В.В. Нешитой,
заведующий кафедрой информационных ресурсов Белорусского государственного университета культуры и искусств,
доктор технических наук, профессор

ГРАФИЧЕСКИЙ МЕТОД ВЫЧИСЛЕНИЯ ГРАНИЦ ЯДРА И ЗОН РАССЕЯНИЯ КНИЖНОГО ФОНДА

В информатике известен закон рассеяния журнальных публикаций С. Бредфорда, который заключается в следующем: «Если научные журналы расположить в порядке убывания числа помещенных в них статей по какому-либо заданному предмету, то в полученном списке можно выделить ядро журналов, посвященных непосредственно этому предмету, и несколько групп или зон, каждая из которых содержит столько же статей, что и ядро. Тогда числа журналов в ядре и последующих зонах будут относиться как $1:n:n^2$» [1, с. 93–94].

Из приведенной формулировки ясно, что закон С. Бредфорда не содержит нужной информации для вычисления границ ядра и зон рассеяния. Для этого требуется знание теоретического рангового закона распределения.

Следовательно, проблема рассеяния журнальных публикаций, книг, упорядоченных по убыванию их выдач, состоит в решении общей задачи – нахождении такого теоретического рангового распределения, которое способно с высокой точностью аппроксимировать все многообразие статистических ранговых распределений. Такое распределение и будет являться наиболее общим (универсальным) законом рассеяния [2].

Эта задача была успешно решена автором путем разработки теории обобщенных распределений [3]. Из этой теории следует, что если ранговое распределение $\mathrm{p_r} = \varphi(r)$ (здесь $\mathrm{p_r}$ – относительная частота выдачи книги с рангом r) представить в другой системе координат – $rp_r = f(\ln r)$, то его график будет иметь три характерные точки: моду C и две точки перегиба A и B. Примем координаты этих точек в качестве границ ядра книг и зон рассеяния.

Для их нахождения по статистическому ранговому распределению без предварительного вычисления теоретического распределения можно предложить простой графический метод, который следует из свойств обобщенных распределений. Суть этого метода покажем на примере закона Вейбулла, функция и плотность распределения которого заданы формулами

$$\mathrm{F(t)} = 1 - \mathrm{e}^{-\alpha \mathrm{t}^{\beta}}, \tag{1}$$

$$\mathrm{p(t)} = \alpha\beta \mathrm{t}^{\beta-1} e^{-\alpha t^{\beta}}. \tag{2}$$

При значениях параметра *β < 1* плотность (2) может с достаточной точностью описывать многие статистические ранговые распределения.

Чтобы получить из рангового распределения полезную информацию, преобразуем плотность (2) к форме *tp(t) = f(lnt):*

$$\mathrm{tp(t)} = \alpha\beta \mathrm{t}^{\beta} e^{-\alpha t^{\beta}} = \alpha\beta e^{\beta \ln t} e^{-\alpha e^{\beta \ln t}}. \tag{3}$$

Введем обозначения: *tp(t) = p(x); lnt = x*. Тогда плотность (3) примет вид

$$\mathrm{p(x)} = \alpha\beta e^{\beta x} e^{-\alpha e^{\beta x}}. \tag{4}$$

Плотность (4) дает новую информацию о ранговом распределении. График этой плотности, т.е. кривая распределения содержит три характерные точки – моду *C* и две точки перегиба *A* и *B,* расположенные на равных расстояниях по обе стороны от моды. Найдем абсциссы этих точек. Продифференцируем плотность (4) по *x* и приравняем первую производную нулю. Решая полученное уравнение, найдем выражение для моды

$$\mathrm{x_C} = \frac{1}{\beta} \ln \frac{1}{\alpha}. \tag{5}$$

В точках перегиба вторая производная равна нулю. Из этого условия для плотности (4) имеем

$$\mathrm{x_A} = x_C - \frac{1}{\beta} \ln \frac{3+\sqrt{5}}{2}, \quad \mathrm{x_B} = x_C + \frac{1}{\beta} \ln \frac{3+\sqrt{5}}{2}.$$

Введем обозначение

$$n = \left(\frac{3+\sqrt{5}}{2}\right)^{\frac{1}{\beta}}. \tag{6}$$

С учетом (6) последние две формулы примут более простой вид

$$\mathrm{x_A} = x_C - \ln n; \tag{7}$$

$$\mathrm{x_B} = x_C + \ln n. \tag{8}$$

РЕПОЗИТОРИЙ БГУКИ

Переходя к распределению Вейбулла, на основании равенств $tp(t) = p(x); \ln t = x$ и формул (5)–(8) найдем

$\ln t_B - \ln t_C = \ln t_C - \ln t_A = \ln n$, откуда $\ln\frac{t_B}{t_C} = \ln\frac{t_C}{t_A} = \ln n$, или

$$\frac{t_B}{t_C} = \frac{t_C}{t_A} = n. \qquad (9)$$

Здесь

$$t_C = \left(\frac{1}{\alpha}\right)^{\frac{1}{\beta}}; \quad t_A = \frac{t_C}{n}; \quad t_B = t_C \cdot n. \qquad (10)$$

Формула (9) может служить уточненной формулировкой закона рассеяния публикаций в смысле С. Бредфорда, хотя она представляет собой новую формулировку закона рассеяния публикаций.

Предположим далее, что ранговое распределение книг хорошо описывается законом Вейбулла с параметрами *α=0,1; β=0,5*. Приведем его к плотности *p(x)*, которая задана формулой (4). Тогда для этой плотности по формулам (5)–(8) найдем: n=6,8541; $x_C = 4{,}6052$; $x_A = 2{,}6804$; $x_B = 6{,}53$.

Рассчитаем по формуле (4) значения плотности *p(x)* с интервалом *Δx=1* и сведем результаты в таблицу.

Значения плотности и тангенса угла наклона касательной к кривой в серединах интервалов

x	*p(x)*	*Δp(x)/Δx*	*x*	*p(x)*	*Δp(x)/Δx*
-2	0,01773		4	0,17646	
		0,01081			0,00369
-1	0,02854		5	0,18015	
		0,01670			-0,04539
0	0,04524		6	0,13476	
		0.02467			-0,07439
1	0.06991		7	0,06037	
		0,03366			-0,04876
2	0,10357		8	0,01161	
		0,03957			-0,01106
3	0,14314		9	0,00055	
		0,03332			

РЕПОЗИТОРИЙ БГУКИ

Построим график плотности *p(x),* т.е. кривую распределения (см. с. 256).

На кривой распределения абсциссу точки *C* легко найти графически путем проведения горизонтальной касательной к кривой (рис. *a*).

Для нахождения абсцисс точек перегиба воспользуемся тем свойством кривой распределения, что в точках А и В первая производная принимает экстремальные значения: в точке А она имеет максимум, а в точке В – минимум. Вычислим тангенс угла наклона кривой к горизонтальной оси на всех интервалах как отношение разности между соседними значениями плотности *p(x)* к ширине интервала *Δx*. Другими словами, найдем приближенные значения первой производной в серединах интервалов и построим график (рис. *b*).

Первая производная *dp(x)/dx* имеет максимум в точке $\mathrm{x_A}$ и минимум в точке x_B. Эти точки легко определить путем проведения горизонтальных касательных к кривой на рис. b.

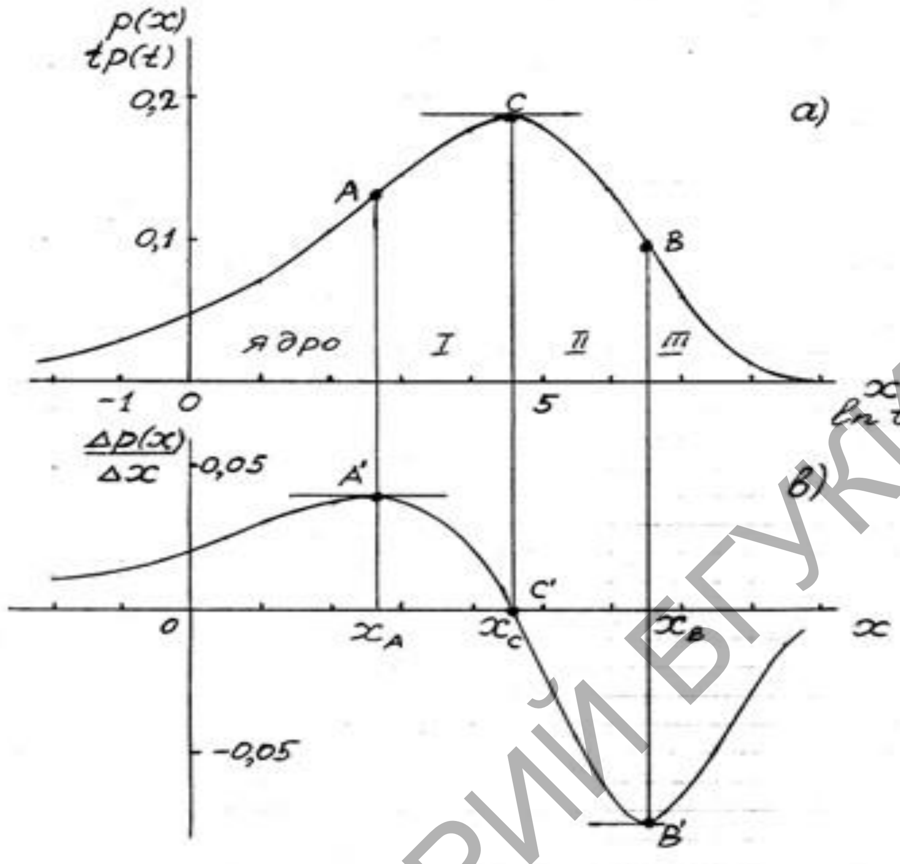
Из построенного графика можно приближенно найти абсциссы трех характерных точек: $\mathrm{x_A} = 2{,}7$; $\mathrm{x_C} = 4{,}6$; $x_B = 6{,}5$.

Переходя к ранговому распределению, находим: $\mathrm{t_A} = e^{x_A} \approx 15$; $t_C = e^{x_C} \approx 99$; $t_B = e^{x_B} \approx 665$. Точные значения для распределения Вейбулла равны: $\mathrm{t_A} = 14{,}59$; $t_C = 100$; $t_B = 685$.

Таким простым методом приближенно могут быть найдены абсциссы трех характерных точек любого статистического рангового распределения без предварительного вычисления теоретического закона распределения.

Значения функции распределения *F(t)* при любом заданном значении ранга t, в том числе в трех характерных точках могут быть вычислены по статистическому ранговому распределению.

РЕПОЗИТОРИЙ БГУИР



Графики плотности распределения (*a*) и ее первой производной (*b*)

Абсцисса точки *A*, т.е. t_A представляет собой ядро книг, а точнее – ядро читательских интересов, поскольку ранговое распределение книг формируется читателями. Функция распределения $F(t_A) = F(x_A)$, т.е. площадь под кривой распределения до точки *A* (рис. а), равна доле книговыдач ядра фонда. Функция $F(t_B) = F(x_B)$ – это доля выдач, приходящаяся на книги от начала частотного списка до точки *B* или, другими словами, оценка вероятности удовлетворения информационных потребностей пользователей книжным фондом объемом t_B. Величина t_B представляет собой оптимальный объем фонда.

Итак, для вычисления графическим методом ядра фонда и зон рассеяния, оптимального объема фонда необходимо и достаточно иметь статистическое ранговое распределение.

---

1. Михайлов, А. И. Основы информатики / А. И. Михайлов, А. И. Черный, Р. С. Гиляревский. – М. : Наука, 1968. – 756 с.

РЕПОЗИТОРИЙ БГУКИ

2. *Нешитой, В. В.* Универсальные законы рассеяния и старения публикаций / В. В. Нешитой // Весн. Беларус. дзярж. ун-та культуры і маст. – 2007. – № 8. – С. 128–133.

3. *Нешитой, В. В.* Элементы теории обобщенных распределений / В. В. Нешитой. – Минск : РИВШ, 2009. – 204 с.

РЕПОЗИТОРИЙ БГУКИ