Учреждение образования "Белорусский государственный университет культуры и искусств"
Кафедра информационных ресурсов

СОГЛАСОВАНО
Зав. кафедрой информационных ресурсов
____________________Р.А.Ровина

СОГЛАСОВАНО
Декан факультета информационно-документных коммуникаций
_____________Н.А.Яцевич

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ПО ДИСЦИПЛИНЕ МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ МЕТОДЫ БИБЛИОТЕЧНО-ИНФОРМАЦИОННОЙ ДЕЯТЕЛЬНОСТИ

для специальности 1-23 01 11 Библиотечно-информационная деятельность (по направлениям).

Автор-составитель: ***В.В.Нешитой,*** доктор технических наук, профессор

Рассмотрено и утверждено
на заседании совета____20 г.,
протокол № _____

Минск 2016

СТРУКТУРА УЧЕБНО-МЕТОДИЧЕСКОГО КОМПЛЕКСА

I. ВВЕДЕНИЕ

Учебно-методический комплекс разработан для студентов *факультета информационно-документных коммуникаций* в соответствии с требованиями образовательного стандарта по специальности 1-23 01 11 Библиотечно-информационная деятельность (по направлениям).

Актуальность изучения дисциплины обусловлена необходимостью использования статистических методов анализа и обработки информации в частности, определения степени использования фонда библиотеки, информационной полноты его комплектования, выявления ядра фонда и зон рассеяния, вычисления оптимального объема фонда и решения ряда других задач. Для этого требуется использование не только известных математических моделей, но и разработка новых моделей (они найдены автором и изложены в его Теории обобщенных распределений, которая защищена докторской диссертацией в ВИНИТИ в 1990г).

Использование таких моделей в учебном процессе и практической работе требует от студента и специалиста ознакомления с базовыми понятиями высшей математики, такими как функция, виды функций, производная функции, производные элементарных функций, правила дифференцирования суммы и разности функций, их произведения, частного, дифференцирование сложных функций. Эти задачи студент должен уметь свободно и грамотно решать.

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Целью изучения дисциплины является формирование у студентов общего подхода к вопросам построения универсальных вероятностных моделей для моделирования библиотечного фонда, описания статистических закономерностей текста и информационных потоков, методов оценивания параметров и овладение современными методами статистического анализа на базе обобщённых распределений и кривых роста.

Задачей изучения дисциплины является обучение студентов навыкам использования различных видов вероятностных моделей текста и информационных потоков, в том числе обобщённых распределений и кривых роста для описания широкого класса статистических распределений и кривых роста в библиотечно-информационной деятельности.

В результате изучения дисциплины студенты должны знать:

– известные и новые методы оценивания параметров;

– сущность различных подходов к вычислению по статистическим данным выравнивающих распределений и кривых роста;

– элементы теории обобщённых распределений;

– математико-статистические модели, которые могут быть использованы в библиотечно-информационной деятельности, в том числе универсальные законы рассеяния и старения публикаций.

Выпускники в пределах своей специальности должны уметь:

– использовать математико-статистические методы для построения моделей библиотечных процессов;

– вычислять законы распределения по статистическим данным;

– вычислять оценки параметров универсальных законов рассеяния и старения публикаций;

– прогнозировать статистические закономерности текста и информационных потоков.

– извлекать информацию из библиотечной статистики.

На изучение дисциплины «Математико-статистические методы библиотечно-информационной деятельности» отводится 34 часа, из них 28 часов аудиторных занятий, в том числе 18 часов лекций, 10 часов практических занятий и 6 часов – самостоятельная работа студентов. Курс рассчитан на один семестр. Форма отчётности и контроля – зачёт.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА ДИСЦИПЛИНЫ

№ п/п	Названия разделов и тем	Количество часов			
		всего аудит	лекц.	практ.	СРС
1.	Функции одной переменной. Производная функции.	4	2	2	
2.	Некоторые понятия теории вероятностей и математической статистики.	2	2		
3.	Методы построения обобщенных непрерывных распределений (в т.ч.по кривым роста новых событий)	4	4		
4.	Классические методы оценивания параметров непрерывных распределений	4	2	2	2
5.	Универсальный метод моментов вычисления закона распределения и оценок параметров	4	2	2	
6.	Общий устойчивый метод вычисления закона распределения и оценок параметров	4	2	2	2
7.	Ранговые распределения в библиотечно-информационной деятельности	4	2	2	2
8.	Построение системы дискретных распределений, оценивание параметров	2	2		
	Всего	**34**	**18**	**10**	**6**

СОДЕРЖАНИЕ (ПРОГРАММА) ДИСЦИПЛИНЫ

Тема 1. Функции одной переменной. Производная функции.

Функция. Способы задания функций. Основные элементарные функции и их графики. Гамма-функция. Понятие эмпирической функции.

Задачи, приводящие к понятию производной. Определение и смысл производной. Касательная. Вычисление производных. Производные элементарных функций. Правила дифференцирования.

Тема 2. Некоторые понятия теории вероятностей и математической статистики.

Предмет теории вероятностей. Случайное событие. Испытание. Относительная частота и вероятность. Основные теоремы теории вероятностей. Непрерывные и дискретные случайные величины и их законы распределения. Числовые характеристики случайных величин.

Предмет математической статистики. Генеральная и выборочная совокупности. Простой статистический ряд. Вариационный ряд. Интервальный ряд распределеня. Гистограмма, полигон. кривая распределения. Плотность и функция распределения.

Тема 3. Методы построения обобщенных непрерывных распределений.

Понятия математического ожидания случайной функции, нового события и кривой роста новых событий. Связь кривой роста с законами распределения вероятностей разных и новых событий. Установление статистической структуры выборки по кривой роста новых событий. Восстановление кривой роста новых событий по статистической структуре выборки. Формулы В.Калинина. Построение непрерывных распределений по методам: К.Пирсона, обобщения, как распределения функций случайных аргументов, Классификация распределений и кривых роста.

Тема 4. Классические методы оценивания параметров непрерывных распределений.

Метод наименьших квадратов. Метод наибольшего правдоподобия. Классический метод моментов. Критерии для установления типа выравнивающей кривой по методу моментов.

Практические занятия

Для выравнивания статистического распределения использовать обобщённые распределения группы А.

Оценить параметры выравнивающего распределения группы А по методу наименьших квадратов.

Установить тип выравнивающего распределения группы Б по показателям U, L.

Найти оценки параметров по классическому методу моментов.

Тема 5. Универсальный метод моментов вычисления закона распределения и оценок параметров

Критерии для установления типа выравнивающей кривой. Вычисление оценок параметров. Номограмма (типы кривых в координатах B1, B2). Расширение трёх систем непрерывных распределений. Центральная предельная теорема теории вероятностей. Простейшее доказательство.

Практические занятия

По статистическому распределению вычислить среднее значение случайной величины, центральные моменты (2 – 4)-го порядков.

Вычислить показатели асимметрии B1 и островершинности B2.

Вычислить критерий В.Нешитого L1.

С помощью номограммы установить тип выравнивающего распределения.

Вычислить оценки параметров по универсальному методу моментов.

Вычислить значения плотности вероятностей и сравнить с эмпирическими значениями.

Тема 6. Общий устойчивый метод вычисления закона распределения и оценок параметров.

Понятие устойчивости. Критерии для установления типа выравнивающей кривой по устойчивому методу. Номограмма (типы кривых в координатах B, H). Вычисление оценок параметров.

Практические занятия

По статистическому распределению вычислить критерии B, H.

С помощью номограммы установить тип выравнивающего распределения.

Вычислить оценки параметров по общему устойчивому методу.

Вычислить значения плотности вероятностей.

Оценить с помощью критерия «хи-квадрат» степень близости выравнивающей кривой и статистического распределения.

Тема 7. Ранговые распределения в библиотечно-информационной деятельности

Построение эмпирической кривой рангового распределения.

Форма представления ранговых распределений.

Критерий однородности ранговых распределений.

Выделение неоднородной части. Законы распределения документальных информационных потоков (ранговые распределения журналов). Универсальные законы старения и рассеяния публикаций. Оптимальный объем библиотечного фонда.

Практические занятия

Рассмотреть 1–2 примера вычисления ранговых законов распределения по статистическим распределениям.

Показать, что вторая система непрерывных распределений В. Нешитого является универсальным законом рассеяния публикаций.

Тема 8. Построение системы дискретных распределений, оценивание параметров.

Построение системы дискретных распределений по системе непрерывных распределений. Построение системы дискретных распределений по кривым роста новых событий на основе формулы В.Калинина. Классификация дискретных распределений. Критерий степени неравномерности появления событий. Ранжирование слов по степени семантической нагрузки. Порядок установления типа выравнивающего распределения и нахождения оценок параметров.

Практические занятия

По статистическим распределениям слов в подвыборках одинакового объёма вычислить типы выравнивающих дискретных распределений, вычислить оценки параметров и ранжировать слова по степени семантической нагрузки.

Установить по статистическому ранговому распределению тип выравнивающего дискретного распределения, найти оценки параметров.

II.ТЕОРЕТИЧЕСКИЙ РАЗДЕЛ

КРАТКИЕ ТЕКСТЫ ЛЕКЦИЙ

Тема 1. Функции одной переменной. Производная функции.

Производные простых функций.

К понятию производной приводят многие задачи, например нахождение уравнения касательной к некоторой кривой в заданной точке, вычисление скорости изменения функции и ускорения в заданной точке, вычисление сечения бруса, имеющего наибольшую жесткость или прочность при его изготовлении из круглого бревна и многие другие.

Геометрический и физический смысл производной – это тангенс угла наклона касательной к кривой в некоторой точке или скорость движения.

Поизводной функции y=f(x) по аргументу x называется предел отношения приращения функции к приращению аргумента при условии, что последнее стремится к нулю. Производная функции y=f(x) обозначается через y' или $f'(x)$.

Операция отыскания производной $f'(x)$ данной функции f(x) называется дифференцированием этой функции. Исходя из определения производной, можно получать формулы для вычисления производных элементарных функций. Найдем для примера производную функции $y = x^2$. Дадим аргументу приращение Δx . Тогда будем иметь

$y + \Delta y = (x + \Delta x)^2 = x^2 + 2x\Delta x + (\Delta x)^2$. Отсюда найдем:

$\Delta y = x^2 + 2x\,\Delta x + (\Delta x)^2 - x^2 = 2x\Delta x + (\Delta x)^2$.

Далее находим предел отношения $\Delta y/\Delta x$ при $\Delta x \to 0$

$\lim \frac{\Delta y}{\Delta x} = 2x + \Delta x = 2x.$

Если найти таким же путем производную от x^3, то получим $(x^3)' = 3x^2$,

а в общем случае производная от $x^n = nx^{n-1}$

Приведем небольшую таблицу производных некоторых функций.

Производная постоянной величины равна нулю, потому что прямая y=c параллельна горизонтальной оси, а тангенс нуля градусов равен нулю.

$c' = 0$,

$x' = 1$, (прямая y=x – это биссектрисса координатного ула, а тангенс 45 градусов равен единице.

$(x^n)' = nx^{n-1}$

$(\ln x)' = \frac{1}{x}$, $x > 0$

$(e^x)' = e^x$

$(a^x)' = a^x \ln a$

$(\sin x)' = \cos x$

$(\cos x)' = -\sin x$

$(\mathrm{tg}x)' = \frac{1}{\cos^2 x}$

$(ctgx)' = -\frac{1}{\sin^2 x}$

Правила дифференцирования

Постоянный множитель необходимо выносить за знак производной, например $y = cx^3$. Тогда

$y' = c(x^3)' = c(3x^2)$.

y=2x; $y' = 2$.

Производная суммы или разности функций равна сумме или разности производных слагаемых:

$y = (2x + 5x^2 + 3)'$

$y' = (2x)' + (5x^2)' + 3' = 2 + 5 * 2x + 0 = 2 + 10x$.

Производная произведения

y=u*v; $y' = u'v + uv'$

Производная частного

$\left(\frac{u}{v}\right)' = \frac{u'v - v'u}{v^2}$.

Производные сложных функций

$y = (ax^5 + bx^3 + c)^3$

$y' = 3(ax^5 + bx^3 + c)^2 \cdot (ax^5 + bx^3 + c)' =$
$= 3(ax^5 + bx^3 + c)^2 \cdot (5ax^4 + 3bx^2)$.

Логарифмическое дифференцирование.

Пример $y = u^v$

Логарифмируем последнее равенство

lny=vlnu

Берем производные от обеих частей последнего равенства

$\frac{y'}{y} = v' \cdot \ln u + v\frac{u'}{u}$

Далее умножаем обе части равенства на y и, заменяя y на u^v, получим

$$y' = vu^{v-1} \cdot u' + u^{v} \cdot \ln u \cdot v' \quad .$$

Тема 2 . Некоторые понятия теории вероятностей и математической статистики

Случайные события и их вероятности

2.1. Случайные события. Испытания. Относительная частота и вероятность

Пусть требуется оценить качество изделий в некоторой партии объемом n. Для этого необходимо над каждым изделием провести наблюдение, т.е. осмотр, измерение, взвешивание и т.д. В теории вероятностей и математической статистике всем этим понятиям соответствует один термин – испытание.

В результате отдельного испытания изделие может быть признано либо годным, либо браком. Возможные **исходы испытания** в данном примере – это **случайные события:** A – годное изделие; B – брак. Эти события называются случайными, потому что заранее нельзя точно предсказать, какое из них наступит при следующем испытании.

Пусть после проверки всей партии изделий объемом n, т.е. после n испытаний, случайное событие A – число годных изделий – появилось n_A раз. Это значит, что относительная частота случайного события A равна

$$w_A = n_A / n \ .$$

Если провести несколько серий испытаний (проверить несколько партий изделий), то относительные частоты в разных сериях будут группироваться около определенного числа, которое называется **вероятностью** случайного события A и обозначается $P(A)$. Как показала практика, с **ростом объема партии изделий *n* относительные частоты теснее группируются около вероятности,** т.е. обнаруживают **устойчивость**.

Устойчивость относительной частоты случайного события является определяющим его свойством, позволяющим использовать относительную частоту как оценку вероятности в различных практических расчетах.

2.2. Виды случайных событий

События, которые непременно происходят при каждом испытании, называются **достоверными**.

События, которые не могут произойти ни при каком испытании, называются **невозможными.**

Вероятность достоверного события равна единице, вероятность невозможного события равна нулю.

Если при осуществлении испытания может наступить хотя бы одно из двух событий А или В, то событие

$$C = A + B$$

называется **суммой**, или **объединением** событий A и B.

Два события A и B называются **несовместными**, если они не могут наступить вместе при одном испытании.

Случайные события образуют **полную группу**, если они попарно несовместны и при любом отдельном испытании непременно должно произойти одно из них.

Сумма вероятностей событий, образующих полную группу, равна единице.

Два случайных события называются **противоположными**, если в одном испытании появление одного из них (A) исключает появление другого ($\bar{A}$ - читается не A).

Сумма вероятностей двух противоположных событий равна единице

$$P(A) + P(\overline{A}) = 1.$$

Противоположные события образуют полную группу.

Если при осуществлении испытания может наступить и событие A, и событие B (совмещение событий A и B), то событие

$$C = A \cdot B$$

называется произведением, или пересечением событий A и B.

Два случайных события называются **независимыми**, если при осуществлении испытаний появление одного из них не изменяет вероятности появления другого.

2.3. Определения вероятности

Классическое определение вероятности события A – отношение числа m элементарных событий (исходов испытаний), благоприятствующих событию A, к общему числу n равновозможных элементарных событий

$$P(A) = \frac{m}{n}.$$

Статистическое определение вероятности

$$P(A) = \frac{n_A}{n},$$

где n_A - частота события A при n испытаниях.

Геометрическая вероятность

$$P(A) = \frac{S_A}{S},$$

где S_A - площадь некоторого замкнутого контура, составляющая часть площади S.

2.4. Основные формулы комбинаторики

Используются для вычисления вероятностей событий.

Перестановки – комбинации из n различных элементов, отличающиеся лишь порядком. Число перестановок вычисляется по формуле

$$P_n = n! = 1.2...n.$$

Число перестановок из n элементов по m, где каждый элемент может использоваться от 0 до m раз, равно

$$P_{n,m} = n^m.$$

Например, при $n = 2$, $m = 8$, $P_{n,m} = 2^8 = 256$.

Размещения - комбинации из n различных элементов по m элементов, которые различаются либо составом элементов, либо их порядком

$$A_n^m = n(n-1)(n-2)...(n-m+1) = \frac{n!}{(n-m)!}.$$

Сочетания – комбинации из n различных элементов по m элементов, различающиеся хотя бы одним элементом

$$C_n^m = \frac{n!}{m!(n-m)!} = \frac{A_n^m}{P_m} = \frac{n(n-1)...(n-m+1)}{1.2...m} = C_n^{n-m}.$$

При этом $C_n^0 = C_n^n = 1$.

Пример.

В партии из N деталей M стандартных. Выбираются n деталей. Требуется найти вероятность того, что m деталей будут стандартными.

Решение. Общее число возможных исходов равно числу способов, которыми можно взять n деталей из N. Это число равно числу сочетаний из N по n, т.е. C_N^n.

Найдем далее число благоприятствующих исходов. Поскольку m стандартных деталей выбираются из общего их числа M, то число таких комбинаций равно C_M^m. Остальные n-m нестандартных деталей выбираются из N–M нестандартных деталей – это C_{N-M}^{n-m} комбинаций. Число благоприятствующих исходов равно произведению $C_M^m C_{N-M}^{n-m}$. Следовательно,

$$P(A = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

Это – известное **гипергеометрическое распределение** .

2.5. Основные теоремы теории вероятностей

2.5.1. Теорема сложения вероятностей (несовместных событий)

Пусть A и B – несовместные события. **Вероятность суммы двух несовместных событий равна сумме вероятностей этих событий**

$$P(A+B) = P(A) + P(B).$$

Для нескольких несовместных событий имеем

$$P(A_1 + A_2 + ... + A_n) = P(A_1) + P(A_2) + ... + P(A_n).$$

Для совместных событий

$$P(A+B) = P(A) + P(B) - P(AB),$$

где $P(AB)$ – вероятность совместного появления событий A и B.

2.5.2. Теорема умножения вероятностей (независимых событий)

Вероятность произведения (совмещения) двух независимых событий равна произведению вероятностей этих событий

$$P(AB)=P(A)P(B).$$

Вероятность произведения двух зависимых событий равна произведению вероятности одного из них на условную вероятность другого, вычисленную при условии, что первое имело место

$$P(AB)=P(A)P(B/A)=P(B)P(A/B).$$

Пример. В урне 2 белых и 3 черных шара. Вынимаем подряд 2 шара. Какова вероятность того, что оба шара белые, т.е. $A=A_1A_2$.

Решение. A_1 – появление белого шара при 1-м испытании; A_2 – появление белого шара при 2-м испытании

$$P(A)=P(A_1)P(A_2/A_1)=\frac{2}{5}\ \ \frac{1}{4}=0{,}1.$$

Следствие теоремы умножения вероятностей.

Вероятность появления хотя бы одного события из событий A_1, $A_2, \ldots, A_n$, независимых в совокупности, равна разности между единицей и произведением вероятностей противоположных событий $\overline{A}_1, \overline{A}_2, \ldots, \overline{A}_n$:

$$P(A)=1-q_1 \cdot q_2 \ldots\ q_n.$$

В частном случае, при $q_1=q_2=\ldots=q_n=q$

$$P(A)=1-q^n.$$

Пример. Вероятности попадания в цель каждого из трех стрелков равны: $p_1=0{,}8$; $p_2=0{,}7$; $p_3=0{,}9$. Найти вероятность хотя бы одного попадания при одном залпе.

Решение. Вероятности промахов равны: $q_1=1-p_1=0{,}2$; $q_2=0{,}3$; $q_3=0{,}1$. Следовательно,

$$P(A)=1-q_1q_2q_3=0{,}994.$$

2.5.3. Формула полной вероятности

Следствием теорем сложения и умножения вероятностей является формула полной вероятности.

Пусть некоторое событие A может произойти вместе с одним из событий $H_1, H_2, \ldots, H_n$, причем последние образуют полную группу несовместных событий. Их называют **гипотезами**.

Приведем без доказательства формулу для вычисления вероятности события A. Она равна сумме произведений вероятности каждой гипотезы на вероятность события при этой гипотезе [2].

$$P(A)=\sum_{i=1}^{n} P(H_i)P(A/H_i).$$

Другими словами, **формула полной вероятности определяет средневзвешенную по всем гипотезам вероятность наступления некоторого события *A***.

Пример. Есть два набора деталей. Вероятность того, что деталь первого набора стандартна, равна 0,8, а второго – 0,9. Найти вероятность того, что взятая наудачу деталь из наудачу взятого набора стандартна, т.е. $P(A)=?$

Решение. Событие H_1 – деталь взята из первого набора; $P(H_1)=1/2$.

Событие H_2 - деталь взята из второго набора; $P(H_2)=1/2$.

Далее, вероятность события A при первой гипотезе $P(A/H_1)=0{,}8$; при второй гипотезе $P(A/H_2)=0{,}9$.

Средневзвешенная вероятность события A по двум гипотезам равна

$$P(A)=0{,}8\cdot 0{,}5+0{,}9\cdot 0{,}5=0{,}85\,.$$

2.5.4. Теорема гипотез (формула Бейеса)

Пусть имеется полная группа несовместных гипотез $H_1,\dots,H_n$, при этом вероятности их равны $P(H_i)$.

Кроме того, известны вероятности некоторого события A, которое может произойти совместно с каждой гипотезой.

Пусть в результате опыта наступило событие A.

Тогда распределение условных вероятностей гипотез при наступлении события A задается формулой Бейеса

$$P(H_i/A)=\frac{P(H_i)P(A/H_i)}{\sum_{i=1}^{n}P(H_i)P(A/H_i)}\,.$$

Рассмотрим пример на формулу полной вероятности и формулу Бейеса, позаимствованный из [3].

Три станка выпускают одинаковые детали (см. табл.).

№ станка	**Дневная выработка (деталей)**	**Вероятност ь гипотезы**	**% бра ка**	**Вероятность брака**
1	$m_1=600$	$P(H_1)=0{,}6$	3	$P_1=P(A/H_1)=0{,}03$
2	$m_2=100$	$P(H_2)=0{,}1$	5	$P_2=P(A/H_2)=0{,}05$
3	$m_3=300$	$P(H_3)=0{,}3$	10	$P_3=P(A/H_3)=0{,}10$

$\sum m_i=1000$

На складе продукция трех станков смешивается. Далее выбирается случайным образом одна деталь.

Требуется:

а) *найти вероятность того, что она бракованная*.

Здесь используется формула полной вероятности.

Вероятности гипотез равны (см. табл.):

$$P(H_1)=\frac{m_1}{m_1+m_2+m_3}=\frac{600}{1000}=0{,}6;$$
$$P(H_2)=0{,}1;\ P(H_3)=0{,}3.$$

Вероятности брака при каждой гипотезе равны:

$$P(A/H_1)=0{,}03;\qquad P(A/H_2)=0{,}05;\qquad P(A/H_3)=0{,}10\,.$$

Тогда

$$P(A)=\sum_{i=1}^{3}P(H_i)P(A/H_i)=0{,}6\cdot 0{,}03+0{,}1\cdot 0{,}05+0{,}3\cdot 0{,}1=0{,}053\,.$$

Формула полной вероятности в данном примере определяет средневзвешенную вероятность брака по трем станкам. Действительно, записав ее в более простом виде, находим

$$P(A) = \frac{m_1 p_1 + m_2 p_2 + m_3 p_3}{m_1 + m_2 + m_3} = \overline{p},$$

где $\mathrm{p}_i = P(A/H_i)$; $\quad \dfrac{m_i}{m_1 + m_2 + m_3} = P(H_i)$.

б) *Найти вероятность того, что случайно отобранная деталь, оказавшаяся бракованной, выпущена первым станком*.

По формуле Бейеса находим

$$P(H_1/A) = \frac{P(H_1)\ P(A/H_1)}{\sum_{i=1}^{n} P(H_i)\ P(A/H_i)} = \frac{0,6 \cdot 0,3}{0,6 \cdot 0,3 + 0,1 \cdot 0,05 + 0,3 \cdot 0,1} = \frac{0,018}{0,053} = 0,34\ .$$

Формула Бейеса в данном примере определяет долю бракованных изделий (в общем объеме брака), изготовленных одним *i*-м станком.

$$P(H_1/A) = \frac{m_1 p_1}{m_1 p_1 + m_2 p_2 + m_3 p_3} = \frac{600 \cdot 0,03}{600 \cdot 0,03 + 100 \cdot 0,05 + 300 \cdot 0,1} = \frac{18}{53} = 0,034\ .$$

То же для других станков

$$P(H_2/A) = \frac{5}{53} = 0,094; \qquad P(H_3/A) = \frac{30}{53} = 0,566\ .$$

2.6. Дискретные случайные величины

Для случайных величин приняты обозначения *X, Y, Z,…*

Возможные значения случайной величины *X* обозначаются строчными буквами $x_1, x_2, \ldots, x_n$.

Дискретной называют случайную величину, которая принимает отдельные изолированные возможные значения с определенными вероятностями (например, число отказавших приборов) в отличие от **непрерывной** случайной величины, которая может принимать все значения из некоторого конечного или бесконечного промежутка (например, время безотказной работы прибора).

Возможные значения прерывных (дискретных) величин могут быть заранее перечислены, а непрерывных – не могут быть перечислены.

2.6.1. Закон распределения вероятностей дискретной случайной величины

Закон распределения случайной величины – это соответствие между возможными значениями случайной величины и их вероятностями.

Его можно задать **таблично, аналитически и графически**:

а) табличная форма закона распределения в виде ряда распределения

X	x_1	x_2	…	x_n
p	p_1	p_2	…	p_n

б) аналитическая форма

$$p_i = f(x_i)\ .$$

в) графическая форма – в виде многоугольника распределения.

На оси абсцисс откладываются значения случайной величины и строятся отрезки, равные по высоте вероятностям. Вершины отрезков для наглядности соединяются ломаной.

2.6.2. Числовые характеристики дискретной случайной величины

Математическое ожидание

Математическое ожидание дискретной случайной величины равно сумме произведений всех ее возможных значений на их вероятности

$$M(X) = x_1 p_1 + x_2 p_2 + ... + x_n p_n .$$

Математическое ожидание есть **неслучайная** (постоянная) величина.

Пример 1. Найти математическое ожидание случайной величины X по ее закону распределения:

X	3	5	2
$\boldsymbol{p}$	0,1	0,6	0,3

Решение. $M(X) = 3 \cdot 0{,}1 + 5 \cdot 0{,}6 + 2 \cdot 0{,}3 = 3{,}9$.

Пример 2. Найти математическое ожидание числа появлений события А в одном испытании, если вероятность события A равна p.

Решение. Случайная величина X – число появлений события A в одном испытании – может принимать два значения: x_1 =1 (событие наступило) с вероятностью p и x_2 = 0 (событие не наступило) с вероятностью $1–p=q$.

Следовательно,

$$M(X) = 1 \cdot p + 0 \cdot q = p ,$$

т.е. **математическое ожидание числа появлений события в одном испытании равно вероятности этого события**.

Оценкой математического ожидания является среднее арифметическое наблюденных значений случайной величины.

Свойства математического ожидания

Приведем без доказательства основные свойства математического ожидания.

1. $M(C)=C$ – математическое ожидание постоянной величины C равно значению самой постоянной.

2. $M(CX)=CM(X)$ – постоянную величину можно выносить за знак математического ожидания.

3. $M(XY)=M(X)M(Y)$ – для двух независимых случайных величин математическое ожидание произведения равно произведению их математических ожиданий.

4. $M(X+Y)=M(X)+M(Y)$ – для двух случайных величин (зависимых или независимых) математическое ожидание суммы равно сумме математических ожиданий слагаемых.

Математическое ожидание числа появлений события A в n независимых испытаниях равно произведению числа испытаний n на вероятность появления события в каждом испытании p, т.е.

$$M(X)=np.$$

Доказательство свойств математического ожидания см., например, в учебниках [2,3]

Дисперсия дискретной случайной величины

Две случайные величины могут иметь одинаковые математические ожидания, но разное рассеяние. Это значит, что математическое ожидание полностью случайную величину не характеризует. Поэтому вводится еще одна числовая характеристика, которая называется **дисперсией** и характеризует рассеяние случайной величины относительно математического ожидания.

Дисперсией дискретной случайной величины *X* называется математическое ожидание квадрата отклонения случайной величины *X* от ее математического ожидания

$$D(X)=M\left[X-M(X)\right]^2.$$

Пример. Случайная величина X имеет распределение

X	1	2	5
p	0,3	0,5	0,2

Требуется вычислить дисперсию.

Имеем:

$$M(X)=1\cdot 0{,}3+2\cdot 0{,}5+5\cdot 0{,}2=2{,}3.$$

$$D(X)=\sum_{i=1}^{3}\left(x_i-M(X)\right)^2 p_i=(1-2{,}3)^2\cdot 0{,}3+(2-2{,}3)^2\cdot 0{,}5+$$
$$+(5-2{,}3)^2\cdot 0{,}2=2{,}01.$$

На практике для вычисления дисперсии используют другую, более удобную формулу

$$D(X)=M(X^2)-\left[M(X)\right]^2.$$

Доказательство:

$$\mathrm{D(X)}=\mathrm{M[X-M(X)]}^2=M\left[X^2-2XM(X)+\left(M(X)\right)^2\right]=$$
$$=M(X^2)-2M(X)M(X)+\left[M(X)\right]^2=M(X^2)-\left[M(X)\right]^2.$$

2.6.3. Свойства дисперсии

1. $D(C)=0$, т.е. дисперсия постоянной величины C равна нулю.

2. $D(CX)=C^2D(X)$ – постоянный множитель можно выносить за знак дисперсии, возведя его в квадрат.

Доказательство:

$$D(CX)=M\left[CX-M(CX)\right]^2=M\left\{C^2\left[X-M(X)\right]^2\right\}=C^2D(X).$$

3. $D(X+Y)=D(X)+D(Y)$ – дисперсия суммы двух случайных величин равна сумме их дисперсий.

Доказательство:

$$D(X+Y)=M(X+Y)^2-[M(X+Y)]^2=M(X^2+2XY+Y^2)-[M(X)+M(Y)]^2=$$
$$=M(X^2)+2M(XY)+M(Y^2)-[(M(X))^2+2M(X)M(Y)+(M(Y))^2].$$

Так как $M(XY)=M(X)M(Y)$, то последнее равенство примет вид

$$D(X+Y)=M(X^2)-[M(X)]^2+M(Y^2)-[M(Y)]^2,$$

откуда

$$D(X+Y)=D(X)+D(Y).$$

Приведем без доказательства еще два свойства дисперсии.

4. $D(C+X)=D(X)$.

5. $D(X-Y)=D(X)+D(Y)$.

Отметим еще одно важное свойство дисперсии.

Дисперсия числа появлений события A в n независимых испытаниях равна

$$D(X)=npq.$$

Доказательство:

Найдем дисперсию числа появлений события А в одном испытании

$$D(X_1)=M(X_1^2)-[M(X_1)]^2.$$
$$M(X_1)=1\cdot p+0(1-p)=p.$$
$$M(X_1^2)=1^2\cdot p+0^2(1-p)=p.$$

Тогда $D(X_1)=p-p^2=p(1-p)=pq$.

Всего n испытаний, следовательно, $D(X)=npq$.

Дисперсия имеет размерность случайность величины в квадрате.

2.6.4. Среднее квадратическое отклонение

Если извлечь из дисперсии квадратный корень, получим **среднее квадратическое отклонение**

$$\sigma(X)=\sqrt{D(X)}.$$

Размерность величины $\sigma(X)$ та же, что и случайной величины X.

Пример. По распределению

X	2	3	10
p	0,1	0,4	0,5

требуется вычислить среднее квадратическое отклонение.

Решение.

$$M(X)=2\cdot 0{,}1+3\cdot 0{,}4+10\cdot 0{,}5=6{,}4.$$
$$M(X^2)=2^2\cdot 0{,}1+3^2\cdot 0{,}4+10^2\cdot 0{,}5=54.$$
$$D(X)=M(X^2)-[M(X)]^2=54-6{,}4^2=13{,}04.$$
$$\sigma(X)=\sqrt{D(X)}=3{,}61.$$

Среднее квадратическое отклонение суммы взаимно независимых случайных величин равно

$$\sigma(X_1+X_2...+X_n)=\sigma(X)=\sqrt{\sigma^2(X_1)+\sigma^2(X_2)+...+\sigma^2(X_n)}$$

Доказательство.

Дисперсия суммы случайных величин равна

$$D(X)=D(X_1)+D(X_2)+...+D(X_n).$$

Тогда

$$\sqrt{D(X)}=\sigma(X)=\sqrt{D(X_1)+...+D(X_n)}=\sqrt{\sigma^2(X_1)+...+\sigma^2(X_n)}.$$

2.6.5. Одинаково распределенные взаимно независимые случайные величины

Рассмотрим n взаимно независимых одинаково распределенных случайных величин $X_1, X_2, ..., X_n$.

Для них среднее арифметическое равно

$$\overline{X}=(X_1+X_2+...+X_n)/n.$$

Докажем три положения [3]:

1. Математическое ожидание среднего арифметического n взаимно независимых одинаково распределенных случайных величин равно математическому ожиданию a каждой из величин

$$M(\overline{X})=a.$$

Доказательство:

$$M(\overline{X})=M\left(\frac{X_1+X_2+...+X_n}{n}\right)=\frac{M(X_1)+M(X_2)+...+M(X_n)}{n}=\frac{na}{n}=a.$$

2. Дисперсия среднего арифметического n взаимно независимых одинаково распределенных случайных величин в n раз меньше дисперсии D каждой из величин:

$$D(\overline{X})=\frac{D(X_i)}{n}.$$

Доказательство:

Так как постоянный множитель можно выносить за знак дисперсии, возведя его в квадрат, то

$$D(\overline{X})=D\left(\frac{X_1+X_2+...+X_n}{n}\right)=\frac{D(X_1)+D(X_2)+...+D(X_n)}{n^2}=\frac{nD(X_i)}{n^2}=\frac{D(X_i)}{n}.$$

3. $\sigma(\overline{X})=\dfrac{\sigma(X_i)}{\sqrt{n}}$ (следует из п.2), т.е. среднее арифметическое n взаимно независимых одинаково распределенных случайных величин имеет значительно меньшее рассеяние (в $\sqrt{n}$ раз), чем каждая отдельная величина.

2.6.6. Моменты (начальные, центральные) дискретной случайной величины

Начальный момент порядка r – это математическое ожидание случайной величины X^r

$$\nu_r=M(X^r)=\sum_{i=1}^{n}x_i^r p_i.$$

Например, начальные моменты первого и второго порядков равны

$$\nu_1=M(X);\ \nu_2=M(X^2).$$

Центральный момент порядка r задается формулой

$$\mu_r = M\left[X - M(X)\right]^r,$$

при этом $\mu_1 = 0; \mu_2 = D(X)$.

Центральный момент второго порядка представляет собой **дисперсию**.

Между начальными и центральными моментами существуют соотношения

$$\mu_2 = \nu_2 - \nu_1^2,$$
$$\mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3,$$
$$\mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4.$$

Следовательно, формула для вычисления дисперсии может быть записана в виде

$$D(X) = \mu_2 = \nu_2 - \nu_1^2.$$

2.6.7. Примеры законов распределения дискретных случайных величин

Гипергеометрическое распределение

$$P_{N,M}(n,m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

В партии из N изделий M стандартных ($M<N$). Из партии отбирают n изделий (**без возврата**).

Случайная величина m – число стандартных изделий среди n отобранных имеет **гипергеометрическое распределение. Оно широко используется в статистических методах контроля качества продукции**.

Биномиальный закон

Если в гипергеометрическом распределении объем партии изделий увеличивать, то гипергеометрическое распределение будет приближаться к **биномиальному закону** ($M/N=p$)

$$P_n(m) = C_n^m p^m (1-p)^{n-m} = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}.$$

Здесь выборка – **с возвращением**!

Закон Пуассона

Следует из биномиального при $n\to\infty$ и малой вероятности p (величина np – постоянная)

$$P_n(m) = \frac{(np)^m}{m! e^{np}},$$

где np – среднее. При $n \le 0,1N$ и q (или p) $\le$ 0,1 закон Пуассона можно использовать вместо гипергеометрического.

2.7. Непрерывные случайные величины

2.7.1 Функция распределения

Функцией распределения (интегральной функцией распределения) случайной величины X называется функция $F(x)$, значения которой равны вероятностям $P(X<x)$

$$\boldsymbol{F(x)=P(X<x)=P(-\infty<X<x).}$$

Из этого определения вытекают следующие свойства функции распределения:

1. $0\le F(x)\le 1$.
2. $F(b)\ge F(a)$ при $b>a$, т.е. функция распределения - неубывающая.
3. $P(a\le X<b)=F(b)-F(a)$.
4. $\lim\limits_{x\to-\infty} F(x)=0,\quad \lim\limits_{x\to\infty} F(x)=1$, если распределение задано на всей числовой оси.
5. Если между двумя случайными величинами X и Y существует функциональная зависимость $Y=\varphi(X)$, причем, с ростом X монотонно растет и Y, то их функции распределения равны

$$F(x)=F(y)=F(\varphi(x)).$$

6. Если с ростом X величина Y монотонно убывает, то

$$F(x)=1-F(y)=1-F(\varphi(x)).$$

2.7.2 Плотность распределения

Плотностью распределения (дифференциальным законом распределения) непрерывной случайной величины X называется первая производная от функции распределения

$$p(x)=F'(x)=\frac{dF(x)}{dx}.$$

График плотности распределения называется кривой распределения.

Свойства плотности распределения.

1. $p(x)\ge 0$.
2. $\int\limits_{-\infty}^{\infty} p(x)dx=1$.
3. $P(a\le X<b)=\int\limits_{a}^{b} p(x)dx.$
4. $F(x)=\int\limits_{-\infty}^{x} p(x)dx.$
5. Если между двумя случайными величинами X и Y существует функциональная зависимость $Y=\varphi(X)$, то взаимосвязь между плотностями распределения $p(x)$ и $p(y)$ задается формулой

$$p(x)=p(y)\left|\frac{dy}{dx}\right|=p(\varphi(x))\left|\frac{d\varphi(x))}{dx}\right| \quad .$$

Действительно, пусть с ростом X растет Y. Тогда $F(x)=F(y)=F(\varphi(x))$ (см. свойства функции распределения). По правилу дифференцирования сложной функции находим

$$p(x)=\frac{dF(x)}{dx}=\frac{dF(y)}{dx}=\frac{dF(y)}{dy}\frac{dy}{dx}=p(y)\frac{dy}{dx}$$

или, поскольку $y = \varphi(x)$,

$$p(x)=\frac{dF(\varphi(x))}{d\varphi(x)}\frac{d\varphi(x)}{dx}=p(\varphi(x))\frac{d\varphi(x)}{dx}.$$

В случае, если с ростом X величина Y убывает, первая производная $dy/dx < 0$, но плотность $p(x) > 0$. Поэтому в общем случае первая производная берется по абсолютной величине.

Рассмотрим примеры.

Пример 1. Из равенства функций распределения $F(x) = F(\ln x)$ требуется найти соотношение между плотностями $p(x)$ и $p(\ln x)$.

Дифференцируя последнее равенство по x, имеем:

$$p(x)=\frac{dF(x)}{dx}=\frac{dF(\ln x)}{d\ln x}\frac{d\ln x}{dx}=\frac{p(\ln x)}{x},$$

откуда $xp(x) = p(\ln x)$.

Пример 2. Задана плотность распределения (**показательный закон**)

$$p(x)=Ne^{-\alpha x} \quad (0<x<\infty).$$

Найти: N – нормирующий множитель; $F(x)$ – функцию распределения; вероятность попадания случайной величины X на интервал $3<x<5$.

Используем свойства плотности:

$$\int_0^\infty p(x)dx=\int_0^\infty Ne^{-\alpha x}dx=\frac{N}{-\alpha}e^{-\alpha x}\Big|_0^\infty=-\frac{N}{\alpha e^{\alpha x}}\Big|_0^\infty=\frac{N}{\alpha}=1.$$

Отсюда $N = \alpha$, т.е.

$$p(x)=\frac{\alpha}{e^{\alpha x}}.$$

Далее функция распределения равна

$$F(x)=\int_0^x p(x)dx=\int_0^x\frac{\alpha}{e^{\alpha x}}=1-\frac{1}{e^{\alpha x}}.$$

Вероятность попадания случайной величины X в заданный интервал равна

$$P(3<x<5)=F(5)-F(3)=\frac{1}{e^{3\alpha}}-\frac{1}{e^{5\alpha}}.$$

Пусть далее некоторая случайная величина Y связана со случайной величиной X зависимостью $Y=1/X$.

Найдем функцию распределения $F(y)$ и плотность $p(y)$.

Так как здесь с ростом X величина Y убывает, то

$$F(y)=1-F(x)=1/e^{\alpha x}.$$

Но $X=1/Y$, поэтому

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

Отсюда плотность *p(y)* равна

$$p(y)=\frac{dF(y)}{dy}=\frac{\alpha}{y^2e^{\alpha/y}}.$$

Плотность *p(y)* можно найти непосредственно по плотности *p(x)*. Поскольку $X=1/Y,\ dx/dy=-1/y^2$, то

$$p(y)=p(x)\left|\frac{dx}{dy}\right|=\frac{\alpha}{e^{\alpha x}}\frac{1}{y^2}=\frac{\alpha}{y^2e^{\alpha/y}}.$$

2.7.3. Числовые характеристики непрерывных случайных величин

Для возможно более полного и всестороннего описания случайных величин используют различные показатели. К ним относятся:

характеристики **положения** – математическое ожидание, мода, медиана;

характеристики **вариации** – дисперсия, среднее квадратическое отклонение, коэффициент вариации;

характеристики **формы распределения** – коэффициенты асимметрии и островершинности, которые выражаются через моменты.

Математическое ожидание случайной величины *X* задается интегралом

$$M(X)=\int_{-\infty}^{\infty}xp(x)dx.$$

Свойства математического ожидания непрерывной случайной величины те же, что и дискретной случайной величины.

Мода – такое значение случайной величины, при котором плотность максимальна.

Медиана *(Me)* случайной величины *X* определяется соотношением

$$P(X<Me)=P(X>Me).$$

Она делит площадь под кривой распределения пополам.

Дисперсия непрерывной случайной величины *X* задается формулой

$$D(X)=M\left[X-M(X)\right]^2=\int_{-\infty}^{\infty}(x-m_x)^2p(x)dx,$$

где $m_x=M(X)$.

Коэффициент вариации

$$V=\frac{\sigma(X)}{m_x}\cdot 100\%$$

- выраженное в процентах отношение среднего квадратического отклонения случайной величины *X* к ее математическому ожиданию.

Центральные моменты *r*-го порядка (*r* = 2,3,4) задаются формулой

$$\mu_r=M\left[X-M(X)\right]^r=\int_{-\infty}^{\infty}(x-m_x)^rf(x)dx.$$

Заметим, что $\mu_0=1;\quad \mu_1=0$.

Коэффициент **асимметрии** (скошенность)

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \text{ или } \beta_1 = \frac{\mu_3}{\mu_2^{3/2}}.$$

Коэффициент **островершинности** (эксцесс)

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \text{ или } \beta_2 = \frac{\mu_4}{\mu_2^2} - 3.$$

2.7.4. Примеры непрерывных распределений

Нормальный закон

Нормальный закон распределения задается плотностью

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad a = M(X), \quad \sigma^2 = D(X).$$

Кривая распределения имеет симметричную колоколообразную форму и характеризуется показателями: $\beta_1=0$; $\beta_2=3$.

Вероятность попадания случайной величины X, распределенной по нормальному закону, на интервал $\alpha < x < \beta$ определяется по формуле

$$P(\alpha < x < \beta) = \Phi\left(\frac{\beta - a}{\sigma}\right) \Phi\left(\frac{\alpha - a}{\sigma}\right),$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} e^{t^2/2} dt$ - функция Лапласа. Здесь величина

$$t = \frac{x - a}{\sigma}$$

представляет собой выраженное в долях «сигма» отклонение случайной величины X от центра распределения a.

В зависимости от значения t вероятность попадания случайной величины X на заданный интервал $m_x \pm t\sigma$ равна:

При t = 1 P = 0,6827.

При t = 2 P = 0,9545.

При t = 3 P = 0,9973.

Таким образом, вероятность выхода значений случайной величины X за пределы 3σ очень мала и равна 1–0,9973=0,0027. Это значит, что из 1000 значений случайной величины X, распределенной по нормальному закону, в среднем только три могут выйти за границы трех стандартных отклонений (правило «трех сигма»). Это «правило» используется во многих практических расчетах, например, при статистическом анализе точности технологических процессов.

Показательный закон

Плотность вероятности и функция распределения задаются формулами

$$p(x) = \alpha e^{-\alpha x}; \; F(x) = 1 - e^{-\alpha x}.$$

Это – один из простейших однопараметрических законов распределения.

2.7.4.2. Закон Вейбулла

Плотность вероятности и функция распределения задаются формулами

$$p(x) = \alpha\beta x^{\beta-1} e^{-\alpha x^{\beta}}; \quad F(x) = 1 - e^{-\alpha x^{\beta}}.$$

Из закона Вейбулла при $\beta = 1$ следует показательный закон, а при $\beta = 2$ - распределение Релея.

Тема 3. Методы построения обобщенных непрерывных распределений

3.1. Построение системы непрерывных распределений методом обобщения

Памятка студенту:
Когда ты хочешь много знать,
Не тратя лишних сил и времени,
Старайся чаще применять
Волшебный метод обобщения!

Рассмотрим три простейших распределения: равномерное, треугольное убывающее и треугольное возрастающее [8.

В случае равномерной плотности функция распределения задается формулой

$$F(t)=\alpha t=1-(1-\alpha t). \qquad (3.1)$$

В случае треугольного убывающего распределения получим

$$F(t)=1-\left(1-\frac{\alpha}{2}t\right)^2 \qquad (3.2)$$

Для треугольного возрастающего распределения имеем

$$F(t)=\alpha t^2=1-(1-\alpha t^2). \qquad (3.3)$$

Обобщим попарно функции распределения (3.1), (3.2) и (3.1), (3.3) путем введения новых параметров.

В первом случае получим

$$F(t)=1-(1-\alpha ut)^{\frac{1}{u}}. \qquad (3.4)$$

Во втором случае

$$F(t)=1-(1-\alpha t^{\beta}). \qquad (3.5)$$

Теперь замечаем, что в формуле (3.1.4.) имеется параметр u, но его нет в формуле (3.1.5). Введем его в последнюю формулу. В результате получим

$$\mathrm{F(t)}=1-(1-\alpha \mathrm{u}t^{\beta})^{\frac{1}{u}} \qquad (3.6)$$

откуда дифференцированием по t найдем плотность распределения

$$\mathrm{p(t)}=\alpha\beta t^{\beta-1}(1-\alpha ut^{\beta})^{\frac{1}{u}-1} \qquad (3.7)$$

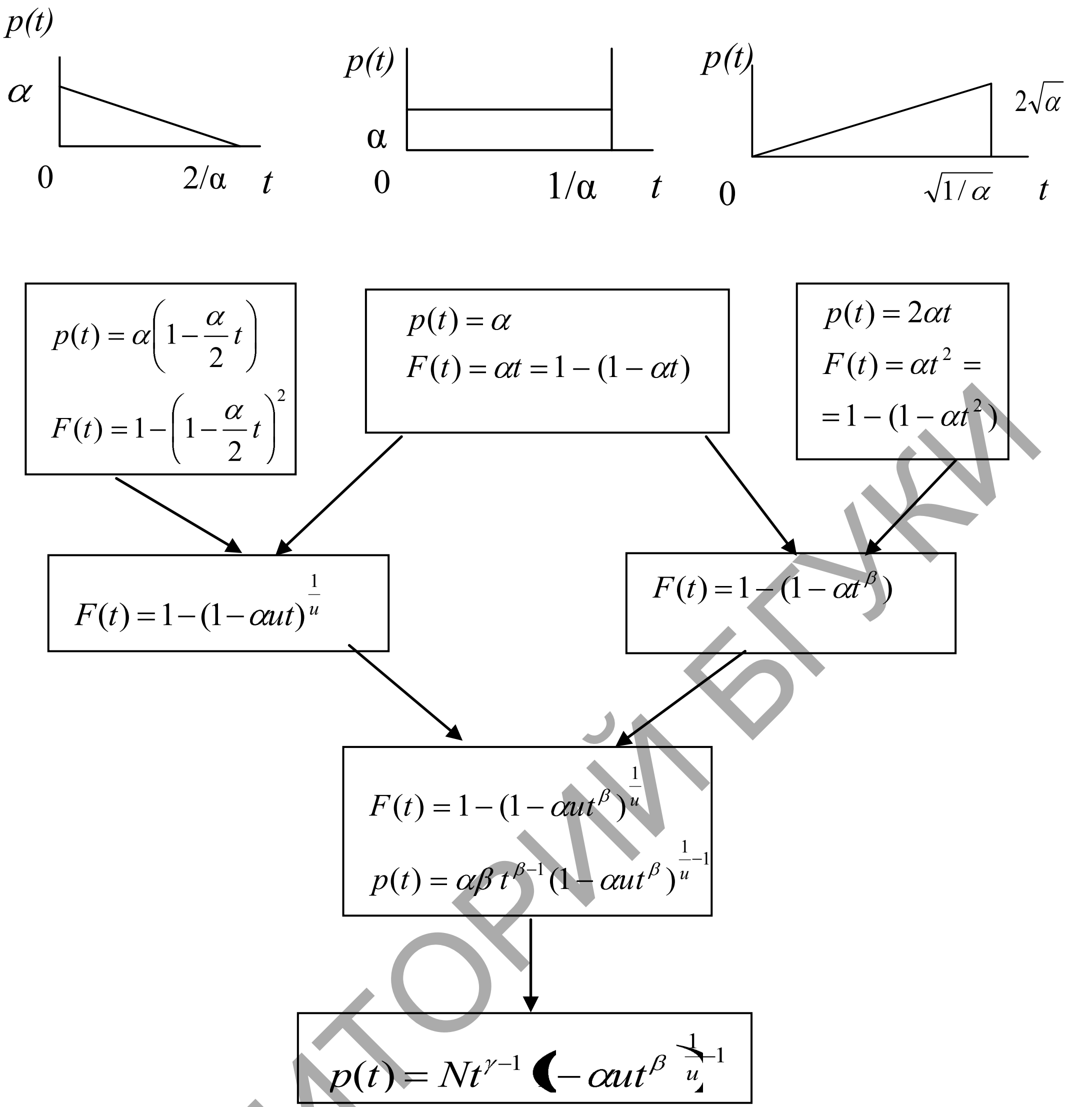


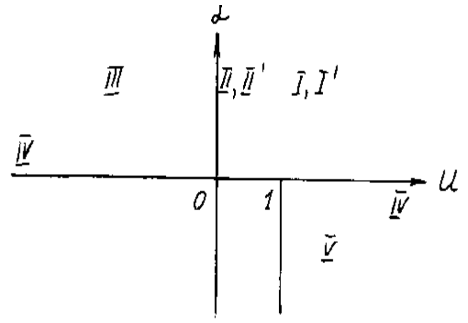
Рис. 3.1 Последовательность обобщения простейших непрерывных распределений.

Последняя плотность может быть еще более расширена за счет введения нового параметра формы. Параметр β в формуле (3.7) используется дважды в качестве показателя степени. Пусть это будут два разных параметра. Тогда вместо (3.7) можем записать [7]

$$p(t)=Nt^{\gamma-1}(1-\alpha ut^{\beta})^{\frac{1}{u}-1}. \qquad (3.8)$$

В итоге получена обобщенная плотность распределения с четырьмя параметрами α, β, γ, u. Нормирующий множитель N выражается через эти параметры из условия нормировки

$$\int_0^{\infty} p(t)dt=1.$$



ис. 3.2 Классификация распределений (типы со штрихом – при $\beta, \gamma < 0$).

3.2. Классификация обобщенных распределений

В зависимости от значений параметров α, u, а также от знака параметров β, γ распределения, заданные обобщенной плотностью (3.8), можно разделить на типы (см. рис. 3.2.).

В таблице 3.1. приведены значения параметров распределений разных типов.

Таблица 3.1 Классификация распределений

Тип кривой	Параметры кривой		
	u	α	$k=\gamma/\beta$
I, I′	$0<u<\infty$	$\alpha > 0$	$0<k<\infty$
II, II′	$u\to 0$		
III	$-\infty<u<\infty$		$0 < k < 1 - \dfrac{1}{u}$
IV	$u\to\pm\infty$	$\alpha u < 0$	
V	$1<u<\infty$	$\alpha < 0$	

Все распределения можно разбить на две большие группы: А и Б.

В группу А входят распределения с параметрами $\beta=\gamma$, или $\gamma/\beta=k=1$. Они задаются формулами (3.6) и (3.7).

В группу Б входят распределения, заданные обобщенной плотностью (3.8). В этом случае функция распределения, т.е. интеграл

$$F(t) = \int_0^t p(t)dt,$$

как правило, не выражается конечным числом элементарных функций.

Отметим, что из плотности (3.8) при $\beta = 2$, $\gamma = 1$ следует группа симметричных распределений.

Симметричны также распределения I типа с параметрами $\beta = 1$, $\gamma=1/u$.

Приведем все существующие типы распределений обеих групп (см. табл. 3.2– 3.4).

Таблица 3.2, Распределения группы А

Тип кривой	Функция распределения	Плотность распределения	Границы кривой
I	$F(t)=1-\left(1-\alpha ut^{\beta}\right)^{\frac{1}{u}}$	$p(t)=\alpha\beta t^{\beta-1}\left(1-\alpha ut^{\beta}\right)^{\frac{1}{u}-1}$	$0<t<(\frac{1}{\alpha u})^{\frac{1}{\beta}}$ $(u>0)$
II	$F(t)=1-e^{-\alpha t^{\beta}}$	$p(t)=\alpha\beta t^{\beta-1}e^{-\alpha t^{\beta}}$	$0<t<\infty$ $(u\to 0)$
III	$F(t)=1-\dfrac{1}{(1-\alpha ut^{\beta})^{-1/u}}$	$p(t)=\dfrac{\alpha\beta t^{\beta-1}}{\left(1-\alpha ut^{\beta}\right)^{1-\frac{1}{u}}}$	$0<t<\infty$ $(u<0)$
I′	$F(t)=\left(1-\dfrac{\alpha u}{t^{\beta}}\right)^{\frac{1}{u}}$	$p(t)=\dfrac{\alpha\beta}{t^{\beta+1}}\left(1-\dfrac{\alpha u}{t^{\beta}}\right)^{\frac{1}{u}-1}$	$(\alpha u)^{1/\beta}<t<\infty$ $(u>0)$
II′	$F(t)=e^{-\alpha/t^{\beta}}$	$p(t)=\dfrac{\alpha\beta}{t^{\beta+1}e^{\alpha/t^{\beta}}}$	$0<t<\infty$ $(u\to 0)$
III′	$F(t)=\dfrac{1}{\left(1-\dfrac{\alpha u}{t^{\beta}}\right)^{-1/u}}$	$p(t)=\dfrac{\alpha\beta}{t^{\beta+1}}\dfrac{1}{\left(1-\dfrac{\alpha u}{t^{\beta}}\right)^{1-\frac{1}{u}}}$	$0<t<\infty$ $(u<0)$

Таблица 3.3, Распределения группы Б

Тип кривой	Плотность распределения	Границы кривой
I	$p(t)=\dfrac{\beta(\alpha u)^{k}\,\Gamma\left(k+\frac{1}{u}\right)}{\Gamma(k)\,\Gamma\left(\frac{1}{u}\right)}t^{k\beta-1}\left(1-\alpha ut^{\beta}\right)^{\frac{1}{u}-1}$	$0<t<(\frac{1}{\alpha u})^{\frac{1}{\beta}}$ $(u>0)$
I′	$p(t)=\dfrac{\beta(\alpha u)^{k}\,\Gamma\left(k+\frac{1}{u}\right)}{\Gamma(k)\Gamma\left(\frac{1}{u}\right)}\dfrac{1}{t^{k\beta+1}}\left(1-\dfrac{\alpha u}{t^{\beta}}\right)^{\frac{1}{u}-1}$	$(\alpha u)^{1/\beta}<t<\infty$ $(u>0)$
II	$p(t)=\dfrac{\beta\alpha^{k}}{\Gamma(k)}t^{k\beta-1}e^{-\alpha t^{\beta}}$	$0<t<\infty$ $(u\to 0)$
II′	$p(t)=\dfrac{\beta\alpha^{k}}{\Gamma(k)}\dfrac{1}{t^{k\beta+1}e^{\alpha/t^{\beta}}}$	$0<t<\infty$ $(u\to 0)$

Тип кривой	Плотность распределения	Границы кривой
III-V	$p(t)=\dfrac{\beta(-\alpha u)^{k}\,\Gamma\left(1-\dfrac{1}{u}\right)}{\Gamma(k)\Gamma\left(1-\dfrac{1}{u}-k\right)}\dfrac{t^{k\beta-1}}{\left(1-\alpha u t^{\beta}\right)^{1-\frac{1}{u}}}$	$0<t<\infty$ $(\alpha u<0)$

Таблица 3.4, Группа симметричных распределений

Тип кривой	Плотность симметричного распределения	Границы кривой
Ic	$p(t)=\dfrac{\sqrt{\alpha u}\,\Gamma\left(\dfrac{1}{2}+\dfrac{1}{u}\right)}{\sqrt{\pi}\,\Gamma\left(\dfrac{1}{u}\right)}\left(1-\alpha u t^{2}\right)^{\frac{1}{u}-1}$	$-\sqrt{\dfrac{1}{\alpha u}}<t<\sqrt{\dfrac{1}{\alpha u}}$
IIc	$p(t)=\sqrt{\dfrac{\alpha}{\pi}}\;e^{-\alpha t^{2}}$	$-\infty<t<\infty$
IIIc-Vc	$p(t)=\dfrac{\sqrt{-\alpha u}\,\Gamma\left(1-\dfrac{1}{u}\right)}{\sqrt{\pi}\;\Gamma\left(\dfrac{1}{2}-\dfrac{1}{u}\right)}\dfrac{1}{\left(1-\alpha u t^{2}\right)^{1-\frac{1}{u}}}$	$-\infty<t<\infty$

3.3. Распределения функций случайного аргумента

Из обобщенной плотности (3.8) можно получить другие распределения как функции случайного аргумента.

Если две случайные величины X, T связаны между собой функциональной зависимостью X=f(T), причем с ростом X растет T, то вероятность P(X < x) = F(x) должна быть равна вероятности P(T < t) = F(t), т.е.

$$F(x) = F(t). \qquad (3.9)$$

Найдем зависимость между плотностями распределения p(x) и p(t).

По правилу дифференцирования сложной функции из (6.4.1) имеем

$$p(x)=\frac{dF(x)}{dx}=\frac{dF(t)}{dt}\frac{dt}{dx}=p(t)\frac{dt}{dx}. \qquad (3.10)$$

Воспользуемся последней формулой для нахождения других обобщенных плотностей.

Пусть между двумя случайными величинами T, X существует взаимосвязь T= eX. Тогда $dt/dx=e^{x}$ и, следовательно,

$$p(x)=p(t)\frac{dt}{dx}=N\,e^{\gamma x}\left(1-\alpha u e^{\beta x}\right)^{\frac{1}{u}-1}. \qquad (3.11)$$

Характерной особенностью этой обобщенной плотности является то, что кривые III-V типов при $k=\dfrac{\gamma}{\beta}=\dfrac{1}{2}\left(1-\dfrac{1}{u}\right)$ являются симметричными.

Если $T = \ln Y$, то таким же путем получим еще одну обобщенную плотность

$$p(y) = \frac{N}{y}(\ln y)^{\gamma-1}\left[1 - \alpha u(\ln y)^{\beta}\right]^{\frac{1}{u}-1}. \qquad (3.12)$$

Кривые распределения, заданные тремя обобщенными плотностями p(x), p(t), p(y), имеют разнообразную форму. Например, для кривой I типа,

заданной плотностью p(t), существуют формы начала и конца кривой, которые представлены ниже.

Рис.3.3 Формы начала кривой в зависимости от значений параметра γ=kβ.

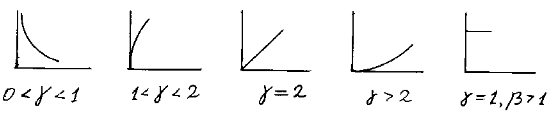
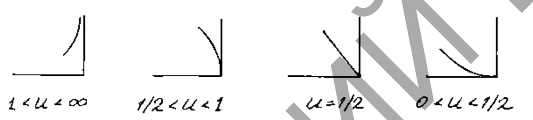


Рис. 3.4 Формы конца кривой в зависимости от значений параметра u.



3.4. Три основные и три дополнительные системы непрерывных распределений В. Нешитого

Полученные выше обобщенные плотности распределения

$$\left.\begin{aligned} p(x) &= Ne^{\gamma x}\left(1 - \alpha u e^{\beta x}\right)^{\frac{1}{u}-1} \\ p(t) &= Nt^{\gamma-1}\left(1 - \alpha u t^{\beta}\right)^{\frac{1}{u}-1} \\ p(y) &= \frac{N(\ln y)^{\gamma-1}}{y}\left[1 - \alpha u(\ln y)^{\beta}\right]^{\frac{1}{u}-1} \end{aligned}\right\} \qquad (3.13)$$

образуют три основные системы непрерывных распределений В. Нешитого.

Вторая основная система непрерывных распределений, заданная плотностью p(t), представлена в таблицах 3. 2 и 3.3.

Введем в плотность p(t) дополнительный параметр сдвига l и перепишем ее в виде

$$p(t) = N(t-l)^{\gamma-1}\left[1 - \alpha u(t-l)^{\beta}\right]^{\frac{1}{u}-1}. \qquad (3.14)$$

На основании плотности (3.4.2) можно получить три дополнительные системы непрерывных распределений.

Первая дополнительная система непрерывных распределений в общем случае задается формулой (3.4.2) при |β| = 1, а в случае симметричных распределений – при β = 2, γ = 1.

Ее легко получить из второй основной системы непрерывных распределений. Для этого достаточно в табл.3.2.3 принять |β| = 1, tβ заменить на $t-l$, а в табл. 3.2.3 заменить величину t2 на $(t-l)^2$.

Для обозначения типов кривых дополнительной системы непрерывных распределений будем использовать двузначный код, записанный арабскими цифрами через точку: 1.1, 1.1′, 2.1 и т.д., где первая цифра обозначает тип кривой, а вторая (единица) указывает на то, что параметр β=1; единица со штрихом соответствует параметру $\beta=-1$.

В большинстве случаев в тексте используется единое обозначение типов, но при необходимости указывается, что |β| = 1.

В таблице 3.4.1 приведены существующие типы первой дополнительной системы непрерывных распределений.

Из симметричных распределений приведен один нормальный закон.

Распределения типа 1.1 при k = 1/u также являются симметричными. Первая дополнительная система непрерывных распределений представляет собой основную часть семейства кривых К. Пирсона.

Таблица 3.5

Первая дополнительная система непрерывных распределений

Тип кривой	Плотность распределения (k=γ/β=γ)	Границы кривой
1.1	$p(t)=\dfrac{(\alpha u)^k\,\Gamma\left(k+\frac{1}{u}\right)}{\Gamma(k)\,\Gamma\left(\frac{1}{u}\right)}(t-l)^{k-1}\left[1-\alpha u(t-l)\right]^{\frac{1}{u}-1}$	$l<t<\dfrac{1}{\alpha u}+l$
1.1′	$p(t)=\dfrac{(\alpha u)^k\,\Gamma\left(k+\frac{1}{u}\right)}{\Gamma(k)\,\Gamma\left(\frac{1}{u}\right)}\dfrac{1}{(t-l)^{k+1}}\left(1-\dfrac{\alpha u}{t-l}\right)^{\frac{1}{u}-1}$	$t>\alpha u+l$
2.1	$p(t)=\dfrac{\alpha^k}{\Gamma(k)}\dfrac{(t-l)^{k-1}}{e^{\alpha(t-l)}}$	$t>l$
2.1′	$p(t)=\dfrac{\alpha^k}{\Gamma(k)}\dfrac{1}{(t-l)^{k+1}e^{\alpha/(t-l)}}$	$t>l$
3.1	$p(t)=\dfrac{(-\alpha u)^k\,\Gamma\left(1-\frac{1}{u}\right)}{\Gamma(k)\Gamma\left(1-\frac{1}{u}-k\right)}\dfrac{(t-l)^{k-1}}{\left[1-\alpha u(t-l)\right]^{1-\frac{1}{u}}}$	$t>l$

Тип кривой	Плотность распределения (k=γ/β=γ)	Границы кривой
IIc	$p(t)=\sqrt{\dfrac{\alpha}{\pi}}e^{-\alpha(t-l)^2}$	$-\infty < t < \infty$

Вторая дополнительная система непрерывных распределений получается из первой при t = ln y и прежних значениях параметров β, γ. При этом обобщенная плотность имеет вид

$$p(y)=\frac{N(\ln y-l)^{\gamma-1}}{y}\left[1-\alpha u(\ln y-l)^{\beta}\right]^{\frac{1}{u}-1}. \qquad (3.15)$$

Третья дополнительная система непрерывных распределений получается из второй при $y=\ln w$

$$p(w)=\frac{N(\ln\ln w-l)^{\gamma-1}}{w\ln w}\left[1-\alpha u(\ln\ln w-l)^{\beta}\right]^{\frac{1}{u}-1}. \qquad (3.16)$$

Тема 4. Классические методы оценивания параметров непрерывных распределений.

Методы оценивания параметров обобщенных непрерывных распределений

При исследовании случайных величин в математической статистике используется выборочный метод. Он заключается в том, что из генеральной совокупности отбирается выборка объёмом, как правило, не менее 100 единиц. При этом она должна правильно отражать пропорции генеральной совокупности, т.е. быть представительной (репрезентативной). Только в этом случае результаты исследования выборки могут быть распространены на всю генеральную совокупность.

Чтобы извлечь информацию из выборки, которая представляет собой простой статистический ряд, необходимо упорядочить все значения исследуемой случайной величины либо по возрастанию, либо по убыванию и построить интервальный ряд распределения или ранжированный ряд. Далее вычисляются числовые характеристики случайной величины и по ним – аппроксимирующий закон распределения и оценки его параметров. Закон распределения является наиболее полной характеристикой случайной величины.

4.1. Метод наименьших квадратов.

Этим методом могут быть найдены оценки параметров распределений группы *A*.

Рассмотрим распределения I – III типов группы *A*. Преобразуем функцию распределения

$$F(t) = 1 - \left(1 - \alpha u t^{\beta}\right)^{\frac{1}{u}}$$

к уравнению прямой

$$\ln \frac{1 - \left[1 - F(t)\right]^{u}}{u} = \ln \alpha + \beta \ln t \quad . \qquad (4.1)$$

Построив по эмпирической функции распределения график зависимости (4.1.1) (при известной оценке параметра *u*) и убедившись, что опытные точки рассеиваются вдоль прямой, по методу наименьших квадратов найдем оценки величин $\ln\alpha, \beta$. Введем обозначения:

$$Y = \ln \frac{1 - \left[1 - F(t)\right]^{u}}{u}, \qquad X = \ln t.$$

Тогда вместо формулы (4.1.1) запишем

$$Y = \ln \alpha + \beta X \ . \qquad (4.2)$$

Оценки параметров $ln\,\alpha, \beta$ (при заданном значении параметра *u*) по методу наименьших квадратов будут равны

$$\beta = \frac{n\Sigma XY - \Sigma X \Sigma Y}{n\Sigma X^{2} - (\Sigma X)^{2}}, \qquad (4.3)$$

$$\ln \alpha = \frac{1}{n}\left(\Sigma Y - \beta \Sigma X\right). \qquad (4.4)$$

Для оценки тесноты связи между переменными *Y, X* при различных значениях параметра *u* вычисляется выборочный коэффициент корреляции

$$r_{y/x} = \frac{n\Sigma XY - \Sigma X \Sigma Y}{\sqrt{n\Sigma X^{2} - (\Sigma X)^{2}}\sqrt{n\Sigma Y^{2} - (\Sigma Y)^{2}}}. \qquad (4.5)$$

В качестве оценки параметра *u* следует принять то его значение, при котором коэффициент корреляции по модулю ближе к единице.

Аналогично приводятся к уравнению прямой функции распределения остальных типов.

Тип II: $F(t) = 1 - e^{-\alpha t^{\beta}}$, $\ln\ln \dfrac{1}{1 - F(t)} = \ln\alpha + \beta \ln t$.

Вводя обозначения $Y = \ln\ln \dfrac{1}{1 - F(t)}$, $X = \ln t$, получим уравнение прямой (4.2).

Тип II′: $F(t) = e^{-\alpha / t^{\beta}}$, $\ln\ln \dfrac{1}{F(t)} = \ln\alpha - \beta \ln t$.

Типы I′, III′: $F(t) = \left(1-\frac{\alpha u}{t^{\beta}}\right)^{\frac{1}{u}}$, $\ln\frac{1-\left[F(t)\right]^{u}}{u} = \ln\alpha - \beta\ln t$.

Из рассмотренных примеров видно, что главная трудность здесь заключается в выборе подходящего значения параметра *u*. Его можно найти путем подбора и вычисления при каждом значении *u* коэффициента корреляции. Однако имеется возможность оценить его более простым и быстрым методом.

Если построить кривую распределения в форме $tp(t) = f(\ln t)$ и график функции распределения $F(t) = \varphi(\ln t)$, то мода $\ln t_c$, т.е. точка, в

которой произведение *tp(t)* максимально, равна

$$\ln t_c = \frac{1}{\beta}\ln\frac{1}{\alpha},$$

откуда $t_c = (1/\alpha)^{1/\beta}$. Подставив значение t_c в функцию распределения, получим [9]

$$F(t_c) = 1-(1-\alpha u t^{\beta})^{\frac{1}{u}} = 1-(1-u)^{\frac{1}{u}}. \qquad (4.6)$$

Последняя формула справедлива для распределений I-III типов группы *A*. Для распределений I′-III′ типов справедливо равенство

$$F(t_c) = (1-u)^{\frac{1}{u}}. \qquad (4.7)$$

В таблице 4.1.1 приведены значения *F(t_c)*, рассчитанные по формулам (4.1.6), (4.1.7).

Таблица 4.1

Значение функции распределения *F(t_c)*

Параметр *u*	$F(t_c)^{I\text{-}III}$	$F(t_c)^{I'\text{-}III'}$	**Тип кривой**
1	1	0	
0,9	0,9226	0,0774	
0,8	0,8663	0,1337	
0,7	0,8209	0,1791	
0,6	0,7828	0,2172	I, I′
0,5	0,7500	0,2500	
0,4	0,7211	0,2789	
0,3	0,6954	0,3046	
0,2	0,6723	0,3277	
0,1	0,6513	0,3487	
0	0,6321	0,3679	II, II′
-0,2	0,5981	0,4019	
-0,4	0,5688	0,4312	
-0,6	0,5431	0,4569	III-III′
-0,8	0,5204	0,4796	
-1,0	0,5000	0,5000	

-1,5	0,4571	0,5429	
-2	0,4226	0,5774	
-2,5	0,3941	0,6059	
-3	0,3700	0,6300	
-4	0,3313	0,6687	
-5	0,3012	0,6988	
-10	0,2132	0,7868	
-20	0,1412	0,8588	
-30	0,1082	0,8918	
-∞	0	1	

На основании полученных результатов можно рекомендовать следующий **порядок установления типа выравнивающего распределения группы *А*** и нахождения оценок параметров на примере плотности *p(t).*

1. Выбрать за начало отсчета значений случайной величины *T* начало кривой распределения.

2. Найти эмпирическую моду $\ln t_c^*$ кривой распределения $tp(t) = p(\ln t)$.

3. Найти эмпирическое значение функции распределения в точке *C* и приравнять теоретическому.

4. С помощью таблицы 4.1.1 определить два значения параметра *u* (в предположении, что выравнивающее распределение относится либо к I-III, либо к I′-III′ типам).

5. По двум значениям параметра *u* определить два типа возможных выравнивающих распределений.

6. Для обоих типов распределений путем построения графиков проверить, ложатся ли опытные точки на прямые.

7. В качестве выравнивающего принять наиболее подходящее распределение.

Таким же образом могут быть найдены оценки параметров распределений группы *A*, заданных плотностями *p(x), p(y).* При этом плотность *p(y)* должна быть приведена к форме $yp(y)\ln y = p(\ln\ln y)$.

4.2. Метод наибольшего правдоподобия

Покажем применение этого метода на примере распределений I типа группы *Б*

$$p(t) = \frac{\beta(\alpha u)^k\,\Gamma\!\left(k+\frac{1}{u}\right)}{\Gamma(k)\,\Gamma\!\left(\frac{1}{u}\right)}\, t^{k\beta-1}\left(-\alpha u t^{\beta}\right)^{\frac{1}{u}-1}, \quad k = \gamma/\beta .$$

Примем в качестве **логарифмической функции правдоподобия** величину $\ln L = M[\ln tp(t)]$ [11].

Вначале логарифмируем плотность *p(t)* (лучше – произведение *tp(t)):*

$$\ln tp(t) = \ln\beta + k\ln\alpha u + \ln\Gamma\left(k+\frac{1}{u}\right) - \ln\Gamma(k)$$
$$-\ln\Gamma\left(\frac{1}{u}\right) + k\beta\ln t + \left(\frac{1}{u}-1\right)\ln\left(1-\alpha u t^{\beta}\right).$$

Далее находим математическое ожидание величины $\ln tp(t)$

$$\ln L = M\left[\ln tp(t)\right] = \ln\beta + k\ln\alpha u + \ln\Gamma\left(k+\frac{1}{u}\right) - \ln\Gamma(k) -$$
$$-\ln\Gamma\left(\frac{1}{u}\right) + k\beta M(\ln t) + \left(\frac{1}{u}-1\right)M\left[\ln\left(1-\alpha u t^{\beta}\right)\right]. \tag{4.8}$$

Уравнения правдоподобия находятся из условий**:**

$$\frac{\partial\ln L}{\partial\alpha}=0;\quad \frac{\partial\ln L}{\partial\beta}=0;\quad \frac{\partial\ln L}{\partial k}=0;\quad \frac{\partial\ln L}{\partial u}=0\ .$$

Приняв обозначение $\frac{d}{dk}\ln\Gamma(k)=\Psi(k)$ для **логарифмической производной гамма-функции**, или иначе **пси-функции**, из (4.2.1) найдем

$$\left.\begin{aligned}
&\frac{k}{\alpha}-(1-u)M\left(\frac{t^{\beta}}{1-\alpha u t^{\beta}}\right)=0\\
&\frac{1}{\beta}+kM(\ln t)-\alpha(1-u)M\left(\frac{t^{\beta}\ln t}{1-\alpha u t^{\beta}}\right)=0\\
&\ln\alpha u+\Psi\left(k+\frac{1}{u}\right)-\Psi(k)+\beta M\left(\ln t\right)=0\\
&\Psi\left(\frac{1}{u}\right)-\Psi\left(k+\frac{1}{u}\right)-M\left[\ln\left(1-\alpha u t^{\beta}\right)\right]=0
\end{aligned}\right\} \tag{4.9}$$

Здесь последнее уравнение приведено к более простой форме с учетом первого уравнения.

Оценки параметров могут быть найдены путем решения системы четырех уравнений правдоподобия – (4.2.2). При этом соответствующие математические ожидания заменяются их оценками, которые вычисляются по статистическому распределению. Однако для нахождения оценок таких величин, как $M\left[t^{\beta}/(1-\alpha u t^{\beta})\right]$ и др. необходимо знать значения параметра β и произведения α*u*, оценки которых предстоит найти. Кроме того, предварительно **необходимо знать тип выравнивающего распределения, а метод наибольшего правдоподобия не предлагает критериев для его установления**.

Эти обстоятельства сильно ограничивают возможности использования метода наибольшего правдоподобия для нахождения оценок параметров обобщенных выравнивающих распределений.

4.3. Классический метод моментов

Метод пригоден для оценивания параметров обобщенных распределений **с параметром** $|\beta|=1$**,** т.е. в случае трех дополнительных систем непрерывных распределений, заданных плотностями (6.5.2) – (6.5.4).

При этом плотности (6.5.3), (6.5.4) должны быть приведены к форме плотности (6.5.2), т.е. представлены в виде

$$yp(y) = p(\ln y),\ \mathrm{w} \ln w\, p(w) = p(\ln\ln w).$$

4.3.1. Распределения I-III типов при β=1.

Рассмотрим обобщенную плотность (3.14) при $|\beta| = 1$, которую запишем в виде [9]

$$p(t) = N(t-a)^{\gamma-1}\left[1-\alpha u(t-a)^{\beta}\right]^{-\frac{1}{u}-1}. \qquad (4.10)$$

Выразим параметры распределения (4.10) через центральные моменты. Для этого представим его в дифференциальной форме (при $|\beta| = 1$)

$$\frac{1}{p(t)}\frac{dp(t)}{dt} = \frac{\alpha\left(1+\gamma u-2u\right)t+1+2a\alpha u-\left(\gamma+a\alpha\gamma u+a\alpha\right)}{\alpha u t^2-\left(1+2a\alpha u\right)t+a(1+a\alpha u)}. \qquad (4.11)$$

Перепишем далее уравнение (4.3.2) в виде

$$\left[\alpha u t^2-\left(1+2a\alpha u\right)t+a(1+a\alpha u)\right]dp(t) =$$
$$= \left[\alpha\left(1+\gamma u-2u\right)t+1+2a\alpha u-\left(\gamma+a\alpha\gamma u+a\alpha\right)\right]p(t)dt.$$

Умножим обе части последнего равенства на t^r и проинтегрируем на бесконечном интервале (левую часть интегрируем по частям). В результате получим

$$t^r\left[\alpha u t^2-\left(1+2a\alpha u\right)t+a\left(1+a\alpha u\right)\right]p(t)\Big|_{-\infty}^{\infty}-\int_{-\infty}^{\infty}\left[(r+2)\alpha u t^{r+1}-\right.$$
$$\left.-(r+1)(1+2a\alpha u)t^r+ra(1+a\alpha u)t^{r-1}\right]p(t)dt =$$
$$= \int_{-\infty}^{\infty}\left\{\alpha(1+\gamma u-2u)t^{r+1}+\left[1+2a\alpha u-\left(\gamma+a\alpha\gamma u+a\alpha\right)\right]t^r\right\}p(t)dt.$$

Здесь первое слагаемое обращается в нуль на концах распределения, поскольку значения плотности $p(t) \to 0$.

Если начало координат перенесено в центр распределения ν_1, то переменная t обозначает отклонение случайной величины от ее среднего значения и поэтому интегралы вида

$$\int_{-\infty}^{\infty} t^r p(t)dt,$$

входящие в последнее уравнение, представляют собой центральные моменты μ_r распределения (7.3.1) при $\beta = 1$.

Следовательно, последнее уравнение можно представить в виде

$$ra(1+a\alpha u)\mu_{r-1}-(2ra\alpha u+\gamma+a\alpha\gamma u+a\alpha)\mu_r+\alpha(1+\gamma u+ru)\mu_{r+1} = r\mu_r. \qquad (4.12)$$

Учитывая, что $\mu_0 = 1, \mu_1 = 0$, из (4.3.3) при r = 0, 1, 2, 3 найдем

$$\left.\begin{array}{llll}
0 & -(\gamma+a\alpha\gamma u+a\alpha) & +0 & =0 \\
\alpha(1+a\alpha u) & -0 & +\alpha(1+\gamma u+u)\mu_2 & =0 \\
0 & -4a\alpha u\mu_2 & +\alpha(1+\gamma u+2u)\mu_3 & =2\mu_2 \\
\alpha(1+a\alpha u)\mu_2 & -6a\alpha u\mu_3 & +\alpha(1+\gamma u+3u)\mu_4 & =3\mu_3
\end{array}\right\} \qquad (4.13)$$

Из первого уравнения системы уравнений (4.13) получим

$$a = -\frac{\gamma}{\alpha(1+\gamma u)} = -\nu_1, \tag{4.14}$$

т.е. **параметр *a* по абсолютной величине равен математическому ожиданию случайной величины *T*.**

Решая далее систему уравнений (4.3.4), найдем **значения параметров распределений I-III типов**

$$\left.\begin{aligned} &Aa^2 + Ba + C = 0;\ a_{12} = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A};\\ &u = -\frac{AD}{6CE};\ \alpha = -\frac{6aCE}{D^2};\quad \gamma = \frac{6a^2E}{D};\ l = \bar{t} + a \end{aligned}\right\} \tag{4.15}$$

где l – параметр сдвига (см. табл. 6.5.1);

$$\left.\begin{aligned} &A = 2\mu_2\mu_4 - 6\mu_2^3 - 3\mu_3^2\\ &B = \mu_3(3\mu_2^2 + \mu_4)\\ &C = \mu_2(4\mu_2\mu_4 - 3\mu_3^2)\\ &D = C - a^2A = 2c + aB = -a(2aA + B)\\ &E = \mu_2\mu_4 - \mu_2^3 - \mu_3^2 \end{aligned}\right\} \tag{4.16}$$

Если разделить величины *A,...,* *E* на μ_2^3 и принять обозначения $\beta_1 = \mu_3^2/\mu_2^3, \beta_2 = \mu_4/\mu_2^2$, **введенные К. Пирсоном для показателей асимметрии и островершинности**, то получим:

$$\left.\begin{aligned} &A^* = 2\beta_2 - 3\beta_1 - 6\\ &B^* = \frac{\mu_3}{\mu_2}(3 + \beta_2)\\ &C^* = \mu_2(4\beta_2 - 3\beta_1)\\ &D^* = C^* - a^2A^* = 2c^* + aB^* = -a(2aA^* + B^*)\\ &E^* = \beta_2 - \beta_1 - 1 \end{aligned}\right\} \tag{4.17}$$

Величины *A*,...,E** могут использоваться в формулах (4.3.6) вместо величин *A,...,E.*

Выразим величины β_1 и β_2 через параметры распределения (4.3.1) при $|\beta| = 1$. Уравнение (4.3.3) с учетом (4.3.5) позволяет записать **рекуррентную формулу для центральных моментов распределений I-III типов**

$$\mu_{r+1} = r\frac{\gamma\mu_{r-1} + \alpha(1-\gamma u)(1+\gamma u)\mu_r}{[\alpha(1+\gamma u)]^2(1+\gamma u + ru)}. \tag{4.18}$$

Из (4.3.9) при $r = 1,2,3$ имеем [9]

$$\left.\begin{aligned} \mu_2 &= \frac{\gamma}{\left[\alpha(1+\gamma u)\right]^2 (1+\gamma u+u)} \\ \mu_3 &= \frac{2(1-\gamma u)\mu_2}{\alpha(1+\gamma u)(1+\gamma u+2u)} \\ \mu_4 &= 3\frac{\gamma\mu_2+\alpha(1-\gamma u)(1+\gamma u)\mu_3}{\left[\alpha(1+\gamma u)\right]^2 (1+\gamma u+3u)} \end{aligned}\right\} \quad (4.19)$$

Выразим с помощью формул (7.3.10) показатели β_1 и β_2 через параметры формы *k, u* распределений I-III типов:

$$\left.\begin{aligned} \beta_1 &= \frac{4(1-\gamma u)^2(1+\gamma u+u)}{\gamma(1+\gamma u+2u)^2} \\ \beta_2 &= 3\frac{1+\gamma u+u}{1+\gamma u+3u}\left[1+\frac{2(1-\gamma u)^2}{\gamma(1+\gamma u+2u)}\right] \end{aligned}\right\} \quad (4.20)$$

Обозначим первый сомножитель в формуле для β_2 через L и назовем его "**критерием *L***" [9]:

$$L = 3\frac{1+\gamma u+u}{1+\gamma u+3u}. \quad (4.21)$$

Величину L можно выразить через показатели β_1, β_2. Используя формулы (7.3.6), (7.3.8), из (7.3.12) найдем

$$L = \frac{4\beta_2-3\beta_1}{4+\beta_1}. \quad (4.22)$$

Из (4.3.13) следует, что при $\beta_1 = 0$ справедливо равенство $L=\beta_2$.

Таким образом, в **случае симметричных распределений критерий *L* есть не что иное, как показатель островершинности**.

Из (4.20) следует, что **критерий *L* в случае распределений I типа** задан **на интервале 1<*L*<3; для распределений II типа *L*=3, а для распределений III типа *L*>3**.

Поскольку показатель асимметрии β_1=0 при $\gamma u = 1$, что видно из (4.27), то **распределения I типа при условии** $\gamma = 1/u$ **являются симметричными**. Для них критерий L (обозначим его L_c) равен

$$L_c = 3\frac{2+u}{2+3u}. \quad (4.23)$$

Из формул (4.18) и (4.3.20) следует, что центральный момент 4-го порядка и критерий L существуют при условии $(\gamma+3)\ u > -1$. А это значит, что по **классическому методу моментов может быть найдена лишь незначительная часть выравнивающих распределений III типа**, для которых выполняется неравенство

$$u > -\frac{1}{\gamma+3}. \quad (4.24)$$

Например, при γ=5 параметр $u > -0{,}125$. Все остальные распределения III типа, а также распределения IV и V типов остаются за пределами применимости классического метода моментов.

Рассмотрим далее распределения II типа.

Тип II. Критерии: $u \to 0$; $L = 3$.

Кривая распределения задается формулой

$$p(t) = \frac{\alpha^{\gamma}}{\Gamma(\gamma)}(t-a)^{\gamma-1}e^{-\alpha(t-a)}, \quad a < t < \infty. \qquad (4.25)$$

Покажем, что для распределений II типа величина

$$A^* = 2\beta_2 - 3\beta_1 - 6 = 0.$$

При $u \to 0$ формулы (4.3.11) дают

$$\beta_1 = \frac{4}{\gamma};\ \beta_2 = 3\left(1+\frac{2}{\gamma}\right) = 3 + 1{,}5\,\beta_1. \qquad (4.26)$$

В общем случае зависимость между показателями β_1 и β_2 выражается формулой

$$\beta_2 = 3\frac{1+\gamma u+u}{1+\gamma u+3u}\left[1+\beta_1\frac{1+\gamma u+u}{2(1+\gamma u+2u)}\right],$$

которая следует из (4.3.11).

Подставляя значения β_1 и β_2 из (4.3.17) в формулу для A^*, получим

$$A^* = 0.$$

Далее, поскольку $D^* = C^* - a^2A^*$, то при $A^* = 0$ имеем равенство $D^* = C^*$ и, следовательно, $D = C$.

С учетом полученных результатов параметры распределения (4.16) равны

$$a = -\frac{C}{B};\ \gamma = \frac{6CE}{B^2};\ \alpha = \frac{6E}{B}. \qquad (4.27)$$

Их можно также выразить через центральные моменты не выше 3-го порядка [9]:

$$a = -\frac{2\mu_2^2}{\mu_3};\ \alpha = \frac{2\mu_2}{\mu_3};\ \gamma = 4\frac{\mu_2^3}{\mu_3^2}. \qquad (4.28)$$

В заключение отметим, что **распределения II типа при $\gamma \to \infty$ приближаются к нормальному закону**, так как при этом условии формулы (4.3.17) дают: $\beta_1 \to 0$, $\beta_2 \to 3$.

Далее, поскольку **для распределений II типа** величина A=0, то **квадратное уравнение $Aa^2+Ba+C=0$ имеет один отрицательный корень.**

В случае распределений I типа квадратное уравнение имеет два корня –один положительный, другой отрицательный**. Для последующих расчетов необходимо принимать тот корень, знак которого противоположен знаку среднего выборочного**.

В случае распределений III типа квадратное уравнение имеет два отрицательных корня. **Меньший из них** (по абсолютной величине) **соответствует распределениям III типа**.

4.3.2. Распределения I′, II′ типов при β=1.

Распределения I′, II′ типов заданы плотностью

$$p(t)=\frac{N}{(t-a)^{\gamma+1}}\left(1-\frac{\alpha u}{t-a}\right)^{\frac{1}{u}-1}, \qquad (4.29)$$

которая получается из (7.3.1) при $\beta=-1,\ \gamma<0$.

Выразим параметры распределения (4.3.20) через их центральные моменты.

Представим плотность (4.28) в дифференциальной форме

$$\frac{1}{p(t)}\frac{dp(t)}{dt}=-\frac{(\gamma+1)t-(\alpha+\alpha\gamma u+a+a\gamma)}{t^2-(2a+\alpha u)t+a(a+\alpha u)}. \qquad (4.30)$$

Далее, поступая как и в предыдущем случае (см. п.7.3.1), придем к следующему уравнению

$$-ra(a+\alpha u)\mu_{r-1}+\left[r(2a+\alpha u)+(1-\gamma)(a+\alpha u)-\alpha\right]\mu_r+\gamma\mu_{r+1}=(r+1)\mu_{r+1}. \qquad (4.31)$$

Полученное уравнение позволяет выразить в явном виде параметры распределения (4.3.20) через его центральные моменты не выше 4-го порядка. Придавая величине r значения $0,1,2,3$ и учитывая, что $\mu_0=1,\ \mu_1=0$, из (4.3.22) получим систему четырех уравнений с четырьмя неизвестными a, α, γ, u [9]:

$$\left.\begin{array}{llll} 0 & +(1-\gamma)(a+\alpha u)-\alpha & +0 & =0 \\ -a(a+\alpha u) & +0 & +\gamma\mu_2 & =2\mu_2 \\ 0 & +2(2a+\alpha u)\mu_2 & +\gamma\mu_3 & =3\mu_3 \\ -3a(a+\alpha u)\mu_2 & +3(2a+\alpha u)\mu_3 & +\gamma\mu_4 & =4\mu_4 \end{array}\right\} \qquad (4.32)$$

Из первого уравнения системы (4.3.23) находим

$$a=-\frac{\alpha(1+\gamma u-u)}{\gamma-1}=-\nu_1. \qquad (4.33)$$

Далее имеем [7]

$$\left.\begin{array}{l} Aa^2+Ba+C=0, \quad l=\bar{t}+a, \\ u=-\dfrac{AD}{6CE};\ \alpha=-\dfrac{6CE}{aA^2};\ \gamma=\dfrac{6E}{A}+1 \end{array}\right\} \qquad (4.34)$$

где величины $A,\dots,E$ задаются формулами (4.7) или (4.8).

Выразим далее центральные моменты распределений I′, II′ типов, заданных плотностью (4.20), через параметры α, γ, u.

Из (4.22) с учетом (4.24) найдем

$$\mu_{r+1}=\frac{r\alpha}{(\gamma-1)(\gamma-1-r)}\left[\frac{\alpha(1+\gamma u-u)}{\gamma-1}\mu_{r-1}+(1+\gamma u-u)\mu_r\right], \qquad (4.35)$$

откуда при r = 1, 2, 3 имеем

$$\left.\begin{aligned}
\mu_2 &= \frac{\alpha^2(1+\gamma u-u)}{(\gamma-1)^2(\gamma-2)} \\
\mu_3 &= \frac{2\alpha(2+\gamma u-u)\mu_2}{(\gamma-1)(\gamma-3)} \\
\mu_4 &= \frac{3\alpha}{(\gamma-1)(\gamma-4)}\left[\frac{\alpha(1+\gamma u-u)}{\gamma-1}\mu_2+(3+\gamma u-u)\mu_3\right]
\end{aligned}\right\} \qquad (4.36)$$

Показатели асимметрии и островершинности на основании (4.27) равны

$$\left.\begin{aligned}
\beta_1 &= \frac{4(\gamma-2)(2+\gamma u-u)^2}{(\gamma-3)^2(1+\gamma u-u)}, \\
\beta_2 &= \frac{3(\gamma-2)}{\gamma-4}\left[1+\frac{2(2+\gamma u-u)^2}{(\gamma-3)(1+\gamma u-u)}\right]
\end{aligned}\right\} \qquad (4.37)$$

Здесь также величина L входит в качестве первого сомножителя в формулу для β_2, т .е.

$$L = 3\frac{\gamma-2}{\gamma-4}, \qquad (4.38)$$

при этом остается справедливой также формула (4.13).

Распределения I′, II′ типов имеют моменты четвертого порядка при $\gamma > 4$. В этом случае критерий $L>3$. Как было показано ранее, для распределений III типа также $L>3$. Но **эти три типа распределений различаются с помощью параметра (критерия) *u*.**

Исследования показали, что при $|\beta|=1$ **кривая распределения I′ типа представляет собой соответствующую кривую III типа, но смещенную вдоль оси абсцисс на величину** $-1/\alpha u > 0$.

Между параметрами распределений I′ и III типов имеются соотношения

$$(\alpha u)' = -\frac{1}{\alpha u};\; u' = \frac{1}{\gamma};\; \gamma' = 1-\frac{1}{u}-\gamma, \qquad (4.39)$$

где штрихом отмечены параметры распределений I′ типа.

Формулы (4.30) позволяют осуществлять переход от распределения III типа к равносильному распределению I′ типа. Обратный переход осуществляется по формулам

$$\gamma = \frac{1}{u'};\; \alpha u = -\frac{1}{(\alpha u)'};\; \frac{1}{\mathrm{u}} = 1-\frac{1}{u'}-\gamma', \qquad (4.40)$$

которые следуют из (4.30).

Для обоих типов кривых корни квадратного уравнения $Aa^2+Ba+C=0$ отрицательны, причем **меньшему по абсолютной величине корню соответствует III тип, а большему – I′ тип распределения. При равных корнях имеем распределение II′ типа.**

Рассмотрим распределения II′ типа.

Тип II′ Критерии: $u\to 0$, $L>3$. Плотность распределения задается формулой

$$p(t)=\frac{\alpha^{\gamma}}{\Gamma(\gamma)}\frac{1}{(t-a)^{\gamma+1}e^{\alpha/(t-a)}},\quad a<t<\infty \tag{4.41}$$

Покажем, что **для распределений II′ типа величина D=0 и дискриминант B^2–4AC = 0**.

Решая совместно три последних уравнения системы (4.3.23) с тремя неизвестными $a(a+\alpha u)$, $2a+\alpha u$, γ и принимая во внимание обозначения (4.3.7), получим $a(a+\alpha u)=\dfrac{C}{A}$, откуда $u=\dfrac{C-a^2A}{\alpha aA}=\dfrac{D}{\alpha aA}$.

Для распределений II′ типа параметр $u\to 0$, следовательно,

$$D=C-a^2A=2C+aB=-a(2aA+B)=0. \tag{4.42}$$

На основании (7.3.33) можем записать

$$a=-\frac{2C}{B}=-\frac{B}{2A}, \tag{4.43}$$

откуда имеем $B^2=4AC$, т.е. дискриминант $B^2-4AC=0$, ч.т.д.

При D=0 параметр α распределений II′ типа на основании (4.25) и (4.33) будет равен

$$\alpha=\frac{3BE}{A^2}. \tag{4.44}$$

Параметры распределений II′ типа можно выразить через моменты не выше 3-го порядка. Из первых трех уравнений системы (4.3.23) при $u\to 0$ имеем:

$$\left.\begin{aligned}&a^2\mu_3+4a\mu_2^2-\mu_2\mu_3=0\\&\gamma=3-4a\frac{\mu_2}{\mu_3}\\&\alpha=-a(\gamma-1)\end{aligned}\right\} \tag{4.45}$$

В заключение отметим, что **распределения II′ типа при $\gamma\to\infty$ приближаются к нормальному**.

При $u\to 0$ формулы (4.28) дают

$$\left.\begin{aligned}&\beta_1=\frac{16(\gamma-2)}{(\gamma-3)^2}\\&\beta_2=\frac{3(\gamma-2)(\gamma+5)}{(\gamma-3)(\gamma-4)}\end{aligned}\right\} \tag{4.46}$$

откуда следует, что при $\gamma\to\infty$ $\beta_1=0$, $\beta_2=3$ как в случае нормального распределения.

Из формул (4.14) можно установить зависимость между показателями β_1 и β_2

$$\beta_2 = 3\frac{\gamma-2}{\gamma-4}\left(1+\beta_1\frac{\gamma-3}{2(\gamma-2)}\right),$$

откуда следует, что при $k\to\infty$ $\beta_2 = 3+1{,}5\beta_2$, т.е. как и в случае распределений II типа.

Покажем далее, что **семейство кривых К. Пирсона является частным случаем системы непрерывных распределений автора**, которая задана обобщенной плотностью (7.3.1) при $|\beta|=1$.

Действительно, если в дифференциальное уравнение (4.2) вместо параметров α, γ, *u* подставить их значения из (4.3.6), то после преобразований будем иметь [9]

$$\frac{1}{p(t)}\frac{dp(t)}{dt} = -\frac{2(A+3E)t+B}{At^2+Bt+C}. \qquad (4.47)$$

Разделив числитель и знаменатель правой части последнего уравнения на величину $2(A+3E)$, получим дифференциальное уравнение К. Пирсона

$$\frac{1}{p(t)}\frac{dp(t)}{dt} = -\frac{t-d}{at^2+bt+c}, \qquad (4.48)$$

множество решений которого составляет семейство кривых распределения К.Пирсона. При этом классификация распределений осуществляется в зависимости от значений дискриминанта b^2-4ac, откуда следует критерий К.Пирсона $k=b^2/4ac$.

В семействе кривых К. Пирсона область выше распределений II′ типа занимают распределения IV типа по К. Пирсону, для которых $0<k<1$, а дискриминант $b^2-4ac<0$.

Таким образом, семейство кривых К. Пирсона является частным случаем системы непрерывных распределений автора, заданной обобщенной плотностью (4.3.1) при $|\beta|=1$.

При $|\beta|\neq 1$ уравнение (4.3.1) можно представить в виде

$$\frac{1}{p(t)}\frac{dp(t)}{dt} = \frac{a_0+a_1t+\ldots+a_mt^m+\ldots}{b_0+b_1t+b_2t^2+\ldots+b_{m+1}t^{m+1}+\ldots}, \qquad (4.49)$$

где коэффициенты a_m, b_m являются функциями параметров α, β, γ, *u*. При целых β величина $m=\beta$; при нецелых β величина $m=\infty$.

В заключение отметим, что после нахождения оценок всех параметров распределения (4.3.1) при $|\beta|=1$ выравнивающие распределения I-III типов должны быть записаны в виде

$$p(t) = N(t-l)^{\gamma-1}\left[-\alpha u(t-l)\right]^{\frac{1}{u}-1}, \qquad (4.50)$$

где l - параметр сдвига. Он равен величине смещения начала выравнивающей кривой распределения относительно начала координат. При этом

$$l = \nu_1^* + a.$$

Распределение I′ типа запишется в виде

$$p(t)=\frac{N}{(t-l)^{\gamma+1}}\left(1-\frac{\alpha u}{t-l}\right)^{\frac{1}{u}-1}. \qquad (4.51)$$

Поскольку здесь $|\beta|=1$, то отношение $k=\gamma/\beta=\gamma$ и, следовательно, в формулах (4.3.41), (4.3.42) параметр $\gamma = k$.

4.3.3. Симметричные распределения Ic-IIIc типов

Рассмотрим симметричные распределения Ic-IIIc типов, заданные плотностью

$$\mathrm{p(t)=N}\left(1-\alpha \mathrm{u}t^2\right)^{\frac{1}{u}-1} \qquad (4.52)$$

или в дифференциальной форме

$$\frac{dp(t)}{dt}=-p(t)\frac{2\alpha(1-u)t}{1-\alpha ut^2}.$$

Запишем последнее уравнение в виде

$$\left(1-\alpha ut^2\right)dp(t)=-2\alpha(1-u)\,tp(t)dt.$$

Умножим обе части полученного равенства на t^r и проинтегрируем на бесконечном интервале. В результате получим

$$-r\mu_{r-1}+\alpha(2+ru)\mu_{r+1}=0. \qquad (4.53)$$

При r=1 и r=3 из (7.3.44) найдем

$$\mu_2=\frac{1}{\alpha(2+u)}, \qquad (4.54)$$

$$\mu_4=\frac{3\mu_2}{\alpha(2+3u)}. \qquad (4.55)$$

Тогда показатель островершинности β_2 будет равен

$$\beta_2=\frac{\mu_4}{\mu_2^2}=3\frac{2+u}{2+3u}=L_c. \qquad (4.56)$$

Последняя формула совпадает с формулой (7.3.14). **Величина β_2 в зависимости от типа распределения принимает значения (при u>–2/3, или $\mu_4<\infty$): $1<\beta_2<3$ - для Ic типа; β_2=3 – для IIc типа (нормального закона); β_2>3 – для IIIc типа.**

Отсюда следует, что **показатели β_1,β_2 могут служить критериями для различения распределений Ic-IIIc типов**.

Выразим параметры симметричных распределений Ic-IIIc типов через их центральные моменты.

В случае нормального закона (тип IIc) оценка параметра α равна

$$\alpha=\frac{1}{2\mu_2}. \qquad (4.57)$$

Оценки параметров α, u распределений Ic, IIIc типов равны

$$\alpha = \frac{3}{4}\frac{\mu_4 - \mu_2^2}{\mu_2\mu_4} = \frac{3}{4}\frac{\beta_2 - 1}{\beta_2\mu_2}, \qquad (4.58)$$

$$u = -\frac{2}{3}\frac{\mu_4 - 3\mu_2^2}{\mu_4 - \mu_2^2} = -\frac{2}{3}\frac{\beta_2 - 3}{\beta_2 - 1}, \qquad (4.59)$$

при этом остается также справедливой общая формула

$$u = -\frac{AD}{6CE}, \qquad (4.60)$$

полученная ранее для распределений I-III, I′, II′ типов при $|\beta| = 1$.

Действительно, поскольку для симметричных распределений показатель $\beta_1 = 0$, то на основании (4.3.8) имеем:

$$A = 2\beta_2 = 6; B = 0; C = 4\beta_2\mu_2; D = 2c; E = \beta_2 - 1.$$

Тогда формула (4.3.51) в этом частном случае примет вид

$$u = -\frac{AD}{6CE} = -\frac{A\cdot 2C}{6\cdot CE} = -\frac{A}{3E} = -\frac{2}{3}\frac{\beta_2 - 3}{\beta_2 - 1},$$

что совпадает с (4.50).

Таким образом, **показатели *L, u* могут служить критериями для классификации как симметричных распределений с параметрами $\boldsymbol{\beta = 2, \gamma = 1}$, так и других распределений I-III, I′, II′ типов с параметром $\boldsymbol{|\beta| = 1}$.**

4.3. Критерии для классификации кривых по методу моментов

Все распределения, рассмотренные в п.4.3.1 – 4.3.3, **могут быть разделены на типы с помощью критериев *L, u***, введенных автором.

Критерий *L* выражается через показатели асимметрии $\boldsymbol{\beta_1}$ и островершинности $\boldsymbol{\beta_2}$

$$L = \frac{4\beta_2 - 3\beta_1}{4 + \beta_1}.$$

При $\beta_1 = 0$ (в случае симметричных распределений) показатель $L = \beta_2$, т.е. он совпадает с показателем островершинности.

Критерий (он же параметр) *u* задается формулой

$$u = -\frac{AD}{6CE} = -\frac{A^* D^*}{6C^* E^*},$$

где величины $A, ..., E$ или $A^*, ...E^*$ вычисляются по формулам (4.7) или (4.8).

На рис. 4.3.1 дана классификация распределений с параметром $|\beta| = 1$, а также симметричных распределений, для которых критерий L задается формулой $L = 3\dfrac{2+u}{2+3u}$.

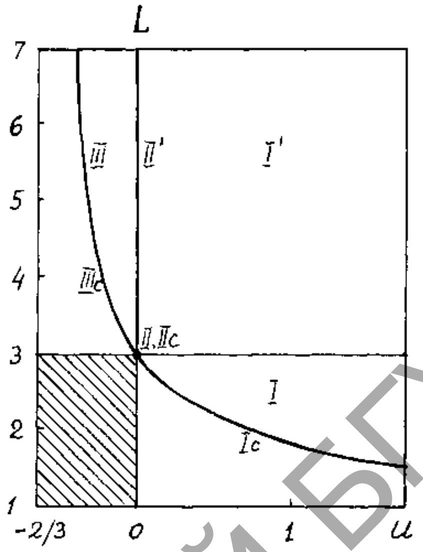


Рис. 4.3.1. Классификация распределений по критериям *u, L*.

Эти же распределения можно классифицировать по критериям β_1, β_2 (рис.4.3.2). В этом случае распределения II типа представлены прямой $\beta_2 = 3 + 1{,}5\beta_1$. Распределения II′ типа – кривой

$$\beta_2 = \frac{3}{32-\beta_1}\left[16 + 13\beta_1 + 2\sqrt{64 + \beta_1(48 + 12\beta_1 + \beta_1^2)}\right]. \qquad (4.61)$$

Последняя формула является следствием равенства $B^2 – 4AC = 0$, справедливого для кривых II′ типа.

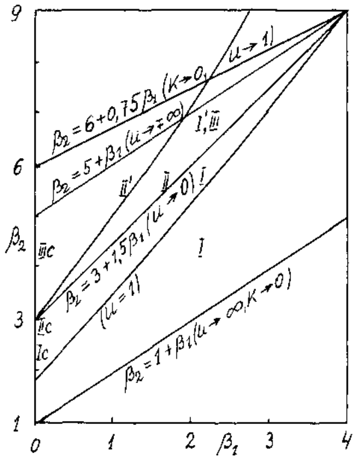


Рис. 4.3.2. Классификация распределений по критериям β_1, β_2.

Распределения III и I′ типов занимают одну и ту же область между распределениями II и II′ типов, поскольку кривая I′ типа при $\beta=1$ представляет собой соответствующую кривую III типа, но смещенную вдоль оси абсцисс на величину $-1/\alpha u>0$.

Распределения I типа занимают область между двумя прямыми - $\beta_2 = 1+\beta_1$ и $\beta_2 = 3+1{,}5\beta_1$, при этом распределения Ic типа (при $\gamma \geq 1, \gamma u = 1$) находятся на интервале $1{,}8<\beta_2<3$ оси ординат. Выше кривой II′ типа находится область распределений, для которых дискриминант $B^2-4AC<0$.

Эту область покрывают распределения III-V типов, принадлежащие трем основным системам непрерывных распределений. На рис. 4.2 последние расположены ниже прямой $\beta_2=6+0{,}75\beta_1$. Распределения IV типа (при $u \to \pm\infty$) лежат на прямой $\beta_2=5+\beta_1$.

Сделаем некоторые **выводы**.

Рассмотренный выше классический метод моментов оценивания параметров имеет существенные недостатки. Во-первых, он применим лишь к распределениям, имеющим моменты вплоть до четвертого порядка, при этом параметр $|\beta|=1$. Во-вторых, метод моментов весьма чувствителен к выбросам на концах статистического распределения, т.е. он не относится к устойчивым методам оценивания параметров.

Как показала практика, этот метод хорошо работает в случае выравнивающих распределений I типа, заданных на ограниченном с обеих сторон интервале, особенно если распределение близко к симметричному.

Для того, чтобы полнее использовать возможности обобщенных распределений по выравниванию статистических рядов распределения, необходимо иметь **общие методы оценивания параметров для распределений всех типов, в том числе не имеющих моментов выше нулевого порядка** (в традиционном их понимании). Методы должны быть общими для трех плотностей (6.2), (6.3), (6.4).

Ниже рассматриваются два таких метода, разработанные автором.

Тема 5. Универсальный метод моментов вычисления закона распределения и оценок параметров

5.1Универсальный метод моментов

За пределами применимости классического метода моментов **остается широкий класс распределений, для которых не существует моментов высоких порядков. Оценки параметров таких распределений могут быть найдены по универсальному методу моментов**, который впервые был описан автором в 1983г.

Основное отличие этого метода от классического метода моментов заключается прежде всего в том, что он применяется к распределениям, заданным обобщенной плотностью *p(x)*. **Другие плотности должны быть приведены к этой форме**. Например, вместо плотности (3.8), которую представим в виде (при $\gamma = k\beta$)

$$p(t) = Nt^{k\beta-1}\left(1-\alpha u t^{\beta}\right)^{\frac{1}{u}-1} \tag{5.1}$$

используется плотность $tp(t) = p(\ln t)$, т.е.

$$tp(t) = p(\ln t) = Ne^{k\beta \ln t}\left(1-\alpha u e^{\beta \ln t}\right)^{\frac{1}{u}-1}. \tag{5.2}$$

Здесь последнее равенство получено из предыдущего путем умножения на *t* обеих его частей и использования записи $e^{\beta \ln t}$ вместо t^{β}, что одно и то же.

Введем далее обозначение *lnt* = *x*. Тогда последнее равенство примет вид

$$p(x) = Ne^{k\beta x}\left(1-\alpha u e^{\beta x}\right)^{\frac{1}{u}-1}, \tag{5.3}$$

т.е. получили обобщенную плотность *p*(*x*).

Если плотность *p(t)* привести к форме $tp(t) = p(\ln t)$, то она будет обладать всеми свойствами плотности *p(x)*.

Плотность

$$p(y) = \frac{N\left(\ln y\right)^{k\beta-1}}{y}\left[1-\alpha u\left(\ln y\right)^{\beta}\right]^{\frac{1}{u}-1} \tag{5.4}$$

также приводится к форме плотности *p*(*x*).

Умножим обе части последней формулы на произведение $y\ln y$, а величину $(\ln y)^{\beta}$ запишем в виде $e^{\beta \ln\ln y}$. В результате получим

$$y\ln y\; p(y) = Ne^{k\beta \ln\ln y}\left(1-\alpha u e^{\beta \ln\ln y}\right)^{\frac{1}{u}-1}. \tag{5.5}$$

Приняв далее обозначения $\ln\ln y = x$, $y\ln y\; p(y) = p(\ln\ln y) = p(x)$, получим плотность (5.3).

Далее так же, как и в классическом методе моментов, центральные моменты μ_2, μ_3, μ_4, а также показатели асимметрии $\beta_1 = \mu_3^2/\mu_2^3$ и островершинности $\beta_2 = \mu_4 / \mu_2^2$ выражаются через параметры обобщенного распределения (5.3). При этом показатели β_1 и β_2 зависят лишь от двух параметров формы ($k = \gamma/\beta$, u) и в зависимости от их значений распределения разделяются на типы.

Приравнивая далее эмпирические значения показателей β_1^*, β_2^* теоретическим β_1, β_2, устанавливаем тип выравнивающей кривой распределения и находим оценки двух параметров формы *k, u.* Оценки двух других параметров – β, α (или произведения αu) вычисляются по простым формулам при известных оценках параметров *k, u* .

Отметим, что статистические центральные моменты *r*-го порядка в зависимости от вида плотности выравнивающего распределения вычисляются по формулам:

– в случае обобщенной плотности $p(x)$

$$\mu_r^* = \frac{1}{M}\sum_{i=1}^{n}\left(x_i - \nu_1^*\right)^r m_i, \text{ где } \nu_1^* = \bar{x} = \frac{1}{M}\sum_{i=1}^{n} x_i m_i \ ;$$

– в случае обобщенной плотности *p(t),* которая приводится к форме $tp(t) = p(\ln t)$,

$$\mu_r^* = \frac{1}{M}\sum_{i=1}^{n}\left(\ln t_i - \nu_1^*\right)^r m_i, \text{ где } \nu_1^* = \overline{\ell n t} = \frac{1}{M}\sum_{i=1}^{n} \ell n t_i m_i \ ;$$

– в случае обобщенной плотности *p(y),* которая приводится к форме $y \ln y p(y) = p(\ln\ln y)$,

$$\mu_r^* = \frac{1}{M}\sum_{i=1}^{n}\left(\ln\ln y_i - \nu_1^*\right)^r m_i,$$

где

$$\nu_1^* = \overline{\ln\ln y} = \frac{1}{M}\sum_{i=1}^{n}\left(\ln\ln y_i\right) m_i .$$

Здесь n – число интервалов группирования статистических данных; m_i – частота i-го интервала; $M = \sum m_i$ - объем выборки.

Это обеспечивает единый порядок установления типа выравнивающего распределения и нахождения оценок параметров для трех основных систем непрерывных распределений.

Эти же моменты используются для оценивания параметров трех дополнительных систем непрерывных распределений, т.е. в случае классического метода моментов.

Рассмотрим для примера распределения III-V типов, заданные плотностью

$$p(x) = \frac{\beta(-\alpha u)^k \Gamma\left(1-\frac{1}{u}\right)}{\Gamma(k)\Gamma\left(1-\frac{1}{u}-k\right)} \frac{e^{k\beta x}}{\left(1-\alpha u e^{\beta x}\right)^{1-\frac{1}{u}}} \qquad (5.6)$$

Как отмечалось выше, этими распределениями можно дополнить первую (дополнительную) систему распределений (см. табл.6.1) при условии $B^2 - 4AC < 0$.

Выразим центральные моменты (2 – 4)-го порядков и начальный момент 1-го порядка (математическое ожидание) через параметры распределения (5.6).

Используя **теорию производящих функций**, для обобщений плотности (5.6) получим:

$$\left.\begin{aligned} \nu_1 &= \frac{1}{\beta}\left[\Psi(k) - \Psi(k') - \ln(-\alpha u)\right] \\ \mu_2 &= \frac{1}{\beta^2}\left[\Psi'(k) + \Psi'(k')\right] \\ \mu_3 &= \frac{1}{\beta^3}\left[\Psi''(k) - \Psi''(k')\right] \\ \mu_4 &= 3\mu_2^2 + \frac{1}{\beta^4}\left[\Psi'''(k) + \Psi'''(k')\right] \end{aligned}\right\} \quad (5.7)$$

где $k'=1-1/u-k$.

Показатели асимметрии β_1 и островершинности β_2 равны

$$\left.\begin{aligned} \beta_1 &= \frac{\left[\Psi''(k) - \Psi''(k')\right]^2}{\left[\Psi'(k) + \Psi'(k')\right]^3} \\ \beta_2 &= 3 + \frac{\Psi'''(k) + \Psi'''(k')}{\left[\Psi'(k) + \Psi'(k')\right]^2} \end{aligned}\right\} \quad (5.8)$$

Заменяя показатели β_1, β_2 их оценками, **из системы двух уравнений (5.4.8) можно найти оценки двух параметров *k, u*, предварительно установив по тем же показателям тип выравнивающей кривой.**

Это – большое преимущество перед методом наибольшего правдоподобия, который требует решения четырех уравнений с четырьмя неизвестными, причем при условии, когда тип распределения заранее задан.

Для нахождения оценок параметров β и α (или произведения α*u*) введем случайную величину *Z*, которая связана со случайной величиной *X* зависимостью $Z = -\alpha u e^{\beta X}$ (см. формулу (5.6)) и рассмотрим ее логарифм

$$\boldsymbol{lnZ = \beta X + ln(-\alpha u)}.$$

Это уравнение является базой для построения универсального метода моментов.

Найдем математическое ожидание логарифма случайной величины *Z*

$$M(\ln Z) = \beta M(X) + \ln(-\alpha u).$$

Из последней формулы следует, что

$$M(X) = \frac{1}{\beta}\left[M(\ln Z) - \ln(-\alpha u)\right],$$

$$M\left[X - M(X)\right]^r = \frac{1}{\beta^r} M\left[\ln Z - M(\ln Z)\right]^r$$

или $$\mu_r = \mu_r^{(Z)} / \beta^r .$$

С учетом полученных равенств первые две формулы из четырех формул (5.4.7) можно переписать в виде

$$\left.\begin{aligned} \nu_1 &= \frac{1}{\beta}\left[\nu_1^{(z)} - \ln(-\alpha u)\right] \\ \mu_2 &= \frac{\mu_2^{(z)}}{\beta^2} \end{aligned}\right\}, \qquad (5.7')$$

где $\nu_1^{(z)} = M(\ln Z) = \Psi(k) - \Psi(-1/u - k)$ - математическое ожидание случайной величины $\ln Z$; $\mu_2^{(z)} = \Psi'(k) + \Psi'\left(1 - \frac{1}{u} - k\right)$ - центральный момент второго порядка случайной величины $\ln Z$;

$\Psi'(k)$ - первая производная пси-функции $\Psi(k) = \frac{d}{dk}\ln\Gamma(k)$.

На основании (7.4.7′) оценки параметра β и произведения α*u* равны

$$\beta = \sqrt{\frac{\mu_2^{(z)}}{\mu_2}}, \qquad (5.9)$$

$$\alpha u = -e^{\nu_1^{(z)} - \beta\nu_1}, \qquad (5.10)$$

где $\nu_1 = M(X)$. При вычислении оценок β и α*u* центральный момент второго порядка случайной величины *X* следует заменить его оценкой μ_2^* (выборочной дисперсией), а *M(X)* – выборочным средним $\nu_1^* = \bar{x}$.

Аналогично выводятся формулы для оценок параметров распределений других типов, заданных плотностью $p(x)$.

Так, в случае распределений II, II′ типов имеем

$$\beta = \sqrt{\frac{\mu_2^{(z)}}{\mu_2}}; \qquad \alpha = e^{\pm\left(\nu_1^{(z)} - \beta\nu_1\right)}, \qquad (5.11)$$

где $\nu_1^{(z)} = \pm\Psi(k)$; $\mu_2^{(z)} = \Psi'(k)$.

Здесь знак "+" относится ко II типу, а "–" – ко II′ типу.

В случае распределений I, I′ типов

$$\beta = \sqrt{\frac{\mu_2^{(z)}}{\mu_2}}; \qquad \alpha u = e^{\pm\left(\nu_1^{(z)} - \beta\nu_1\right)}, \qquad (5.12)$$

где

$$\nu_1 = \pm\left[\Psi(k) - \Psi\left(k + \frac{1}{u}\right)\right],$$
$$\mu_2 = \Psi'(k) - \Psi'\left(k + \frac{1}{u}\right).$$

При расчетах по универсальному методу моментов необходимо уметь вычислять с заданной точностью значения **гамма-функции и ее логарифмических производных**. Ниже приводятся приближенные формулы для их вычисления.

Логарифм гамма-функции находится по формуле

$$\left.\begin{aligned} &\ln\Gamma(x) \approx \frac{\ln 2\pi}{2} - \sum_{S=x}^{x+n} \ln s + \left(x + n + \frac{1}{2}\right)\ln(x+n) - \\ &-(x+n) + \frac{1}{12(x+n)} - \frac{1}{360(x+n)^3} + \frac{1}{1260(x+n)^5} + \ldots \end{aligned}\right\} \quad (5.13)$$

Логарифмические производные гамма-функции на основании (5.13) **равны**:

$$\left.\begin{aligned} &\Psi(x) = \frac{d}{dx}\ln\Gamma(x) \approx -\sum_{s=x}^{x+n} \frac{1}{s} + \ln(x+n) + \frac{1}{2(x+n)} - \frac{1}{12(x+n)^2} + \\ &+ \frac{1}{120(x+n)^4} - \frac{1}{252(x+n)^6} + \ldots \\ &\Psi'(x) = \sum_{s=x}^{x+n} \frac{1}{s^2} + \frac{1}{x+n} - \frac{1}{2(x+n)^2} + \frac{1}{6(x+n)^3} - \frac{1}{30(x+n)^5} + \ldots \\ &\Psi''(x) = -2\sum_{s=x}^{x+n} \frac{1}{s^3} - \frac{1}{(x+n)^2} + \frac{1}{(x+n)^3} - \frac{1}{2(x+n)^4} + \frac{1}{6(x+n)^6} + \ldots \\ &\Psi'''(x) = 6\sum_{s=x}^{x+n} \frac{1}{s^4} + \frac{2}{(x+n)^3} - \frac{3}{(x+n)^4} + \frac{2}{(x+n)^5} - \frac{1}{(x+n)^7} + \ldots \end{aligned}\right\} \quad (5.14)$$

Для облегчения различных **расчетов в Приложении 1 дана таблица значений функций** $\Gamma(x), \Psi(x), \Psi'(x)$ при $x = 0{,}1 \div 1$ с шагом 0,01.

Точность приведенных формул тем выше, чем больше сумма $x+n$. Для приближенных расчетов на калькуляторе можно принять $n=2$, а при более точных расчетах на ПЭВМ – $n=4$.

Рассчитаем для разных типов распределений значения показателей β_1, β_2 при различных значениях параметров *k, u*. Далее в системе координат (β_1, β_2) отметим области для распределений разных типов (см. рис.5.1).

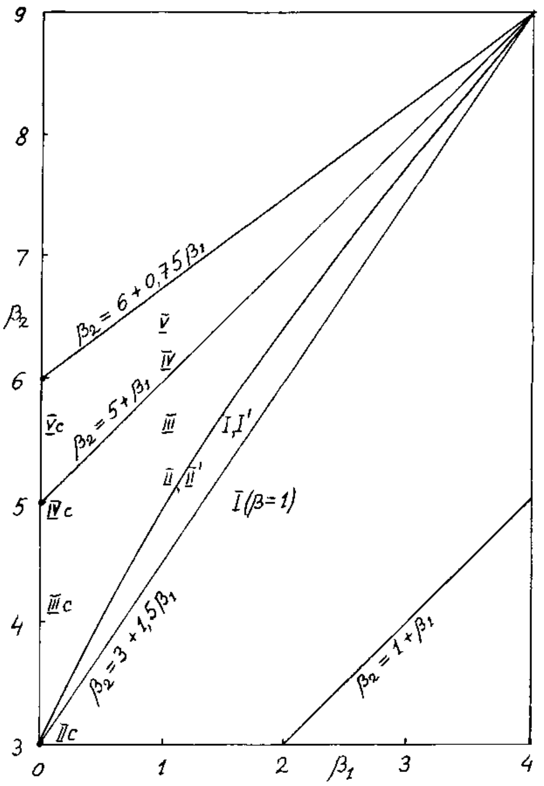


Рис. 5.1. Классификация распределений, заданных обобщенной плотностью $p(x) = Ne^{k\beta x}\left(1-\alpha u e^{\beta x}\right)^{\frac{1}{u}-1}$, по критериям β_1, β_2.

Как видно из рисунка, распределения II, II′ типов представлены кривой, распределения IV типа – прямой $\beta_2 = 5 + \beta_1$. Распределения V типа лежат ниже прямой $\beta_2 = 6 + 0.75\beta_1$, а распределения I, I′ типов – выше прямой $\beta_2 = 3 + 1.5\beta_1$

Симметричные распределения III типа с параметрами формы $k = 0,5(1-1/u)$ представлены отрезком $3 < \beta_2 < 6$ оси ординат**. С ростом параметра *k* распределения IIIс типа, а также II, II′ типов приближаются к нормальному закону**, для которого $\beta_1 = 0, \quad \beta_2 = 3$.

В заключение отметим, что ниже прямой $\beta_2 = 3 + 1,5\beta_1$ находится область распределений I типа, заданных плотностью (4.10) с параметром β=1.

Для быстрого установления типа выравнивающей кривой и нахождения оценок параметров *k, u* по методу моментов автором построена **номограмма** (Приложение 2). **Она строилась для распределений с левосторонней асимметрией**, у которых центральный момент 3-го порядка $\mu_3 < 0$.

Если статистическое распределение имеет правостороннюю асимметрию ($\mu_3 > 0$), то в случае распределений III-V типов вначале с помощью номограммы находятся оценки параметров *u, k′,* затем вычисляется оценка параметра *k*:

$$k = 1 - \frac{1}{u} - k'. \qquad (5.15)$$

Аналогично для распределений I типа (β=1) при $\mu_3 > 0$ вначале по номограмме находятся оценки параметров *u′, k′*, затем вычисляются оценки параметров *u, k*:

$$u = \frac{1}{k'};\quad k = \frac{1}{u'}. \qquad (5.16)$$

Построенная номограмма состоит из двух частей. Верхняя часть (выше прямой $\beta_2 = 3 + 1.5\beta_1$) относится к распределениям с плотностью *p(x)* или к распределениям с плотностью *p(t), p(y),* которые приведены соответственно к форме $tp(t) = p(\ln t);\, y \ln y\, p(y) = p(\ln \ln y)$.

Нижняя часть номограммы относится к распределениям I типа с плотностью *p(t)* при β=1 (т.е. типа 1.1) и является продолжением верхней части. Прямой $\beta_2 = 3 + 1{,}5\beta_1$ при $\mu_3 > 0$ представлены распределения второго типа с параметром $\beta = 1$ (гамма-распределения).

Это дает возможность расширить основные системы непрерывных распределений за счет включения в них распределений типов 1.1 и 2.1, которые относятся к дополнительным системам непрерывных распределений (с параметром β=1).

Тогда **первая (основная) система непрерывных распределений SRN1 в общем случае будет включать три обобщенные плотности**

$$\left.\begin{aligned} p(x) &= N e^{k\beta x}\left(1 - \alpha u e^{\beta x}\right)^{\frac{1}{u}-1} \\ p(t) &= N (t-l)^{k-1}\left[1 - \alpha u (t-l)\right]^{\frac{1}{u}-1} \\ p(t) &= N\left[1 - \alpha u (t-\bar{t})^2\right]^{\frac{1}{u}-1} \end{aligned}\right\} \qquad (5.17)$$

Первая система непрерывных распределений включает две группы симметричных распределений: одна из них (типы IIIc-Vc) задана плотностью *p(x)* при $k = 0{,}5(1-1/u)$, другая (типы Ic-Vc) – плотностью *p(t)* при $\beta = 2,\ \gamma = 1$. Кроме того, симметричные распределения Ic типа описываются также плотностью *p(t)* с параметром сдвига *l* при *ku* = 1.

Первая система непрерывных распределений может быть также задана двумя плотностями (без последней) или даже одной плотностью *p*(*x*).

Аналогично во **вторую основную систему непрерывных распределений SRN2 войдут обобщенные плотности**

$$\left.\begin{aligned} p(t) &= N t^{k\beta-1}\left(1 - \alpha u t^{\beta}\right)^{\frac{1}{u}-1} \\ p(y) &= \frac{N(\ln y - l)^{k-1}}{y}\left[1 - \alpha u(\ln y - l)\right]^{\frac{1}{u}-1} \\ p(y) &= \frac{N}{y}\left[1 - \alpha u(\ln y - \overline{\ln y})^2\right]^{\frac{1}{u}-1} \end{aligned}\right\} \qquad (5.18)$$

которые получены из первой системы как распределения функций случайных аргументов: $X=lnT$ - для первой плотности; $T=lnY$ – для двух других плотностей.

Наконец, **в третью основную систему непрерывных распределений SNR3 войдут обобщенные плотности**

$$\left.\begin{aligned} p(y) &= \frac{N(\ln y)^{k\beta-1}}{y}\left[1-\alpha u(\ln y)^{\beta}\right]^{-\frac{1}{u}-1} \\ p(w) &= \frac{N(\ln\ln w-l)^{k-1}}{w\ln w}\left[1-\alpha u(\ln\ln w-l)^{k}\right]^{-\frac{1}{u}-1} \\ p(w) &= \frac{N}{w\ln w}\left[1-\alpha u(\ln\ln w-\overline{\ln\ln w})^{2}\right]^{-\frac{1}{u}-1} \end{aligned}\right\} \qquad (5.19)$$

Вторая и третья основные системы непрерывных распределений также могут быть заданы либо двумя плотностями (без третьей), либо одной первой плотностью распределения.

Для нахождения оценок параметров трех основных систем непрерывных распределений по методу моментов автором созданы **программы** SNR1MM, SNR2MM, SNR3MM .

Номограмма, представленная в Приложении 2, **остается справедливой для трех основных систем непрерывных распределений**.

5.4.1. Законы распределения суммы независимых случайных величин

Системы непрерывных распределений, заданные обобщенными плотностями, а также методы оценивания параметров, доведенные до программной реализации, позволяют более просто решать различные задачи.

Пусть, например, требуется установить закон распределения суммы n независимых одинаково распределенных случайных величин $X=X_1+X_2+...+X_n$. Среднее каждой случайной величины равно ν_1.

Распределение случайной величины X_i может быть задано как **аналитически**, так и **таблично**.

Поэтому для нахождения **закона распределения суммы *n* независимых случайных величин, т.е. композиции *n* распределений**, можно использовать **общий метод**.

Для этого достаточно вычислить моменты суммы *n* независимых случайных величин $\nu_{1(n)}$, $\mu_{2(n)}$, $\mu_{3(n)}$, $\mu_{4(n)}$, а также показатели $\beta_{1(n)}$ и $\beta_{2(n)}$ по известным моментам случайной величины X_i.

Далее по методу моментов (универсальному или классическому) с помощью программы **устанавливается тип выравнивающей кривой и находятся оценки параметров**.

Пусть моменты случайной величины X_i известны. Обозначим их соответственно ν_1, μ_2, μ_3, μ_4.

Тогда **среднее суммы n независимых случайных величин будет равно**

$$\nu_{1(n)} = \sum_{i=1}^{n} \nu_{1i} \quad . \qquad (5.20)$$

Если случайные величины X_i равны и подчиняются одному и тому же закону распределения, то

$$\nu_{1(n)} = n\nu_1 . \qquad (5.21)$$

Найдем далее центральный момент второго порядка суммы *n* независимых одинаково распределенных случайных величин $\mu_{2(n)}$.

Начнем с рассмотрения суммы двух независимых случайных величин:

$$\mu_2(X+Y) = M\left[(X+Y) - (m_x + m_y)\right]^2 = M\left[(X-m_x) + (Y-m_y)\right]^2 ,$$

где $m_x = \nu_1(x),\ \ m_y = \nu_1(y)$.

Обозначим для краткости $X - m_x = x, Y - m_y = y$. Тогда

$$\mu_2(X+Y) = M(x+y)^2 = M(x^2 + 2xy + y^2) = M(x^2) + 2M(xy) + M(y^2).$$

Поскольку *M(xy) = M(x)M(y),* последнее выражение можно представить в виде

$$\mu_2(X+Y) = M(X-m_x)^2 + 2M(X-m_x)M(Y-m_y) + M(Y-m_y)^2,$$

или

$$\mu_2(X+Y) = \mu_2(X) + 2\mu_1(X)\mu_1(Y) + \mu_2(Y) .$$

Но центральный момент первого порядка равен нулю. Поэтому второе слагаемое здесь равно нулю, и последняя формула примет вид

$$\mu_2(X+Y) = \mu_2(X) + \mu_2(Y) . \qquad (5.22)$$

На основании рассмотренного примера можно сформулировать следующее **правило**: **при возведении в *r*-ю степень суммы случайных величин** $x=X-m_x$; $y=Y-m_y,\ldots$ **в итоге следует учесть только те члены, которые не содержат первых степеней сомножителей, так как их математические ожидания равны нулю.**

Используя это правило, **найдем центральный момент второго порядка суммы трех случайных величин**

$$\mu_2(X+Y+Z) = M(x+y+z)^2,$$

где $x=X-m_x;\ y=Y-m_{y;}\ z=Z-m_z$.

Итак,

$$\mu_2(X+Y+Z) = M(x+y+z)^2 = M(x^2 + y^2 + z^2 + \ldots).$$

Здесь не записаны члены, математические ожидания которых равны нулю. Следовательно,

$$\mu_2(X+Y+Z) = \mu_2(X) + \mu_2(Y) + \mu_2(Z) . \qquad (5.23)$$

В случае суммы *n* независимых одинаково распределенных случайных величин

$$\mu_{2(n)} = n\mu_2 . \qquad (5.24)$$

Найдем далее выражение для центрального момента третьего порядка суммы *n* независимых одинаково распределенных случайных величин.

Рассмотрим вначале **сумму двух независимых случайных ве**личин

$$\mu_3(X+Y) = M\left[(x+y)^3\right] = M\left[x^3 + 3x^2y + 3xy^2 + y^3\right],$$

откуда

$$\mu_3(X+Y) = \mu_3(X) + \mu_3(Y). \qquad (5.25)$$

Аналогично **для суммы трех случайных величин** имеем

$$\mu_3(X+Y+Z) = M\left[(x+y+z)^3\right] = M\left[(x+y+z)^2(x+y+z)\right] =$$
$$= M\left[(x^2+y^2+z^2+2xy+2xz+2yz)(x+y+z)\right] = M\left[x^3+y^3+z^3+...\right]$$

Остальные члены в квадратных скобках равны нулю.

Таким образом,

$$\mu_3(X+Y+Z) = \mu_3(X) + \mu_3(Y) + \mu_3(Z). \qquad (5.26)$$

Для **суммы *n* независимых одинаково распределенных случайных величин**

$$\mu_{3(n)} = n\mu_3. \qquad (5.27)$$

И, наконец, **найдем выражение для центрального момента четвертого порядка суммы *n* независимых одинаково распределенных случайных величин.**

Начнем с **суммы двух случайных величин**

$$\mu_4(X+Y) = M(x+y)^4,$$

где по-прежнему $x = X - m_x$; $y = Y - m_y$. Итак,

$$\mu_4(X+Y) = M\left[(x+y)^4\right] = M\left[(x+y)^3(x+y)\right] =$$
$$= M\left[(x^3+3x^2y+3xy^2+y^3)(x+y)\right] = M(x^4+3x^2y^2+3x^2y^2+y^4+...) =$$
$$= M(x^4) + 6M(x^2)M(y^2) + M(y^4).$$

Отсюда имеем

$$\mu_4(X+Y) = \mu_4(X) + \mu_4(Y) + 6\mu_2(X)\mu_2(Y). \qquad (5.28)$$

Если $X=Y$, то $\mu_{4(n=2)} = 2\mu_4 + 6\mu_2^2$. (5.29)

Найдем далее центральный момент четвертого порядка **суммы трех случайных величин**

$$\mu_4(X+Y+Z) = M\left[(x+y+z)^4\right] = M\left[(x+y+z)(x+y+z)\right]^2 =$$
$$= M(x^2+y^2+z^2+2xy+2xz+2yz)^2 =$$
$$= M(x^4+y^4+z^4+6x^2y^2+6x^2z^2+6y^2z^2+...),$$

откуда

$$\mu_4(X+Y+Z) = \mu_4(X) + \mu_4(Y) + \mu_4(Z) +$$
$$+ 6\mu_2(X)\mu_2(Y) + 6\mu_2(X)\mu_2(Z) + 6\mu_2(Y)\mu_2(Z). \qquad (5.30)$$

Если $X=Y=Z$, то

$$\mu_{4(n=3)} = 3\mu_4 + 18\mu_2^2. \qquad (5.31)$$

На основании полученных ранее формул можно записать **общее выражение для центрального момента 4-го порядка суммы *n* независимых одинаково распределенных случайных величин.**

$$\mu_{4(n)} = n\mu_4 + 3n(n-1)\mu_2^2. \qquad (5.32)$$

Действительно, произведение $3n(n-1)$ при n=2 равно 6, а при n=3 равно 18.

Таким образом, **моменты суммы *n* независимых одинаково распределенных случайных величин $X=\sum X_i$ равны**

$$\left.\begin{aligned} \nu_{1(n)} &= n\nu_1 \\ \mu_{2(n)} &= n\mu_2 \\ \mu_{3(n)} &= n\mu_3 \\ \mu_{4(n)} &= n\mu_4 + 3n(n-1)\mu_2^2 \end{aligned}\right\} \qquad (5.33)$$

и легко вычисляются по моментам отдельной случайной величины X_i. Далее по известным моментам можно найти выравнивающее распределение суммы n независимых одинаково распределенных случайных величин.

При этом **найденное выравнивающее распределение может совпадать с композицией законов распределения слагаемых** (например, в случае n показательных законов), **но может и не совпадать с ней** (например, если случайные величины распределены по закону равномерной плотности). Это связано с тем, что **моменты не определяют полностью распределения.**

5.4.2. Центральная предельная теорема для трех систем непрерывных распределений

Используя **теорию производящих функций**, можно показать, что для первой системы непрерывных распределений, заданной обобщенной плотностью

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

моменты суммы n независимых одинаково распределенных случайных величин $X=\sum X_i$ связаны с моментами случайной величины X_i формулами (7.4.33).

Производящая функция в этом случае есть

$$M\left[e^{(X-\nu_1)t}\right],$$

где t – вспомогательный параметр.

Выразим на основании формул (5.4.33) **показатели асимметрии и островершинности распределения суммы *n* независимых одинаково распределенных случайных величин** $\beta_{1(n)}$ **и** $\beta_{2(n)}$ **через аналогичные показатели отдельной случайной величины X_i, т.е.** β_1 **и** β_2:

$$\left.\begin{aligned} \beta_{1(n)} &= \frac{\mu_{3(n)}^2}{\mu_{2(n)}^3} = \frac{n^2\mu_3^2}{n^3\mu_2^3} = \frac{\beta_1}{n} \\ \beta_{2(n)} &= \frac{\mu_{4(n)}}{\mu_{2(n)}^2} = 3 + \frac{\beta_2 - 3}{n} \end{aligned}\right\} \qquad (5.34)$$

Из формул (5.4.34) немедленно **следует центральная предельная теорема теории вероятностей (для первой системы непрерывных распределений)**:

распределение суммы *n* независимых одинаково распределенных случайных величин $X = \sum X_i$ **с ростом *n* приближается к нормальному закону**

$$p(x) = \sqrt{\frac{\alpha}{\pi}}\, e^{-\alpha (x-\nu_{1(n)})^2}, \tag{5.35}$$

для которого $\beta_1=0$, $\beta_2=3$. При этом на номограмме (Приложение 2) точка с координатами $(\beta_{1(n)}, \beta_{2(n)})$ с ростом *n* перемещается по прямой от точки (β_1, β_2) исходного распределения (случайной величины X_i) к точке (0,3) нормального закона, оценки параметров которого равны

$$\left.\begin{aligned} \nu_{1(n)} &= n\nu_1 \\ \alpha &= \frac{1}{2\mu_{2(n)}} = \frac{1}{2n\mu_2} \end{aligned}\right\}. \tag{5.36}$$

Формулы (5.4.33), (5.4.34) позволяют также переходить от распределения суммы *n* независимых одинаково распределенных случайных величин к распределению отдельной случайной величины X_i.

Полученные выше результаты остаются в силе и для обобщенной плотности

$$p(t) = N t^{k\beta-1}\left(1-\alpha u t^{\beta}\right)^{\frac{1}{u}-1},$$

т.е**. в случае второй системы непрерывных распределений, если ее привести к форме плотности *p*(*x*)**, т.е. представить в виде

$$t p(t) = N e^{k\beta \ln t}\left(1-\alpha u e^{\beta \ln t}\right)^{\frac{1}{u}-1}.$$

Моменты случайной величины $\ln T_i$ будут задаваться формулами

$$\nu_1 = M(\ln T_i),$$

$$\mu_r = M(\ln T_i - \nu_1)^r.$$

Формулировка центральной предельной теоремы несколько изменится: **распределение суммы логарифмов *n* независимых одинаково распределенных случайных величин** $\ln T = \sum \ln T_i$ **с ростом *n* приближается к нормальному закону, а произведение *n* случайных величин $T=T_1T_2\ldots T_n$ – к логарифмически нормальному закону**

$$p(t) = \sqrt{\frac{\alpha}{\pi}}\,\frac{1}{t}\, e^{-\alpha(\ln t-\nu_{1(n)})^2}, \tag{5.37}$$

для которого $\beta_1=0$, $\beta_2=3$. Оценки параметров $\nu_{1(n)}$, α задаются формулами (7.4.36).

И, наконец, **в случае третьей системы непрерывных распределений**, заданных обобщенной плотностью

$$p(Y)=\frac{N(\ln Y)^{k\beta-1}}{Y}\left[1-\alpha u(\ln Y)^{\beta}\right]^{\frac{1}{u}-1},$$

полученные выше результаты остаются справедливыми, если ее также привести к форме плотности $p(x)$, т.е. представить в виде

$$Yp(Y)\ln Y=Ne^{k\beta\ln\ln Y}\left(1-\alpha u e^{\beta\ln\ln Y}\right)^{\frac{1}{u}-1}.$$

Тогда моменты случайной величины $\ln\ln Y_i$ будут задаваться формулами

$$\nu_1=M(\ln\ln Y_i),$$

$$\mu_r=M(\ln\ln Y_i-\nu_1)^r.$$

Центральная предельная теорема сформулируется в виде: распределение суммы двойных логарифмов n независимых одинаково распределенных случайных величин $\ln\ln Y=\sum\ln\ln Y_i$ с **ростом n приближается к нормальному закону, произведение** $\ln Y=\ln Y_1\ln Y_2\ldots\ln Y_n$ **– к логарифмически нормальному закону, а величина** $Y=e^{\ln Y_1\ln Y_2\ldots\ln Y_n}$ **– к двойному логарифмически нормальному закону**

$$p(Y)=\sqrt{\frac{\alpha}{\pi}}\frac{1}{Y\ln Y}e^{-\alpha(\ln\ln Y-\nu_{1(n)})^2}, \qquad (5.38)$$

для которого $\beta_1=0$, $\beta_2=3$.

Здесь также оценки параметров $\nu_{1(n)}$, α задаются формулами (5.36).

5.4.3. Законы распределения среднего выборочного

Рассмотрим *n* случайных величин, распределенных по одному и тому же закону, заданному, например, обобщенной плотностью *p(x)* (или *p(t)*, или *p(y)*).

Найдем закон распределения среднего выборочного $\bar{x}$, (или $\overline{\ln T}$, или $\overline{\ln\ln Y}$) по заданному закону распределения случайной величины *X* (или $\ln T$, или *lnlnY*).

Для решения этой задачи достаточно найти центральные моменты (2–4)-го порядков среднего арифметического *n* независимых одинаково распределенных случайных величин и вычислить показатели асимметрии и островершинности. Далее по универсальному или классическому методу моментов с помощью программы легко устанавливается тип искомого распределения и вычисляются оценки его параметров.

Известно, что **математическое ожидание среднего арифметического *n* независимых случайных величин с одинаковыми средними равно среднему отдельной случайной величины**

$$M(\bar{X})=\nu_1. \qquad (5.39)$$

Найдем далее **центральные моменты выборочного среднего**. Для этого вначале докажем, что постоянный множитель можно выносить за знак центрального момента r-го порядка, возведя его в r-ю степень:

$$\mu_r(CX) = C^r \mu_r(X). \qquad (5.40)$$

Действительно,

$$\mu_r(CX) = M\left[CX - M(CX)\right]^r = C^r M[X - M(X)]^r = C^r \mu_r(X).$$

На основании (7.4.40) имеем:

$$\mu_2(\bar{X}) = m_{2(n)} = \mu_2\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{\mu_2(X_1 + X_2 + ... + X_n)}{n^2}.$$

Поскольку **для *n* независимых одинаково распределенных случайных величин справедливо равенство** (5.4.24)

$$\mu_{2(n)} = n\mu_2,$$

то из предыдущего выражения получим

$$\mu_2(\bar{X}) = m_{2(n)} = \frac{n\mu_2}{n^2} = \frac{\mu_2}{n}. \qquad (5.41)$$

Найдем центральный момент третьего порядка выборочного среднего. На основании равенства (5.4.40) можем записать

$$\mu_3(\bar{X}) = m_{3(n)} = \mu_3\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n^3}\mu_3(X_1 + X_2 + ... + X_n).$$

Поскольку $\mu_{3(n)} = n\mu_3$ (см. формулу (5.4.27)), то из предыдущего выражения получим

$$\mu_3(\bar{X}) = m_{3(n)} = \frac{n\mu_3}{n^3} = \frac{\mu_3}{n^2}. \qquad (5.42)$$

Для центрального момента четвертого порядка среднего выборочного можем записать

$$\mu_4(\bar{X}) = m_{4(n)} = \mu_4\left(\frac{X_1 + X_2 + ... + X_n}{n}\right) = \frac{1}{n^4}\mu_4(X_1 + X_2 + ... + X_n).$$

Поскольку для суммы n независимых одинаково распределенных случайных величин справедливо равенство (5.32)

$$\mu_{4(n)} = n\mu_4 + 3n(n-1)\mu_2^2,$$

то из предыдущей формулы получим

$$\mu_4(\bar{X}) = m_{4(n)} = \frac{1}{n^3}\left[\mu_4 + 3(n-1)\mu_2^2\right]. \qquad (5.43)$$

Таким образом, **моменты среднего арифметического *n* независимых одинаково распределенных случайных величин выражаются через моменты отдельной случайной величины посредством формул**

$$
\left.\begin{aligned}
M(\bar{X}) &= \nu_1 \\
m_{2(n)} &= \frac{\mu_2}{n} \\
m_{3(n)} &= \frac{\mu_3}{n^2} \\
m_{4(n)} &= \frac{1}{n^3}\left(\mu_4 + 3(n-1)\mu_2^2\right)
\end{aligned}\right\} \qquad (5.44)
$$

Из формул (5.33) и (5.44) найдем взаимосвязь между моментами **суммы** и **среднего арифметического** *n* независимых одинаково распределенных случайных величин:

$$
\left.\begin{aligned}
M(\bar{X}) &= \frac{\nu_{1(n)}}{n} = \nu_1 \\
m_{r(n)} &= \frac{\mu_{r(n)}}{n^r}
\end{aligned}\right\} \qquad (5.45)
$$

Из (7.4.45) следует, что **показатели асимметрии** β_1 **и островершинности** β_2 **для распределений суммы и среднего арифметического *n* независимых одинаково распределенных случайных величин одни и те же**.

Зная центральные моменты, а также показатели асимметрии и островершинности, по универсальному или классическому методу моментов нетрудно найти закон распределения среднего арифметического и, следовательно, вычислить доверительные границы для среднего выборочного при заданной доверительной вероятности и любом заданном значении *n*. **С ростом *n* распределение среднего выборочного** $\bar{x}$ **(или** $\overline{\ln T}$**, или** $\overline{\ln \ln Y}$**) приближается к нормальному закону (центральная предельная теорема в случае первой системы непрерывных распределений**).

В случае второй системы - распределение среднего геометрического случайной величины *T*

$$
\bar{T}_{геом.} = \left(T_1 T_2 \ldots T_n\right)^{\frac{1}{n}} = e^{\frac{\Sigma \ln Ti}{n}} = e^{\overline{\ln T}}
$$

с ростом *n* приближается к логарифмически нормальному закону.

В случае третьей системы - распределение среднего геометрического логарифмов отдельных случайных величин

$$
\overline{\ln Y}_{геом} = \left(\ln Y_1 \cdot \ln Y_2 \ldots \ln Y_n\right)^{\frac{1}{n}}
$$

с ростом *n* приближается к логарифмически нормальному закону, а величина

$$
Y = e^{\left(\ln Y_1 \cdot \ln Y_2 \ldots \ln Yn\right)^{\frac{1}{n}}}
$$

- к двойному логарифмически нормальному закону.

При n > 100 распределение среднего выборочного $\bar{x}$ (или $\overline{\ln T}$, или $\overline{\ln \ln Y}$) можно считать нормальным.

Действительно, если распределение случайной величины X (или $\ln T$, или $\ln \ln Y$) характеризуется показателями $0 < \beta_1 < 4$, $3 < \beta_2 < 9$ (в случае трех основных систем непрерывных распределений), то распределение среднего арифметического n = 100 независимых одинаково распределенных случайных величин на основании формул (7.4.34) будет иметь показатели

$$0 < \beta_1(n) < 0{,}04; \quad 3 < \beta_{2(n)} < 3{,}06, \tag{5.46}$$

т.е. близкие к нормальному закону.

Отметим, что формулы (5.34) и (5.44) позволяют переходить от распределения среднего арифметического n независимых одинаково распределенных случайных величин к распределению отдельной случайной величины.

Полученные результаты вскрывают характер изменения показателей $\beta_{1(n)}$, $\beta_{2(n)}$ с изменением числа n независимых одинаково распределенных случайных величин: **с ростом *n* точка с координатами ($\beta_{1(n)}$; $\beta_{2(n)}$) приближается к точке (0; 3) по кратчайшему пути, т.е. по прямой**

$$\beta_{2(n)} = 3 + \frac{\beta_2 - 3}{\beta_1} \beta_{1(n)}, \tag{5.47}$$

которая следует из (5.4.34).

При этом степень приближения к нормальному закону при больших n зависит от исходных значений показателей β_1, β_2 случайной величины X_i.

Установим минимально необходимое значение *n*, при котором распределение среднего арифметического *n* независимых одинаково распределенных случайных величин можно считать нормальным.

Рассмотрим рис. 5.4.2. На нем точкой $A(\beta_1; \beta_2)$ обозначено распределение случайной величины X_i.

Распределение среднего арифметического n независимых одинаково распределенных случайных величин представлено точкой $B(\beta_{1(n)}; \beta_{2(n)})$.

Заштрихованная площадь – это область нормального закона. Она ограничена двумя вертикальными и двумя наклонными отрезками прямых, тангенс угла наклона которых к горизонтальной оси принят равным 1,75.

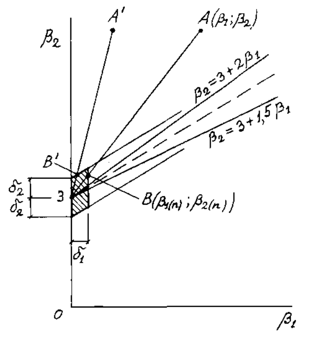


Рис. 5.4.2. Фрагмент номограммы в области нормального закона

Область нормального закона может быть задана и другими способами.

Если при некотором значении *n* точка *B* (или *B′*) попадает на границу или внутрь заштрихованной области, то закон распределения среднего арифметического можно считать нормальным.

Пусть точка *B* находится на правой вертикальной границе области нормального закона (см. рис. 5.4.2). При этом условии минимально необходимое значение *n* можно найти из формулы

$$\beta_{1(n)} = \frac{\beta_1}{n}$$

при $\beta_{1(n)} = \delta_1$:

$$n_1 = \frac{\beta_1}{\delta_1}. \qquad (5.48)$$

Если точка *B* займет положение *B′*, то она будет являться точкой пересечения двух наклонных прямых, которые задаются уравнениями

$$\beta_{2(n)} = 3 + \frac{\beta_2 - 3}{n}; \; \beta_{2(n)} = 3 + \delta_2 + 1{,}75\beta_{1(n)} = 3 + \delta_2 + 1{,}75\frac{\beta_1}{n}.$$

Приравнивая правые части этих уравнений, найдем

$$n_2 = \frac{\left|\beta_2 - 3 - 1{,}75\beta_1\right|}{\delta_2}. \qquad (5.49)$$

Из двух полученных значений *n* выбираем большее.

Если в формулах (5.48), (5.49) принять $\delta_1 = \delta_2 = 0{,}05$, то значения параметра *u* выравнивающих распределений по абсолютной величине будут несколько меньше 0,02 (для нормального закона *u*→0).

Рассмотрим *пример*.

Пусть случайная величина X_i описывается распределением II′ типа:

$$p(x) = \frac{\alpha^k}{\Gamma(k)} \frac{1}{x^{k+1} e^{\alpha/x}} \qquad (5.50)$$

с параметрами: α = 12; *k* = 5. Тогда теоретические моменты будут равны (см. формулы (5.3.24), (5.3.27) при *u*→0, γ = *k*):

$$\nu_1 = \frac{\alpha}{k-1} = 3,$$

$$\mu_2 = \frac{\alpha^2}{(k-1)^2(k-2)} = 3,$$

$$\mu_3 = \frac{4\alpha\mu_2}{(k-1)(k-3)} = 18,$$

$$\mu_4 = \frac{3\alpha}{(k-1)(k-4)}\left(\frac{\alpha\mu_2}{k-1} + 2\mu_3\right) = 405.$$

Показатели асимметрии и островершинности на основании (5.37) равны:

$$\beta_1 = \frac{16(k-2)}{(k-3)^2} = 12; \;\; \beta_2 = \frac{3(k-2)(k+5)}{(k-3)(k-4)} = 45$$

Найдем по формулам (5.48), (5.49) минимально необходимое значение n (объем выборки), при котором распределение среднего арифметического n независимых одинаково распределенных случайных величин можно считать нормальным.

При $\delta_1 = \delta_2 = 0{,}05$ имеем:

$$n_1 = \frac{\beta_1}{\delta_1} = \frac{12}{0{,}05} = 240; \quad n_2 = \frac{\beta_2 - 3 - 1{,}75\beta_1}{\delta_2} = \frac{45 - 3 - 1{,}75 \cdot 12}{0{,}05} = 420 .$$

Принимаем $n = 420$.

На практике часто ограничиваются небольшими значениями n $(n = 25 \div 30)$.

Пусть n = 25. Тогда распределение среднего арифметического n независимых одинаково распределенных случайных величин будет иметь моменты (см. формулу 7.4.44):

$$m_{2(n)} = \frac{\mu_2}{n} = \frac{3}{25} = 0{,}12,$$

$$m_{3(n)} = \frac{\mu_3}{n^2} = \frac{18}{25^2} = 0{,}0288,$$

$$m_{4(n)} = \frac{1}{n^3}\left(\mu_4 + 3(n-1)\mu_2^2\right) = \frac{1}{25^3}\left(405 + 3 \cdot 24 \cdot 3^2\right) = 0{,}067392 .$$

Показатели асимметрии и островершинности равны:

$$\beta_{1(n)} = \frac{m_{3(n)}^2}{m_{2(n)}^3} = \frac{0{,}0288^2}{0{,}12^3} = 0{,}48,$$

$$\beta_{2(n)} = \frac{m_{4(n)}}{m_{2(n)}^2} = \frac{0{,}067392}{0{,}12^2} = 4{,}68 .$$

Контроль:

$$\beta_{1(n)} = \frac{\beta_1}{n} = \frac{12}{25} = 0{,}48; \quad \beta_{2(n)} = 3 + \frac{\beta_2 - 3}{n} = 3 + \frac{45 - 3}{25} = 4{,}68 .$$

Выравнивающее распределение среднего арифметического при n=25 может быть описано обобщенной плотностью

$$p(x) = Ne^{\gamma x}\left(1 - \alpha u e^{\beta x}\right)^{\frac{1}{u} - 1}$$

(точнее, распределением III типа первой системы непрерывных распределений) с параметрами

$\alpha u = -8{,}036098\ E - 7$; $\beta = 5{,}062564$; $\gamma = 8{,}556006$; $u = -0{,}668744$ и нормирующим множителем N = 3,194777E–10.

Зададим доверительную вероятность P = 0,9973 и найдем по одной из программ (например, $SNR11M97$) доверительный интервал для среднего арифметического: $2{,}038459 < \bar{x} < 4{,}50299$. Его ширина составляет $7{,}114489\ S(\bar{x})$.

Для сравнения по той же программе найдем доверительный интервал при условии справедливости нормального закона: . Его ширина составляет $6S(\bar{x})$, т.е. ошибка в определении границ дове$1{,}960769 < \bar{x} < 4{,}039222$ рительного интервала по нормальному закону оказалась существенной.

Полученные выше **формулы (5.48) и (5.49) можно использовать также для вычисления минимально необходимого значения *n*, при котором распределение среднего арифметического логарифмов *n* независимых одинаково распределенных случайных величин можно считать нормальным**.

Используя универсальный метод моментов, для рассмотренной выше плотности II' типа с параметрами $\alpha = 12, k = \gamma = 5,\ \beta = 1$, приведенной к форме $xp(x) = p(\ln x)$, найдем:

$$M\left[\ln(x)\right] = \nu_1 = -\frac{1}{\beta}\left[\Psi(k) - \ln\alpha\right] = 0{,}978787\ ;$$

$$\mu_2 = \frac{1}{\beta^2}\Psi'(k) = 0{,}22132\ ;$$

$$\mu_3 = -\frac{1}{\beta^3}\Psi''(k) = 0{,}04879\ ;$$

$$\mu_4 = 3\mu_2^2 + \frac{1}{\beta^4}\Psi'''(k) = 0{,}16838\ .$$

Показатели асимметрии и островершинности равны:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = 0{,}2196\ ;\ \beta_2 = \frac{\mu_4}{\mu_2^2} = 3{,}4375\ .$$

Далее по формулам (7.4.48) и (7.4.49) имеем (при $\delta_1 = \delta_2 = 0{,}05$)

$$n_1 = \frac{\beta_1}{\delta_1} = 4{,}392\ ;\ n_2 = \frac{\beta_2 - 3 - 1{,}75\beta_1}{\delta_2} = 1{,}064\ .$$

Принимаем $n = 5$.

Итак, если случайная величина X задана плотностью (7.4.50), то распределение среднего арифметического логарифма случайной величины X близко к нормальному закону уже при $n = 5$, а распределение среднего арифметического самой случайной величины X близко к нормальному закону при значительно большей величине n ($n = 420$).

Если же центральные моменты высоких порядков не существуют (например, $\mu_4 \to \infty$), то и величина $n \to \infty$, т.е. распределение среднего арифметического случайной величины X ни при каком n не приближается к нормальному закону.

Тема 6. Общий устойчивый метод

Проверка показала, что универсальный метод моментов в принципе решает задачу оценивания параметров обобщенных распределений. Однако **существенным его недостатком является неустойчивость, поскольку эмпирические моменты высоких порядков (μ_3^*, μ_4^*) сильно зависят от значений частот на концах распределения**.

Поэтому автором обобщенных распределений был разработан **общий устойчивый метод оценивания параметров [10-13], который по точности не уступает методу наибольшего правдоподобия, но значительно проще последнего.**

Здесь так же, как и в случае универсального метода моментов, вводятся **два показателя – асимметрии *B* и островершинности *H*, которые зависят от двух параметров формы** $k=\gamma/\beta, u$. По этим показателям устанавливается тип выравнивающей кривой распределения и находятся оценки параметров *k, u*. Оценки двух других параметров рассчитываются по простым формулам.

Достоинством метода является его устойчивость, т.е. он мало чувствителен к выбросам на концах статистического распределения.

К недостаткам его следует отнести то, что для оценивания параметров выравнивающей кривой **он требует группирования статистических данных**, так же как и метод наибольшего правдоподобия.

Если обобщенное распределение задано плотностью $p(x)$, то показатели *B, H* равны

$$\left.\begin{aligned} B &= M\left[p(x)(x-M(x))\right] = f(k,u) \\ H &= S_3/S_1^3 = f(k,u) \end{aligned}\right\}, \qquad (6.1)$$

где

$$S_r = M\left[p(x)^r\right] = f(\beta,k,u). \qquad (6.2)$$

Исследования показали, что величина H задана на интервале $\sqrt{2} < H < 2$, а величина B – на интервале $-1/4<B<1/4$.

Вычислим для разных типов распределений значения показателей *B*, *H* при различных значениях параметров *k, u.* Далее построим **номограмму** (Приложение 3). **Она справедлива для трех основных систем непрерывных распределений, заданных первыми плотностями.** При этом **они должны быть приведены к форме плотности *p*(*x*).**

На номограмме распределения II, II′ и IV типов представлены кривыми. Типы I, I′ , III, V занимают определенные области. Симметричные распределения IIIc, Vc типов представлены отрезками на оси OH: для IIIc типов $\sqrt{2} < H < \pi^2/6$; для Vc типа $\pi^2/6 < H < 2$. Распределения IVc типа представлены точкой $H = \pi^2/6$. Распределения IIc типа также представлены точкой $H = \sqrt{2}$.

На **номограмме изображены области распределений с левосторонней асимметрией, для которых** $0 < B_1 < 1/4$. Сюда относится часть распределений III-V типов при $0<k<(1-1/u)/2$, а также распределения I, II типов. При этом распределения приведены к форме плотности $p(x)$.

Распределения I′, II′ типов, а также часть распределений III-V типов при $(1-1/u)/2<k<1-1/u$ имеют **правостороннюю асимметрию**. Для них $-1/4<B<0$, причем для распределений I, II и I′, II′ типов справедливы равенства: $B'=-B,\, H'=H$.

Таким образом, **показатели *B, H* однозначно определяют тип распределения, приведенного к форме плотности *p*(*x*). Более того, с помощью этих показателей могут быть найдены оценки параметров *u, k* непосредственно из номограммы.**

Для распределений III-V типов при $B<0$ из номограммы вначале находятся оценки параметров *k′, u* (при $B>0$), затем вычисляется величина $k=1-1/u-k'$.

Оценка параметра β для всех типов равна [11]

$$\beta=\frac{S_1}{S_1^{(z)}}\,. \tag{6.3}$$

Тогда $\gamma = k\beta$.

Оценки параметра α для распределений II, II′ типов и произведения αu для остальных типов равны [12]:

$$\left.\begin{aligned}
&\text{Типы I, I}':\quad \alpha u = e^{\pm\left(\nu_1^{(z)}-\beta\nu_1\right)}\\
&\text{Типы II, II}':\quad \alpha = e^{\pm\left(\nu_1^{(z)}-\beta\nu_1\right)}\\
&\text{Типы III - V}:\quad \alpha u = -e^{\nu_1^{(z)}-\beta\nu_1}
\end{aligned}\right\} \tag{6.4}$$

где в зависимости от типа распределения величины $\nu_1^{(z)}$ и $S_1^{(Z)}$ рассчитываются по формулам:

Типы I, I′:

$$\left.\begin{aligned}
&\nu_1^{(z)}=\pm\left[\Psi(k)-\Psi\left(k+\frac{1}{u}\right)\right]\\
&S_1^{(z)}=\frac{1}{2\sqrt{\pi}}\,\frac{2\left(k+\dfrac{1}{u}\right)-1}{\dfrac{2}{u}-1}\cdot\frac{g(k)\,g\left(\dfrac{1}{u}\right)}{g\left(k+\dfrac{1}{u}\right)}
\end{aligned}\right\} \tag{6.5}$$

Типы II, II′:

$$\nu_1^{(Z)}=\pm\,\Psi(k);\quad S_1^{(Z)}=\frac{g(k)}{2\sqrt{\pi}} \tag{6.6}$$

Типы III-V:

$$\left.\begin{aligned} \nu_1^{(z)} &= \Psi(k) - \Psi\left(1 - \frac{1}{u} - k\right) \\ S_1^{(z)} &= \frac{1}{2\sqrt{\pi}} \frac{g(k)\, g\left(1 - \frac{1}{u} - k\right)}{g\left(1 - \frac{1}{u}\right)} \end{aligned}\right\} \qquad (6.7)$$

Величина

$$g(k) = \frac{\Gamma\left(k + \frac{1}{2}\right)}{\Gamma(k)} \qquad (6.8)$$

может быть вычислена по приближенным формулам:

- при $x > 4$

$$g(x) \approx \sqrt{x}\, e^{-\frac{1}{8x} + \frac{1}{192x^3} - \frac{1}{640x^5} + \ldots}; \qquad (6.9)$$

- при $0<x<4$

$$g(x) = \frac{g(x+n)}{\prod\limits_{i=1}^{n}\left[1 + \frac{1}{2(x+i-1)}\right]}, \qquad (6.10)$$

где $n \geq 4$;

$$g(x+n) = \sqrt{x+n}\, e^{-\frac{1}{8(x+n)} + \frac{1}{192(x+n)^3} - \ldots} \qquad (6.11)$$

Для облегчения расчетов в Приложении 1 приводятся также значения функции $g(x)$.

Для установления типа выравнивающей кривой распределения и нахождения оценок параметров по общему устойчивому методу достаточно найти значения статистических показателей ν_1^*, S_1^*, B^*, H^* и приравнять их соответствующим теоретическим. Эти показатели для каждой системы непрерывных распределений вычисляются по-своему. Но **номограмма применима ко всем трем системам непрерывных распределений.**

Оценки статистических показателей в случае выравнивающих распределений, заданных плотностью $p(x)$, вычисляются по формулам:

$$\left.\begin{aligned} &\nu_1^* = \bar{x} = \sum_{i-1}^{n} x_i p_i h_i \\ &S_1^* = \sum_{i-1}^{n} p_i^2 h_i,\ S_3^* = \sum_{i-1}^{n} p_i^4 h_i \\ &B_1^* = \sum_{i-1}^{n} x_i p_i^2 h_i - \nu_1^* S_1^*;\ \mathrm{H}^* = \frac{S_3^*}{(S_1^*)^3} \end{aligned}\right\} \qquad (6.12)$$

где $p_i = m_i/(Mh_i)$ – эмпирическая плотность распределения; m_i – наблюденная частота случайной величины X в i-ом интервале $(i = 1,2...,n); M = \sum_{i=1}^{n} m_i$ - наблюденная частота во всех n интервалах (объем выборки); h_i – ширина i-го интервала; x_i – значение случайной величины X в середине i-го интервала.

Формулы (6.12) можно выразить через абсолютные частоты m_i:

$$\left.\begin{aligned} &\nu_1^* = \bar{x} = \sum_{i-1}^{n} x_i \frac{m_i}{M} \\ &S_1^* = \sum_{i-1}^{n} \left(\frac{m_i}{M}\right)^2 \frac{1}{h_i}; S_3^* = \sum_{i-1}^{n} \left(\frac{m_i}{M}\right)^4 \frac{1}{h_i^3} \\ &B^* = \sum_{i-1}^{n} x_i \left(\frac{m_i}{M}\right)^2 \frac{1}{h_i} - \nu_1^* S_1^*; H^* = \frac{S_3^*}{(S_1^*)^3} \end{aligned}\right\} \tag{6.13}$$

Показатель островершинности H^* при $h_i = \text{const}$ примет вид

$$H^* = M^2 \frac{\sum_{i=1}^{n} m_i^4}{\left(\sum_{i=1}^{n} m_i^2\right)^3}, \tag{6.14}$$

т**.е. ширина интервала не входит в формулу (6.14**). Отсюда следует вывод, что **ширину интервала группирования статистических данных лучше принимать постоянной** (по крайней мере для распределений, близких к симметричным).

Если выравнивающее распределение задано обобщенной плотностью *p(t),* статистические показатели рассчитываются по формулам:

$$\left.\begin{aligned} &\nu_1^* = \overline{\ln t} = \sum_{i-1}^{n} \ell n t_i \frac{m_i}{M}; \quad S_1^* = \sum_{i-1}^{n} t_i \left(\frac{m_i}{M}\right)^2 \frac{1}{h_i} \\ &S_3^* = \sum_{i-1}^{n} t_i^3 \left(\frac{m_i}{M}\right)^4 \frac{1}{h^3}; \; H^* = \frac{S_3^*}{\left(S_1^*\right)^3} \\ &B^* = \sum_{i-1}^{n} t_i \ln t_i \left(\frac{m_i}{M}\right)^2 \frac{1}{h_i} - \nu_1^* S_1^* \end{aligned}\right\} \tag{6.15}$$

При $h_i = \text{const}$

$$H^* = M^2 \frac{\sum_{i=1}^{n} t_i^3 m_i^4}{\left(\sum_{i=1}^{n} t_i m_i^2\right)^3}. \tag{6.16}$$

Для установления типа выравнивающей кривой и нахождения оценок параметров по общему устойчивому методу автором созданы **программы SNR1,SNR2,SNR3 .**

В заключение отметим, что **общий устойчивый метод основан на взаимосвязи между законами распределения случайных величин *X* и *Z*.**

Запишем обобщенную плотность $p(x)$

$$p(x) = Ne^{k\beta x}\left(1 - \alpha u e^{\beta x}\right)^{\frac{1}{u}-1}.$$

Пусть для определенности параметр $u > 0$.

Введем случайную величину

$$Z = \alpha u e^{\beta X}. \tag{6.17}$$

Тогда плотность $p(z)$ будет равна

$$p(z) = p(x)\frac{dx}{dz}.$$

Поскольку на основании (6.17)

$$x = \frac{1}{\beta}(\ln z - \ln \alpha u),$$

то

$$\frac{dx}{dz} = \frac{1}{\beta z},\ p(z) = \frac{p(x)}{\beta z}, \tag{6.18}$$

откуда имеем замечательное равенство

$$\beta z p(z) = p(x). \tag{6.19}$$

На его базе строится общий устойчивый метод оценивания параметров.

Поскольку плотность $p(z)$ является функцией двух параметров формы $k = \gamma/\beta, u$, то последняя формула позволяет ввести критерии, зависящие от этих двух параметров.

Запишем на основании формулы (6.19) следующее равенство:

$$\beta^r M\left[(zp(z))^r\right] = M\left[p^r(x)\right].$$

Введем обозначения

$$M\left[(zp(z))^r\right] = S_r^{(z)};\ M\left[p^r(x)\right] = S_r.$$

Тогда последнее равенство перепишется в виде

$$\beta^r S_r^{(z)} = S_r. \tag{6.20}$$

Формула (7.5.20) позволяет найти значение параметра β (например, при $r = 1$), **а также получить критерий островершинности**, зависящий от двух параметров *k, u.* Для этого необходимо взять отношение S_2 / S_1^2 либо S_3 / S_1^3. Последнее оказалось наиболее подходящим.

Таким путем был получен показатель островершинности *H*.

Показатель асимметрии В найден из условия, чтобы для симметричных распределений он был равен нулю и в то же время использовал ранее введенные величины. Такой показатель может иметь вид

$B = M[xp(x)] - M(x)M[p(x)]$ или

$$\boldsymbol{B = M[p(x)(x - M(x))]}.$$

Покажем, что он зависит от двух параметров *k, u*.

Поскольку $p(x) = \beta z p(z),\ x = \frac{1}{\beta}(\ln z - \ln \alpha u)$, то

$$B = M\left[\beta z p(z)\left(\frac{1}{\beta}\ln z - \frac{1}{\beta}M(\ln z)\right)\right] = M\left[z p(z)(\ln z - M(\ln z))\right] = f(k,u).$$

По показателям *B, H* строится номограмма, позволяющая устанавливать тип выравнивающей кривой распределения и находить оценки параметров *k, u*. Оценка параметра β вычисляется по величинам S_1 и $S_1^{(z)}$. Оценка параметра α или произведения α*u* вычисляется по тем же формулам, что и в случае универсального метода моментов.

Если в качестве показателей асимметрии и островершинности использовать величины

$$B = F(x_c) - 0{,}5,\ H = \frac{p(x_c)}{M\left[p(x)\right]},$$

где x_c – мода, то можно построить аналогичную номограмму для установления типа выравнивающей кривой распределения и нахождения в первом приближении оценок параметров *k, u* по координатам одной характерной точки *C* и среднему значению плотности *p*(*x*).

Тема 7. Ранговые распределения в библиотечно-информационной деятельности.

7.1 Ранговые распределения

Статистические данные, полученные в результате наблюдения, представляют собой простой статистический ряд. Чтобы извлечь из этого ряда информацию, его упорядочивают либо по возрастанию значений случайной величины, либо по убыванию. В обоих случаях получим вариационный (ранжированный) ряд.

Статистические распределения, в том числе ранговые, широко используются в научных исследованиях. Анализ этих распределений позволяет ученым совершать открытия. Так, в физике была введена постоянная Планка, в химии Д.И.Менделеевым построена Периодическая система элементов, в информатике С.Бредфордом сформулирован закон рассеяния публикаций по периодическим изданиям.

При ранжировании статистических данных открываются возможности извлечения новой информации, изучения структуры выборки, вычисления различных показателей, наиболее полно характеризующих исследуемую случайную величину.

Наличие обобщенных распределений для описания статистических вариационных рядов открывает перед исследователем новые перспективы.

Ранговые распределения находят широкое применение в информатике, математической лингвистике, социологии, библиотечном деле и других отраслях знания.

Рассмотрим, например, частотный словарь. В таком словаре разные слова упорядочены по убыванию (точнее, по невозрастанию) частоты их употребления в текстах, на базе которых построен словарь. Порядковый номер слова и есть его ранг.

В качестве другого примера можно привести ранговое распределение журналов по некоторой отрасли знания (например, по химии и химической технологии), упорядоченных по убыванию числа помещенных в них статей по заданному предмету.

Для описания ранжированных рядов необходимо использовать такие теоретические распределения, которые обладают теми же свойствами, что и ранжированные ряды. Спрашивается, откуда взять распределения, пригодные для выравнивания статистических ранговых распределений? Чтобы решить эту проблему, необходимо либо разработать теорию ранговых распределений, либо использовать ранее построенные обобщенные распределения. Среди множества частных случаев этих распределений найдутся такие, которые с достаточной точностью могут описывать статистические ранговые распределения.

7.2. Форма представления ранговых распределений

Статистическое ранговое распределение можно представить в виде обычной гистограммы, которую можно аппроксимировать непрерывной убывающей кривой распределения. Для большей наглядности статистического рангового распределения строят график зависимости $\ln p_r = f(\ln r)$, где p_r – относительная частота слова частотного словаря с рангом r или доля статей по заданному предмету в журнале с рангом r.

Однако принятая форма представления ранговых распределений несет слишком мало информации о статистическом распределении. На таком графике колебания частот мало заметны, поскольку последние изображены в логарифмическом масштабе. Кроме того, такое преобразование кривой распределения не имеет вероятностного смысла.

В связи с вышесказанным целесообразно перейти к другой форме представления ранговых распределений, а именно, $rp_r = f(lnr)$ [8]. По оси ординат будем откладывать произведение ранга слова (журнала) на его относительную частоту (или долю статей), а по оси абсцисс – натуральный логарифм ранга.

График зависимости $rp_r = f(lnr)$ имеет принципиальные преимущества перед традиционной формой представления ранговых распределений. Во-первых, он описывается первой системой непрерывных распределений (плотностью $p(x)$). В данном случае $rp_r = p(x)$; $lnr = x$. Во-вторых, на такой кривой видны колебания самих частот (по оси ординат), а не их логарифмов. В-третьих, статистические ранговые распределения однородных случайных величин имеют одновершинную кривую распределения (этим свойством обладает обобщенная плотность $p(x)$ при $u<1$). Это позволяет устанавливать однородность или неоднородность статистических ранговых распределений, выделять неоднородную часть, а также решать другие задачи [8].

7.3. Универсальный закон рассеяния публикаций

Глубокое изучение любой дисциплины, допускающей применение количественных методов исследования, должно сопровождаться построением и использованием математических или вероятностно-статистических моделей. Так, в информатике и математической лингвистике широко известны такие математические модели, как закон Дж.Ципфа

$$p_r = \frac{k}{r^{\gamma}}, \tag{7.1}$$

применяемый для описания ранговых распределений слов частотного словаря, а также журналов, упорядоченных по убыванию числа помещенных в них статей по заданному предмету; закон С.Бредфорда рассеяния публикаций; закон старения публикаций и др. К сожалению, каждый из этих законов, как правило, используется сам по себе, без взаимосвязи с другими законами, т.е. без указания его места в ряду других, более общих законов распределения.

Такими общими законами могут служить построенные автором системы непрерывных распределений, поскольку они включают как частные случаи множество известных распределений, в том числе указанные выше.

С помощью обобщенных распределений можно описать практически любое статистическое распределение, если оно представляет собой однородную совокупность значений непрерывной случайной величины.

Так, **первая система** хорошо описывает распределение первоисточников по числу цитирований в зависимости от года издания (закон старения публикаций), а также распределение технологических погрешностей, распределение работников некоторой организации по возрасту.

Вторая система описывает ранговые распределения журналов, упорядоченных по убыванию числа помещенных в них статей по заданному предмету. Из этой же системы выводится математически точная формулировка закона рассеяния публикаций в смысле С.Бредфорда. Она описывает также распределение слов словаря, фраз и предложений по длине, распределение работающих по уровню заработной платы.

Третья система описывает ранговые распределения знаменательных (полнозначных) слов частотного словаря, а также частотных словарей дескрипторов, терминов.

Четвертая система описывает распределение простых чисел.

Закон Ципфа входит как частный случай во вторую и третью системы непрерывных распределений. Закон Вейбулла, который также используется в математической лингвистике и информатике, относится ко второй системе распределений группы А. Из второй системы следуют основные распределения семейства К.Пирсона.

Таким образом, в результате разработки теории обобщенных распределений информатика и математическая лингвистика приобрели мощный математический аппарат, позволяющий решать множество задач на более высоком уровне.

Решим на базе обобщенных распределений наиболее важные проблемы в информатике: дадим математически точную формулировку закона рассеяния публикаций в смысле С.Бредфорда, а также установим в самом общем виде законы рассеяния и старения публикаций.

В 1948г. С.Бредфорд дал окончательную формулировку открытого им в 1934г. закона рассеяния публикаций в периодических изданиях. Приведем формулировку этого закона, позаимствованную из книги [10, с.178,179]: «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 в последующих зонах будут относиться как $1 : n : n^2$».

Несмотря на некоторую неопределенность этой формулировки, С.Бредфорду удалось отразить в ней суть закона рассеяния публикаций, по крайней мере в первом приближении.

Все последующие попытки других исследователей по совершенствованию модели С.Бредфорда оказались безуспешными. И это закономерно, поскольку исследователи строили свои модели в основном на законе Ципфа и предположении о равенстве числа статей в ядре журналов и зонах рассеяния.

Математически точную формулировку закона рассеяния можно дать лишь на базе универсальных распределений, которые с высокой точностью описывают статистические ранговые распределения журналов по различным отраслям знания. Эти распределения для каждого статистического вариационного ряда имеют свои параметры.

Исследования показали [16], что статистические ранговые распределения журналов, упорядоченных по убыванию числа помещенных в них статей по заданному предмету, хорошо аппроксимируются обобщенной плотностью

$$p(t)=Nt^{k\beta-1}\left(1-\alpha u t^{\beta}\right)^{\frac{1}{u}-1} \qquad (7.2)$$

или в более общем случае - второй системой непрерывных распределений, заданной тремя обобщенными плотностями (3.4.18).

Однако убывающая кривая «ранг – относительная частота», т.е. $p_r = f(r)$ не имеет никаких особых точек, которые позволили бы дать математически точную формулировку закона рассеяния публикаций. Поэтому автором введена другая форма представления ранговых распределений, а именно: $rp_r = f(lnr)$. Ранее было показано, что убывающая кривая распределения $p_r = f(r)$ после ее приведения к форме $rp_r = f(lnr)$ в случае однородной выборки превращается в одновершинную кривую, которая описывается плотностью

$$p(x)=Ne^{k\beta x}\left(1-\alpha u e^{k\beta x}\right)^{\frac{1}{u}-1}. \qquad (7.3)$$

Другими словами, такое преобразование распределений второй системы сводит их к распределениям первой системы, т.е. плотность $p(t)$ преобразуется к плотности $p(x)$. Действительно, если умножить обе части плотности (7.2) на величину t и записать выражение t^{β} в виде $e^{\beta \ln t}$, что одно и то же, то из плотности $p(t)$ получим плотность $p(x)$. График этой плотности, т.е. кривая распределения имеет три характерные точки: моду С и две точки перегиба А и В. При этом точки перегиба расположены на равных расстояниях от моды С – и в этом по нашему мнению состоит суть закона рассеяния в толковании Бредфорда! **Примем эти точки в качестве границ ядра и зон рассеяния**.

Итак, для плотности $p(x)$ имеем

$$x_C - x_A = x_B - x_C. \qquad (7.4)$$

Учитывая взаимосвязи между первой и второй системами непрерывных распределений, т.е. $x = lnt$, $p(x) = tp(t)$, для плотности $p(t)$ можем записать

$$lnt_C - lnt_A = lnt_B - lnt_C, \qquad (7.5)$$

откуда имеем равенство

$$\frac{t_C}{t_A}=\frac{t_B}{t_C}=n. \qquad (7.6)$$

Точки А, С, В делят все журналы в ранжированном ряду на четыре части: ядро и три зоны рассеяния. Количество журналов, входящих в ядро, определяется равенством $t_Я = t_A$. Количество журналов в первой зоне $t_I = t_C - t_A$; во второй зоне $t_{II} = t_B - t_C$. Остальные журналы относятся к III зоне: $t_{III} > t_B$.

Теперь можно дать математически точную формулировку закона рассеяния публикаций. Она несколько отличается от формулировки Бредфорда ($t_Я : t_I : t_{II} = 1 : n : n^2$).

Из формулы (7.6) следует, что между количеством наименований журналов **от начала частотного списка до точек А, С, В** имеется соотношение:

$$t_A : t_C : t_B = t_A(1 : n : n^2) \qquad (7.7)$$

В то же время между количеством наименований журналов **в ядре и последующих зонах** имеется другое соотношение (при $(t_Я = t_A)$)

$$t_Я : t_I : t_{II} = t_A(1 : (n-1) : (n-1)n.. \qquad (7.8)$$

Как видим, формулировка Бредфорда является комбинацией из двух точных формул (7.7) и (7.8). При этом из закона Бредфорда неясно, как определяется число журналов, образующих ядро, какая доля статей содержится в нем, сколько может быть зон рассеяния, чему равна величина *n*. Обобщенная плотность *p(t)* дает возможность однозначно ответить на все эти вопросы.

Журналы, входящие в ядро, содержат долю статей, равную функции распределения в точке А, т.е. $F(t_A)$. 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 - $1 - F(t_B)$.

Количество зон рассеяния, как правило, равно трем. Но при определенных значениях параметров аппроксимирующей плотности *p(t)* оно может быть меньше.

На базе плотности *p(t)* нетрудно найти координаты трех характерных точек и вычислить величину *n*. Абсциссы точек *A* и *B* можно рассчитать при известных значениях величин t_C и *n*.

Мода t_C находится из условия $dtp(t)/dlnt = 0$ и в общем случае для распределений I-V типов равна

$$t_C = \left(\frac{k}{\alpha(1 + ku - u)} \right)^{1/\beta}. \qquad (7.9)$$

Величина *n* задается формулой

$$n = \left[1 + \frac{1 - u \mp \sqrt{[4k(1+ku-u) + (1-u)](1-u)}}{2k(1+ku-u)} \right]^{1/\beta}. \qquad (7.10)$$

В формуле (7.10) в числителе знак «минус» относится к распределениям 5-го типа. Поскольку такие ранговые распределения встречаются весьма редко, во многих статьях автора этот знак опущен.

Абсциссы точек перегиба вычисляются по формулам:

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

Рассмотрим один частный случай. Ранговые распределения журналов часто описываются законом Вейбулла с функцией распределения

$$F(t) = 1 - e^{-\alpha t^{\beta}} , \qquad (7.11)$$

которая следует из формулы (2.2.6) при $u \to 0$. Тогда из формул (7.9) и (7.10) при $k = 1$ имеем равенства:

$$t_c = \left(\frac{1}{\alpha}\right)^{\frac{1}{\beta}} , \quad n = \left(\frac{3+\sqrt{5}}{2}\right)^{\frac{1}{\beta}} . \qquad (7.12)$$

При этом значения функции распределения в трех характерных точках независимо от значений параметров равны: $F(t_A) = 0.3175$; $F(t_C) = 0.6321$; $F(t_B) = 0.9271$. Это значит, что в ядро журналов входит 32% от всех статей по данному предмету; в ядро и первую зону рассеяния – 63% статей, а в ядро и первые две зоны – 93% статей. По зонам рассеяния доли статей распределяются так: первая зона рассеяния содержит 31% статей; вторая зона – 30% статей. На третью зону приходится лишь 7% статей. Между числом наименований журналов в ядре и последующих зонах справедливо общее соотношение (7.8).

Отсюда следует, что для более полного удовлетворения информационных потребностей специалистов справочно-информационный фонд должен комплектоваться по крайней мере теми журналами, которые образуют ядро и первые две зоны рассеяния. Количество таких журналов равно t_B, при этом полнота комплектования фонда $F(t_B) = 0.93$ (под полнотой комплектования фонда понимается вероятность удовлетворения запросов потребителей информации этим фондом). Величина t_B может характеризовать некоторый оптимальный объем справочно-информационного фонда с точки зрения полноты его комплектования при ограниченных материальных ресурсах.

Дальнейшие исследования показали [16], что любая обобщенная плотность, входящая во вторую систему непрерывных распределений, дает закон рассеяния в виде формул (7.7), (7.8), **но при этом размеры ядра и зон рассеяния, величина *n*, а также доли статей в ядре и зонах рассеяния различны** (последние у С.Бредфорда одинаковы) и в общем случае зависят как от вида распределения, так и от значений параметров.

Поскольку наиболее полной характеристикой случайной величины является ее закон распределения, в данном случае рангового, то **наиболее общим и универсальным законом рассеяния публикаций является вторая система непрерывных распределений, заданная тремя обобщенными плотностями** (3.13). Именно обобщенные плотности позволяют наиболее точно описывать статистические ранговые распределения журналов, вычислять накопленную долю статей в заданном числе журналов в ранжированном ряду, в том числе в характерных точках A, C, B, вычислять координаты этих точек и величину *n*, входящую в закон рассеяния. Именно обобщенные плотности позволяют дать математически точную формулировку закона рассеяния публикаций в виде формул (7.7), (7.8), справедливых для всех убывающих распределений второй системы. Другими словами, в качестве универсального закона рассеяния публикаций выступает вторая система непрерывных распределений, а закон рассеяния в виде формул (7.7), (7.8) является лишь следствием свойств ранговых распределений, частным случаем универсального закона, отражающим **соотношение между абсциссами характерных точек на кривой распределения. Значения же функции распределения в характерных точках в этих формулах не задействованы**. Поэтому не зная теоретического распределения с его значениями параметров, нельзя вычислить число журналов, входящих в ядро и зоны рассеяния, величину n, а также долю статей в ядре и зонах рассеяния, которая выражается через функцию распределения.

Если же за основу принять формулировку закона рассеяния С.Бредфорда, то нетрудно придти к выводу, что «не существует универсальной математической модели, пригодной для описания распределения публикаций и журналов вне зависимости от их тематической принадлежности» [35, с. 57].

Цитата приведена из Справочника библиографа 2002г. издания. Далее авторы Справочника иллюстрируют сказанное известной таблицей, в которой приводятся ранги журналов, публикующих статьи по химии и химической технологии, и процент статей от общего числа библиографированных.

Но эти табличные данные очень хорошо описываются законом Вейбулла, который следует из обобщенной плотности (7.2) при $k=1,\ u \to 0$, т.е.

$$p(t) = \alpha\beta t^{\beta-1} e^{-\alpha t^{\beta}}. \qquad (7.13)$$

Функция распределения $F(t)$, обозначающая в данном случае суммарную долю библиографированных статей из t первых журналов частотного списка, задается формулой (7.11), при этом параметры закона Вейбулла приближенно равны: $\alpha = 0{,}036$; $\beta = 0{,}53$ [13].

Обобщенная плотность $p(t)$, заданная формулой (7.2), и ее частный случай – распределение Вейбулла (7.13) дают один и тот же закон рассеяния в виде формул (7.7), (7.8), но значения функции распределения в трех характерных точках в обоих случаях различны и зависят от вида распределения и значений параметров.

Таким образом, универсальная математическая модель, о которой говорится в Справочнике, существует и задается обобщенной плотностью $p(t)$, а в общем случае – второй системой непрерывных распределений. Более того, из этих распределений как следствие их свойств вытекают формулы (7.7) и (7.8), которые **уточняют закон рассеяния С.Бредфорда**.

Удовлетворить же требованиям закона С.Бредфорда в виде $t_Я : t_I : t_{II} = 1 : n : n^2$ при условии, что число статей в ядре журналов и зонах рассеяния одинаково, действительно не может никакое теоретическое распределение.

Остается только признать в качестве универсального закона рассеяния публикаций вторую систему непрерывных распределений (по крайней мере ту ее часть, которая хорошо описывает ранговые распределения). Тогда все противоречия снимаются.

Обобщенные распределения позволяют также вычислять число журналов, содержащих заранее определенную долю статей по заданному предмету. Например, в случае распределений группы А I-III типов с функцией распределения

$$F(t) = 1 - \left(1 - \alpha u t^{\beta}\right)^{\frac{1}{u}}, \tag{7.14}$$

имеем

$$t = \left\{ \frac{1}{\alpha u} \left[1 - \left(1 - F(t)\right)^{u}\right] \right\}^{\frac{1}{\beta}}, \tag{7.15}$$

где t – ранг журнала; $F(t)$ – накопленная относительная частота статей по заданному предмету в t журналах.

В частном случае, при $u \to 0$, из формулы (7.14) имеем распределение Вейбулла, из которого находим

$$t = \left(\frac{1}{\alpha} \ln \frac{1}{1 - F(t)} \right)^{\frac{1}{\beta}}. \tag{7.16}$$

В случае распределений группы Б эта задача решается с помощью соответствующих компьютерных программ.

Поскольку вторая система непрерывных распределений хорошо описывает статистические распределения работающих по уровню заработной платы, ее можно также использовать для естественной и объективной классификации работающих с помощью характерных точек с целью оптимизации уровня подоходного налога, который может быть различным для разных зон.

Необходимо также отметить, что вторая система непрерывных распределений, в частности, плотность $p(t)$ хорошо описывает такие статистические распределения, как слов по длине (в словаре), фраз по количеству словоупотреблений, словосочетаний по длине, терминов по длине и др., а также ранговые распределения научных сотрудников по продуктивности [10].

В заключение следует дать обоснование использования трех характерных точек для вывода математически точной формулировки закона рассеяния публикаций в смысле Бредфорда.

Точки А, С, В являются особыми точками кривой распределения, заданной плотностью $p(x)$. Между ординатами трех характерных точек существует соотношение

$$\lambda = \frac{\left[p(x_C)\right]^2}{p(x_A)\,p(x_B)} = \left(\frac{1-2u}{1-u}\right)^{1-\frac{1}{u}} = \left(1 + \frac{1}{1-1/u}\right)^{1-\frac{1}{u}}, \tag{7.17}$$

из которого видно, что показатель λ зависит лишь от одного параметра формы u, т.е. он является идентификатором типа кривой распределения. В зависимости от значений показателя λ распределения, заданные плотностью $p(x)$**,** можно разделить на типы. Для I,I´ типов $e < \lambda < \infty$ (при $0 < u < 1/2$); для II,II´ типов $\lambda = e$; для III типа $2 < \lambda < e$; для IV типа $\lambda = 2$; и наконец для V типа $1 < \lambda < 2$. Таким образом, по ординатам трех характерных точек А, С, В может быть однозначно установлен тип кривой распределения и найдена оценка параметра *u* из соотношения (7.17) при известном λ.

Анализ плотности *p(x)* показывает, что у распределений II-V, II′ 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 $0 < u < 1$. У распределений I′ типа точка перегиба *A* существует при $0 < u < 1/2$, точки *B, C* – при $0 < u < 1$.

Если кривая рангового распределения, приведенная к форме плотности *p(x),* имеет три характерные точки, то для такого распределения существует ядро и три (не более!) зоны рассеяния.

7.4. Универсальный закон старения публикаций

Закон старения публикаций заключается в том, что число ссылок на публикации в зависимости от их года издания вначале резко растет, затем убывает с увеличением срока давности издания. Максимальное число ссылок приходится на публикации одно-двухлетней давности.

Для описания этого закона предлагалось множество математических моделей, но задача так и не была решена (по той же причине, что и в случае закона рассеяния публикаций, т.е. из-за отсутствия подходящего универсального распределения).

Исследования автора показали, что распределение числа ссылок на публикации в зависимости от года их издания хорошо описывается первой системой непрерывных распределений, в частности, обобщенной плотностью *p(x)* [10], где *x* – год издания. Если за начало отсчета принять текущий год ($x = 0$), то для предыдущего года будем иметь $x = -1$ и т.д. Обобщенная плотность распределения *p(x)* обладает тем свойством, что значения случайной величины *X* могут быть как положительными, так и отрицательными.

Таким образом, **наиболее общим и универсальным законом старения публикаций является первая система непрерывных распределений**. Обобщенные плотности позволяют наиболее точно описывать статистические распределения, вычислять накопленную долю ссылок на публикации по любому заданному интервалу времени их издания, вычислять координаты трех характерных точек, как и в случае закона рассеяния, а также вычислять другие показатели, интересующие исследователя.

Абсциссы трех характерных точек для плотности *p(x)* задаются формулами (в случае распределений I-V типов)

$$x_c = \frac{1}{\beta} \ln \frac{k}{\alpha(+ku-u)}, \qquad (7.18)$$

$$x_{A,B} = x_C \mp \ln n, \qquad (7.19)$$

где величина *n* рассчитывается по прежней формуле (7.10).

7.5. Ранговые распределения лексических единиц

В случае однородной совокупности лексических единиц (слов, словосочетаний, терминов, дескрипторов) их ранговые распределения хорошо описываются второй и третьей системами непрерывных распределений [10], которая задана тремя обобщенными плотностями (3.13). Для вычисления типа выравнивающей кривой и оценок ее параметров статистическое распределение необходимо привести к форме плотности *p(t)* либо *p(x)* и воспользоваться соответствующей компьютерной программой.

Характерные точки кривых распределения могут быть использованы как естественные границы различных зон лексических единиц (служебных слов, общеупотребительной лексики, отраслевой, межотраслевой).

В итоге можно сделать вывод, что обобщенные распределения являются универсальными законами распределения не только теории вероятностей и математической статистики, но и информатики, математической лингвистики, экономики и других областей знания. При использовании обобщенных распределений исчезают ранее существовавшие барьеры на пути к новому знанию. Например, для нахождения наилучшей аппроксимирующей кривой **не требуется выдвигать гипотезы о виде закона распределения**. Система непрерывных распределений выбирается в зависимости от свойств случайной величины, а тип распределения и оценки параметров вычисляются по статистическому распределению. При этом вычисленная кривая распределения является наилучшей (разумеется, для принятого метода оценивания параметров). В случае однородности статистической совокупности оба метода – универсальный метод моментов и устойчивый метод – дают очень близкие значения оценок параметров аппроксимирующего распределения. Наиболее точные оценки параметров получаются в случае симметричного или близкого к нему статистического распределения, приведенного к форме плотности *p(x)*.

Универсальные законы старения и рассеяния публикаций, а также ранговые распределения лексических единиц, заданные соответственно плотностями *p(x), p(t), p(y),* являются фундаментальными закономерностями информатики, математической лингвистики, информетрии, библиотековедения, а также менеджмента качества.

Каждый инженер должен владеть методами теории обобщенных распределений, применение которой при контроле качества продукции позволит поднять на новый качественный уровень технологические процессы, но для этого предварительно необходимо включить эту теорию в учебный процесс в первую очередь во всех технических учебных заведениях. Гуманитариям она тоже будет весьма полезна.

Считаю своим долгом предупредить всех, кто пытается в научных исследованиях использовать «закон Ципфа». Этого делать ни в коем случае нельзя, потому что такого закона не существует. Какое бы вы ни взяли статистическое ранговое распределение (например, частотный словарь или журналы, упорядоченные по убыванию числа помещенных

в них статей по некоторой заданной теме, или по убыванию числа обращений к ним) – в системе координат $\mathrm{rp_r} = f(\ln r)$ вы всегда получите колоколообразную кривую с модой (точкой на горизонтальной оси, в которой плотность максимальна) и двумя точками перегиба, которые расположены на равных расстояниях по обе стороны от моды. Эти точки отделяют выпуклую часть кривой распределения от вогнутой и наоборот, вогнутой от выпуклой. Координаты этих точек используются мною в качестве естественных границ ядра и зон рассеяния. По закону же Ципфа должна получиться горизонтальная прямая, на которой никаких характерных точек нет. Использование теории обобщенных распределений автора позволит вам решить практически любые задачи: это и вычисление закона распределения, его прогнозирование, анализ и прогнозирование временных рядов, вычисление дискретных распределений и связанных с ними кривых роста, прогнозирование дискретных распределений на выборку заданного объема и т.д.

Тема 8. Построение системы дискретных распределений

8.1. Построение системы дискретных распределений по кривым роста новых событий.

Моделирование кривой роста и статистической структуры словаря ключевых слов

Для аппроксимации статистических зависимостей между количеством произведенных испытаний и количеством наступивших разных событий автором разработана система кривых роста, заданная двухпараметрической формулой [18]

$$y=\frac{1}{\alpha u}\left[1-\left(1-\alpha(u-1)x\right)^{\frac{u}{u-1}}\right], \qquad (8.1)$$

где у – количество наступивших разных событий (разных слов, в том числе ключевых, наименований книг, запросов и т.д.); х – количество произведенных испытаний (объем выборки в словоупотреблениях, число книговыдач, число абоненто-запросов и т.д.).

Последняя формула включает систему кривых роста, которые можно разделить на типы в зависимости от значений параметра u.

При u > 0 имеем кривые роста I типа. Они задаются формулой (8.1). В частности, при $u \to 1$ из (8.1) следует формула

$$y=\frac{1}{\alpha}\left(1-\frac{1}{e^{\alpha x}}\right). \qquad (8.2)$$

При $u \to 0$ из (1) следует кривая II типа

$$y=\frac{1}{\alpha}\ln(1+\alpha x). \qquad (8.3)$$

При u < 0 имеем кривую III типа. Она задается той же формулой (8.1).

Между кривой роста разных событий и статистической структурой выборки существует взаимосвязь, установленная В.М. Калининым [3]

$$y_m=(-1)^{m+1}\frac{x^m}{m!}\frac{d^m y}{dx^m}. \qquad (8.4)$$

По формуле (4) можно рассчитать частотный спектр, или статистическую структуру выборки, т.е. количество событий с частотой появления 1, 2,…, m раз, если задана кривая роста разных событий y = f(x).

Формулы (8.1) и (8.4) позволяют также построить систему дискретных распределений.

Построение системы дискретных распределений

Распределения I типа (u > 0).

Продифференцируем выражение (8.1) m раз по x и подставим m-ю производную в (8.4). В результате получим формулу, позволяющую вычислять число событий с частотой m, т.е. y_m при числе испытаний x

$$y_m = \frac{y_{m=0}}{m!}\left(\frac{\alpha u x}{1+\alpha(1-u)x}\right)^m \prod_{i=0}^{m-1}\left[1+i\left(\frac{1}{u}-1\right)\right], \quad m=1,2,\ldots, \tag{8.5}$$

где

$$y_{m=0} = \frac{1}{\alpha u}\left[1+\alpha(1-u)x\right]^{\frac{u}{u-1}}. \tag{8.6}$$

В данном случае число разных событий, наступающих при x испытаниях, ограничено: $0<y<1/\alpha u$, причем, $1/\alpha u = n$ (величина n – это число разных событий, составляющих полную группу; сумма вероятностей этих событий равна единице).

Разделив величину y_m на n, получим выражение для вероятности наступления событий ровно m раз при x испытаниях: $p_m = y_m / n$ (при этом удобно разделить на n величину $y_{m=0}$):

$$p_m = \frac{p_{m=0}}{m!}\left(\frac{\alpha u x}{1+\alpha(1-u)x}\right)^m \prod_{i=0}^{m-1}\left[1+i\left(\frac{1}{u}-1\right)\right], \quad m=1,2,\ldots, \tag{8.7}$$

где

$$p_{m=0} = \left[1+\alpha(1-u)x\right]^{\frac{u}{u-1}}. \tag{8.8}$$

Исследования показали, что частными случаями распределения I типа (7) являются: биномиальное – при u > 1; Пуассона – при u → 1; отрицательное биномиальное – при 0 < u < 1 (в том числе геометрическое распределение – при u = ½).

Распределения II типа (u → 0). В данном случае кривая роста разных событий задается формулой (8.3), на основании которой и формулы В.М. Калинина (8.4) имеем

$$y_m = \left(\frac{\alpha x}{1+\alpha x}\right)^m \frac{1}{\alpha m}, \quad m=1,2,\ldots \tag{8.9}$$

Разделив (8.9) на (8.3), получим

$$\frac{y_m}{y} = p_m = \left(\frac{\alpha x}{1+\alpha x}\right)^m \frac{1}{m\ln(1+\alpha x)} \qquad (8.10)$$

Последнее распределение известно как распределение Фишера по логарифмическому ряду и находит широкое применение в биологии [10].

Распределения III типа (- ∞ < u < ∞). Кривая роста разных событий задается общей формулой (8.1). Из (8.1) и (8.4) имеем

$$y_m = \frac{y_{m=1}}{m!}\left(\frac{-\alpha u x}{1+\alpha(1-u)x}\right)^{m-1}\prod_{i=1}^{m-1}\left[i\left(1-\frac{1}{u}\right)-1\right], \quad m=2,3,..., \qquad (8.11)$$

где

$$y_{m=1} = x\left[1+\alpha(1-u)x\right]^{\frac{1}{u-1}}. \qquad (8.12)$$

Разделив y_m на y, получим выражение для вероятности p_m

$$p_m = \frac{p_{m=1}}{m!}\left(\frac{-\alpha u x}{1+\alpha(1-u)x}\right)^{m-1}\prod_{i=1}^{m-1}\left[i\left(1-\frac{1}{u}\right)-1\right], \quad m=2,3,..., \qquad (8.13)$$

где

$$p_{m=1} = \frac{-\alpha u x}{(1+\alpha(1-u)x)\left[1-(1+\alpha(1-u)x)^{\frac{u}{1-u}}\right]}. \qquad (8.14)$$

Оценивание параметров дискретных распределений

Для установления типа аппроксимирующего дискретного распределения введем критерий [10]

$$HD = \frac{\frac{x}{y}\ln\frac{x}{y_{m=1}}}{\frac{x}{y_{m=1}}-1}. \qquad (8.15)$$

При HD = 1 выравнивающее распределение относится ко II-му типу. При HD < 1 – к I-му типу. При HD > 1 – к III-му типу.

Далее рассчитываются оценки параметров α, u. Их можно найти по методу моментов. В случае распределений I типа

$$\alpha = \sum_{m\ge 1}\left(\frac{m}{x}\right)^2 y_m - \frac{1}{x},$$

(8.16)

$$u = \frac{1}{\alpha n} \ , \qquad (8.17)$$

где $x = \sum_{m \geq 1} m y_m , \quad n = \sum_{m \geq 0} y_m .$

В случае распределений II типа оценка единственного параметра α находится методом простых итераций по формуле, которая следует из (8.3)

$$\alpha_{i+1} = \frac{1}{y} \ln (1 + \alpha_i x) \qquad (8.18)$$

где $y = \sum_{m \geq 1} y_m$; α_i - значение параметра α на предыдущем шаге итерации. В качестве первого приближения можно принять $\alpha_1 = 1/y_{m=1}$.

В случае распределений III типа (а также I типа) оценка параметра u может быть найдена методом итераций по формуле

$$u_{i+1} = (1 - u_i) \frac{x}{y} \frac{1 - \left(\frac{y_{m=1}}{x} \right)^{u_i}}{\left(\frac{x}{y_{m=1}} \right)^{1-u_i} - 1} . \qquad (8.19)$$

Тогда оценка параметра α равна

$$\alpha = \frac{1}{u y} \left[1 - \left(\frac{y_{m=1}}{x} \right)^{u} \right] = \frac{1}{(1-u) x} \left[\left(\frac{x}{y_{m=1}} \right)^{1-u} - 1 \right] . \qquad (8.20)$$

Таким образом, для оценивания параметров α, u достаточно знать три величины: x, y, $y_{m=1}$.

Отметим, что формулы (8.16), (8.19), (8.20) справедливы для распределений трех типов.

При известных оценках параметров α, u вычисляются теоретические значения y_m и сравниваются со статистическими. Расчет осуществляется по рекуррентной формуле

$$y_{m+1} = y_m \frac{\alpha x [1 + m (1 - u)]}{[1 + \alpha (1 - u) x] (m + 1)}, \qquad (8.21)$$

которая справедлива для распределений всех трех типов. Вначале по формуле (8.12) вычисляется количество событий с частотой m = 1, т.е. $y_{m=1}$. Далее по формуле (8.21) последовательно находятся значения y_{m+1} при m = 1, 2 и т.д. В случае распределений I типа дополнительно вычисляется величина $y_{m=0}$ по формуле (8.6).

Кривая роста и статистическая структура словаря ключевых слов

Построенная система дискретных распределений, взаимосвязанная с системой кривых роста разных событий, позволяет легко решать многие задачи. Рассмотрим пример.

За некоторое время эксплуатации БелРАСНТИ при индексировании документов по автомобильному транспорту было употреблено $y = 3786$ разных ключевых слов при общей их частоте употребления $x = 147644$. Количество ключевых слов с частотой $m = 1$ составило $y_{m=1} = 1518$. По этим трем величинам требуется рассчитать:

- тип аппроксимирующего дискретного закона распределения;
- оценки параметров u, α дискретного распределения и кривой роста разных ключевых слов;
- кривую роста разных ключевых слов.
- частотный спектр ключевых слов, т.е. количество ключевых слов с частотой употребления 1, 2, … , m раз.

Установим тип аппроксимирующего дискретного распределения. Для этого вычислим по формуле (15) критерий HD. Он оказался равным 1,854.

Поскольку HD>1, то искомое распределение относится к III типу. Далее по формулам (8.19), (8.20) вычисляем оценки параметров u, α: u = -0,601736; α = 0,00645759. Частотный спектр описывается формулой (8.21).

Кривая роста разных ключевых слов описывается уравнением (1), которое при найденных оценках параметров u, α примет вид

$$y = 257.35\left[(1+0.0103435x)^{0.375677} - 1\right] \qquad (8.22)$$

Рассчитанные по формуле (8.22) значения y приведены в таблице 8.1, графа 3. В той же таблице в графе 2 даны значения y, восстановленные по частотному спектру ключевых слов с помощью формулы В.М.Калинина [3]

$$y = y_0 - \sum_{m\geq 1}\left(1-\frac{x}{x_0}\right)^m y_m, \qquad (8.23)$$

где x, y – текущие значения объемов выборки и словаря ($x < x_0$; $y < y_0$; $x_0 = 147644$; $y_0 = 3786$). В таблице приводится также расчетное количество ключевых слов с частотой употребления один и два раза. Все расчеты выполнены по программе автора SDR99.

Таблица 8.1.

Зависимость количества разных ключевых слов y от объема выборки x

Объем выборки x	Количество разных ключевых слов y		Количество ключевых слов с частотой	
	По частотному спектру	По формуле (8.22)	Один раз $Y_{m=1}$	Два раза $Y_{m=2}$
1	2	3	4	5
10000	1222	1218	549	170
20000	1671	1654	714	222
30000	1988	1967	833	259
40000	2240	2220	928	289
50000	2454	2436	1010	315
80000	2963	2955	1205	376
100000	3239	3236	1311	409
147644	3786	3786	1518	474
200000		4274	1702	531
500000		6135	2401	749
1000000		8037	3116	972
1500000		9401	3628	1133
2000000		10503	4042	1262

Анализ данных таблицы 8.1 показывает, что максимальная относительная ошибка формулы (8.22) составила около 1%, и это при условии, когда оценки параметров u, α вычислялись всего лишь по трем величинам: x, y, $y_{m=1}$.

Формула (8.1) позволяет прогнозировать рост словаря ключевых слов в зависимости от количества заиндексированных документов D, а также полноту словаря, что важно знать при ведении информационно-поискового тезауруса. Для этого достаточно в формуле (8.1) заменить величину x на произведение Dh, где h – глубина индексирования. Ее можно оценить количеством ключевых слов, приходящихся в среднем на один поисковый образ документа.

Полноту словаря ключевых слов будем измерять вероятностью непоявления нового ключевого слова в точке с координатами (x;y) кривой роста, т.е. функцией распределения вероятностей новых событий $\overline{F}(y)$, при этом $\overline{F}(y)=\overline{F}(x)$. Новым будем считать любое слово при первом его употреблении для индексирования документа.

Полнота словаря рассчитывается по формуле [10]

$$\overline{F}(y)=\overline{F}(x)=1-\frac{dy}{dx}, \qquad (8.24)$$

где первая производная dy/dx равна вероятности появления нового слова.

Дифференцируя выражение (8.1) и подставляя первую производную в (8.24), получим

$$\overline{F}(y)=1-(1-\alpha u y)^{\frac{1}{u}},$$

$$\overline{F}(x)=1-(1-\alpha(u-1)x)^{\frac{1}{u-1}}.$$

Последние две формулы позволяют вычислять значения величин y, x, при которых будет достигнута заданная полнота $\overline{F}(y)=\overline{F}(x)$:

$$y=\frac{1}{\alpha u}\left[1-(1-\overline{F}(y))^{u}\right],$$

$$x=\frac{1}{\alpha(u-1)}\left[1-(1-\overline{F}(x))^{u-1}\right].$$

Полноту словаря объемом y можно также выразить через число ключевых слов с частотой употребления один раз [10]

$$\overline{F}(y)=1-\frac{y_{m=1}}{x},$$

где $y_{m=1}$ можно взять из частотного словаря ключевых слов. Эта формула следует из (8.24) и (8.4) при m = 1.

В таблице 8.2 приведены частоты употребления и количество ключевых слов с указанной частотой. Статистические и расчетные данные, вычисленные по программе SDR99, достаточно близки между собой.

Система дискретных распределений, взаимосвязанная с системой кривых роста разных событий, может быть использована во всех тех случаях, когда речь идет о последовательности независимых испытаний и частота появления разных событий подчиняется одному из дискретных законов, описанных в настоящей работе.

Использование системы непрерывных распределений [4] наряду с системой дискретных распределений, а также кривых роста и компьютерных программ, т.е. использование теории обобщенных распределений в целом позволяет описать все многообразие статистических распределений и кривых роста, которые встречаются в библиотечно-информационной деятельности.

С помощью математико-статистических моделей, наиболее точно аппроксимирующих статистические закономерности, из библиотечной (и любой другой) статистики может быть извлечена наиболее полная, объективная и ценная информация. При этом теория требует наличия определенных статистических данных. Так, при статистическом учете количества книговыдач, количества абоненто-запросов и т.д. совершенно необходимо вести учет количества разных наименований выданных книг, разных запросов и т.д. Только в этом

Таблица 8.2 Количество ключевых слов y_m с заданной частотой m

Частота m	Количество слов по факту		Количество слов по расчету	
	Y_m	Сумма Y_m	Y_m	Сумма Y_m
1	2	3	4	5
1	1518	1518	1518	1518
2	450	1968	473,6	1991,6
3	229	2197	256,2	2247,8
4	144	2341	168,0	2415,8
5	136	2477	121,7	2537,5
6	119	2596	93,7	2631,2
7	77	2673	75,3	2706,5
8	66	2739	62,3	2768,8
9	72	2811	52,7	2821,5
10	55	2866	45,4	2866,9
11	47	2913	39,7	2906,6
12	38	2951	35,2	2941,8
13	41	2992	31,4	2973,2
14	27	3019	28,3	3001,5
15	25	3044	25,7	3027,2
16 и >	742	3786	758,8	3786

случае может быть построена статистическая кривая роста разных событий. Анализ такой кривой дает объективную информацию, необходимую при решении различных задач. Это оптимизация комплектования фонда, оценка его полноты, анализ использования, оценка состояния и прогнозирование.

Использование системы непрерывных распределений позволяет вычислять наилучшее аппроксимирующее распределение, в том числе ранговое, находить универсальные законы рассеяния и старения публикаций. На базе универсального закона рассеяния можно дать математически точную формулировку закона Бредфорда, вычислить границы ядра и зон рассеяния, доли статей в каждой зоне.

По всем разделам теории обобщенных распределений автором созданы компьютерные программы, которые апробированы на большом статистическом материале в течение длительного времени – с 1990г.

Использование теории обобщенных распределений гарантирует высокую экономическую эффективность статистических методов во всех практических приложениях, в том числе в библиотечно-информационной деятельности, в системах управления качеством, в научных исследованиях.

Возьмем из табл. 8.2 расчетные данные о частоте ключевых слов и их количестве, вычисленном по дискретному распределению 3-го типа автора. По этим данным построим график зависимости $LNYm = f(LNm)$.

Из построенного графика видно, что эта зависиттмость близка к уравнению прямой, т.е. фактически имеем **закон Лотки.** Но теоретическая указанная зависимость имеет точку перегиба. И, следовательно, в окрестности этой точки действительно можно провести прямую, но только на некотором ограниченном с двух сторон от этой точки интервале. При увеличении частоты m точки все дальше рассеиваются от указанной прямой. и не ложатся на прямую. В доказательсто этого факта приведем еще один график (рис 8.2) из учебного пособия автора[9]. В этом случае статистические данные полные – учтены словосочетания с частотами от1 до 107.

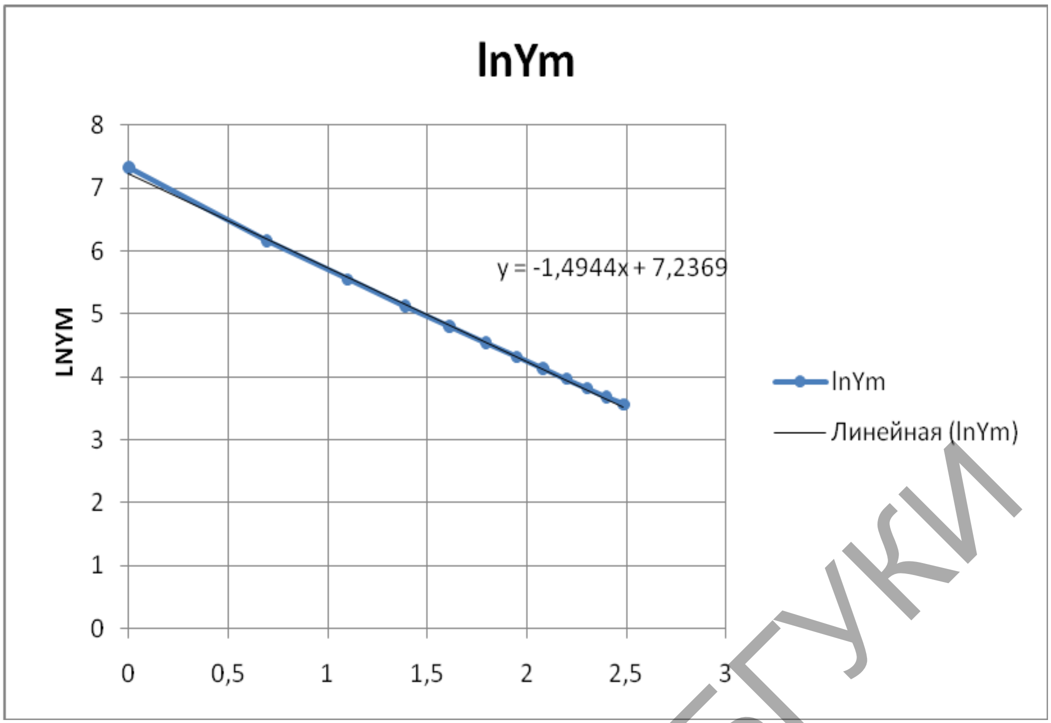


Рисунок 8.1 Диаграмма Лотки $\mathrm{Ym} = \dfrac{\mathrm{Y}_1}{m^{1,4944}}$; $\mathrm{LNYm} = \mathrm{LNY}_1 - 1{,}4944\ LNm$

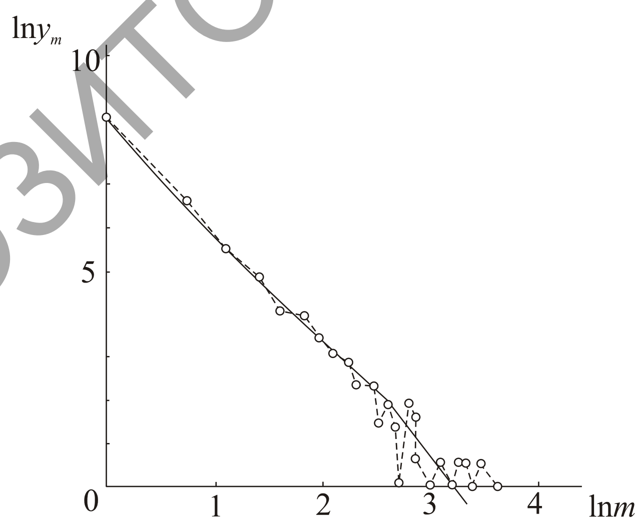


Рис. 8.2. Распределение словосочетаний в английском газетном тексте по дискретному распределению 3-го типа

ВОПРОСЫ К ЗАЧЁТУ

Раздел I. Кривая роста новых событий и ее исследование

1. Понятие математического ожидания случайной функции, нового события и кривой роста новых событий.
2. Связь кривой роста с законами распределения вероятностей разных и новых событий.
3. Формула В.М.Калинина для расчёта статистической структуры выборки по кривой роста новых событий.
4. Формула В.М.Калинина для восстановления кривой роста новых событий по статистической структуре выборки.
5. Порядок построения системы кривых роста и непрерывных распределений новых событий.

Раздел II.Системы непрерывных распределений

6. Методы построения универсальных (обобщённых) непрерывных распределений.
7. Семейство кривых К.Пирсона.
8. Три системы непрерывных распределений В.Нешитого.
9. Распределения групп А и Б.
10. Классификация распределений.
11. Ранговые распределения. Закон Ципфа в семействе ранговых распределений.
12. Характерные точки кривых распределения и связь их с законами рассеяния публикаций.
13. Методы оценивания параметров: метод моментов.
14. Метод наибольшего правдоподобия.
15. Метод наименьших квадратов.
16. Универсальный метод моментов) (В.Нешитого).
17. Общий устойчивый метод (В.Нешитого).
18. Применение системы непрерывных распределений в библиотечно-информационной деятельности и лингвистике.
19. Универсальный закон рассеяния публикаций.
20. Универсальный закон старения публикаций.

Раздел III. Система дискретных распределений

21. Методы построения системы дискретных распределений.
22. Классификация дискретных распределений.
23. Порядок вычисления по статистическим данным дискретного закона распределения и оценок параметров.
24. Критерий степени неравномерности появления событий.
25. Прогнозирование кривой роста новых событий и частотного спектра.
26. Расчёт достоверной части частотного словаря.

СПИСОК ЛИТЕРАТУРЫ

1.Белоногов, Г. Г. О некоторых статистических закономерностях в русской письменной речи. /Г.Г. Белоногов // Вопросы языкознания. –1962. – №1. – С.100-101.

2. Вентцель, Е.С. Теория вероятностей / Е. С. Вентцель. –М.: Наука, 1969 – 576 с.

3. Герасимович, А. И., Матвеева, Я. И. – Математическая статистика / А.И. Герасимович, Я.И. Матвеева. – Минск : Вышэйшая школа. 1978. – 200 с.

4. Гмурман, В.Е. Теория вероятностей и математическая статистика / В.Е.Гмурман – М.: Высшая школа, 1977. –479 с.

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

6. Михайлов, А.И. Научные коммуникации и информатика / А. И. Михайлов, А. И. Черный, Р. С. Гиляревский. – М.: Наука, 1976. – 436 с.

7. Мицевич, Т.А. Исследование структуры потоков научно-технической информации по машиностроению / Т.А. Мицевич // НТИ. Серия 2. – 1975. - №5. – С. 3-16.

8. Нешитой, В.В. Форма представления ранговых распределений / В.В.Нешитой // Ученые записки Тартуского гос. ун-та. – 1987. – Вып. 774. – С.123 -134.

9. Нешитой, В.В. Исследование статистических закономерностей текста и информационных потоков: дис…докт.техн. наук. Мн., 1987. – 505 с.

10.Нешитой, В.В. Методы статистического анализа на базе обобщенных распределений: учеб-метод. пособие / В.В. Нешитой. – Мн.: Веды, 2001. – 168 с.

11.Нешитой, В.В. Универсальные законы рассеяния и старения публикаций / В.В. Нешитой // Веснік Бел. дзярж. ун-та культ. і маст. – 2007. - №8. – С. 128-133

12.Нешитой, В.В.Моделирование кривой роста и статистической структуры словаря ключевых слов / В.В.Нешитой // Веснік Беларус. дзярж. ун-та культ. і мастацтв. – 2008. – №9. – С. 123-132.

13.Нешитой, В.В. Статистическое моделирование библиотечного фонда/ В.В.Нешитой // НТБ – Москва, 2009. – С.36-46.

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

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

17.Нешитой, В.В. Методы статанализа в библиотечной деятель- ности: вычисление непрерывных распределений : учеб: метод. пособие / В. В. Нешитой. – Минск: Бел. гос. ун-т культуры и искусств, 2010. –61 с.

18.Нешитой, В. В. Методы статанализа в библиотечно-информационной деятельности: вычисление дискретных распределений и кривых роста : учеб.-метод. пособие / В. В. Нешитой. – Минск: РИВШ, 2012. –134 с.

19.Нешитой, В. В. Статистические методы анализа использования библиотечного фонда / В. В. Нешитой, Б В. Петренко – Вестник Библиотечной Ассамблеи Евразии РГБ. – Москва, 2013. №2 – С.84-86.

Петренко, Б. В. Применение закона Вейбулла для расчета полноты комплектования справочно-информационного фонда / Б. В. Петренко, В. В. Нешитой // Проблемы оптимального комплектования и использования справочно-информационного фонда для принятия решений / Общество «Знание» Украинской ССР. – Киев, 1974. – С. 6-8.

21.Поллард Дж. Справочник по вычислительным методам статистики / Дж. Поллард; пер. с англ. – М.: Финансы и статистика, 1982. –344 с.

22.Хайтун, С. Д. Наукометрия. Состояние и перспективы/ С. Д. Хайтун.-М: Наука, 1983.-344 с.

23.Bradford, S. C. Documentation.-London, 1948.-156 p.

Brookes, B.C. The Derivation and Application of the Bradford-Zipf Distribution // Journal of Documentation. - 1968. - v.24. - N4. - P.247-265.

25.Brookes, B.C. Bradford's law and the bibliography of science // Nature. - 1969. - N9. - P. 953-956.

26.Vickeri, B C. Bradford's law of scattering // Journal of Documentation. – 1948. – V.4. – N3. – P. 198-203.

27.Zipf, G.K. Human behaviour and the principle of least offort. - Cambridge,1949.

III. Самостоятельная работа студентов
(подготовка к практическим занятиям и зачету)

Методичесие указания по выполнению практических заданий

К теме 1. Функции одной переменной. Производная функции.

Изучить тему по конспекту или школьному учебнику «Алгебра и начала анализа». Можно воспользоваться услугами Интернета, рассмотреть примеры, решенные в аудитории. Но главное – выучить таблицу производных и правила дифференцирования сложных функций.

К теме 2. Некоторые понятия теории вероятностей и математической статистики

См. конспект и список литературы №№ 1–3.(основная литература).

К теме 3. Методы построения обобщенных непрерывных распределений См. учеб.-метод пособие, №5 с. 48-61.

Изучить тему по конспекту и учебно-методическому пособию,

электронный вариат которого можно взять у автора теории обобщенных распределений НВВ.

К теме 4. Классические методы оценивания параметров непрерывных распределений

См. учеб.-метод пособие, №5 с. 69-84. рекомендации те же.

К теме 5. Универсальный метод моментов вычисления закона распределенитя и оценок параметров.

См. учеб.-метод пособие, №7 с. 87-96.рекомендации те же.

К теме 6. Общий устойчивый метод вычисления закона распределения и оценок параметров.

См. учеб.-метод пособие,№5 с. 110-117.рекомендации те же.

К теме 7. Ранговые распределения в библиотечно-информационной деятельности.

См. учеб.-метод пособие №5 БГУКИ с. 154-166, а также статью

Нешитой В.В. Форма представления ранговых распределений // Учёные записки Тартуского гос. ун-та. – 1987. – Вып. 774. – С. 123 – 134. Можно обратиться к репозиторию БГУКИ и внимательно просмотреть статьи Нешитого.

К теме 8. Построение системы дискретных распределений, оценивание параметров дискретных распределений.

См. в списке литературы № 9 РИВШ С. 25 -32; 46 – 54, а также

Нешитой В.В. Статистическое моделирование библиотечного фонда / В.В.Нешитой // НТБ – Москва, 2009. – №12. – С., а также в репозитории БГУКИ статья В.Нешитого:

Моделирование кривой роста и статистической структуры словаря ключевых слов.

Контроль самостоятельной работы студентов (СРС)

Эту работу обязательно выполнить всем студентам ФИДК с помощью табличного процессора EXCEL.

Применение систем непрерывных распределений в библиотечно-информационной деятельности.

Ранговое распределение журналов задано таблицей

Ранг журнала *t*	Доля статей в *t* журналах
50	0,33
100	0,46
500	0,81
1000	0,92

В предположении, что статистическое распределение описывается законом Вейбулла, вычислить по методу наименьших квадратов оценки параметров α и β.

Вычислить координаты характерных точек *A, C, B*.

Вычислить по закону Вейбулла долю статей, содержащихся в ядре журналов, в первой, второй и третьей зонах рассеяния, а также количество журналов в ядре и зонах рассеяния.

Методические указания. Для вычисления оценок параметров α и β закона Вейбулла преобразуем функцию распределения Вейбулла

$$F(t) = 1 - \frac{1}{e^{\alpha t^{\beta}}}$$

к линейному виду

$$\ln \ln \frac{1}{1 - F(t)} = \ln \alpha + \beta \ln t$$

или

$Y = A + \beta X$,

где

$$Y = \ln\ln\frac{1}{1-F(t)}; \quad A = \ln\alpha; \quad X = \ln t.$$

Оценки параметров *β, A* вычисляются по формулам

$$\beta = \frac{\overline{XY} - \overline{X}\cdot\overline{Y}}{\overline{X^2} - (\overline{X})^2}; \quad A = \overline{Y} - \beta\overline{X}.$$

Оценка параметра α равна

$$\alpha = e^A.$$

Для вычисления координат трех характерных точек используются формулы, справедливые в случае закона Вейбулла

$$t_C = \left(\frac{1}{\alpha}\right)^{\frac{1}{\beta}}; \quad n = \left(\frac{3+\sqrt{5}}{2}\right)^{\frac{1}{\beta}}; \quad t_A = \frac{t_C}{n}; \quad t_B = t_C \cdot n; \quad F(t_A) = 0.3175; \quad F(t_C) = 0.6321; \quad F(t_B) = 0.9271$$

Рейтинговый регламент

по курсу «Математико-статистические методы библиотечно-информационной деятельности» для студентов ФИДК 2-го курса

За каждое посещение лекций и практических занитий студенту начисляется 5 баллов, а за каждый пропуск без уважительной причины – Снимается 5баллов.

1. Посещение лекций 5*18 =90
2. Посещение практических занятий 5*10 = 50
3. Неподготовленность к практическим занятиям – минус 5 баллов

Всего по курсу баллов 140

К зачету допускаются студенты, набравшие 80% баллов от их общего количества,т.е 140*0.8 =90 баллов.

Студенты, пропустившие занятия без уважительной причины, на зачете должны показать конспект лекции с восстановленной пропущенной темой, а пропустившие практическое занятие – должны самостоятельно выполнить практическое задание.

На зачете обязательно предъявлять выполненную самостоятельную работу.

Без ее предъявления студент не допускается к зачету.

ПРИЛОЖЕНИЯ

Приложение 1

Таблица значений функций

$\Gamma(x)$, $\Psi(x) = d\ln(\Gamma(x))/dx$, $\Psi'(x)$, $g(x) = \Gamma(x+0.5)/\Gamma(x)$

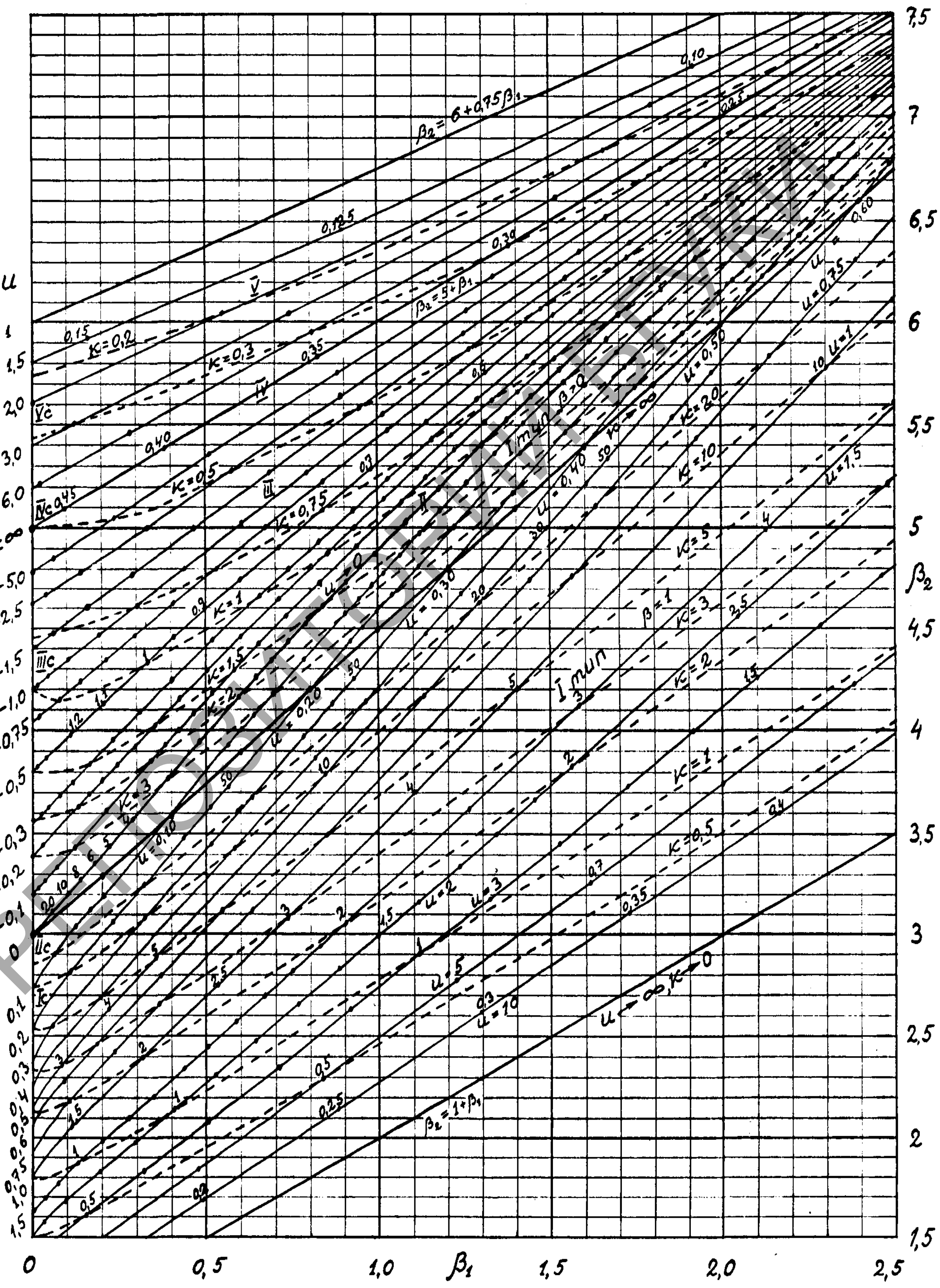
x	**Γ(x)**	**Ψ(x)**	**Ψ´(x)**	**g(x)**
0.01	99.43246	- 100.5609	10001.62	0.01748
0.02	49.44216	- 50.54479	2501.598	0.0345
0.03	32.78495	- 33.86225	1112.687	0.05108
0.04	24.46092	- 25.51327	626.5537	0.06724
0.05	19.47007	- 20.49785	401.5324	0.08301
0.06	16.14571	- 17.14929	279.2893	0.09839
0.07	13.77357	- 14.75333	205.5729	0.11342
0.08	11.99654	- 12.9528	157.7215	0.12811
0.09	10.6162	- 11.54929	124.909	0.14248
0.10	9.51349	- 10.42375	101.4333	0.15653
0.11	8.61267	- 9.500419	84.05954	0.17029
0.12	7.86323	- 8.72879	70.8414	0.18377
0.13	7.23023	- 8.07388	60.55102	0.19698
0.14	6.68867	- 7.51072	52.38271	0.20993
0.15	6.22027	- 7.02099	45.79	0.22263
0.16	5.81126	- 6.59095	40.39171	0.23509
0.17	5.45116	- 6.21009	35.9153	0.24732
0.18	5.13181	- 5.87024	32.1618	0.25934
0.19	4.84676	- 5.56494	28.98315	0.27115
0.20	4.59084	- 5.28904	26.26738	0.28275
0.21	4.35988	-5.03834	23.92849	0.29416
0.22	4.15048	-4.80944	21.89961	0.30538
0.23	3.9598	-4.59949	20.12804	0.31642
0.24	3.7855	-4.40616	18.57186	0.32729
0.25	3.62561	-4.22745	17.19733	0.33799
0.26	3.47845	-4.0617	15.97709	0.34853
0.27	3.3426	-3.90747	14.88874	0.35891
0.28	3.21685	-3.76355	13.91381	0.36914
0.29	3.10014	-3.62887	13.03697	0.37923
0.30	2.99157	-3.50252	12.24537	0.38917
0.31	2.89033	-3.38371	11.52821	0.39898
0.32	2.79575	-3.27174	10.87638	0.40865
0.33	2.7072	-3.16599	10.28209	0.4182
0.34	2.62416	-3.06593	9.73868	0.42762
0.35	2.54614	-2.97107	9.240461	0.43693
0.36	2.47273	-2.88099	8.78248	0.44612
0.37	2.40355	-2.7953	8.36047	0.45519
0.38	2.33825	-2.71367	7.97072	0.46415
0.39	2.27655	-2.63579	7.60996	0.47301
0.40	2.21816	-2.56138	7.27536	0.48176

0.41	2.16284	-2.4902	6.9644	0.49042
0.42	2.11037	-2.42202	6.67487	0.49897
0.43	2.06055	-2.35664	6.40481	0.50743
0.44	2.01319	-2.29387	6.15249	0.5158
0.45	1.96813	-2.23354	5.91635	0.52408
0.46	1.92523	-2.17549	5.69502	0.53227
0.47	1.88432	-2.11959	5.48725	0.54037
0.48	1.8453	-2.06571	5.29194	0.54839
0.49	1.80805	-2.01372	5.10809	0.55633
0.50	1.77245	-1.96351	4.9348	0.56419
0.51	1.73841	-1.91499	4.77126	0.57197
0.52	1.70584	-1.86805	4.61673	0.57968
0.53	1.67465	-1.82262	4.47054	0.58731
0.54	1.64477	-1.77862	4.3321	0.59488
0.55	1.61612	-1.73596	4.20084	0.60237
0.56	1.58864	-1.69458	4.07627	0.60979
0.57	1.56226	-1.65441	3.95792	0.61715
0.58	1.53693	-1.6154	3.84538	0.62444
0.59	1.51259	-1.57749	3.73826	0.63167
0.60	1.48919	-1.54062	3.63621	0.63884
0.61	1.46669	-1.50475	3.5389	0.64594
0.62	1.44504	-1.46983	3.44604	0.65299
0.63	1.4242	-1.43581	3.35734	0.65997
0.64	1.40413	-1.40267	3.27256	0.6669
0.65	1.38479	-1.37035	3.19146	0.67378
0.66	1.36616	-1.33883	3.11381	0.68059
0.67	1.3482	-1.30806	3.03943	0.68736
0.68	1.33088	-1.27803	2.96812	0.69407
0.69	1.31418	-1.24869	2.89972	0.70073
0.70	1.29805	-1.22002	2.83405	0.70734
0.71	1.28249	-1.192	2.77097	0.7139
0.72	1.26747	-1.1646	2.71034	0.72041
0.73	1.25296	-1.13779	2.65203	0.72688
0.74	1.23895	-1.11155	2.59591	0.7333
0.75	1.22542	-1.08586	2.54188	0.73967
0.76	1.21233	-1.0607	2.48982	0.746
0.77	1.19969	-1.03606	2.43965	0.75228
0.78	1.18747	-1.0119	2.39125	0.75852
0.79	1.17565	-0.98823	2.34456	0.76472
0.80	1.16423	-0.96501	2.29947	0.77087
0.81	1.15318	-0.94223	2.25593	0.77699
0.82	1.14249	-0.91988	2.21385	0.78306
0.83	1.13216	-0.89795	2.17317	0.78909
0.84	1.12216	-0.87642	2.13382	0.79509
0.85	1.11248	-0.85527	2.09574	0.80105
0.86	1.10312	-0.8345	2.05887	0.80697

0.87	1.09407	-0.81409	2.02317	0.81285
0.88	1.08531	-0.79403	1.98858	0.8187
0.89	1.07683	-0.77431	1.95505	0.82451
0.90	1.06863	-0.75493	1.92254	0.83028
0.91	1.06069	-0.73586	1.891	0.83602
0.92	1.05301	-0.7171	1.8604	0.84173
0.93	1.04559	-0.69865	1.83069	0.8474
0.94	1.0384	-0.68049	1.80184	0.85305
0.95	1.03145	-0.66261	1.77381	0.85865
0.96	1.02473	-0.64501	1.74657	0.86423
0.97	1.01823	-0.62768	1.7201	0.86978
0.98	1.01195	-0.6106	1.69435	0.87529
0.99	1.00587	-0.59379	1.6693	0.88077
1.00	1	-0.57722	1.64493	0.88623

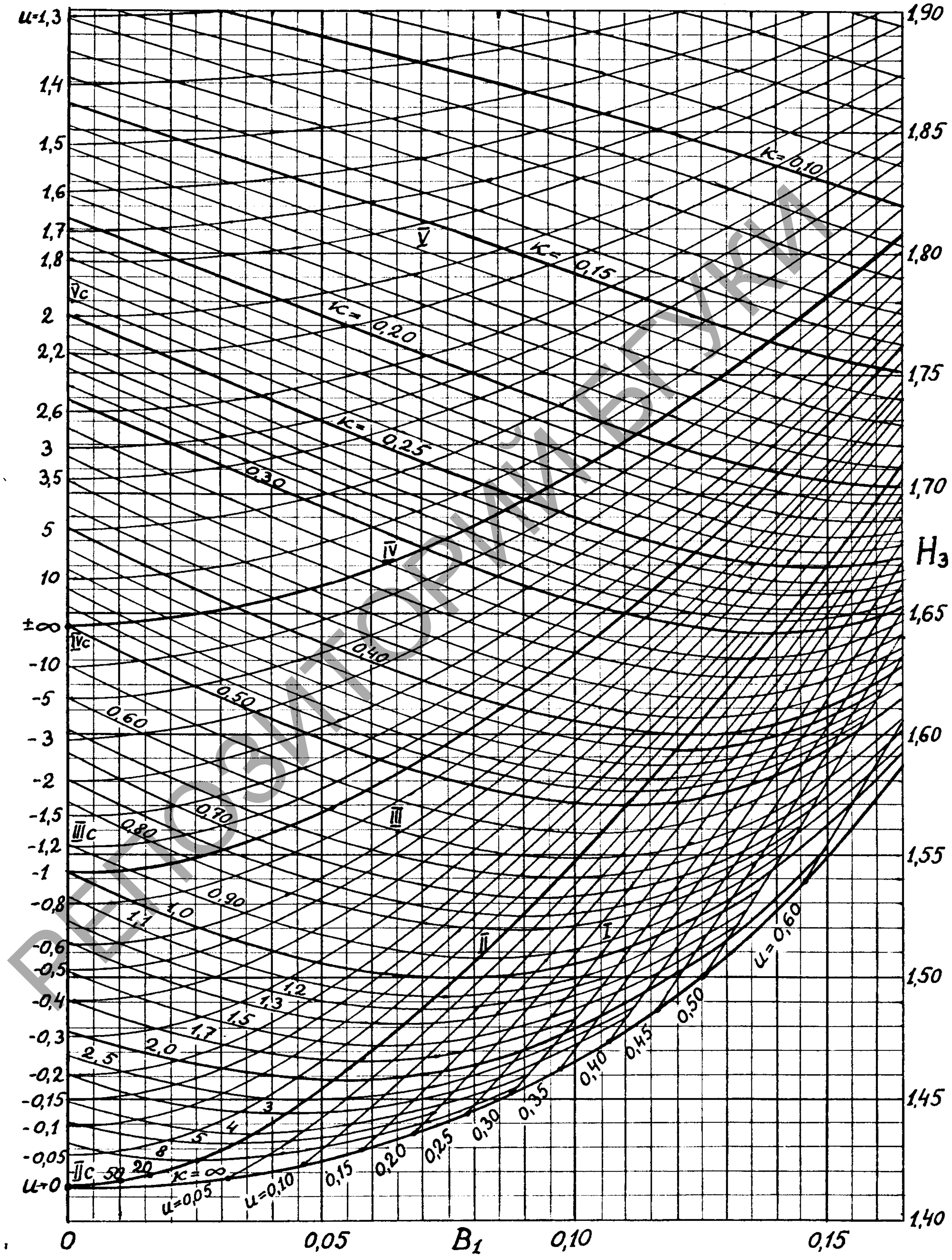
Приложение2

Номограмма для установления типа аппроксимирующего распределения и оценок параметров k, u по методу моментов



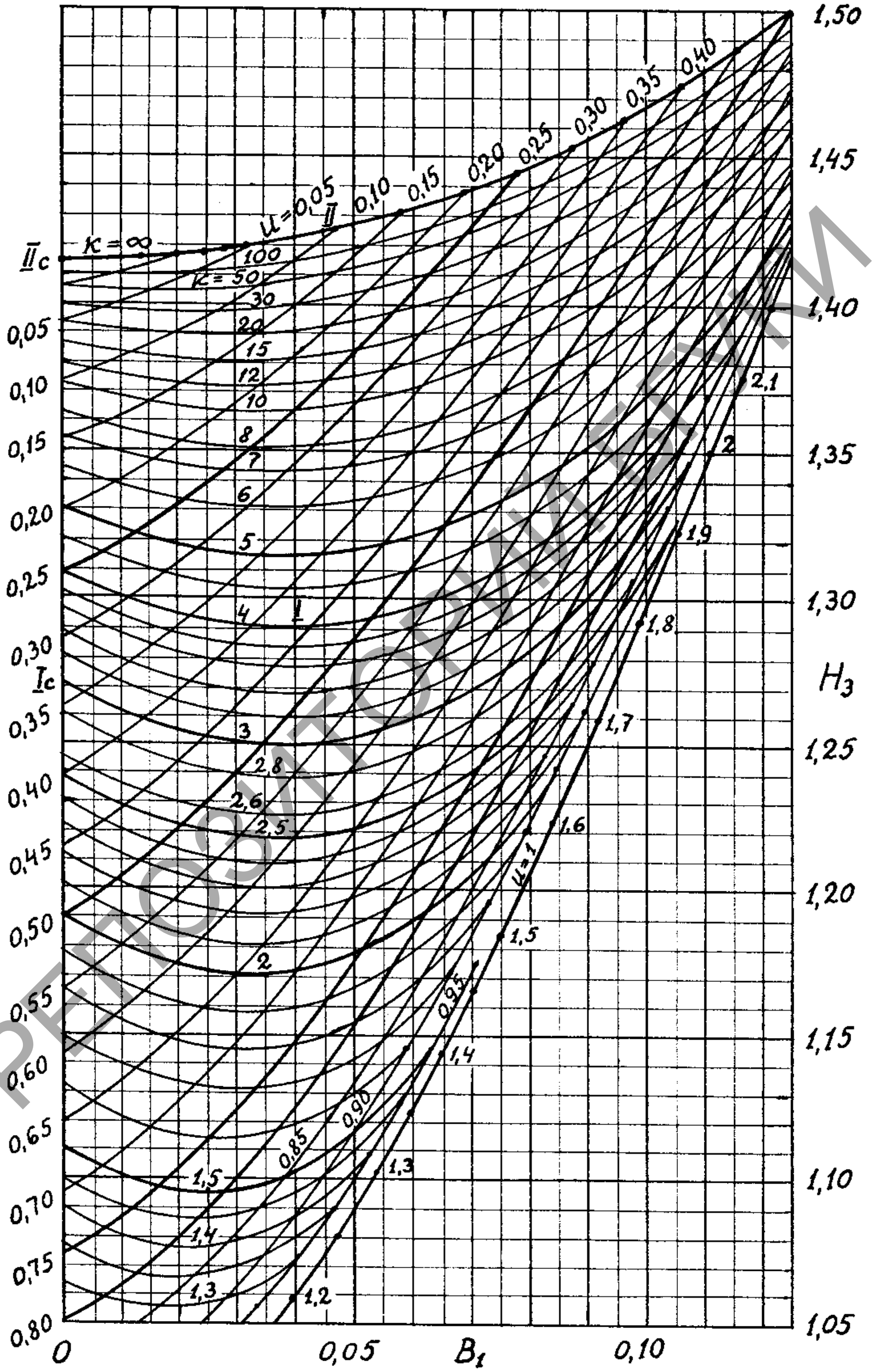
Приложение 3

Номограмма для установления типа аппроксимирующего распределения и оценок параметров k, u по общему устойчивому методу



Приложение 3А

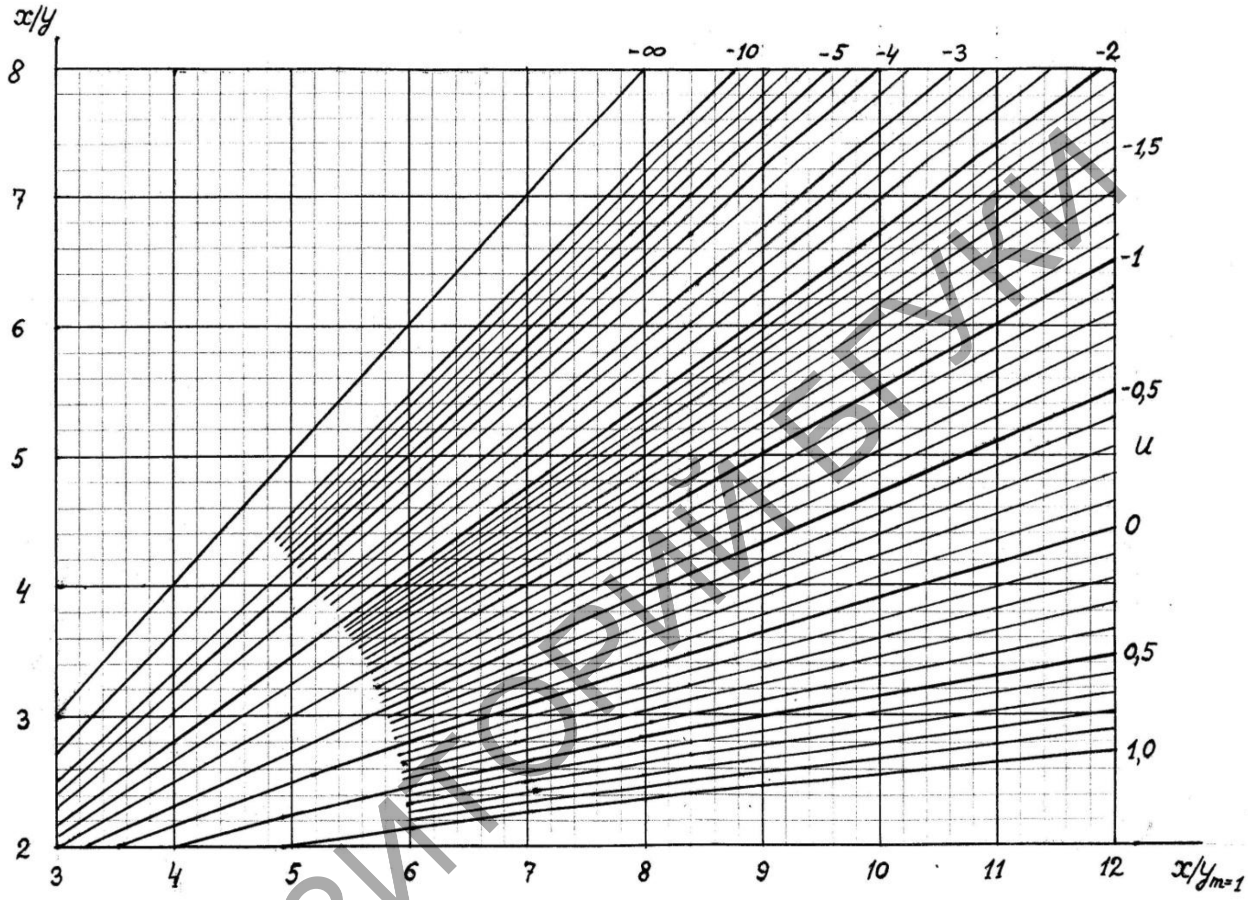
Номограмма для установления типа аппроксимирующего распределения и оценок параметров k,u по устойчивому методу при β=1.



Приложение 4

Номограмма для установления типа дискретного распределения, кривой роста и оценки параметра и по значениям величин

x/y, $x/y_{m=1}$



Приложение 5

Номограмма для установления типа дискретного распределения, кривой роста и оценки параметра u по значениям величин

y/x, $y_{m=1}/x$