

В.В.Нешитой, доктор технических наук,
*Белорусский государственный
университет культуры и искусств*

СТАТИЧЕСКАЯ И ДИНАМИЧЕСКАЯ ПОЛНОТА КОМПЛЕКТОВАНИЯ БИБЛИОТЕЧНОГО ФОНДА

Полнота комплектования библиотечного фонда является одной из важнейших его характеристик. Высокая полнота комплектования фонда гарантирует высокую степень удовлетворения информационных потребностей пользователей. Ниже пойдет речь об информационной полноте комплектования при двух подходах к ее определению.

Статическая полнота комплектования

Первый подход основывается на ранговых распределениях. Если все события, составляющие полную группу, упорядочить по убыванию, а точнее – по невозрастанию частоты их появления в некоторой выборке достаточно большого объема, приписать каждому событию порядковый номер, то получим статистическое ранговое распределение разных событий. Порядковый номер события является его рангом. В качестве примеров можно привести следующие статистические ранговые распределения:

- распределение лексических единиц (словоформ, лексем, терминов) по частоте их употребления в текстах большого объема;
- распределение периодических изданий по числу помещенных в них статей по заданному предмету;
- распределение книг по числу их выдач за длительный период времени.

Все статистические ранговые распределения имеют общие свойства: с ростом ранга частота событий (абсолютная или относительная) убывает, а количество событий, имеющих одинаковую частоту, растет. Статистическое ранговое распределение можно представить в виде таблицы или графика.

В таблице обычно приводятся следующие величины:

- ранг события r или интервал рангов – для событий с одинаковой частотой; абсолютная частота события или группы событий m ; количество событий с данной частотой y_m , при этом $\sum_{m \geq 1} y_m = Y$, где Y – количество наступивших разных событий; произведение my_m ; накопленная абсолютная частота событий $\sum my_m$,

при этом $\sum m_{y_m} = X$, где X – объем выборки; относительная частота событий $p_r = m_r/X$; накопленная относительная частота событий от начала частотного списка до ранга r , т.е. эмпирическая функция распределения $F(r) = \sum p_r$. Функция распределения изменяется от нуля до единицы.

Для описания широкого многообразия статистических распределений, в том числе ранговых, автором разработаны три (основные) системы непрерывных распределений [1]:

$$p(x) = Ne^{kx} (1 - aue^{bx})^{\frac{1}{u}-1}; \quad (1)$$

$$p(t) = Nt^{kb-1} (1 - aut^b)^{\frac{1}{u}-1}; \quad (2)$$

$$p(y) = \frac{N}{y} (\ln y)^{kb-1} (1 - a(\ln y)^b)^{\frac{1}{u}-1}, \quad (3)$$

где N – нормирующий множитель; α , β , k , u – параметры, которые вычисляются по статистическому распределению.

При различных исследованиях в социально-гуманитарной сфере могут быть использованы плотности $p(t)$, $p(y)$, которые заданы формулами (2), (3). Плотность $p(t)$ описывает множество статистических распределений, в том числе ранговых: периодических изданий, упорядоченных по убыванию числа помещенных в них статей по заданному предмету; книг – по числу выдач; научных сотрудников – по продуктивности и др. Плотность $p(y)$ описывает статистические ранговые распределения лексических единиц (словоформ, лексем, словосочетаний, терминов).

Выясним взаимосвязь между первой и второй системами непрерывных распределений. Первая система обладает уникальным свойством: графики плотности $p(x)$, т.е. кривые распределения, при значениях параметра $u < 1/2$ имеют моду x_C и две точки перегиба x_A и x_B , расположенные на равных расстояниях от моды. Эти точки можно использовать для разделения рангового распределения, заданного плотностью $p(t)$, на зоны. Но для этого ее необходимо представить в такой системе координат, чтобы она преобразовалась в плотность $p(x)$ и приобрела все свойства последней, т.е. имела бы моду и две точки перегиба.

Если по вертикальной оси отложить произведение $tp(t)$, а по горизонтали – $\ln t$, то вторая система распределений приобретет

свойства первой системы [2], т.е. будет иметь моду и две точки перегиба. Отсюда вытекает правило: чтобы для убывающего рангового распределения найти характерные точки, его необходимо привести к форме плотности $p(x)$, т.е. изобразить графически в системе координат $rp(r) = f(\ln r)$. Тогда ранговое распределение будет иметь моду $\ln r_C$ и две точки перегиба $\ln r_A$ и $\ln r_B$, которые находятся на равных расстояниях от моды: $\ln r_C - \ln r_A = \ln r_B - \ln r_C$. Отсюда следует соотношение

$$(4) \quad \frac{r_C}{r_A} = \frac{r_B}{r_C} = n,$$

которое может быть принято в качестве закона рассеяния публикаций.

В литературных источниках утверждается, что закон рассеяния публикаций Бредфорда вытекает из закона Ципфа. Но последний является частным случаем обобщенной плотности $p(t)$ при $b < 0, u = 1$. Представленный в форме $tp(t) = f(\ln t)$, он не имеет характерных точек. Поэтому на базе закона Ципфа невозможно получить закон рассеяния публикаций Бредфорда.

Точки A, C, B делят все журналы в ранжированном ряду на четыре части: ядро и три зоны рассеяния. Количество журналов, входящих в ядро, определяется равенством $t_J = t_A$. Количество журналов в первой зоне равно разности $t_I = t_C - t_A$; во второй зоне $t_{II} = t_B - t_C$. Остальные журналы относятся к третьей зоне: $t_{III} > t_B$. При этом количество журналов от начала частотного списка до точки C в n раз больше количества журналов в ядре. Количество журналов до точки B в n раз больше их количества до точки C и в n^2 раз больше, чем в ядре.

Поэтому формулу (4) можно представить в виде $t_A : t_C : t_B = t_A (1 : n : n^2)$ или $t_J : t_I : t_{II} = t_J [1 : (n - 1) : (n - 1)n]$.

Последние формулы не содержат информации о количестве журналов и долях статей в каждой зоне, сколько может быть зон рассеяния, чему равна величина n . Вся эта информация содержится в обобщенной плотности $p(t)$, поэтому распределение (2) по праву является универсальным законом рассеяния публикаций.

Журналы, входящие в ядро, содержат долю статей, равную функции распределения $F(t_A) = \int_0^{t_A} p(t) dt$. Аналогично доля статей в журналах, входящих в ядро и первую зону рассеяния, составляет $F(t_C)$, и т.д. Следовательно, доля статей в первой зоне рассеяния

составляет $F(t_C) - F(t_A)$; во второй зоне $F(t_B) - F(t_C)$, а в третьей зоне $-1 - F(t_B)$.

Мода t_C находится из условия $dtp(t)/dln t = 0$ и в общем случае для распределений $I-V$ типов равна [1, 4]

$$t_C = \left(\frac{k}{a(1+ku-u)} \right)^{1/b}. \quad (5)$$

Величина n задается формулой

$$n = \left[1 + \frac{1-u + \sqrt{[4k(1+ku-u) + (1-u)](1-u)}}{2k(1+ku-u)} \right]^{1/b}. \quad (6)$$

Абсциссы точек перегиба вычисляются по формулам:

$$t_A = t_C/n; \quad t_B = t_C \cdot n. \quad (7)$$

С целью более рационального комплектования фонда в него следует включать те издания, которые образуют ядро и первые две зоны рассеяния. Количество таких изданий равно t_B , при этом полнота комплектования фонда задается функцией распределения $F(t_B)$. Например, в случае справедливости закона Вейбулла она равна 0,93. Величина t_B может характеризовать некоторый оптимальный объем фонда, а величина $F(t_B)$ – информационную полноту комплектования фонда объемом t_B , т. е. вероятность удовлетворения информационных потребностей пользователей этим фондом. В то же время величина $F(t_B)$ – это доля книговыдач, приходящаяся на фонд объемом t_B . Функция распределения $F(t)$ представляет собой статическую полноту комплектования. Она не зависит от числа книговыдач, но зависит от рангового закона распределения книг по числу их выдач за длительный период времени и поэтому является относительно устойчивым показателем полноты комплектования фонда.

Динамическая полнота комплектования

Второй подход к определению информационной полноты комплектования библиотечного фонда основывается на кривой роста новых выданных книг, т. е. ранее не выдававшихся от начала учета книговыдач.

Пусть вероятности выдач каждой книги заданы и равны $p_1, p_2, \dots, p_K, \dots, p_N$, где N – объем фонда. Будем считать, что книги упорядочены по убыванию вероятностей. Тогда статистическое ранговое распределение книг можно описать плотностью $p(t)$. Количество выданных книг y (имеется в виду количество разных

наименований) в зависимости от числа книговыдач x можно рассчитать по формуле [4]

$$y = \int_0^N \left(1 - \frac{1}{e^{xp(t)}} \right) dt. \quad (8)$$

Тогда вероятность выдачи новой (ранее не выдававшейся) книги будет равна первой производной от кривой роста (8)

$$P(A'') = \frac{dy}{dx}, \quad (9)$$

а накопленная вероятность y выданных книг будет равна функции распределения $F(y) = 1 - \frac{dy}{dx}$, (10)

которая представляет собой динамическую информационную полноту комплектования фонда. С ростом x она увеличивается.

На базе формулы В.Калинина [5] по кривой роста (8) можно рассчитать частотный спектр, т.е. количество книг, выданных один раз, два раза, ... m раз, при x книговыдачах

$$y_m = (-1)^{m+1} \frac{x^m}{m!} \frac{d^m y}{dx^m}. \quad (11)$$

Из этой формулы при $m=1$ имеем $y_{m=1} = x(dy/dx)$, откуда $dy/dx = y_{m=1} / x$. С учетом последнего равенства функция распределения (10) запишется в виде $F(y) = 1 - \frac{y_{m=1}}{x}$.

По этой формуле может быть оценена динамическая полнота комплектования фонда при известных значениях книговыдач x и книг, выданных один раз $y_{m=1}$.

Исследования показывают, что при $y=r$ справедливо неравенство $F(y) < F(r)$, т.е. динамическая полнота меньше статической. Это связано с тем, что разные книги выдаются лишь в среднем по убыванию вероятностей, в то время как статическая полнота рассчитывается по ранговому распределению, когда книги упорядочены по невозрастанию вероятностей.

1. Нешитой, В.В. Элементы теории обобщенных распределений / В.В.Нешитой. – Мн.: РИВШ, 2009. – 204 с.

2. Нешитой, В.В. Форма представления ранговых распределений / В.В.Нешитой // Учен. зап. Тартуского гос. ун-та. – 1987. – Вып. 774. – С. 123–134.

3. 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.

4. *Нешитой, В.В.* Математико-статистические методы анализа в библиотечно-информационной деятельности: учеб.-метод. пособие / В.В.Нешитой. – Мн.: БГУ культуры и искусств, 2009. – 203 с.

5. *Калинин, В.М.* Некоторые статистические законы математической лингвистики / В.М. Калинин // Проблемы кибернетики. – М.: Наука, 1964. – Вып. 11. – С. 246–255.

РЕПОЗИТОРИЙ БГУКИ