

Распределение экстремальных значений

1. Историческая справка

Изучение свойств распределений экстремальных значений в течение долгого времени находилось несколько в стороне от основных направлений статистической теории распределений. Дело в том, что на ранней стадии создания статистической теории основное внимание уделялось проблемам подгонки кривых распределения, и лишь значительно позже — развитию теории статистического вывода. В настоящее время теория распределения экстремальных значений является составной частью многих естественнонаучных дисциплин. Упомянем в связи с этим изучение таких явлений как ливни, ураганы, наводнения, загрязнение атмосферы и коррозия, а также тонкие математические результаты, касающиеся точечных случайных процессов и регулярно меняющихся функций. Распределениями экстремальных значений первоначально интересовались абстрактные вероятностники, да специалисты в прикладных областях — инженеры и гидрологи. Только с недавних пор эти распределения вошли в сферу существенных интересов специалистов по статистике. Хронологически первые сведения о существовании семейства распределенных экстремальных значений связаны с работой 1709 г. Николая Бернулли, где обсуждается распределение координаты точки, наиболее удаленной от начала отсчета, из n точек, случайно расположенных на отрезке длины t . См. об этом также в книге Gumbel (1958).

Теория распределений экстремальных значений по запросам астрономии восходит, по-видимому, к решению задачи об отбраковке и использовании резко уклоняющихся наблюдаемых значений. Ранние статьи Fuller (1914) и Griffith (1920), посвященные этой теме, весьма специальные как по постановке прикладных задач, так и по примененным математическим методам. Начало систематического изучения теории можно, вероятно, отнести к статье Bortkiewicz (1922), где изучается размах выборки из нормальной популяции. Об этом уже говорилось в гл. 13 и, как там было отмечено, дальнейшие результаты последовали достаточно быстро. С современной точки зрения следует указать на важность статьи Bortkiewicz (1922), поскольку в ней впервые ясно сформулирована проблема нахождения *распределения наибольшего значения* в последовательности случайных величин. Буквально через год von Mises (1923) вычислил значение математического ожидания, а Dodd (1923) вычислил медиану распределения и обсудил проблемы, возникающие для

негауссовских порождающих распределений. Очень близко к вопросам, рассматриваемым в настоящей главе, примыкает результат Fréchet (1927), нашедшего асимптотическое распределение наибольших значений. В следующем году появилась работа Fisher and Tippet (1928), независимо получивших близкие результаты для одной и той же проблемы. Еще Fréchet (1927) выделил один из классов предельных распределений наибольшей из порядковых статистик. Fisher and Tippet (1928) показали, что распределение экстремальных значений принадлежит одному из трех возможных типов. Еще раньше Tippet (1925) изучил поведение функции распределения и моментов наибольшего выборочного значения при возрастании объема выборки из нормального распределения. Von Mises (1936) нашел простые достаточные условия слабой сходимости функции распределения наибольшего из значений независимых нормальных случайных величин к каждому из трех типов распределений, ранее указанных в статье Fisher and Tippet (1928). Через семь лет Гнеденко (1943) опубликовал фундаментальное исследование по теории распределений экстремальных значений и установил необходимые и достаточные условия слабой сходимости функции распределения экстремальных значений к соответствующим предельным функциям. Усовершенствованное изложение результатов Б. В. Гнеденко содержится в статье de Haan (1970). Классическая статья Гнеденко (1943) приведена полностью в первом томе книги *Основные достижения статистики (Breakthroughs in Statistics)* и снабжена предисловием R. L. Smith с анализом влияния результатов Б. В. Гнеденко на дальнейшее развитие теории распределений экстремальных значений.

За теоретическими исследованиями 20-х — середины 30-х гг. в 30-х и в 40-х гг. XX в. последовал целый ряд работ, связанных с приложениями распределений экстремальных значений. Сюда относятся исследования продолжительности человеческой жизни, интенсивности радиоактивного излучения [Gumbel (1937a, b)], прочности материалов [Weibull (1939)], наводнений [Gumbel (1941, 1944, 1945, 1949a)] и [Rantz and Riggs (1949)]. Сейсмические явления исследуются в статье Nordquist (1945), ливневые осадки — в работе Potter (1949).

Существенный вклад в прикладные исследования, связанные с распределениями экстремальных значений, принадлежат Гумбелю. Большинство полученных им прикладных результатов приводится в монографии Gumbel (1958), которая является расширенным вариантом брошюры Gumbel (1954). Многочисленные приложения распределений экстремальных значений приводятся также в п. 14 настоящей главы.

Библиография в конце главы насчитывает около 350 названий. Столь значительное число работ, однако, составляет лишь небольшую часть от общего числа публикаций, связанных с заявленной темой. Даже библиография в монографии Gumbel (1958), не включающая публикаций последних 35 лет, содержит гораздо больше ссылок. Столь большое число работ свидетельствует не только об актуальности и практической важности тематики, но также об отсутствии координации усилий исследователей, что зачастую приводит к повторному (и даже многократному) переоткрытию известных результатов, появляющихся в разных изданиях.

2. Введение

К семейству *распределений экстремальных значений* обычно относят следующие три типа семейств:

Тип 1:

$$\Pr[X \leq x] = \exp \{-e^{-(x-\xi)/\theta}\}. \quad (22.1)$$

Тип 2:

$$\Pr[X \leq x] = \begin{cases} 0, & x < \xi, \\ \exp \left\{ - \left(\frac{x-\xi}{\theta} \right)^{-k} \right\}, & x \geq \xi. \end{cases} \quad (22.2)$$

Тип 3:

$$\Pr[X \leq x] = \begin{cases} \exp \left\{ - \left(\frac{\xi-x}{\theta} \right)^k \right\}, & x < \xi, \\ 1, & x \geq \xi. \end{cases} \quad (22.3)$$

Здесь ξ , $\theta > 0$ и $k > 0$ — параметры. Распределения, получающиеся при замене случайной величины X на $-X$, также относятся к *распределениям экстремальных значений*.

Из приведенных трех семейств чаще всего в качестве распределения экстремальных значений упоминается первое. Некоторые авторы даже считают (22.1) единственным таким распределением. Учитывая это, а также то, что распределения (22.2) и (22.3) приводятся к типу 1 с помощью простых преобразований

$$Z = \log(X - \xi) \quad \text{и} \quad Z = -\log(\xi - X)$$

соответственно, мы в этой главе будем, большей частью, рассматривать свойства первого семейства. Заметим также, что распределение типа 3 после перехода от X к $-X$ будет *семейством распределений Вейбулла*. Эти распределения изучаются в гл. 21, так что здесь нет необходимости заниматься ими в деталях.

Конечно, типы 1 и 2 также тесно связаны с распределениями Вейбулла, это объясняется приведенными соотношениями между Z и X . Распределения типа 1 иногда называют *логвейбулловскими распределениями*, см. например, в работах White (1964, 1969).

Распределения первого семейства иногда также называют *дважды экспоненциальными*, что объясняется формулой (22.1). Однако мы не используем этот термин, чтобы избежать путаницы с распределением Лапласа (гл. 24), тоже, хотя и редко, называемым дважды экспоненциальным.

Термин «экстремальные значения» в названии семейства распределения объясняется тем, что такие распределения получаются как предельные при $n \rightarrow \infty$ наибольшего значения из n независимых одинаково распределенных случайных величин (см. п. 3). Замена X на $-X$ приведет к распределению наименьшего значения. Уже сказано, что они тоже относятся к распределениям экстремальных значений, поэтому нет смысла разбирать их отдельно.

Хотя распределения связаны с экстремальными значениями, следует иметь в виду два обстоятельства: (1) они не описывают *всех* экстремальных значений (например, для конечного числа случайных величин или выборок конечного объема) и (2) их можно использовать как обычные распределения безотносительно к свойству экстремальности.

В связи с последним замечанием отметим, что распределения типа 1 аппроксимируются распределением Вейбулла при больших значениях его параметра c , см. формулу (21.3). Кроме того, если X имеет распределение типа 1, то $Z = \exp [-(X - \xi)/\theta]$ распределено по показательному закону с плотностью

$$p_Z(z) = e^{-z}, \quad 0 \leq z.$$

О терминологии. Распределение типа 2 называют также *распределением типа Фреше*, типа 3 — *распределением типа Вейбулла*, типа 1 — *распределением типа Гумбеля*. Как уже отмечено, распределения типа Фреше и типа Вейбулла получаются одно из другого при изменении знака случайной величины. К типу 1 принадлежит используемое в демографии *распределение типа Гомперца*, впервые упомянутое в 1825 году и применявшееся в течение почти ста лет до появления работ Фреше и Типпета, обнаруживших связь распределений Гомперца с семейством распределений экстремальных значений. Этот факт, однако, не часто упоминается в литературе. Подробнее мы обсудим это в п. 8.

Распределения (22.1)–(22.3) появились в литературе независимо друг от друга. Однако они являются представителями одного и того же более общего семейства функций распределения

$$\Pr[X \leq x] = e^{-\{1 + [(x - \xi)/\theta]/\alpha\}^{-\alpha}}, \quad 1 + \frac{x - \xi}{\alpha \theta} > 0, \\ -\infty < \alpha < \infty, \quad \theta > 0. \quad (22.4)$$

Если $\alpha > 0$, то (22.4) совпадает с (22.2). Если $\alpha < 0$, то (22.4) совпадает с (22.3). Наконец, если $\alpha \rightarrow \infty$ или $\alpha \rightarrow -\infty$, то (22.4) стремится к распределению типа 1. Поэтому распределение (22.4) можно считать *обобщением распределения экстремальных значений*. Иногда его называют *распределением фон Мизеса* или *распределением фон Мизеса–Дженкинса*. Подробнее об этом распределении см. в п. 15.

Достаточно полный обзор распределений экстремальных значений приведен в работе Mann and Singpurwalla (1982). Аналогичное исследование распределения Гумбеля содержится в работе Tiago de Oliveira (1983).

3. Предельные распределения экстремумов

Распределения экстремальных значений получаются как предельные распределения наибольшего (наименьшего) из значений независимых одинаково распределенных непрерывных случайных величин при бесконечном увеличении их числа или, что то же самое, наибольшего (наименьшего) выборочного значения при бесконечном увеличении объема выборки из непрерывного распределения.

Чтобы получить невырожденное предельное распределение, следует «уменьшать» наибольшее выборочное значение, вводя линейное преобразование с коэффициентами, зависящими от объема выборки. Это аналогично нормировке (например, как в формулировке центральной предельной теоремы; см. гл. 13 п. 2), однако, вообще говоря, не ограничиваются только последовательностями линейных преобразований.

Пусть $X_1, X_2, \dots, X_n$ — независимые одинаково распределенные случайные величины с общей плотностью

$$p_{X_j}(x) = f(x), \quad j = 1, 2, \dots, n.$$

Тогда функция распределения случайной величины $X'_n = \max(X_1, X_2, \dots, X_n)$ имеет вид

$$F_{X'_n}(x) = [F(x)]^n, \quad (22.5)$$

где

$$F(x) = \int_{-\infty}^x f(t) dt.$$

Ясно, что для любого фиксированного x при стремлении n к бесконечности

$$\lim_{n \rightarrow \infty} F_{X'_n}(x) = \begin{cases} 1, & \text{если } F(x) = 1, \\ 0, & \text{если } F(x) < 1. \end{cases}$$

Даже если получается собственное распределение, оно «тривиально» и не представляет интереса. Чтобы получить содержательный результат, мы должны найти предельное распределение последовательности преобразованных, в некотором смысле «уменьшенных» значений, таких как $\{a_n X'_n + b_n\}$, где a_n и b_n могут зависеть от n , но не от x .

Чтобы различать $F(x)$ и функцию распределения наибольшего значения преобразованной («уменьшенной») случайной величины, обозначим последнюю $G(x)$. Если имеется Nn величин: $X_1, X_2, \dots, X_{Nn}$, то наибольшая из них есть также наибольшая из N величин:

$$\max(X_{(j-1)n+1}, X_{(j-1)n+2}, \dots, X_{jn}), \quad j = 1, 2, \dots, N.$$

Следовательно, функция $G(x)$ должна удовлетворять уравнению

$$[G(x)]^N = G(a_N x + b_N). \quad (22.6)$$

Это уравнение выведено в статьях Fréchet (1927) и Fisher and Tippet (1928). Его иногда называют *постулатом стабильности*.

Распределения типа 1 получаются при $a_N = 1$; распределения типов 2 и 3 — при $a_N \neq 1$. В этом последнем случае

$$x = a_N x + b_N \Leftrightarrow x = b_N(1 - a_N)^{-1},$$

и из (22.6) следует, что $G(b_N(1 - a_N)^{-1})$ должно быть равно 0 или 1. Тип 2 соответствует значению 1, тип 3 — значению 0.

Рассмотрим более подробно случай $a_N = 1$, соответствующий типу 1. В этом случае уравнение (22.6) принимает вид

$$[G(x)]^N = G(x + b_N). \quad (22.7)$$

Так как $G(x + b_N)$ также должно удовлетворять (22.6), то

$$[G(x)]^{NM} = [G(x + b_N)]^M = G(x + b_N + b_M). \quad (22.8)$$

В силу (22.6)

$$[G(x)]^{NM} = G(x + b_{NM}), \quad (22.9)$$

поэтому

$$b_N + b_M = b_{NM},$$

следовательно,

$$b_N = \theta \log N, \quad \text{где } \theta — \text{константа.} \quad (22.10)$$

Дважды логарифмируя почленно равенство (22.7), учитывая, что $G \leq 1$ и используя выражение (22.10) для b_N , записываем:

$$\log N + \log\{-\log G(x)\} = \log\{-\log G(x + \theta \log N)\}. \quad (22.11)$$

Положим

$$h(x) = \log\{-\log G(x)\}.$$

При увеличении аргумента $\theta \log N$ величина $h(x)$ убывает на $\log N$. Следовательно,

$$h(x) = h(0) - \frac{x}{\theta}. \quad (22.12)$$

Поскольку $h(x)$ убывает по x , то $\theta > 0$. Из (22.12) получаем:

$$-\log G(x) = \exp\left[-\frac{x - \theta h(0)}{\theta}\right] = \exp\left(-\frac{x - \xi}{\theta}\right),$$

где обозначено $\xi = \theta \log(-\log G(x))$. Таким образом,

$$G(x) = \exp\left\{-e^{-(x-\xi)/\theta}\right\},$$

что совпадает с (22.1). Мы опускаем вывод распределений типов 2 и 3, отсылая читателя к работам Galambos (1978, 1987).

В большой статье Гнеденко (1943) установлена связь между свойствами исходного распределения $[F(x)]$ в наших обозначениях и типом предельного распределения. Найденные Б. В. Гнеденко условия относятся к поведению $F(x)$ при больших (малых) x , если речь идет о наибольших (наименьших) значениях случайных величин. Оказалось, что при одном и том же исходном распределении наибольшее и наименьшее значения могут иметь предельные распределения, относящиеся к разным типам.

Приведем резюме результатов Б. В. Гнеденко.

Распределение типа 1. Определим X_α равенством

$$F(X_\alpha) = \alpha.$$

Условие сходимости к *распределению типа 1*:

$$\lim_{n \rightarrow \infty} n \left[1 - F\left(X_{1-n^{-1}} + y \left(X_{1-(ne)^{-1}} - X_{1-n^{-1}}\right)\right) \right] = e^{-y}. \quad (22.13)$$

Условие сходимости к *распределению типа 2*:

$$\lim_{x \rightarrow \infty} \frac{1 - F(x)}{1 - F(cx)} = c^k, \quad c > 0, \quad k > 0. \quad (22.14)$$

Условие сходимости к *распределению типа 3*:

$$\lim_{x \rightarrow 0^-} \frac{1 - F(cx + \omega)}{1 - F(x + \omega)} = c^k, \quad c > 0, \quad k > 0, \quad (22.15)$$

где $F(\omega) = 1$, $F(x) < 1$ при $x < \omega$.

Б. В. Гнеденко доказал также, что приведенные условия являются необходимыми и достаточными и что *не существует иных распределений*, удовлетворяющих постулату стабильности. Другая интерпретация этих условий приводится в работе Clough and Kotz (1965). Они рассмотрели специальную систему массового обслуживания, в которой возникают распределения экстремальных значений.

Среди распределений, удовлетворяющих условию сходимости к типу 1 (22.13), назовем нормальное, экспоненциальное и логистическое. Условию сходимости к типу 2 (22.14) удовлетворяет распределение Коши. Сходимость к типу 3 имеет место для распределений, сосредоточенных на ограниченной сверху части числовой оси.

Результаты Б. В. Гнеденко обобщались разными авторами. Н. В. Смирнов исследовал предельное поведение порядковых статистик как фиксированного, так и возрастающего порядков. В статье Смирнов (1952) полностью классифицированы предельные типы и области их притяжения при исследовании поведения максимального члена последовательности. Предположение об одинаковом распределении случайных величин заменено более слабым в статье Juncosa (1949). Watson (1954) исследовал наибольший член стационарной последовательности случайных величин с зависимостью от каждой из m предыдущих. При слабых ограничениях предельные распределения те же, что и в случае независимости. В статье Bergman (1962) изучены наборы перестановочных случайных величин и выборки случайного объема. В работе Harris (1970) классические результаты распространяются на одну из моделей теории надежности: последовательную систему заменяемых элементов. Weinstein (1973) обобщил основополагающий результат Б. В. Гнеденко, рассмотрев асимптотику распределения экспоненциального типа, если порождающей функцией распределения является $V(x) = 1 - e^{-x^v}$, $x \geq 0$. Он показал, что

$$\lim_{n \rightarrow \infty} V^n \left\{ \left(x_n^v + \frac{u}{d_n} \right)^{1/v} \right\} = e^{-e^{-u}}, \quad v > 0,$$

тогда и только тогда, когда

$$\lim_{n \rightarrow \infty} n \left[1 - V \left\{ \left(x_n^v + \frac{u}{d_n} \right)^{1/v} \right\} \right] = e^{-u}, \quad v > 0,$$

где

$$\begin{aligned} V(x_n) &= 1 - \frac{1}{n}, \\ V\left(x_n + \frac{1}{c_n}\right) &= 1 - \frac{1}{ne}, \\ d_n &= \frac{c_n}{v|x_n|^{v-1}}, \\ x^v &\equiv |x|^v \operatorname{sgn}(x). \end{aligned}$$

Результат Б. В. Гнеденко получается при $v = 1$. В работе Jeruchim (1976) содержится предупреждение, что в прикладных исследованиях следует внимательно относиться к оценке параметра v .

Выполнение необходимых и достаточных условий (22.13)–(22.15) не всегда легко установить. В этих случаях могут оказаться полезными достаточные условия, найденные фон Мизесом для абсолютно непрерывных распределений в статье von Mises (1936).

Для распределения типа 1. Если функция $r(x) = f(x)/[1 - F(x)]$ отлична от нуля и дифференцируема в точке $F^{-1}(1)$ (или при достаточно больших x при $F^{-1}(x) = \infty$), то достаточным условием сходимости к распределению типа 1 является

$$\lim_{x \rightarrow F^{-1}(1)-0} \frac{d}{dx} \left(\frac{1}{r(x)} \right) = 0. \quad (22.16)$$

Для сходимости к распределению типа 2 достаточно выполнения неравенства $r(x) > 0$ при всех достаточно больших x и существования предела

$$\lim_{x \rightarrow \infty} xr(x) = \alpha, \quad \alpha > 0. \quad (22.17)$$

Достаточным условием сходимости к распределению типа 3 является выполнение неравенства $F^{-1}(1) < \infty$ и существование предела

$$\lim_{x \rightarrow F^{-1}(1)-0} \{F^{-1}(1) - x\} r(x) = \alpha, \quad \alpha > 0. \quad (22.18)$$

В статье de Naan (1976) приводится простое доказательство этих условий. Напомним, что функция $r(x) = f(x)/[1 - F(x)]$, входящая в (22.16)–(22.18), есть *интенсивность отказа*, называемая также *функцией риска* (см. гл. 1 и п. В2).

Выбор (не однозначный) нормировочных констант a_N и $b_N > 0$ зависит от типа предельного распределения. Наиболее подходящие значения a_N и $b_N > 0$ даются следующими формулами.

Тип 1.

$$\begin{aligned} a_N &= F^{-1} \left(1 - \frac{1}{N} \right), \\ b_N &= F^{-1} \left(1 - \frac{1}{Ne} \right) - F^{-1} \left(1 - \frac{1}{N} \right). \end{aligned} \quad (22.19)$$

Тип 2.

$$\begin{aligned} a_N &= 0, \\ b_N &= F^{-1} \left(1 - \frac{1}{N} \right). \end{aligned} \quad (22.20)$$

Тип 3.

$$\begin{aligned} a_N &= F^{-1}(1), \\ b_N &= F^{-1}(1) - F^{-1} \left(1 - \frac{1}{N} \right). \end{aligned} \quad (22.21)$$

Аналогичные результаты для предельных распределений наименьших значений последовательностей случайных величин получаются очевидными преобразованиями.

Перечислим несколько, по нашему мнению, удачных книг, где изучаются разные аспекты теории экстремальных значений и статистические приложения.

В книгах David (1981) и Arnold, Balakrishnan and Nagaraja (1992) приводится компактное изложение асимптотической теории распределений экстремальных значений. В работах Galambos (1978, 1987), Resnik (1987) и Leadbetter, Lindgren and Rootzén (1983) излагаются уточнения и развитие теории. В книге Reiss (1988) обсуждаются различные вопросы, связанные со скоростью сходимости распределений экстремальных значений и порядковых статистик к предельным распределениям. Castillo (1988) развивает результаты, полученные в работе Gumbel (1958) и излагает статистические аспекты теории экстремальных значений. В работе Harter (1978) приводится аннотированная библиография по теории распределений экстремальных значений.

Обозначим $F_X(x; \xi, \theta)$ функцию распределения выборочного минимума, принадлежащую к типу 1 распределений экстремальных значений:

$$F_X(x; \xi, \theta) = 1 - e^{-e^{(x-\xi)/\theta}}, \quad \theta > 0, \quad \xi \in \mathbb{R}.$$

Пусть, далее, $G_X(x; a, b, c)$ — трехпараметрическое распределение Вейбулла:

$$G_X(x; a, b, c) = \begin{cases} 0, & x < c, \\ 1 - e^{-[(x-c)/b]^a}, & x \geq c, \end{cases}$$

где $a, b > 0, c \in \mathbb{R}$. Davidovich (1992) получил оценки разности между двумя функциями распределения. Он, в частности, показал, что

$$F_X\left(x; b+c, \frac{b}{a}\right) - G_X(x; a, b, c) < \begin{cases} e^{-a}, & x < c, \\ \frac{2e^{-2}}{a-2}, & c \leq x < c+2b, \\ e^{a-2a}, & x \geq c+2b. \end{cases}$$

Отсюда видно, что если $a \rightarrow \infty, b \rightarrow \infty$ и $c \rightarrow -\infty$ так, что $b+c \rightarrow d, |d| < \infty$ и $\frac{b}{a} \rightarrow f, 0 < f < \infty$, то приведенное распределение Вейбулла равномерно аппроксимирует распределение минимальных значений для $\xi = d, \theta = f$.

Нетрудно показать, что если независимые случайные величины $Y_1, Y_2, \dots$ одинаково показательно распределены (см. гл. 19, п. 1):

$$\Pr[Y \leq y] = 1 - e^{-y}, \quad y > 0, \quad (22.22)$$

и L имеет усеченное в нуле распределение Пуассона (см. гл. 4, п. 10):

$$\Pr[L = l] = \frac{(e^\lambda - 1)^{-1} \lambda^l}{l!}, \quad l = 1, 2, \dots, \quad (22.23)$$

то случайная величина

$$X = \max(Y_1, Y_2, \dots, Y_l)$$

имеет распределение, относящееся к распределениям экстремальных значений: функция распределения случайной величины X равна

$$\Pr[X \leq x] = (e^\lambda - 1)^{-1} \exp\left(\left\{\lambda (1 - e^{-x})\right\}\right) = c \exp(-\lambda e^{-x}). \quad (22.24)$$

Подобным же образом распределение Фреше получается с помощью распределений Парето (см. гл. 20), а распределение Вейбулла — из степенного распределения (гл. 20). В статье Sibuya (1967) предложен датчик псевдослучайных чисел, порождающий распределение экстремальных значений, и основанный на соотношениях (22.22) и (22.23), приводящих к равенству (22.24).

4. Функции распределения и моменты

В этом пункте рассматривается только распределение типа 1, определенное формулой (22.1). Из этой формулы находится плотность

$$p_X(x) = \theta^{-1} e^{-(x-\xi)/\theta} \exp \left[-e^{(x-\xi)/\theta} \right]. \quad (22.25)$$

Если $\xi = 0$ и $\theta = 1$ или, что то же самое, рассматривается $Y = (X - \xi)/\theta$, то получается *стандартная форма*

$$p_Y(y) = \exp(-y - e^{-y}). \quad (22.26)$$

В п. 1 отмечено, что случайная величина $Z = \exp[-(X - \xi)/\theta] = e^{-Y}$ имеет экспоненциальное распределение:

$$p_Z(z) = e^{-z}, \quad z \geq 0.$$

Следовательно,

$$E \left[e^{t(X-\xi)/\theta} \right] = E \left[Z^{-t} \right] = \Gamma(1 - t)$$

при $t < 1$. Заменяя t на θt , получаем производящую функцию моментов случайной величины X :

$$E \left[e^{tX} \right] = e^{t\xi} \Gamma(1 - \theta t), \quad \theta |t| < 1. \quad (22.27)$$

Производящая функция семиинвариантов для X имеет вид

$$\psi(t) = \xi t + \log \Gamma(1 - \theta t). \quad (22.28)$$

Семиинварианты случайной величины X таковы:

$$\kappa_1(X) = E[X] = \xi - \theta\psi(1) = \xi + \gamma\theta \approx \xi + 0.57722\theta, \quad (22.29)$$

где γ — постоянная Эйлера,

$$\kappa_r(X) = (-\theta)^r \psi^{(r-1)}(1), \quad r \geq 2. \quad (22.30)$$

В частности,

$$\sigma^2 = \text{var}(X) = \frac{1}{6} \pi^2 \theta^2 \approx 1.64493 \theta^2, \quad (22.31)$$

$$\sigma = \text{Std. dev.}(X) \approx 1.28255 \theta. \quad (22.31)'$$

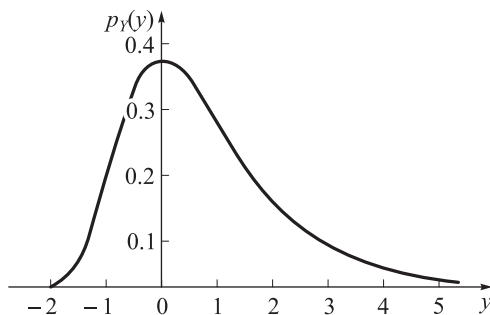


Рис. 22.1. Стандартная плотность распределения типа 1: $p_Y(y) = e^{-y} \exp(-e^{-y})$

Асимметрия и эксцесс равны соответственно

$$\alpha_3^2(X) = \beta_1(X) \approx 1.29857 \quad \text{и} \quad \alpha_4(X) = \beta_2(X) = 5.4. \quad (22.32)$$

Константы ξ и θ суть параметры сдвига и масштаба соответственно. Графики плотностей (22.25) имеют одну и ту же форму.

Рассматриваемое распределение унимодально. Мода находится в точке $x = \xi$, абсциссы точек перегиба суть

$$x = \xi \pm \theta \log \left[\frac{1}{2} (3 + \sqrt{5}) \right] \approx \xi \pm 0.96242\theta. \quad (22.33)$$

При $0 < p < 1$ непосредственно из (22.1) получается p -квантиль, определяемая равенством $F(x_p) = p$:

$$X_p = \xi - \theta \log(-\log p). \quad (22.34)$$

Отсюда находим нижнюю квартиль $X_{0.25}$, медиану $X_{0.5}$ и верхнюю квартиль $X_{0.75}$:

$$X_{0.25} = \xi - \theta \log(\log 4) \approx \xi - 0.32663\theta, \quad (22.35)$$

$$X_{0.5} = \xi - \theta \log(\log 2) \approx \xi + 0.36661\theta, \quad (22.36)$$

$$X_{0.75} = \xi - \theta \log(-\log 0.75) \approx \xi + 1.24590\theta. \quad (22.37)$$

Формула (22.34) позволяет находить квантили с помощью карманного калькулятора. Большая часть стандартного распределения (22.26) сосредоточена в интервале $(-2; 7)$. Для распределения (22.1) вероятность попадания в интервал $(\xi - 2\theta; \xi + 7\theta)$ равна 0.998. Таким образом, 99.8% распределения сосредоточено в интервале $(\mu - 2.0099\sigma; \mu + 5.0078\sigma)$, где μ — математическое ожидание, σ — стандартное отклонение. Более детально свойства распределения описаны в статье Lehman (1963).

График плотности стандартного распределения (22.26) приводится на рис. 22.1. Форма этой кривой напоминает логнормальную плотность при $e^{\sigma^2} = 1.1325$ (в обозначениях гл. 14). Значения β_1 и β_2 этой логнормальной плотности равны 1.300 и 5.398 соответственно [ср. с (22.32)]. В табл. 22.1 сравниваются значения соответствующих стандартных функций распределения.

Таблица 22.2 содержит проценти́ли стандартного распределения с нулевым математическим ожиданием и единичной дисперсией, что отвечает

ТАБЛИЦА 22.1

Сравнение стандартных функций распределения

x	$F(x)$	
	Распределение экстремальных значений типа 1 ^a	Логарифмически нормальное распределение ^b
−2.0	0.00068	0.00022
−1.5	0.02140	0.01959
−1.0	0.1321	0.1342
−0.5	0.3443	0.3471
0.0	0.5704	0.5700
0.5	0.7440	0.7423
1.0	0.8558	0.8546
1.5	0.92237	0.92096
2.0	0.95774	0.95792
2.5	0.97752	0.97730
3.0	0.98810	0.98837
3.5	0.99371	0.99389
4.0	0.99668	0.99677

$$^a F(x) = \exp[-\exp(-1.28524x - 0.57722)].$$

$$^b F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u(x)} \exp(-u^2/2) du, \text{ где } u(x) = 6.5277 \lg(x + 2.74721) - 2.68853.$$

ТАБЛИЦА 22.2

Процентили стандартизированного распределения экстремальных значений типа 1

α	Процентиль	α	Процентиль
0.0005	−2.0325	0.75	0.5214
0.0001	−1.9569	0.9	1.3046
0.0025	−1.8460	0.95	1.8658
0.005	−1.7501	0.975	2.4163
0.01	−1.6408	0.99	3.1367
0.025	−1.4678	0.995	3.6791
0.05	−1.3055	0.9975	4.2205
0.1	−1.1004	0.999	4.9355
0.25	−0.7074	0.9995	5.4761
0.5	−0.1643		

[...]