Совокупность преимуществ УВОСП позволяет преподавателю более оперативно повышать уровень своих компетенций и, следовательно, повышать качество обучения.

Махнач, Г. В. Автоматизация учетно-вспомогательных операций в преподавательской деятельности / Г. В. Махнач // Междунар. конгресс по информатике: информац. системы и технологии, Респ. Беларусь, Минск, 4–7 нояб. 2013 г. / редкол.: С. В. Абломейко (отв. ред.), В. В. Казаченок (отв. ред.) [и др.]. – Минск : БГУ, 2013. – С. 236–240.

В. В. Нешитой,
профессор кафедры информационных ресурсов,
доктор технических наук, профессор

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ АНАЛИЗА БИБЛИОТЕЧНОЙ ДЕЯТЕЛЬНОСТИ КАК ПРОФЕССИОНАЛЬНАЯ КОМПЕТЕНТНОСТЬ СТУДЕНТОВ-ГУМАНИТАРИЕВ

При исследовании степени использования библиотечного фонда нельзя обойти статистические методы, ибо только они позволяют ответить на многие вопросы, в том числе: насколько востребован тот или иной документ, в какую группу он входит – в часто или редко запрашиваемых. Для исследования этих и других подобных вопросов необходимо провести статистический анализ использования библиотечного фонда: для этого нужно отобрать достаточно большое количество изданных документов (не < 1000–5000 шт.) – упорядочить их по убыванию частоты спроса, то есть построить статистическое ранговое распределение, а главное – в итоге найти теоретическое ранговое распределение и вычислить оценки его параметров по статистическому распределению. При этом статистическое ранговое распределение необходимо строить в системе координат $\mathrm{rp_r} = \varphi(\ln r)$ – по вертикальной оси откладываются произведения рангов книг r на их относительные частоты $\mathrm{p_r}$, которые являются оценками вероятностей выдачи книг с рангом r. Построенная таким образом кривая распределения имеет три характерные точки: моду $\ln r_c$, (в которой произведение rp_r максимально) и две точки перегиба $\mathrm{lnr_A}, \mathrm{lnr_B}$, которые отделяют выпуклую часть кривой от вогнутой и наоборот, вогнутую от выпуклой. Они расположены на равных расстояниях по обе стороны от моды. Примем эти точки в качестве границ ядра (публикаций, журналов, книг и других объектов, упорядоченных по убыванию некоторого признака). Здесь уместно отметить, что при формулировке своего закона рассеяния С. Бредфорд отмечал только, что «можно выделить ядро и несколько групп или зон журналов, каждая из которых содержит столько же статей, что и ядро». Здесь он не указывает, по какому правилу можно выделить ядро и зоны рассеяния и откуда могут взяться эти зоны. Другими словами, он не исследует и не использует свойства статистических ранговых распределений. По этой причине попытки некоторых исследователей по выделению зон ранговых распределений по Бредфорду практически всегда заканчиваются неудачей. Из теории обобщенных распределений автора настоящей статьи следуют важные выводы, которые обосновывают универсальные законы старения и рассеяния публикаций. Использование этой теории позволяет вычислять закон распределения, в том числе рангового, по статистическому ряду. На базе найденного закона легко решаются многие задачи.

Рассмотрим первую и вторую системы непрерывных распределений, которые заданы четырехпараметрическими плотностями вида:

$$p(x) = Ne^{k\beta x}(1-\alpha u e^{\beta x})^{\frac{1}{u}-1}, \quad -\infty < x < \infty; \tag{1}$$

$$p(t) = Nt^{k\beta-1}(1-\alpha u t^{\beta})^{\frac{1}{u}-1}, \qquad 0 < t < \infty, \tag{2}$$

где α, β, k, u – параметры распределений; N – нормирующий множитель.

Плотность (1) обладает тем замечательным свойством, что при $u < 1/2$ кривые распределения, заданные этой плотностью, имеют моду и две точки перегиба, расположенные на равных расстояниях по обе стороны от моды. Эта плотность является универсальным законом старения публикаций. Она не может аппроксимировать статистические ранговые распределения. Для этих задач наиболее подходит плотность (2), то есть вторая система непрерывных распределений. Отметим, что обе плотности взаимосвязаны. Так, вторая система непрерывных распределений может быть получена из первой при $x = \ln t$ по известной формуле $p(t) = p(x)dx/dt$, на основании которой и равенства $x = \ln t$ имеем плотность (2).

Отсюда следует, что при известном универсальном законе старения легко находится формула (в виде распределения) для универсального закона рассеяния и наоборот, по закону рассеяния находится закон старения. Однако несмотря на взаимосвязь между этими законами они не были открыты до разработки теории обобщенных распределений. Более того, не были найдены даже частные случаи этих законов, например, распределение Вейбулла в качестве закона рассеяния публикаций, функция и плотность вероятностей которого задаются формулами:

$$F(t) = 1 - e^{-\alpha t^{\beta}}, \quad p(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}}. \tag{3}$$

Поскольку этот закон весьма простой, его целесообразно проверять в первую очередь при отыскании подходящего рангового распределения. Для этого функцию распределения необходимо преобразовать к линейному виду

$$\ln\ln(1/(1 - F(t))) = \ln\alpha + \beta\ln t. \tag{4}$$

Принимая $Y = \ln\ln(1/(1 - F(t)))$, $X = \ln t$, из (4) получим $Y=\ln\alpha + \beta X$.

Для проверки применимости закона Вейбулла необходимо по статистической функции распределения вычислить значения X, Y и построить график зависимости Y=f(X). Если эмпирические точки расположатся вдоль прямой (4), то далее по методу наименьших квадратов вычисляются оценки параметров α и β этой прямой:

$$\beta = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{\overline{X^2} - \left(\overline{X}\right)^2}, \quad \alpha = \exp\left(\overline{Y} - \beta\overline{X}\right). \tag{5}$$

В приведенных формулах черта сверху обозначает, что берутся средние значения соответствующих величин.

Абсциссы точек *A, C, B* для закона Вейбулла вычисляются по формулам:

$$t_C = (1/\alpha)^{1/\beta}; \quad t_A = t_C/n; \quad t_B = t_C \cdot n; \quad n = (3 + \sqrt{5})/2)^{1/\beta}. \tag{6}$$

Значения функции распределения в этих точках при любых значениях параметров α и β соответственно равны:

$$F(t_A) = 0{,}31748;\ F(t_C) = 0{,}63212;\ F(t_B) = 0{,}92705. \qquad (7)$$

Рассмотрим в качестве примера кумулятивное число статей в сериальных изданиях, отраженных в выпуске журнала «Математика» (1997–2010), в зависимости от ранга [3]. Общее число источников равно 3 799, число статей в них – 257 960. Автор отмеченной статьи не нашел теоретического рангового закона распределения. Использование второй системы непрерывных распределений привело к однозначному решению – эти статистические данные с высокой точностью описываются законом Вейбулла, который следует из плотности (2) при $u \to 0$, k=1 (см. график). Параметры закона Вейбулла равны: α=0,023349, β=0,687331. Далее по формулам (6) вычисляются абсциссы характерных точек: n=4,56; $t_C = 236{,}598$; $t_A = 58{,}333$; $t_B = 959{,}641$. Накопленные доли статей в этих точках задаются формулами (7). Таким образом, в ядро входят 58 журналов, которые содержат 31,7 % статей. Оптимальный объем фонда, то есть ядро и первые две зоны рассеяния, составляют 960 журналов, на которые приходится 92,7 % статей, на остальные – 2 839 журналов, то есть самую обширную третью зону рассеяния, приходится лишь 7,3 % статей.

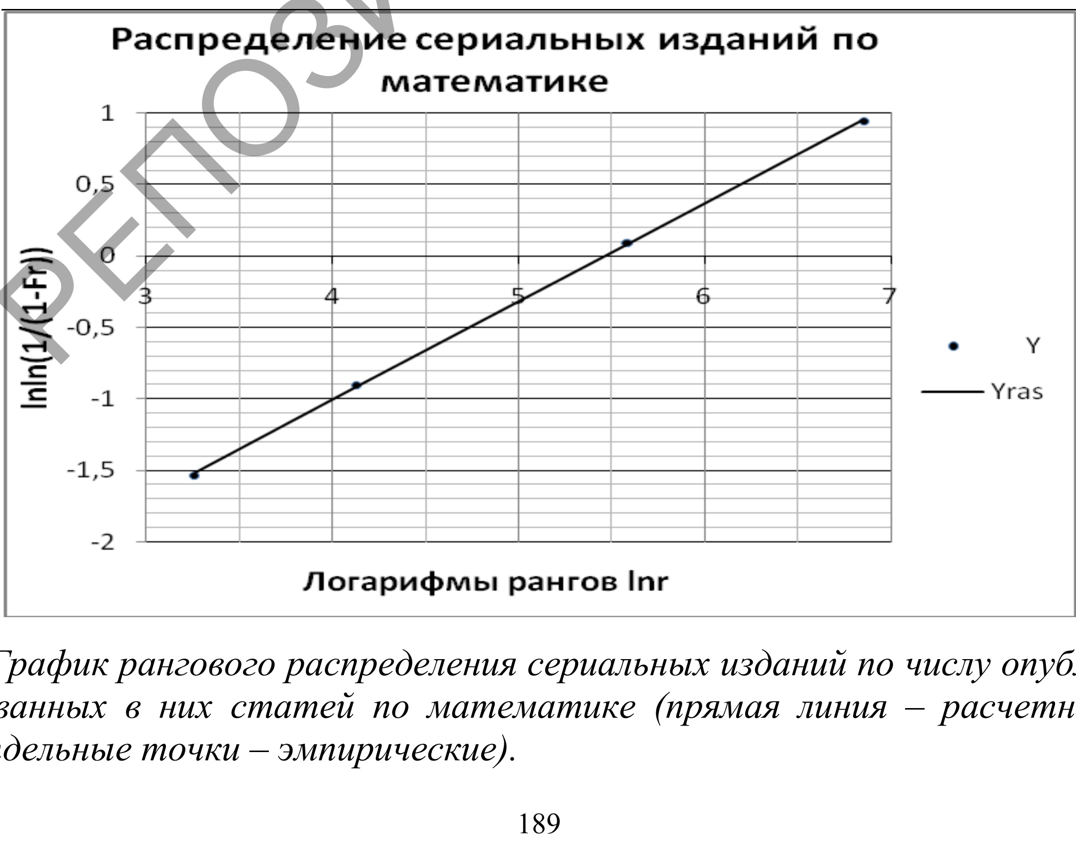

График рангового распределения сериальных изданий по числу опубликованных в них статей по математике (прямая линия – расчетная, отдельные точки – эмпирические).

В заключение следует отметить, что под статистические исследования в разных областях знаний, в том числе информетрии, наукометрии, библиотечно-информационной деятельности, математической лингвистики, в системах менеджмента качества, экономике, прогностике и многих других с разработкой теории обобщенных распределений подведена серьезная научная математическая база, которая при ее использовании поднимет на новый, значительно более высокий уровень исследования в отмеченных областях знания. В то же время для овладения этой теорией достаточно иметь прочные знания по школьному курсу математики и дисциплине [2] «Математико-статистические методы библиотечно-информационной деятельности», которую студенты изучают в БГУКИ. Студент должен быть уверен, что с помощью теории обобщенных распределений он по статистическому распределению, которое получено на базе однородной выборки достаточно большого объема, практически всегда вычислит теоретический закон распределения, в том числе рангового, и оценки его параметров (даже без использования компьютерной программы – тип распределения и оценки двух параметров формы *k, u* он возьмет из номограммы, а параметры *α, β* рассчитает по формулам, которые приведены в учебных пособиях автора теории, а также множестве статей.

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

2. *Нешитой, В. В.* Математико-статистические методы анализа в библиотечно-информационной деятельности : учеб.-метод. пособие / В. В. Нешитой. – Минск : БГУКИ, 2009. – 202 с.

3. *Шамаев, В. Г.* Инфометрическое исследование документального потока по физико-математическим и некоторым другим наукам, отраженным в журнале ВИНИТИ РАН / В. Г. Шамаев. – НТИ. Сер. 2. – 2011. – № 1. – С. 24–30.