

УДК 519.22./25

В.Ю. Дубницький, А.И. Ходырев

Харьковский институт банковского дела Университета банковского дела НБУ, Харьков

ОЦЕНКА ПАРАМЕТРОВ РАНГОВЫХ РАСПРЕДЕЛЕНИЙ: РАСПРЕДЕЛЕНИЕ МАНДЕЛЬБРОТА-ЦИПФА

Получены оценки параметра распределения Ципфа-Мандельброта методом моментов и методом максимума правдоподобия.

Ключевые слова: ранговые распределения, распределение Ципфа, распределение Ципфа-Мандельброта, метод моментов, метод максимума правдоподобия.

Постановка проблемы

Классические распределения, используемые в статистике, относятся к классу частотных. Меньшее распространение получили так называемые ранговые распределения, нашедшие применение в математической лингвистике [1], теории фракталов [2], социологии [3], биологии [4], экономике [5].

Анализ последних исследований и публикаций. Подробно основные понятия ранговых распределений изложены в работе [6]. Сохраняя принятые в ней обозначения, коротко изложим основные определения этой теории.

Частотной интегральной формой выборочного распределения назовем условия вида

$$\sum_{x_0}^I n(x) = N, \quad (1)$$

где x_0 , I – минимальные и максимальные значения случайной величины $X = x$ в выборке объема N ; $n(x)$ – частота появления случайной величины x . Частотная функция распределения примет вид:

$$F(x) = \frac{1}{N} \sum_{x_0}^x n(\xi). \quad (2)$$

Ранговая дифференциальная форма выборочно-статистического распределения отвечает условию

$$r = \sum_x^I n(\xi), \quad 1 \leq r \leq N.$$

Ранговая интегральная форма аналогичного распределения имеет вид:

$$X(r) = \sum_1^r x(\xi); \quad \sum_1^N x(\xi) = G. \quad (3)$$

Ранг r – это порядковый номер данного значения случайной величины x , когда эти значения расположены в порядке убывания частоты их встречаемости $n(x)$.

В качестве примера получения рангового распределения из исходного частотного рассмотрим задачи №№ 635, 636 из работы [7]. В этой работе показано, что распределение частот в задаче № 635 не соответствует закону Гаусса (нормальному), в задаче № 636 – соответствует.

Исходные данные для этих задач приведены в табл. 1 и 2.

Столбцы 2, 3, 6, 7 взяты непосредственно из условия задачи. Используя критерий χ^2 , следует принять гипотезу о непротиворечивости исходных данных нормальному распределению. В столбцах 4, 8 исходная совокупность проранжирована по убыванию, то есть, если $x_i \geq x_j$, то $r_i \leq r_j$. В том случае, когда $f_i = f_j$, то ранги r_i , r_j определяют по схеме со вмещенных рангов [8].

Гистограмма частотного распределения для данных, приведенных в табл. 1, показана на рис. 1, Гистограмма рангового распределения, соответствующего приведенному частотному, показана на рис. 2.

Таблица 1

Исходные данные для получения рангового распределения, соответствующего условию задачи № 635

i	Переменная x_i	Частота f_i	Ранг r_i	i	Переменная x_i	Частота f_i	Ранг r_i
1	2	3	4	5	6	7	8
1	5	15	8	6	15	21	6
2	7	26	2,5	7	17	24	5
3	9	25	4	8	19	20	7
4	11	30	1	9	21	13	9
5	13	26	2,5	–	–	–	–

Исходные данные для получения рангового распределения, соответствующего условию задачи № 636

i	Переменная x_i	Частота f_i	Ранг r_i	i	Переменная x_i	Частота f_i	Ранг r_i
1	2	3	4	5	6	7	8
1	0,3	6	10	7	1,5	21	6
2	0,5	9	8	8	1,7	24	5
3	0,7	26	2,5	9	1,9	20	7
4	0,9	25	4	10	2,1	8	9
5	1,1	30	1	11	2,3	5	11
6	1,3	26	2,5	–	–	–	–

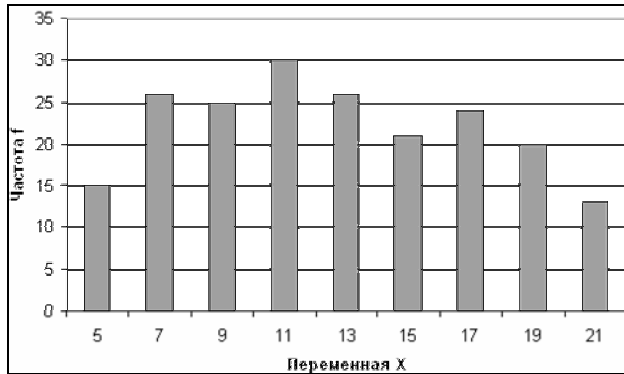


Рис. 1. Гистограмма частотного распределения для данных, приведенных в табл. 1

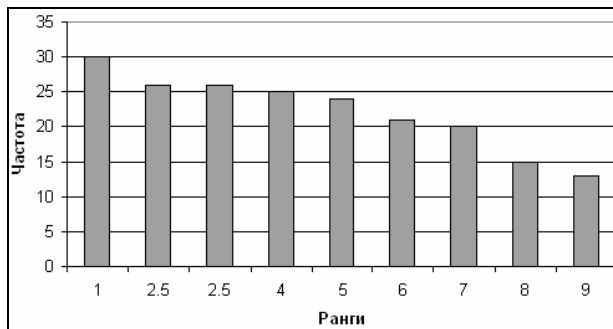


Рис. 2. Гистограмма рангового распределения для данных, приведенных в табл. 1

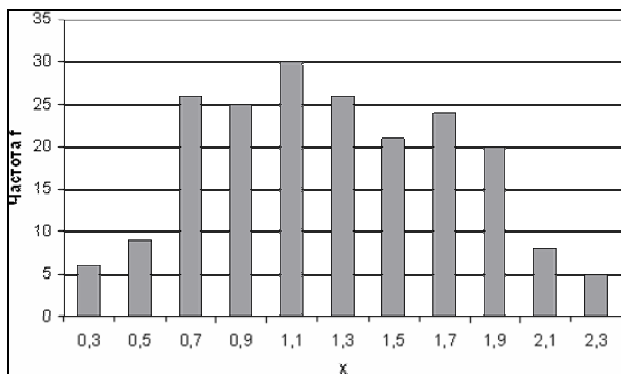


Рис. 3. Гистограмма частотного распределения для данных, приведенных в табл. 2

Гистограмма частотного распределения для данных, приведенных в табл. 2 показана на рис. 3,

гистограмма рангового распределения, соответствующего приведенному частотному, показана на рис. 4.

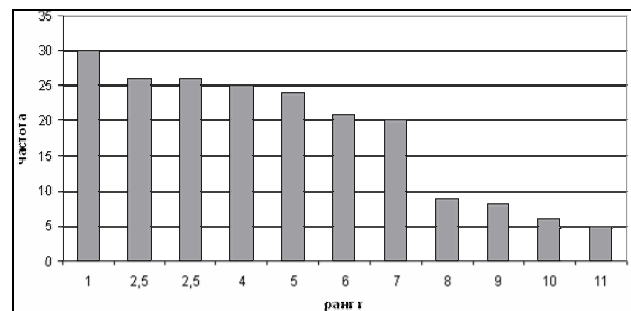


Рис. 4. Гистограмма рангового распределения для данных, приведенных в табл. 2

Ранговые распределения для условий каждой из задач с высокой точностью могут быть записаны в виде:

$$f_i = \frac{A}{r_i^\gamma}, \quad (4)$$

где r_i – ранг i -ой частоты; f_i – ее численное значение; A – эмпирические константы. Их определение выполняли в два этапа. На первом этапе для зависимости

$$\ln f_i = \ln A - \gamma \ln r_i. \quad (5)$$

Величины $\ln A$ и γ определяли стандартным методом наименьших квадратов. На втором этапе полученные значения использовали в качестве начальных для процедуры нелинейной регрессии. Полученные результаты определения величин A и γ приведены в табл. 3.

Таблица 3

Результаты определения параметров рангового закона распределения

Оцениваемые параметры		Задача	
		635	636
A	Линеаризованная форма	3,5402	3,8913
	Нелинейная форма	32,1756	34,9271
γ	Линеаризованная форма	0,3298	0,7280
	Нелинейная форма	0,2702	0,4300

Изложение результатов исследования

Рассмотрим распределение Ципфа в виде, приведенном в [9, с. 53], то есть в случае, когда случайная величина X непрерывна.

Плотность этого распределения имеет вид:

$$f(x) = \frac{c}{x^{1+\alpha}}; \quad x \geq x_0 > 0. \quad (5)$$

Тогда функцию распределения определим как

$$F(x) = \int_{x_0}^x \frac{c}{x^{1+\alpha}} dx = C \left(\frac{1}{\alpha x_0^\alpha} - \frac{1}{\alpha x^\alpha} \right); \quad x \geq x_0 > 0. \quad (6)$$

Из условия нормировки

$$\int_{x_0}^{\infty} \frac{c}{x^{1+\alpha}} dx = 1, \quad (7)$$

следует, что

$$c = \alpha x_0^\alpha. \quad (8)$$

Тогда математическое ожидание m_x случайной величины X , распределенной по закону Ципфа в виде (5), примет вид:

$$m_x = \int_{x_0}^{\infty} x \frac{\alpha x_0^\alpha}{x^{1+\alpha}} dx = \frac{\alpha x_0}{\alpha - 1}; \quad \alpha > 1. \quad (9)$$

Второй начальный момент:

$$M[x^2] = \int_{x_0}^{\infty} x^2 \frac{\alpha x_0^\alpha}{x^{1+\alpha}} dx = \frac{\alpha x_0^2}{\alpha - 2}; \quad \alpha > 2. \quad (10)$$

Дисперсия данного распределения с учетом (9) и (10):

$$\sigma_x^2 = M[x^2] - m_x^2 = \frac{\alpha x_0^2}{(\alpha - 2)(\alpha - 1)^2}; \quad \alpha > 0. \quad (11)$$

Заменяя в (9) и (11) значения математического ожидания m_x и дисперсии σ^2 их оценками \bar{x} и S^2 соответственно, получим систему уравнений вида

$$\begin{cases} \bar{x} = \frac{\alpha x_0}{\alpha - 1}; \\ S^2 = \frac{\alpha x_0^2}{(\alpha - 2)(\alpha - 1)}. \end{cases} \quad (12)$$

$$\quad (13)$$

Решив эту систему относительно α и x_0 , получим их оценки $\hat{\alpha}$ и \hat{x}_0 . На первом шаге разрешим (12) относительно α :

$$\hat{\alpha} = \frac{\bar{x}}{\bar{x} - x_0}. \quad (14)$$

Подставим (10) в (9) и получим, что

$$x_0^{(1)} = \frac{\sqrt{(\bar{x})^4 + 4(S^2)^2} - (\bar{x})^2 + 2S^2}{2\bar{x}}; \quad (15)$$

$$x_0^{(2)} = \frac{(\bar{x})^4 + 4(S^2)^2 + (\bar{x})^2 - 2S^2}{2\bar{x}}. \quad (16)$$

Из решений (15) и (16) выбираем то, при котором $x_0^{(k)}$, $k = 1, 2$ положительно, затем из (14) получим оценку $\hat{\alpha}$ величины α .

Если величина x_0 известна из физического смысла задачи, а на практике чаще всего так и бывает, то оценку параметра α можно получить методом максимума правдоподобия.

В этом случае для плотности вида (5) функция правдоподобия примет вид:

$$L = \prod_{i=1}^n \frac{\alpha x_0^\alpha}{x_i^{1+\alpha}} = \frac{(\alpha x_0^\alpha)^n}{\prod_{i=1}^n x_i^{1+\alpha}}. \quad (17)$$

Тогда

$$\ln L = n \ln(\alpha x_0^\alpha) - (1 + \alpha) \sum_{i=1}^n \ln x_i. \quad (18)$$

Следовательно,

$$\frac{d \ln L}{d \alpha} = n \left(\frac{1}{\alpha} + \ln x_0 \right) - \sum_{i=1}^n \ln x_i = 0. \quad (19)$$

Таким образом, оценка

$$\hat{\alpha} = \frac{n}{\sum \ln x_i - n \ln x_0}. \quad (20)$$

В некоторых из указанных выше работ, например в [2], использовано расширение этого закона, известное как закон Мандельброта-Ципфа, плотность которого имеет вид:

$$f(x) = \begin{cases} \frac{C}{(x+B)^{1+\alpha}}, & x \geq x_0 > 0; \\ 0, & x < x_0, \end{cases} \quad (21)$$

однако параметры этого закона находят, решая задачу нелинейной регрессии.

Определим параметры закона Мандельброта-Ципфа методом моментов. Из условия нормировки

$$f(x) = \int_{x_0}^{\infty} \frac{c}{(x+B)^{1+\alpha}} dx = 1 \quad (22)$$

следует, что

$$C = \alpha (x_0 + B)^\alpha. \quad (23)$$

Тогда функция распределения примет вид:

$$F(x) = \int_{x_0}^x \frac{\alpha (x_0 + B)^\alpha}{(x+B)^{\alpha+1}} dx = 1 - \left(\frac{x_0 + B}{x+B} \right). \quad (24)$$

Математическое ожидание случайной величины X , распределенной по закону Мандельброта-Ципфа, вычислим из условия

$$M[x] = \int_{x_0}^{\infty} \frac{x \alpha (x_0 + B)^\alpha}{(x+B)^{\alpha+1}} dx = \frac{\alpha x_0 + B}{\alpha - 1}, \quad \alpha > 1. \quad (25)$$

Второй начальный момент

$$M[x] = \int_{x_0}^{\infty} \frac{x\alpha(x_0+B)^\alpha}{(x+B)^{\alpha+1}} dx = \frac{2(B+x_0^2)}{\alpha-2} - \frac{2B(B+x_0)}{\alpha-1} + x_0^2, \quad \alpha > 2. \quad (26)$$

Следовательно, дисперсию закона Мандельброта-Ципфа определим из условия

$$D[x] = M[x^2] - M[x]^2 = \frac{\alpha(x_0+B)^2}{(\alpha-2)(\alpha-1)^2}, \quad \alpha > 2. \quad (27)$$

Таким образом, оценки параметров α и B можно определить, решив систему уравнений

$$\begin{cases} M[X] = \frac{\alpha x_0 + B}{\alpha - 1}; \\ D[X] = \frac{\alpha(x_0 + B)^2}{(\alpha - 2)(\alpha - 1)^2}. \end{cases} \quad (28)$$

Определим параметры закона Мандельброта-Ципфа методом максимума правдоподобия. Функцию правдоподобия представим в виде

$$L(x) = \prod_{i=1}^n \frac{\alpha(x_0+B)^\alpha}{(x_i+B)^{\alpha+1}} = \frac{(\alpha(x_0+B)^\alpha)^n}{\prod_{i=1}^n (x_i+B)^{\alpha+1}}. \quad (29)$$

Логарифм функции правдоподобия примет вид:

$$\ln L = n \ln \alpha + n\alpha \ln(x_0+B) - (\alpha-1) \sum_{i=1}^n \ln(x_i+B). \quad (30)$$

Параметры α и x_0 определим в результате решения системы уравнений:

$$\begin{cases} \frac{\partial L}{\partial \alpha} = \frac{n}{\alpha} + n \ln(x_0+B) - \sum_{i=1}^n \ln(x_i+B) = 0; \\ \frac{\partial L}{\partial B} = \frac{n\alpha}{x_0+B} - (\alpha+1) \cdot \sum_{i=1}^n (x_i+B)^{-1} = 0. \end{cases} \quad (31)$$

В том случае, когда случайная величина $X = 1, 2, \dots$ дискретна и распределена по закону Ципфа, то определение параметров этого закона усложняется.

Вероятность Pr того, что возможное значение случайной величины $X = x$, определим так [10]:

$$Pr[X = x] = x^{-(\rho+1)}, \quad x = 1, 2, \dots; \quad \rho > 0. \quad (32)$$

Исходя из условия нормировки

$$\sum_{x=1}^{\infty} Pr[X = x] = 1, \quad (33)$$

получим, что

$$p(X = x) = \frac{1}{\zeta(\rho+1)} x^{-(\rho+1)}, \quad x = 1, 2, \dots, \quad \rho > 0, \quad (34)$$

где

$$\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha} -$$

дзета-функция Римана[11].

Функция распределения в этом случае примет вид:

$$F(x) = \begin{cases} 0, & x \leq 1; \\ \frac{1}{\zeta(\rho+1)} \sum_{i=1}^k i^{-(\rho+1)} & k < x \leq k+1; \\ \zeta(\rho+1) & k = 1, 2, \dots \end{cases} \quad (35)$$

В работе [10] показано, что математическое ожидание случайно величины X имеет вид

$$\bar{x} = \frac{\zeta(\rho)}{\zeta(\rho+1)}, \quad \rho > 1 \quad (36)$$

и, следовательно, есть функция параметра ρ .

В этой же работе приведены оценки ρ^* параметра ρ , полученные методом моментов и методом максимального правдоподобия.

Оценку $\hat{\rho}$ параметра ρ методом моментов получают, решив уравнение:

$$\frac{\zeta(\rho^*)}{\zeta(\rho^*+1)} = \bar{x}^*, \quad (37)$$

где \bar{x}^* – оценка среднего значения величины X , полученная по статистическим данным.

Для получения оценки ρ параметра ρ^* по методу максимального правдоподобия следует решить уравнение

$$-\frac{\zeta'(\rho^*+1)}{\zeta(\rho^*+1)} = \frac{1}{n} \sum_{i=1}^k \ln x_i. \quad (38)$$

Для облегчения процесса поиска корней уравнений (37) и (38) в работе [10] приведены вспомогательные таблицы. С целью автоматизации процесса поиска корня уравнения (37) или (38) нами получены аппроксимации указанных таблиц. Для уравнения (37) аппроксимирующее уравнение получено в виде зависимости

$$\rho^* = \frac{1}{\alpha - \frac{b}{\bar{x}}}. \quad (39)$$

Для уравнения (38) аппроксимирующее уравнение получено в виде зависимости

$$\rho^* = -a + \frac{b}{w}, \quad (40)$$

где $w = \frac{1}{n} \sum_{i=1}^k \ln x_i$.

В том случае, когда величина $w > 4$, в работе [10] рекомендована формула:

$$\rho = \frac{\ln 2}{1 + 2^{w+1}}. \quad (41)$$

Таблиця 4

Коефіцієнти апроксимуючого уравнения для оценки ρ^* параметра ρ по методу моментов

Интервал изменения величины \bar{x}	Параметры эмпирического уравнения		Средняя абсолютная ошибка аппроксимации
	а	в	
1	2	3	4
1,01 ... 1,20	1,6350	1,4495	0,012
1,21 ... 1,40	1,1521	0,8902	$0,6 \cdot 10^{-3}$
1,41 ... 1,60	1,0809	0,7932	$0,1 \cdot 10^{-3}$
1,61 ... 1,80	1,0495	0,7432	$0,1 \cdot 10^{-3}$
1,81 ... 2,00	1,0311	0,7097	$0,19 \cdot 10^{-3}$

Таблиця 5

Коефіцієнти апроксимуючого уравнения для оценки ρ^* параметра ρ по методу максимума правдоподобия

Интервал изменения величины w	Параметры эмпирического уравнения		Средняя абсолютная ошибка аппроксимации
	а	в	
1	2	3	4
0,1...1(0,1)	-0,45	0,989	0,02
1,1... 2(0,1)	-0,2385	0,7959	0,0025
2,2 ... 4(0,2)	-0,125	0,5681	0,0038

Выводы

1. Показан способ получения рангового распределения из частотного.
2. Получены оценки параметра распределения Ципфа-Мандельброта методом моментов и методом максимума правдоподобия.
3. Построены апроксимирующие выражения для определения параметра распределения Ципфа-Мандельброта методом моментов и методом максимума правдоподобия по специальным таблицам.

Список литературы

1. Орлов Ю.К. Обобщенный закон Ципфа-Мандельброта и частотные структуры единиц уровней / Ю.К. Орлов // *Вычислительная лингвистика*. – М.: Наука, 1976. – С. 179-202.
2. Шредер М. Фракталы, хаос, степенные законы / М. Шредер. – Ижевск.: РХД, 2005. – 528 с.
3. Фурашев В.Н. Моделирование информационно-электронных процессов / В.Н. Фурашев, Д.В. Ландэ, С.М. Брайчевский. – К.: НИЦПИ АПрН Украины, 2007. – 544 с.
4. Суховельский В.Г. Методы количественной оценки видовой разнообразия сообществ / В.Г. Суховельский,

Ю.Н. Баранчиков // *Энтомологические исследования в Сибири*. – 1998. – Вып. 1. – С. 44-54.

5. Андреев М.Ю. Моделирование деятельности современной российской банковской системы / М.Ю. Андреев, Н.П. Пильник, И.Г. Поспелов // *Экономический журнал высшей школы экономики*. – 2009. – № 2. – С. 143-171.
6. Хайтун С.Д. Наукометрия: Состояние и перспективы / С.Д. Хайтун. – М.: Наука, 1983. – 280 с.
7. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике / В.Е. Гмурман. – М.: Высшая школа, 1979. – 400 с.
8. Кендалл М. Ранговые корреляции / М. Кендалл. – М.: Статистика, 1975. – 216 с.
9. Переверзев Е.С. Случайные сигналы в задачах оценки состояния технических систем / Е.С. Переверзев, Ю.Ф. Даниев, Г.П. Филей. – К.: Наукова думка, 1992. – 252 с.
10. Вадзинский Р.Н. Справочник по вероятностным распределениям / Р.Н. Вадзинский. – СПб.: Наука, 2001. – 295 с.
11. Уиттекер Э.Т. Курс современного анализа: в 2т. / Э.Т. Уиттекер, Дж.Н. Ватсон. – М.: Физматгиз, 1963. – Т. 2. – 512 с.

Поступила в редколлегию 27.04.2010

Рецензент: д-р физ.-мат. наук, проф. М.В. Новожилова, Харьковский государственный технический университет строительства и архитектуры, Харьков.

ОЦІНКА ПАРАМЕТРІВ РАНГОВОГО РОЗПОДІЛУ МАНДЕЛЬБРОТА-ЦІПФА

В.Ю. Дубницький, О.І. Ходирев

Одержані оцінки параметра розподілу Ціпфа-Мандельброта методом моментів та методом максимуму правдоподібності.

Ключові слова: рангові розподіли, розподіл Ціпфа, розподіл Ціпфа-Мандельброта, метод моментів, метод максимуму правдоподібності.

EVALUATION OF MANDELBRÖT-ZIPF RANK DISTRIBUTION PARAMETERS

V.Y. Dubnitsky, A.I. Khodyrev

Evaluations of Zipf-Mandelbrot distribution parameter were obtained by method of moments and method of maximum likelihood.

Keywords: rank distributions, Zipf distribution, Zipf-Mandelbrot distribution, method of moments, method of maximum likelihood.