

УДК 311

О предсказании поведения «хвостов» распределений и оценке «ожидаемых непредвиденных» потерь при управлении рисками

ISSN 1812-5220
© Проблемы анализа риска, 2017

А. А. Быков,
Российское научное общество
анализа риска,
г. Москва

Аннотация

В работе рассматриваются основные аспекты асимптотической теории вероятностей экстремальных событий, позволяющие при количественных оценках рисков решать важные практические задачи: 1) обоснованно предсказывать асимптотическое поведение «хвостов» распределений на основе статистической обработки имеющихся данных; 2) рассчитывать ожидаемое значение превышения порогового значения потерь (предсказывать величину «ожидаемых непредвиденных» потерь). Кроме того, в работе демонстрируется эффективность использования технологий графического статистического анализа, основанная на использовании интервалов равной вероятности вместо классических интервалов равной длины при статистической проверке гипотез и выборе законов распределения.

Ключевые слова: управление риском, оценка риска, прогнозирование непредвиденных убытков, асимптотическая теория вероятностей экстремальных событий, статистика экстремальных значений, графические статистические технологии.

Содержание

Введение

1. О роли центральной предельной теоремы. Почему так «популярны» нормальное и логарифмически нормальное распределения
2. Проверка гипотезы о нормальном законе распределения. Классический статистический метод с использованием критерия Хи-квадрат
3. Интервалы равной длины или равной вероятности?
4. Интервалы равной вероятности
5. Процедура построения графиков квантилей в общем случае
6. Основные тестовые распределения для практических приложений
7. Главные результаты классической асимптотической теории экстремальных значений
8. Моделирование надпороговых значений: полупараметрический подход
9. Моделирование надпороговых значений: параметрический подход

Заключение

Литература

Введение

Количественная оценка риска при управлении рыночными, кредитными и в ряде случаев операционными рисками осуществляется на практике с использованием методологии Value-At-Risk (VaR). При этом в рамках методологии используются количественные метрики риска, определяющие уровень потерь по риску, который возможен в течение определенного временного интервала с заданной доверительной вероятностью. Методология применяется к расчету таких показателей, как Earnings at Risk — EaR, Market VaR, Credit VaR, Operational VaR, для оценки рисков, присущих инвестиционным проектам (Net Present VaR), при установлении допустимых уровней риска, расчете резервов и в ряде других практических задач в области управления рисками.

Методология VaR применяется также при расчете т.н. экономического капитала — объема капитала, необходимого для покрытия непредвиденных потерь (убытков), и его количественных показателей для установления максимального убытка, который компания может допустить при заданном доверительном интервале и временном горизонте. Непредвиденные потери являются потенциальными потерями компании при уровне значимо-

сти (например, 99,99%), превышающем доверительный уровень (например, 97,5%) — наибольший прогнозируемый уровень ожидаемых потерь (рис. 1).

В данной работе будут рассмотрены вопросы, связанные с моделированием функции распределения потерь (убытков), прогнозированием поведения «хвостов» распределения, а также расчетом «ожидаемых непредвиденных» потерь.

1. О роли центральной предельной теоремы. Почему так «популярны» нормальное и логарифмически нормальное распределения

Среди законов распределения, с которыми мы встречаемся в практических (естественнонаучных, технических, социально-экономических) приложениях, нормальное распределение играет особую роль.

«Популярность» нормального распределения связана с выводами центральной предельной теоремы, суть которой заключается в том, что если X_1, \dots, X_n — взаимно независимые одинаково распределенные случайные величины, имеющие математическое ожидание $\mu = 0$ и дисперсию $\sigma^2 = 1$, то при $n \rightarrow \infty$ распределение нормированных сумм



Рис. 1. Иллюстрация ожидаемых и непредвиденных потерь на графике плотности распределения (Деклерк, Шибав, 2008)

$S_n^* = \frac{X_1 + \dots + X_n}{\sqrt{n}}$ стремится к нормальному распределению $N(0,1)$ с плотностью $f = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Это одна из формулировок. В более общем виде: распределение центрированных и нормированных сумм н.о.р.с.в. стремится к нормальному распределению $N(0,1)$ при конечных математических ожиданиях и дисперсиях (Феллер, 1984; Смирнов, Дуинин-Барковский, 1969). Обобщение центральной предельной теоремы на случай неодинаково распределенных величин и ослабление требования конечной дисперсии было осуществлено Линдбергом (Lindeberg, 1922).

Следует учитывать тот факт, что на случайную величину действуют факторы, отклоняющие ее в ту или иную сторону. С математической точки зрения это отклонение можно представить в виде следующих вариантов.

- Аддитивное действие случайных факторов, когда $X_i = X_0 + \Delta x_1 + \Delta x_2 + \dots + \Delta x_n$, где Δx_j может быть как больше нуля, так и меньше для всех $j = 1 \div n$. При этом $X_i - X_0 = \sum_{j=1}^n \Delta x_j$.

Поскольку Δx_j — величина случайная, то $X_i - X_0$ представляет собой сумму большого числа случайных величин. Следовательно, согласно центральной предельной теореме величина $y = X_i - X_0$ асимптотически приближается к нормальному распределению.

- Мультипликативное действие случайных факторов, когда $X_i = X_0 \cdot k_1 \cdot k_2 \cdot \dots \cdot k_n$, где k_j может быть как больше, так и меньше единицы для всех

$j = 1 \div n$. При этом $\ln \frac{X_i}{X_0} = \ln \prod_{j=1}^n k_j = \sum_{j=1}^n \ln k_j$.

Поскольку k_j — случайная величина, то $\ln k_j$ — тоже есть величина случайная. Таким образом,

$\ln \frac{X_i}{X_0} = \ln X_i - \ln X_0$ представляет собой сумму большого числа случайных величин и, согласно центральной предельной теореме, асимптотически приближается к нормальному распределению. Следовательно, величина $y = \frac{X_i}{X_0}$ подчинена логарифмически нормальному распределению (Иванов, 2005).

2. Проверка гипотезы о нормальном законе распределения. Классический статистический метод с использованием критерия Хи-квадрат

Продemonстрируем классическую, рекомендуемую учебными изданиями, например (Вуколов, Ефимов, Земсков, 1984), процедуру проверки гипотезы о нормальности распределения с использованием критерия χ^2 (Хи-квадрат).

Используем иллюстративные данные — выборку, состоящую из 88 чисел, с размахом (ω), равным 43,2. Элементы выборки объединим в группы. Для этого интервал, содержащий все элементы выборки, разбиваем на 8 частичных непересекающихся интервалов, имеющих одинаковую длину (b), равную 5,4. После того как частичные интервалы выбраны, определяем частоты — количество n_i^* элементов выборки, попавших в i -й интервал. Результаты группировки представлены в табл. 1.

Полигон и гистограмма частот группированной выборки представлены на рис. 2 и 3 соответственно.

Для проверки гипотезы о нормальности распределения значений используем критерий χ^2 . Процедура применения критерия χ^2 для проверки гипотезы состоит из следующих этапов (Вуколов, Ефимов, Земсков, 1984):

- 1) по выборке наблюдений случайной величины X находим оценки неизвестных параметров предполагаемого закона распределения $F(x)$;

- 2) определяем частоты $n_i, i = 1, 2, \dots, r$, с которыми каждое значение или группа значений встречается

Результаты группировки Таблица 1
иллюстративных статистических данных

№ интервала i	Границы интервала	Середина интервала x_i^*	Частота n_i^*
1	16,2—21,6	18,9	3
2	21,6—27	24,3	3
3	27—32,4	29,7	7
4	32,4—37,8	35,1	9
5	37,8—43,2	40,5	21
6	43,2—48,6	45,9	22
7	48,6—54	51,3	12
8	54—59,4	56,7	11

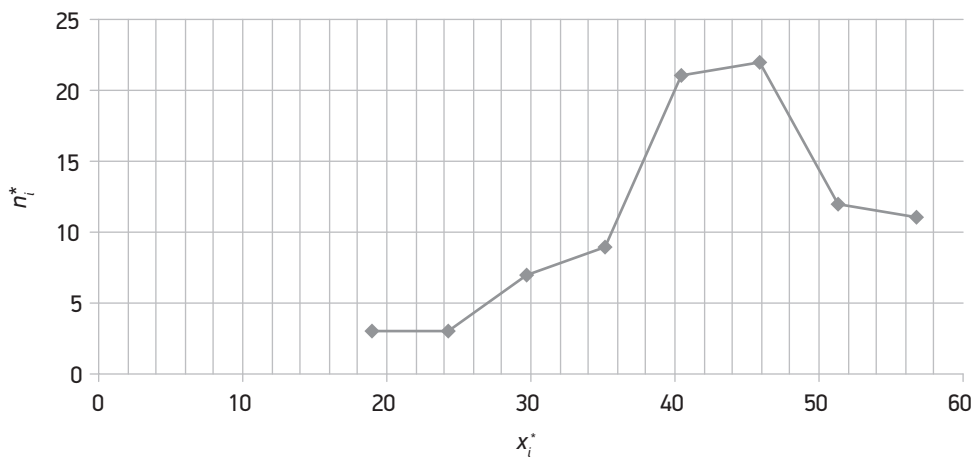


Рис. 2. Полигон частот группированной выборки

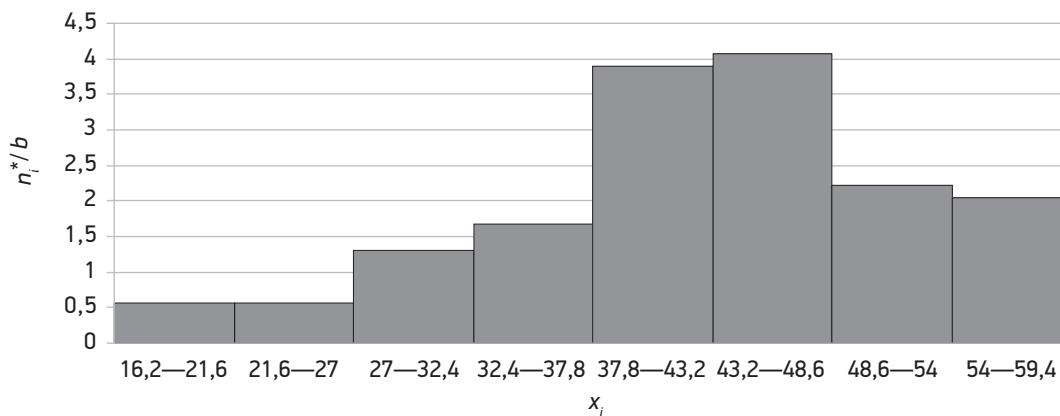


Рис. 3. Гистограмма частот группированной выборки

в выборке. Очевидно, что $\sum_{i=1}^r n_i = n$, где n — число элементов выборки;

3) используя предполагаемый закон распределения $F(x)$, вычисляем вероятности $p_i, i = 1, 2, \dots, r$, с которыми случайная величина X принимает каждое значение, или вероятность появления группы значений. Очевидно, что $\sum_{i=1}^r p_i = 1$;

4) вычисляем выборочное значение статистики критерия:

$$\chi_B^2 = \sum_{i=1}^r \frac{(n_i^* - np_i)^2}{np_i}. \quad (2.1)$$

При этом необходимым является условие $np_i \geq 5$;

5) статистическое решение: гипотеза не противоречит выборке наблюдений на заданном уровне α , если $\chi_B^2 < \chi_{1-\alpha}^2(r - l - 1)$, где l — число пара-

метров распределения $F(x)$, которые оцениваются по выборке; если же $\chi_B^2 \geq \chi_{1-\alpha}^2(r - l - 1)$, то гипотеза отклоняется.

Необходимые для проверки гипотезы значения x_i^* и n_i^* представлены в табл. 2.

В качестве среднего значения генеральной совокупности используем среднее значение показателя, равное 44,4. С учетом того, что среднее по генеральной совокупности известно, среднееквадратическое отклонение вычисляем по формуле $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$,

где x_i — значение показателя; \bar{x} — среднее значение показателя; n — количество значений показателей.

В результате расчета получаем: $\sigma = 9,74$.

Результаты вычислений, необходимые для проверки гипотезы о нормальном распределении группированных данных

Таблица 2

x_i^*	n_i^*	z_i^*	$f(z_i^*)$	$np_i = \frac{nb}{\sigma} f(z_i^*)$	np_i	$n_i^* - np_i$	$\frac{(n_i^* - np_i)^2}{np_i}$
18,9	3	2,62	0,0129	0,629	9,138	3,862	1,632
24,3	3	2,06	0,0468	2,283			
29,7	7	1,51	0,1276	6,226			
35,1	9	0,96	0,2516	12,276	12,276	-3,276	0,874
40,5	21	0,4	0,3683	17,969	17,969	3,031	0,511
45,9	22	0,15	0,3945	19,248	19,248	2,752	0,393
51,3	12	0,71	0,3101	15,13	15,13	-3,13	0,648
56,7	11	1,26	0,1804	8,802	8,802	2,198	0,549

Результаты вычислений, необходимые для проверки гипотезы о нормальном распределении группированных данных, сведены в табл. 2, в которой в первом столбце приведены значения середин интервалов x_i , а во втором — значения соответствующих им частот n_i^* , $i = 1, 2, \dots, 8$. В третьем столбце приведены нормированные значения z_i^* , рассчитанные по следующей формуле:

$$z_i^* = \frac{|x_i^* - \bar{x}|}{\sigma}.$$

В четвертом столбце представлены значения плотности $f(z_i^*)$ нормального распределения $N(0, 1)$, определенные по таблице значений функции плотности нормального распределения.

В пятый столбец занесены значения np_i , вычисленные по формуле

$$np_i = \frac{nb}{\sigma} f(z_i^*).$$

Шестой столбец содержит значения np_i после объединения первых трех интервалов, так как для $i = 1$ и $i = 2$ значения $np_i < 5$. Наконец, седьмой и восьмой столбцы служат для расчета выборочного значения статистики критерия χ_B^2 .

В соответствии с формулой (2.1) $\chi_B^2 = 4,607$.

Так как по выборке определены оценки двух параметров, то $l = 2$, число степеней свободы равно $6 - 2 - 1 = 3$. Уровень значимости $\alpha = 0,01$. Согласно таблице квантилей Хи-квадрат распределения $\chi_p^2(l)$,

$\chi_{0,99}^2(3) = 11,3$. Таким образом, в связи с тем что $\chi_B^2 < \chi_{0,99}^2(3)$, гипотеза о нормальном распределении значений показателя верна.

Далее кратко представим основные преимущества графической техники, применяемой в статистике экстремальных значений, по сравнению с классическими методами статистического анализа при определении законов распределения случайных величин.

3. Интервалы равной длины или равной вероятности?

Рассмотрим специальные статистические средства, которые применяются для того, чтобы ответить на важный вопрос: «Обеспечивает ли выбранная модель распределения правдоподобное соответствие имеющемуся распределению случайной переменной?»

На первый взгляд кажется, что простейший процесс проверки заключается в сравнении теоретической и наблюдаемой плотностей распределения. В действительности же непосредственной связи между теоретической и наблюдаемой плотностями распределения нет, и вообще *определенной* наблюдаемой плотности не существует. Чтобы получить какую-то наблюдаемую плотность распределения классическими (общепринятыми) статистическими методами, кратко представленными в предыдущем разделе, мы должны выбрать некоторый класс интервалов разбиения и подсчитать

число наблюдений, содержащихся в соответствующих интервалах.

Здесь имеют место два произвольных шага:

- 1) выбор длины интервала и
- 2) выбор начальной точки деления.

Влияние первого шага хорошо известно, а о влиянии второго упоминается редко. Это обуславливается также и тем, что, к сожалению, имеется обычай публиковать только сгруппированные данные.

Неоднозначность понятия наблюдаемого распределения приводит к серьезным недостаткам при общепринятой проверке непрерывных распределений при помощи критерия χ^2 , так как только из-за сдвига начальной точки деления мы будем получать различные значения χ^2 . Для одних и тех же наблюдений, одинаковых теоретических предпосылок, одних и тех же параметров и даже одной и той же длины интервалов. При фактическом вычислении величины χ^2 используется только одно наблюдаемое распределение. Другие приводят к иным значениям этого критерия, которые могут отличаться в пределах выбранного уровня значимости.

Путь, свободный от этой неоднозначности, состоит в замене интервалов равной длины интервалами равной вероятности. Из этого подхода вытекают методы сравнения, основанные на сопоставлении каждого индивидуального наблюдения с соответствующей теоретической величиной.

4. Интервалы равной вероятности

Здесь и в дальнейшем предполагаем, что у нас есть выборка X_1, \dots, X_n одинаково распределенных независимых случайных величин с функцией распределения $P(X_i \leq x) = F(x)$.

В действительности мы не знаем вероятности $F(X_k)$ для каждого наблюдаемого (некоторого k -го) значения X_k . Однако N значений X , наблюдаемых в выборке, как можно показать (см. Смирнов, Дуин-Барковский, 1969), в среднем делят ось x так, что $(N + 1)$ полученных промежутков отвечают равным приростам функции $F(x)$.

Другими словами, расположив наблюдения в порядке неубывания и получив таким образом вариационный ряд величин X

$$X_1 \leq X_2 \leq \dots \leq X_N$$

для вероятностей $F(x)$ в соответствующих точках справедливо:

$$F(X_1) < F(X_2) < \dots < F(X_k) < \dots < F(X_N).$$

Так же как и сами X_k , они будут случайными величинами. Но, как можно показать (Смирнов, Дуин-Барковский, 1969), математические ожидания приростов функции $F(x)$ в каждом из интервалов $(-\infty, X_1], (X_1, X_2], \dots, (X_{N-1}, X_N], (X_N, \infty)$ будут равны между собой, т.е.

$$E[F(X_1) - 0] = E[F(X_2) - F(X_1)] = E[F(X_3) - F(X_2)] = \dots = E[F(X_N) - F(X_{N-1})].$$

Каждый из этих приростов равен $\frac{1}{N+1}$, поэтому

мы будем иметь:

$$E[F(x_1)] = \frac{1}{N+1}; E[F(x_2)] = \frac{2}{N+1}; \dots; E[F(x_N)] = \frac{N}{N+1}.$$

Далее, можно показать, что

$$Var[F(x_k)] = \frac{k(N-k+1)}{(N+1)^2(N+2)} < \frac{1}{4(N+2)}$$

и, следовательно, $F(x_k)$ по вероятности сходится к $E[F(x_k)]$ при $N \rightarrow \infty$.

Поэтому, приравнявая $F(x_k) = \frac{k}{N+1}$, мы сделаем

случайную ошибку, почти наверное сколь угодно малую, если N достаточно велико.

Далее, для закона распределения $\varphi(y)$, где $y = \alpha(x - \beta)$, причем α и β — параметры распределения, определим отвечающие $\frac{k}{N+1}$ значения

$y = Y_k$ такие, что $\varphi(Y_k) = \frac{k}{N+1}$. Мы получим последовательность $Y_1 < Y_2 < \dots < Y_k < \dots < Y_N$ такую, что в координатах (x, y) точки $Q(Y_k, X_k)$, где $k = 1, 2, \dots, N$, лежат вблизи теоретической прямой $y = \alpha(x - \beta)$ или $x = \frac{y}{\alpha} + \beta$ и отклоняются от нее лишь в силу наличия случайных ошибок.

Если бы величина X точно следовала бы закону $F(x) = \varphi(y)$ и мы точно знали бы всякий раз, какому значению вероятности $F(x)$ отвечает наблюдаемое значение x , то имели бы точную линейную зависимость $y = \alpha(x - \beta)$ или $x = \frac{y}{\alpha} + \beta$. На графике она

изобразилась бы прямой линией. Параметры α и β без труда определились бы, например, по координатам (y_0, x_0) и (y_1, x_1) двух каких-либо точек такого графика. Отсюда и вытекает основная идея построения графиков квантилей.

Таким образом, преимущество данного метода по сравнению с обычными способами построения кривых заключается в том, что метод преобразует теоретическую кривую $(F(x), x)$ в прямую линию. Если на квантиль-диаграмме наблюдается линейный участок, то прямая может быть построена с использованием, например, классического алгоритма метода наименьших квадратов.

Степень правдоподобия гипотезы может быть измерена посредством коэффициента корреляции или детерминации. Значение коэффициента детерминации или корреляции следует рассматривать только как статистический критерий соответствия модели эмпирическим данным. Основываясь на данных показателях, можно провести формальную проверку согласия или правдоподобия гипотезы. Гипотеза отвергается, если его значение слишком сильно отличается от единицы или, что эквивалентно, ниже, чем некоторое табулированное критическое значение. Но вопрос о том, правильно или неправильно выбрана функция распределения, может быть во многих случаях решен простой визуальной проверкой. Если достигнуто достаточно хорошее согласие, то предсказание (в ограниченных пределах) может быть сделано путем продолжения прямой.

5. Процедура построения графиков квантилей в общем случае

Таким образом, идея построения графиков квантилей, их еще называют *квантиль-квантиль графики* или *диаграммы* (сокращенно *QQ-графики* или *QQ-диаграммы*), основана на том факте, что для основных классов распределений квантили $Q(p)$ линейно связаны с соответствующими стандартными квантилями из этого класса распределений. Поскольку линейность на графике может быть легко замечена визуально и/или определена количественно посредством построения линии регрессии, такой подход идеально годится для того, чтобы ответить на поставленный в начале раздела вопрос: «Обеспечивает ли выбранная модель распределения правдо-

подобное соответствие имеющемуся распределению случайной переменной?»

Суммируя вышеизложенное, можно очертить общие идеи, принципы и схему построения графиков квантилей (Beirlant, Taugels, Vynckier, 1996). А именно, чтобы построить график квантилей на вероятностной бумаге, следует произвести следующие действия.

- Сначала данные наблюдения располагают в порядке неубывания, получая вариационный ряд величин

$$X_1^* \leq X_2^* \leq \dots \leq X_n^*.$$

- Каждому i -му члену (и соответственно X_i) этой последовательности ставится в соответствие число $\frac{i}{n+1} = p_i$ оценивающее отвечающую ему вероятность $F(x_i)$.

- Вводится предположение о некотором гипотетическом распределении и характеризуется функциональное соотношение между теоретическим квантилем $Q(p)$ гипотетического распределения в зависимости от вычисляемого (по p) эмпирического значения.

- Затем по функции теоретического квантиля $Q(p)$ рассчитываются соответствующие значения эмпирического квантиля

$$\hat{Q}_n(p_i) = Q_n \frac{i}{n+1}$$

как функции, задающей для данного значения p ($0 < p < 1$) наименьшее значение, слева от которого располагаются по крайней мере $100p$ процентов данных. Эта функция аппроксимирует соответствующую *теоретическую функцию квантилей* Q , определяемую как обратная функция от функции распределения.

- Строится график (в виде возрастающей функции) оценочных квантилей X_i^* в зависимости от соответствующих эмпирических квантилей $Q_n\left(\frac{i}{n+1}\right)$.

Оценочные значения квантилей будем отображать на вертикальной оси, эмпирические теоретические квантили — на горизонтальной.

- Если принятая гипотеза о модели распределения оправдывается, то точки $(X_i^*, \hat{Q}_n(p_i))$ будут лежать вблизи прямой.

6. Основные тестовые распределения для практических приложений

Среди множества вероятностных распределений есть такие, которые *наиболее часто встречаются* в силу действия вероятностных законов. Нормальное (и логнормальное) распределение употребляется, поскольку предсказывается центральной предельной теоремой, о чем говорилось выше. Дело в том, что интересующие нас *случайные величины могут часто рассматриваться как суммы большого числа независимых между собой слагаемых*, каждое из которых имеет лишь незначительные размеры по сравнению со всей суммой. Но в таком случае мы находимся как раз в условиях применимости центральной предельной теоремы и можем ожидать, что распределение данной величины мало отклоняется от нормальной формы. Статистические исследования и в самом деле констатировали приближенную *нормальность распределения в достаточно широком классе случаев*.

Заметим еще, что в некоторых случаях рассматриваемая величина X оказывается распределенной асимметрично, но некоторая функция от нее, например $\lg X$, X^2 , X^3 и т.д., приближенно следует нормальному закону. Такое функциональное преобразование часто оказывается полезным при изучении конкретных распределений. Кроме того, в статистической практике мы оперируем часто такими функциями от случайных величин (например, средней арифметической, медианой, моментами выборки), которые при большом объеме выборки оказываются, опять-таки приближенно, нормально распределенными. Этот факт имеет очень большое значение для статистической теории и практики.

Поэтому в соответствии с общими вероятностными закономерностями исторически так сложилось, что нормальное и логнормальное распределения занимают главный класс модельных вероятностных распределений, для которых применяется техника графиков квантилей. Экспоненциальное распределение играет не менее, если не более важную роль, в том числе для анализа экстремальных значений. Для экстремальных событий имеются свои характерные классы вероятностных распределений (Risk of extreme, 1992; Evaluating risks, 1994; Reiss, 2001). Поэтому в качестве тестовых законов распределений на практике часто рекомендуют использовать нормальный и логнормальный законы распределения, предсказывае-

мые центральной предельной теоремой (Феллер, 1984; Смирнов, Дуинин-Барковский, 1969), а также законы распределения экстремального типа (Гумбель, 1965; Лидбеттер, Линдгрэн, Ротсен, 1989) — Вейбулла, Гумбея, Фреше, Парето и экспоненциальное распределение. Более подробно распределения экстремального типа будут рассмотрены далее (см. также Акимов, Быков, Щетинин, 2009; Быков, 2014).

6.1. Экспоненциальное (показательное) распределение с параметром λ ($x > 0$)

Стандартный пример этого класса распределений записывается в виде:

$$1 - F_\lambda(x) = \exp(-\lambda x).$$

Функция квантиля для экспоненциального распределения имеет простой вид:

$$Q_\lambda(p) = -\frac{1}{\lambda} \ln(1 - p), \text{ для } p \in (0, 1).$$

Для нескольких или всех ($i = 1, 2, \dots, n$) значений $p_i \in (0, 1)$:

$$p = \frac{1}{n+1}, \frac{2}{n+1}, \dots, \frac{n}{n+1}$$

на квантиль-диаграмме наносятся точки с координатами

$$\left(-\ln \left(1 - \frac{i}{n+1} \right), x_i^* \right).$$

6.2. Распределение Вейбулла

Экспоненциальное распределение принадлежит к классу Вейбулловых распределений:

$$1 - F(x) = \exp(-\lambda x^\tau), x > 0.$$

Выражение для функции квантиля

$$Q(p) = \left(-\frac{1}{\lambda} \ln(1 - p) \right)^{1/\tau}, 0 < p < 1.$$

Для значений $p = \frac{i}{n+1}$, $i = 1, 2, \dots, n$ на квантиль-диаграмме наносятся точки с координатами

$$\left(\ln \left(-\ln \left(1 - \frac{i}{n+1} \right) \right), \ln(x_i^*) \right).$$

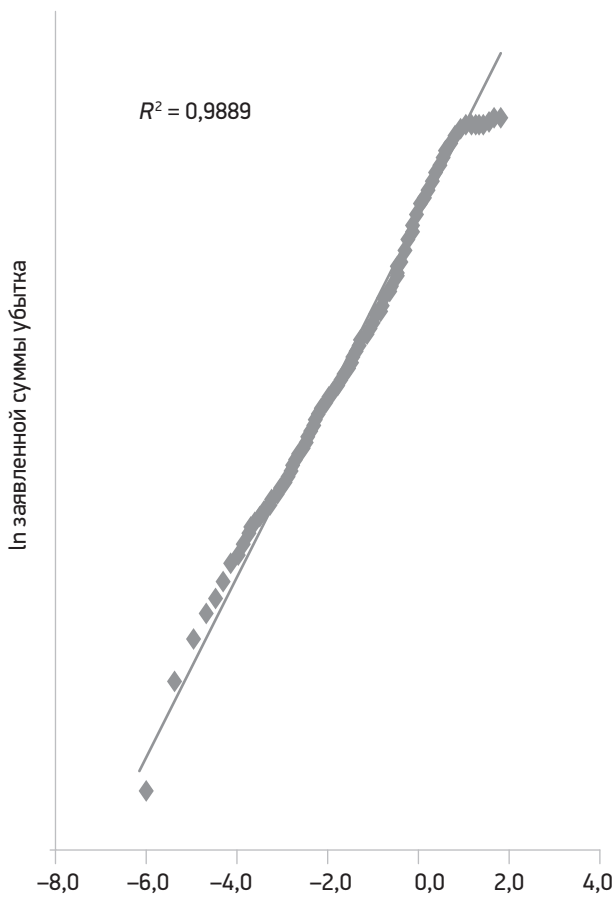


Рис. 4. Иллюстрация квантиль-диаграммы Вейбулла суммы заявленных убытков по иллюстративной программе страхования (Быков, 2014)

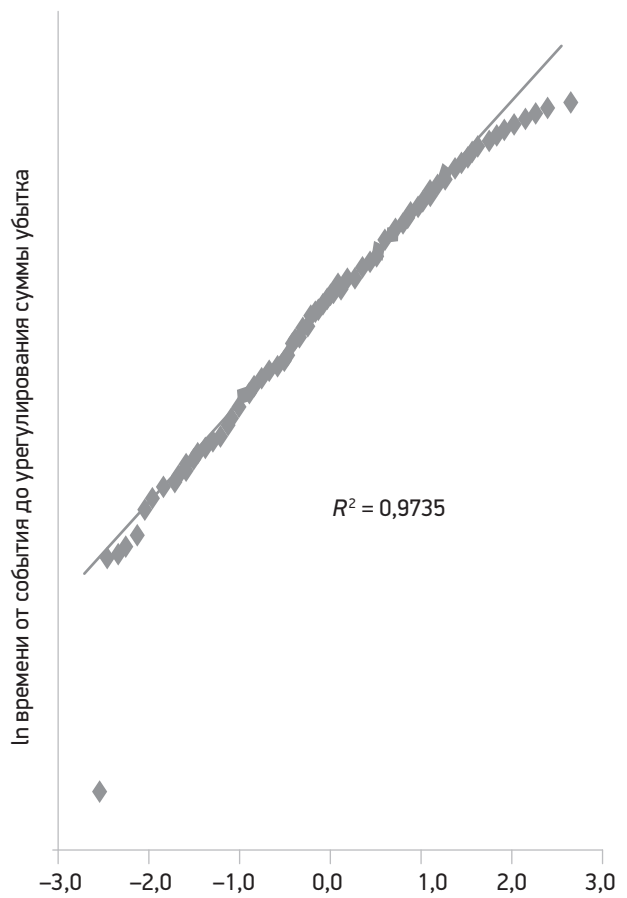


Рис. 5. Иллюстрация квантиль-диаграммы логнормального распределения времени урегулирования убытков по иллюстративной программе страхования (Быков, 2014)

6.3. Нормальное и логнормальное распределение

Нормальность распределения может быть проверена через общеизвестные линейные отношения между квантилями $Q(p)$ любого нормального распределения $N(\mu, \sigma^2)$ и соответствующими квантилями $\Phi^{-1}(p)$ стандартного нормального распределения в виде:

$$Q(p) = \mu + \sigma \Phi^{-1}(p).$$

Функция $\Phi^{-1}(p)$ содержится в любом полном статистическом справочнике или числовом пакете программ.

Для значений $p = \frac{i}{n+1}$, $i = 1, 2, \dots, n$ на квантиль-

диаграмме наносятся точки с координатами

$$\left(\Phi^{-1}\left(\frac{i}{n+1}\right), x_i^* \right).$$

Для проверки на соответствие логнормальному распределению строится график

$$\left(\Phi^{-1}\left[\frac{i}{n+1}\right], \ln(x_i^*) \right), i = 1, 2, \dots, n.$$

6.4. Распределение Парето

Распределение Парето с параметром $\alpha > 0$ (называется индексом Парето):

$$1 - F_Y(y) = y^{-\alpha} \text{ для } y > 1,$$

получается из экспоненциального (показательного) распределения с параметром α : $1 - F_X(x) = \exp(-\alpha x)$ посредством применения экспоненциального преобразования:

$$y = g(x) = \exp(x).$$

Квантиль-диаграммы в этом случае строятся с использованием следующей координатной плоскости:

$$\left(-\ln\left(1 - \frac{i}{n+1}\right), \ln(x_i^*) \right).$$

6.5. Распределение Гумбеля и распределение Фреше

Первый тип из максимум-устойчивых асимптотических распределений экстремальных значений — двойное экспоненциальное или так называемое *распределение Гумбеля* G_0 :

$$G_0(y) = \exp(-e^{-x}) \text{ для } x \in \mathbb{R}.$$

Квантиль-диаграммы строятся с использованием координатной плоскости:

$$\left(-\ln\left(-\ln\left(\frac{i}{n+1}\right)\right), x_i^* \right).$$

Второй тип из максимум-устойчивых асимптотических распределений экстремальных значений — так называемое *распределение Фреше*:

$$F(x) = \exp(-x^{-\alpha}) \text{ для } x, \alpha > 0.$$

Квантиль-диаграммы строятся с использованием координатной плоскости:

$$\left(-\ln\left(-\ln\left(\frac{i}{n+1}\right)\right), \ln(x_i^*) \right).$$

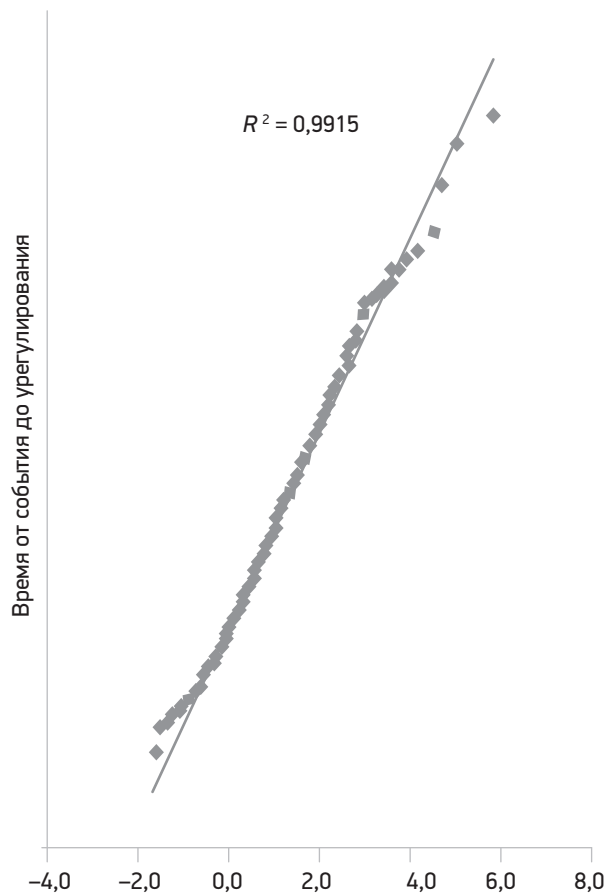


Рис. 6. Иллюстрация квантиль-диаграммы распределения Гумбеля времени урегулирования убытков по иллюстративной программе страхования имущества (Быков, 2014)

7. Главные результаты классической асимптотической теории экстремальных значений

Основная задача *статистической теории экстремальных значений* состоит в анализе наблюдаемых экстремумов и предсказании тех экстремумов, которые могут иметь место при последующих наблюдениях. Экстремумы не являются фиксированными величинами — это новые случайные величины, зависящие от исходного распределения и от объема выборки. Однако выявлены и некоторые их свойства, не зависящие от распределения. Основные результаты асимптотической теории вероятностей

экстремальных значений кратко представим в данном разделе.

Классическая асимптотическая теория вероятностей экстремальных значений в основном имеет дело с независимыми и одинаково распределенными случайными величинами и со свойствами распределения их максимума.

Пусть ξ_1, ξ_2, \dots — последовательность независимых и одинаково распределенных (н.о.р.) случайных величин (с.в.) и M_n — максимум первых n из этих величин, т.е.

$$M_n = \max(\xi_1, \xi_2, \dots, \xi_n). \quad (5.1)$$

Для н.о.р.с.в. функцию распределения (ф.р.) M_n можно определить в точном виде:

$$P\{M_n \leq x\} = P\{\xi_1 \leq x, \xi_2 \leq x, \dots, \xi_n \leq x\} = F^n(x), \quad (5.2)$$

где F обозначает общую ф.р. для ξ_i .

Большая часть классической теории экстремальных значений имеет дело с распределением M_n , особенно с его свойствами при $n \rightarrow \infty$. В частности, значительная часть «Статистики экстремальных значений», как озаглавлена книга (Гумбель, 1965), имеет дело с распределением M_n в целом ряде типовых случаев и с множеством родственных вопросов (например, относительно разных порядковых статистик, размаха и т.д.).

В классической аксиоматической теории вероятностей в центральной предельной теореме асимптотически нормальное распределение суммы многих н.о.р. случайных величин получают независимо от того, какова их исходная ф.р. (Феллер, 1984; Смирнов, Дунин-Барковский, 1969). Фактически, чтобы применять асимптотическую теорию, вовсе не обязательно знать эту ф.р. очень точно. Подобная ситуация имеет место и в теории экстремальных значений. Невырожденное асимптотическое распределение M_n (соответствующим образом нормализованного) обязательно должно принадлежать одному из трех единственно возможных общих семейств независимо от исходной ф.р. F . Кроме того, нет никакой необходимости знать ф.р. F полностью, чтобы определить, к какой предельной форме (если таковая существует) она приводит, т.е. к какой «области притяжения» она принадлежит. В действительности это определяется только поведением функции $F(x)$ для больших x , так что

об асимптотических свойствах максимума можно сказать многое, основываясь лишь на довольно ограниченной информации о свойствах ф.р. F .

Главные достижения классической теории, относящиеся к распределениям, можно условно свести к двум основным результатам (Лидбеттер, Линдгрэн, Ротсен, 1989).

7.1. Типы экстремальных распределений

Один из них — фундаментальный результат, называемый *теоремой об экстремальных типах* или теоремой о типах экстремальных распределений, — описывает возможные формы предельного распределения максимумов M_n при линейных нормализациях в последовательностях независимых и одинаково распределенных случайных величин. Точнее говоря, этот основной классический результат утверждает, что если для некоторых последовательностей нормализующих констант $a_n > 0$, b_n случайная величина $a_n(M_n - b_n)$ имеет невырожденную предельную функцию распределения $G(x)$:

$$P\{a_n(M_n - b_n) \leq x\} \xrightarrow{\omega} G(x) \quad (5.3)$$

(здесь имеется в виду сходимость в точках непрерывности ф.р. G , хотя в действительности все представляющие интерес ф.р. G непрерывны), то эта функция G должна иметь одну из трех единственно возможных форм.

Какие же именно ф.р. G могут встречаться в качестве такого предела? Возможные предельные распределения в классической теории отождествляются с классом распределений, обладающих определенным свойством устойчивости — так называемыми максимум-устойчивыми распределениями. Другими словами, все возможные невырожденные ф.р. G , которые могут встречаться в качестве пределов в (5.3), образуют в точности класс *максимум-устойчивых* распределений, и каждое максимум-устойчивое распределение G имеет (с точностью до преобразований сдвига и масштаба) одну из следующих трех параметрических форм, обычно называемых тремя *распределениями экстремальных значений*:

Тип I:

$$G(x) = \exp(-e^{-x}), \quad -\infty < x < \infty.$$

Тип II:

$$G(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & \text{для некоторого } \alpha > 0, x > 0. \end{cases}$$

Тип III:

$$G(x) = \begin{cases} \exp(-(-x)^\alpha), & \text{для некоторого } \alpha > 0, x \leq 0, \\ 1, & x > 0. \end{cases}$$

Именно эти три типа распределения экстремальных значений впервые были обнаружены Фреше (Fréchet, 1927), а центральный результат — теорема об экстремальных типах — впервые был получен Фишером и Типпетом (Fisher, Tippett, 1928) и позднее был доказан в полной общности Гнеденко (Gnedenko, 1943). Впоследствии доказательство этого результата для н.о.р. случайных величин, используя новый, более простой подход, было предложено де Ханом (Naan de, 1976) и затем было распространено на случаи зависимости.

Функцию распределения того же типа, что и $\exp(-e^{-x})$ (т.е. функцию, имеющую вид $\exp\{-e^{-(ax+b)}\}$ для некоторых $a > 0, b$), мы будем называть ф.р. типа I. Аналогично мы будем говорить, что ф.р. является ф.р. типа II (или типа III), если она имеет форму $G(ax+b)$, где G — ф.р. экстремальных значений типа II (типа III). Поскольку параметр a может изменяться, то распределения типа II и типа III в действительности являются семействами типов в рамках узкого понимания термина «тип». Одна-

ко использование обычной привычки ссылаться на три типа экстремальных значений, очевидно, не приведет к каким-либо недоразумениям.

Этот результат играет важную роль не только в случае н.о.р.с.в., но и при наличии зависимости, где он уже никоим образом не столь тривиален, но все еще сохраняет силу при надлежащих условиях и весьма важен для дальнейшего развития теории как для случая н.о.р. случайных величин, так и для зависимых последовательностей.

Распределения $G(x)$ можно выразить в единой параметрической форме, называемой обобщенным распределением экстремальных величин: $H_\xi(x)$ — функция распределения такая, что при всех x , удовлетворяющих соотношению $0 < H_\xi(x) < 1$, справедливо равенство

$$H_\xi(x) = \begin{cases} \exp\{-(1+\xi x)^{-1/\xi}\} & : \xi \neq 0 \\ \exp\{-e^{-x}\} & : \xi = 0 \end{cases}, \quad 1 + \xi x > 0, \quad (5.4)$$

где ξ — вещественное число.

Плотности распределения $H_\xi(x)$ при различных значениях параметра ξ изображены на рис. 7.

Параметр ξ называют параметром формы функции распределения H . Так как для любой случайной величины X с ф.р. F_X и констант $\mu \in R$ и $\sigma > 0$ функция распределения $\tilde{X} = \mu + \sigma X$ задана

$$F_{\tilde{X}}(x) = F_X\left(\frac{x - \mu}{\sigma}\right),$$

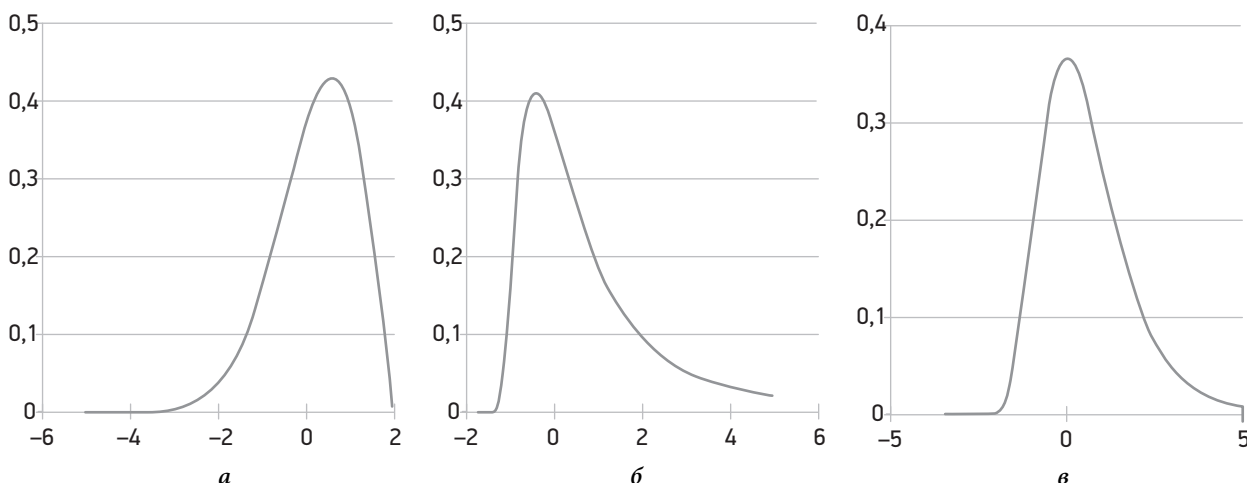


Рис. 7. Плотности обобщенного распределения экстремальных величин H_ξ . (а) с $\xi = -0,5$; (б) с $\xi = 0,5$; (в) с $\xi = 0$

мы можем добавить параметры локализации и масштаба к указанной выше параметризации и рассмотреть

$$H_{\xi, \mu, \sigma}(x) = H_{\xi}\left(\frac{x - \mu}{\sigma}\right).$$

Если теоремы о сходимости, приведенные в следующем подразделе, справедливы, то функция распределения F принадлежит области притяжения функции H , и используют обозначение $F \in MDA(H)$. Большинство распределений в статистике лежит в области притяжения функции H с некоторым значением параметра ξ .

Если $F \in MDA(H_{\xi})$ и $\xi = 0$ или $F \in MDA(H_{\xi})$ и $\xi < 0$, то говорят, что функция F имеет тонкий или короткий «хвост» соответственно. Тонкохвостые ($\xi = 0$) распределения включают в себя нормальное, экспоненциальное, гамма- и логнормальное распределения. Короткохвостые ($\xi < 0$) распределения имеют конечную правую границу и включают в себя равномерное и бета- распределения. Тяжелохвостые распределения, которые лежат в области притяжения распределения Фреше, $F \in MDA(H_{\xi})$ с $\xi > 0$ особенно интересны в области исследования экстремальных величин. В этом классе находятся такие распределения, как Стьюдента- t , α -устойчивые и Парето.

Заметим, что если случайная величина X имеет функцию распределения F , где $F \in MDA(H_{\xi})$, $\xi > 0$, тогда все моменты EX^{β} бесконечны для $\beta > 1/\xi$.

7.2. Сходимость к экстремальным распределениям

Другим центральным результатом, используемым при построении теории, является общая теорема Хинчина (см.: Лидбеттер, Линдгрэн, Ротсен, 1989) о сходимости функций распределения и теоремы о сходимости вероятностей $P\{M_n \leq u_n\}$.

Второй основной результат дает простое необходимое и достаточное условие, при котором для заданной последовательности постоянных $\{u_n\}$ сходится последовательность вероятностей $P\{M_n \leq u_n\}$. Иными словами, результат о сходимости указывает условия сходимости последовательности вероятностей $P\{M_n \leq u_n\}$, где $\{u_n\}$ — произвольная последовательность вещественных постоянных. (В случае когда выполняется (5.3), такая сходимость сохраняется для всех членов семейства последова-

тельств $\{u_n = x/a_n + b_n\}$, где x принимает все возможные вещественные значения.)

Важность его обнаруживается в классической теории областей притяжения трех типов экстремальных значений.

В силу (5.2) соотношение (5.3) можно записать в виде

$$F^n(a_n^{-1}x + b_n) \xrightarrow{\omega} G(x), \quad (5.5)$$

где символ $\xrightarrow{\omega}$ обозначает, как и выше, сходимость в точках непрерывности предельной функции. Если соотношение (5.5) выполняется для некоторых последовательностей $\{a_n > 0\}$ и $\{b_n\}$, то говорят, что ф.р. F принадлежит к области притяжения (для максимумов) закона G , и обозначают $F \in D(G)$.

Может случиться так, что для заданной ф.р. F вовсе не существует такой ф.р. G , что $F \in D(G)$. Это означает просто, что максимум M_n не имеет невырожденного предельного распределения ни при какой линейной нормализации (обычным примером этого является распределение Пуассона). С другой стороны, пределы $P\{M_n \leq u_n\}$ могут существовать и для других заслуживающих внимания последовательностей u_n , не обязательно имеющих вид $x/a_n + b_n$ и даже не зависящих от параметра x .

Выведены необходимые и достаточные условия, определяющие, какое именно из возможных предельных распределений реализуется (если таковое имеется), т.е. условия, при которых $F \in D(G)$. Эти условия содержатся, например, в работе (Loynes, 1965), где можно найти доказательства их достаточности.

Конечно, важно знать, какой именно (если хотя бы какой-то) из трех типов предельных законов применим, когда каждая с.в. ξ_n имеет заданную ф.р. F . Для каждого типа известны необходимые и достаточные условия, относящиеся к поведению «хвоста» $1 - F(x)$ при возрастании x . Сформулируем эти условия, опуская доказательства необходимости и/или достаточности, поскольку последние довольно длинны и не являются нашей основной целью. Соответствующие доказательства можно найти в работах (Лидбеттер, Линдгрэн, Ротсен, 1989; Gnedenko, 1943; Naan de, 1976).

Сначала приведем несколько *очень простых и полезных достаточных условий*, применимых

в случае, когда ф.р. F имеет плотность f . Эти условия принадлежат Мизесу, а их простые доказательства приведены в работе (Наан де, 1976). В последующих выкладках обозначим $x_F = \sup \{x; F(x) < 1\}$ для правой концевой точки произвольной ф.р. F . Иначе говоря, $F(x) < 1$ для всех $x < x_F$ и $F(x) = 1$ для всех $x \geq x_F$.

Теорема 7.2.1. *Предположим, что ф.р. F н.о.р. случайных величин, образующих последовательность $\{\xi_n\}$, абсолютно непрерывна и имеет плотность f . Тогда приводимые ниже условия являются достаточными для того, чтобы ф.р. F принадлежала соответствующей области притяжения:*

Тип I: f имеет отрицательную производную f' для всех x в некотором интервале (x_0, x_F)

$(x_F \leq \infty), f(x) = 0$ для $x \geq x_F$ и

$$\lim_{t \uparrow x_F} \frac{f'(t)(1 - F(t))}{f^2(t)} = -1.$$

Тип II: $f(x) > 0$ для всех конечных $x \geq x_0$ и

$$\lim_{t \rightarrow \infty} \frac{tf(t)}{1 - F(t)} = \alpha > 0.$$

Тип III: $f(x) > 0$ для всех x в некотором конечном интервале $(x_0, x_F), f(x) = 0$ для $x > x_F$ и

$$\lim_{t \uparrow x_F} \frac{(x_F - t)f(t)}{1 - F(t)} = \alpha > 0.$$

Теорема 7.2.2. *Необходимыми и достаточными условиями принадлежности ф.р. F случайных величин, образующих последовательность н.о.р.с.в. $\{\xi_n\}$, каждому из трех типов являются (в порядке сложности):*

Тип II: $x_F = \infty$ и

$$\lim_{t \rightarrow \infty} (1 - F(tx)) / (1 - F(t)) = x^{-\alpha},$$

$\alpha > 0$, для каждого $x > 0$.

Тип III: $x_F = \infty$ и

$$\lim_{h \downarrow 0} (1 - F(x - xh)) / (1 - F(x_F - xh)) = x^\alpha,$$

$\alpha > 0$, для каждого $x > 0$.

Тип I: существует такая строго положительная функция $g(t)$, что

$$\lim_{t \uparrow x_F} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x}$$

для всех вещественных x .

В действительности можно показать, что когда имеет место предел типа I, то

$$\int (1 - F(u)) du < \infty,$$

и в качестве g можно взять функцию, определяемую соотношением

$$g(t) = \int_t^{x_F} (1 - F(u)) du / (1 - F(t)) \text{ для } t < x_F.$$

Следствие 7.2.3. *Константы a_n, b_n в сходимости $P\{a_n(M_n - b_n) < x\} \rightarrow G(x)$ в каждом из указанных выше случаев могут быть выбраны следующим образом:*

Тип II: $a_n = y_n^{-1}, b_n = 0$.

Тип III: $a_n = (x_F - y_n), b_n = x_F$.

Тип I: $a_n = [g(y_n)]^{-1}, b_n = y_n$

где $y_n = F^{-1}(1 - 1/n) = \inf \{x; F(x) \geq 1 - 1/n\}$.

Отметим, что приведенные выше критерии применимы к любой ф.р. в каждой области притяжения независимо от того, является ли предел конкретным представляющим тип распределением $G(x)$, перечисленным в теореме об экстремальных типах, или любой другой ф.р. $G(ax + b)$ этого типа. Действительно, если предел имеет вид $G(ax + b)$, то и $G(x)$ также будет пределом при соответствующем изменении нормализующих констант, так как если

$$P\{a_n(M_n - b_n) \leq x\} \rightarrow G(ax + b),$$

то ясно, что

$$P\{\alpha_n(M_n - \beta_n) \leq x\} \rightarrow G(x),$$

с $\alpha_n = aa_n, \beta_n = b_n - b / (aa_n)$.

8. Моделирование надпороговых значений: полупараметрический подход

Современный подход к моделированию и анализу показателей с высокой волатильностью предполагает описание не только их наибольших (максимальных) значений, но и значений, превосходящих некоторый заранее известный верхний порог. Такой подход называют методом надпороговых значений (*Peaks Over Threshold (POT)*).

Первый, полупараметрический подход, основан на оценке экстремального индекса Хилла (Hill, 1975), работах Даниэльсона и де Вриза (Danielsson, de Vries, 1997; Danielsson et al, 2001).

Второй подход полностью параметрический, основан на аппроксимации функции распределения

надпороговых значений с помощью обобщенного распределения Парето (см., например: Акимов, Быков, Щетинин, 2009; McNeil, Saladin, 1997; Embrechts, Resnick, Samorodnitsky, 1999).

В данном разделе рассмотрим первый метод несколько более подробно.

Предположим, что F_X — функция распределения наблюдений за некоторый период времени, где мы хотели бы оценить квантиль высокого порядка. Пусть распределение больших значений X имеет вид:

$$P(X > x) = cx^{-\alpha}, \alpha > 0, x > x_0. \quad (8.1)$$

Полупараметрический подход использует оценку Хилла для параметра α вместе с порядковыми статистиками эмпирических данных, которая затем используется для вычисления квантиля.

Пусть $X^{(1)} \geq X^{(2)} \geq \dots \geq X^{(n)}$ — порядковая статистика выборки эмпирических данных объема n , в предположении, что они независимы и одинаково распределены с функцией распределения F_X .

Если X является случайной величиной типа Парето на «хвосте» и $X^{(k+1)}$ — наибольшая порядковая статистика, то для $x > X^{(k+1)}$ справедливо

$$\frac{\bar{F}_X(x)}{\bar{F}_X(X^{(k+1)})} = \left(\frac{x}{X^{(k+1)}}\right)^{-\alpha}.$$

Значение эмпирической функции распределения $\bar{F}_X(X^{(k+1)}) = k/n$ предполагает следующую оценку F_X на верхнем «хвосте»:

$$\bar{F}_X(x) = 1 - \frac{k}{n} \left(\frac{x}{X^{(k+1)}}\right)^{-\tilde{\alpha}}, \text{ для } x > X^{(k+1)}.$$

Обращая это отношение для заданного $q = \bar{F}_X(x)$, получим выражение для квантиля

$$x_q(X) = X^{(k+1)} \left(\frac{n}{k}(1-q)\right)^{-1/\tilde{\alpha}}. \quad (8.2)$$

Значение q должно быть большим, а именно $q = \bar{F}_X(x) > \bar{F}_X(X^{(k+1)}) = 1 - k/n$.

Мы получили оценку для квантиля, но она зависит от k через $X^{(k+1)}$, размера выборки n и параметра $\tilde{\alpha}$. Чтобы вычислить α , Хилл [16] предложил следующую ее оценку $\tilde{\alpha}^{(Hill)}$ в виде

$$\tilde{\alpha}_{k,n}^{(Hill)} = \left(\frac{1}{k} \sum_{i=1}^k (\ln X^{(i)} - \ln X^{(k+1)})\right)^{-1}. \quad (8.3)$$

В случае независимых и одинаково распределенных случайных величин, а также в случае стационарных случайных последовательностей в работах Даниэльсона, Пенга и де Вриза (Danielsson, de Vries, 1997; Danielsson et. al, 2001) доказано, что оценка

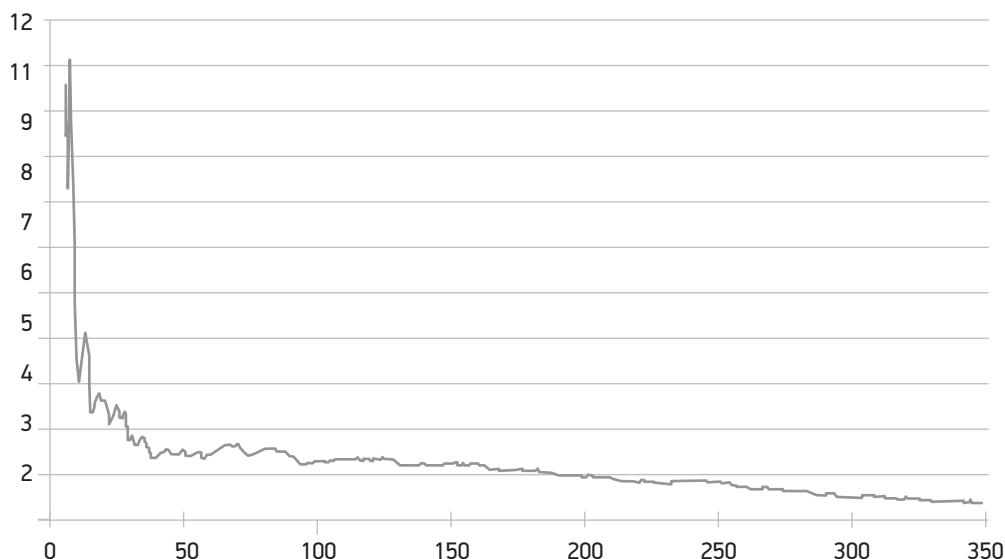


Рис. 8. График Хилла $\left\{ \left(k, \tilde{\alpha}_{k,n}^{(Hill)} \right) : k = 1, \dots, n-1 \right\}$

$\tilde{\alpha}^{(Hill)}$ обладает свойствами непрерывности и асимптотической нормальности.

Чтобы построить оценку $x_q(X)$, мы должны выбрать уровень порога $X^{(k+1)}$ или, что то же самое, значение k . Традиционно выбор k производят визуально, используя для этого график оценки Хилла

$\left\{ \left(k, \tilde{\alpha}_{k,n}^{(Hill)} \right) : k = 1, \dots, n-1 \right\}$. Выбор k осуществляется

внутри отрезка стабильности величины (8.3). Величина $\tilde{\alpha}_{n-1,n}^{(Hill)}$ вычисляется обычно как оценка максимального правдоподобия для α . Пример графика оценки Хилла изображен на рис. 8.

Отметим, что в работе С. Резника и К. Старицы (Resnick, Starica, 1997) предложен альтернативный, графический метод, названный *AltHill*-график, обладающий определенным преимуществом, поскольку он вытянут по правую сторону, что соответствует малым значениям k , это чаще упрощает выбор k .

9. Моделирование надпороговых значений: параметрический подход

Напомним, что теоретическая проблема моделирования экстремальных величин состоит в необходимости построения математической модели их надпороговых значений, когда нам известна некоторая цензурированная выборка $(X_1, \dots, X_n), X_i > u, i = 1, \dots, n$.

Будем рассматривать экстремальные величины, превышающие некоторое значение $u > 0$.

Значения $X - u$ будем в дальнейшем называть эксцессами.

Определение 9.1. Пусть случайная величина X имеет функцию распределения F с верхней границей $\omega(F)$. Для любого порога $u < \omega(F)$ определим функцию распределения эксцессов как

$$F_u(x) = P(X - u \leq x | X > u) \text{ для } 0 \leq X < \omega(F) - u, \quad (9.1)$$

следовательно, функция среднего значения эксцессов случайной величины X

$$e_x(u) = E(X - u | X > u). \quad (9.2)$$

Заметим, что для $0 \leq x < x_F - u$ мы можем выразить $F_u(x)$ через F :

$$F_u(x) = \frac{F(u+x) - F(u)}{1 - F(u)},$$

а функция среднего значения эксцессов $e_x(u)$ может быть выражена через функцию распределения эксцессов следующим образом:

$$e_x(u) = \int_0^{x_F - u} x dF_u(x).$$

Определение 9.2. Обобщенное распределение Парето:

$$G_{\xi,\beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta} \right)^{-1/\xi} & : \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta} \right) & : \xi = 0 \end{cases}, \quad (9.3)$$

где $\beta > 0, x \geq 0$, когда $\xi \geq 0$ и $0 \leq x \leq -\beta / \xi$, когда $\xi < 0$.

Иногда в определении $G_{\xi,\beta}(x)$ необходимо ввести параметр локализации μ , тогда мы будем рассматривать обобщенное распределение Парето в виде $G_{\xi,\beta,\mu}(x) = G_{\xi,\beta}(x - \mu)$. Распределение является тяжелохвостым в случае $\xi > 0$. Распределения (9.3) при $\beta = 1$ изображены на рис. 9.

Справедлива следующая теорема А. Балкема, де Хаана (Galambos, Kotz, 1978).

Теорема 9.3. Пусть случайная величина X имеет функцию распределения F . Тогда для любого $\xi \in R$ функция распределения F принадлежит области притяжения экстремальных величин $X \in MDA(H_\xi)$, если и только если

$$\limsup_{u \uparrow x_F} \sup_{0 < x < x_F - u} \left| F_u(x) - G_{\xi,\beta(u)}(x) \right| = 0$$

для некоторой положительной функции $\beta(u)$, где $G_{\xi,\beta(u)}(x)$ — функция обобщенного распределения Парето (9.3).

Для моделирования надпороговых экстремальных значений финансовых показателей воспользуемся результатами и выводами, вытекающими из теоремы 9.3.

Она утверждает, что функция распределения эксцессов F_u может быть аппроксимирована обобщенным распределением Парето при значениях порога u , близких к правой границе функции распределения F .

Чтобы продемонстрировать, как это может быть использовано, заметим, что вместо выражения (9.1) мы можем записать

$$\bar{F}(x) = \bar{F}(u) \bar{F}_u(x - u) \quad (9.4)$$

для $x > u$.

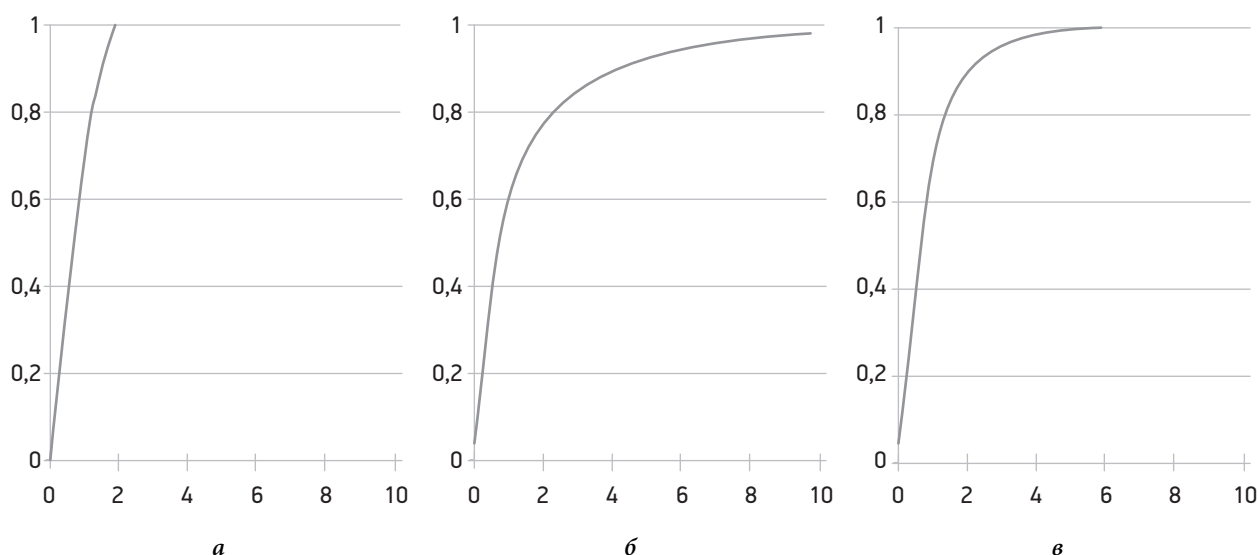


Рис. 9. Функция распределения $G_{\xi,\beta}$ со значением параметра $\beta = 1$. (а) $\xi = -0,5$; (б) $\xi = 0,5$; (в) $\xi = 0$, которые соответствуют распределению Парето с $\alpha = 2$

Полагая, что значение u достаточно велико, мы можем аппроксимировать F_u функцией $G_{\xi,\beta(u)}$ и использовать эмпирическую оценку для $\tilde{F}(u)$:

$$\tilde{F}(u) = \frac{N_u}{n}, \text{ где } N_u = \sum_{i=1}^n 1_{\{X_i > u\}},$$

где n — общее количество наблюдений. Верхний «хвост» $F(x)$ может быть оценен как

$$\tilde{F}(x) = 1 - \tilde{F} = 1 - \frac{N_u}{n} \left(1 + \xi \frac{x-u}{\beta} \right)^{-1/\xi} \quad (9.5)$$

для всех $x > u$.

Это позволяет экстраполировать условную функцию распределения эксцессов за пределы присутствующих в выборке данных, что невозможно в случае использования эмпирической оценки $F(x)$ для $x > u$.

Параметры ξ и β функции распределения $G_{\xi,\beta(u)}$ могут быть получены, например, методом максимального правдоподобия при фиксированном пороге u . Члены выборки для метода максимального правдоподобия: $X_{i_1} - u, \dots, X_{i_k} - u$, где X_{i_1}, \dots, X_{i_k} — наблюдения, превышающие порог u .

Для выбора значения порога u часто используют графический метод, используя график средних эксцессов $(u, \tilde{e}_X(u))$. Графический метод анализа

средних эксцессов опирается на следующую теорему (Drees, De Haan, Resnick, 2000).

Теорема 9.4. Предположим, что случайная величина X имеет обобщенное распределение Парето с параметрами $\xi < 1$ и β . Тогда для $u < x_F$

$$e_X(u) = \frac{\beta + \xi u}{1 - \xi}, \beta + \xi u > 0.$$

Ограничение $\xi < 1$ связано с тем, что тяжелохвостое распределение должно хотя бы иметь конечное среднее.

Если значение порога u достаточно велико, так, что F_u приблизительно совпадает с $G_{\xi,\beta}$, то по теореме 9.4 график $(u, e_X(u))$ линейен.

График средних эксцессов используется как графический метод проверки соотношения между возможными значениями порога u и функцией среднего значения эксцессов. Он состоит в анализе $e_X(u)$ и выборе u на участке линейности $e_X(u)$.

Обычно вместо $e_X(u)$ строят выборочную оценку:

$$\tilde{e}_X(u) = \frac{\sum_{i=1}^n (X_i - u)^+}{\sum_{i=1}^n 1_{\{X_i > u\}}}$$

в зависимости от порога u (рис. 10).

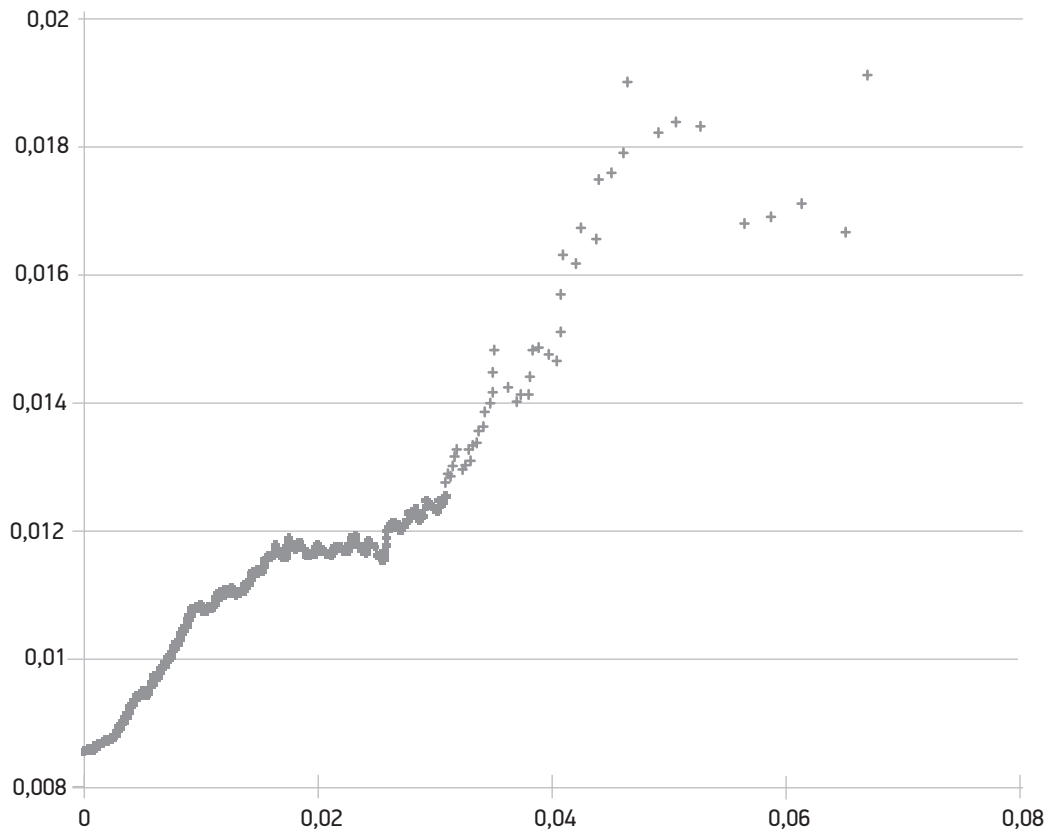


Рис. 10. График выборочных средних эксцессов ($u, \hat{e}_x(u)$)

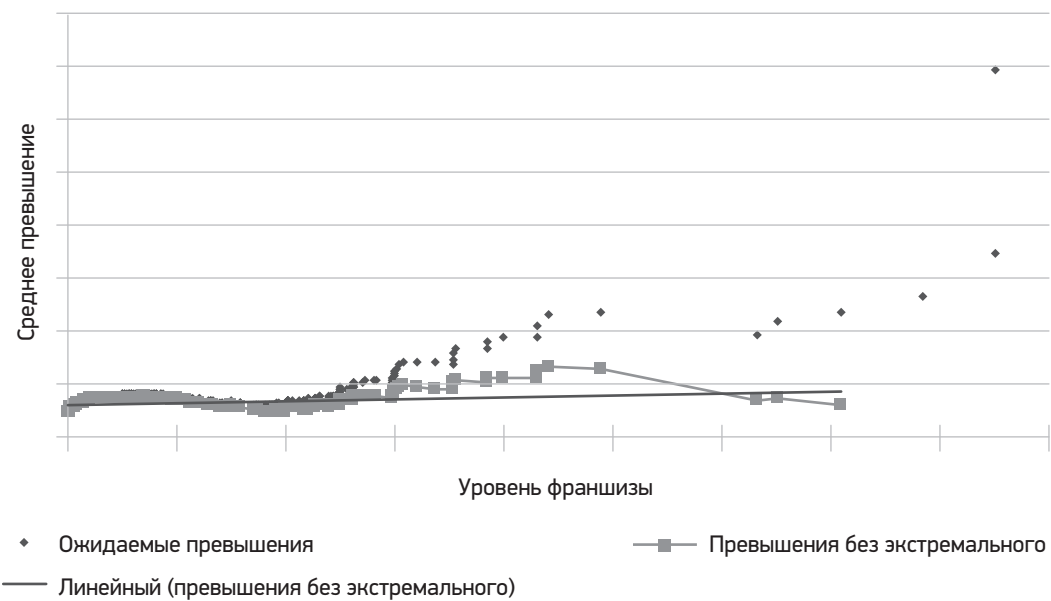


Рис. 11. Ожидаемые превышения определенного уровня (франшизы) заявленных убытков по иллюстративной программе страхования имущества (Быков, 2014)

На возможность в ряде актуарных приложений использования данного метода указывает также зависимость ожидаемых превышений определенного уровня (франшизы) заявленных убытков (рис. 11). В случае экспоненциального распределения данные должны группироваться вокруг горизонтальной прямой.

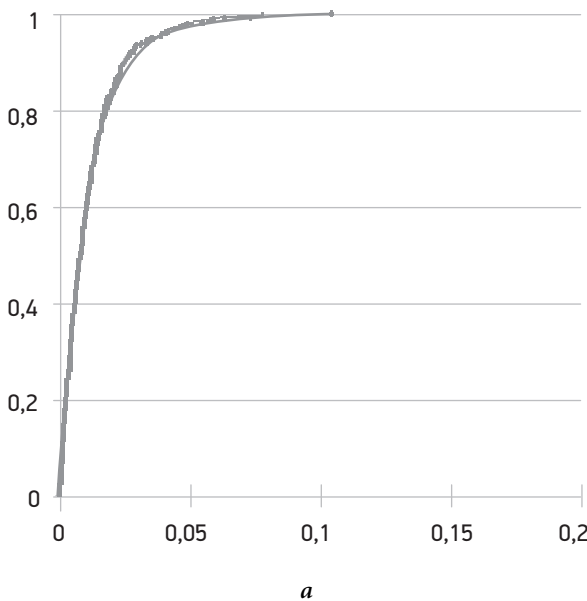
После использования графика средних эксцессов для выбора значения порога мы можем получить оценку для «хвоста» распределения, используя (9.5).

Чтобы вычислить величину квантиля $x_q(X)$ при условии $x_q(X) > u$, мы можем просто обратить выражение (9.4) для оценки «хвоста» и получить (рис. 12):

$$x_q(X) = u + \frac{\tilde{\beta}}{\xi} \left(\left(\frac{n}{N_u} (1-q) \right)^{-\xi} - 1 \right). \quad (9.6)$$

Найдем оценку для ожидаемого превышения, используя модель (9.3). Используя определение ожидаемого превышения S_q и функции среднего значения эксцессов, получим:

$$S_q(X) = x_q(X) + e_X(x_q(X)).$$



Так как функция распределения эксцессов F_u может быть аппроксимирована моделью $G_{\xi, \beta(u)}$ с $\xi < 1$, то по теореме 9.4 получим значение $x_q(X) > u$:

$$S_q(X) = x_q(X) + \frac{\beta + \xi(x_q(X) - u)}{1 - \xi} = \frac{\beta + x_q(X) - \xi u}{1 - \xi}$$

и следующую оценку для ожидаемого превышения порогового значения u

$$\hat{S}_q(X) = \frac{\tilde{x}_q}{1 - \xi} + \frac{\tilde{\beta} - \tilde{\xi}u}{1 - \xi}. \quad (9.7)$$

Заключение

Таким образом, в работе рассматриваются основные теоретические аспекты асимптотической теории вероятностей экстремальных событий, знание которых позволит риск-менеджерам на практике эффективно решать задачи в области количественных оценок и прогнозирования рисков.

1. Демонстрируются преимущества графических статистических методов анализа и обработки данных по сравнению с классическими подходами, основанными, в частности, на использовании критерия Хи-квадрат и обработке данных с применением

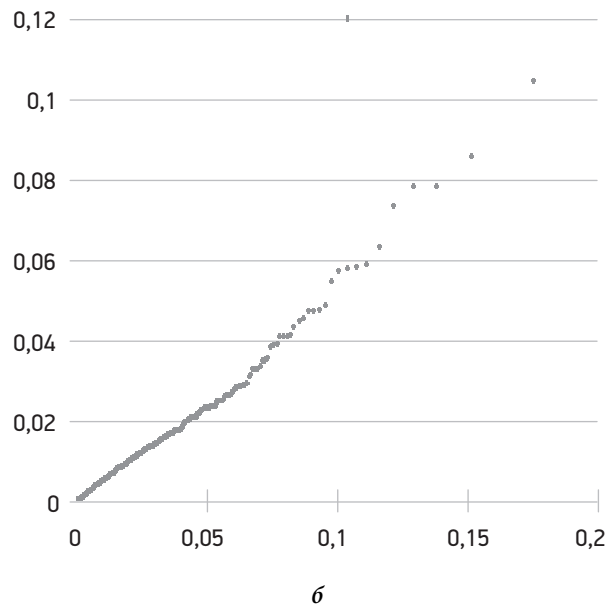


Рис. 12. (а) аппроксимация (гладкая кривая) эмпирических данных (точки); (б) QQ-график выборочных квантилей против квантилей, приближенных с помощью $G_{\xi, \tilde{\beta}}$ распределения

технологии построения полигона частот и гистограмм распределений. Для одних и тех же наблюдений, одинаковых теоретических предпосылок, одних и тех же параметров и даже одной и той же длины интервалов только из-за сдвига начальной точки деления можно получить различные значения χ^2 .

Путь, свободный от этой неоднозначности, состоит в замене интервалов равной длины интервалами равной вероятности. Из этого подхода вытекают методы сравнения, основанные на сопоставлении каждого индивидуального наблюдения с соответствующей теоретической величиной. Идея построения графиков квантилей основана на том факте, что для основных классов распределений квантили $Q(p)$ линейно связаны с соответствующими стандартными квантилями из этого класса распределений. При этом линейность на графике может быть легко замечена визуально и/или определена количественно посредством построения линии регрессии. Поэтому рассмотренные специальные статистические графические средства эффективно применяются для обеспечения правдоподобного соответствия выбранной модели распределения имеющемуся распределению случайной переменной.

2. Среди множества вероятностных распределений есть такие, которые *наиболее часто встречаются* в силу действия вероятностных законов. В соответствии с общими вероятностными закономерностями исторически так сложилось, что нормальное и логнормальное распределения занимают главный класс модельных вероятностных распределений, для которых применяется техника графиков квантилей. Для экстремальных событий имеются свои характерные классы вероятностных распределений. Поэтому в качестве тестовых законов распределений на практике рекомендуется использовать нормальный и логнормальный законы распределения, предсказываемые центральной предельной теоремой, а также законы распределения экстремального типа — Вейбулла, Гумбеля, Фреше, Парето и экспоненциальное распределение.

3. Основная задача статистической теории экстремальных значений состоит в анализе наблюдаемых экстремумов и предсказании экстремальных значений, которые могут иметь место при последующих наблюдениях. Экстремумы не являются фик-

сированными величинами — это новые случайные величины, зависящие от исходного распределения и от объема выборки. Однако выявлены и некоторые их свойства, не зависящие от распределения.

3.1. Основные классические результаты асимптотической теории вероятностей экстремальных значений кратко представлены в работе. Один из них — фундаментальный результат, называемый *теоремой об экстремальных типах*, или теоремой о типах экстремальных распределений, — описывает три возможные формы предельного распределения максимумов при линейных нормализациях в последовательностях независимых и одинаково распределенных случайных величин. Три возможных типа распределения экстремальных значений можно выразить в единой параметрической форме, называемой обобщенным распределением экстремальных величин.

3.2. Другим центральным результатом является общая теорема о сходимости функций распределения и теоремы о сходимости вероятностей $P\{M_n \leq u_n\}$. Второй основной результат дает простое необходимое и достаточное условие, при котором для заданной последовательности постоянных $\{u_n\}$ сходится последовательность вероятностей $P\{M_n \leq u_n\}$. Важность его обнаруживается в классической теории областей притяжения трех типов экстремальных значений, поскольку важно знать, какой именно (если хотя бы какой-то) из трех типов предельных законов применим, когда каждая с.в. ξ_n имеет заданную ф.р. F . Выведены необходимые и достаточные условия, определяющие, какое именно из возможных предельных распределений реализуется (если таковое имеется), т.е. для каждого типа известны необходимые и достаточные условия, относящиеся к поведению «хвоста» $1 - F(x)$ при возрастании x .

4. Таким образом, классические результаты асимптотической теории вероятностей экстремальных значений позволяют обоснованно предсказывать асимптотическое поведение «хвостов» распределений на основе статистической обработки имеющихся данных. Не обладая представительным статистическим рядом данных об экстремальных убытках, тем не менее можно утверждать:

1) что «хвосты» распределения гипотетически могут подчиняться одному из трех асимптотических типов распределений;

2) обладая знаниями о характере закона распределения F «вокруг средних значений», предполагая, что F абсолютно непрерывна и имеет плотность f , и используя теоремы о сходимости, можно определить тип экстремального распределения и его параметры.

5. Еще один практически важный аспект связан с возможностью предсказания значений, превосходящих некоторый заранее известный верхний порог. Кратко представлены основные подходы к моделированию надпороговых значений, условно разделенные на полупараметрический и параметрический подходы. Второй подход основан на аппроксимации функции распределения надпороговых значений с помощью обобщенного распределения Парето, поскольку согласно доказанным теоремам функция распределения эксцессов F_u может быть аппроксимирована обобщенным распределением Парето при значениях порога u , близких к правой границе функции распределения F_u . Параметры распределения могут быть определены с использованием графика средних эксцессов $e_X(u)$, поскольку если значение порога u достаточно велико, так, что F_u приблизительно совпадает с обобщенным распределением Парето $G_{\xi, \beta}$, то график $(u, e_X(u))$ линеен. Определив параметры обобщенного распределения Парето, по приведенным в работе формулам можно оценить ожидаемое значение превышения порогового значения, например, потерь (другими словами, предсказывать величину «ожидаемых непредвиденных» потерь).

Литература

- Акимов В.А., Быков А.А., Щетинин Е.Ю. Введение в статистику экстремальных значений и ее приложения. М.: ФГУ ВНИИ ГОЧС (ФЦ), 2009. 524 с.
- Быков А.А. Статистический анализ урегулирования убытков по программам имущественного страхования: рекомендации для страхователей и риск-менеджеров крупных компаний. М.: Газпром ВНИИГАЗ, 2014. 242 с.
- Быков А.А. Статистические методы прогнозирования риска чрезвычайных ситуаций / Под ред. чл.-корр. РАН Б.Н. Порфирьева. М.: Анкил, 2014. 156 с.
- Вуколов Э.А., Ефимов А.В., Земсков В.Н. Сборник задач по математике для втузов. Специальные курсы. М.: Наука. Главная редакция физико-математической литературы, 1984 г. 608 с.
- Гумбель Э. Статистика экстремальных значений. М.: Мир, 1965.
- Деклерк Д., Шибяев В. Расчет экономического капитала как показатель зрелости системы управления рисками // Управление рисками № 1 (59). 2008. С. 21—24.
- Иванов В.И. Статистическое обоснование отсутствия порога действия канцерогенных факторов на популяционном уровне // Проблемы анализа риска. 2005. Т. 2. № 3. С. 276—278.
- Лидбеттер М., Линдгрэн Г., Ротсен Х. Экстремумы случайных последовательностей и процессов. М.: Мир, 1989. 392 с.
- Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. М.: Наука. 1969. 512 с.
- Феллер В. Введение в теорию вероятностей и ее приложения. М.: Мир, 1984. 738 с.
- Beirlant J., Taugels J.L., Vynckier P. Practical Analysis of Extreme Values. Leuven Universitu Press, Belgium, 1996.
- Danielsson J., C. de Vries, Beyond the Sample: Extreme quantile and probability estimation. Preprint, LSE, 1997.
- Danielsson J., L. De Haan, L. Peng, C. G. De Vries, Using a bootstrap method to choose the sample fraction in the tail index estimation, Journal of Multivariate Analysis, 76. 2001. P. 226—248 .
- Drees H., L. De Haan, S. Resnick, How to make a Hill plot, The Annals of Statistics. 2000. 28(1). P. 254—274.
- Embrechts P., S.I. Resnick, Samorodnitsky, Extreme value theory as a risk management tool, North American Actuarial Journal, 1999. 3. P. 30—41.
- Evaluating risks of extreme events for univariate loss functions. Journal of Water Resorces, 1994, 120(3). 382—399.
- Fréchet M. Sur la loi de probabilité de l'écart maximum. Ann. Soc. Math. Polon., 1927, 6. 93—116.
- Fisher R.A., Tippett L.H.C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Proc. Camb. Phil. Soc., 1928, 24. 180—190.
- Galambos J., S. Kotz, Characterizations of probability distributions. Lecture notes in Mathematics. New York: Springer-Verlag, 1978.
- Gnedenko B.V. Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math., 1943, 44, 423—453.
- Haan de L. Sample extremes: an elementary introduction. Statist. Neerlandica, 1976, 30. 161—172.
- Hill B.M. A simple general approach to inference about the tail of a distribution. Annals of Statistics, 1975. 3(5): 1163—1174.

- Lindeberg J.W. Math. Zeit. № 15 (1922). P. 211—235.
- Loynes R.M. Extreme values in uniformly mixing stationary stochastic processes. Ann. Math. Statist. 1965. 36. 993—999.
- McNeil J.A., Saladin T. The peaks over threshold method for estimating high quantiles of loss distributions, In Proceedings of the 28th International ASTIN Colloquium, 1997.
- Reiss R.D., Thomas M. Statistical analysis of extreme values with applications to insurance, finance and other fields. Springer, 2001.
- Resnick S., Starica C. Smoothing the Hill estimator, Advances in Applied Probability. 1997. 29.
- Risk of extreme events in a multiobjective framework. Water Resource Bulletin, 1992, 28(1), 201—209.

Сведения об авторе

Быков Андрей Александрович: доктор физико-математических наук, профессор, заслуженный деятель науки Российской Федерации. Вице-президент Российского научного общества анализа риска. Главный редактор журнала «Проблемы анализа риска», начальник отдела Аппарата Правления ПАО «Газпром»

Область научных интересов: анализ и управление риском, прикладные области актуарной математики и теории риска, теории вероятностей и статистики экстремальных значений, математической экономики и экологии, теории безопасности человека и окружающей среды

Публикации: автор более 300 научных трудов, из них более 200 опубликованных, в их числе 19 монографий, 2 учебных пособия, нормативные и методические документы

Контактная информация:

Адрес: 125167, г. Москва, ул. Восьмого Марта 4-я, д. 6а

Тел.: +7 (495) 787-52-26

E-mail: journal@dex.ru