выполнена нами самостоятельно будет точно указываться. Тем не менее мы не претендуем на авторство конкретных утверждений, примеров и обоснований, а также многочисленных понятий из теории риска, используемых в данной работе.

* Номера [1], [2] и [3] в списке литературы соответственно.

2. Резюме

Модель индивидуального иска обычно базируется на следующих упрощающих принципах:

1. Анализируется фиксированный относительно короткий промежуток времени (так чтобы не учитывать инфляцию и возможные доходы от различных инвестиций).
2. Число договоров страхования фиксировано и неслучайно.
3. Плата за страховку целиком вносится в начале анализируемого промежутка времени.
4. Наблюдается каждый отдельный договор страхования вместе с известными статистическими свойствами связанной с ним величины индивидуального иска.
5. Исключаются катастрофические несчастные случаи, которые могут повлечь иски сразу по нескольким договорам (независимость индивидуальных исков).

В рамках данной модели общее число исков в портфеле договоров страхования является случайной величиной. Эта величина моделируется как сумма всех исков по отдельным договорам, которые предполагаются независимыми. Естественным образом возникает много вопросов, связанных с этой случайной величиной: как она распределена? К какому классу случайных величин относится? Каким образом можно посчитать её параметры? Как это сделать наиболее удобно и просто? Постараемся ответить на эти вопросы, описав несколько методов, позволяющих достигнуть определенных результатов в исследуемой области.

3. Свёртка

В модели индивидуального риска особый интерес представляет распределение суммарного иска S с фиксированным числом n страховых договоров:

$$S = X_1 + X_2 + \dots + X_n, \quad (1)$$

где X_i , $i = 1, 2, \dots, n$, иск от i -го договора. Предполагая, что все иски X_i — это независимые случайные величины, распределение их суммы может быть посчитано с помощью *свёртки*. Традиционно в теории вероятностей свёртка — это математическая операция, позволяющая вычислять функцию распределения суммы двух независимых случайных величин X и Y по следующей формуле:

$$F_{X+Y}(s) = P(X + Y \leq s) = \int_{-\infty}^{+\infty} F_Y(s-x) dF_X(x) =: F_X * F_Y(s) \quad (2)$$

Функцию распределения $F_X * F_Y(s)$ обычно называют *свёрткой функций распределения* $F_X(\cdot)$ и $F_Y(\cdot)$. Применяя формулу (2) несколько раз, мы можем посчитать функцию

распределения суммы любого числа слагаемых. Причём совершенно неважно в каком порядке мы производим свёртку, результат от этого не зависит:

$$(F_X * F_Y) * F_Z = F_X * (F_Y * F_Z) = F_X * F_Y * F_Z \quad (3)$$

Таким образом, функция распределения суммы n независимых одинаково распределённых случайных величин с функцией распределения F есть просто n -кратная свертка функции F , которую мы обозначим следующим образом:

$$F * F * \dots * F =: F * n. \quad (4)$$

Хочется заметить, что свертка имеет довольно большое значение в теории риска в целом. В настоящее время благодаря мощным вычислительным машинам представляется возможным находить функцию распределения суммарного числа исков непосредственно, используя именно свёртку. Здесь также стоит добавить, что в этом случае нередко применяется преобразование Фурье характеристической функции. Ввиду вышеизложенных причин многие методы аппроксимации сегодня утратили свое значение. Тем не менее это совсем не означает, что теперь они подлежат забвению. Напротив, для того чтобы получить максимально полное и объективное представление о модели индивидуального риска необходимо тщательно изучить все основные методы.

4. Аппроксимация

Альтернативный подход основан на аппроксимации распределения величины суммарного иска S . Предположим, что S является суммой достаточно большого числа случайных величин. Тогда в соответствии с центральной предельной теоремой, мы можем приблизить искомое распределение гауссовой величиной, имеющей такое же среднее и такую же дисперсию как S . Кажущуюся простота этого метода компенсируется его недостатками. Во-первых, тяжело точно сказать какое именно число исков является достаточно большим, для применения теоремы. Во-вторых, метод дает довольно грубую оценку хвостов распределения. И, наконец, третий центральный момент суммарного числа исков обычно больше 0, в то время как для нормально распределённой величины он равен 0 (простое упражнение по теории вероятностей, для которого достаточно знать производящую функцию моментов для $X \sim N(a, \sigma^2)$, $m_X(t) = \exp\left(at + \frac{\sigma^2 t^2}{2}\right)$ и её связь с

моментами: $E(X^k) = \frac{d^k}{dt^k} m_X(t) \Big|_{t=0}$).

В качестве альтернативы центральной предельной теореме опишем несколько альтернативных методов аппроксимации: *гамма – приближение* (англ. gamma approximation), *нормальное степенное приближение* (англ. normal power (NP) approximation) и многие другие. Данные аппроксимации имеют ряд преимуществ в сравнении с приближением, базирующимся на центральной предельной теореме. Во – первых, они оказываются гораздо более точными. Во – вторых, их недостатки совсем

незначительны в сравнении с ошибками, которые возникают вследствие неточной оценки третьего центрального момента случайной величины.

4.1. Гамма-приближение

Часто распределение величины суммарного иска оказывается очень похожим на гамма-распределение. Как было замечено, его плотности свойственно небольшое смещение вправо, кроме того, величина суммарного иска принимает неотрицательные значения и имеет «вершину» (наиболее часто встречающееся значение). Кроме стандартных параметров α и β ($G(\alpha, \beta)$) мы добавим третью степень свободы, то есть рассмотрим *сдвинутое гамма-распределение* ($G(x - x_0; \alpha, \beta)$). Иными словами, нас интересует распределение случайной величины $\eta = x_0 + \gamma_{\alpha, \beta}$, где $\gamma_{\alpha, \beta} \sim G(\alpha, \beta)$. Ниже для наглядности приведены графики плотности гамма величины и её функции распределения, а также код, посредством которого эти графики реализованы.

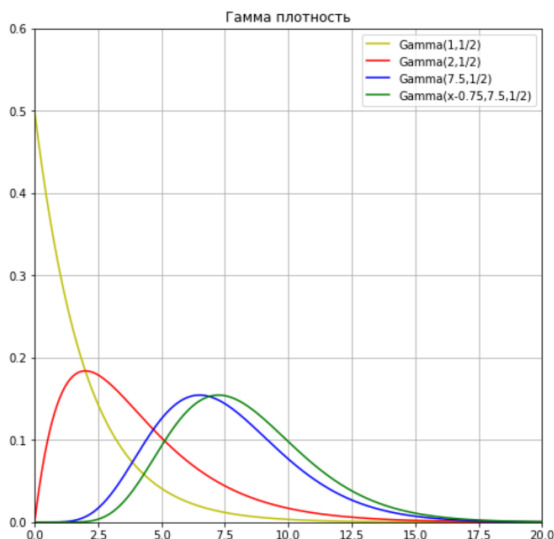


Рис. 4.1.1. Гамма плотность.

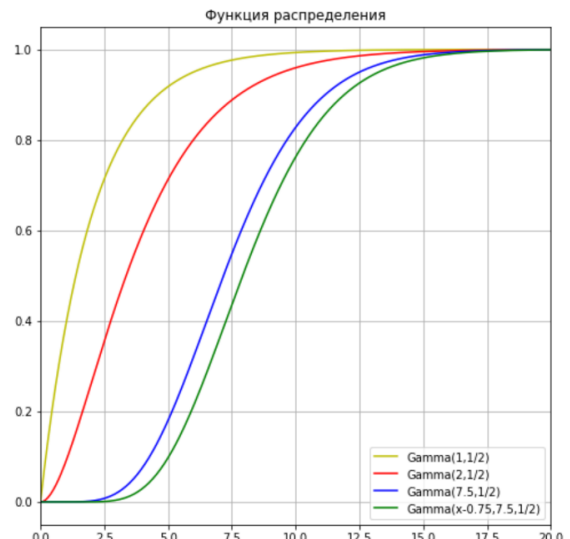


Рис. 4.1.3. Распределение гамма величины.

```
import numpy as np
import scipy.stats as stats
from matplotlib import pyplot as plt
plt.figure(figsize = (8, 8))
x = np.linspace (0, 100, 2000)
y1 = stats.gamma.pdf(x, a=1, scale=2)
y2 = stats.gamma.pdf(x, a=2, scale=2)
y3 = stats.gamma.pdf(x, a=7.5, scale=1)
y4 = stats.gamma.pdf(x-0.75, a=7.5, scale=1)
plt.plot(x, y1, "y-", label='Gamma(1,1/2)')
plt.plot(x, y2, "r-", label='Gamma(2,1/2)')
plt.plot(x, y3, "b-", label='Gamma(7.5,1/2)')
plt.plot(x, y4, "g-", label='Gamma(x-0.75,7.5,1/2)')
plt.legend(loc='upper right')
plt.title('Гамма плотность')
plt.ylim([0,0.55])
plt.xlim([0,20])
plt.show()
```

Рис. 4.1.2. Код для графиков плотностей.

```
import numpy as np
import scipy.stats as stats
from matplotlib import pyplot as plt
plt.figure(figsize=(8,8))
x=np.linspace(0,100,2000)
y1=stats.gamma.cdf(x, a=1, scale=2)
y2=stats.gamma.cdf(x, a=2, scale=2)
y3=stats.gamma.cdf(x, a=7.5, scale=1)
y4=stats.gamma.cdf(x-0.75, a=7.5, scale=1)
plt.plot(x, y1, 'y-', label='Gamma(1,1/2)')
plt.plot(x, y2, 'r-', label='Gamma(2,1/2)')
plt.plot(x, y3, 'b-', label='Gamma(7.5,1/2)')
plt.plot(x, y4, 'g-', label='Gamma(x-0.75,7.5,1/2)')
plt.legend(loc = "lower right")
plt.title('Функция распределения')
plt.ylim([-0.05,1.05])
plt.xlim([0,20])
plt.show()
```

Рис. 4.1.4. Код для графиков функций распределения.

Заметим, что

$$F_{\eta}(s) = P(x_0 + \gamma_{\alpha, \beta} \leq s) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \int_0^{s-x_0} y^{\alpha-1} e^{-\beta y} dy, \quad s - x_0 \geq 0 \quad (5)$$

Найдем моменты сдвинутого гамма-распределения, используя п. ф. м.*
 $\left(1 - \frac{t}{\beta}\right)^{-\alpha}$, $t < \beta$. Опуская вычисления, получаем:

$$E(\gamma_{\alpha, \beta}) = \frac{\alpha}{\beta}, \quad E(\gamma_{\alpha, \beta}^2) = \frac{\alpha(\alpha+1)}{\beta^2}, \quad E(\gamma_{\alpha, \beta}^3) = \frac{\alpha(\alpha+1)(\alpha+2)}{\beta^3}.$$

Приравнявая моменты величины η и величины S , имеем следующую систему:

$$E(\eta) = E(x_0 + \gamma_{\alpha, \beta}) = x_0 + E(\gamma_{\alpha, \beta}) = x_0 + \frac{\alpha}{\beta} = E(S),$$

$$Var(\eta) = Var(x_0 + \gamma_{\alpha, \beta}) = Var(\gamma_{\alpha, \beta}) = E(\gamma_{\alpha, \beta}^2) - (E(\gamma_{\alpha, \beta}))^2 = \frac{\alpha}{\beta^2} = Var(S),$$

$$E(\eta - E(\eta))^3 = E(\gamma_{\alpha, \beta} - E(\gamma_{\alpha, \beta}))^3 = 2\nu_1^3 - 3\nu_1\nu_2 + \nu_3 = 2\left(\frac{\alpha}{\beta}\right)^3 - 3\left(\frac{\alpha}{\beta}\right)\left(\frac{\alpha(\alpha+1)}{\beta^2}\right) + \frac{\alpha(\alpha+1)(\alpha+2)}{\beta^3} = \frac{2\alpha}{\beta^3} = E(S - E(S))^3.$$

Решив её, получим уравнения, связывающие параметры гамма величины с её моментами:

$$i. \quad \alpha = 4 \frac{(Var(S))^3}{(E(S - E(S)))^3} = \frac{4}{\gamma^2},$$

$$ii. \quad \beta = 2 \frac{Var(S)}{E(S - E(S))^3} = \frac{2}{\gamma\sigma},$$

$$iii. \quad x_0 = E(S) - 2 \frac{(Var(S))^2}{E(S - E(S))^3} = \mu - 2 \frac{\sigma}{\gamma},$$

где μ, σ^2, γ - математическое ожидание, дисперсия и коэффициент асимметрии величины S соответственно.

Подчеркнем, что в силу соотношения (i) величина $\gamma = \frac{2}{\sqrt{\alpha}}$, поэтому одно из условий применения данной аппроксимации (положительность коэффициента асимметрии величины суммарного иска S) выполнено. В монографии [2] следующее замечание (см. страницу 4 раздела 4) сформулировано без каких-либо комментариев, однако

* Здесь и далее п. ф. м. – производящая функция моментов.

приведенное нами доказательство восполняет этот пробел. Заметим, что если $\gamma \downarrow 0$, то имеет место нормальное приближение:

▷ Действительно, рассмотрим хар. ф. * гамма-распределения:

$$\psi_{\gamma, \beta}(t) = \left(1 - \frac{it}{\beta}\right)^{-\alpha},$$

для простоты положим $x_0 = 0$. В силу соотношений (i) – (iii) имеем:

$$\left(1 - \frac{it}{\beta}\right)^{-\alpha} = \left(1 - \frac{it\gamma\sigma}{2}\right)^{-\frac{4}{\gamma^2}} = \exp\left(-\frac{4}{\gamma^2} \left(\ln\left(1 - \frac{it\gamma\sigma}{2}\right)\right)\right),$$

Благодаря формуле Тейлора при $\gamma \downarrow 0$ получаем:

$$\begin{aligned} \left(1 - \frac{it}{\beta}\right)^{-\alpha} &= \exp\left(-\frac{4}{\gamma^2} \left(-\frac{it\gamma\sigma}{2} + \frac{t^2\gamma^2\sigma^2}{4 \cdot 2!} + \bar{o}(\gamma^2)\right)\right) = \exp\left(\frac{2\sigma}{\gamma} it - \frac{\sigma^2}{2} t^2 + \bar{o}(\gamma^2)\right) = \\ &= \exp\left(\mu it - \frac{\sigma^2}{2} t^2 + \bar{o}(\gamma^2)\right). \end{aligned}$$

Таким образом,

$$\lim_{\gamma \rightarrow 0} \psi_{\gamma, \beta}(t) = \lim_{\gamma \rightarrow 0} \exp\left(\mu it - \frac{\sigma^2}{2} t^2 + \bar{o}(\gamma^2)\right) = \exp\left(\mu it - \frac{\sigma^2}{2} t^2\right) = \psi_X(t),$$

где $X \sim N(\mu, \sigma^2)$. ◁

В рассматриваемой нами статье [2] нижеследующее утверждение **4.1.2** приводится без доказательства, но в настоящей работе мы приведем полное и исчерпывающее объяснение. Прежде чем перейти к нему, докажем небольшое утверждение:

Утв. 4.1.1. Пусть $\chi^2(k)$ - распределение хи-квадрат с k степенями свободы, $\gamma_{\alpha, \beta} \sim G(\alpha, \beta)$. Тогда $\gamma_{\alpha, \beta} = \frac{1}{2\beta} \chi^2(2\alpha)$.

Док-во: ▷ Вспомним, чему равна хар. ф. распределения хи-квадрат:

$$\psi_{\chi^2(k)}(t) = (1 - 2it)^{-\frac{k}{2}},$$

в таком случае

$$\psi_{\frac{1}{2\beta} \chi^2(2\alpha)}(t) = E\left(e^{it \frac{1}{2\beta} \chi^2(2\alpha)}\right) = E\left(e^{i \left(\frac{t}{2\beta}\right) \chi^2(2\alpha)}\right) = \left(1 - 2it \frac{1}{2\beta}\right)^{-\alpha} = \left(1 - \frac{it}{\beta}\right)^{-\alpha} = \psi_{\gamma_{\alpha, \beta}}(t). \triangleleft$$

Теперь мы готовы доказать следующее приближение:

* Здесь и далее хар. ф. – характеристическая функция.

Утв. 4.1.2. Если $Y \sim G(\alpha, \beta), \alpha \geq \frac{1}{4}$, то приблизительно $\sqrt{4\beta Y} - \sqrt{4\alpha - 1} \sim N(0, 1)$.

Док-во: ▷ Доказательство будет опираться на следующий факт*:

$$\frac{X_k - \mu_k}{\sigma_k} \xrightarrow{d} N(0, 1), k \rightarrow \infty,$$

где X_k имеет хи-распределение с k степенями свободы (то есть $X_k^2 \sim \chi^2(k)$). Для среднего и дисперсии величины X_k справедливы следующие соотношения:

$$\mu_k = \sqrt{2} \Gamma\left(\frac{k+1}{2}\right) / \Gamma\left(\frac{k}{2}\right), \quad \sigma_k^2 = k - \mu_k^2.$$

Положим $\tilde{X}_k = \sqrt{2}X_k$, тогда математическое ожидание этой величины будет равно $\tilde{\mu}_k = \sqrt{2}\mu_k$, а дисперсия $\tilde{\sigma}_k^2 = 2k - \tilde{\mu}_k^2$. Заметим, что с ростом k величины $\tilde{\mu}_k \approx \sqrt{2k-1}$ и $\tilde{\sigma}_k^2 \approx 1$, причем чем больше степеней свободы, тем точнее указанные соотношения. Чтобы убедиться в том, что данное замечание действительно имеет место, посмотрим на графики:

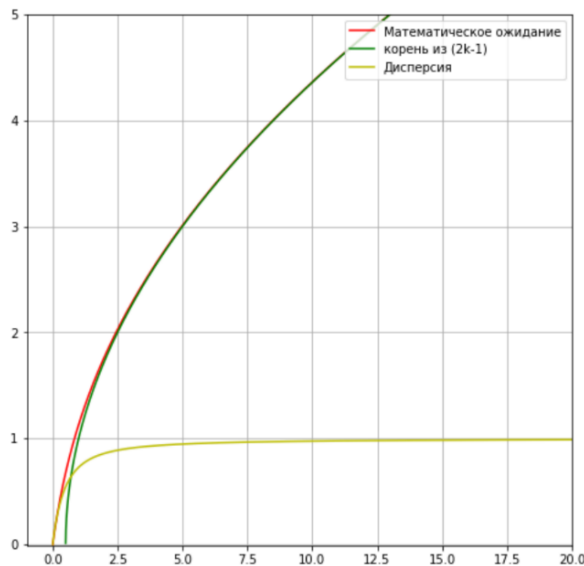


Рис. 4.1.5. Графики зависимостей параметров величины \tilde{X}_k от k .

```
import math
import numpy as np
from matplotlib import pyplot as plt
from scipy.special import gamma, factorial
plt.figure(figsize=(8,8))
x=np.linspace(0,100,10000)
z1=(2*(gamma((x+1)/2)))/gamma(x/2)
z2=(2*x-1)**(1/2)
z3=2*x-((2*(gamma((x+1)/2)))/gamma(x/2))**2
plt.plot(x, z1, 'r-', label='Математическое ожидание')
plt.plot(x, z2, 'g-', label='корень из (2k-1)')
plt.plot(x, z3, 'y-', label='Дисперсия')
plt.legend(loc = "upper right")
plt.ylim([-0.01,5])
plt.xlim([-1,20])
grid1 = plt.grid(True)
plt.show()
```

Рис. 4.1.6. Код для реализации графиков параметров величины \tilde{X}_k .

Таким образом, сделанное замечание справедливо и следовательно, мы можем утверждать, что при достаточно больших k величина $\tilde{X}_k \sim N(\sqrt{2k-1}, 1)$. Далее из предыдущего утверждения получаем, что

$$Y = \frac{1}{2\beta} \chi^2(2\alpha),$$

* Подтверждение можно найти в книге [4] из списка литературы.

следовательно, $\sqrt{4\beta Y} = \sqrt{2\chi^2(2\alpha)} \sim N(\sqrt{4\alpha-1}, 1)$ и поэтому $\sqrt{4\beta Y} - \sqrt{4\alpha-1} \sim N(0, 1)$. \triangleleft

Итак, пусть мы аппроксимируем величину S сдвинутым гамма-распределением. То есть $\eta = x_0 + \gamma_{\alpha, \beta}$, где $\gamma_{\alpha, \beta} \sim G(\alpha, \beta)$, тогда как несложно заметить $\gamma_{\alpha, \beta} = \eta - x_0$ и по утверждению **4.1.2** имеем: $\sqrt{4\beta(\eta - x_0)} - \sqrt{4\alpha-1} \sim N(0, 1)$.

Поэтому $\Phi(y) \approx P(\sqrt{4\beta(S - x_0)} - \sqrt{4\alpha-1} \leq y) = P(\sqrt{4\beta(S - x_0)} \leq y + \sqrt{4\alpha-1})^*$.

Так как мы рассматриваем $y \geq 0$, то можем возвести обе части неравенства в квадрат:

$$\begin{aligned} P(\sqrt{4\beta(S - x_0)} \leq y + \sqrt{4\alpha-1}) &= P(4\beta(S - x_0) \leq y^2 + 2y\sqrt{4\alpha-1} + 4\alpha-1) = \\ &= P\left(S - x_0 \leq \frac{y^2}{4\beta} + \frac{y\sqrt{4\alpha-1}}{2\beta} + \frac{4\alpha-1}{4\beta}\right) = P\left(S \leq x_0 + \frac{1}{4\beta}(y^2-1) + y\sqrt{\frac{\alpha}{\beta^2} - \frac{1}{4\beta^2}} + \frac{\alpha}{\beta}\right) = \\ &= P\left(S \leq \mu - \frac{2\sigma}{\gamma} + \frac{\gamma\sigma}{4 \cdot 2}(y^2-1) + y\sqrt{\sigma^2 - \frac{\gamma^2\sigma^2}{16}} + \frac{2\sigma}{\gamma}\right) = P\left(\frac{S-\mu}{\sigma} \leq \frac{\gamma}{8}(y^2-1) + y\sqrt{1 - \frac{\gamma^2}{16}}\right) = \\ &= P\left(\frac{S-\mu}{\sigma} \leq y + \frac{\gamma}{8}(y^2-1) - y\left(1 - \sqrt{1 - \frac{\gamma^2}{16}}\right)\right) \end{aligned}$$

Таким образом,

$$\Phi(y) \approx P\left(\frac{S-\mu}{\sigma} \leq y + \frac{\gamma}{8}(y^2-1) - y\left(1 - \sqrt{1 - \frac{\gamma^2}{16}}\right)\right), y \geq 0. \quad (6)$$

Видим, что в правой части неравенства появилась некая корректирующая добавка к y , а именно $\frac{\gamma}{8}(y^2-1) - y\left(1 - \sqrt{1 - \frac{\gamma^2}{16}}\right)$. Это и есть «плата» за асимметричность распределения величины S . Глядя на неё нетрудно убедиться, что при стремлении коэффициента асимметрии к нулю ($\gamma \rightarrow 0$) вся эта добавка тоже стремится к нулю.

4.2. Нормальное степенное приближение

Ещё одно приближение, которое использует три момента для аппроксимации случайной величины называется нормальным степенным приближением. Данная аппроксимация очень похожа на изложенную выше (см формулу (6)). Только корректирующая добавка в этом случае выглядит не так громоздко. Нормальное приближение базируется на *разложении Эджуорта*[†] (англ. Edgeworth series). Его мы

* За $\Phi(y)$ здесь и далее мы будем обозначать функцию распределения стандартной нормальной величины.

[†] Фрэнсис Исидор Эджуорт (1845 – 1926) – ирландский британский экономист и статистик.

опишем более подробно чуть позже*, а пока просто воспользуемся результатами этого разложения. Так, если Z - случайная величина с нулевым математическим ожиданием и единичной дисперсией, то имеет место следующее равенство:

$$G(x) = \Phi(x) - a_3\Phi^{(3)}(x) + a_4\Phi^{(4)}(x) + \frac{a_5^2}{2}\Phi^{(6)}(x) + \dots,$$

где $a_k = \frac{1}{k!} \left. \frac{d^k (\log m_Z(t))}{dt^k} \right|_{t=0}$. Используя результаты статьи [2] (см. пункт 6, формула (6.2)),

выпишем несколько первых коэффициентов в явном виде:

$$a_0 = a_1 = 0, \quad a_2 = \frac{1}{2}, \quad a_3 = \frac{1}{6}\gamma_Z^3,$$

В силу данного разложения и метода Ньютона (более подробно об этом можно почитать в книге BEARD, R. E., PENTIKAINEN, T. and PESONEN, E. (1984) Risk Theory Chapman and Hall, pp. 108-111†) получаем, что $Z \approx Y + \frac{\gamma_Z}{6}(Y^2 - 1)$.

Итак, пусть $E(S) = \mu, Var(S) = \sigma^2, \gamma_S = \gamma$, центрируем и нормируем величину S , чтобы получившаяся величина имела нулевое среднее и единичную дисперсию. Тогда имеем:

$$\Phi(s) = P(Y \leq s) \approx P\left(Y + \frac{\gamma}{6}(Y^2 - 1) \leq s + \frac{\gamma}{6}(s^2 - 1)\right) = P\left(\frac{S - \mu}{\sigma} \leq s + \frac{\gamma}{6}(s^2 - 1)\right), s \geq 1. \quad (7)$$

$$P\left(\frac{S - \mu}{\sigma} \leq x\right) = P\left(Y + \frac{\gamma_Z}{6}(Y^2 - 1) \leq x\right) \approx \Phi\left(\sqrt{\frac{9}{\gamma^2} + \frac{6x}{\gamma}} + 1 - \frac{3}{\gamma}\right), x \geq 1. \quad (8)$$

Из вида этих формул ясно, что первая больше подходит для работы с квантилями, в то время как вторая удобнее для аппроксимации функции распределения величины суммарного иска S . Теперь рассмотрим пару конкретных примеров:

Пример 4.2.1. Пусть величина S имеет среднее 10000, коэффициент вариации 1000 и асимметрию 1. Вычислим необходимый капитал для того чтобы покрыть величину суммарного иска с вероятностью 95%.

▷ По формуле (7) получаем: $P\left(\frac{S - \mu}{\sigma} \leq s + \frac{\gamma}{6}(s^2 - 1)\right) \approx \Phi(s) = 0.95$, с помощью

функции НОРМ.СТ.ОБР из Microsoft Excel, которая возвращает обратное значение функции распределения стандартной гауссовой величины, заключаем, что $s = 1.645$. В таком случае требуемая сумма равна

$$E(S) + \sigma_s \left(1.645 + \frac{\gamma}{6}((1.645)^2 - 1)\right) = 10000 + 1000 \left(1.645 + \frac{1}{6}((1.645)^2 - 1)\right) \approx 11929$$

* См. стр. 21.

† [5] в списке литературы.

Таким образом, страховой компании будет достаточно суммы в 11929 условных единиц, чтобы выполнить свои обязательства перед страхователями. <

Пример 4.2.2. Пусть величина суммарного иска S имеет такие же параметры, как и в предыдущем примере. Выбирая разные приближения (нормальное степенное, гамма-приближение, центральную предельную теорему), вычислим и сравним между собой вероятности, с которыми страховая компания не сможет выполнить свои обязательства, если её капитал составляет 13000 условных единиц.

▷ Используя нормальное степенное приближение, по формуле (8) получаем:

$$P(S > 13000) = P\left(\frac{S - \mu}{\sigma} > 3\right) \approx 1 - \Phi\left(\sqrt{9 + 6 \cdot 3 + 1} - 3\right) = 1 - \Phi(2.29) = 0.011$$

Теперь посчитаем эту вероятность, обратившись к сдвинутому гамма-приближению:

$$P(S > 13000) = P(x_0 + \gamma_{\alpha, \beta} > 13000) = 1 - \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^{13000 - x_0} y^{\alpha-1} e^{-\beta y} dy,$$

При помощи соотношений (i)-(iii) посчитаем нужные нам параметры:

$$\alpha = \frac{4}{\gamma^2} = 4, \quad \beta = \frac{2}{\gamma\sigma} = 0.002, \quad x_0 = \mu - \frac{2\sigma}{\gamma} = 10000 - 2000 = 8000.$$

$$\text{Тогда } P(S > 13000) = 1 - \frac{(0.002)^4}{\Gamma(4)} \int_0^{5000} y^3 e^{-0.002y} dy = 1 - 0.989664 \approx 0.010.^*$$

Наконец воспользуемся центральной предельной теоремой:

$$P(S > 13000) = P\left(\frac{S - \mu}{\sigma} > 3\right) = 1 - \Phi(3) \approx 0.0014^+$$

Легко видеть, что центральная предельная теорема, как и утверждалось, дает довольно грубую оценку искомой вероятности. Гамма-приближение и нормальная аппроксимация оказались точнее и при этом неплохо соотносились друг с другом. <

4.3. Приближение ортогональными многочленами

В этом разделе мы опишем метод, позволяющий аппроксимировать плотность случайной величины. Итак, пусть нам дана функция $f(x)$ с носителем (интервалом) $I \subset (-\infty, +\infty)$ (допускается и нестрогое включение), тогда она может быть приближена следующим образом:

* Данный интеграл был посчитан с помощью программы WolframAlpha, (<https://www.wolframalpha.com/>).

+ $\Phi(3)$ посчитано с помощью функции НОРМ.СТ.РАСП из Microsoft Excel, которая возвращает значение функции распределения стандартной гауссовой величины в данной точке.

Возьмем непрерывную функцию $w(x)$, такую что $w(x) > 0, x \in I$. В случае, если концы интервала не ограничены, то накладывается дополнительное условие: необходимо существование интеграла $\int \pi(x)w(x)dx$ для любого многочлена $\pi(x)$. В этом и следующих пунктах данного раздела интеграл берётся по носителю I . Пусть $\pi_0(x), \pi_1(x), \dots$ - последовательность ортогональных многочленов относительно функции $w(x)$, то есть таких что

$$\int \pi_i(x)\pi_k(x)w(x)dx = 0, \text{ при } i \neq k,$$

где $\pi_k(x)$ имеет степень k . Обычно в качестве $\pi_0(x)$ берётся единица, а следующие члены последовательности находятся рекурсивно. Также положим

$$C_k = \int \pi_k^2(x)w(x)dx, \quad k = 0, 1, 2, \dots \quad (9)$$

Наконец, довольно непростая теория математического анализа позволяет нам утверждать, что для гладкой функции $f(x)$ справедливо указанное ниже соотношение:

$$f(x) = A_0\pi_0(x)w(x) + A_1\pi_1(x)w(x) + \dots,$$

где коэффициенты A_0, A_1, \dots определяются следующим образом:

$$\int \pi_k(x)f(x)dx = \int \pi_k(x) \sum_{i=0}^{\infty} A_i\pi_i(x)w(x)dx = A_k \int \pi_k^2(x)w(x)dx = A_k C_k,$$

$$A_k = \frac{\int \pi_k(x)f(x)dx}{C_k}.$$

Их можно интерпретировать как линейную функцию первых k моментов функции $f(x)$. Поэтому если $f(x)$ является плотностью некоторой случайной величины S , то формула для коэффициентов A_k переписется в ином виде:

$$A_k = \frac{E(\pi_k(S))}{C_k}.$$

Таким образом, плотность $f(x)$ может быть аппроксимирована по следующей формуле:

$$f(x) \approx A_0\pi_0(x)w(x) + \dots + A_n\pi_n(x)w(x).$$

Преимущество данного метода состоит в том, что совсем необязательно знать точную формулу для функции $f(x)$, достаточно лишь знать её моменты вплоть до некоторого порядка n .

Следующие два раздела рассматривают более частные случаи.

4.4. Приближение Гамма-функцией Бауэrsa

Пусть S - случайная величина с плотностью $f(x)$, принимающая неотрицательные значения. Для начала аппроксимируем эту плотность уже знакомой нам гамма-плотностью, сохраняя среднее и дисперсию. Тогда получаем, что $E(S) = \frac{\alpha}{\lambda}$, $Var(S) = \frac{\alpha}{\lambda^2}$, где α и λ параметры указанного гамма-распределения, которые могут быть выражены следующим образом: $\lambda = \frac{E(S)}{Var(S)}$, $\alpha = \frac{(E(S))^2}{Var(S)}$.

Для удобства введём нормированную случайную величину $Z = \beta \cdot S$, где $\beta = \frac{E(S)}{Var(S)}$ и пусть $g(z), z > 0$ плотность этой случайной величины. Заметим, что для Z параметр λ будет равен единице (что объясняет почему данная случайная величина названа нормированной). Действительно, $E(Z) = E(\beta \cdot S) = \frac{(E(S))^2}{Var(S)}$ и $Var(Z) = Var(\beta \cdot S) = \frac{(E(S))^2}{(Var(S))^2} Var(S) = \frac{(E(S))^2}{Var(S)}$, поэтому $\frac{E(Z)}{Var(Z)} = 1$. Кроме того, данные выкладки позволяют убедиться в том, что $\alpha = E(Z) = Var(Z)$.

Теперь обратимся к результатам подраздела 4.3. Для приближения $g(z)$ в качестве носителя I возьмём интервал $(0, +\infty)$, а непрерывную положительную функцию $w(x)$ положим равной $\frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}$. Для многочленов будем использовать следующие функции:

$$L_k(x) = (-1)^k x^{1-\alpha} e^x \frac{d^k}{dx^k} (x^{k+\alpha-1} e^{-x}), \quad (10)$$

где параметр α как и в определении $w(x)$ равен $E(Z) = Var(Z)$. Отметим, что указанные выше функции $L_k(x)$ называются полиномами Лагерра*. Напишем первые четыре из них в явном виде:

$$L_0(x) = 1, \quad L_1(x) = x - \alpha, \quad L_2(x) = x^2 - 2(\alpha+1)x + \alpha(\alpha+1),$$

$$L_3(x) = x^3 - 3(\alpha+2)x^2 + 3(\alpha+2)(\alpha+1)x - (\alpha+2)(\alpha+1)\alpha.$$

Подробно докажем следующее утверждение, которое было опущено в анализируемой нами монографии [2]:

* Эдмóн Николá Лагéрр (1834 – 1886) – французский математик, член Парижской академии наук.

Утв. 4.4.1. Полиномы Лагерра ортогональны относительно функции $w(x)$. Кроме того, для коэффициентов C_k справедлива следующая формула: $C_k = k! \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)}$, $k = 0, 1, \dots$

▷ Докажем сразу для полинома произвольной степени n . Для начала рассмотрим интеграл:

$$\frac{1}{\Gamma(\alpha)} \int_0^{+\infty} x^n \cdot x^{\alpha-1} e^{-x} L_n(x) dx = \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} x^n (-1)^n \frac{d^n}{dx^n} (x^{n+\alpha-1} e^{-x}) dx,$$

проинтегрировав его по частям n раз имеем:

$$\begin{aligned} \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} x^n \cdot x^{\alpha-1} e^{-x} L_n(x) dx &= \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} (-1)^n n! (-1)^n \frac{d^{n-n}}{dx^{n-n}} (x^{n+\alpha-1} e^{-x}) dx = \\ &= \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} (-1)^{2n} n! (x^{n+\alpha-1} e^{-x}) dx = \frac{n!}{\Gamma(\alpha)} \int_0^{+\infty} x^{n+\alpha-1} e^{-x} dx = n! \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)}. \end{aligned}$$

Обратим внимание на то, что если взять степень переменной x меньше n , то интеграл будет равен нулю. Действительно,

$$\frac{1}{\Gamma(\alpha)} \int_0^{+\infty} x^m x^{\alpha-1} e^{-x} L_n(x) dx = (-1)^{m+n} m! \int_0^{+\infty} \frac{d^{n-m}}{dx^{n-m}} (x^{n+\alpha-1} e^{-x}) dx = (-1)^{m+n} m! \left. \frac{d^{n-m-1}}{dx^{n-m-1}} (x^{n+\alpha-1} e^{-x}) \right|_0^{+\infty} = 0,$$

при дифференцировании выражения в скобках $n-m-1$ раз, мы получим многочлен, умноженный на e^{-x} , причём минимальная степень переменной x этого многочлена равна $\alpha > 0$. Поэтому при подстановке нуля мы получим также ноль (ведь многочлен обратится в ноль и его произведение с $e^{-0} = 1$ даст снова ноль). При подстановке бесконечности за счёт экспоненты в отрицательной степени мы также получим ноль.

Используя правило Лейбница, запишем эквивалентное представление для многочлена Лагерра n -ой степени:

$$\begin{aligned} L_n(x) &= (-1)^n x^{1-\alpha} e^x \frac{d^n}{dx^n} (x^{n+\alpha-1} e^{-x}) = (-1)^n x^{1-\alpha} e^x \sum_{k=0}^n C_n^k (x^{n+\alpha-1})^{(n-k)} (e^{-x})^{(k)} = \\ &= (-1)^n x^{1-\alpha} e^x \sum_{k=0}^n C_n^k \frac{(n+\alpha-1)!}{(k+\alpha-1)!} x^{k+\alpha-1} (-1)^k e^{-x} = (-1)^n \sum_{k=0}^n C_n^k \frac{(n+\alpha-1)!}{(k+\alpha-1)!} x^k (-1)^k. \end{aligned}$$

Поэтому при подстановке в формулу $\int_0^{+\infty} L_m(x) L_n(x) w(x) dx$ полинома $L_m(x)$, $m < n$,

получим ноль. В силу симметрии это распространяется на все многочлены различных степеней. Таким образом, мы доказали ортогональность.

Теперь получим формулу для C_n . Подставляем $L_n(x)$ в $C_n = \int_0^{+\infty} L_n(x) L_n(x) w(x) dx$ и

меняем местами знаки суммирования и интеграла:

$$\begin{aligned}
C_n &= (-1)^n \sum_{k=0}^n \int_0^{+\infty} C_n^k \frac{(n+\alpha-1)!}{(k+\alpha-1)!} x^k (-1)^k L_n(x) w(x) dx = \\
&= (-1)^n \sum_{k=0}^n \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} C_n^k \frac{(n+\alpha-1)!}{(k+\alpha-1)!} x^k (-1)^k x^{\alpha-1} e^{-x} L_n(x) dx.
\end{aligned}$$

Исходя из замечаний, сделанных ранее, все слагаемые, где степень x меньше n обращаются в ноль. Поэтому:

$$\begin{aligned}
C_n &= (-1)^n \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} C_n^n \frac{(n+\alpha-1)!}{(n+\alpha-1)!} x^n (-1)^n x^{\alpha-1} e^{-x} L_n(x) dx = \\
&= \frac{(-1)^{2n}}{\Gamma(\alpha)} \int_0^{+\infty} x^n \cdot x^{\alpha-1} e^{-x} L_n(x) dx = n! \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)},
\end{aligned}$$

и формула для коэффициентов C_k доказана. \triangleleft

Для коэффициентов A_k имеем следующие соотношения (доказательство можно найти в статье [2] на странице 57 раздела 4 главы 4):

$$A_0 = 1, \quad A_1 = A_2 = 0, \quad A_3 = \frac{\Gamma(\alpha)}{3! \Gamma(\alpha+3)} (g_3 - (\alpha+2)(\alpha+1)\alpha),$$

где $g_k = E(Z^k)$.

Таким образом, довольно грубая аппроксимация плотности случайной величины $Z = \beta \cdot S$ примет вид:

$$g(z) \approx w(z) + A_3 L_3(z) w(z),$$

для функции распределение имеем:

$$G(z) \approx W(z) + A_3 \cdot \int_0^z L_3(x) w(x) dx,$$

где $W(z)$ есть функция распределения гамма величины с параметрами α и $\lambda=1$

$\left(W(z) = \frac{1}{\Gamma(\alpha)} \int_0^z x^{\alpha-1} e^{-x} dx \right)$, значения которой можно найти в специальных таблицах и статистических пакетах. Используя формулу (10), а также явное выражение для функции $w(x)$, можем переписать интеграл $\int_0^z L_3(x) w(x) dx$ следующим образом:

$$\int_0^z L_3(x) w(x) dx = \int_0^z (-1)^3 x^{1-\alpha} e^x \frac{d^3}{dx^3} (x^{3+\alpha-1} e^{-x}) \frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)} dx = -\frac{1}{\Gamma(\alpha)} \int_0^z \frac{d^3}{dx^3} (x^{2+\alpha} e^{-x}) dx =$$

$$\begin{aligned}
&= -\frac{1}{\Gamma(\alpha)} \cdot \frac{d^2}{dx^2} \left(x^{2+\alpha} e^{-x} \right) \Big|_0^z = \frac{-1}{\Gamma(\alpha)} \cdot \left((2+\alpha)(1+\alpha) x^\alpha e^{-x} - 2(2+\alpha) x^{1+\alpha} e^{-x} + x^{2+\alpha} e^{-x} \right) \Big|_0^z = \\
&= \frac{-z^\alpha e^{-z}}{\Gamma(\alpha)} \left((2+\alpha)(1+\alpha) - 2(2+\alpha)z + z^2 \right).
\end{aligned}$$

В таком случае, подставив найденное ранее значение коэффициента A_3 , для приближения интересующей нас функции распределения $G(z)$ имеем:

$$\begin{aligned}
G(z) \approx W(z) - \frac{\Gamma(\alpha)}{3!\Gamma(\alpha+3)} \left(g_3 - (\alpha+2)(\alpha+1)\alpha \right) \frac{z^\alpha e^{-z}}{\Gamma(\alpha)} \left((2+\alpha)(1+\alpha) - 2(2+\alpha)z + z^2 \right) = W(z) - \\
- \frac{z^\alpha e^{-z}}{6} \left(g_3 - (\alpha+2)(\alpha+1)\alpha \right) \left(\frac{(2+\alpha)(1+\alpha)}{\Gamma(\alpha+3)} - \frac{2(2+\alpha)z}{\Gamma(\alpha+3)} + \frac{z^2}{\Gamma(\alpha+3)} \right)
\end{aligned}$$

Применяя известное свойство гамма функции ($\Gamma(\alpha+1) = \alpha \cdot \Gamma(\alpha)$), можем переписать это выражение в более простом виде:

$$G(z) \approx W(z) - z^\alpha e^{-z} \left(\frac{1}{\Gamma(\alpha+1)} - \frac{2z}{\Gamma(\alpha+2)} + \frac{z^2}{\Gamma(\alpha+3)} \right) \cdot A,$$

где $A = \frac{1}{6} \left(g_3 - (\alpha+2)(\alpha+1)\alpha \right).$

В итоге для получения желаемой аппроксимации функции распределения S остается только подставить в качестве аргумента функции $G(z)$ $z = \beta \cdot x$. Действительно, $F_S(x) = P(S < x) = P(\beta \cdot S < \beta \cdot x) = P(Z < \beta \cdot x) = G(\beta \cdot x).$

4.5. Аппроксимация Грама – Шарлье*

Пусть случайная величина S имеет среднее μ и дисперсию σ^2 . В этот раз будем аппроксимировать её при помощи плотности стандартного нормального распределения. В прошлом пункте мы рассматривали нормированную величину $Z = \beta \cdot S$, в этом поступим похожим образом, а именно, рассмотрим $Z = \frac{S - \mu}{\sigma}$. Тогда $E(Z) = 0$ и $Var(Z) = 1$. Аналогично предыдущему подразделу за $g(z)$ и $G(z)$ обозначим соответственно плотность и функцию распределения величины Z . Вновь будем активно использовать результаты пункта 4.3 и в качестве носителя I возьмём всю числовую прямую $(-\infty, +\infty)$, а в качестве непрерывной положительной функции $w(x)$ плотность стандартной

* Йёрген Педерсон Грам (1850 – 1916) – датский актуарий и математик.
Карл Вильгельм Людвиг Шарлье (1862 – 1934) – шведский астроном.

нормальной величины $\left(\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right)$. Что касается многочленов, то в этом случае будем использовать полиномы Эрмита*. Они задаются по следующему правилу:

$$H_k(x) = \frac{\phi^{(k)}(x)}{\phi(x)}, \quad k = 0, 1, \dots$$

Выпишем несколько в явном виде:

$$H_0(x) = 1, \quad H_1(x) = -x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = -(x^3 - 3x), \quad H_4(x) = x^4 - 6x^2 + 3.$$

Ровно, как и в случае с полиномами Лагерра доказательства ортогональности многочленов Эрмита и явного вида формулы C_n в монографии [2] опущены. Поэтому мы приведем их в данной работе, однако, чтобы сосредоточиться на более важных вещах, мы всё же сошлемся на один результат. Так, для полиномов Эрмита справедливо следующее

представление: $H_n(x) = (-1)^n \sum_{j=0}^{\lfloor n/2 \rfloor} (-1)^j \frac{n!}{j!(n-2j)!} \frac{x^{n-2j}}{2^j}$. Доказательство этого факта можно

найти на странице 4 раздела 1.3 главы 1 статьи, указанной под номером [6] в списке литературы.

Утв. 4.5.1. Многочлены Эрмита ортогональны относительно функции $w(x) = \phi(x)$.

Док-во: ▷ Зафиксируем произвольное натуральное число n и рассмотрим интеграл

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^m \cdot e^{-\frac{x^2}{2}} H_n(x) dx, \quad \text{где } m < n, m \in \mathbb{Z}_+.$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^m \cdot e^{-\frac{x^2}{2}} H_n(x) dx = \int_{-\infty}^{+\infty} x^m \cdot \left(e^{-\frac{x^2}{2}} \right)^{(n)} dx = x^m \cdot \left(e^{-\frac{x^2}{2}} \right)^{(n-1)} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} m \cdot x^{m-1} \cdot \left(e^{-\frac{x^2}{2}} \right)^{(n-1)} dx$$

Произведя интегрирование по частям, заметим, что первое слагаемое равно 0.

Действительно, $\lim_{x \rightarrow \infty} \left(x^m \cdot e^{-\frac{x^2}{2}} \right) = 0$ для любого целого m . Следовательно, проинтегрировав ещё $m-1$ раз получим:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^m \cdot e^{-\frac{x^2}{2}} H_n(x) dx = (-1)^m m! \int_{-\infty}^{+\infty} \left(e^{-\frac{x^2}{2}} \right)^{(n-m)} dx = (-1)^m m! \left(e^{-\frac{x^2}{2}} \right)^{(n-m-1)} \Big|_{-\infty}^{+\infty} = 0$$

Теперь возьмем два полинома различных степеней $H_n(x)$ и $H_m(x)$ (пусть для определенности m всё также меньше n) и посмотрим на следующий интеграл:

* Шарль Эрмит (1822 – 1901) – французский математик, признанный лидер математиков Франции во второй половине 19-го века.

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} H_m(x) H_n(x) dx.$$

По доказанной ранее лемме $H_m(x)$ - многочлен степени m , причем $m < n$. Подставив явный вид для этого полинома и поменяв местами знаки суммирования и интеграла, получим сумму из нулевых слагаемых. Итак, $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} H_m(x) H_n(x) dx = 0$. \triangleleft

Утв. 4.5.2. Для коэффициентов C_k справедливо следующее соотношение:

$$C_k = k!, \quad k = 0, 1, \dots$$

Док-во: \triangleright В соответствии с формулой (9) коэффициенты C_k могут быть найдены из соотношения: $C_k = \int_{-\infty}^{+\infty} H_k(x) H_k(x) w(x) dx$. $H_k(x) = (-1)^k \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \frac{(k+1)!}{j!(k-2j)!} \frac{x^{k-2j}}{2^j}$, поэтому исходя из замечания в начале предыдущего утверждения после подстановки этого выражения вместо $H_k(x)$ и перестановке знаков суммирования и интегрирования, получаем, что только одно слагаемое не окажется равным нулю:

$$\begin{aligned} C_k &= \int_{-\infty}^{+\infty} (-1)^k \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \frac{k!}{j!(k-2j)!} \frac{x^{k-2j}}{2^j} H_k(x) w(x) dx = \\ &= (-1)^k \sum_{j=0}^{\lfloor k/2 \rfloor} (-1)^j \frac{k!}{j!(k-2j)!} \int_{-\infty}^{+\infty} \frac{x^{k-2j}}{2^j} H_k(x) w(x) dx = (-1)^k \int_{-\infty}^{+\infty} x^k H_k(x) w(x) dx = \\ &= (-1)^k \int_{-\infty}^{+\infty} x^k \frac{\phi^{(k)}(x)}{\phi(x)} w(x) dx = (-1)^k \int_{-\infty}^{+\infty} x^k \phi^{(k)}(x) dx = (-1)^k \int_{-\infty}^{+\infty} x^k \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right)^{(k)} dx \end{aligned}$$

Проинтегрировав по частям k раз получаем:

$$C_k = (-1)^{2k} \int_{-\infty}^{+\infty} k! \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \right)^{(k-k)} dx = k! \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = k! \triangleleft$$

Используя результаты статьи [2] (см. страницу 59 раздел 5 главы 4) запишем первые коэффициенты A_k :

$$A_0 = 1, \quad A_1 = A_2 = 0, \quad A_3 = -\frac{g_3}{3!}, \quad A_4 = \frac{g_4 - 3}{4!}.$$

Таким образом, плотность величины Z может быть приближена в соответствии с следующим выражением:

$$g(z) = w(z) + A_3 H_3(z) w(z) + A_4 H_4(z) w(z) = \phi(z) + A_3 \phi^{(3)}(z) + A_4 \phi^{(4)}(z) + \dots, \quad (11)$$

для функции распределения имеем:

$$G(z) = \Phi(z) + A_3 \Phi^{(3)}(z) + A_4 \Phi^{(4)}(z) + \dots \quad (12)$$

Произведя замену $z = \frac{x-\mu}{\sigma}$ получаем приближение для функции распределения S . В самом деле,

$$F_s(x) = P(S < x) = P\left(\frac{S-\mu}{\sigma} < \frac{x-\mu}{\sigma}\right) = P\left(Z < \frac{x-\mu}{\sigma}\right) = G\left(\frac{x-\mu}{\sigma}\right),$$

и поэтому $F(x) \approx \Phi\left(\frac{x-\mu}{\sigma}\right) + A_3 \Phi^{(3)}\left(\frac{x-\mu}{\sigma}\right) + A_4 \Phi^{(4)}\left(\frac{x-\mu}{\sigma}\right)$.

Пример 4.5.1. Пусть случайная величина S имеет составное пуассоновское распределение с параметрами λ и $F(x)$. Тогда применяя описанную выше аппроксимацию, имеем:

$$g(z) = \phi(z) - \lambda^{-\frac{1}{2}} \frac{p_3}{3! p_2^{\frac{3}{2}}} \phi^{(3)}(z) + \lambda^{-1} \frac{p_4}{4! p_2^2} \phi^{(4)}(z) + \dots,$$

$$G(z) = \Phi(z) - \lambda^{-\frac{1}{2}} \frac{p_3}{3! p_2^{\frac{3}{2}}} \Phi^{(3)}(z) + \lambda^{-1} \frac{p_4}{4! p_2^2} \Phi^{(4)}(z) + \dots,$$

где $p_k = \int x^k dF(x)$.

► Обоснование этого факта имеется на странице 59 в разделе 5 главы 4 статьи [2]. ◀

Заметим, что данное приближение имеет существенный недостаток. Как было замечено в начале этого пункта* левая часть формулы (12) с ростом среднего числа исков (то есть λ) сходится по распределению к величине, имеющей стандартное нормальное распределение. В соответствии с этим можно показать, что для всех $k \geq 3$ коэффициенты A_k стремятся к нулю при $\lambda \rightarrow \infty$. Это ясно видно на примере формул, указанных выше, для A_3 и A_4 . Однако скорость этой сходимости оказывается не совсем удовлетворительной и в некоторой степени противоречит нашим ожиданиям. Так, например, коэффициент A_5 будет иметь порядок роста $O(\lambda^{-\frac{3}{2}})$, и всё на первый взгляд хорошо, ведь он стремится к нулю быстрее A_4 , который в свою очередь осуществляет это быстрее A_3 . Тем не менее следующий по порядку коэффициент разрушает это свойство: A_6 будет иметь порядок роста снова равный $O(\lambda^{-1})$. Дальнейшее исследование показывает, что коэффициенты A_7 и A_9 будут иметь аналогичный порядок роста. Всё это указывает на то, что указанные коэффициент сходятся довольно медленно и вносят ощутимый вклад в аппроксимацию. Поэтому если мы зафиксируем некоторый номер и отбросим оставшийся хвост ряда, то получим довольно грубое приближение. То есть применение данной аппроксимации будет оправдано, если мы рассматриваем достаточно большой начальный отрезок ряда, что на

* См. страницу 5.

практике не всегда может быть осуществимо, либо требует больших вычислительных и временных затрат.

В связи с возникшей проблемой имеет смысл рассмотреть разложение Эджуорта, которое в некотором роде решает данный вопрос. Здесь наблюдается совсем другая картина: бесконечный ряд, вообще говоря, может даже не сходиться, в то время как его усеченная часть дает хорошую аппроксимацию рассматриваемого распределения.

4.6. Разложение Эджуорта

Ранее мы уже пользовались этим разложением*, теперь пришло время рассмотреть его более подробным образом. Пусть случайная величина Z как и в предыдущем пункте есть центрированная и нормированная величина S (то есть $E(Z)=0$ и $Var(Z)=1$). Обозначим за $M(t)$ п. ф. м. случайной величины Z и запишем ряд Тейлора для её логарифма в нуле:

$$\log M(t) = a_0 + a_1 t + a_2 t^2 + \dots, \quad (13)$$

$$\text{где } a_k = \frac{1}{k!} \left. \frac{d^k \log M(t)}{dt^k} \right|_{t=0}.$$

Ранее мы уже находили коэффициенты $a_i, i=0,1,2,3$, вновь используя результаты статьи [2], представленные на странице 60 раздела 6 главы 4, заключаем, что $a_4 = \frac{g_4 - 3}{24}$. Следующий шаг состоит в возведении формулы (13) в экспоненту и вынесении первого члена разложения:

$$M(t) = \exp(a_2 t^2 + a_3 t^3 + a_4 t^4 + \dots) = \exp\left(\frac{t^2}{2}\right) \cdot \exp(a_3 t^3 + a_4 t^4 + \dots).$$

Применяя формулу Тейлора для экспоненты, можем переписать это соотношение в ином виде:

$$M(t) = \exp\left(\frac{t^2}{2}\right) \cdot \left(1 + a_3 t^3 + a_4 t^4 + \frac{1}{2} a_3^2 t^6 + \dots\right). \quad (14)$$

Несложным упражнением по теории вероятностей является доказательство следующего тождества: $t^k \cdot e^{\frac{t^2}{2}} = (-1)^k \int_{-\infty}^{+\infty} e^{tx} \phi^{(k)}(x) dx, k \in \mathbb{N}$ (необходимо лишь уметь интегрировать по частям и знать, что интеграл Пуассона равен корню из π).

* См. страницу 13.

Таким образом, п. ф. м. стандартной нормальной плотности в точности равна $e^{\frac{t^2}{2}}$, п. ф. м. $-\phi'(x)$ есть $t \cdot e^{\frac{t^2}{2}}$ и так далее. В силу соотношения (14) это означает, что $M(t)$ равна п. ф. м. для $\phi(x) - a_3\phi'''(x) + a_4\phi^{(4)}(x) + \frac{a_3^2}{2}\phi^{(6)}(x) + \dots$. Поскольку существует взаимно-однозначное соответствие между распределениями и хар. ф., а п. ф. м. в свою очередь выражается через характеристическую функцию, заключаем, что для плотности величины Z справедливо следующее соотношение:

$$g(z) = \phi(z) - a_3\phi'''(z) + a_4\phi^{(4)}(z) + \frac{a_3^2}{2}\phi^{(6)}(z) + \dots \quad (15)$$

Такое представление для функции $g(z)$ и называется разложением Эджуорта. В математике такой ряд обычно называют *формальным* рядом, поскольку во многих случаях он расходится. Но с точки зрения аппроксимации, его усеченная часть может давать достаточно хорошее приближение. Разложение для функции распределения случайной величины Z получается из формулы (15) заменой всех ϕ на Φ :

$$G(z) = \Phi(z) - a_3\Phi^{(3)}(z) + a_4\Phi^{(4)}(z) + \frac{a_3^2}{2}\Phi^{(6)}(z) + \dots$$

Пример 4.6.1. Пусть случайная величина S как и в примере 4.5.1 имеет составное пуассоновское распределение с параметрами λ и $F(x)$. Тогда разложение Эджуорта дает следующий результат:

$$g(z) = \phi(z) - \lambda^{\frac{1}{2}} \cdot \frac{P_3}{6p_2^{\frac{3}{2}}} \phi'''(z) + \lambda^{-1} \left(\frac{P_4}{24p_2^2} \phi^{(4)}(z) + \frac{P_3^2}{72p_2^3} \phi^{(6)}(z) \right) + \dots,$$

$$G(z) = \Phi(z) - \lambda^{\frac{1}{2}} \cdot \frac{P_3}{6p_2^{\frac{3}{2}}} \Phi^{(3)}(z) + \lambda^{-1} \left(\frac{P_4}{24p_2^2} \Phi^{(4)}(z) + \frac{P_3^2}{72p_2^3} \Phi^{(6)}(z) \right) + \dots,$$

где сделав замену $z = \frac{x - \lambda p_1}{\lambda p_2}$, получаем аппроксимацию для функции распределения случайной величины S (как и до этого $p_k = \int x^k dF(x)$).

► Подробное доказательство можно найти на стр. 62 раздела 6 главы 4 монографии [2]. ◀

Сделаем несколько важных выводов касательно данной аппроксимации:

1. Если в формуле (15) использовать в качестве усеченного ряда больше слагаемых, то в результате аппроксимирующая плотность для функции $g(z)$ может быть отрицательной, что противоречит самому определению плотности. Естественно, в этих случаях указанное приближение неприменимо. Однако стоит отметить, что данное замечание справедливо, как для аппроксимации гамма функцией Бауэrsa, так и для приближения Грама-Шарлье.

2. Практический опыт показывает, что разложение Эджуорта даёт хорошее приближение для той части функции распределения, которой соответствуют наиболее вероятные значения случайной величины, в то время как для хвостов оно часто не совпадает с реальными данными. Для решения этой проблемы используется метод Эшера*, который будет рассмотрен в следующем разделе.
3. В силу того, что разложение Эджуорта, вообще говоря, расходится, совсем необязательно, что качество аппроксимации улучшится с увеличением числа слагаемых в рассматриваемом отрезке ряда.

4.7. Аппроксимация Эшера

Пусть S некоторая случайная величина с функцией распределения $F(x)$ и п. ф. м. $M(t)$. Каждому произвольному действительному числу h поставим в соответствие случайную величину \bar{S}_h , функция распределения $\bar{F}_h(x)$ и п. ф. м. $\bar{M}_h(t)$ которой будут задаваться по следующему правилу:

$$d\bar{F}_h(x) = \frac{e^{hx} dF(x)}{M(h)} \quad (16)$$

или

$$\bar{F}_h(x) = \int_{-\infty}^x e^{hy} d \frac{F(y)}{M(h)} \quad (17)$$

Функцию $\bar{F}_h(x)$ принято называть *преобразованием Эшера* функции $F(x)$ с параметром h . Таким образом, мы имеем бесконечное семейство преобразований (ведь h может быть любым действительным числом). С помощью формулы (17) можно показать, что для п. ф. м. величины \bar{S}_h верно следующее соотношение:

$$\bar{M}_h(t) = \frac{M(t+h)}{M(h)}. \quad (18)$$

$$\text{Действительно, } \bar{M}_h(t) = \int_{-\infty}^{+\infty} e^{tx} d\bar{F}_h(x) = \frac{1}{M(h)} \int_{-\infty}^{+\infty} e^{tx} e^{hx} dF(x) = \frac{M(t+h)}{M(h)}.$$

Следующий пример показывает, как работает преобразование Эшера с известными распределениями. Отметим, что в рассматриваемой статье [2] обоснование этого примера опущено, но мы восполним этот пробел и предоставим детальное обоснование.

Пример 4.7.1. Если величина S имеет пуассоновское распределение с параметром λ , то её преобразование Эшера \bar{S}_h снова будет пуассоновской величиной, но уже с другим параметром λe^h . В случае, когда S есть экспоненциально распределенная случайная

* Ф. Эшер – шведский актуарий.

величина с параметром λ , её преобразование Эшера будет иметь то же распределение с параметром $\lambda - h$ (в этом случае преобразование Эшера определено только при $h < \lambda$).

▷ Итак, пусть $S \sim Poiss(\lambda)$, тогда для п. ф. м. имеем $M(t) = \exp(\lambda(e^t - 1))$, в силу

$$\text{соотношения (18): } \bar{M}_h(t) = \frac{M(t+h)}{M(h)} = \frac{\exp(\lambda(e^{t+h} - 1))}{\exp(\lambda(e^h - 1))} = \exp(\lambda e^{t+h} - \lambda e^h) = \exp(\lambda e^h (e^t - 1)).$$

Следовательно, $\bar{S}_h \sim Poiss(\lambda e^h)$.

Почти аналогично проводится доказательство для $S \sim Exp(\lambda)$. Действительно,

$$M(t) = \left(1 - \frac{t}{\lambda}\right)^{-1}, \quad \text{поэтому} \quad \bar{M}_h(t) = \frac{\left(1 - \frac{t+h}{\lambda}\right)^{-1}}{\left(1 - \frac{h}{\lambda}\right)^{-1}} = \frac{\lambda - h}{\lambda - t - h} = \left(\frac{\lambda - h - t}{\lambda - h}\right)^{-1} = \left(1 - \frac{t}{\lambda - h}\right)^{-1},$$

откуда заключаем, что $S \sim Exp(\lambda - h)$. ◁

Пример 4.7.2. Пусть случайная величина S имеет составное пуассоновское распределение с параметрами λ и п. ф. м. $m(t)$. Тогда \bar{S}_h снова имеет составное пуассоновское распределение с параметрами $\lambda \cdot m(h)$ и п. ф. м., которая есть преобразование Эшера исходной.

▷ Обоснование содержится в статье [2] на стр. 63 раздела 7 главы 6. ◁

Обратим внимание на одно важное свойство преобразования Эшера, которым мы будем пользоваться позже при выводе аппроксимации: математическое ожидание $E(\bar{S}_h)$ есть возрастающая функция по переменной h . Не лишним будет сказать, что объяснение этого факта монография [2] не содержит, но мы предоставим необходимое доказательство:

поскольку $E(\bar{S}_h) = \bar{M}'(0) = \frac{M'(h)}{M(h)}$ то для производной по h имеем:

$$\frac{dE(\bar{S}_h)}{dh} = \frac{M''(h)M(h) - (M'(h))^2}{(M(h))^2} = \bar{M}''(0) - (\bar{M}'(0))^2 = Var(S) \geq 0, \quad \text{следовательно,}$$

среднее не убывает по h . Будем рассматривать случай, когда дисперсия величины суммарного иска положительна, и тогда $E(\bar{S}_h)$ строго возрастает по h .

Предположим, что мы хотим найти $F(x) = P(S < x)$ для некоторого конкретного значения x . Такая ситуация вполне понятна, например, владелец страховой компании, хочет понять, насколько некоторая сумма x надежна в смысле покрытия величины суммарного иска. Если такая вероятность окажется равной, скажем 95%, то с достаточно большой уверенностью можно говорить о том, что суммы x хватит для осуществления всех выплат по поступившим искам. Основная идея данной аппроксимации состоит в

нахождении числа h такого, что $E(\bar{S}_h) = \frac{M'(h)}{M(h)} = x$. В силу указанного выше свойства функции среднего, такое h существует.

Следующий шаг состоит в применении приближения Эджуорта к плотности \bar{f} величины \bar{S}_h :

$$\bar{f}(y) \approx \phi(z) - a_3 \phi^{(3)}(z),$$

где $z = (y - E(\bar{S}_h)) / (Var(\bar{S}_h))^{1/2} = (y - x) / (Var(\bar{S}_h))^{1/2}$, поскольку разложения применимо для величины с нулевым средним и единичной дисперсией. Как было замечено ранее

$$a_3 = \frac{g_3}{6} = \frac{\sigma^{-3} E((\bar{S}_h - E(\bar{S}_h))^3)}{6} = \frac{E((\bar{S}_h - E(\bar{S}_h))^3)}{6(Var(\bar{S}_h))^{3/2}}, \text{ поэтому для плотности } \bar{S}_h \text{ имеем:}$$

$$\bar{f}(y) \approx \phi(z) - \frac{E((\bar{S}_h - E(\bar{S}_h))^3)}{6(Var(\bar{S}_h))^{3/2}} \phi^{(3)}(z).$$

В силу формулы **(16)** $f(y) = M(h) e^{-hy} \bar{f}(y)$, проинтегрировав это равенство от x до $+\infty$, получаем:

$$1 - F(x) = M(h) \int_x^{+\infty} e^{-hy} \bar{f}(y) dy, \quad (19)$$

или

$$F(x) = M(h) \int_{-\infty}^x e^{-hy} \bar{f}(y) dy \quad (20)$$

На первый взгляд, кажется, что между указанными формулами почти нет разницы и неважно в какую из них подставлять выражение для $\bar{f}(y)$. Однако на практике этот выбор имеет большое значение. Нюанс скрывается во втором выводе предыдущего раздела: аппроксимация Эджуорта немного искажает данные о хвостах распределения. В связи с этим, выбор производится на основе знака числа h . Если $h > 0$ (поскольку $E(\bar{S}_h)$ возрастает, то в этом случае $x = E(\bar{S}_h) > E(\bar{S}_0) = M'(0) = E(S)$), то предпочтительнее оказывается формула **(19)**. Действительно, для y , лежащих в некоторой окрестности x , разложение Эджуорта обеспечивает хорошее приближение. В то время как для больших значений переменной y , как было отмечено, хвосты не соответствуют действительности. Но благодаря множителю e^{-hy} , который оказывается очень близким к нулю в силу положительного значения h , это искажение сглаживается, и аппроксимация оказывается более точной. Таким образом, для получения *приближения Эшера* при $h > 0$ рациональнее оказывается использовать именно формулу **(19)**:

$$1 - F(x) \approx M(h) \int_x^{+\infty} \exp\left(-h\left(z \cdot \left(\text{Var}(\bar{S}_h)\right)^{\frac{1}{2}} + x\right)\right) \phi(z) dz -$$

$$-M(h) \int_x^{+\infty} \exp\left(-h\left(z \cdot \left(\text{Var}(\bar{S}_h)\right)^{\frac{1}{2}} + x\right)\right) \frac{E\left(\left(\bar{S}_h - E(\bar{S}_h)\right)^3\right)}{6\left(\text{Var}(\bar{S}_h)\right)^{\frac{3}{2}}} \phi^{(3)}(z) dz,$$

что можно переписать в следующем виде:

$$1 - F(x) \approx M(h) e^{-hx} \left(E_0(u) - \frac{E\left(\left(\bar{S}_h - E(\bar{S}_h)\right)^3\right)}{6\left(\text{Var}(\bar{S}_h)\right)^{\frac{3}{2}}} E_3(u) \right) \quad (21)$$

где $u = h\left(\text{Var}(\bar{S}_h)\right)^{\frac{1}{2}}$ и $E_k(u) = \int_0^{+\infty} e^{-uz} \phi^{(k)}(z) dz$, $k = 0, 1, \dots$ (такие функции принято называть функциями Эшера).

Поскольку $e^{-uz} \phi(z) = e^{-uz} e^{-\frac{z^2}{2}} = e^{\frac{u^2}{2}} e^{-\frac{(z+u)^2}{2}} = e^{\frac{u^2}{2}} \phi(z+u)$, то сделав замену в интеграле, получаем следующее выражение для $E_0(u)$:

$$E_0(u) = \int_0^{+\infty} e^{-uz} \phi(z) dz = e^{\frac{u^2}{2}} \int_0^{+\infty} \phi(z+u) dz = e^{\frac{u^2}{2}} \int_u^{+\infty} \phi(t) dt = e^{\frac{u^2}{2}} (1 - \Phi(u)).$$

Для получения рекуррентного соотношения для функций Эшера проинтегрируем по частям общую формулу для $E_k(u)$:

$$E_k(u) = \int_0^{+\infty} e^{-uz} \phi^{(k)}(z) dz = e^{-uz} \phi^{(k-1)}(z) \Big|_0^{+\infty} + u \int_0^{+\infty} e^{-uz} \phi^{(k-1)}(z) dz = -\phi^{(k-1)}(0) + u E_{k-1}(u).$$

Поэтому

$$E_1(u) = -\frac{1}{\sqrt{2\pi}} + u E_0(u),$$

$$E_2(u) = u E_1(u) = -\frac{u}{\sqrt{2\pi}} + u^2 E_0(u),$$

$$E_3(u) = \frac{1}{\sqrt{2\pi}} + u E_2(u) = \frac{1-u^2}{\sqrt{2\pi}} + u^3 E_0(u).$$

Стоит отметить, что для функций Эшера существуют специальные таблицы, хотя в силу развития вычислительной техники их применение на практике в последнее время существенно сократилось.

Следующий пример вместе с доказательством содержится в разделе 7 главы 4 монографии [2] на стр. 65:

Пример 4.7.3. В условиях предыдущего примера для величины S имеем следующее приближение:

$$1 - F(x) \approx e^{\lambda(m(t)-1)-hx} \left(E_0(u) - \frac{m'''(h)}{6\lambda^{1/2}(m''(h))^{3/2}} E_3(u) \right),$$

где $u = h(\lambda m''(h))^{1/2}$.

5. Связь между моделями индивидуального и коллективного риска

5.1. Описание модели индивидуального иска

Как уже было сказано, в модели индивидуального риска величина суммарного иска по всему портфелю представляется в виде суммы исков по всем договорам страхования (считается что все договора пронумерованы от 1 до n):

$$S^{ind} = X_1 + X_2 + \dots + X_n,$$

Пусть p_i , как и раньше, есть вероятность того, что i -ый договор не привел ни к одному иску. За q_i обозначим вероятность возникновения *ровно* одного иска по i -му договору. Будем считать, что возможность подачи двух или более исков по одному договору страхования исключена, то есть $p_i + q_i = 1$. Данное предположение может иметь смысл, например, в контексте страхования жизни, где q_i будет означать вероятность смерти страхователя в течение года. В рамках данной модели удобно ввести следующие случайные величины:

- D_1, \dots, D_n , где $D_i = \begin{cases} 1, & i\text{-ый договор привел к иску} \\ 0, & \text{иначе} \end{cases}$
- C_1, \dots, C_n , где C_i есть величина иска по i -му договору (при условии, что $D_i = 1$)

Тогда величина суммарного иска по всему портфелю из n договоров может быть записана следующим образом:

$$S^{ind} = D_1 C_1 + \dots + D_n C_n \quad (22)$$

Обозначим за $P_i(\cdot)$ функцию распределения C_i , то есть P_i это условное распределение величины иска от i -го договора страхования, при условии, что по нему произошел страховой случай. Тогда для функции распределения $D_i C_i$ имеем:

$$F_i(x) = P(D_i C_i \leq x) = P(D_i C_i \leq x | D_i = 0)P(D_i = 0) + P(D_i C_i \leq x | D_i = 1)P(D_i = 1) = P(0 \leq x)p_i + P_i(x)q_i = p_i I(x) + P_i(x)q_i,$$

где $I(x) = 0$, если $x < 0$ и $I(x) = 1$ в противном случае.

Теперь, чтобы подсчитать функцию распределения величины S^{ind} , предположим, что договора страхования независимы между собой, то есть что случайные величины $D_1, \dots, D_n, C_1, \dots, C_n$ независимы. Тогда, используя уже знакомую нам свертку, можем подсчитать функцию распределения величины суммарного иска:

$$F = F_1 * \dots * F_n.$$

Отметим, что мы можем найти моменты величины S и без её функции распределения. Так, например:

$$E(S^{ind}) = \sum_{i=1}^n E(D_i C_i) = \sum_{i=1}^n E(D_i) E(C_i) = \sum_{i=1}^n (1 \cdot q_i + 0 \cdot p_i) E(C_i) = \sum_{i=1}^n q_i E(C_i) \quad (23)$$

Для дисперсии величины суммарного иска справедливо следующее соотношение, обоснование которого содержится на стр. 49 раздела 1 главы 4 статьи [2]:

$$Var(S^{ind}) = \sum_{i=1}^n q_i Var(C_i) + \sum_{i=1}^n p_i q_i (E(C_i))^2 \quad (24)$$

Заметим, что если вернуться к контексту страхования жизни и предположить, что в случае отхода души страхователя в мир иной страховщик выплачивает выгодоприобретателю некоторую сумму z_i * формулы (23) и (24) примут более простой вид:

$$E(S^{ind}) = \sum_{i=1}^n q_i z_i, \quad Var(S^{ind}) = \sum_{i=1}^n p_i q_i z_i^2. \quad (25)$$

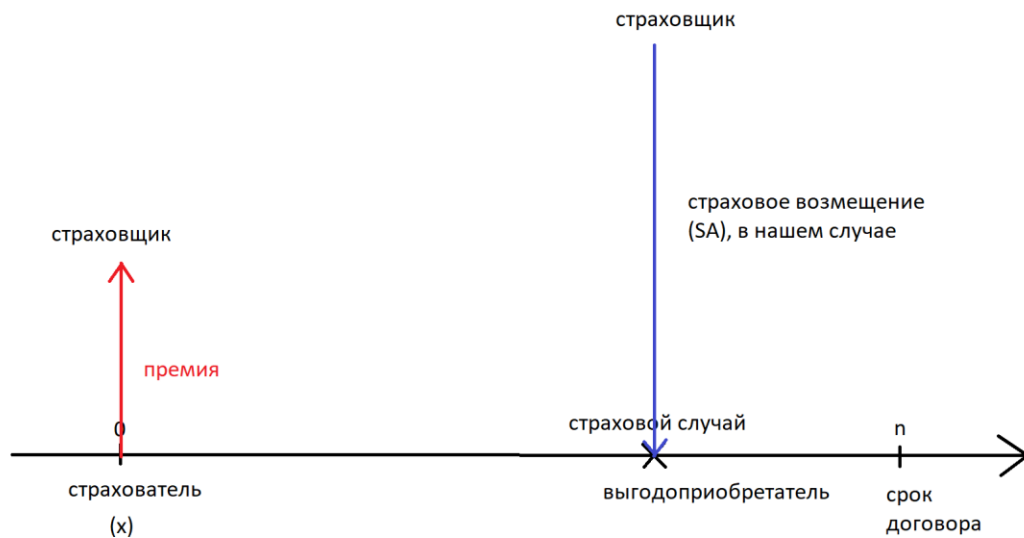


Рис. 5.1.1. Простейшее временное страхование жизни

* Данная ситуация представлена на рисунке 5.1.1.

Действительно, ведь в описанном случае C_i является уже не случайной, а детерминированной величиной и поэтому $E(C_i) = z_i$ и $Var(C_i) = 0$. Тем не менее даже в чуть более сложной ситуации, когда, например, страховщик выплачивает двойное страховое возмещение, если страхователь умирает из-за несчастного случая, величина C_i уже будет являться случайной величиной в обычном смысле этого слова.

Пример 5.1.1. Представим себе, что небольшая канадская компания заключила групповой договор страхования на 14 своих сотрудников. Актуарию, нанятому страховой, необходимо посчитать первые три момента и коэффициент асимметрии величины суммарного иска. Для этого он выбрал таблицу смертности канадского института актуариев за 1969 – 1975 года. Каждый работник застрахован на сумму, равную его зарплате, округленную до целого числа тысяч долларов (в большую сторону), вся информация представлена в таблице 5.1.1.

Работник	Возраст	Пол	Страховая выплата	Вероятность смерти
1	20	М	\$ 15 000,00	0.00149
2	23	М	\$ 16 000,00	0.00142
3	27	М	\$ 20 000,00	0.00128
4	30	М	\$ 28 000,00	0.00122
5	31	М	\$ 31 000,00	0.00123
6	46	М	\$ 18 000,00	0.00353
7	47	М	\$ 26 000,00	0.00394
8	49	М	\$ 24 000,00	0.00484
9	64	М	\$ 60 000,00	0.02182
10	18	F	\$ 14 000,00	0.00050
11	22	F	\$ 17 000,00	0.00050
12	26	F	\$ 19 000,00	0.00054
13	37	F	\$ 30 000,00	0.00103
14	55	F	\$ 55 000,00	0.00479
	Всего		\$ 373 000,00	

Таблица 5.1.1.

► В нашем случае z_i есть страховая выплата по i -му работнику, а q_i - вероятность смерти i -го работника (соответственно $p_i = 1 - q_i$). Поэтому

$$E(S^{ind}) = \sum_{i=1}^{14} q_i z_i = 2054.41,$$

Для третьего центрального момента имеем:

$$\begin{aligned} E\left(\left(S^{ind} - E(S^{ind})\right)^3\right) &= E\left(\left(S^{ind}\right)^3\right) - 3 \cdot E(S^{ind}) \cdot E\left(\left(S^{ind}\right)^2\right) + 2\left(E(S^{ind})\right)^3 = E\left(\left(S^{ind}\right)^3\right) - \\ &- 3 \cdot E(S^{ind}) \left(Var(S^{ind}) + \left(E(S^{ind})\right)^2\right) + 2\left(E(S^{ind})\right)^3 = E\left(\left(S^{ind}\right)^3\right) - 3E(S^{ind})Var(S^{ind}) - \left(E(S^{ind})\right)^3 = \end{aligned}$$

$$= \sum_{i=1}^{14} q_i z_i^3 - 3E(S^{ind})Var(S^{ind}) - (E(S^{ind}))^3.$$

Итак, $E\left((S^{ind} - E(S^{ind}))^3\right) = 5.4687849 \cdot 10^{12}$. Следовательно, для коэффициента асимметрии имеем:

$$\gamma = \frac{E\left((S^{ind} - E(S^{ind}))^3\right)}{\sigma^3} = 5.26734515.$$

Это означает, что плотность величины суммарного S^{ind} иска смещена вправо, что делает большие величины исков более вероятными. ◁

5.2. Описание модели коллективного риска

Что касается модели коллективного риска, то здесь договора индивидуального страхования не рассматриваются по отдельности. Поступающие иски не связываются с конкретными договорами, они анализируются как результат суммарного иска компании. Иными словами, величины X_i в этой модели, это не иск от i -го договора, а i -ый по порядку реально поступивший иск (величины X_i предполагаются независимыми и одинаково распределенными). Очень важно четко различать и понимать, что стояло за X_i раньше (в модели индивидуального иска) и что за ними кроется теперь. Ещё раз, до этого X_i обозначала величину индивидуального иска по i -му договору, то есть эта величина характеризовала иски *привязанные* к конкретному i -му договору страхования. И вообще говоря, X_i могли меняться при переходе от одного договора к другому. В рамках новой модели иски *не ассоциируются* с конкретными договорами страхования, а рассматриваются во всем портфеле целиком.

В связи с этим величина суммарного иска S^{coll} является суммой *случайного* числа слагаемых (ведь число исков за рассматриваемый промежуток времени есть случайная величина), то есть $S^{coll} = X_1 + \dots + X_\nu$, где ν - число исков за конкретный промежуток (при этом мы требуем, чтобы величина иска X и их число ν были независимы). Одним из самых распространенных предположений в рамках модели коллективного риска является предположение о том, что величина S^{coll} имеет составное пуассоновское распределение с параметрами λ (среднее число исков за анализируемый промежуток времени) и функцией распределения $F(x)$ индивидуального иска X .

Утв. 5.2.1. Для среднего и дисперсии величины S^{coll} справедливы следующие формулы:

$$\bullet \quad E(S^{coll}) = E(X)E(\nu) \quad (26)$$

$$\bullet \quad Var(S^{coll}) = Var(\nu)(E(X))^2 + Var(X)E(\nu) \quad (27)$$

▷ Обоснование этого утверждения содержится на стр. 50 раздела 1 главы 4 научного труда [2]. ◀

В том же месте указанной монографии приводится расчёт этих величин в случае, когда S^{coll} имеет составное пуассоновское распределение:

$$E(S^{coll}) = \lambda \cdot E(X), \quad Var(S^{coll}) = \lambda \cdot (E(X))^2 + \lambda \cdot Var(X) = \lambda \cdot E(X^2) \quad (28)$$

Из формул легко заметить, что параметр λ действительно характеризует среднее число исков.

5.3. Связь между моделями

Соотнесем полученные результаты с моделью индивидуального иска, описанной в начале этого раздела. Как уже было сказано величина D_i принимает значение 1, если по i -му договору был подан иск и значение 0 в противном случае. Поэтому среднее число исков за фиксированный промежуток времени может быть посчитано следующим образом: $E(D_1 + \dots + D_n) = \sum_{i=1}^n E(D_i) = \sum_{i=1}^n q_i$. В силу данной причины, естественно, положить

$$\lambda = \sum_{i=1}^n q_i. \quad (29)$$

Теперь разберемся с функцией распределения поступившего иска X , как было сказано ранее распределение исков может меняться при переходе от одного договора страхования к другому, в связи с этим важно учесть, к какому именно договору мы обращаемся. Поэтому

$$F(x) = P(X \leq x) = \sum_{i=1}^n P(X \leq x | \text{иск принадлежит } i\text{-му договору}) \cdot P(\text{иск принадлежит } i\text{-му договору}) = \sum_{i=1}^n P_i(x) \cdot \frac{q_i}{\lambda}, \quad (30)$$

где величина $\frac{q_i}{\lambda}$ характеризует долю исков по i -му договору среди общего числа исков.

Таким образом, параметры составного пуассоновского распределения выбираются так, чтобы среднее число исков было одинаковым в обоих моделях, а функция распределения поданного иска, была равна среднему взвешенному функций распределения величин индивидуальных исков по каждому договору страхования, с весами равными доле, которой занимает иск по конкретному договору, среди всех исков в портфеле. В силу данных соотношений можем переписать формулы (26) и (27) в следующем виде:

$$E(S^{coll}) = \sum_{i=1}^n q_i E(C_i), \quad Var(S^{coll}) = \sum_{i=1}^n q_i Var(C_i) + \sum_{i=1}^n q_i (E(C_i))^2. \quad (31)$$

В анализируемой нами статье выкладки, описывающее этот переход опущены, однако мы приведем подробное объяснение:

Док-во:▷ по формуле (28) имеем $E(S^{coll}) = \lambda \cdot E(X) = \lambda \cdot \int_0^{\infty} x dF(x)$, подставим вместо $F(x)$ выражение, которые мы нашли выше и получим:

$$E(S^{coll}) = \lambda \cdot \int_0^{\infty} x d \left(\sum_{i=1}^n P_i(x) \cdot \frac{q_i}{\lambda} \right) = \lambda \sum_{i=1}^n \frac{q_i}{\lambda} \cdot \int_0^{\infty} x dP_i(x) = \sum_{i=1}^n q_i E(C_i).$$

Далее, $Var(S^{coll}) = \lambda \cdot E(X^2)$, проделывая аналогичные преобразования (и заменяя x на x^2) получаем, что $Var(S^{coll}) = \sum_{i=1}^n q_i E(C_i^2) = \sum_{i=1}^n q_i Var(C_i) + \sum_{i=1}^n q_i (E(C_i))^2$.

Сравним полученные данные с соответствующими результатами для модели индивидуального риска:

- $E(S^{ind}) = \sum_{i=1}^n q_i E(C_i) = E(S^{coll})$
- $Var(S^{ind}) = \sum_{i=1}^n q_i Var(C_i) + \sum_{i=1}^n p_i q_i (E(C_i))^2 \leq \sum_{i=1}^n q_i Var(C_i) + \sum_{i=1}^n q_i (E(C_i))^2 = Var(S^{coll})$

Видим, что средняя величина суммарного иска не изменилась, в то время как дисперсия этой величины при переходе к модели коллективного иска увеличилась. Поэтому этот переход является в некотором смысле осторожным шагом.

Составное пуассоновское распределение, которое довольно часто описывает распределение величины S^{coll} , может быть получено как предельное распределение величины суммарного иска в модели индивидуального риска. Докажем этот факт, используя производящие функции моментов. Так, пусть $M_i(t)$ обозначает п. ф. м. величины C_i , а $M(t)$ - п. ф. м. поступившего иска X . Сперва найдем п. ф. м. $D_i C_i$:

$$\begin{aligned} M_{D_i C_i}(t) &= E(e^{t D_i C_i}) = E\left((e^{t C_i})^{D_i}\right) = E(e^{t C_i}) \cdot P(D_i = 1) + E\left((e^{t C_i})^0\right) \cdot P(D_i = 0) = \\ &= M_i(t) q_i + p_i = 1 + q_i (M_i(t) - 1). \end{aligned}$$

Поэтому для п. ф. м. величины суммарного иска имеем: $M_{S^{ind}}(t) = E\left(e^{t \sum_{i=1}^n D_i C_i}\right) = \prod_{i=1}^n E(e^{t D_i C_i}) = \prod_{i=1}^n (1 + q_i (M_i(t) - 1))$. Далее, заменим i -ый договор страхования на k независимых одинаковых договоров, таких что вероятность подачи иска по каждому из них равна $\frac{q_i}{k}$, а функция распределения величины этого иска (при условии

что он был подан) также равна $P_i(x)$. Тогда формула для п. ф. м. величины S^{ind} примет вид: $M_{S^{ind}}(t) = \prod_{i=1}^n \left(1 + \frac{q_i}{k} (M_i(t) - 1)\right)^k$. Посмотрим на это выражение при $k \rightarrow \infty$:

$$\left(1 + \frac{q_i}{k} (M_i(t) - 1)\right)^k \rightarrow e^{q_i(M_i(t)-1)},$$

то есть $\lim_{k \rightarrow \infty} \prod_{i=1}^n \left(1 + \frac{q_i}{k} (M_i(t) - 1)\right)^k = \prod_{i=1}^n e^{q_i(M_i(t)-1)} = e^{\sum_{i=1}^n q_i M_i(t) - \lambda} = e^{\lambda \left(\sum_{i=1}^n \frac{q_i}{\lambda} M_i(t) - 1\right)}$. Теперь покажем,

что $M(t)$ равна в точности сумме, находящейся в степени экспоненты. Действительно,

$$M(t) = E(e^{tX}) = \int_0^\infty e^{tx} dF(x) = \sum_{i=1}^n \int_0^\infty e^{tx} \frac{q_i}{\lambda} dP_i(x) = \sum_{i=1}^n \frac{q_i}{\lambda} M_i(t). \text{ Итак, получаем, что при}$$

$k \rightarrow \infty$ п. ф. м. $M_{S^{ind}}(t) \rightarrow e^{\lambda(M(t)-1)}$. Остается заметить, что предельная величина есть п. ф.

м. величины суммарного иска S^{coll} (обоснование этого факта почти в точности повторяет доказательство, приведенное на стр. 68 главы 5 книги [7]). Таким образом, распределение величины суммарного иска в модели индивидуального риска при увеличении числа договоров страхования стремится к составному пуассоновскому распределению. Подчеркнем, что в изложенной нами конструкции мы, по сути, обобщили хорошо известное свойство пуассоновского распределения, а именно то, что оно является пределом для биномиального распределения.

5.4. Примеры

Представим себе портфель, содержащий 31 договор страхования жизни, со свойствами, указанными в таблице 5.4.1*:

Портфель из 31 договора страхования					
Вероятность иска	Страховое возмещение				
	1	2	3	4	5
0.03	2	3	1	2	-
0.04	-	1	2	2	1
0.05	-	2	4	2	2
0.06	-	2	2	2	1
Среднее число исков для данного возмещения	0.06	0.35	0.43	0.36	0.20

Таблица 5.4.1.

* Страховое возмещение в данном случае указано в условных единицах.

Так, например, мы имеем 2 договора страхования с страховым возмещением 1 и вероятностью возникновения страхового случая 0.03; 3 договора с величиной возмещения 2 и с вероятностью возникновения иска 0.03 и так далее. Заметим, что последняя строчка в этой таблице вычислялась следующим образом: вероятность возникновения иска умножалась на число соответствующих договоров. Так, для договора из первого столбика имеем: $0.03 \cdot 2 = 0.06$. Также из таблицы мы можем найти какую долю составляют иски, которые привели к определенной страховой выплате, от общего числа. Так, доля исков по которым была выплачена сумма 2 равна $0.35 / 1.4 = 0.25$. Используя формулы (23) и (24) можем найти среднее и дисперсию величины суммарного иска:

$$E(S^{ind}) = 0.06 + 2 \cdot 0.35 + 3 \cdot 0.43 + 4 \cdot 0.36 + 5 \cdot 0.2 = 4.49,$$

$$\begin{aligned} Var(S^{ind}) = & 2 \cdot (0.03 \cdot 0.97 + 0.05 \cdot 0.95 \cdot 4 + 0.06 \cdot 0.94 \cdot 4 + 0.04 \cdot 0.96 \cdot 9 + 0.06 \cdot 0.94 \cdot 9 + \\ & + 0.03 \cdot 0.97 \cdot 16 + 0.04 \cdot 0.96 \cdot 16 + 0.05 \cdot 0.95 \cdot 16 + 0.06 \cdot 0.94 \cdot 16 + 0.05 \cdot 0.95 \cdot 25) + 3 \cdot (0.03 \cdot 0.97 \cdot 4) + \\ & + 4 \cdot (0.05 \cdot 0.95 \cdot 9) + 1 \cdot (0.04 \cdot 0.96 \cdot 4 + 0.03 \cdot 0.97 \cdot 9 + 0.04 \cdot 0.96 \cdot 25 + 0.06 \cdot 0.94 \cdot 25) = 15.3003^*. \end{aligned}$$

Теперь найдем функцию распределения величины суммарного иска. Для этого необходимо будет произвести операцию свёртки 31 раз (ведь у нас ровно столько договоров индивидуального страхования). Сперва для наглядности посмотрим на графики функций распределения величин индивидуального иска, они изображены на рисунке 5.4.1. Отметим, что мы обозначили договор, расположенный в таблице 5.4.1 в первом столбике и на первой строчке за «Договор 1.1»; полис, находящийся в 4 столбике и на 4 строчке за «Договор 4.4» и так далее.

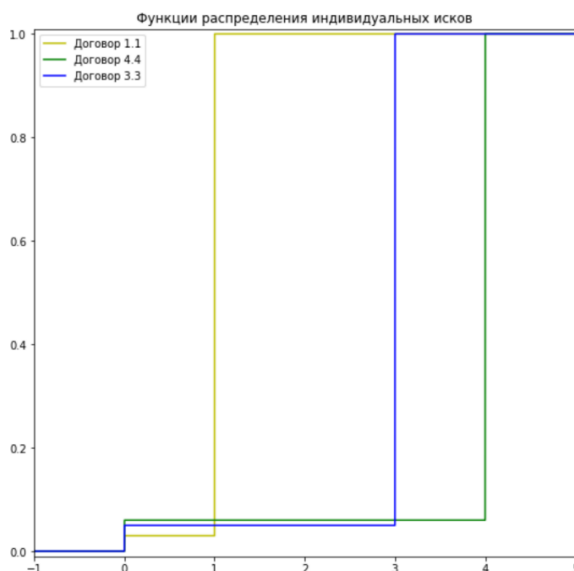


Рис. 5.4.1. Функции распределения индивидуальных исков.

```
plt.figure(figsize = (9,9))
x = np.linspace (-1,5,10000)
y1 = 0.03*np.heaviside(x, 0)+0.97*np.heaviside(x-1, 0)
plt.plot(x, y1, "y-", label = 'Договор 1.1')
y2 = 0.06*np.heaviside(x, 0)+0.94*np.heaviside(x-4, 0)
plt.plot(x, y2, "g-", label = 'Договор 4.4')
y2 = 0.05*np.heaviside(x, 0)+0.95*np.heaviside(x-3, 0)
plt.plot(x, y2, "b-", label = 'Договор 3.3')
plt.legend(loc = 'upper left')
plt.title('Функции распределения индивидуальных исков')
plt.ylim([-0.01,1.01])
plt.xlim([-1,5])
plt.show()
```

Рис. 5.4.2. Код для графиков функций распределения.

* Слагаемые в сумме сгруппированы по числу эквивалентных договоров в портфеле.

Результаты для функции распределения суммарного иска показаны в таблице 5.4.2. Код, с помощью которого они были достигнуты представлен на рисунке 5.4.3.

Далее, перейдем к модели коллективного риска, обратившись к формулам (29), (30). Тогда $\lambda = 1.4$, ведь среднее число исков есть в точности сумма нижней строчки в таблице 5.4.1: $0.06 + 0.35 + 0.43 + 0.36 + 0.20 = 1.4$. Как было замечено ранее, в рамках данного метода $E(S^{ind}) = \sum_{i=1}^n q_i E(C_i) = E(S^{coll})$, поэтому ожидаемая величина суммарного иска останется равной 4.49.

Распределение величины суммарного иска в рамках модели индивидуального риска		
х	Вероятность, P(x)	Значение функции распределения, F(x)
0	0.238195	0.238195
1	0.014734	0.252929
2	0.087734	0.340663
3	0.113183	0.453846
4	0.110709	0.564555
5	0.096327	0.660882
6	0.061548	0.722431
7	0.069022	0.791453
8	0.054817	0.846270
9	0.043147	0.889417
10	0.030107	0.919525

Таблица 5.4.2.

```
import numpy as np
def h(a):
    return np.convolve(a,np.convolve(a,a))
#Инициализируем наши данные. Для этого заводим массив для каждого договора индивидуального страхования.
#В данном случае позиция в массиве означает величину страховой выплаты.
a1=[0.97,0.03]
b1, b2, b3, b4=[0.97,0,0.03], [0.96,0,0.04], [0.95,0,0.05], [0.94,0,0.06]
c1, c2, c3, c4=[0.97,0,0,0.03], [0.96,0,0,0.04], [0.95,0,0,0.05], [0.94,0,0,0.06]
d1, d2, d3, d4=[0.97,0,0,0,0.03], [0.96,0,0,0,0.04], [0.95,0,0,0,0.05], [0.94,0,0,0,0.06]
z2, z3, z4=[0.96,0,0,0,0.04], [0.95,0,0,0,0.05], [0.94,0,0,0,0.06]
#Выполняем свёртку функций распределения величин индивидуальных исков, не забывая про число договоров.
x1=np.convolve(a1,a1)
x2=[h(b1), b2, np.convolve(b3,b3), np.convolve(b4,b4)]
x3=[c1, np.convolve(c2,c2), np.convolve(c3,c3), np.convolve(c4,c4)]
x4 = [np.convolve(d1,d1), np.convolve(d2,d2),np.convolve(d3,d3),np.convolve(d4,d4)]
x5 = [z2,np.convolve(z3,z3), z4]
y1 = [np.convolve(x2[0], x2[1]), np.convolve(x2[2], x2[3])]
y2 = [np.convolve(x3[0], x3[1]), np.convolve(x3[2], x3[3])]
y3 = [np.convolve(x4[0], x4[1]), np.convolve(x4[2], x4[3])]
y4 = np.convolve(x5[0], np.convolve(x5[1],x5[2]))
z = np.convolve(np.convolve(y1[0],y1[1]), np.convolve(y2[0],y2[1]))
v = np.convolve(np.convolve(y3[0],y3[1]), y4)
w = np.convolve(z, np.convolve(x1,v))
#Выводим массив вероятностей для значений величины суммарного иска от 0 до 10.
print(w[0:11])
for i in range(1,98):
    w[i]+=w[i-1]
#Выводим массив значений функции распределения суммарного иска в точках 0,1,2,...,10.
print(w[0:11])
```

Рис. 5.4.3. Код для подсчёта функции распределения суммарного иска.

Теперь посчитаем дисперсию суммарного коллективного иска: по формуле (31)

$Var(S^{coll}) = \sum_{i=1}^n q_i Var(C_i) + \sum_{i=1}^n q_i (E(C_i))^2$, однако в нашем случае величины C_i детерминированы, поэтому первая сумма в этой формуле окажется равной нулю. Итак,

$$Var(S^{coll}) = 2 \cdot (0.03 + 0.05 \cdot 4 + 0.06 \cdot 4 + 0.04 \cdot 9 + 0.06 \cdot 9 + 0.03 \cdot 16 + 0.04 \cdot 16 + 0.05 \cdot 16 + 0.06 \cdot 16 + 0.05 \cdot 25) + 3 \cdot 0.03 \cdot 4 + 4 \cdot 0.05 \cdot 9 + 1 \cdot (0.04 \cdot 4 + 0.03 \cdot 9 + 0.04 \cdot 25 + 0.06 \cdot 25) = 16.09.$$

Выведенная нами теория не обманула, дисперсия суммарного иска при переходе к коллективной модели действительно увеличилась.

Покажем, что в рамках полученной модели коллективного риска предположение о том, что суммарный иск имеет составное распределение Пуассона в самом деле имеет место и является неплохой аппроксимацией реальных данных. Так, разделим наши договора на 5 групп – по величинам страховых выплат и предположим, что в каждой группе число исков моделируется распределением Пуассона с соответствующим параметром из нижней строчки таблицы 5.4.1. Тогда вероятность того, что договор из первой группы не привёл иску есть $e^{-0.06} = 0.941765$, привёл к одному - $\frac{0.06}{1!} e^{-0.06} = 0.056506$, к двум - $\frac{0.06^2}{2!} e^{-0.06} = 0.001695$, к трём и более – пренебрежимо мала (что хорошо соотносится с реальной ситуацией, ведь договора из первой группы просто не могут произвести более двух исков). Распределение вероятностей по числу исков для всех групп отражено в таблице 5.4.3.

Распределение количества исков по договорам страхования					
х	Размер страховых выплат				
	1	2	3	4	5
0	0.941765	0.704688	0.650509	0.0697676	0.818731
1	0.056506	0	0	0	0
2	0.001695	0.246641	0	0	0
3	0.000034	0	0.279719	0	0
4	0.000001	0.043162	0	0.251163	0
5	-	0	0	0	0.163746
6	-	0.005036	0.060140	0	0
7	-	0	0	0	0
8	-	0.000441	0	0.045209	0
9	-	0	0.008620	0	0
10	-	0.000031	0	0	0.016375

Таблица 5.4.3.

Ясно, что договора из второй категории не могут привести к выплатам нечетных сумм, ибо величины исков равны строго двум. Именно по этой причине на этих местах в таблице находятся нули. Ровно из тех же соображений нули стоят и на других местах. Сейчас настало время уложить в голове всё, что мы имеем на данный момент и сделать

дальнейшие выводы. Сначала мы работали с портфелем договоров страхования в рамках модели индивидуального риска. Затем решили перейти к модели коллективного риска, чтобы аппроксимировать величину суммарного иска. Отталкиваясь от новой модели, мы имеем пять договоров страхования, число исков, производимых каждым из них, имеет распределение Пуассона с соответствующими параметрами, а величины исков по каждому полису детерминированы. Чтобы найти интересующее нас распределение, остаётся произвести свёртку функций распределений по каждому договору.

Для большей наглядности будем производить эту операцию последовательно, то есть сначала осуществим её для договоров из первой и второй группы, затем из первой, второй и третьей и так далее. Чтобы добиться максимального понимания в производимых нами операциях, покажем, как выглядят графики функций распределения договоров из 1-й группы и из 2-й, и что с ними происходит после свёртки (см. рисунок 5.4.4). Все промежуточные результаты вместе с интересующим нас распределением суммарного иска указаны в таблице 5.4.4.

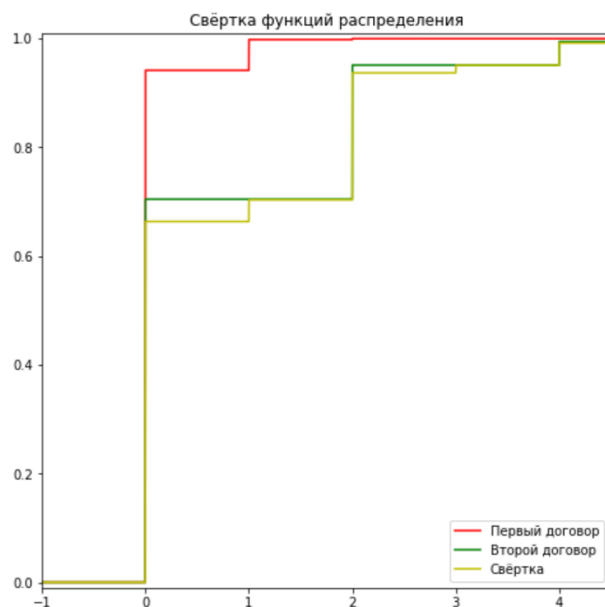


Рис. 5.4.5. Свёртка функций распределения договоров страхования.

```
import math
import numpy as np
from matplotlib import pyplot as plt
plt.figure(figsize=(8,8))
x=np.linspace(-1,5,10000)
y1=(0.941765*np.heaviside(x,0)+0.056506*np.heaviside(x-1,0)
+0.001695*np.heaviside(x-2,0)+0.000034*np.heaviside(x-3,0)
+0.000001*np.heaviside(x-4,0))
y2=(0.704688*np.heaviside(x,0)+0.246641*np.heaviside(x-2,0)
+0.043162*np.heaviside(x-4,0))
y3=(0.663650*np.heaviside(x,0)+0.039819*np.heaviside(x-1,0)
+0.233572*np.heaviside(x-2,0)+0.013961*np.heaviside(x-3,0)
+0.041066*np.heaviside(x-4,0))
plt.plot(x, y1, 'r-', label='Первый договор')
plt.plot(x, y2, 'g-', label='Второй договор')
plt.plot(x, y3, 'y-', label='Свёртка')
plt.legend(loc = "lower right")
plt.title('Свёртка функций распределения')
plt.ylim([-0.01,1.01])
plt.xlim([-1,4.5])
plt.show()
```

Рис. 5.4.4. Код для графика свёртки функций распределения.

Сравнивая итоговые данные для функции распределения суммарного иска в таблицах 5.4.2 и 5.4.4 можем сделать вывод о том, что наша аппроксимация дает достаточно хорошие результаты. Почти все значения функции совпали с точностью до сотых долей.

Итак, примеры, разобранные в этом разделе, позволяют лишний раз убедиться в важности свёртки и составного Пуассоновского распределения в моделях коллективного и индивидуального исков. Как было показано – свёртка двух распределений вычисляется довольно простым образом (все в конечном счёте сводится к умножению и суммированию положительных чисел, что несложно осуществить на вычислительной машине), а составное Пуассоновское распределение небеспочвенно используется в качестве приближения для функции распределения величины суммарного иска, ведь оно действительно даёт хорошую аппроксимацию.

Подсчёт распределения суммарного иска					
x	Промежуточные вычисления			Распределение сумм. иска	Функция распределения сумм. иска
	1+2	1+2+3	1+2+3+4	1+2+3+4+5	
0	0.663651	0.431711	0.301194	0.246597	0.246597
1	0.039819	0.025903	0.018072	0.014796	0.261393
2	0.233472	0.151876	0.105960	0.086753	0.348146
3	0.013961	0.194717	0.135850	0.111224	0.459370
4	0.041067	0.037853	0.134839	0.110397	0.569766
5	0.002447	0.066899	0.053179	0.092859	0.662625
6	0.004816	0.046950	0.070901	0.061008	0.723633
7	0.000286	0.014068	0.058721	0.065427	0.789060
8	0.000424	0.015001	0.039490	0.054577	0.843637
9	0.000025	0.007924	0.023502	0.041321	0.884958
10	0.000030	0.002912	0.020690	0.030579	0.915537

Таблица 5.4.4.

```

import math
import numpy as np
from matplotlib import pyplot as plt
a1=[0.941765, 0.056506, 0.001695, 0.000034, 0.000001]
a2=np.zeros(201)
a3=np.zeros(201)
a4=np.zeros(201)
a5=np.zeros(201)
#Заполняем созданные массивы вероятностями возникновения определенного числа исков.
for i in range (0,201):
    if i%2==0:
        a2[i]=(((0.35)**(i/2)/math.factorial(i/2)))*math.exp(-0.35)
    if i%3==0:
        a3[i]=(((0.43)**(i/3)/math.factorial(i/3)))*math.exp(-0.43)
    if i%4==0:
        a4[i]=(((0.36)**(i/4)/math.factorial(i/4)))*math.exp(-0.36)
    if i%5==0:
        a5[i]=(((0.20)**(i/5)/math.factorial(i/5)))*math.exp(-0.20)
#Производим свёртки.
a1pa2=np.convolve(a1,a2)
a1pa2pa3=np.convolve(a1pa2,a3)
a1pa2pa3pa4=np.convolve(a1pa2pa3,a4)
a1pa2pa3pa4pa5=np.convolve(a1pa2pa3pa4,a5)
#Вычисляем значения функции распределения.
for i in range (1, len(a4)):
    a1pa2pa3pa4pa5[i]+=a1pa2pa3pa4pa5[i-1]
print(a1pa2pa3pa4pa5)

```

Рис. 5.4.6. Код для подсчёта функции распределения величины суммарного иска в случае, когда она имеет составное пуассоновское распределение.

6. Ошибки и недостатки

К сожалению, при подсчёте распределения величины суммарного иска нельзя обойтись без различных ошибок и неточностей. Во-первых, это связано с округлением возможных сумм исков по договорам страхования до определенной денежной единицы. Во-вторых, не стоит забывать о естественном свойстве аппроксимации – она дает лишь приблизительные значения. Конечно, всегда можно увеличить вычислительное время и стремиться к минимальным округлениям, но идеального результата нам не добиться никогда. И в-третьих, в самом начале мы сделали крайне важное предположение о независимости величин индивидуальных исков. Почти вся наша дальнейшая стройная теория базировалась на этом факте. К сожалению, на практике это требование нередко бывает нарушено, что не оставляет нам шансов на применение достигнутых результатов. Кроме того, необходимо четко понимать и осознавать, что мы рассматривали именно математическую *модель*, которая порой может плохо соотноситься с реальными данными. Именно поэтому следует пользоваться достигнутыми результатами крайне осторожно и полагаться в первую очередь на соображения здравого смысла. Также стоит сделать очевидное замечание – при исследовании величины суммарного иска лучше подключать как можно больше различных методов и инструментов, только так можно добиться наиболее полной и объективной картины.

7. Заключение

Итак, модель индивидуального риска и различные методы и инструменты её исследования были детально описаны. Мы рассмотрели некоторые виды аппроксимации и сравнили их между собой. Так, например, было установлено, что гамма-приближение и нормальное степенное приближение дают более точные результаты нежели просто центральная предельная теорема; аппроксимация Эшера исправляет некоторые недостатки разложения Эджуорта, которое в свою очередь совершенствует приближение Грама-Шарлье. Что касается связи между моделями индивидуального и коллективного рисков, то мы установили, что переход к модели коллективного риска является в некотором смысле осторожным шагом, поскольку дисперсия величины суммарного иска увеличивается, в то время как среднее не меняется. Также используя свойства обеих моделей, мы выяснили, что при определенных условиях величина суммарного иска в модели индивидуального риска имеет составное пуассоновское распределение. Но самое главное – нам удалось дополнить изложение анализируемых научных статей, доказав многие утверждения самостоятельно (утв. 4.1.1 и утв. 4.1.2, посвященные гамма-приближению; утверждения об ортогональности и явном виде коэффициентов многочленов Лагерра и Эрмита – утв. 4.4.1, утв. 4.5.1, утв. 4.5.2; пример 4.7.1, показывающий как действует преобразование Эшера на величины с известными распределениями), восполнив некоторые пробелы в рассуждениях (замечание на стр. 8 о нормальном приближении; вывод формулы (31) на стр. 32; пробел в обосновании аппроксимации Эшера на стр. 24), и полностью сопроводив раздел 5.4, описывающий важные примеры в рамках моделей индивидуального и коллективного рисков, необходимыми графиками (рис. 5.4.1, рис. 5.4.5) и расчётами (рис. 5.4.2 – 5.4.4, рис. 5.4.6), реализованными посредством языка программирования Python.

Хочется заметить, что нами была изучена достаточно большая часть теории, но ещё большая часть вышла за рамки нашего анализа. К сожалению, многие важные вопросы, так и не были освещены в данной работе. Во-первых, было замечено, что качество аппроксимации Эджуорта необязательно улучшится, если увеличить число слагаемых в рассматриваемом отрезке ряда. В связи с этим возникает естественный вопрос, а какое число слагаемых использовать предпочтительнее? Во-вторых, как мы выяснили, приближение Эшера достаточно хорошо сглаживает хвосты исследуемого распределения и дает более точные результаты, но насколько оно повышает эту самую точность? В-третьих, практически центральным вопросом была вероятность разорения, очень много примеров было посвящено именно ей, однако вопрос выбора этой вероятности не обсуждался. Какое значение взять «лучше» или «безопаснее», 5% или может быть 2%? Каждый из указанных вопросов действительно очень важен и заслуживает отдельного исследования.

8. Список литературы

- [1] Dhaene, J.; Vyncke, D. Encyclopedia of actuarial science, The individual risk model.
- [2] Gerber, H. U. (1979). An Introduction to Mathematical Risk Theory, Wharton School, University of Pennsylvania.
- [3] Panjer, H. H., Willmot, G. E. (1992), Insurance Risk Models.
- [4] Lenz H.-J., Wilrich P.-Th., Frontiers in Statistical Quality Control 5.
- [5] Beard, R.E., Pentikäinen, T., Pesonen, E. (1984) Risk Theory. The Stochastic Basis of Insurance, 3-rd ed., Chapman and Hall, pp. 108-111, London etc.
- [6] Keith Y. Patarroyo (2020), A digression on Hermite polynomials.
- [7] Фалин Г. И., (1994) Математический анализ рисков в страховании.
- [8] Kuon, S.; Radtke, M.; Reich, A. (1993). An appropriate way to switch from the individual risk model to the collective one.