

## ИССЛЕДОВАНИЕ РАНГОВЫХ РАСПРЕДЕЛЕНИЙ

В. В. Нешитой

В информатике и математической лингвистике чрезвычайно широкое распространение получили так называемые ранговые распределения, примерами которых являются: распределение разных слов по частоте их употребления в текстах (слова упорядочены по невозрастанию частот), распределение периодических изданий по количеству помещенных в них статей, распределение научных сотрудников по продуктивности.

В настоящей статье ставятся следующие задачи:

- выработать наиболее целесообразную форму представления ранговых распределений,
- установить критерий однородности ранговых распределений,
- рассчитать необходимый объем выборки (для установления типа выравнивающей кривой распределения),
- найти общую формулу для закона Брэдфорда,
- установить место закона Ципфа в системе непрерывных распределений,
- проверить соответствие закона Ципфа опытным данным,
- дать метод оценивания параметров выравнивающих распределений,
- найти выравнивающие распределения для некоторых частных случаев.

Перечисленные задачи решаются на основе свойств обобщающих систем непрерывных распределений, построенных автором.

### 1. ТРИ СИСТЕМЫ НЕПРЕРЫВНЫХ РАСПРЕДЕЛЕНИЙ И НЕКОТОРЫЕ ИХ СВОЙСТВА

Пусть одна из систем непрерывных распределений задается обобщенной плотностью вида

$$p(t) = N t^{\nu-1} (1 - \alpha u t^{\beta})^{\frac{1}{u}-1}, \quad (1)$$

где  $N$  — нормирующий множитель,  $\alpha$  — масштабный параметр,  $\beta, \gamma, u$  — параметры формы,  $t$  — расстояние от начала отсчета значений случайной величины  $T$ .

Нормирующий множитель  $N$  выражается через гамма-функцию и, соответственно, равен:

— для распределений I, I' типов (классификацию распределений см. в [1])

$$N = \frac{\beta \Gamma\left(\frac{\gamma}{\beta} + \frac{1}{u}\right)}{\left(\frac{1}{\alpha u}\right)^{\nu/\beta} \Gamma\left(\frac{\gamma}{\beta}\right) \Gamma\left(\frac{1}{u}\right)} \quad (\alpha, \beta, \gamma, u > 0); \quad (2)$$

— для распределений II, II' типов

$$N = \frac{\beta \alpha^{\nu/\beta}}{\Gamma\left(\frac{\gamma}{\beta}\right)} \quad (\alpha, \beta, \gamma > 0, u \rightarrow 0); \quad (3)$$

— для распределений III—V типов

$$N = \frac{\beta \Gamma\left(1 - \frac{1}{u}\right)}{\left(-\frac{1}{\alpha u}\right)^{\nu/\beta} \Gamma\left(\frac{\gamma}{\beta}\right) \Gamma\left(1 - \frac{1}{u} - \frac{\gamma}{\beta}\right)} \quad (\alpha u < 0, 0 < \frac{\gamma}{\beta} < 1 - \frac{1}{u}). \quad (4)$$

Данная система распределений предоставляет широкие возможности для описания ранговых (невозрастающих) распределений, поскольку кривые распределения при определенных значениях  $\beta, \gamma, u$  являются невозрастающими. Формы кривых распределения разных типов исследованы в работе [2].

Пусть, далее, случайная переменная  $X$  связана со случайным аргументом  $T$  зависимостью  $X = \ln T$ . Тогда плотность распределения  $p(x)$  выразится формулой

$$p(x) = N e^{\nu x} (1 - \alpha u e^{\beta x})^{\frac{1}{u}-1}, \quad (5)$$

где нормирующий множитель  $N$ , в зависимости от типа кривой распределения, определяется формулами (2)—(4).

Рассмотрим основные свойства кривых распределения, принадлежащих данной системе. Плотность (5) содержит два параметра формы  $k = \gamma/\beta, u$ , масштабный параметр  $\beta$  и параметр сдвига  $\alpha$ . Кривые распределения в данном случае имеют фактически одну форму: левая часть кривой при  $x \rightarrow -\infty$  асимптотически приближается к оси абсцисс, правая часть может либо асимптотически приближаться к оси абсцисс при  $x \rightarrow \infty$ , либо опускаться на нее под определенным углом, если величина  $x$  ограничена справа. Кривые распределения, заданные плотностью (5), имеют три характерных точки:  $A, C, B$ .  $A$  и  $B$  — точки перегиба,  $C$  — мода. Абсциссы этих точек, соответственно, равны

$$x_C = \frac{1}{\beta} \ln \frac{k}{\alpha(1 + ku - u)}, \quad (6)$$

$$x_{A,B} = \frac{1}{\beta} \ln \frac{2k(1 + ku - u) + (1 - u) \mp \sqrt{4k(1 + ku - u) + (1 - u)}(1 - u)}{2\alpha(1 + ku - u)^2}. \quad (7)$$

Из (6), (7) следует соотношение

$$x_B - x_C = x_C - x_A = \frac{1}{\beta} \ln R, \quad (8)$$

где

$$R = e^{f(x_B - x_C)} = e^{f(x_C - x_A)} = 1 + \frac{1-u + \sqrt{[4k(1+ku-u) + (1-u)](1-u)}}{2k(1+ku-u)}, \quad (9)$$

т. е. точки перегиба  $A$  и  $B$  располагаются симметрично относительно моды  $C$ .

Между ординатами трех характерных точек существует соотношение

$$\frac{[p(x_C)]^2}{p(x_A)p(x_B)} = \left(\frac{1-2u}{1-u}\right)^{1-\frac{1}{u}} = \left(1 + \frac{1}{1-\frac{1}{u}}\right)^{1-\frac{1}{u}} = \lambda, \quad (10)$$

из которого видно, что величина  $\lambda$  является функцией одного параметра формы  $u$ . Распределения, заданные плотностью (5), можно — в зависимости от значений  $\lambda$  — разделить на типы: I,  $I' - \lambda > e \approx 2,72$ ; II,  $II' - \lambda = e$ ; III —  $2 < \lambda < e$ ; IV —  $\lambda = 2$ ; V —  $1 < \lambda < 2$ . Таким образом, по координатам трех характерных точек  $A$ ,  $C$ ,  $B$  может быть установлен тип кривой распределения и найдено значение параметра  $u$  (из соотношения (10) при заданном  $\lambda$ ). Однако на практике это реализовать трудно, и тип выравнивающей кривой распределения мы будем устанавливать другим способом.

Анализ формул (6)–(10) показывает, что у распределений II–V типов существуют все три точки  $A$ ,  $C$ ,  $B$ . У распределений I типа точка перегиба  $B$  существует при  $0 < u < 1/2$ , точки  $A$ ,  $C$  — при  $0 < u < 1$ .

После рассмотрения основных свойств распределений, заданных плотностью (5), можно сделать некоторые выводы: если эмпирическое распределение, полученное в результате статистической обработки однородных случайных величин, принадлежит рассмотренному классу распределений, то оно должно обладать всеми их свойствами (т. е. должно иметь одну четко выраженную моду, одну или две точки перегиба, расположенные на равных расстояниях по обе стороны от моды, и т. д.). Но какое отношение к рассмотренным распределениям имеют ранговые невозрастающие распределения?

Ответ на поставленный вопрос заключается в следующем: ранговые невозрастающие распределения можно привести к форме распределений с плотностью (5) и тогда они проявят все свойства последних!

Запишем плотность (1) в виде

$$tp(t) = Ne^{v \ln t} (1 - \alpha u e^{\beta \ln t})^{\frac{1}{u} - 1}. \quad (11)$$

Обозначив, далее,  $tp(t)$  через  $p(x)$ ,  $\ln t$  через  $x$ , из (11) будем иметь (5).

Итак, чтобы привести распределение (в том числе ранговое), заданное плотностью (1), к форме (5), необходимо построить по имеющимся значениям  $t$ ,  $p(t)$  график зависимости  $tp(t)$  от  $\ln t$ . Построенная в данной системе координат кривая будет представлять собой кривую распределения, заданную плотностью (5). Для этой кривой распределения будут выполняться условия (6)–(10) (с учетом равенств  $tp(t) = p(x)$ ,  $\ln t = x$ ). Например, равенство (6) примет вид:

$$x_C = \ln t_C = \frac{1}{\beta} \ln \frac{k}{\alpha(1+ku-u)},$$

откуда

$$t_C = \left[ \frac{k}{\alpha(1+ku-u)} \right]^{1/\beta}, \quad (12)$$

Таким же путем найдем из (7):

$$t_{A,B} = \left[ \frac{2k(1+ku-u) + (1-u) \mp \sqrt{[4k(1+ku-u) + (1-u)](1-u)}}{2\alpha(1+ku-u)^2} \right]^{1/\beta}. \quad (13)$$

Условие (8) переписывается в виде

$$\ln t_B - \ln t_C = \ln t_C - \ln t_A = \frac{1}{\beta} \ln R,$$

или

$$\frac{t_B}{t_C} = \frac{t_C}{t_A} = R^{1/\beta}, \quad (14)$$

где  $R$  вычисляется по формуле (9).

Выражение (14) представляет собой не что иное как закон рассеяния информации Брэдфорда, полученный как следствие свойств кривых распределения, заданных обобщенной плотностью (1). Из (9) следует, что величина  $R$  зависит от двух параметров формы  $k$ ,  $u$ . Отношения же  $t_B/t_C$ ,  $t_C/t_A$  зависят еще от параметра  $\beta$ . Интервал  $0 < t < t_A$  представляет собой ядро, а зонами рассеяния являются:  $t_A < t < t_C$  — первая зона;  $t_C < t < t_B$  — вторая зона;  $t > t_B$  — третья зона. В случае распределений II–V типов существует ядро и три зоны рассеяния; в случае распределений I типа количество зон рассеяния может быть меньше.

Таким образом, количество зон рассеяния, а также значения  $R^{1/\beta}$  зависят от типа выравнивающей кривой распределения (и, следовательно, от значений  $\beta$ ,  $k$ ,  $u$ ), которая описывает множество журналов, упорядоченных по невозрастанию количества помещенных в них статей по данному вопросу.

Для того чтобы проверить, выполняются ли условия (6)–(10) для рангового распределения  $p_r = f(r)$ , где  $p_r$  — относительная частота события (журнала, словоформы и т. д.) с рангом  $r$ , достаточно построить график зависимости  $rp_r$  от  $\ln r$ . В случае однородности рангового распределения должна получиться одновершинная кривая, при этом точки перегиба  $A$  и  $B$  должны расположиться симметрично относительно моды  $C$ . При выполнении этих условий может быть найдено выравнивающее распределение, заданное плотностью (1). Из этого же графика по координатам трех характерных точек приближенно можно оценить размеры ядра и зон рассеяния. Более точные координаты указанных точек могут быть рассчитаны по приведенным выше формулам при известном выравнивающем распределении.

Рассмотрим, далее, класс непрерывных логарифмических распределений, заданных плотностью

$$p(y) = \frac{N}{y} (\ln y)^{\nu-1} [1 - \alpha u (\ln y)^\beta]^{\frac{1}{u} - 1}. \quad (15)$$

Данное распределение следует из (1) при  $T = \ln Y$ . Если предполагается, что эмпирическое ранговое распределение может быть описано плотностью (15), то оно должно быть приведено к форме (5) путем построения графика зависимости  $rp_r \ln r$  от  $\ln \ln r$ . Если предположение верно, то для построенной кривой будут выполняться условия (6)–(10) (с учетом равенств  $rp_r \ln r = p(x)$ ,  $\ln \ln r = x$ ).

Отметим, что обобщенная плотность (15) при  $u \rightarrow 0$ ,  $\beta = 2$ ,  $\nu = 1$ ,  $\alpha > 0$  переходит в плотность логнормального закона

$$p(y) = \sqrt{\frac{\alpha}{\pi}} \frac{1}{y e^{\alpha (\ln y)^2}} = \sqrt{\frac{\alpha}{\pi}} \frac{1}{y^{1+\alpha \ln y}} \quad (0 < y < \infty).$$

Укажем еще на один частный случай при  $\alpha > 0$ ,  $\beta$ ,  $\nu$ ,  $u = 1$ :

$$p(y) = \frac{\alpha}{y} (1 < y < e^{1/\alpha}).$$

Последнее распределение представляет собой непрерывный аналог закона Ципфа:  $p_r = C/r$ . Следовательно, закон Ципфа относится к семейству логарифмических распределений и является весьма частным его случаем.

## 2. ОЦЕНКА ПАРАМЕТРОВ

Рассмотрим два показателя, характеризующих островершинность и асимметрию обобщенных распределений, приведенных к форме (5).

В качестве показателя островершинности может служить величина [3]:

$$H_r = \frac{S_r}{S_1^r}, \quad (16)$$

где

$$S_r = M[p(x)]^r. \quad (17)$$

Величина  $S_r$  является функцией трех параметров  $\beta, k, u$ . Отношение же  $S_r/S_1^r$  зависит от двух параметров  $k, u$ . В частности, при  $u \rightarrow 0, 0 < k < \infty$  (т. е.)

в случае распределений II типа)  $\sqrt{\frac{2r}{r+1}} < H_r < \frac{2r}{r+1}$ . Принимая  $r=3$ , будем иметь  $\sqrt{2} < H_3 < 2$ .

В качестве показателя асимметрии обобщенных распределений, приведенных к форме (5), может служить величина, задаваемая формулой

$$B_1 = M[p(x)(x - M(x))]. \quad (18)$$

Величина  $B_1$  также зависит от двух параметров формы  $k, u$ . В случае распределений II—V, II' типов она определена на интервале  $-0,25 < B_1 < 0,25$ .

Формулы (16)—(18), выраженные через плотность (1), примут вид

$$H_r = \frac{S_r}{S_1^r}, \quad S_r = M[tp(t)]^r,$$

$$B_1 = M[tp(t)(\ln t - M(\ln t))],$$

поскольку  $p(x) = tp(t), x = \ln t$ .

На основании формул (16)—(18) были рассчитаны таблицы\* значений  $B_1, H_3$  при различных значениях параметров  $k, u$ ; затем был построен график (см. рис. 1). С помощью рис. 1 при известных эмпирических значениях показателей  $B_1, H_3$  (обозначим их  $B_1^*, H_3^*$ ) могут быть оценены параметры  $k, u$ .

Оценки параметров  $\beta, \alpha$  (или произведения  $\alpha u$ ) находятся по эмпирическим значениям величин  $S_1, v_1$  (т. е.  $S_1^*, v_1^*$ ):

$$\beta = \frac{S_1^*}{S_1^{(z)}} \text{ — для всех типов,} \quad (19)$$

$$\left. \begin{aligned} \alpha u &= e^{\pm(v_1^{(z)} - \beta v_1^*)} \text{ — для I, I' типов,} \\ \alpha &= e^{\pm(v_1^{(z)} - \beta v_1^*)} \text{ — для II, II' типов,} \\ \alpha u &= -e^{v_1^{(z)} - \beta v_1^*} \text{ — для III—V типов,} \\ h^\beta &= -\frac{1}{\alpha u} = \frac{1}{e^{v_1^{(z)} - \beta v_1^*}} \text{ — для IV типа,} \end{aligned} \right\} \quad (20)$$

\* Программы для расчета таблиц на мини-ЭВМ СМ-4 написаны сотрудником Вычислительного центра Белорусского политехнического института И. А. Бухваловой.

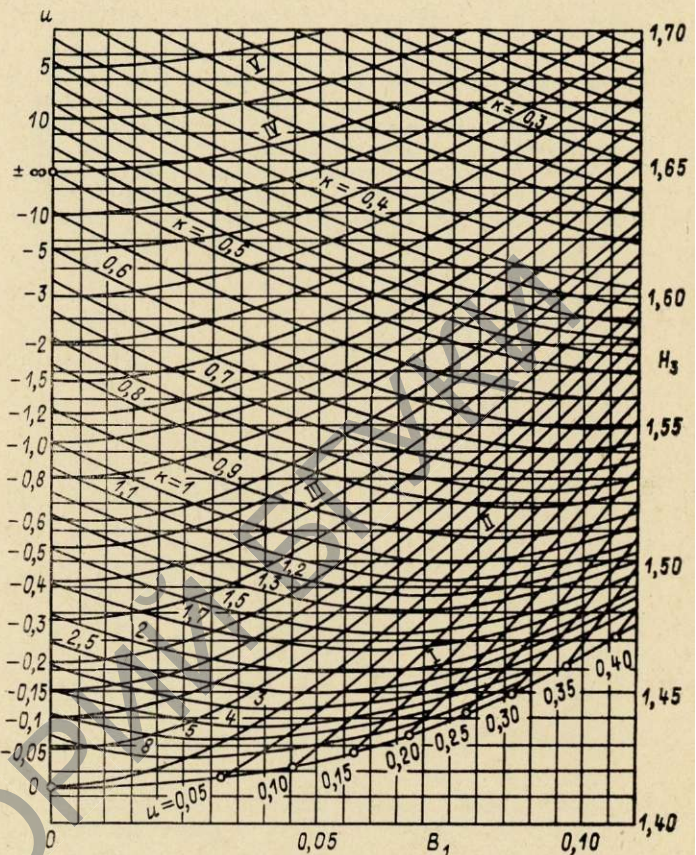


Рис. 1. График для установления типа выравнивающей кривой распределения и оценки параметров  $u, k$

В зависимости от вида выравнивающего распределения (см. плотности (1), (5), (15)) в качестве эмпирической величины  $S_1^*$  принимается одна из величин  $tp(t), p(x), yp(y) \ln y$ ; аналогично для величины  $v_1^*$  имеем:  $\ln t, x, \ln \ln y$ .

Величины  $S_1^{(z)}, v_1^{(z)}$  в зависимости от типа кривой распределения вычисляются по формулам:

— в случае распределений I, I' типов

$$S_1^{(z)} = \frac{\Gamma(k + \frac{1}{u})}{\Gamma(k) \Gamma(\frac{1}{u})} \frac{\Gamma(2k) \Gamma(\frac{2}{u} - 1)}{\Gamma[2(\frac{1}{u} + k) - 1]}, \quad (21)$$

$$\begin{aligned} v_1^{(z)} &= \pm \left( \sum_{s=1}^{k-1} \frac{1}{s} - \sum_{s=1}^{k+\frac{1}{u}-1} \frac{1}{s} \right) = \\ &= \pm \left( \psi(k) - \psi\left(k + \frac{1}{u}\right) \right), \end{aligned} \quad (22)$$

где  $\psi(k) = \sum_{s=1}^{k-1} \frac{1}{s} - C = \frac{d}{dk} \ln \Gamma(k)$  — логарифмическая

производная гамма-функции ( $C \approx 0,577216$  — постоянная Эйлера);

— в случае распределений II, II' типов

$$S_1^{(z)} = \frac{\Gamma(2k)}{2^{2k} [\Gamma(k)]^2}, \quad (23)$$

$$v_1^{(z)} = \pm \left( \sum_{s=1}^{k-1} \frac{1}{s} - C \right) = \pm \psi(k); \quad (24)$$

— в случае распределений III-V типов

$$S_1^{(z)} = \left[ \frac{\Gamma\left(1 - \frac{1}{u}\right)}{\Gamma(k) \Gamma\left(1 - \frac{1}{u} - k\right)} \right]^2 \times \frac{\Gamma(2k) \Gamma\left[2\left(1 - \frac{1}{u} - k\right)\right]}{\Gamma\left[2\left(1 - \frac{1}{u}\right)\right]}, \quad (25)$$

$$v_1^{(z)} = \sum_{s=1}^{k-1} \frac{1}{s} - \sum_{s=1}^{k-1} \frac{\left(1 - \frac{1}{u} - k\right)^{-1}}{s} = \psi(k) - \psi\left(1 - \frac{1}{u} - k\right). \quad (26)$$

Величина  $\ln \Gamma(k)$  и, следовательно,  $\Gamma(k)$  с любой точностью может быть вычислена по формуле, которая следует из ряда Стирлинга:

$$\ln \Gamma(k) \approx \frac{\ln 2\pi}{2} - \sum_{s=k}^{k+n} \ln s + \left(k+n + \frac{1}{2}\right) \ln(k+n) - (k+n) + \frac{1}{12(k+n)} - \frac{1}{360(k+n)^3} + \dots, \quad (27)$$

а величина  $\psi(k) = \frac{d}{dk} \ln \Gamma(k)$  — по формуле

$$\psi(k) \approx - \sum_{s=k}^{k+n} \frac{1}{s} + \ln(k+n) + \frac{1}{2(k+n)} - \frac{1}{12(k+n)^2} + \frac{1}{120(k+n)^4} - \dots, \quad (28)$$

где  $n$  — целое. При  $k+n=5$  последние две формулы дают погрешность, по абсолютной величине не превышающую значения  $2,54 \cdot 10^{-7}$ .

На рис. 1 представлены распределения с левосторонней асимметрией, для которых  $0 < B_1 < 0,25$ , и симметричные ( $B_1=0$ ), при этом распределения приведены к форме (5). Распределения с правосторонней асимметрией характеризуются значениями  $-0,25 < B_1 < 0$ . В этих случаях оценки  $k, u$  находятся при положительных значениях  $B_1$ . Кроме того, в случае распределений III типа при  $-0,25 < B_1 < 0$  из рис. 1 сначала находят оценки параметров  $k', u$ , затем вычисляется величина  $k$  по формуле  $k = 1 - \frac{1}{u} - k'$ .

### 3. ВЫРАВНИВАНИЕ СТАТИСТИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Рассмотрим в качестве примера ранговое распределение научных сотрудников по продуктивности, т. е. по числу опубликованных ими статей за определенный период [4]. Обозначим через  $m_r$  количество статей научного сотрудника с рангом  $r$ . Из опыта имеем

$$1 \leq r \leq 404, \quad 1 \leq m_r \leq 202 \quad (\text{статей}), \quad \sum_{r=1}^{404} m_r = X = 2770 \quad (\text{статей}).$$

На рис. 2 приведен график зависимости  $rp_r$  от  $\ln r$  (отдельные точки), построенный по эмпирическим данным. Здесь  $p_r = m_r/X$ . При построении графика в области малых частот выбирались ранги, удовлетворяющие условию  $m_r - m_{r+1} = 1$ , при этом учитывались средние относительные частоты  $p_r = (m_r + m_{r+1})/2X$ . Для наивысшего ранга  $r=404$  (обозначим его через  $Y$ ) имеем  $p_Y = 1/2X$ , поскольку  $m_Y = 1, m_{Y+1} = 0$ .

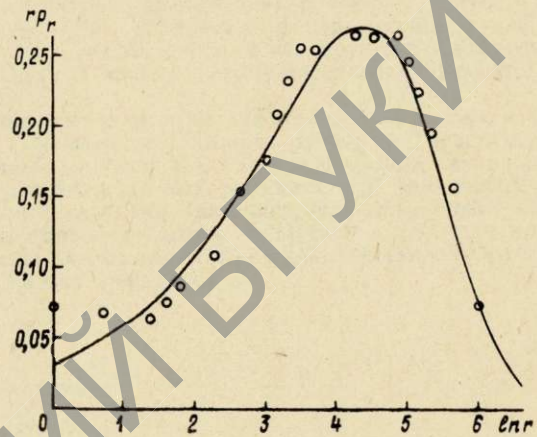


Рис. 2. Распределение научных сотрудников по продуктивности

Следовательно, ордината крайней справа точки по построению равна

$$Y \bar{p}_Y = \frac{Y}{2X}. \quad (29)$$

Сделаем некоторые выводы из построенного графика.

1. Крайняя справа точка в зависимости от объема выборки  $X$  занимает вполне определенное место на кривой распределения. С ростом величины  $X$  после достижения наивысшего положения она смещается вправо и приближается к горизонтальной оси. Объем выборки можно считать достаточным (для установления типа выравнивающей кривой распределения), если ордината этой точки меньше наперед заданного числа  $\varepsilon$ , т. е.

$$Y \bar{p}_Y = \frac{Y}{2X} \leq \varepsilon. \quad (30)$$

Величину  $\varepsilon$  целесообразно принимать из условия  $0,01 \leq \varepsilon \leq 0,1$ .

2. Известно, что закон Цифа может быть записан в виде  $rp_r = C$ . Если бы распределение научных сотрудников по продуктивности описывалось последней формулой, то на рис. 2 мы имели бы прямую  $rp_r = \text{const}$ , т. е. статистическое распределение, приведенное к форме (5), было бы равномерным. Опыт показывает, что ни одно из ранговых распределений в данной системе координат не является равномерным. А это означает, что закон Цифа не подтверждается опытными данными.

3. Исследуемая выборка является достаточно однородной, так называемый график имеет вполне определенную моду, если пренебречь некоторым разбросом точек вблизи моды и в самом начале кривой. Однако первые точки соответствуют, по крайней мере, двум

сотрудникам с повышенной продуктивностью.

Следовательно, график зависимости  $rp_r$  от  $\ln r$  позволяет решить вопрос об однородности рангового распределения, а в случае неоднородности — разбить совокупность на две более однородные части.

Найдем выравнивающую кривую распределения, заданную обобщенной плотностью (1).

Поскольку статистическое распределение приведено к форме (5), то можно воспользоваться формулами (16)—(18). Разобьем интервал, на котором задано статистическое распределение, на равные отрезки  $\Delta \ln r_i = \Delta x_i = 0,5$  ( $i=1, 2, \dots, 12$ ). В середине каждого отрезка из графика найдем плотность распределения  $(rp_r)_i = p(x_i)$ . Тогда по приведенным ниже формулам, которые являются эмпирическими аналогами формул (16)—(18), найдем

$$v_1^* = \bar{x} = \overline{\ln r} = \sum_{i \geq 0} x_i p(x_i) \Delta x_i = 3,50175;$$

$$S_1^* = \sum_{i \geq 1} [p(x_i)]^2 \Delta x_i = 0,180575;$$

$$S_3^* = \sum_{i \geq 1} [p(x_i)]^4 \Delta x_i = 0,009139;$$

$$H_3^* = S_3^* / (S_1^*)^3 = 1,5521;$$

$$B_1^* = \sum_{i \geq 1} (x_i - \bar{x}) [p(x_i)]^2 \Delta x_i =$$

$$= \sum_{i \geq 1} \bar{x}_i [p(x_i)]^2 \Delta x_i - v_1^* S_1^* = 0,08665.$$

Приравняв эмпирические значения показателей асимметрии  $B_1^*$  и островежности  $H_3^*$  соответствующим теоретическим, с помощью рис. 1 находим оценки параметров  $u, k$  выравнивающей кривой распределения:  $u \approx -0,2, k \approx 0,6$ . Выравнивающее распределение относится к III типу. При известных оценках  $u, k$  по формулам (25)—(28) находим:  $S_1^{(z)} = 0,1706, v_1^{(z)} = -3,1316$ . Тогда по формулам (19), (20), (4) для распределения III типа будем иметь:  $\beta = S_1^* / S_1^{(z)} = 1,0585, \alpha u = -e^{v_1^{(z)} - \beta v_1^*} = -1/932,81, N = 1/31,643$ . Выравнивающее распределение задается формулой (при  $\gamma = k\beta = 0,6351$ ):

$$p(t) = \frac{1}{31,643 t^{0,3649} \left(1 + \frac{t^{1,0585}}{932,81}\right)^6} \quad (31)$$

На рис. 2 представлен график кривой распределения (31), приведенной к форме (11), который достаточно точно выравнивает статистическое распределение.

Дальнейшие исследования показали, что ранговое распределение словоформ по радиотехнической специальности [5] (объем выборки  $X=406\,000$  словоупотреблений, объем словаря  $Y=28\,722$  словоформы) может быть описано обобщенной плотностью (1) с параметрами  $\alpha=2,2693, \beta=0,02478, \gamma=0,49566, u=1/3$  и нормирующим множителем  $N=0,4307$ ; распределение абонентов БелРАСНТИ, ранжированных по количеству полученных ими библиографических описаний документов (сигнальной информации) в течение одного квартала, — той же плотностью с параметрами  $\alpha=1,375, \beta=0,129, \gamma=1,29, u=0,4$  и нормирующим множителем  $N=0,0925$ . Последние два распределения относятся к I типу.

В заключение отметим, что обобщенные плотности (1), (5), (15) позволяют решать различные практические задачи более точно и в более общем виде, чем при использовании отдельных распределений, являющихся их частными случаями (например, при распределении Вейбулла, которое относится ко II типу и на рис. 1 представлено одной лишь точкой с координатами  $B_1=0,0767, H_3=1,5$  или  $u \rightarrow 0, k=1$ ).

#### ЛИТЕРАТУРА

1. Нешиной В. В. Система непрерывных распределений в информатике и лингвистике. — НТИ. Сер. 2, 1984, № 3, с. 1—6.
2. Нешиной В. В. Построение системы непрерывных распределений / БелНИИНТИ. — Минск, 1979. — 193 с. — Рукопись деп. в БелНИИНТИ 30.07.80, № 174.
3. Нешиной В. В. Оценивание параметров обобщенных распределений / БелНИИНТИ. — Минск, 1983. — 96 с. — Рукопись деп. в БелНИИНТИ 11.03.84, № 857.
4. Козачков Л. С. О некоторых проблемах релевантности в информатике и науковедении. — НТИ, Сер. 2, 1969, № 8, с. 3—11.
5. Межлумова А. Б. Лексический минимум по радиотехнической специальности. — Минск, 1972.

Статья поступила в редакцию 5 июня 1984 г.