

УДК 351.861

А.А. Быков

Приложения асимптотической теории вероятностей экстремальных значений к прогнозированию риска экстремальных чрезвычайных ситуаций

Аннотация

В работе продемонстрированы некоторые простые возможные приложения асимптотической теории вероятностей экстремальных значений для прогнозирования риска экстремальных ЧС. Приводятся основные классические результаты асимптотической теории экстремальных значений, демонстрируются методы обработки статистических данных о ЧС с использованием графических способов построения квантиль-диаграмм. Приведены примеры построения графиков квантилей. Рассмотрены вопросы приложения классических результатов асимптотической теории экстремальных значений к анализу многолетних данных по ЧС, получения законов распределения параметров ЧС (на примерах максимальных длительностей, максимальных площадей и количества пожаров). Рассмотрены вопросы графического анализа периодов повторяемости для целей прогнозирования. Показано как, зная законы распределения количества пожаров и затрат на ликвидацию пожаров и применяя классическую технику актуарных расчетов, спрогнозировать (или запланировать, например, на будущий год) размер необходимых затрат на ликвидацию пожаров так, чтобы его было достаточно для ликвидации пожаров с гарантией 95 %. В заключение рассмотрен вопрос о том, можно ли, используя графическую технику предложить упрощенный подход для введения статистического критерия экстремальной ЧС?

Ключевые слова: риск; прогнозирование риска; экстремальная чрезвычайная ситуация; статистика экстремальных значений.

Содержание

Введение

1. Главные результаты классической асимптотической теории экстремальных значений
2. Специальные графические статистические средства: графики квантилей
3. Статистическая обработка данных многолетних наблюдений
4. На какие другие вопросы можно получить ответы с использованием приведенной технологии

Заключение

Литература

Введение

Экстремальные чрезвычайные ситуации (ЧС) характеризуются большим размером последствий для населения и окружающей среды. К таким ЧС согласно Классификации ЧС природного и техногенного характера (Постановление Правительства Российской Федерации от 21.05.2007 г. № 304) можно отнести ЧС регионального или межрегионального характера с размером материального ущерба 50—500 млн рублей или ЧС федерального характера с размером материального ущерба свыше 500 млн рублей. Однако, вероятность наступления таких ЧС очень мала, из-за чего возникают теоретические проблемы прогнозирования риска экстремальных ЧС.

Проблеме оценки риска экстремальных ЧС, тяжелых аварий, природных катастроф долгое время не придавалось большого значения, и она оставалась практически не исследованной до конца 80-х годов. Ряд крупных аварий и катастроф, произошедших в последние десятилетия, привел к изменению отношения к проблеме исследования риска экстремальных ЧС как природного, так и техногенного характера с большими значениями ущерба. Появились работы, указывающие на возможность существенно более высоких оценок вероятности наступления экстремальных стихийных явлений и техногенных катастроф, по сравнению с ранее прогнозируемыми значениями. Одновременно, отчетливо обозначилась ограниченность традиционной методологии и методов оценки риска ЧС при их ис-

пользовании для случая крупных техногенных аварий и катастрофических природных явлений.

Возникает задача поиска специальных прикладных методов определения и прогнозирования риска экстремальных ЧС с использованием достижений асимптотической теории вероятностей экстремальных значений, которая за последние годы достигла существенного прогресса в теории, оставаясь при этом слабо адаптированной для практических приложений.

Обычно в прикладных задачах статистику ЧС представляют в частотном виде, изучая динамические закономерности частот, и меньшее внимание уделяется вероятностным распределениям произошедших событий и анализу закономерностей им присущих. Но именно такой прикладной статистический анализ вероятностного распределения реализации чрезвычайных ситуаций позволит точнее оценить риск экстремальных ЧС на основе теорем о сходимости к предельным распределениям, а исследование многолетних данных позволяет более точно осуществлять прогнозирование риска таких ЧС.

1. Главные результаты классической асимптотической теории экстремальных значений

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.

$$M_n = \max(\xi_1, \xi_2, \dots, \xi_n). \quad (1)$$

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

$$P\{M_n \leq x\} = P\{\xi_1 \leq x, \xi_2 \leq x, \dots, \xi_n \leq x\} = F^n(x). \quad (2)$$

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 Гумбеля [1]) 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 обычной 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. 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 [2].

Один из них — фундаментальный результат, называемый теоремой об экстремальных типах или теорема 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 $G(x)$

$$P\{a_n(M_n - b_n) \leq x\} \xrightarrow{w} G(x). \quad (3)$$

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

Тип II: $G(x) = \begin{cases} 0, & x \leq 0, \\ \exp(-x^{-\alpha}), & \text{при } \alpha > 0, x > 0, \end{cases}$

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

Именно эти три типа «распределения экстремальных значений» впервые были обнаружены Фреше [3], а центральный результат — теорема об

экстремальных типах — впервые был получен Фишером и Типпетом [4] и позднее был доказан в полной общности Гнеденко [5]. В последствии доказательство этого результата для н.о.р. случайных величин, используя новый, более простой подход, было предложено де Ханом [6], и затем было распространено на случаи зависимости.

Другим центральным результатом, используемым при построении теории, является общая теорема Хинчина [2] 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 (1), такая сходимость сохраняется для всех членов семейства последовательностей $\{u_n = x/a_n + b_n\}$, где x принимает все возможные вещественные значения).

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

В силу (2) соотношение (3) можно записать в виде:

$$F^n(a_n^{-1}x + b_n) \xrightarrow{\omega} G(x). \quad (4)$$

где символ $\xrightarrow{\omega}$ обозначает, как и выше, сходимость в точках непрерывности предельной функции. Если соотношение (4) 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 $F \in D(G)$.

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 в работе [7], где можно найти доказательства их достаточности.

Приведем несколько очень простых и полезных достаточных условий, применимых в случае, когда ф.р. F имеет плотность f .

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

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

Таким образом не обладая представительным статистическим рядом данных об экстремальных ЧС, тем не менее можно утверждать:

1) что «хвосты» распределения гипотетически будут 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

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 .

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 и для дальнейшего развития теории как для случая н.о.р. случайных величин, так и для зависимых последовательностей.

2. Специальные графические статистические средства: графики квантилей

В этом разделе очень кратко 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 «наблюденной» плотностей распределения. В действительности же непосредственной связи между теоретической и наблюдаемой плотностями распределения нет, и вообще *определенной* «наблюденной» плотности не существует. Чтобы получить какую-то «наблюденную» плотность распределения, обычно выбирают некоторый класс интервалов разбиения и подсчитывают число наблюдений, содержащихся в соответствующих интервалах. Здесь имеют место два произвольных шага: выбор длины интервала и выбор начальной точки деления. Влияние первого шага хорошо известно, а о влиянии второго упоминается редко. Это обуславливается тем, что, к сожалению, имеется обычай публиковать только сгруппированные данные. Неоднозначность понятия «наблюденного» 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 приведут 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.

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 $P(X_i \leq x) = F(x)$.

Конечно, 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, в среднем делят ось x так, что $(N+1)$ полученных промежутков отвечают равным приростам функции $F(x)$.

Другими словами, расположив наблюдения в порядке неубывания, мы получим вариационный ряд величин X

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

Тогда вероятности $F(x)$ в соответствующих точках будут:

$$F(x_1) < F(x_2) < \dots < F(x_k) < F(x_N).$$

Так же как и сами X_k , они будут случайными величинами. Но, как можно показать, 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

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

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 навверное сколь угодно малую, если N достаточно велико.

Далее для закона распределения $\phi(y)$, где $y = \alpha(x - \beta)$, причем α и β — параметры распределения, определим отвечающие $\frac{k}{N+1}$ значения $y = y_k$ такие, что $\left(\phi(y_k) = \frac{k}{N+1}\right)$. 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 $x = \beta + \frac{y}{\alpha}$ и отклоняются от нее лишь в силу наличия случайных ошибок.

Если бы величина X точно следовала бы закону $F(x) = \phi(y)$ и мы точно знали бы всякий раз, какому значению вероятности $F(x)$ отвечает наблюдаемое значение x , то имели бы точную линейную зависимость $y = \alpha(x - \beta)$. 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 с использованием вероятностной бумаги.

Преимущество вероятностной бумаги по сравнению с обычными способами построения кривых заключается в том, что она преобразует теоретическую кривую

$$(F(x), x)$$

в прямую линию.

Таким образом, если на квантиль-диаграмме наблюдается линейный участок, то прямая может быть построена с использованием, например, классического алгоритма метода наименьших квадратов, в частности прямая линия с наклоном b и нулевым смещением минимизирует следующую сумму квадратов

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

что приводит к хорошо известной формуле:

$$\hat{b} = \frac{\sum_{i=1}^n x_i^* q_i}{\sum_{i=1}^n q_i^2},$$

где

$$q_i = -\ln \left(1 - \frac{i}{n+1} \right), i = 1, 2, \dots, n.$$

Степень правдоподобия гипотезы может быть измерена посредством коэффициента корреляции:

$$r_Q = \frac{\sum_{i=1}^n (x_i^* - \bar{x})(q_i - \bar{q})}{\sqrt{\sum_{i=1}^n (q_i - \bar{q})^2 \sum_{i=1}^n (x_i^* - \bar{x})^2}},$$

где

$$\bar{x} = n^{-1} \sum_{i=1}^n x_i^*, \bar{q} = n^{-1} \sum_{i=1}^n q_i.$$

Значение коэффициента корреляции всегда удовлетворяет неравенству $r_Q < 1$. Кроме того, $r_Q = 1$

только в том случае, если все точки лежат на прямой. Значение коэффициента корреляции следует рассматривать только как статистический критерий соответствия модели эмпирическим данным. Основываясь на показателе r_Q , 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.

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.

Процедура построения графиков квантилей в общем случае

Таким образом, идея построения графиков квантилей, или 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 правдоподобное соответствие имеющемуся распределению случайной переменной?"

Суммируя вышеизложенное, можно очертить общие идеи, принципы и схему построения графиков квантилей [8]. А именно, чтобы построить график квантили 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 теоретической квантилью $Q(p)$ гипотетического распределения в зависимости от вычисляемого (по p) эмпирического значения;

затем по функции теоретической квантили $Q(p)$ рассчитываются соответствующие значения эмпирической квантили $\hat{Q}_n(p_i) = Q\left(\frac{i}{n+1}\right)$ 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 квантили* 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 квантили 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.

3. Статистическая обработка данных многолетних наблюдений

Рассмотрим вопросы, как можно классическую теорию экстремальных значений применять к анализу многолетних данных по чрезвычайным ситуациям природного и техногенного характера, с примерами статистической обработки данных многолетних наблюдений по лесным пожарам и пожарам в жилом секторе.

Для того, чтобы применить технику статистической обработки данных в приложении к лесным пожарам, необходимо сначала построить ряды многолетних данных по максимальным значениям предполагаемых для обработки показателей. Для демонстративных целей мы выбрали максимальные длительности и максимальные площади лесных пожаров. Сначала продемонстрируем процедуру статистической обработки данных по максимальным длительностям пожаров. В табл. 1 данные по максимальным длительностям представлены последовательно по годам и в последней колонке представлен построенный на их основе вариационный ряд данных x_k ($k = 1, 2, \dots, N$) максимальных по годам значений длительности пожаров.

Теперь необходимо сопоставить данные наблюдений с теоретической функцией распределения и оценить параметры последней. Исходя из выводов, представленных выше, в качестве теоретической функции будем использовать второй тип распределения для максимумов $P_{II,n}(x)$ — распределение Фреше: $P_{II,n}(x) \equiv F(x) = \exp(-x^{-\alpha})$ для $x, \alpha > 0$.

Квантиль диаграммы в этом случае (Фреше) строятся с использованием координатной плоскости (рис. 1):

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

Искомое уравнение будет иметь вид:

$$\ln x = 0,64y + 1,374.$$

Отсюда вытекает закон распределения максимальных длительностей лесных пожаров:

$$F(x) = \exp[(0,2531x)^{-1,5625}].$$

Регрессионную прямую наносим на график (рис. 1), продолжая ее в область с наибольшими и

Таблица 1

Максимальные длительности (сут.) лесных пожаров в Тверской области с 1990 по 2005 годы

№ п/п	Год	Максимальная длительность, сут.	Вариационный ряд максимальных длительностей, сут.
1	1990	4	2
2	1991	2	2
3	1992	18	3
4	1993	3	3
5	1994	10	4
6	1995	15	4
7	1996	16	4
8	1997	4	4
9	1998	3	6
10	1999	8	6
11	2000	6	6
12	2001	6	8
13	2002	6	10
14	2003	4	15
15	2004	4	16
16	2005	2	18

наименьшими значениями. Находим, например, что значению вероятности $P_{п,п}(x) = \Phi_n(y) = 0,95$ соответствует значение $\ln x = 3,7$, что дает максимальную длительность пожаров $[x] \approx 40$ суток.

Таким образом, в Тверской области максимально возможная длительность лесного пожара в случае его возникновения может достигать значения $[x] \approx 40$ суток.

Аналогично получены другие законы распределения для параметров лесных пожаров.

Закон распределения максимальных площадей (в га) лесных пожаров в Тверской области:

$$F(x) = \exp[(0,0444x)^{-0,893}].$$

Логарифмически нормальный закон распределения количества пожаров наилучшим образом описывает имеющиеся данные по Тверской области (коэффициент корреляции 0,994 очень близок к 1). Результаты статистической обработки данных представлены на рис. 2. Находим, например, что значению вероятности $\Phi_n(y) = 0,95$ соответствует по графику значение $\ln x \approx (7,7 - 7,8)$, что дает максимальное количество пожаров $[x] \approx (2300 - 2400)$.

Таким образом, в Тверской области максимально возможное количество лесных пожаров в год может достигать значения $[x] \approx (2300 - 2400)$.

Закон распределения, в данном случае задаваемый плотностью распределения, количества лесных пожаров в Тверской области выглядит следующим образом:

$$f(x) = \frac{0,275}{x} \exp[-1,039 \ln^2(x / 228,4)].$$

Другие примеры представлены на рис. 3 и рис. 4.

Закон распределения максимального количества погибших при пожарах в жилом секторе (Тверской области):

$$F(x) = \exp[-\exp[-1,37x + 5,06]].$$

Распределение «внешних издержек» аварий на атомных станциях (рис. 4) относится к классу Вейбулловых распределений.

«Внешние издержки» — это расходы, связанные с мероприятиями непосредственно направленными на защиту населения:

- эвакуация;
- временные убежища и питание;
- компенсационные выплаты;
- сельскохозяйственные ограничения;

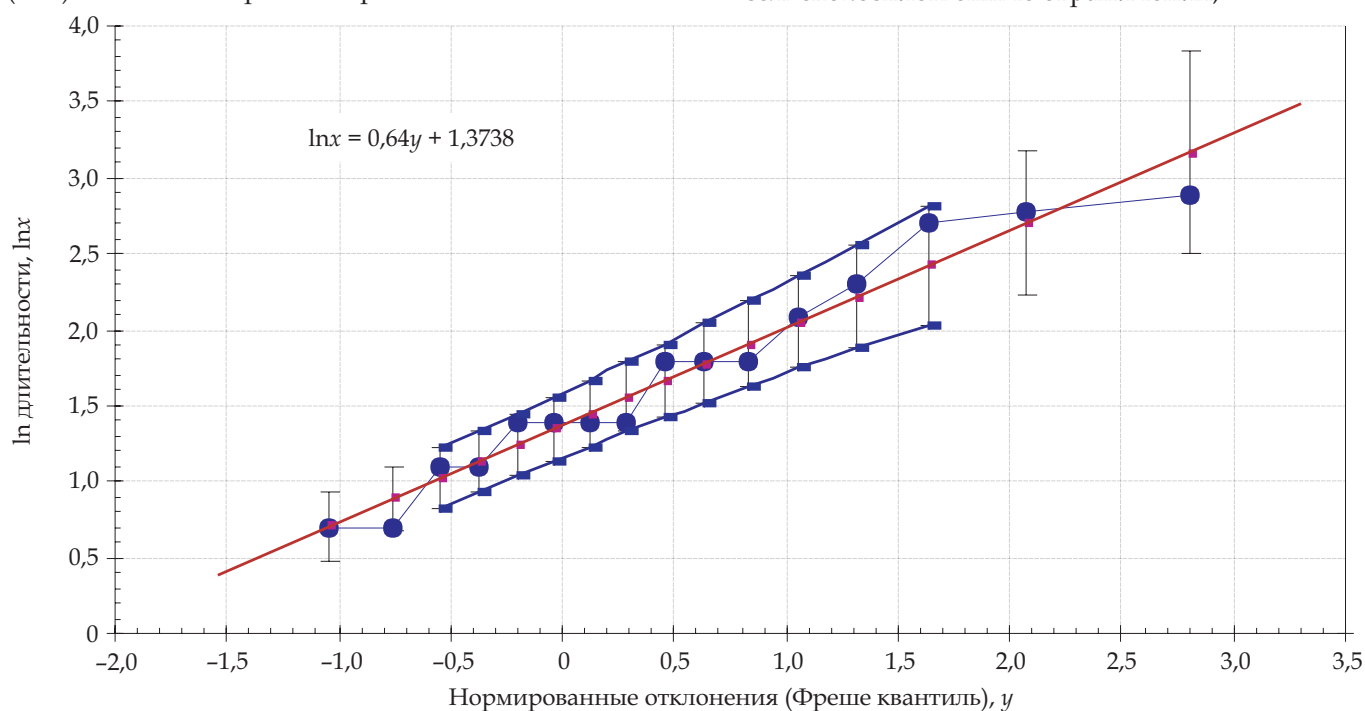


Рис. 1. Квантиль диаграмма максимальной длительности пожаров в Тверской области с 1990 по 2005 годы

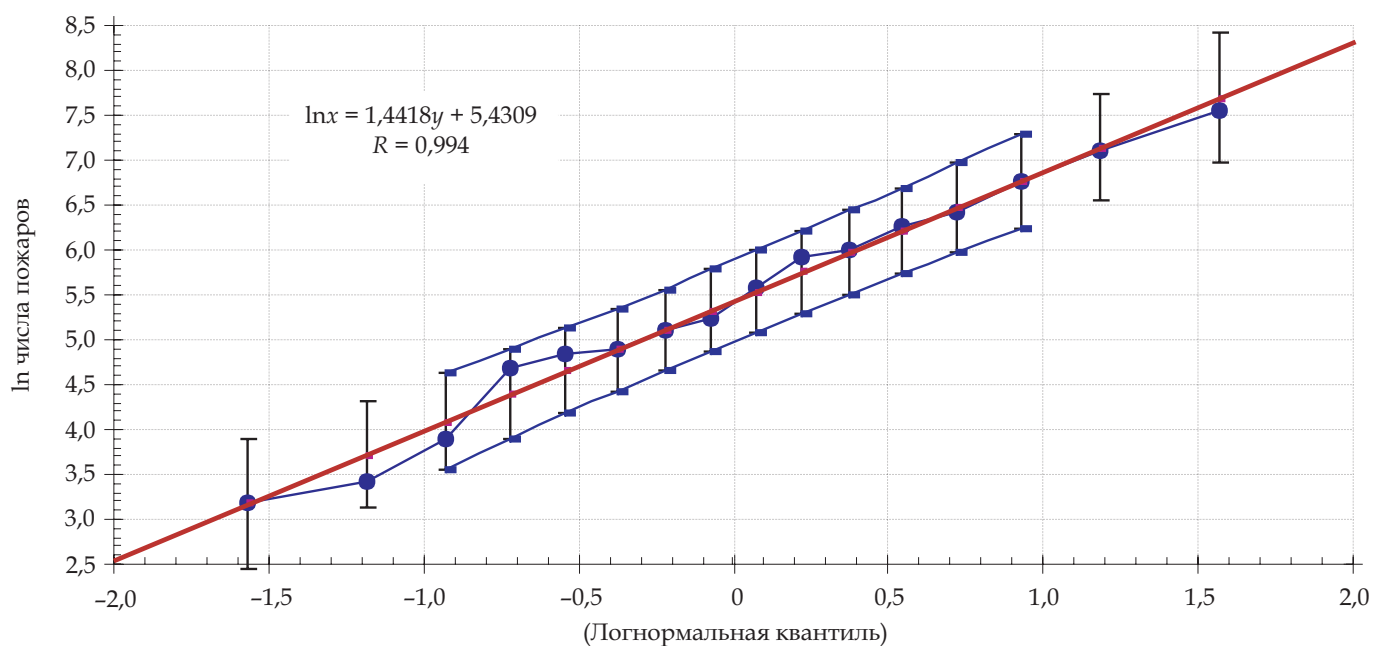


Рис. 2. Квантиль диаграмма \ln количества (числа) лесных пожаров в Тверской области

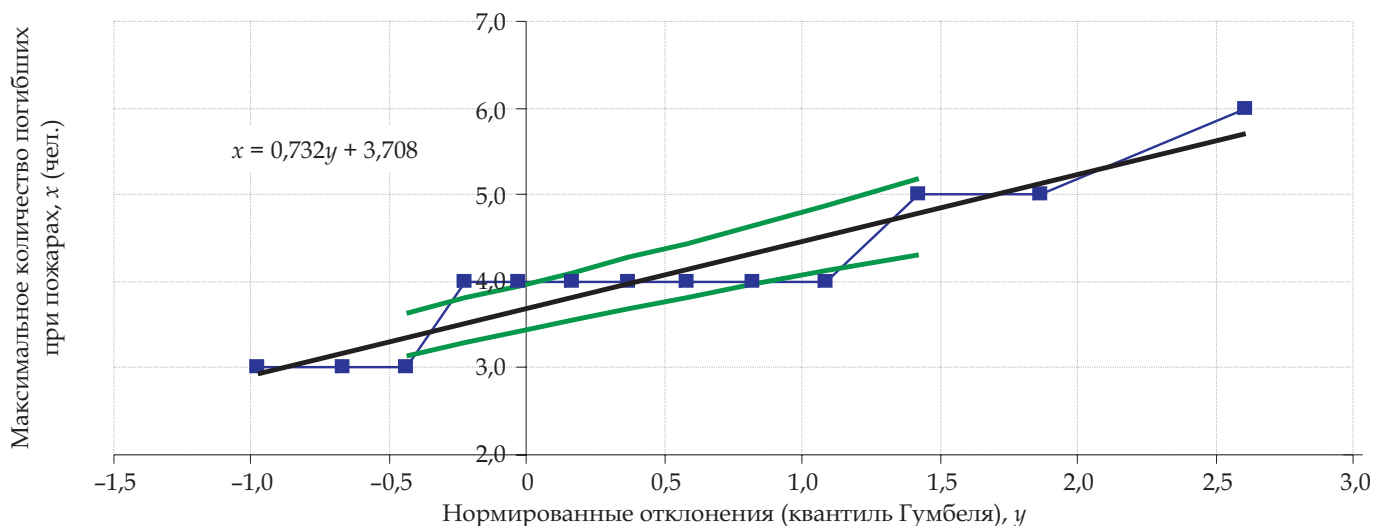


Рис. 3. Квантиль диаграмма максимального количества погибших при пожарах в жилом секторе

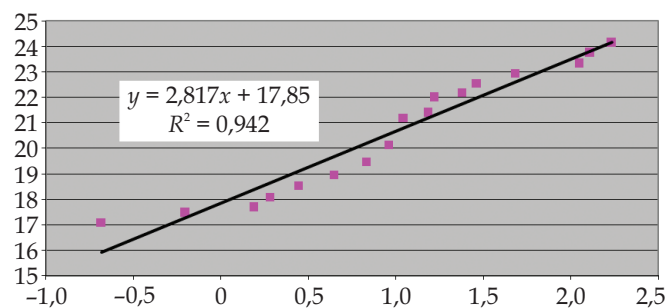


Рис. 4. Квантиль диаграмма распределения Вейбулла «внешних издержек» (без учета радиационно-индуцированных последствий) для АС Mühleberg

дезактивация территории.

Издержки радиационно-индуцированных последствий — это, главным образом, затраты на социальное, в том числе и медицинское, обеспечение граждан, получивших в результате радиационной аварии ущерб здоровью, и их потомков.

Класс Вейбулловых распределений:

$$1 - F(x) = \exp(-\lambda x^\tau)$$

Квантиль диаграммы в этом случае строятся с использованием координатной плоскости (рис. 4):

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

Наклон этого графика можно будет использовать как приближение параметра $1/\tau$. Параметр τ может интерпретироваться как параметр рассеяния, чем меньше τ , тем более широкой становится форма распределения.

Параметр λ может быть оценен через точку пересечения с осью ординат.

Для данных, представленных на рис. 4, распределение Вейбулла приобретает вид:

$$1 - F(x) = \exp(-0,0018x^{0,355}) \text{ для } x > 0.$$

Математическое ожидание и дисперсия определяется формулами:

$$E(x) = \lambda^{-\frac{1}{\tau}} \Gamma\left(1 + \frac{1}{\tau}\right)$$

и

$$\sigma_x^2 \equiv D(x) = \lambda^{-\frac{2}{\tau}} \left[\Gamma\left(1 + \frac{2}{\tau}\right) - \Gamma^2\left(\frac{1}{\tau} + 1\right) \right].$$

Гамма-функция:

$$\Gamma(y) = \int_0^{\infty} x^{y-1} e^{-x} dx.$$

В используемом примере (рис. 4):

$$\sigma = \sqrt{D} \approx 10^9 \$,$$

$$E \approx 2,58 \cdot 10^8 \$.$$

С точки зрения прогнозирования важными элементами являются:

установление динамики изменения законов распределения, например, параметров пожаров (год от года, рис. 5, рис. 6): длительностей, площадей и др. Анализируя представленные на рис. 5 и рис. 6 данные можно заключить, что средняя площадь пожаров в жилом секторе и дисперсия, а также средняя длительность пожаров в жилом секторе и дисперсия с каждым годом постепенно уменьшаются;

определение периодов повторяемости (табл. 2 и рис. 7). Наиболее часто следует ожидать количество пожаров из диапазона значений от 37 до 39 в год, ожидаемый период которых составляет минимальную величину 2 года. Максимальное количество пожаров которое может превышать значение 50 пожаров в год следует ожидать 1 раз в 17 лет. В то же время это значение может быть близким к 6 годам, что требует учета при планировании сил и средств, а также финансового резерва для ликвидации пожаров.

4. На какие другие вопросы можно получить ответы с использованием приведенной технологии

Вопрос 1: Можно ли, зная законы распределения количества пожаров в жилом секторе (рис. 7) и затрат на ликвидацию пожаров (рис. 8), спрогнозировать (или запланировать, например, на будущий год) размер необходимых затрат C_{95} на ликвидацию пожаров так, чтобы его было достаточно для ликвидации пожаров с гарантией 95 %?

Ответ: Да, можно, если дополнительно применить к полученным результатам на основе асимптотической теории экстремальных значений классическую технику актуарных расчетов.

На основе методов актуарной математики можно получить следующую формулу для расчета C_{95} :

$$C_{95} = \mu_V \cdot \mu_N + 1,645 \sqrt{\mu_N \cdot \sigma_V^2 + \mu_V \cdot \sigma_N^2},$$

где μ_V, μ_N — математические ожидания, σ_V^2, σ_N^2 — дисперсии затрат на ликвидацию V и числа пожаров N соответственно.

Вопрос 2: Можно ли, используя графическую технику предложить упрощенный подход для введения статистического критерия экстремальной ЧС?

Продemonстрируем основную идею подхода к ответу на поставленный вопрос.

График экспоненциальной квантиль-диаграммы эмпирических данных по погибшим в результате взрывов и пожаров в России в 1991–1996 гг. (рис. 9) имеет квазилинейный участок вплоть до значения числа погибших меньше или равно 7:

$$X_i \leq 7.$$

При этом значении наблюдается изгиб на графике.

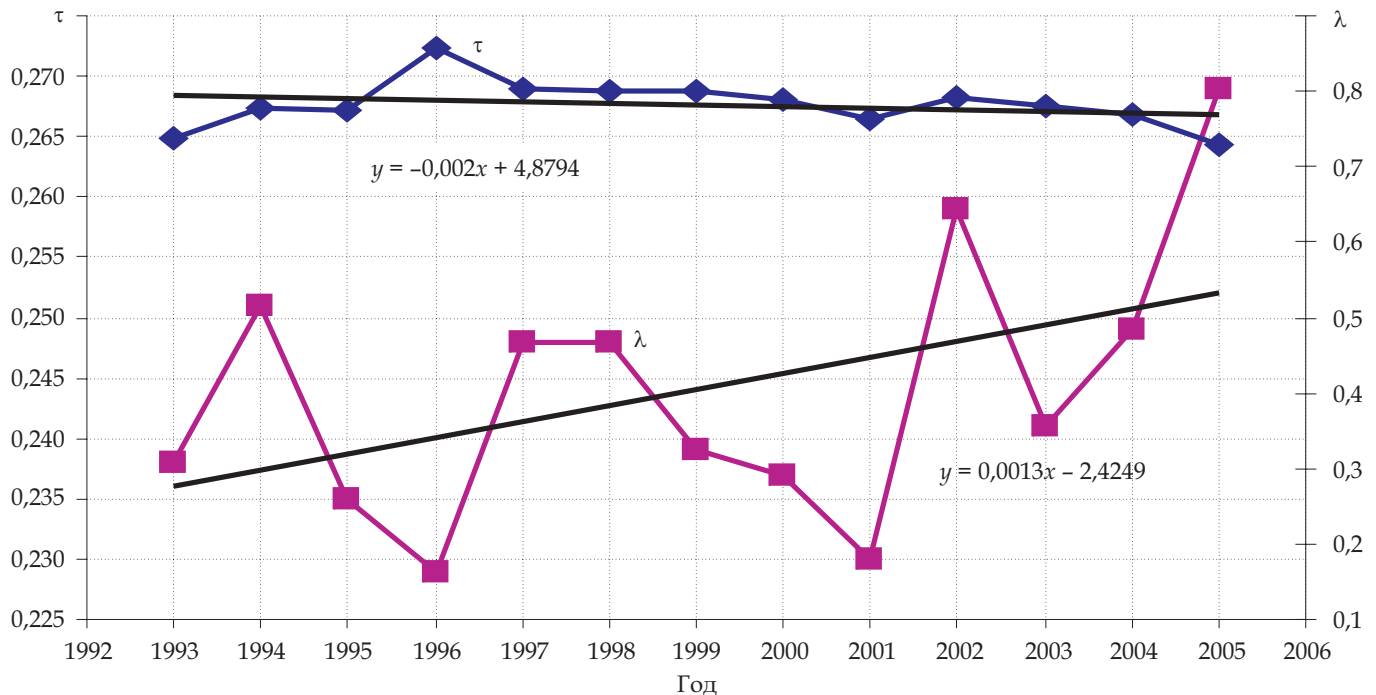


Рис. 5. Годовая динамика параметров распределения Вейбулла по площади пожаров в жилом секторе Тверской области

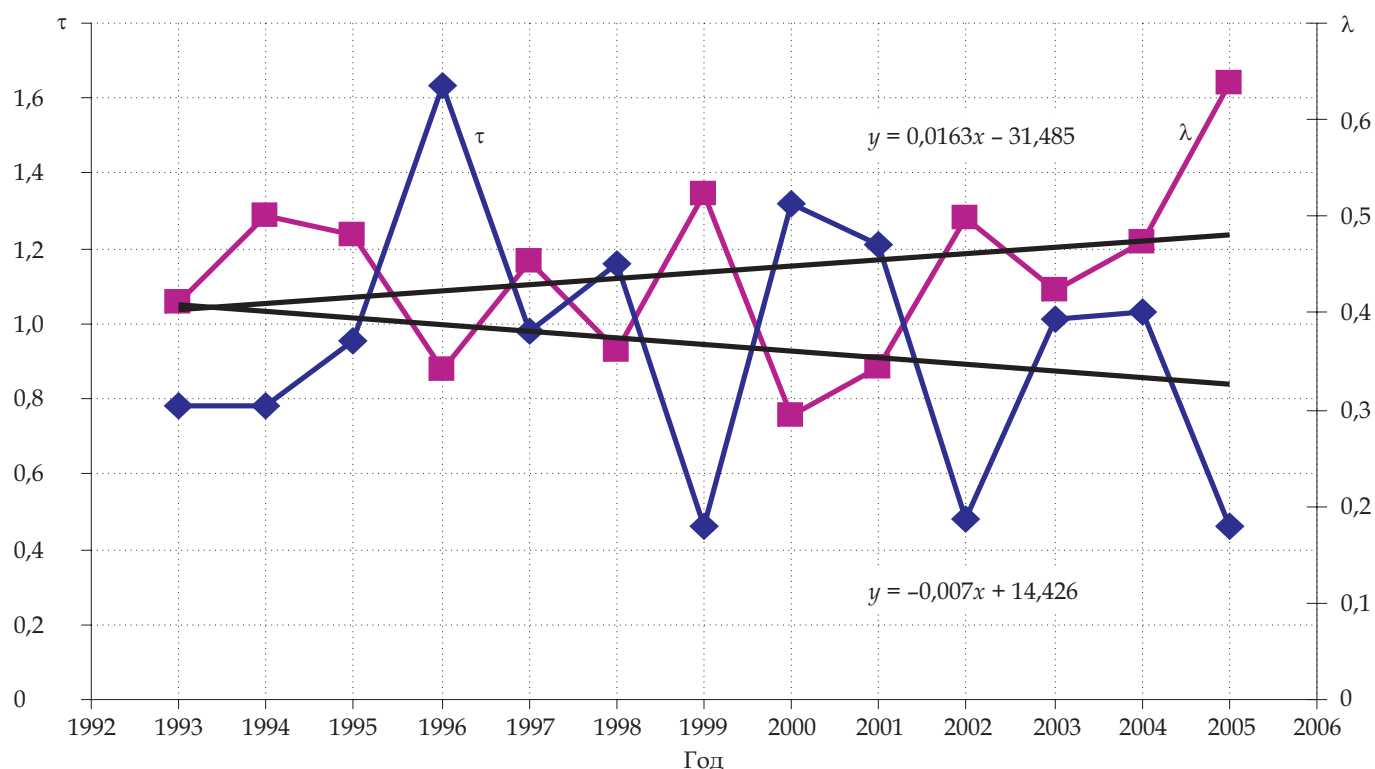


Рис. 6. Годовая динамика параметров распределения Вейбулла по площади пожаров в жилом секторе Тверской области

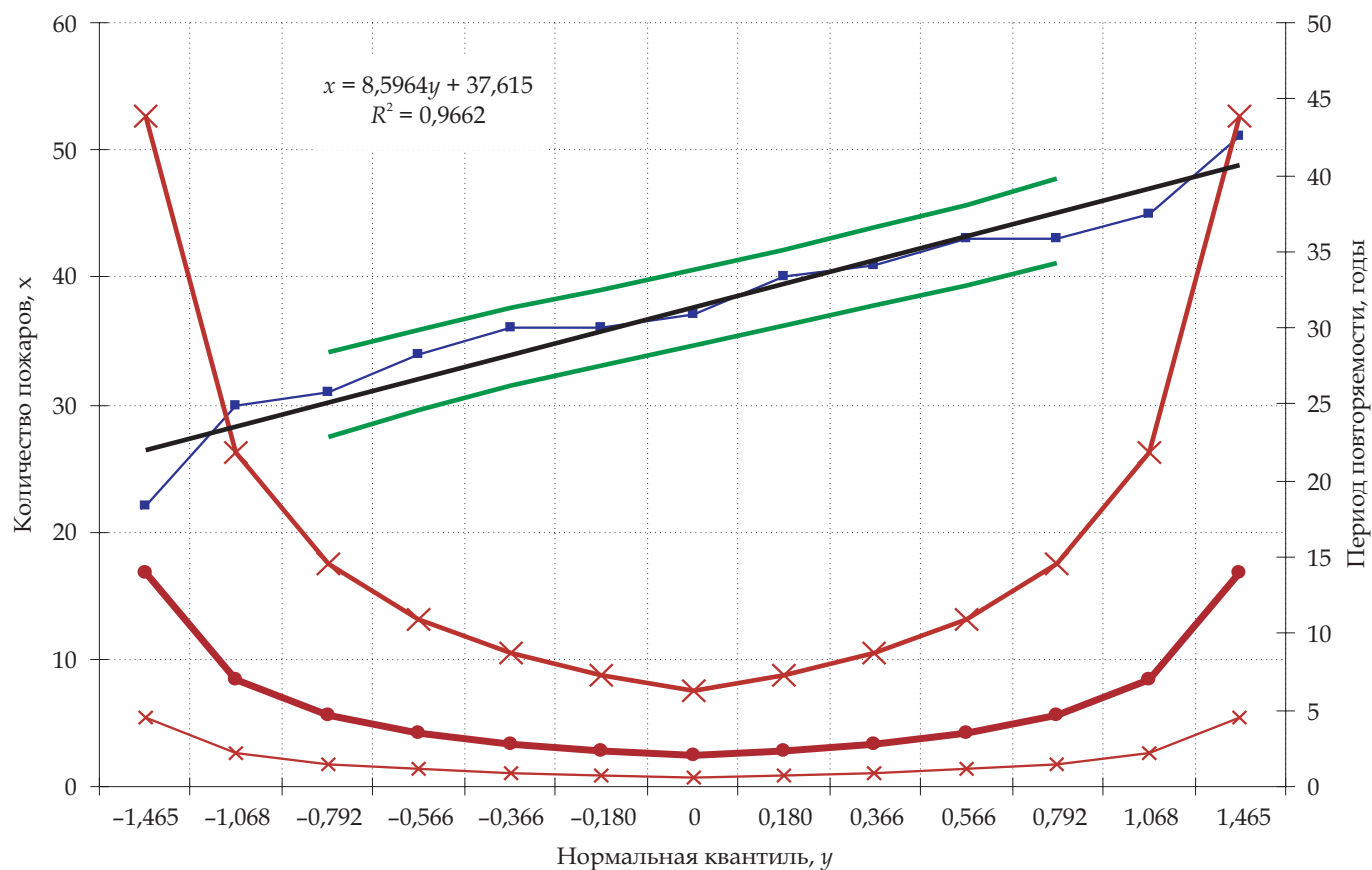


Рис. 7. Квантиль диаграмма и периоды повторяемости $T_1(x)$ (левая ветвь) и $T(x)$ (правая ветвь) количества пожаров в жилом секторе Тверской области

Примечание. Закон распределения: $f(x) = \frac{1}{0,116\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x + 4,38}{0,116}\right)^2\right)$.

Таблица 2

Диапазон числа пожаров и ожидаемый период их повторяемости

Диапазон числа пожаров		Ожидаемый период повторяемости, год	Диапазон периода повторяемости, год	
< 25	> 50	17	6	55
25 – 28	48 – 50	12	4	39
28 – 30	46 – 48	6	1,5	19,5
30 – 32	44 – 46	4,5	1	13,5
32 – 36	40 – 44	3	0,66	9
36 – 37	39 – 40	2,5	0,5	7
37 – 39		2	0,5	7

Этот факт возможно использовать в дальнейшем в качестве основной идеи метода введения статистического критерия экстремальной чрезвычайной ситуации.

Таким образом, основная идея метода установления статистического критерия экстремальной чрезвычайной ситуации основана на исследовании квазилинейных участков на квантиль диаграммах и областей нелинейности.

Подобный «изгиб» мы наблюдаем на многих квантиль диаграммах по исследованию статистики погибших и пострадавших при взрывах и пожарах. Другими словами, при выборе различных распределений и проверке на соответствие статистическим данным с использованием графической техники построения квантиль диаграмм, мы будем наблюдать квазилинейные участки и изгибы на некоторых графиках в определенном диапазоне данных.

Этот факт можно попытаться объяснить с позиций законов теории вероятностей следующим образом. Как правило, значения случайных величин, недалеко отстоящие от среднего («центральные» значения) подчиняются закономерностям, предсказываемым на основе центральной предельной теоремы и ряда других классических законов теории вероятностей. В то же время экстремальным значениям свойственны свои закономерности, с характерными асимптотическими вероятностными распределениями. Поэтому наблюдаемый на графиках, в первую очередь экспоненциальной, нормальной и логнормальной квантиль диаграмм, а также графиках квантилей Вейбулла, Парето, Гумбеля и Фреше, «изгибы» вероятно будут свидетельствовать о переходе от центральных значений к экстремальным.

Отсюда можно сделать следующий вывод: «**Область изгиба на графиках квантилей можно рассматривать как область значений, в которой находится статистический количественный критерий экстремальной чрезвычайной ситуации**». Правда, за счет графической техники установления такого критерия он имеет некоторый элемент неопределенности. В частности из рассматриваемого графика экспоненциальной квантили число погибших более 7 может быть взято за основу количественного критерия экстремальной чрезвычайной

ситуации в жилищном секторе при пожарах и взрывах.

Для более точного введения подобного критерия на основе техники построения квантиль диаграмм необходим более тщательный анализ графиков квантилей различных распределений (а также их хвостов), в первую очередь — указанных выше (на достоверной статистике). Процедура может быть следующей. Первоначально определяется диапазон значений (например, числа пострадавших или погибших при взрывах и пожарах) исследуемой переменной на различных графиках квантилей. Например, из построенных нами графиков видно, что точки «изгиба» наблюдаются в диапазоне 5 — 15 погибших и 50 — 55 пострадавших при пожарах и взрывах на квантиль диаграммах различных распределений. Отсюда следует, что количественных критерий экстремальной ЧС по погибшим или пострадавшим в результате пожаров и взрывов на производстве и в жилищном секторе может быть установлен в этих диапазонах. Именно таким способом из графиков квантилей можно, например, установить следующие количественные значения статистического критерия экстремальной ЧС:

для числа погибших при взрывах и пожарах — 8,

для числа пострадавших при взрывах и пожарах — 30,

для совокупности погибших и пострадавших при взрывах и пожарах — 50.

Совершенно аналогично можно установить отдельные количественные критерии экстремальной ЧС для жилищного сектора и производств, т.е. техногенной сферы.

Для уточнения диапазона и количественного критерия экстремальной ЧС дополнительно строится серия графиков квантилей распределений хвостовых значений или значений, превышающих определенный уровень. С некоторого значения или области значений исследуемой переменной квантиль диаграммы приобретают линейный вид. Это значение с.в. или диапазон значений можно использовать для уточнения количественного статистического критерия экстремальной ЧС.

Заключение

Представленные выше и многие другие приложения асимптотической теории вероятностей эк-

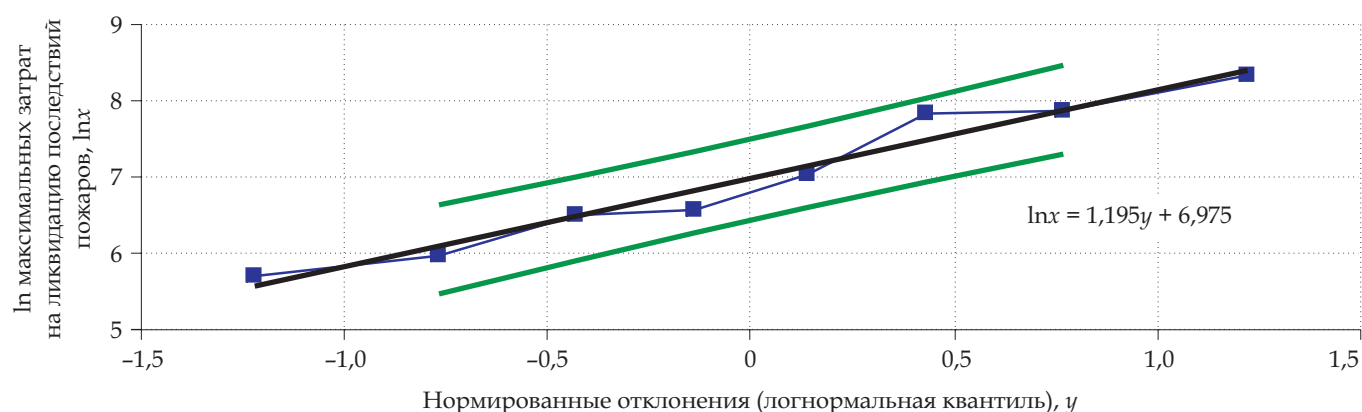


Рис. 8. Квантиль диаграмма \ln максимальных затрат на ликвидацию пожаров в жилом секторе

Примечание. Закон распределения: $f(x) = \frac{0,133}{x} \exp\left(-0,35 \ln^2\left(\frac{x}{1069,6}\right)\right)$.

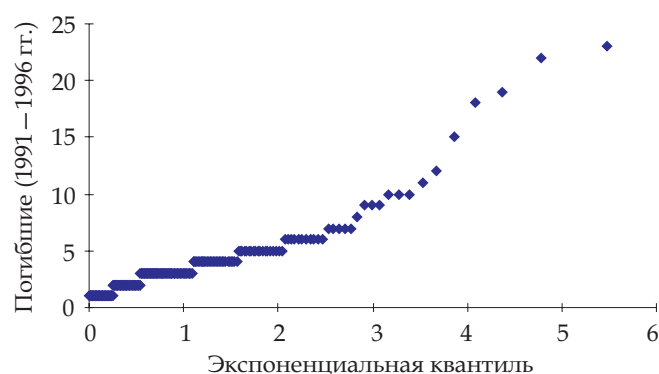


Рис. 9. Экспоненциальная квантиль-диаграмма эмпирических данных по погибшим в результате взрывов и пожаров в России в 1991 – 1996 гг.

стремальных значений с применением более «изошренных» методов достаточно подробно описаны в изданной в 2009 году монографии: «Введение в статистику экстремальных значений и ее приложения». Однако, прикладные возможности асимптотической теории вероятностей экстремальных значений гораздо шире описанных в монографии, далеко не до конца изучены и достаточно перспективны в области прогнозирования риска ЧС.

Литература

1. Гумбель Э. Статистика экстремальных значений. М.: Мир, 1965.
2. Лидбеттер М., Линдгрен Г., Ротсен Х. Экстремумы случайных последовательностей и процессов. М.: Мир, 1989. 392 с.

3. Фреше (Frechet M.) Sur la loi de probabilité de l'écart maximum. Ann. Soc. Math. Polon., 6, 93 – 116, 1927.
4. Фишер, Туннем (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., 24, 180 – 190, 1928.
5. Гнеденко Б.В. (Gnedenko B.V.) Sur la distribution limite du terme maximum d'une série aléatoire. Ann. Math., 44, 423 – 453, 1943.
6. Хан де (Haan de L.) Sample extremes: an elementary introduction. Statist. Neerlandica, 30, 161 – 172, 1976.
7. Loynes R.M. Extreme values in uniformly mixing stationary stochastic processes. Ann. Math. Statist., 36, 993 – 999, 1965.
8. Beirlant J., Taugels J.L., Vynckier P. (1996) Practical Analysis of Extreme Values. Leuven Universitu Press, Belgium.
9. Акимов В.А., Быков А.А., Щетинин Е.Ю. Введение в статистику экстремальных значений и её приложения. Монография; МЧС России. М.:ФГУ ВНИИ ГОЧС (ФЦ), 2009. 524 с.

Октябрь 2011 г.

Сведения об авторе:

Быков Андрей Александрович; ООО «Газпром ВНИИГАЗ»; e-mail: A_Bykov@vniigaz.gazprom.ru; 142717, Московская обл., Ленинский р-н, пос. Развилка; д.ф.-м.н.; профессор; заслуженный деятель науки Российской Федерации; начальник лаборатории «Экономические механизмы управления риском»; главный редактор журнала «Проблемы анализа риска»; тел. +7(498) 657-42-44.