Глава 1

Распределения ущерба

Цели главы

К концу данной главы вы будете уметь:

- описывать и использовать простые функции ущерба
- описывать и применять свойства стандартных функций потерь

§1 Введение

Компаниям, занимающимся общими видами страхования, необходимо исследовать опыт страховых возмещений и применять математические средства для различных целей, включающих в себя:

- определение ставки страхового взноса (т.е. решение какой именно размер премии необходимо взимать с владельцев страхового полиса)
- резервирование (т.е определение величины средств, которые должны быть сохранены для покрытия страховых исков)
- пересмотр договоров перестрахования
- проверку платежеспособности (т.е. определение финансового положение компании)

В данной главе мы рассмотрим распределения ущерба, которые являются математическим методом моделирования индивидуальных исков.

Будут введены несколько новых статистических распределений, и мы увидим каким образом они могут быть применены для понимания имеющихся данных об исках. Затем соответствующие распределения будут

будут использованы для оценки вероятностей и для вычисления премий за общие виды страхования.

Теорию, разработанную здесь, мы используем в следующей главе при рассмотрении эффектов от применения перестрахования и полисов превышения ущерба, а также в главе 3 (коллективные иски).

Практические аспекты установки размеров премии за общие виды страхования рассмотрены в разделе G

Краткий конспект главы начинается с введения терминологии, используемой при рассмотрении распределений для долгосрочного страхования. Затем сами распределения рассматриваются более подробно. Далее следует раздел, посвященный композициям распределений. Заметим, что обозначения, используемые в кратком конспекте слегка отличаются от используемых в остальной части курса. Экзаменационные вопросы могут быть представлены в различном виде. Тем не менее, экзаменаторы точно объясняют свои обозначения.

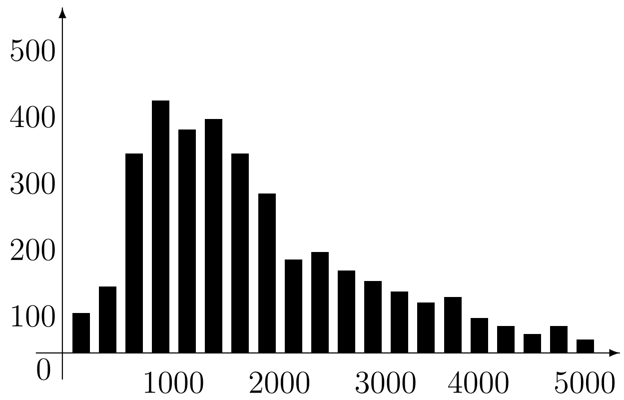
§2 Распределения ущерба

2.1 Типичные распределения ущерба

Индивидуальные потери

При общем страховании иски обычно удовлетворяются по принципу возмещения ущерба, т.е. размер выплаты равен сумме, необходимой для замены утраченного или восстановления поврежденного имущества. Таким образом величина страхового возмещения будет изменяться от иска к иску в зависимости от природы и серьезности причиненного ущерба.

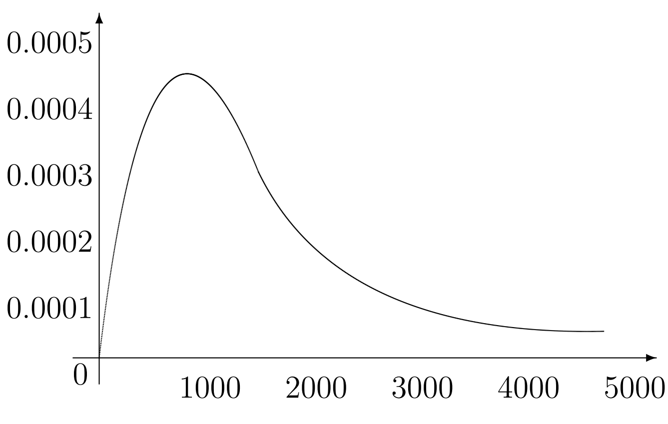
Иски по общим видам страхования изучаются путем построения гистограммы. Например, ниже представлена гистограмма, показывающая количество исков определённой величины в типичном портфеле, включающем в себя страхование автотранспорта.



Основная часть выплат приходится на относительно небольшие суммы, например помятые крылья или украденные автомагнитолы. В то же время иногда присутствуют и значительно большие выплаты, например компенсации людям, которые были травмированы.

Величины индивидуальных исков могут быть описаны математически путем рассмотрения *частот* выплат того или иного размера. Это определяет распределение ущерба, которое является функцией плотности распределения величины индивидуального иска. Распределение ущерба имеет обычные свойства функции плотности распределения, т.е. интеграл по области определения равен 1 и вероятности могут быть найдены путем интегрирования.

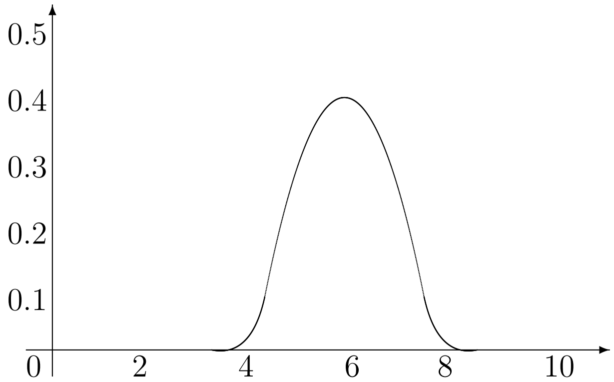
Например, распределение ущерба, определенное функцией, представленной на графике ниже, может являться хорошей моделью для распределения исков по автострахованию, которое показано на графике выше.



Обычно, распределения ущерба имеют положительную асимметрию. Наибольшее число исков приходится на центральную часть, но в то же время присутствует небольшое число гораздо более значительных исков вытягивающих 'хвост' распределения.

Агрегированные потери

Мы также заинтересуемся агрегированными распределениями ущерба (т.е. суммарной величиной исков). Например, на графике ниже показана модель распределения вероятностей для агрегированных исков в каждом месяце для страхового портфеля, рассмотренного выше.



2.2 Моменты и производящая функция моментов

Если мы рассматриваем конкретное распределение ущерба, то можно вычислить моменты величины индивидуального иска, а также значение дисперсии. Одним из способов сделать это является использование производящей функции моментов.

Пример 1.1 Величина индивидуального иска моделируется усеченным экспоненциальным распределением с плотностью равной:

$$f_X(x) = \lambda e^{-\lambda(x-M)},\ X > M$$

Найти производящую функцию моментов и вывести из неё формулы для математического ожидания и дисперсии величины индивидуального иска.

Решение Производящая функция моментов случайной величины X по определению есть:

$$M_X(t) = E(e^{tX}) = \int\limits_M^\infty e^{tX} \lambda e^{-\lambda(x-M)} dx = \lambda e^{\lambda M} \int\limits_M^\infty e^{-(\lambda-t)x} dx$$

Вычисляя интеграл (при $t < \lambda$), имеем:

$$M_X(t) = \lambda e^{\lambda M}\left[\frac{e^{-(\lambda-t)x}}{-(\lambda-t)}\right]_M^{\infty} = \lambda e^{\lambda M}\left[0 + \frac{e^{-(\lambda-t)M}}{\lambda - t}\right] = \frac{\lambda}{\lambda - t}e^{tM}$$

В данном случае наиболее простым способом вычисления моментов является представление производящей функции в виде ряда, а затем использование её свойств:

$$M_X(t) = \left(1 - \frac{t}{\lambda}\right)^{-1} e^{tM} = \left(1 + \frac{t}{\lambda} + \frac{t^2}{\lambda^2} + \cdots\right)\left(1 + tM + \frac{t^2M^2}{2} + \cdots\right)$$

Т.е.

$$1 + tE(X) + \frac{t^2}{2}E(X^2) + \ldots = 1 + t\left(\frac{1}{\lambda} + M\right) + t^2\left(\frac{1}{\lambda^2} + \frac{M}{\lambda} + \frac{M^2}{2}\right) + \cdots$$

Сравнивая коэффициенты при t и t^2, находим, что:

$$E(X) = \frac{1}{\lambda} + M$$

и

$$\frac{1}{2}E(X^2) = \frac{1}{\lambda^2} + \frac{M}{\lambda} + \frac{M^2}{2} \Rightarrow E(X^2) = 2\left(\frac{1}{\lambda^2} + \frac{M}{\lambda} + \frac{M^2}{2}\right)$$

Таким образом, дисперсия равна:

$$Var(X) = E(X^2) - [E(X)]^2 = \left(\frac{2}{\lambda^2} + \frac{2M}{\lambda} + M^2\right) - \left(\frac{1}{\lambda^2} + \frac{2M}{\lambda} + M^2\right) = \frac{1}{\lambda^2}$$

(Также эти формулы можно было получать иначе, дифференцируя производящую функцию моментов и полагая $t = 0$)

В случае распределений, не имеющих простого выражения для производящей функции, моменты нужно находить путем интегрирования или при помощи таблиц.

Пример 1.2 Размер индивидуального иска имеет логнормальное распределение с плотностью

$$f_X(x) = \frac{1}{x\sigma\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right],\ x > 0$$

Найти среднюю величину иска.

Решение Средняя величина иска является математическим ожиданием случайной величины X и определяется как:

$$E(X) = \int_0^{\infty} x\frac{1}{x\sigma\sqrt{2\pi}}exp\left[-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right]dx$$

Использование подстановки $z = \frac{\log x - \mu}{\sigma}$ (откуда $dz = \frac{dx}{x\sigma}$ и $x = e^{\mu+\sigma z}$) приводит к следующему выражению:

$$E(X) = \int\limits_{-\infty}^{\infty} e^{\mu+\sigma z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz$$

Мы можем вычислить этот интеграл, выделяя полный квадрат:

$$E(X) = \int\limits_{-\infty}^{\infty} e^{\mu} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z^2-2\sigma z+\sigma^2)+\frac{1}{2}\sigma^2} dz = e^{\mu+\frac{1}{2}\sigma^2} \int\limits_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-\sigma)^2} dz$$

Подынтегральная функция в последнем выражении является плотностью случайной величины имеющей нормальное распределение $N(\sigma, 1)$. Таким образом, последний интеграл равен 1, и мы получаем, что:

$$E(X) = e^{\mu+\frac{1}{2}\sigma^2}$$

Вопрос для самоподготовки 1.1. Вывести формулу для производящей функции моментов стандартного нормального распределения

Вопрос для самоподготовки 1.2. Случайная величина X имеет распределение *логгамма* с параметрами α и λ, если $\log X$ имеет распределение $\Gamma(\alpha, \lambda)$. Вывести формулу для средней величины иска, если случайная величина индивидуального иска имеет распределение $LogGamma(\alpha, \lambda)$

2.3 Стандартные распределения ущерба

Мы уже встречались с некоторыми стандартными статистическими распределениями, которые могут быть использованы для моделирования исков. Например, гамма и логнормальное распределения весьма часто оказываются подходящими. Иногда и более простые распределения, такие как экспоненциальное (которое является частным случаем гамма распределения) или нормальное могут оказаться вполне подходящими.

Статистиками были предложены несколько видов распределений, являющихся наиболее подходящими для моделирования отдельных типов ущерба при общем страховании. Три из этих типов распределений кратко описаны ниже. Они включены в справочные таблицы в Приложении. Простое выражение для производящей функции моментов этих распределений отсутствует.

1. **Распределение Парето**[1]

   Распределение Парето представлено в двух формах:

[1]Вилфредо Парето (1848-1923) был итальянским экономистом и социологом

- **Распределение Парето (двухпараметрическая форма)**
  Функция плотности распределения имеет следующий вид:

$$f_X(x) = \frac{\alpha\lambda^\alpha}{(\lambda + x)^{\alpha+1}}$$

- **Обобщенное распределение Парето (трехпараметрическая форма)**
  Функция плотности распределения имеет следующий вид:

$$f_X(x) = \frac{\Gamma(\alpha + k)\lambda^\alpha x^{k-1}}{\Gamma(\alpha)\Gamma(k)(\lambda + x)^{\alpha+k}},\ x > 0$$

Двухпараметрическое распределение является частным случаем обобщенного распределения Парето при $k = 1$.

Вопрос для самоподготовки 1.3. Проверить формулу для математического ожидания двухпараметрическое распределения Парето, представленную в "Таблицах". Для каких значений параметров верна данная формула?

2. **Распределение Бурра**[2]

Распределение Бурра является обобщением двухпараметрического распределения Парето, в котором переменная x заменена на x^γ. Функция плотности распределения имеет следующий вид:

$$f_X(x) = \frac{\alpha\gamma\lambda^\alpha x^{\gamma-1}}{(\lambda + x^\gamma)^{\alpha+1}},\ x > 0$$

Вопрос для самоподготовки 1.4. Проверить, что формула для математического ожидания распределения Бурра, приведенная в "Таблицах сводится к формуле для математического ожидания распределения Парето при $\gamma = 1$

3. **Распределение Вейбулла**

Распределение Вейбулла является обобщением экспоненциального распределения, в котором переменная x заменена x^γ. Функция плотности распределения имеет следующий вид:

$$f_X(x) = c\gamma x^{\gamma-1} e^{-cx^\gamma},\ x > 0$$

[2] Бурр предложил это распределение в 1942г.

Вопрос для самоподготовки 1.5. Вывести формулу для математического ожидания распределения Вейбулла.

Вопрос для самоподготовки 1.6. Проверить, что формула для дисперсии распределения Вейбулла, приведенная в "Таблицах сводится к формуле для дисперсии экспоненциального распределения при $\gamma = 1$.

Функции плотности двухпараметрического распределения Парето, распределения Бурра и распределения Вейбулла могут быть проинтегрированы. Таким образом функции распределения имеют простой вид и вероятности могут быть легко найдены.

Пример 1.3 Найти формулу для медианы двухпараметрического распределения Парето. Начертите график медианы и математического ожидания как функций параметра α и прокомментируйте его.

Решение По определению, медианой m является точка, в которой $P(X \leqslant m) = \frac{1}{2}$. Полагая $\alpha > 1$, будем иметь:

$$\frac{1}{2} = \int_0^m \alpha\lambda^\alpha(\lambda + x)^{-\alpha-1}dx = -\lambda^\alpha \frac{1}{(\lambda + x)^\alpha}\bigg|_0^m = 1 - \left(\frac{\lambda}{\lambda + m}\right)^\alpha$$

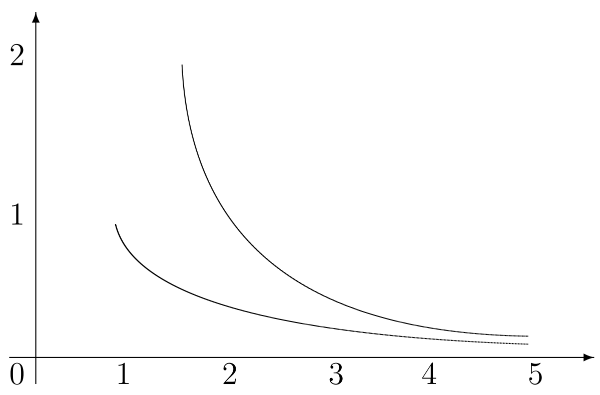
Данная формула может быть переписана в следующем виде:

$$m = \lambda(\sqrt[\alpha]{2} - 1)$$

Математическое ожидание равно:

$$\mu = \frac{\lambda}{\alpha - 1}$$

Графики для m/λ (нижний график) и μ/λ (верхний график) для значений $\alpha > 1$ показывают, что математическое ожидание всегда больше медианы, т.е. распределение Парето всегда имеет положительную асимметрию.



Вопрос для самоподготовки 1.7. Вычислить долю исков в размере от 2500 до 5000, если величина индивидуального иска подчиняется распределению Вейбулла с параметрами $c = 0.00001, \gamma = 1.5$

Вопрос для самоподготовки 1.8. Вычислить долю исков превышающих 300000, если величина индивидуального иска (измеренная в тысячах) подчиняется обобщенному распределению Парето с параметрами $\alpha = 5, \lambda = 200, k = 2$

2.4 Влияние инфляции на величину страховых выплат

Пусть для моделирования возможных исков, входящих в состав страхового портфеля, используется гамма распределение. Какое влияние оказывает инфляция на на распределение величины индивидуального риска?

Пусть случайная величина X индивидуального иска в году 0 подчиняется гамма распределению $\Gamma(\alpha, \lambda)$. Пусть инфляция составляет 10% в год, тогда случайная величина иска в году 1 будет равна $Y = 1.1X$. Мы можем вычислить функцию плотности распределения этой случайной величины путем преобразования интеграла. Используя замену переменных $y = 1.1x$, будем иметь:

$$\int_0^\infty \frac{1}{\Gamma(\alpha)}\lambda^\alpha x^{\alpha-1}e^{\lambda x}dx = \int_0^\infty \frac{1}{\Gamma(\alpha)}\lambda^\alpha \left(\frac{y}{1.1}\right)^{\alpha-1} e^{-\lambda y/1.1}\frac{dy}{1.1} =$$

$$= \int_0^\infty \frac{(\lambda/1.1)^\alpha}{\Gamma(\alpha)} y^{\alpha-1}e^{-\frac{\lambda}{1.1}y}dy$$

Сгруппировав члены в подынтегральном выражении, можно заметить, что оно является функцией плотности распределения, соответствующей гамма распределению $\Gamma(\alpha, \lambda/1.1)$, следовательно, случайная величина Y имеет это распределение. Заметим, что математическое ожидание гамма распределения $\Gamma(\alpha, \lambda/1.1)$ равно $\frac{1.1\alpha}{\lambda}$, следовательно, средняя величина иска увеличится на 10%, как и ожидалось.

Вопрос для самоподготовки 1.9. Насколько возрастет дисперсия величины индивидуального иска?

Вопрос для самоподготовки 1.10. Пусть величина индивидуального иска подчиняется распределению Парето $Pareto(\alpha, \lambda)$. Какому распределению будет подчиняться величина индивидуального иска, если инфляция составила $100k\%$?

2.5 Краткое изложение полученных результатов

Вы должны уметь вычислять вероятности и находить формулы для математического ожидания и дисперсии для всех стандартных функций распределения ущерба. Также вам необходимо уметь находить формулы для производящих функций некоторых распределений. Если нахождение некоторых из них вызывает у вас проблемы, то ниже приведен список результатов, которые вы должны уметь доказывать, с рекомендациями о том, каким способом можно это сделать.

Вы не должны заучивать содержимое этой таблицы наизусть. Гораздо важнее понимание используемых математических методов, так что вы сможете использовать их при решении конкретных задач, предложенных на экзамене.

Распределение	Величина	Способ вычисления
$\Gamma(\alpha, \lambda)$	Вероятности	Использовать распределение χ^2
	Производящая функция моментов	Привести интеграл к виду гамма распределения $\Gamma(\alpha, \lambda - t)$
	Математическое ожидание и дисперсия	Использовать производящую функцию моментов или... Привести интеграл к виду гамма распределения
$Exp(\lambda)$	Вероятности	Прямое интегрирование
	Функция распределения	Так же как и вероятности

Распределение	Величина	Способ вычисления
	Производящая функция моментов	Привести интеграл к виду экспоненциального распределения $Exp(\lambda - t)$
	Математическое ожидание и дисперсия	Проинтегрировать по частям или представить производящую функцию моментов в виде ряда
$Pareto(\alpha, \lambda)$	Вероятности	Выполнить замену $u = \lambda + x$
	Функция распределения	Так же как и вероятности
	Математическое ожидание и дисперсия	Выполнить замену $u = \lambda + x$, или интегрировать по частям, или представить x в числителе как $(\lambda + x) - \lambda$ и вычислить два получившихся интеграла
Обобщенное $Pareto(\alpha, \lambda, k)$	Вероятности	Выполнить замену $u = \lambda + x$ (подходит только в случае "хороших"значений k)
	Функция распределения	Так же как и вероятности
	Математическое ожидание и дисперсия	Привести интеграл к виду обобщенного распределения Парето с параметрами $\alpha - 1, \lambda, k + 1$ (для вычисления математического ожидания) либо $\alpha - 2, \lambda, k + 2$ (для вычисления дисперсии)
$N(\mu, \sigma^2)$	Вероятности	Использовать таблицы
	Производящая функция моментов	Выполнить замену $z = \frac{x-\mu}{\sigma}$, а затем выделить полный квадрат под знаком интеграла
$LogN(\mu, \sigma^2)$	Вероятности	Использовать соответствующее нормальное распределение. Например, $P(X > k) = P(N(\mu, \sigma^2) > \log k)$
	Математическое ожидание и дисперсия	Использовать производящую функцию моментов соответствующего нормального распределения или выполнить замену $z = \frac{\log x - \mu}{\sigma}$ и проинтегрировать по частям

Распределение	Величина	Способ вычисления
$Weibull(c, \gamma)$	Вероятности	Выполнить замену $u = cx^{\gamma}$
	Функция распределения	Так же как и вероятности
	Математическое ожидание и дисперсия	Выполнить замену $u = cx^{\gamma}$ и привести конечный интеграл к виду гамма распределения
$Burr(\alpha, \lambda, \gamma)$	Вероятности	Выполнить замену $u = \lambda + x^{\gamma}$
	Функция распределения	Так же как и вероятности
	Математическое ожидание и дисперсия	Выполнить замену $u = \lambda + x^{\gamma}$ и привести конечный интеграл к виду гамма распределения

2.6 Смешанные распределения

До настоящего момента мы подразумевали, что все возможные иски из страхового портфеля будут иметь одно и то же распределение, которое характеризуется одним (возможно неизвестным) набором параметров. Это предположение нереалистично в том смысле, что различные виды страховых полисов могут вести к искам, величины которых подчиняются различным распределениям.

Будем рассматривать портфель, включающий в себя полисы страхования автотранспорта. Предполагаем, что размер возможного иска по каждому полису имеет экспоненциальное распределение. Более реалистичным предположением является то, что параметр этого экспоненциального распределения не фиксирован и будет изменяться от полиса к полису. В этом случае суммарное распределение ущерба будет являться смесью ущерба из большого диапазона различных экспоненциальных распределений, каждое из которых со своим собственным значением параметра. Возникает вопрос: каково суммарное распределение ущерба от страхового портфеля? Ответ на него будет зависеть от того, каким образом различные значения параметра распределены среди владельцев полисов. Вместо рассмотрения λ в качестве фиксированного параметра, нам теперь необходима модель распределения значений λ по всему страховому портфелю.

Пусть λ_i является параметром экспоненциального распределения для i-го владельца полиса, и суммарное распределение λ_i является гамма распределением $\Gamma(\alpha, \sigma)$. Можем ли мы найти распределение суммарных потерь?

Мы знаем, что условное распределение $X|\lambda$ является экспоненциальным $Exp(\lambda)$ и что λ имеет гамма распределение $\Gamma(\alpha, \sigma)$. Взвешивая индивидуальные распределения с помощью функции плотности значений параметра, получим следующее безусловное распределение:

$$P(X = x) = \int\limits_0^\infty f_{X|\lambda}(x) f_\lambda(\lambda) d\lambda$$

Найдем безусловное распределение, 'исключая интегрированием' параметр λ. Имеем:

$$P(X = x) = \int\limits_0^\infty \lambda e^{-\lambda x} \frac{1}{\Gamma(\alpha)} \delta^\alpha \lambda^{\alpha-1} e^{-\delta\lambda} d\lambda$$

Приводя интеграл к виду функции распределения для гамма распределения $\Gamma(\alpha + 1, \delta + x)$, будем иметь:

$$P(X = x) = \frac{\Gamma(\alpha+1)\delta^\alpha}{\Gamma(\alpha)(\delta+x)^{\alpha+1}} \int\limits_0^\infty \frac{1}{\Gamma(\alpha+1)} (\delta+x)^{\alpha+1} \lambda^\alpha e^{-(\delta+x)\lambda} d\lambda$$

Но интеграл просто равен 1, таким образом получаем, что:

$$P(X = x) = \frac{\alpha\delta^\alpha}{(\delta+x)^{\alpha+1}}, \ x > 0$$

Мы получили функцию плотности распределения $Pareto(\alpha, \delta)$. Таким образом можно сказать, что когда экспоненциальные распределения ущерба смешиваются при помощи гамма распределения, то итоговым распределением будет распределения Парето $Pareto(\alpha, \delta)$.

Вопрос для самоподготовки 1.11. Ежегодное количество исков по индивидуальным полисам, входящим в страховой портфель, имеет распределение Пуассона с параметром θ. Изменение θ в зависимости от полиса моделируется в предположении, что индивидуальные значения θ имеют гамма распределение $\Gamma(\alpha, \delta)$ на страховом портфеле. Вычислить смешанное распределение ежегодного числа исков по каждому из полисов в страховом портфеле.

Вопрос для самоподготовки 1.12. Число исков по индивидуальным полисам страхового портфеля имеет распределение $B(n, p)$. Изменение параметра p подчиняется бета распределению $Beta(\alpha, \beta)$. Найти смешанное распределение.

§3 Подбор подходящего распределения ущерба

3.1 Выбор распределения

Выбор распределения, подходящего для моделирования определенного страхового портфеля, включает в себя следующие шаги:

1. Определение общего вида распределения размеров иска (например, подсчет числа исков, величина которых попадает на тот или иной интервал, и построение гистограммы).

2. Выбор семейства распределений, имеющих схожую форму.

3. Оценка параметров распределения (например методом моментов или методом максимального правдоподобия)

4. Проверка гипотезы для определения того, является ли выбранное распределение и его параметры хорошей моделью для рассматриваемых возможных исков (например, используя критерий χ^2)

При выборе подходящего семейства распределений для описания распределения ущерба необходимо учесть следующие факторы:

- Цель
  Пригодность каждого конкретного распределения будет зависеть от цели анализа. Например, если распределение используется для вычисления базовой премии, то наиболее важной задачей будет являться выбор распределения, наиболее точно описывающего диапазон значений, на который приходится основная часть исков. В то же время, если распределение используется для обзора планов перестрахования или обдумывания эффектов от очень больших исков, то приоритетной задачей будет являться выбор распределения, наиболее точно описывающего верхнюю границу исков.

- Общий вид
  Распределение ущерба должно соответствовать рассматриваемому распределению исков с учетом интервалов допустимых значений случайных величин, дисперсии, асимметрии и общего вида. Например, нормальное распределение обычно не является подходящим для моделирования величины индивидуального иска, т.к. оно всегда является симметричным, в то время как большинство рассматриваемых исков имеют положительную асимметрию. Так же и экспоненциальное распределение не является предпочтительным при

описании небольших исков вследствие того, что функция плотности распределения убывает монотонно, вместо того, чтобы иметь "горб в центре"как у многих других распределений.

- "Хвосты"
  Вследствие того что очень большие риски относительно редки, достаточно сложно быть уверенным в точности вида распределения ущерба на верхней границе. В то же время крайне важно не занизить оценку числа исков, попадающих на эту часть распределения, т.к. именно они по определению включают в себя наибольшую часть средств. Функции плотности большинства стандартных статистических распределении включают в себя экспоненту (например, гамма распределение содержит $e^{-\lambda x}$). Вследствие того что эти распределение имеют экспоненциальный "хвост они убывают очень быстро. Другие распределения, такие как Парето или Бурра, содержат x в некоторой степени. Вследствие того что эти распределения имеют полиномиальные хвосты, они убывают с меньшей скоростью. Распределения ущерба с полиномиальными "хвостами"могут быть более пригодными для моделирования больших исков, потому что их "тяжелые хвосты"уменьшают риск недооценки частоты больших исков.

3.2 Выбор значений параметров

Как только вид моделирующего распределения выбран, необходимо решить каким образом оценивать значение каждого параметра. Мы рассмотрим три метода определения величин параметров:

1. Метод моментов
2. Оценка методом максимального правдоподобия
3. Метод процентилей

Далее приводится краткое описание этих методов.

Метод моментов

Для получения оценки параметра по методу моментов, необходимо приравнять выборочные и теоретические начальные моменты. Например, если бы мы пытались оценить значение одного параметра, то решали бы уравнение:

$$E(X) = \frac{\sum x_i}{n}$$

Т.е. приравнивали бы первые начальные моменты.

Если бы мы пытались найти оценки для двух параметров (например, если бы мы подбирали гамма распределение и нам нужно было бы найти оценки для α и λ), то решали бы систему уравнений:

$$\begin{cases} E(X) = \dfrac{\sum x_i}{n} \\ E(X^2) = \dfrac{\sum x_i^2}{n} \end{cases}$$

В действительности в двухпараметрическом случае оценки обычно получаются приравниванием выборочных и теоретических математических ожиданий и дисперсий. В том случае, если в знаменателе выборочной дисперсии стоит n, мы будем иметь те же самые оценки, которые были бы получены приравниванием первых двух начальных моментов. В более общем случае, мы можем использовать столько уравнений вида $E(X^k) = \frac{\sum x_i^k}{n}, k = 1, 2, \ldots$, сколько нам необходимо для вычисления оценок значимых параметров.

Пример 1.4 Основываясь на анализе прошедших исков, страховая компания считает, что средняя величина индивидуального иска в выделенной категории в следующем году составит 5000, а стандартное отклонение 7500. Необходимо оценить долю исков, размер которых превысит 25000, если величина индивидуального иска имеет логнормальное распределение.

Решение Вычисление формул для математического ожидания и стандартного отклонения логнормального распределения приводит к следующим выражениям:

$$e^{\mu+\frac{1}{2}\sigma^2} = 5000,\ e^{\mu+\frac{1}{2}\sigma^2}\sqrt{e^{\sigma^2}-1} = 7500$$

Разделим второе выражение на первое. Будем иметь:

$$\sqrt{e^{\sigma^2}-1} = \frac{7500}{5000} = 1.5 \Rightarrow \sigma^2 = 1.179$$

Теперь мы можем вычислить значение μ:

$$\mu = \ln 5000 - \frac{1}{2}1.179 = 7.928$$

Доля исков, превышающих 25000, является всего лишь вероятностью того, что размер индивидуального иска превысит значение 25000:

$$P(X > 25000) = P(\ln X > \ln 25000) = P\big(N(7.928, 1.179) > \ln 25000\big) =$$
$$= P\left(N(0,1) > \frac{\ln 25000 - 7.928}{\sqrt{1.179}}\right) = 1 - \Phi(2.025) = 0.021$$

Т.е. размер 2.1% исков превысит 25000.

Вопрос для самоподготовки 1.13. Повторите вычисления, считая, что величина иска подчиняется распределению Парето. Прокомментируйте свой ответ.

Оценка максимального правдоподобия

Нахождение оценки максимального правдоподобия включает в себя следующие шаги:

1. Выпишите функцию правдоподобия. Если правдоподобие основано на наборе известных значений $x_1, x_2, \ldots, x_n$, то функция правдоподобия примет вид $f(x_1)f(x_2)\cdots f(x_n)$, где $f(x)$ является функцией плотности распределения (или функцией вероятности в дискретном случае), параметры которого оцениваются.

2. Прологарифмируйте. Это упростит вычисления.

3. Продифференцируйте функцию правдоподобия по каждому неизвестному параметру и приравняйте полученное(-ые) выражение(-я) к нулю.

4. Решите итоговое(-ые) уравнение(-я) для нахождения оценки максимального правдоподобия

5. Путем нахождения второй производной проверьте, что значения, которые вы нашли, *максимизируют* функцию правдоподобия.

Пример 1.5 Страховая компания моделирует стоимость ремонта застрахованных автомобилей, попавших в аварию, используя экспоненциальное распределение. Найти оценку максимального правдоподобия средней стоимости, если средняя стоимость ремонта составила 2200 и было отремонтировано 1000 автомобилей.

Решение Пусть $X_1, X_2, \ldots, X_n$ означает индивидуальную стоимость ремонта (где $n = 1000$).

$$L = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i} = \lambda^n e^{-\lambda n \overline{x}}$$

Где $\overline{x} = \frac{1}{n}\sum\limits_{i=1}^{n} x_i$ означает среднюю величину иска.

Для нахождения оценки максимального правдоподобия, нам необходимо вычислить значение λ, максимизирующее величину L или, иначе, значение, максимизирующее $\ln L$:

$$\ln L = n \ln \lambda - \lambda n \overline{x}$$

Продифференцируем это выражения для нахождения стационарных точек:

$$\frac{\partial}{\partial\lambda}\ln L = \frac{n}{\lambda} - n\overline{x}$$

Приравнивая левую часть к нулю, будем иметь:

$$\widehat{\lambda} = \frac{1}{\overline{x}} \Rightarrow \widehat{\lambda} = 1/2200$$

Вторая производная $\frac{\partial^2}{\partial\lambda^2}\ln L = -\frac{n}{\lambda^2} < 0$, а это означает, что найденная нами стационарная точка является точкой максимума.

Математическое ожидание экспоненциального распределения $Exp(\lambda)$ равно $1/\lambda$. Используя тот факт, что оценка максимального правдоподобия функции равна функции оценки, получаем, что оценка максимального правдоподобия для средней величины иска равна:

$$1/\widehat{\lambda} = \overline{x} = 2200$$

Вопрос для самоподготовки 1.14. Случайная величина X имеет распределение Бурра с параметрами $\gamma = 2, \lambda = 500$.Показать, что оценка максимального правдоподобия параметра α, основанная на случайной выборке $x_1, x_2, \ldots, x_n$, равна $\widehat{\alpha} = \frac{n}{\sum\ln(500+x_i^2)-n\ln 500}$ и вычислить её значение, если была получена следующая выборка: $52, 109, 114, 163, 181$

Метод процентилей

Этот метод включает в себя установление равенства между выборочным и теоретическим процентилями. Выбор процентилей будет зависеть от числа оцениваемых параметров. В однопараметрическом случае обычно используются выборочная и теоретическая медианы. В двухпараметрическом случае могут быть использованы верхняя и нижняя квартили.

Пример 1.6 Используйте метод процентилей для вычисления параметров распределения Вейбулла, основываясь на следующей случайной выборке (значения были отсортированы по возрастанию). Величины исков выражены в тысячах.

0.1	0.5	2.2	4.1	28.1
0.2	0.7	2.6	5.9	30.0
0.2	0.9	2.9	6.2	49.2
0.3	1.3	3.2	12.1	63.8
0.4	1.8	3.3	15.2	118.0

Решение Так как мы оцениваем два параметра, то будем использовать верхнюю и нижнюю квартили для нахождения оценок c и γ в распределении Вейбулла.

Функция распределения имеет следующий вид:

$$\int_0^x c\gamma t^{\gamma-1}e^{-ct^\gamma}dt = -e^{-ct^\gamma}\Big|_0^x = 1 - e^{-cx^\gamma}$$

Таким образом верхняя квартиль распределения Вейбулла — значение x, являющееся решением уравнения:

$$0.75 = 1 - e^{-cx^\gamma} \Rightarrow x = \left(\frac{\ln 4}{c}\right)^{\frac{1}{\gamma}}$$

Решение уравнения $0.25 = 1 - e^{-cx^\gamma}$ для нахождения нижней квартили даст нам значение:

$$x = \left(\frac{\ln 4/3}{c}\right)^{\frac{1}{\gamma}}$$

Исходя из имеющихся 25 выборочных значений, выборочные квартили будут соответствовать значению $\frac{1}{4} * 25 + \frac{1}{2} = 6.75$, которое равно 0.65, и значению $\frac{3}{4} * 25 + \frac{1}{2} = 19.25$, которое равно 12.875. Таким образом, решая систему уравнений:

$$\begin{cases} \left(\frac{\ln 4}{c}\right)^{\frac{1}{\gamma}} = 12.875 \\ \left(\frac{\ln 4/3}{c}\right)^{\frac{1}{\gamma}} = 0.65 \end{cases}$$

мы получим искомые оценки:

$$\widehat{\gamma} = 0.526, \widehat{c} = 0.361$$

3.3 Проверка качества выбранного распределения

Одним из способов проверки того, может ли выбранное распределение ущерба служить хорошей моделью для рассматриваемых величин исков, является критерий χ^2.

Пример 1.7 Анализ стоимости ремонта в примере 1.5 дает нам следующие значения в различных интервалах:

0-1000: 200	1000-2000: 3000	2000-3000: 250
3000-4000: 150	4000-5000: 100	5000+: 0

Используем эту информацию для проверки, является ли экспоненциальное распределение хорошей моделью для стоимости индивидуального ремонта.

Решение Мы проверяем:

H_0: Стоимость имеет экспоненциальное распределение.

H_1: Стоимость имеет не экспоненциальное распределение.

Для применения критерия χ^2 нам необходимо вычислить ожидаемые значения, т.е. наиболее вероятные значения в каждом интервале, при условии, что цена подчиняется экспоненциальному распределению. Используя нашу оценку параметра $\lambda = 1/2200$, вероятность того, что цена индивидуального ремонта попадет на интервал 2000-3000, может быть вычислена следующим образом:

$$\int_{2000}^{3000} \lambda e^{\lambda x} dx = -e^{\lambda x}\Big|_{2000}^{3000} = 0.1472$$

Тогда ожидаемое число исков на этом интервале равно $1000 * 0.1472 = 147.2$ Ожидаемые значения для всех интервалов могут быть вычислены аналогично:

$$365.3, 231.8, 147.2, 93.4, 59.3, 103.0$$

Теперь можно вычислить значение статистики χ^2:

$$\chi^2 = \frac{(200-365.3)^2}{365.3} + \frac{(300-231.8)^2}{231.8} + \cdots + \frac{(0-103)^2}{103} = 331.89$$

У нас 6 интервалов, но мы установили равенство для итоговых значений и оценили один параметр. Таким образом у нас $6 - 1 - 1 = 4$ степени свободы. Вычисленное значение χ^2 намного превышает 14.86 — наибольшее значение для распределения χ^2 с 4 степенями свободы при доверительной вероятности 99.5%. Таким образом мы можем отвергнуть гипотезу H0 с абсолютной уверенностью и сделать вывод, что стоимость ремонтных работ не подчиняется экспоненциальному распределению.

Этот вывод также подтверждается следующим наблюдением: если бы значения имели экспоненциальное распределение, то мы ожидали бы увидеть монотонное убывание от интервала к интервалу. В то же время, мы видим, что значение в первом интервале на 100 меньше, чем во втором.

§4 Вычисление премий

4.1 Частоты исков

Фактическая частота исков для группы полисов страхования — это среднее количество исков на полис:

$$\text{Фактическая частота исков} = \frac{\text{Число исков}}{\text{Среднее число полисов}}$$

Ожидаемая частота исков для группы страховых полисов — это ожидаемое число исков на полис.

Пример 1.8 За прошедшие 5 лет страховая компания удовлетворила 7000 исков, касающихся молодых водителей со старыми автомобилями. Среднее число застрахованных водителей в данной группе за этот период составило 5000. Вычислите ежегодную частоту исков для данной категории водителей.

Решение Фактическая частота исков за 5 лет составила $7000/5000 = 1.4$, что соответствует ежегодной частоте исков $1.4/5 = 0.28$, т.е. 28% в год.

§5 Рисковые премии и офисные премии

Рисковая премия за полис общих видов страхования равна ожидаемой величине исков:

Рисковая премия = Ожидаемая частота исков ∗ Средняя величина иска

Пример 1.9 Вычислите рисковую премию на следующий год, если в предыдущем примере страховщик ожидает, что средняя величина выплат для данной категории водителей составит 1500.

Решение Рисковая премия на следующий год равна:

Ожидаемая частота исков ∗ Средняя величина иска $= 0.28 * 1500 = 420$
Т.е. 420 в год.

Офисная премия—реальная премия, взимаемая страховщиком. Теоретическая офисная премия может быть вычислена путем изменения рисковой премии (её увеличения) с учетом издержек, комиссии, желаемой прибыли и различных непредвиденных расходов и (её уменьшения) с учетом инвестиционной прибыли.

Пример 1.10 Вычислите теоретическую офисную премию, которую должна взимать компания в следующем году с упомянутой в предыдущем примере категории водителей, если:

- издержки на урегулирование каждого иска равны 100
- комиссия составляет 20% от офисной премии
- компания хочет иметь прибыль в размере 5% от величины офисных премий

Инвестиционный доход не учитывается.

Решение Теоретическая офисная премия на следующий год должна быть:

$$P = 420 + 100 * 0.28 + 0.2P + 0.05P \Rightarrow P = 448/0.75 = 597$$

Т.е 597 в год.

Вопрос для самоподготовки 1.15. Используя информацию, приведенную ниже, вычислите офисную премию для полисов в этот страховом портфеле:

Среднее число действительных страховых полисов в год	6200
Общее число исков, полученных за последние 10 лет	3150
Предполагаемое (текущее) распределение величины иска	$\Gamma(50, 0.02)$
Предполагаемые расходы по урегулированию иска (каждого)	50
Надбавка на прибыль и комиссия	35%

Вопрос для самоподготовки 1.16. Назовите три причины, по которым вычисления в предыдущем вопросе для самоподготовки могут не дать на практике адекватного результата.

Вопрос для самоподготовки 1.17. При исследовании величин страховых возмещений, уплаченных компанией по страховым полисам, входящим в портфель, были получены следующие результаты:

$0 - 500$:	120
$500 - 1000$:	386
$1000 - 1500$:	490
$1500 - 2000$:	322
$2000 - 2500$:	62

Для того чтобы проверить, является ли нормальное распределение адекватной моделью для представленных величин исков, из имеющихся данных были вычислены оценки μ и σ^2, а также соответствующие ожидаемые частоты для каждой категории: $104, 327, 564, 309, 76$. Проверьте, является ли нормальное распределение хорошей моделью для данных исков.