

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО  
ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ЯДЕРНЫЙ УНИВЕРСИТЕТ «МИФИ»

**Свойства распределений случайных величин**

Учебное пособие

Москва 2022

УДК 53.08:519.2(075)  
ББК 30.10:22.17 Я7  
Х86

Хохлов С.С., Дмитриева А.Н. **Свойства распределений случайных величин: Учебное пособие.** – М.: НИЯУ МИФИ, 2022. 122 с.

Рассмотрены свойства распределений случайных величин, наиболее часто встречающихся при анализе экспериментальных данных, а также приведены примеры работы с этими распределениями.

Пособие предназначено для студентов первого и второго курсов магистратуры, осваивающих дисциплины «Практикум по методам анализа данных экспериментальных комплексов» и «Моделирование установок и экспериментов в области космических лучей», а также проходящих производственную практику (научно-исследовательскую работу).

Составлено в соответствии с образовательными стандартами НИЯУ МИФИ: 14.04.02 «Ядерная физика и технологии».

**Рецензент д-р физ.-мат. наук Р.П. Кокоулин.**

ISBN 978-5-7262-2891-4 ©Национальный  
исследовательский ядерный  
университет «МИФИ», 2022

# Содержание

Предисловие	4
1. Случайные величины и их распределения	5
2. Биномиальное распределение	39
3. Распределение Пуассона	47
4. Равномерное распределение	71
5. Распределение Гаусса	73
6. Распределение Коши	90
7. Гамма-распределение	93
8. Производящая функция моментов	97
9. Распределение $\chi^2$	106
Заключение	115
Рекомендуемый список литературы	120
Рекомендуемый список интернет-источников	121

# Предисловие

Статистический анализ данных является неотъемлемой частью любого научного исследования. При этом качество получаемых результатов нередко зависит от того, к какому классу и виду распределений, по предположению исследователя, относится анализируемая случайная величина. Таким образом, чем больше информацией о свойствах распределений владеет исследователь, тем проще и точнее он получает искомые результаты.

При изучении материалов данного пособия нет особой необходимости пытаться запомнить наизусть все представленные выкладки, так как они приведены здесь скорее в качестве справочного материала. Однако авторы считают, что в результате изучения этого пособия молодой исследователь должен усвоить основные характеристики описанных распределений и методы их применения на практике. С этой целью в пособии рассмотрены конкретные примеры, которые могут быть адаптированы под многие встречающиеся в научной работе задачи.

Авторы благодарят выпускников магистратуры и сотрудников Научно-образовательного центра НЕВОД НИЯУ МИФИ М.Б. Амельчакова, Т.А. Каретникову, С.С. Тимакова, И.Ю. Трошина, Е.П. Хомчука, М.Ю. Целиненко и В.В. Шутенко за помощь в подборе материалов, советы и конструктивную критику.

# 1. Случайные величины и их распределения

Базовым понятием при анализе экспериментальных данных является понятие случайной величины.

*Случайная величина* (случайная переменная, случайное значение) – это величина, принимающая с определенными вероятностями в зависимости от случая те или иные значения.

Случайная величина может быть дискретной или непрерывной. Дискретная случайная величина – это случайная величина, множество значений которой счетно, т.е. каждому значению можно присвоить номер. Количество таких номеров называется числом исходов, оно может равняться конкретному числу, например 2, для биномиального распределения или быть бесконечным, например, для распределения Пуассона. Непрерывная случайная величина имеет бесконечное множество значений, которые невозможно пронумеровать.

В качестве примеров дискретной случайной величины можно привести количество событий, регистрируемых детектором за фиксированный интервал времени; суммарный спин группы элементарных частиц, энергию электрона в атоме водорода, количество планет в звездной системе. Примеры непрерывных случайных величин: температура объекта, энергия свободной частицы, масса звезды и т.д.

Если дискретная случайная величина  $x_i$  появляется с вероятностью  $p_i$ :  $P(x_i) = p_i$ , то набор всех значений  $P(x_i)$  называется распределением дискретной случайной величины  $x$ . Для описания распределения дискретной случайной величины может использоваться как аналитическое выражение, так и табличное представление. При этом, если слу-

чайная величина равна номеру исхода  $x_n = n$ , то будем использовать обозначение  $P_n = P(x_n = n)$ .

Одним из важнейших свойств распределения дискретной случайной величины является условие нормированности (вероятность того, что случайная величина примет какое-либо значение, равна единице):

$$\sum_{i=1}^{\infty} P(x_i) = 1. \quad (1.1)$$

Поскольку непрерывная случайная величина может принимать несчетное число значений, то невозможно каждому из этих значений присвоить какую-либо, даже очень маленькую, вероятность. Однако можно рассчитать вероятность того, что непрерывная случайная величина находится в каком-то интервале значений.

Пусть  $F(a)$  – вероятность того, что непрерывная случайная величина  $x$  находится в интервале  $-\infty < x < a$ :

$$P(-\infty < x < a) = F(a).$$

Тогда функцию  $F(x)$  называют **функцией вероятности**, а функцию

$$f(x) = \frac{dF}{dx}$$

называют **функцией плотности вероятности**. Таким образом, если известна функция плотности вероятности, то вероятность того, что случайная величина находится в пределах  $-\infty < x < a$ :

$$P(-\infty < x < a) = F(a) = \int_{-\infty}^a f(x)dx.$$

Вероятность того, что случайная величина находится в интервале  $a < x < b$ :

$$\begin{aligned} P(a < x < b) &= P(-\infty < x < b) - P(-\infty < x < a) = \\ &= F(b) - F(a) = \int_a^b f(x)dx. \end{aligned}$$

Стоит отметить, что при этом вероятность наблюдения непрерывной случайной величины, равной какому-то фиксированному значению, равна нулю:

$$P(x = x_0) = \int_{x_0}^{x_0} f(x)dx = 0.$$

Это свойство часто называют *парадоксом нулевой вероятности*.

Функция вероятности и функция плотности вероятности обладают следующими свойствами:

1.  $\lim_{x \rightarrow \infty} F(x) = 1$ ,
2.  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,
3.  $F(x) \geq 0$ ,
4.  $f(x) \geq 0$ .

Функция плотности вероятности нормирована:

$$\int_{-\infty}^{\infty} f(x)dx = 1. \quad (1.2)$$

Следующим важнейшим понятием теории вероятности является *математическое ожидание* – среднееарифметическое значение случайной величины  $x$  или любой функции от случайной величины  $g(x)$ .

Математическое ожидание может быть оценено для выборки из  $N$  случайных величин:

$$E(g(x)) = \frac{\sum_{i=1}^N g(x_i)}{N},$$

для дискретного распределения:

$$E(g(x)) = \sum_{i=1}^{\infty} g(x_i)P(x_i),$$

для непрерывного распределения:

$$E(g(x)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Математическое ожидание от функции  $g(x) = x^n$  называют начальным моментом  $n$ -го порядка:

$$\mu_n = E(x^n) = \overline{x^n} = \frac{\sum_{i=1}^N x_i^n}{N},$$

$$\mu_n = E(x^n) = \overline{x^n} = \sum_{i=1}^{\infty} x_i^n P(x_i),$$

$$\mu_n = E(x^n) = \overline{x^n} = \int_{-\infty}^{\infty} x^n f(x)dx.$$

Как правило, при обсуждении выборки или распределения случайной величины в первую очередь анализируют первый и второй начальные моменты, т.е. среднее и среднеквадратичное значения.

Среднее значение (математическое ожидание случайной величины  $x$ ) рассчитывается по формулам:

$$\mu_1 = E(x) = \bar{x} = \frac{\sum_{i=1}^N x_i}{N}, \quad (1.3)$$

$$\mu_1 = E(x) = \bar{x} = \sum_{i=1}^{\infty} x_i P(x_i), \quad (1.4)$$

$$\mu_1 = E(x) = \bar{x} = \int_{-\infty}^{\infty} x f(x) dx. \quad (1.5)$$

Средний квадрат случайной величины рассчитывается как

$$\mu_2 = E(x^2) = \overline{x^2} = \frac{\sum_{i=1}^N x_i^2}{N}, \quad (1.6)$$

$$\mu_2 = E(x^2) = \overline{x^2} = \sum_{i=1}^{\infty} x_i^2 P(x_i), \quad (1.7)$$

$$\mu_2 = E(x^2) = \overline{x^2} = \int_{-\infty}^{\infty} x^2 f(x) dx. \quad (1.8)$$

Кроме начальных моментов для описания распределений используют центральные моменты. Центральным моментом  $n$ -го порядка называют математическое ожидание от функции  $g(x) = (x - \bar{x})^n$ .

Центральный момент 2-го порядка называют дисперсией случайной величины:

$$D(x) = E((x - E(x))^2). \quad (1.9)$$

Данная формула не очень удобна для использования, преобразуем ее. При этом учтем, что  $E(x) = \bar{x}$  является константой, и, поскольку математическое ожидание от константы равно самой константе  $E(C) = C$ , то и  $E(E(x)) = E(x) = \bar{x}$ .

$$\begin{aligned}
 D(x) &= E((x - E(x))^2) = \\
 &= E(x^2 - 2xE(x) + E(x)^2) = \\
 &= E(x^2) - E(2xE(x)) + E(E(x)^2) = \\
 &= E(x^2) - 2E(x)E(x) + E(x)^2 = \\
 &= E(x^2) - 2E(x)^2 + E(x)^2 = E(x^2) - E(x)^2 = \\
 &= \mu_2 - \mu_1^2 = \overline{x^2} - \bar{x}^2.
 \end{aligned} \tag{1.10}$$

Среднеквадратичным отклонением, или стандартным отклонением  $\sigma$ , называют величину, квадрат которой равен дисперсии:

$$\begin{aligned}
 D(x) &= \sigma^2, \\
 \sigma &= \sqrt{D(x)} = \sqrt{(x - \bar{x})^2} = \sqrt{x^2 - \bar{x}^2}.
 \end{aligned}$$

При кратком описании параметров распределения, кроме среднего значения и стандартного отклонения, часто приводят наиболее вероятное значение и ширину распределения на полувысоте.

Наиболее вероятное значение (мода) – это координата максимума распределения:

$$\begin{aligned}
 P(n_{\text{вер.}}) &= P_{\text{макс.}}, \\
 f(x_{\text{вер.}}) &= f_{\text{макс.}}.
 \end{aligned}$$

Ширина распределения на полувысоте (Full Width at Half Maximum, *FWHM*) определяется как разница координат полувысот распределения:

$$\begin{aligned}
 f(x_{\text{л.}}) &= \frac{f_{\text{макс.}}}{2}, \quad f(x_{\text{пр.}}) = \frac{f_{\text{макс.}}}{2}, \\
 FWHM &= x_{\text{пр.}} - x_{\text{л.}}.
 \end{aligned}$$

Также для описания распределений используют различные квантили. **Квантиль** – это значение, которое случайная величина не превышает с заданной вероятностью  $\eta$ .

Таким образом, для дискретного распределения значение  $x_k$  является  $\eta$ -м квантилем, если выполняется условие:

$$\sum_{i=1}^{i=k} P(x_i) = \eta.$$

Для непрерывного распределения величина  $x = a$  является  $\eta$ -м квантилем, если:

$$\int_{-\infty}^a f(x)dx = F(a) = \eta.$$

Если вероятность задана в процентах, то квантиль называется процентилем; если вероятность задана в десятых долях – квартили называют децинтилями. Первым, вторым и третьим квартилями называют 0.25-квантиль, 0.5-квантиль, 0.75-квантиль.

Медианой (серединой) распределения называют 0.5-квантиль (соответствует 50-му процентилю, 5-му децинтилю, 2-му квартилю).

Медиана выборки – такое значение, при котором половина элементов выборки не меньше него, а другая половина – не больше.

Для дискретного распределения значение  $x_m$  является медианным, если:

$$\sum_{i=1}^{i=m} P(x_i) = 0.5.$$

Для непрерывного распределения величина  $x_m$  является медианой, если:

$$\int_{-\infty}^{x_m} f(x)dx = F(x_m) = 0.5.$$

### **Пример 1.1. Определение параметров ряда измерений**

При повторяющихся измерениях выходного напряжения высоковольтного источника питания были получены следующие значения: 1403, 1407, 1399, 1404, 1403, 1398, 1408 и 1405 В. Определить среднее напряжение, его среднеквадратичное отклонение и медианное значение.

Найдем среднее напряжение:

$$\begin{aligned}\bar{x} &= \frac{\sum x_i}{N} = \\ &= \frac{1403 + 1407 + 1399 + 1404 + 1403 + 1398 + 1408 + 1405}{8} = \\ &= \frac{11227}{8} = 1403.375 \text{ В.}\end{aligned}$$

Среднеквадратичное отклонение можно определить двумя путями. В первом случае вычисляем среднее для квадратов отклонение от среднего значения:

$$\begin{aligned}\sigma &= \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} = \\ &= \sqrt{\frac{(1403 - 1403.375)^2 + 3.625^2 + \dots + (1.625)^2}{8}} = \\ &= \sqrt{\frac{85.87}{8}} = 3.28 \text{ В.}\end{aligned}$$

При втором способе определения среднеквадратичного отклонения сначала находим среднеквадратичное значение напряжения:

$$\begin{aligned} \overline{x^2} &= \frac{\sum x_i^2}{N} = \\ &= \frac{1403^2 + 1407^2 + 1399^2 + 1404^2 + \dots + 1405^2}{8} = \\ &= \frac{15755777}{8} = 1969472.125 \text{ В}^2, \\ \sigma &= \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{1969472.125 - 1403.375^2} = \\ &= \sqrt{10.74} = 3.28 \text{ В}. \end{aligned}$$

Для определения медианы выстроим величины по нарастанию: 1398, 1399, 1403, 1403, 1404, 1405, 1407, 1408 В. Медианным значением должно служить центральное, однако в нашем случае имеется четное число измерений равное восьми, поэтому в качестве медианного берем среднее между четвертым и пятым выстроенных по порядку напряжений:

$$x_{\text{мед.}} = \frac{1403 + 1404}{2} = 1403.5 \text{ В}.$$

### Пример 1.2. Составление и анализ гистограмм

В результате измерения отклика фотоэлектронного умножителя было получено 100 значений заряда сигнала в пКл:

110, 131, 130, 147, 110, 111, 172, 129, 108, 137, 93, 137, 119, 162, 143, 107, 115, 111, 147, 128, 100, 180, 143, 129, 152, 143, 100, 116, 152, 156, 140, 115, 168, 144, 154, 119, 100, 141, 140, 176, 118, 141, 118, 137, 141, 127, 162, 131, 121, 142, 151, 119, 159, 141, 125, 137, 121, 152, 154, 152, 118, 127, 138, 125, 128, 130, 118, 135, 114, 158, 167, 170, 124, 153, 99, 150, 138, 113, 156, 145, 137, 130, 104, 124, 165, 150, 111, 147, 153, 133, 166, 132, 111, 141, 159, 112, 120, 148, 150, 129.

Построить распределение зарядов и оценить его параметры.

Имея набор измерений, можно рассчитать среднее значение, дисперсию, среднеквадратичное отклонение и медианное значение.

Рассчитаем среднее:

$$\begin{aligned}\bar{x} &= \frac{\sum x_i}{N} = \frac{110 + 131 + 130 + \dots + 148 + 150 + 129}{100} = \\ &= \frac{13492}{100} = 134.92 \text{ пКл.}\end{aligned}$$

Дисперсию можно рассчитывать по определению:

$$\begin{aligned}D(x) &= E((x - \bar{x})^2) = \frac{\sum (x_i - \bar{x})^2}{N} = \\ &= \frac{(110 - 134.92)^2 + (131 - 134.92)^2 + \dots + (-5.9)^2}{100} = \\ &= \frac{37857}{100} = 378.6 \text{ пКл}^2.\end{aligned}$$

Также дисперсию можно рассчитать через средний квадрат:

$$\begin{aligned}\overline{x^2} &= \frac{\sum x_i^2}{N} = \frac{110^2 + 131^2 + 130^2 + \dots + 150^2 + 129^2}{100} = \\ &= \frac{1858198}{100} = 18581.98 \text{ пКл}^2, \\ D(x) &= \overline{x^2} - \bar{x}^2 = 18581.98 - 134.92^2 = 378.6 \text{ пКл}^2.\end{aligned}$$

Зная дисперсию, вычисляем среднеквадратичное отклонение:

$$\sigma = \sqrt{D(x)} = \sqrt{378.6} = 19.5 \text{ пКл}.$$

Чтобы найти медианное значение, нужно выстроить измеренные значения по возрастанию:

93, 99, 100, 100, 100, 104, 107, 108, 110, 110, 111, 111, 111, 111, 112, 113, 114, 115, 115, 116, 118, 118, 118, 118, 119, 119, 119, 120, 121, 121, 124, 124, 125, 125, 127, 127, 128, 128, 129, 129, 129, 130, 130, 130, 131, 131, 132, 133, 135, 137, 137, 137, 137, 137, 138, 138, 140, 140, 141, 141, 141, 141, 141, 142, 143, 143, 143, 144, 145, 147, 147, 147, 148, 150, 150, 150, 151, 152, 152, 152, 152, 153, 153, 154, 154, 156, 156, 158, 159, 159, 162, 162, 165, 166, 167, 168, 170, 172, 176, 180.

Количество измерений  $N = 100$  – четное число, поэтому в качестве медианного берется среднее значение между  $x_{50} = 137$  пКл и  $x_{51} = 137$  пКл. Таким образом получаем, что  $x_{\text{мед.}} = \frac{x_{50} + x_{51}}{2} = \frac{137 + 137}{2} = 137$  пКл.

Однако полученные значения не дают представлений о форме распределения измеренной случайной величины; для этих целей проводят гистограммирование.

Измеренные значения объединяют по интервалам определенной длины, подсчитывая, сколько значений попадет в каждый интервал. Длина интервала называется **шагом гистограммы**. Например, выбираем шаг гистограм-

мы, равный  $s = 1$  пКл, началом каждого интервала выбираем полуцелое число. Таким образом получаем интервалы  $(92.5, 93.5]$ ,  $(93.5, 94.5]$ ,  $(94.5, 95.5]$ , ...  $(178.5, 179.5]$ ,  $(179.5, 180.5]$ . После этого подсчитываем, сколько значений попало в каждый интервал. В интервал  $(92.5, 93.5]$  попадет одно событие; в интервалы  $(93.5, 94.5]$ ,  $(94.5, 95.5]$ ,  $(95.5, 96.5]$ ,  $(96.5, 97.5]$ ,  $(97.5, 98.5]$  – ноль событий; в интервал  $(98.5, 99.5]$  – одно событие; в интервал  $(99.5, 100.5]$  – три события и т.д. Итоговую гистограмму (табл. 1.1) записывают, используя центры интервалов.

Полученное распределение количества событий по заряду при шаге гистограммирования  $s = 1$  пКл представлено на рис. 1.1 слева вверху. В полученной гистограмме много пустых бинов, что мешает восприятию, для улучшения представления необходимо увеличить шаг гистограммы. Рассмотрим варианты с  $s = 5$ , 10 и 20 пКл. При шаге  $s = 5$  пКл получаем гистограмму, приведенную в табл. 1.2 слева и показанную на рис. 1.1 справа вверху. При шаге  $s = 10$  пКл получаем гистограмму, приведенную в табл. 1.2 в центре и показанную на рис. 1.1 слева внизу. При шаге  $s = 20$  пКл получаем гистограмму, приведенную в табл. 1.2 справа и показанную на рис. 1.1 справа внизу.

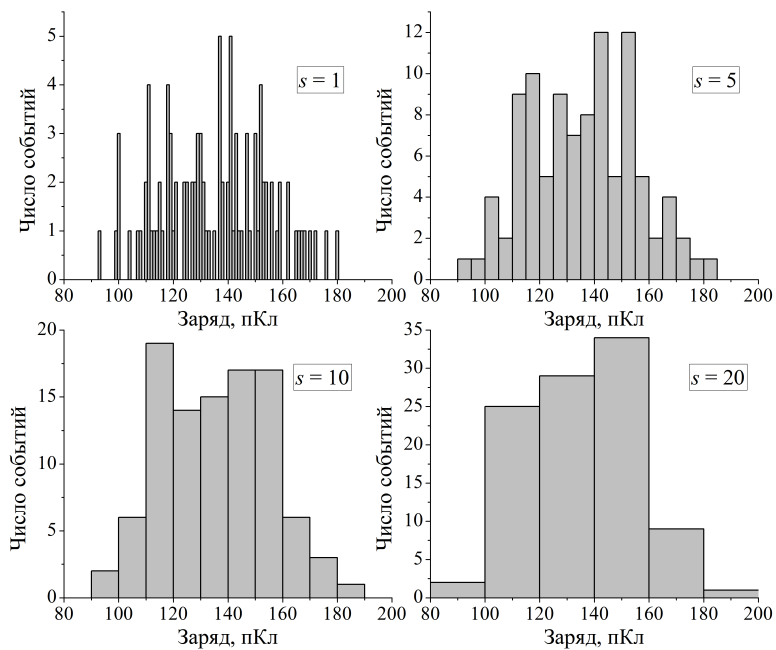


Рис. 1.1. Распределения событий по заряду при шаге гистограммирования  $s = 1, 5, 10$  и  $20$  пКл

Таблица 1.1. Распределение событий по заряду при шаге гистограммирования  $s = 1$  пКл:  $x_i$  – заряд сигнала (в пКл),  $N_i$  – количество событий

$x_i$	$N_i$	$x_i$	$N_i$	$x_i$	$N_i$	$x_i$	$N_i$
93	1	115	2	137	5	159	2
94	0	116	1	138	2	160	0
95	0	117	0	139	0	161	0
96	0	118	4	140	2	162	2
97	0	119	3	141	5	163	0
98	0	120	1	142	1	164	0
99	1	121	2	143	3	165	1
100	3	122	0	144	1	166	1
101	0	123	0	145	1	167	1
102	0	124	2	146	0	168	1
103	0	125	2	147	3	169	0
104	1	126	0	148	1	170	1
105	0	127	2	149	0	171	0
106	0	128	2	150	3	172	1
107	1	129	3	151	1	173	0
108	1	130	3	152	4	174	0
109	0	131	2	153	2	175	0
110	2	132	1	154	2	176	1
111	4	133	1	155	0	177	0
112	1	134	0	156	2	178	0
113	1	135	1	157	0	179	0
114	1	136	0	158	1	180	1

Таблица 1.2. Распределения событий по заряду при шаге гистограммирования  $s = 5, 10$  и  $20$  пКл:  $x_i$  – заряд сигнала (в пКл),  $N_i$  – количество событий

$s = 5$ пКл		$s = 10$ пКл		$s = 20$ пКл	
$x_i$	$N_i$	$x_i$	$N_i$	$x_i$	$N_i$
92.5	1	95	2	90	2
97.5	1	105	6	110	25
102.5	4	115	19	130	29
107.5	2	125	14	150	34
112.5	9	135	15	170	9
117.5	10	145	17	190	1
122.5	5	155	17		
127.5	9	165	6		
132.5	7	175	3		
137.5	8	185	1		
142.5	12				
147.5	5				
152.5	12				
157.5	5				
162.5	2				
167.5	4				
172.5	2				
177.5	1				
182.5	1				

Данные, хранящиеся в виде гистограмм, меньше по объему, чем исходные данные. Однако чем больше шаг гистограммы, тем с меньшей точностью рассчитываются параметры распределения.

При шаге  $s = 1$  пКл:

$$\begin{aligned}\bar{x} &= \frac{\sum x_i N_i}{\sum N_i} = \\ &= \frac{93 \cdot 1 + 94 \cdot 0 + 95 \cdot 0 + \dots + 178 \cdot 0 + 179 \cdot 0 + 180 \cdot 1}{1 + 0 + 0 + \dots + 0 + 0 + 1} = \\ &= \frac{13492}{100} = 134.92 \text{ пКл.}\end{aligned}$$

$$\begin{aligned}\overline{x^2} &= \frac{\sum x_i^2 N_i}{\sum N_i} = \\ &= \frac{93^2 \cdot 1 + 94^2 \cdot 0 + 95^2 \cdot 0 + \dots + 179^2 \cdot 0 + 180^2 \cdot 1}{1 + 0 + 0 + \dots + 0 + 1} = \\ &= \frac{1858198}{100} = 18581.98 \text{ пКл}^2.\end{aligned}$$

$$D(x) = \overline{x^2} - \bar{x}^2 = 18581.98 - 134.92^2 = 378.6 \text{ пКл}^2.$$

$$\sigma = \sqrt{D(x)} = \sqrt{378.57} = 19.5 \text{ пКл.}$$

Для поиска медианного значения по гистограмме необходимо взять половину от суммы значений всех бинов  $\sum N/2 = 50$ , после чего, начиная с самого левого или с самого правого бина, считать накопленную сумму, пока не получим искомую величину. После суммирования бина  $x_1 = 136$  накопленная сумма равна  $N_1 = 49$ , в бине с  $x_2 = 137$  лежит 5 значений, т.е. после суммирования этого бина накопленная сумма будет равна  $N_2 = 53$ . Следовательно,  $x_{\text{мед.}} = x_2 = 137$  пКл.

Значения получились полностью совпадающими с расчетом по негистограммированному ряду, поскольку шаг гистограммы не превышает минимальную разницу между измеренными величинами.

Оценим среднее значение и среднеквадратичное отклонение для гистограммы с  $s = 5$  пКл.

$$\begin{aligned}\bar{x} &= \frac{\sum x_i N_i}{\sum N_i} = \\ &= \frac{92.5 \cdot 1 + 97.5 \cdot 1 + 102.5 \cdot 4 + \dots + 177.5 \cdot 1 + 182.5 \cdot 1}{1 + 1 + 4 + \dots + 1 + 1} = \\ &= \frac{13555}{100} = 135.55 \text{ пКл.}\end{aligned}$$

$$\begin{aligned}\overline{x^2} &= \frac{\sum x_i^2 N_i}{\sum N_i} = \\ &= \frac{92.5^2 \cdot 1 + 97.5^2 \cdot 1 + 102.5^2 \cdot 4 + \dots + 177.5^2 \cdot 1 + 182.5^2 \cdot 1}{1 + 1 + 4 + \dots + 1 + 1} = \\ &= \frac{1875575}{100} = 18755.75 \text{ пКл}^2.\end{aligned}$$

$$D(x) = \overline{x^2} - \bar{x}^2 = 18755.75 - 135.55^2 = 382 \text{ пКл}^2.$$

$$\sigma = \sqrt{D(x)} = \sqrt{381.95} = 19.5 \text{ пКл.}$$

При определении медианного значения при суммировании до бина  $x_1 = 132.5$  имеем накопленную сумму  $N_1 = 48$  и при  $x_2 = 137.5$   $N_2 = 56$ . Таким образом,  $x_{\text{мед.}} = x_2 = 137.5$  пКл.

Для гистограммы с шагом  $s = 10$  пКл:

$$\begin{aligned}\bar{x} &= \frac{\sum x_i N_i}{\sum N_i} = \\ &= \frac{95 \cdot 2 + 105 \cdot 6 + 115 \cdot 19 + \dots + 165 \cdot 6 + 175 \cdot 3 + 185 \cdot 1}{2 + 6 + 19 + \dots + 6 + 3 + 1} = \\ &= \frac{13580}{100} = 135.8 \text{ пКл.}\end{aligned}$$

$$\begin{aligned} \overline{x^2} &= \frac{\sum x_i^2 N_i}{\sum N_i} = \\ &= \frac{95^2 \cdot 2 + 105^2 \cdot 6 + 115^2 \cdot 19 + \dots + 175^2 \cdot 3 + 185^2 \cdot 1}{2 + 6 + 19 + \dots + 3 + 1} = \\ &= \frac{1882900}{100} = 18829.0 \text{ пКл}^2. \end{aligned}$$

$$D(x) = \overline{x^2} - \bar{x}^2 = 18829 - 135.8^2 = 387.4 \text{ пКл}^2.$$

$$\sigma = \sqrt{D(x)} = \sqrt{387.4} = 19.7 \text{ пКл}.$$

Накопленная сумма вместе с бином  $x_1 = 125$  равна  $N_1 = 41$ , если сложить эту сумму со следующим бином  $x_2 = 135$ , получаем  $N_2 = 56$ . Следовательно,  $x_{\text{мед.}} = x_2 = 135$  пКл.

При шаге гистограммы  $s = 20$  пКл:

$$\begin{aligned} \bar{x} &= \frac{\sum x_i N_i}{\sum N_i} = \\ &= \frac{90 \cdot 2 + 110 \cdot 25 + 130 \cdot 29 + 150 \cdot 34 + 170 \cdot 9 + 190 \cdot 1}{2 + 25 + 29 + 34 + 9 + 1} = \\ &= \frac{13520}{100} = 135.2 \text{ пКл}. \end{aligned}$$

$$\begin{aligned} \overline{x^2} &= \frac{\sum x_i^2 N_i}{\sum N_i} = \\ &= \frac{90^2 \cdot 2 + 110^2 \cdot 25 + 130^2 \cdot 29 + \dots + 170^2 \cdot 9 + 190^2 \cdot 1}{2 + 25 + 29 + \dots + 9 + 1} = \\ &= \frac{1870000}{100} = 18700 \text{ пКл}^2. \end{aligned}$$

$$D(x) = \overline{x^2} - \bar{x}^2 = 18700 - 135.2^2 = 421.0 \text{ пКл}^2.$$

$$\sigma = \sqrt{D(x)} = \sqrt{421} = 20.5 \text{ пКл.}$$

Сумма значений бинов, включая  $x_1 = 110$ , равна  $N_1 = 27$ ; суммируя следующий бин  $x_2 = 130$ , получаем  $N_2 = 56$ . Следовательно,  $x_{\text{мед.}} = x_2 = 130$  пКл. Поскольку шаг гистограммы достаточно большой, медианное значение можно оценить более точно, аппроксимируя значения линейной зависимостью. При этом необходимо учитывать, что в бине  $x_1 = 110$  пКл учтены все значения до  $x_1 + s/2 = 120$  пКл, а в следующем бине учтены все значения до  $x_2 + s/2 = 140$  пКл. Тогда

$$\begin{aligned} x_{\text{мед.}} &= x_1 + \frac{s}{2} + (x_2 - x_1) \frac{\sum N_i - N_1}{N_2 - N_1} = \\ &= 110 + \frac{20}{2} + (130 - 110) \frac{50 - 27}{56 - 27} = 133.8 \text{ пКл.} \end{aligned}$$

Как видно из проведенных вычислений, рассчитанные средние значения и среднеквадратичные отклонения для гистограмм с шагом  $s = 1, 5$  и  $10$  пКл практически совпадают с прямым расчетом для ряда значений. При расчете среднеквадратичного отклонения при шаге гистограммы  $s = 20$  пКл появляется заметное отклонение от первоначального значения. Это происходит из-за того, что шаг гистограммы становится сопоставимым с реальным среднеквадратичным отклонением.

Таким образом, можно сделать вывод, что оптимальным шагом гистограммирования в данном случае является  $s = 10$  пКл.

### Пример 1.3. Определение характеристик дискретного распределения

При калибровке одного из фотоумножителей черенковского водного детектора НЕВОД был получен спектр его откликов на события с вертикальными мюонами. Спектр приведен в табл. 1.3 и показан на рис. 1.2. Определить основные характеристики данного распределения.

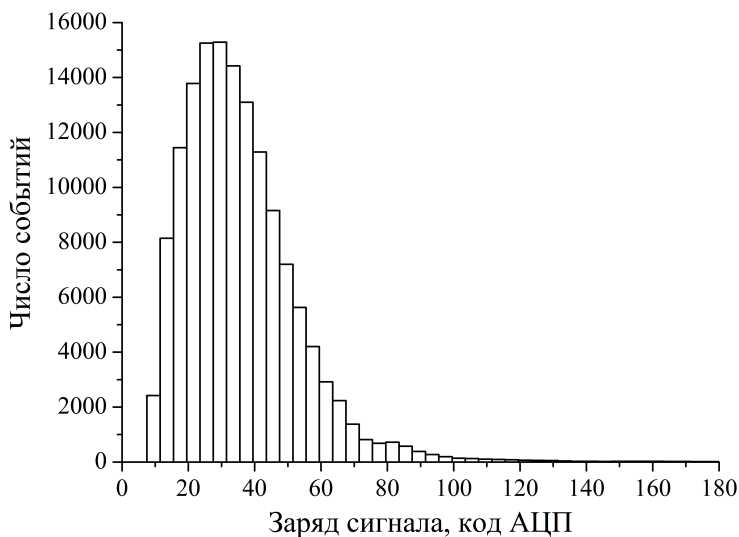


Рис. 1.2. Пример дискретного распределения

Приведенный спектр является типичным примером дискретного распределения. Событий, в которых отклик ФЭУ составлял 0, 1, 2 или 3 кода амплитудно-цифрового преобразователя (АЦП), помещались в первый бин, события с откликом 4, 5, 6 или 7 – во второй бин и т.д. События с откликами более 300 кодов АЦП учитывались в последнем бине распределения.

Таблица 1.3. Пример дискретного распределения:  $x_i$  – заряд сигнала (в кодах АЦП),  $N_i$  – количество событий

$x_i$	$N_i$	$x_i$	$N_i$	$x_i$	$N_i$
0-3	0	100-103	142	200-203	7
4-7	1	104-107	127	204-207	3
8-11	2 421	108-111	103	208-211	7
12-15	8 146	112-115	92	212-215	7
16-19	11 444	116-119	84	216-219	6
20-23	13 785	120-123	66	220-223	8
24-27	15 251	124-127	62	224-227	6
28-31	15 283	128-131	56	228-231	7
32-35	14 427	132-135	36	232-235	5
36-39	13 096	136-139	26	236-239	4
40-43	11 291	140-143	26	240-243	5
44-47	9 157	144-147	20	244-247	5
48-51	7 200	148-151	23	248-251	3
52-55	5 626	152-155	23	252-255	3
56-59	4 206	156-159	27	256-259	3
60-63	2 922	160-163	22	260-263	3
64-67	2 238	164-167	20	264-267	3
68-71	1 377	168-171	21	268-271	7
72-75	817	172-175	14	272-275	1
76-79	683	176-179	13	276-279	2
80-83	719	180-183	13	280-283	1
84-87	576	184-187	6	284-287	3
88-91	385	188-191	4	288-291	2
92-95	272	192-195	4	292-295	0
96-99	196	196-199	12	296-299	60

Определим среднее значение распределения. При подсчете будем учитывать средний отклик в каждом бине (1.5 кода АЦП – в первом, 5.5 – во втором и т.д.):

$$\begin{aligned}\bar{x} &= \frac{\sum x_i N_i}{\sum N_i} = \\ &= \frac{1.5 \cdot 0 + 5.5 \cdot 1 + 9.5 \cdot 2421 + \dots + 293.5 \cdot 0 + 297.5 \cdot 60}{0 + 1 + 2421 + \dots + 0 + 60} = \\ &= \frac{5122955}{142722} = 35.89 \text{ кода АЦП.}\end{aligned}$$

Для того чтобы определить среднеквадратичное отклонение, вычислим среднеквадратичное значение отклика:

$$\begin{aligned}\overline{x^2} &= \frac{\sum x_i^2 N_i}{\sum N_i} = \\ &= \frac{1.5^2 \cdot 0 + 5.5^2 \cdot 1 + 9.5^2 \cdot 2421 + \dots + 297.5^2 \cdot 60}{0 + 1 + 2421 + \dots + 60} = \\ &= \frac{235704029}{142722} = 1651.49 \text{ квадратных кодов АЦП.}\end{aligned}$$

Теперь можем рассчитать среднеквадратичное отклонение:

$$\begin{aligned}\sigma &= \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{1651.49 - 35.89^2} = \\ &= \sqrt{1651.49 - 1288.09} = 19.1 \text{ кода АЦП.}\end{aligned}$$

Определим наиболее вероятный отклик. Максимальное количество событий (15283) наблюдается в бине с зарядами 28–31. Таким образом, наиболее вероятным значением заряда можно считать середину бина, т.е. 29.5 кода АЦП. Стоит отметить, что наиболее вероятное для данного распределения меньше среднего значения, что объясняется несимметричностью данного распределения – наличием правого «хвоста».

Теперь определим полную ширину на половине высоты (*FWHM*). Как уже было отмечено, максимальное количество событий в бине равно  $N_{\text{макс.}} = 15283$ . Это значение соответствует высоте распределения. Тогда полувысота распределения равна  $N_{\text{макс.}}/2 = 7641.5$ .

Двигаясь влево от наиболее вероятного, находим два бина, значения в которых максимально приближены к полувысоте:

$$\begin{array}{ll} 8 - 11 & 2421 \\ 12 - 15 & 8146 \end{array}$$

Левую границу искомого диапазона можно вычислить линейным фитированием:

$$\begin{aligned} x_{\text{л.}} &= x_i + (x_{i+1} - x_i) \frac{N_{\text{макс.}}/2 - N_i}{N_{i+1} - N_i} = \\ &= 9.5 + (13.5 - 9.5) \frac{7641.5 - 2421}{8146 - 2421} = 13.1 \text{ кода АЦП.} \end{aligned}$$

Теперь, двигаясь вправо от наиболее вероятного, находим два бина, значения в которых максимально приближены к полувысоте:

$$\begin{array}{ll} 44 - 47 & 9157 \\ 48 - 51 & 7200 \end{array}$$

Определяем правую границу:

$$\begin{aligned} x_{\text{пр.}} &= x_i + (x_{i+1} - x_i) \frac{N_{\text{макс.}}/2 - N_i}{N_{i+1} - N_i} = \\ &= 45.5 + (49.5 - 45.5) \frac{7641.5 - 9157}{7200 - 9157} = 48.6 \text{ кода АЦП.} \end{aligned}$$

Ширина распределения на полувысоте равна:

$$FWHM = x_{\text{пр.}} - x_{\text{л.}} = 48.6 - 13.1 = 35.5 \text{ кода АЦП.}$$

Построим функцию вероятности распределения. Для этого последовательно складываем количество событий в бинах и заносим результаты в табл. 1.4. Как видно из табл. 1.3, в бине  $0 - 3$  имеется 0 событий, т.е. 0 событий имеют заряд сигнала меньше или равный 3 кодам АЦП. В бине  $4 - 7$  имеется 1 событие, следовательно, имеется  $0 + 1 = 1$  сигнал с зарядом, не превышающим 7 кодов АЦП, заносим это значение в табл. 1.4 во вторую колонку в строчку напротив числа 7. В бине  $8 - 11$  имеется 2421 событие, т.е. всего имеется  $0 + 1 + 2421 = 2422$  сигнала с зарядом меньше или равным 11 кодов АЦП. В бине  $12 - 15$  в табл. 1.3 имеется 8146 событий, поэтому напротив значения 15 записываем  $0 + 1 + 2421 + 8146 = 10568$  и т.д. В последнем бине распределения имеется 60 событий, количество событий, с зарядом не более 299 кодов АЦП равно  $0 + 1 + 2421 + \dots + 60 = 142722$ .

Для расчета значений функции вероятности необходимо разделить значения из второй колонки на общее число событий в гистограмме, т.е. на 142722. Соответствующие значения функции вероятности приведены в третьей колонке табл. 1.4. График функции вероятности приведен на рис. 1.3.

Значения функции вероятности позволяют анализировать различные квантили распределения. Например, чтобы определить 95-й процентиль, необходимо найти решение уравнения:

$$F(x_{95}) = 0.95.$$

При  $x = 63$  вероятность равна 0.94068, а при  $x = 67$  вероятность равна 0.95636. Проведя линейную аппроксимацию, получаем:

$$\begin{aligned}
 x_{95} &= 63 + (67 - 63) \frac{0.95 - 0.94068}{0.95636 - 0.94068} = \\
 &= 63 + (67 - 63) \frac{0.00932}{0.01568} = 65.4 \text{ кода АЦП.}
 \end{aligned}$$

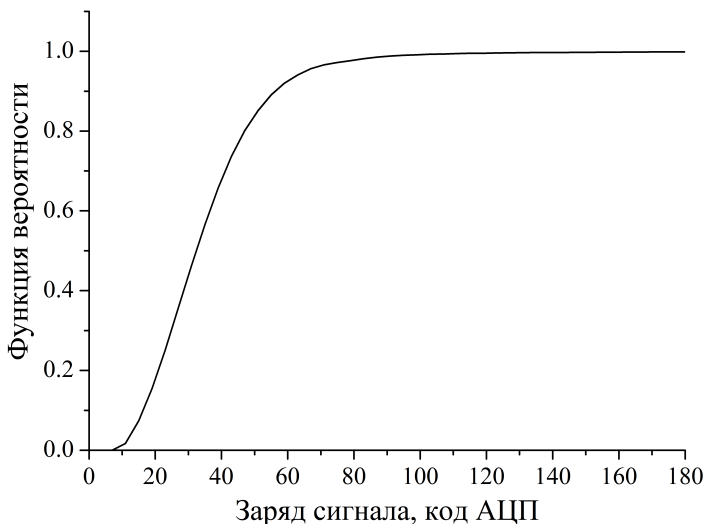


Рис. 1.3. Функция вероятности для примера 1.3

Следовательно, несмотря на то, что в распределении имеются сигналы с зарядом до 299 кодов АЦП, 95 % всех сигналов имеют заряд менее 65.4 кода АЦП.

Найдем квантили распределения, которые соответствуют 25-му, 50-му, 75-му и 100-му процентиям.

Таблица 1.4. Функция вероятности распределения  
для примера 1.3

Заряд сигнала $x_i$ , код АЦП	Количество событий, не превышающих $x_i$	Значения функции вероятности
3	0	0
7	1	0.00001
11	2422	0.01697
15	10568	0.07405
19	22012	0.15423
23	35797	0.25082
27	51048	0.35767
31	66331	0.46476
35	80758	0.56584
39	93854	0.65760
43	105145	0.73671
47	114302	0.80087
51	121502	0.85132
55	127128	0.89074
59	131334	0.92021
63	134256	0.94068
67	136494	0.95636
71	137871	0.96601
75	138688	0.97174
79	139371	0.97652
83	140090	0.98156
87	140666	0.98559
91	141051	0.98829
95	141323	0.99020
99	141519	0.99157
103	141661	0.99257
107	141788	0.99346

Продолжение табл. 1.4

Заряд сигнала $x_i$ , код АЦП	Количество событий, не превышающих $x_i$	Значения функции вероятности
111	141891	0.99418
115	141983	0.99482
119	142067	0.99541
123	142133	0.99587
127	142195	0.99631
131	142251	0.99670
135	142287	0.99695
139	142313	0.99713
143	142339	0.99732
147	142359	0.99746
151	142382	0.99762
155	142405	0.99778
159	142432	0.99797
163	142454	0.99812
167	142474	0.99826
171	142495	0.99841
175	142509	0.99851
179	142522	0.99860
183	142535	0.99869
187	142541	0.99873
191	142545	0.99876
195	142549	0.99879
199	142561	0.99887
203	142568	0.99892
207	142571	0.99894
211	142578	0.99899
215	142585	0.99904
219	142591	0.99908

Окончание табл. 1.4

Заряд сигнала $x_i$ , код АЦП	Количество событий, не превышающих $x_i$	Значения функции вероятности
223	142599	0.99914
227	142605	0.99918
231	142612	0.99923
235	142617	0.99926
239	142621	0.99929
243	142626	0.99933
247	142631	0.99936
251	142634	0.99938
255	142637	0.99940
259	142640	0.99943
263	142643	0.99945
267	142646	0.99947
271	142653	0.99952
275	142654	0.99952
279	142656	0.99954
283	142657	0.99955
287	142660	0.99957
291	142662	0.99958
295	142662	0.99958
299	142722	1

Первый квартиль находится из уравнения

$$\begin{aligned} F(x_{Q_1}) &= F(x_{25}) = 0.25, \\ x_{Q_1} = x_{25} &= 19 + (23 - 19) \frac{0.25 - 0.15423}{0.25082 - 0.15423} = \\ &= 19 + (23 - 19) \frac{0.09577}{0.09659} = 23.0 \text{ кода АЦП.} \end{aligned}$$

Второй квартиль, также называемый медианой, определяется как

$$\begin{aligned} F(x_{Q_2}) &= F(x_{\text{мед.}}) = F(x_{50}) = 0.50, \\ x_{Q_2} = x_{\text{мед.}} = x_{50} &= 31 + (35 - 31) \frac{0.5 - 0.46476}{0.56584 - 0.46476} = \\ &= 31 + (35 - 31) \frac{0.03524}{0.10108} = 32.4 \text{ кода АЦП.} \end{aligned}$$

Третий квартиль:

$$\begin{aligned} F(x_{Q_3}) &= F(x_{75}) = 0.75, \\ x_{Q_3} = x_{75} &= 43 + (47 - 43) \frac{0.75 - 0.73671}{0.80087 - 0.73671} = \\ &= 43 + (47 - 43) \frac{0.01329}{0.06416} = 43.8 \text{ кода АЦП.} \end{aligned}$$

Четвертый квартиль не требует вычислений, он равен последнему значению заряда, при котором функция вероятности принимает значение 1:

$$\begin{aligned} F(x_{Q_4}) &= F(x_{100}) = 1.0, \\ x_{Q_4} = x_{100} &= 299 \text{ кодов АЦП.} \end{aligned}$$

На основе вычисленных квартилей можно утверждать, что в интервалах  $[0, 23.0]$ ,  $(23.0, 32.4]$ ,  $(32.4, 43.8]$  и  $(43.8, 299.0]$  кодов АЦП лежит одинаковое число событий.

Наиболее важным из квартилей является второй – это медиана распределения. Половина входящих в распределение величин имеют значения меньше медианного, вторая половина – имеют значения больше медианного.

Медианное значение, наряду со средним значением и наиболее вероятным значением, часто используют для описания распределения. Обратим внимание, что для данного распределения эти базовые величины не совпадают между собой:  $\bar{x} = 35.9$  кода АЦП,  $x_{\text{вер.}} = 29.5$ ,  $x_{\text{мед.}} = 32.4$  кода АЦП.

### Пример 1.4. Определение параметров непрерывного распределения

Распределение электронов по энергиям имеет вид треугольника и проходит через точки  $(1 \text{ ГэВ}, 0)$   $(2 \text{ ГэВ}, h)$   $(4 \text{ ГэВ}, 0)$ , где  $h$  – параметр (рис. 1.4). Определить среднее значение, дисперсию, среднеквадратичное отклонение, ширину на полувысоте, наиболее вероятное и медианное значения распределения.

Для начала определимся с размерностью по оси ординат. По ней откладывается плотность вероятности, следовательно, ее размерность обратна размерности физической величины, отложенной по оси абсцисс, т.е.  $[\text{ГэВ}^{-1}]$ .

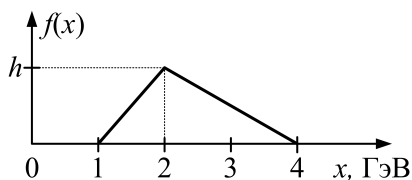


Рис. 1.4. Распределение электронов по энергиям

Для начала запишем плотность вероятности данного распределения, имеющего два линейных участка. Первый участок должен проходить через точки  $(1, 0)$  и  $(2, h)$ . Для получения параметров прямой  $y = k_1 \cdot x + b_1$  составляем систему уравнений:

$$\begin{cases} 0 = k_1 \cdot 1 + b_1, \\ h = k_1 \cdot 2 + b_1. \end{cases}$$

Решив систему, получаем значения для коэффициентов  $k_1 = h [\Gamma \text{ЭВ}^{-2}]$ ,  $b_1 = -h [\Gamma \text{ЭВ}^{-1}]$ .

Теперь найдем параметры прямой на втором участке между точками  $(2, h)$  и  $(4, 0)$ :

$$\begin{cases} h = k_2 \cdot 2 + b_2, \\ 0 = k_2 \cdot 4 + b_2. \end{cases}$$

Решив систему, получаем  $k_2 = -0.5h [\Gamma \text{ЭВ}^{-2}]$ ,  $b_2 = 2h [\Gamma \text{ЭВ}^{-1}]$ .

Таким образом, плотность вероятности имеет вид

$$f(x) = \begin{cases} hx - h, & 1 \Gamma \text{ЭВ} < x < 2 \Gamma \text{ЭВ}, \\ -0.5hx + 2h, & 2 \Gamma \text{ЭВ} < x < 4 \Gamma \text{ЭВ}. \end{cases}$$

Параметр  $h$  определим из условия нормировки:

$$\begin{aligned} \int_{x_{\text{мин.}}}^{x_{\text{макс.}}} f(x) dx &= 1, \\ \int_1^2 f_1(x) dx + \int_2^4 f_2(x) dx &= 1, \\ \int_1^2 (hx - h) dx + \int_2^4 (-0.5hx + 2h) dx &= 1, \\ \left( h \frac{x^2}{2} - hx \right) \Big|_1^2 + \left( -0.5h \frac{x^2}{2} + 2hx \right) \Big|_2^4 &= 1, \\ \left( h \frac{2^2}{2} - 2h - h \frac{1^2}{2} + h \right) + & \\ + \left( -0.5h \frac{4^2}{2} + 8h + 0.5h \frac{2^2}{2} - 4h \right) &= 1, \end{aligned}$$

$$\begin{aligned}
 h \frac{3}{2} - h - h \frac{6}{2} + 4h &= 1, \\
 h \frac{3}{2} &= 1, \\
 h &= \frac{2}{3}.
 \end{aligned}$$

Таким образом, получаем отнормированную плотность вероятности:

$$f(x) = \begin{cases} \frac{2}{3}x - \frac{2}{3}, & 1 \Gamma \text{ЭВ} < x < 2 \Gamma \text{ЭВ}, \\ -\frac{1}{3}x + \frac{4}{3}, & 2 \Gamma \text{ЭВ} < x < 4 \Gamma \text{ЭВ}. \end{cases}$$

Найдем среднее значение:

$$\begin{aligned}
 \bar{x} = E(x) &= \int_{x_{\text{мин.}}}^{x_{\text{макс.}}} x f(x) dx = \int_1^2 x f_1(x) dx + \int_2^4 x f_2(x) dx = \\
 &= \int_1^2 x \left( \frac{2}{3}x - \frac{2}{3} \right) dx + \int_2^4 x \left( -\frac{1}{3}x + \frac{4}{3} \right) dx = \\
 &= \left( \frac{2}{3} \frac{x^3}{3} - \frac{2}{3} \frac{x^2}{2} \right) \Big|_1^2 + \left( -\frac{1}{3} \frac{x^3}{3} + \frac{4}{3} \frac{x^2}{2} \right) \Big|_2^4 = \\
 &= \left( \frac{2 \cdot 2^3}{3 \cdot 3} - \frac{2 \cdot 2^2}{3 \cdot 2} - \frac{2 \cdot 1^3}{3 \cdot 3} + \frac{2 \cdot 1^2}{3 \cdot 2} \right) + \\
 &\quad + \left( -\frac{1 \cdot 4^3}{3 \cdot 3} + \frac{4 \cdot 4^2}{3 \cdot 2} + \frac{1 \cdot 2^3}{3 \cdot 3} - \frac{4 \cdot 2^2}{3 \cdot 2} \right) = \\
 &= \left( \frac{14}{9} - \frac{6}{6} \right) + \left( -\frac{56}{9} + \frac{48}{6} \right) = \frac{5}{9} + \frac{16}{9} = \frac{7}{3} \Gamma \text{ЭВ}.
 \end{aligned}$$

Для вычисления дисперсии рассчитаем средний квадрат:

$$\begin{aligned}
 \overline{x^2} = E(x^2) &= \int_{x_{\text{мин.}}}^{x_{\text{макс.}}} x^2 f(x) dx = \int_1^2 x^2 f_1(x) dx + \int_2^4 x^2 f_2(x) dx = \\
 &= \int_1^2 x^2 \left( \frac{2}{3}x - \frac{2}{3} \right) dx + \int_2^4 x^2 \left( -\frac{1}{3}x + \frac{4}{3} \right) dx =
 \end{aligned}$$

$$\begin{aligned}
&= \left( \frac{2x^4}{3 \cdot 4} - \frac{2x^3}{3 \cdot 3} \right) \Big|_1^2 + \left( -\frac{1x^4}{3 \cdot 4} + \frac{4x^3}{3 \cdot 3} \right) \Big|_2^4 = \\
&= \left( \frac{2 \cdot 2^4}{3 \cdot 4} - \frac{2 \cdot 2^3}{3 \cdot 3} - \frac{2 \cdot 1^4}{3 \cdot 4} + \frac{2 \cdot 1^3}{3 \cdot 3} \right) + \\
&\quad + \left( -\frac{1 \cdot 4^4}{3 \cdot 4} + \frac{4 \cdot 4^3}{3 \cdot 3} + \frac{1 \cdot 2^4}{3 \cdot 4} - \frac{4 \cdot 2^3}{3 \cdot 3} \right) = \\
&= \left( \frac{30}{12} - \frac{14}{9} \right) + \left( -\frac{240}{12} + \frac{224}{9} \right) = -\frac{210}{12} + \frac{210}{9} = \\
&= \frac{-3 \cdot 210 + 4 \cdot 210}{36} = \frac{210}{36} = \frac{35}{6} \Gamma \text{ЭВ}^2.
\end{aligned}$$

Таким образом, дисперсия данного распределения

$$D(x) = \overline{x^2} - \bar{x}^2 = \frac{35}{6} - \left( \frac{7}{3} \right)^2 = \frac{35 \cdot 3 - 49 \cdot 2}{18} = \frac{7}{18} \Gamma \text{ЭВ}^2.$$

Среднеквадратичное отклонение

$$\sigma = \sqrt{D(x)} = \sqrt{\frac{7}{18}} \Gamma \text{ЭВ} = \frac{\sqrt{14}}{6} \Gamma \text{ЭВ}.$$

Наиболее вероятное значение соответствует максимуму распределения, т.е. вершине треугольника. Таким образом,  $x_{\text{вер.}} = 2 \Gamma \text{ЭВ}$ .

Теперь определим ширину на полувывоте. Максимум плотности вероятности при  $x_{\text{вер.}} = 2 \Gamma \text{ЭВ}$  равен  $f_{\text{макс.}} = 2/3 \Gamma \text{ЭВ}^{-1}$ . Для определения левой границы полувывоты необходимо решить уравнение

$$\begin{aligned}
\frac{1}{2} f_{\text{макс.}} &= f(x_{\text{л.}}) = f_1(x_{\text{л.}}), \\
\frac{1}{2} \cdot \frac{2}{3} &= \frac{2}{3} x_{\text{л.}} - \frac{2}{3},
\end{aligned}$$

откуда получаем  $x_{\text{л.}} = 1.5 \Gamma \text{ЭВ}$ .

Для определения правой границы решаем уравнение

$$\begin{aligned} \frac{1}{2} f_{\text{макс.}} &= f(x_{\text{пр.}}) = f_2(x_{\text{пр.}}), \\ \frac{1}{2} \cdot \frac{2}{3} &= -\frac{1}{3} x_{\text{пр.}} + \frac{4}{3}, \end{aligned}$$

откуда получаем  $x_{\text{пр.}} = 3$  ГэВ.

Таким образом, ширина распределения на полувысоте

$$FWHM = x_{\text{пр.}} - x_{\text{л.}} = 3 - 1.5 = 1.5 \text{ ГэВ.}$$

Теперь определим медианное значение. Это можно сделать из уравнения

$$\int_{x_{\text{мин.}}}^{x_{\text{мед.}}} f(x) dx = \int_{x_{\text{мед.}}}^{x_{\text{макс.}}} f(x) dx = 0.5.$$

Под первой частью распределения ( $1 \text{ ГэВ} < x < 2 \text{ ГэВ}$ ) лежит  $1/3$  всех событий, под второй частью ( $2 \text{ ГэВ} < x < 4 \text{ ГэВ}$ ) лежат оставшиеся  $2/3$  событий. Поэтому медианное значение удобнее искать через интегрирование функции  $f_2(x)$ :

$$\begin{aligned} \int_{x_{\text{мед.}}}^{x_{\text{макс.}}} f_2(x) dx &= 0.5, \\ \int_{x_{\text{мед.}}}^4 \left( -\frac{1}{3}x + \frac{4}{3} \right) dx &= \left( -\frac{1}{3} \frac{x^2}{2} + \frac{4}{3}x \right) \Big|_{x_{\text{мед.}}}^4 = \frac{1}{2}, \\ -\frac{1}{3} \frac{4^2}{2} + \frac{4}{3}4 + \frac{1}{3} \frac{x_{\text{мед.}}^2}{2} - \frac{4}{3}x_{\text{мед.}} &= \frac{1}{2}, \\ x_{\text{мед.}}^2 - 8x_{\text{мед.}} + 13 &= 0, \\ x_{\text{мед.}} &= \frac{8 \pm \sqrt{12}}{2} = 4 \pm \sqrt{3}. \end{aligned}$$

Медианное значение должно лежать левее  $x_{\text{макс.}} = 4$  ГэВ, откуда получаем

$$x_{\text{мед.}} = 4 - \sqrt{3} = 2.27 \text{ ГэВ.}$$

## 2. Биномиальное распределение

Биномиальное распределение является дискретным распределением только с двумя исходами, один из которых обычно называют благоприятным, а второй – неблагоприятным. Пусть проводится  $N$  независимых испытаний, в каждом из которых случайное событие может появиться с вероятностью  $p$  (благоприятный исход). Тогда число появлений благоприятных событий  $n$  в данной серии испытаний имеет биномиальное распределение.

Допустим, имеется детектор элементарных частиц, срабатывающий с вероятностью  $p$ . Будем считать срабатывание детектора при пролете частицы через него благоприятным исходом, несрабатывание – неблагоприятным. Как вычислить вероятность того, что детектор сработает  $n$  раз в  $N$  случаях? Вероятность того, что детектор сработает в первом событии, равна  $p$ , а вероятность того, что детектор сработает в  $n$  первых случаях, равна

$$P_{\text{благ. исход}} = p_1 \cdot p_2 \cdot \dots \cdot p_{n-1} \cdot p_n = p^n.$$

Соответственно, в оставшихся случаях события должны пройти с неблагоприятным исходом, вероятность которого равна  $1 - p$ :

$$P_{\text{неблаг. исход}} = (1 - p_{n+1}) \cdot (1 - p_{n+2}) \times \dots \\ \times (1 - p_{N-1}) \cdot (1 - p_N) = (1 - p)^{N-n}.$$

Таким образом, вероятность того, что детектор первые  $n$  раз сработает, а оставшиеся разы не сработает, равна

$$P_{n,N} = p^n \cdot (1 - p)^{N-n}.$$

Однако такая комбинация срабатываний-несрабатываний не является единственно возможной. Нам важно, чтобы благоприятный исход случился  $n$  раз, и не имеет значения, в какой последовательности идут благоприятные и

неблагоприятные исходы. Для учета всех возможных комбинаций используется число сочетаний.

$$C_N^n = \frac{N!}{n!(N-n)!}.$$

Таким образом, получаем формулу биномиального распределения (примеры приведены на рис. 2.1):

$$P_n = C_N^n \cdot p^n \cdot (1-p)^{N-n}. \quad (2.1)$$

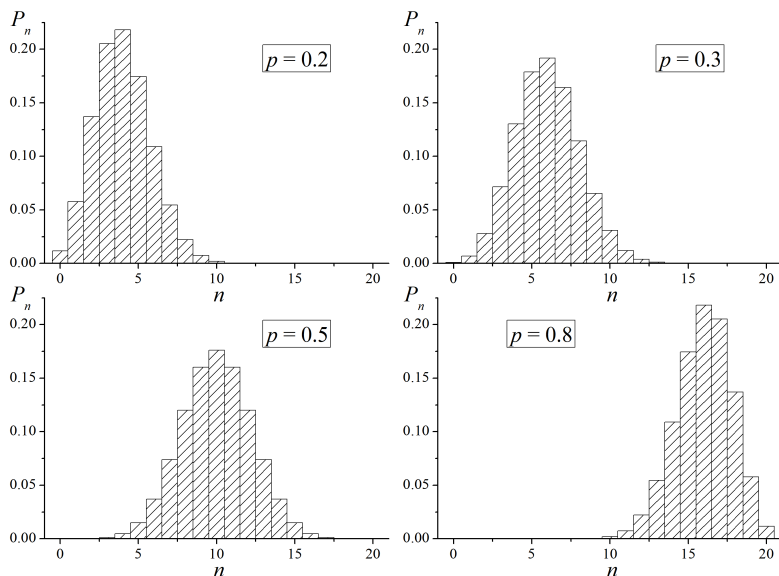


Рис. 2.1. Пример биномиального распределения для  $N = 20$

Биномиальное распределение является нормированным, для проверки этого можно воспользоваться биномом Ньютона:

$$(a + b)^N = \sum_{n=0}^{n=N} C_N^n a^n b^{N-n}.$$

Тогда можно записать, что:

$$\sum_{n=0}^{n=N} P_n = \sum_{n=0}^N C_N^n p^n (1-p)^{N-n} = (p + 1 - p)^N = 1,$$

т.е. биномиальное распределение нормировано.

Вычислим основные параметры биномиального распределения. Среднее значение (математическое ожидание):

$$\begin{aligned} E(n) = \bar{n} &= \sum_{n=0}^{n=N} n P_n = \sum_{n=1}^N n C_N^n p^n (1-p)^{N-n} = \\ &= \sum_{n=1}^{n=N} n \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} = \\ &= \sum_{n=1}^{n=N} \frac{N!}{(n-1)!(N-n)!} p^n (1-p)^{N-n} \implies \\ \stackrel{n-1=m}{\implies} & \sum_{m=0}^{m=N-1} \frac{N(N-1)!}{m!(N-m-1)!} p^{m+1} (1-p)^{N-m-1} \implies \\ \stackrel{N-1=M}{\implies} & \sum_{m=0}^{m=M} \frac{NM!}{m!(M-m)!} p^{m+1} (1-p)^{M-m} = \\ &= Np \sum_{m=0}^{m=M} C_M^m p^m (1-p)^{M-m} = Np. \end{aligned} \tag{2.2}$$

Вычислим дисперсию биномиального распределения, при этом воспользуемся формулой

$$D_n = \overline{n^2} - \bar{n}^2.$$

Второе слагаемое практически вычислено, осталось найти первое:

$$\begin{aligned}
 \overline{n^2} &= E(n^2) = \sum_{n=0}^{n=N} n^2 P_n = \\
 &= \sum_{n=1}^N n^2 \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} = \quad (2.3) \\
 &= \sum_{n=1}^N n \frac{N!}{(n-1)!(N-n)!} p^n (1-p)^{N-n} \implies \\
 &\stackrel{n-1=m}{\implies} \sum_{m=0}^{m=N-1} (m+1) \frac{N!}{m!(N-m-1)!} p^{m+1} (1-p)^{N-m-1} = \\
 &= Np \sum_{m=0}^{m=N-1} (m+1) \frac{(N-1)!}{m!(N-m-1)!} p^m (1-p)^{N-m-1} \implies \\
 &\stackrel{N-1=M}{\implies} Np \sum_{m=0}^{m=M} (m+1) \frac{M!}{m!(M-m)!} p^m (1-p)^{M-m} = \\
 &= Np \left( \sum_{m=0}^M m C_M^m p^m (1-p)^{M-m} + \sum_{m=0}^M C_M^m p^m (1-p)^{M-m} \right).
 \end{aligned}$$

Обе эти суммы нам уже встречались, первая – при подсчете среднего значения, вторая – при проверке нормировки. Стоит, однако, обратить внимание, что в первой сумме суммирование проводится до  $m = M$ . Таким образом, имеем

$$\overline{n^2} = Np(Mp + 1) = Np((N-1)p + 1).$$

Найдем дисперсию, подставив найденные величины из формул (2.2) и (2.3):

$$\begin{aligned}
 D &= Np((N-1)p + 1) - N^2 p^2 = \\
 &= N^2 p^2 - Np^2 + Np - N^2 p^2 = Np(1-p).
 \end{aligned}$$

Рассмотрим примеры использования биномиального распределения.

### Пример 2.1. Парадокс дней рождения

Для группы из 25 человек рассчитать ожидаемое количество пар людей с совпадающими днями рождения (предполагая, что дни рождения случайно равномерно распределены в течение года).

Вероятность того, что у произвольно выбранного человека сегодня день рождения равна  $p = 1/365$ . Вероятность того, что в группе имеются два человека, у которых сегодня день рождения равна

$$P_{\text{сегодня}} = C_{25}^2 p^2 (1-p)^{25-2}.$$

Однако дни рождения могут совпасть не только сегодня, но и в любой день года, число которых  $L = 1/p = 365$ .

$$\begin{aligned} N_{\text{пар.}} &= L \cdot P_{\text{сегодня}} = LC_{25}^2 p^2 (1-p)^{25-2} = \\ &= C_{25}^2 p^1 (1-p)^{23} = \frac{25!}{2!23!} p (1-p)^{23} = \\ &= \frac{25 \cdot 24}{2} \frac{1}{365} \left(1 - \frac{1}{365}\right)^{23} = 0.77. \end{aligned}$$

### Пример 2.2. Овербукинг

Для повышения эффективности работы многие авиакомпании пользуются овербукингом – продают на рейс немного больше билетов, чем количество посадочных мест на борту. Оценим вероятность того, что кому-нибудь из 158 пассажиров, купивших билет на рейс со 156 посадочными местами, не хватит места, если вероятность визита пассажира составляет  $p = 0.94$ .

Для авиакомпании неблагоприятной является ситуация, когда из 158 пассажиров, купивших билет, на рейс явятся 157 или 158. Вероятности этих событий равны

$$\begin{aligned}
P_{157} &= C_{158}^{157} p^{157} (1-p)^{158-157} = \frac{158!}{157!1!} (0.94)^{157} (1-0.94) = \\
&= 5.7 \cdot 10^{-4}, \\
P_{158} &= C_{158}^{158} p^{158} (1-p)^{158-158} = \frac{158!}{158!0!} (0.94)^{158} = \\
&= 0.6 \cdot 10^{-4}.
\end{aligned}$$

Таким образом, вероятность «неудобной» для авиакомпании ситуации возникает с вероятностью менее  $P < 0.001$ , и «лишние» пассажиры появляются только на одном из 1500 рейсов (на которые можно продать 3000 «лишних» билетов). Недовольство пассажира в таком случае компенсируется из полученной прибыли.

### **Пример 2.3. Расчет эффективности мюонного годоскопа УРАГАН**

Мюонный годоскоп УРАГАН состоит из 8 плоскостей газоразрядных трубок. Эффективность регистрации мюона каждой плоскостью составляет  $p = 0.95$ . Определить эффективность регистрации мюона годоскопом, если условием его срабатывания является срабатывание любых 4 плоскостей из 8.

Триггерное условие будет выполнено, если сработает не менее 4 плоскостей:

$$\begin{aligned}
P &= \sum_{n=4}^8 C_8^n p^n (1-p)^{8-n} = \frac{8!}{4!4!} 0.95^4 0.05^4 + \frac{8!}{5!3!} 0.95^5 0.05^3 + \\
&+ \frac{8!}{6!2!} 0.95^6 0.05^2 + \frac{8!}{7!1!} 0.95^7 0.05^1 + \frac{8!}{8!0!} 0.95^8 0.05^0 = 0.99998.
\end{aligned}$$

### Пример 2.4. Расчет эффективности супермодуля детектора ДЕКОР

Супермодуль детектора ДЕКОР также состоит из 8 плоскостей газоразрядных трубок. Эффективность регистрации мюона каждой плоскостью составляет  $p = 0.95$ . Однако, в отличие от годоскопа УРАГАН, триггерным условием срабатывания супермодуля является одновременное срабатывание как минимум двух четных и как минимум двух нечетных плоскостей. Определить эффективность регистрации мюона супермодулем ДЕКОР.

Чтобы выработался триггер супермодуля, необходимо выполнение условия как по четным, так и по нечетным плоскостям:

$$P = P_{\text{четн.}} = P_{\text{нечетн.}}$$

В супермодуле по четыре четных и нечетных плоскостей. Таким образом, группа считается сработавшей, если в ней сработали от 2 до 4 плоскостей:

$$\begin{aligned} P_{\text{четн.}} = P_{\text{нечетн.}} &= \sum_{n=2}^4 C_4^n p^n (1-p)^{4-n} = \frac{4!}{2!2!} 0.95^2 0.05^2 + \\ &+ \frac{4!}{3!1!} 0.95^3 0.05^1 + \frac{4!}{4!0!} 0.95^4 0.05^0 = 0.9995. \end{aligned}$$

Таким образом, общая эффективность супермодуля

$$P = 0.9995^2 = 0.999.$$

### Пример 2.5. Оценка эффективности детектора и ее погрешности

Детектор из  $N = 1000$  событий зарегистрировал  $n = 673$ . Определить эффективность детектора и оценить ее погрешность.

Эффективность детектора  $\varepsilon$  рассчитывается по формуле  $\varepsilon = n/N$  и является оценкой вероятности  $p$  биномиального распределения:

$$E(\varepsilon) = E(n/N) = E(n)/N = Np/N = p.$$

Для оценки погрешности эффективности необходимо вычислить ее дисперсию:

$$D(\varepsilon) = D\left(\frac{n}{N}\right) = \frac{D(n)}{N^2} = \frac{Np(1-p)}{N^2} = \frac{p(1-p)}{N} = \frac{\varepsilon(1-\varepsilon)}{N}.$$

Таким образом, погрешность эффективности

$$\sigma_\varepsilon = \sqrt{D(\varepsilon)} = \sqrt{\frac{\varepsilon(1-\varepsilon)}{N}}.$$

Таким образом, для нашего примера

$$\varepsilon = 673/1000 = 0.673, \\ \sigma_\varepsilon = \sqrt{0.673 \cdot (1 - 0.673)/1000} = 0.0148.$$

Получаем, что оцененная эффективность детектора равна  $67.3 \pm 1.5$  %.

### 3. Распределение Пуассона

Данное распределение впервые было исследовано в 1837 г. французским математиком, механиком и физиком С. Пуассоном (1781–1840) и является одним из важнейших дискретных распределений.

События, которые могут быть описаны данным распределением, подчиняются трем правилам пуассоновского потока событий:

- 1) события независимы друг от друга;
- 2) в малом промежутке времени  $\delta t$  может произойти не более одного события;
- 3) вероятность реализации события в малом промежутке времени пропорциональна величине этого промежутка времени  $P \sim \delta t$ , т.е.  $P_1 = \mu \delta t$ .

Три условия пуассоновского потока событий позволяют вывести распределение Пуассона. Будем ожидать события в пределах временного промежутка от 0 до  $t + \delta t$ , который можно разбить на два промежутка: «большой» – от 0 до  $t$  и бесконечно малый – от  $t$  до  $t + \delta t$  (рис. 3.1).

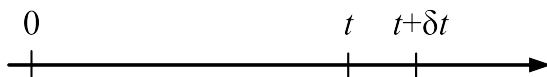


Рис. 3.1. Пояснение к выводу распределения Пуассона

Рассмотрим вероятность того, что на всем промежутке в промежутке от 0 до  $t + \delta t$  не произойдет ни одного события. Промежутки являются независимыми друг от друга,

поэтому вероятность отсутствия событий во всем промежутке от 0 до  $t + \delta t$  может быть представлена как произведение вероятностей отсутствия событий в обоих промежутках  $(0, t)$  и  $(t, \delta t)$ :

$$P_0(0, t + \delta t) = P_0(0, t) \cdot P_0(t, t + \delta t). \quad (3.1)$$

Вероятность того, что в промежутке  $(t; t + \delta t)$  не произойдет события, равна

$$P_0(t, t + \delta t) = 1 - \mu \delta t. \quad (3.2)$$

Подставляя (3.2) в (3.1), получаем

$$P_0(0, t + \delta t) = P_0(0, t) \cdot (1 - \mu \delta t).$$

После преобразования получаем

$$\frac{P_0(0, t + \delta t) - P_0(0, t)}{\delta t} = -\mu \cdot P_0(0, t).$$

Устремив  $\delta t$  к нулю ( $\delta t \rightarrow 0$ ), получаем дифференциальное уравнение

$$\frac{dP_0}{dt} = -\mu \cdot P_0.$$

Разделив переменные, получаем

$$\frac{dP_0}{P_0} = -\mu \cdot dt.$$

После интегрирования имеем

$$\ln P_0 = -\mu t + \ln C,$$

откуда получаем

$$P_0 = C \cdot e^{-\mu t}.$$

Константу  $C$  можно определить из условия, что в нулевой момент времени вероятность того, что еще ни одно событие не произошло, равна единице:

$$P_0(0, 0) = 1.$$

Откуда следует, что константа  $C = 1$ .

Тогда конечная формула того, что в промежутке  $(0; t)$  не произойдет ни одного события, равна

$$P_0 = e^{-\mu t}.$$

Теперь получим вероятность того, что в промежутке  $(0; t + \delta t)$  произойдет одно событие. Это может быть в двух случаях: либо событие произойдет в промежутке  $(0; t)$ , а в промежутке  $(t; t + \delta t)$  не произойдет; либо события не будет в промежутке  $(0; t)$ , а будет в промежутке  $(t; t + \delta t)$ . Итоговая вероятность определяется суммой вероятностей двух возможных вариантов:

$$P_1(0, t + \delta t) = P_0(0, t) \cdot P_1(t, t + \delta t) + P_1(0, t) \cdot P_0(t, t + \delta t).$$

Подставив вероятности для бесконечно малых промежутков времени, получаем

$$P_1(0, t + \delta t) = P_0(0, t) \cdot \mu \delta t + P_1(0, t) \cdot (1 - \mu \delta t).$$

Преобразуем к производной:

$$\frac{P_1(0, t + \delta t) - P_1(0, t)}{\delta t} = P_0(0, t) \cdot \mu - P_1(0, t) \cdot \mu.$$

Устремив  $\delta t$  к нулю, получаем дифференциальное уравнение

$$\frac{dP_1}{dt} = P_0 \cdot \mu - P_1 \cdot \mu.$$

Подставим известную вероятность  $P_0$ , получаем

$$\frac{dP_1}{dt} = \mu \cdot e^{-\mu t} - P_1 \cdot \mu. \quad (3.3)$$

Данное дифференциальное уравнение является неоднородным. Одним из способов его решения является метод варьирования постоянной. Для этого необходимо сначала получить решение для однородного уравнения

$$\frac{dP_1}{dt} = -P_1 \cdot \mu,$$

а именно:

$$P_1 = C \cdot e^{-\mu t}.$$

Для получения решения неоднородного дифференциального уравнения нужно предположить, что константа в решении однородного уравнения сама является функцией от времени:

$$P_1 = C(t) \cdot e^{-\mu t}.$$

Решение такого вида необходимо подставить в неоднородное уравнение (3.3), откуда получаем новое дифференциальное уравнение для функции  $C(t)$ :

$$C'(t) \cdot e^{-\mu t} - \mu C(t) \cdot e^{-\mu t} = \mu \cdot e^{-\mu t} - \mu C(t) \cdot e^{-\mu t}.$$

Сократив, получаем выражение

$$C'(t) = \mu,$$

откуда

$$C(t) = \mu \cdot t + C_1.$$

В начальный момент времени вероятность события равна нулю:

$$P_1(0, 0) = 0,$$

откуда следует, что константа  $C_1 = 0$ .

Тогда конечная формула того, что в промежутке  $(0; t)$  произойдет одно событие, равна

$$P_1 = \mu t \cdot e^{-\mu t}.$$

Теперь оценим вероятность того, что в промежутке от 0 до  $t + \delta t$  произойдет  $k$  событий. При этом необходимо учитывать, что согласно второму условию пуассоновского потока в бесконечно малом промежутке времени  $\delta t$  может произойти не более одного события. Таким образом, либо в интервале  $(0; t)$  мы увидим  $k - 1$  событие, и одно событие попадет в бесконечно малый интервал  $(t; t + \delta t)$ ; либо все  $k$  событий попадают в «большой» временной интервал  $(0; t)$ , а в бесконечно малом интервале  $(t; t + \delta t)$  не будет наблюдаться ни одного события.

Таким образом, получаем

$$P_k(0, t + \delta t) = P_{k-1}(0, t) \cdot P_1(t, t + \delta t) + P_k(0, t) \cdot P_0(t, t + \delta t).$$

Подставляем значения вероятностей для бесконечно малого интервала:

$$P_k(0, t + \delta t) = P_{k-1}(0, t) \cdot \mu \delta t + P_k(0, t) \cdot (1 - \mu \delta t).$$

Преобразовав, получаем

$$\frac{P_k(0, t + \delta t) - P_k(0, t)}{\delta t} = P_{k-1}(0, t) \cdot \mu - P_k(0, t) \cdot \mu.$$

Устремим  $\delta t$  к нулю, получаем неоднородное дифференциальное уравнение

$$\frac{dP_k}{dt} = P_{k-1} \cdot \mu - P_k \cdot \mu.$$

Решением однородного дифференциального уравнения

$$\frac{dP_k}{dt} = -P_k \cdot \mu$$

является функция

$$P_k = C_k \cdot e^{-\mu t}.$$

Для получения решения неоднородного дифференциального уравнения варьируем постоянную  $C_k$ :

$$P_k = C_k(t) \cdot e^{-\mu t}.$$

Подставляем данное решение и получаем уравнение для функции  $C_k(t)$ :

$$C'_k(t) \cdot e^{-\mu t} - \mu C_k(t) \cdot e^{-\mu t} = -\mu C_k(t) \cdot e^{-\mu t} + \mu \cdot P_{k-1}.$$

Сократив, получаем выражение

$$C'_k(t) = \mu P_{k-1} \cdot e^{\mu t}.$$

Поскольку уже известно выражение для вероятности одного события  $P_1$ , то можно найти функцию  $C_2(t)$ :

$$C'_2(t) = \mu P_1 \cdot e^{-\mu t}.$$

Проинтегрировав, получаем

$$C_2(t) = \frac{1}{2} \mu^2 t^2 + C_2^*.$$

Из начального условия ( $P_2(0) = 0$ ) получаем, что  $C_2^* = 0$ . Тогда формула для  $P_2$ :

$$P_2 = \frac{(\mu t)^2}{2!} \cdot e^{-\mu t}.$$

Найдем функцию  $C_3(t)$ :

$$C_3' = \mu \frac{(\mu t)^2}{2!} \cdot e^{\mu t} \cdot e^{-\mu t}.$$

Берем интеграл и получаем

$$C_3(t) = \frac{(\mu t)^3}{3!} + C_3^*.$$

Из начального условия ( $P_3(0) = 0$ ) получаем, что  $C_3^* = 0$ . Тогда формула для  $P_3$ :

$$P_3 = \frac{(\mu t)^3}{3!} \cdot e^{-\mu t}. \quad (3.4)$$

Применив математическую индукцию, можем записать выражение для вероятности при  $k = n$ :

$$P_n = \frac{(\mu t)^n}{n!} \cdot e^{-\mu t}. \quad (3.5)$$

Распределение Пуассона может быть получено из биномиального распределения. Для этого отрезок времени  $(0, t)$  надо разделить на большое количество малых отрезков. Тогда вероятность благоприятного события внутри малого отрезка может быть записана как

$$p = \mu \delta t = \lim_{N \rightarrow \infty} \frac{\mu t}{N}.$$

Рассмотрим, как в этом случае может быть преобразовано выражение для вероятности по биномиальному распределению:

$$\begin{aligned}
P_n &= \frac{N!}{n!(N-n)!} p^n (1-p)^{N-n} = \\
&= \lim_{N \rightarrow \infty} \frac{N!}{n!(N-n)!} \left(\frac{\mu t}{N}\right)^n \left(1 - \frac{\mu t}{N}\right)^{N-n} = \\
&= \lim_{N \rightarrow \infty} \frac{N(N-1)(N-2)\dots(N-n+1)}{n!} \left(\frac{\mu t}{N}\right)^n \frac{(1 - \frac{\mu t}{N})^N}{(1 - \frac{\mu t}{N})^n} = \\
&= \lim_{N \rightarrow \infty} \frac{N^n (1 - \frac{1}{N})(1 - \frac{2}{N})\dots(1 - \frac{n-1}{N})}{n!} \left(\frac{\mu t}{N}\right)^n \frac{(1 - \frac{\mu t}{N})^N}{(1 - \frac{\mu t}{N})^n} = \\
&= \frac{(\mu t)^n}{n!} \lim_{N \rightarrow \infty} \frac{(1 - \frac{1}{N})(1 - \frac{2}{N})\dots(1 - \frac{n-1}{N})}{(1 - \frac{\mu t}{N})^n} (1 - \frac{\mu t}{N})^N = \\
&= \frac{(\mu t)^n}{n!} \lim_{N \rightarrow \infty} \frac{(1 - \frac{1}{N})(1 - \frac{2}{N})\dots(1 - \frac{n-1}{N})}{(1 - \frac{\mu t}{N})^n} \cdot \lim_{N \rightarrow \infty} \left(1 - \frac{\mu t}{N}\right)^N.
\end{aligned}$$

Каждая из скобок первого предела стремится к единице, поэтому первый предел равен единице:

$$\begin{aligned}
P_n &= \frac{(\mu t)^n}{n!} \lim_{N \rightarrow \infty} \left(1 - \frac{\mu t}{N}\right)^N = \\
&= \frac{(\mu t)^n}{n!} \lim_{N \rightarrow \infty} \left( \left(1 - \frac{\mu t}{N}\right)^{-\frac{N}{\mu t}} \right)^{-\mu t}.
\end{aligned}$$

Далее воспользуемся вторым замечательным пределом:

$$\lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x = e.$$

Получаем выражение для распределения Пуассона:

$$P_n = \frac{(\mu t)^n}{n!} \lim_{N \rightarrow \infty} \left( \left(1 - \frac{\mu t}{N}\right)^{-\frac{N}{\mu t}} \right)^{-\mu t} = \frac{(\mu t)^n}{n!} e^{-\mu t}.$$

Данный вывод не является математически строгим, однако демонстрирует тесную связь биномиального распределения и распределения Пуассона.

Если временной интервал, в котором наблюдаются события, фиксирован, можно сделать замену  $\lambda = \mu t$  и получить распределение Пуассона (примеры для разных  $\lambda$  приведены на рис. 3.2):

$$P_n = \frac{\lambda^n}{n!} \cdot e^{-\lambda}. \quad (3.6)$$

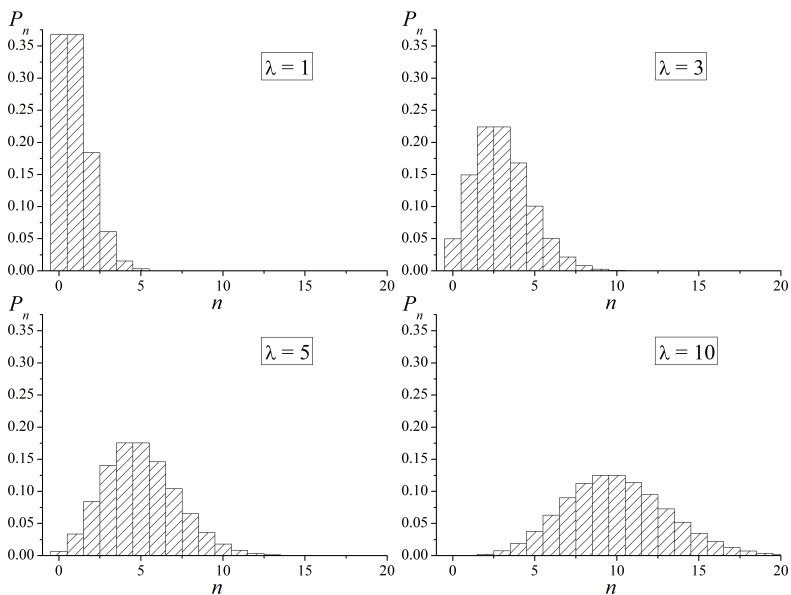


Рис. 3.2. Пример распределения Пуассона для разных  $\lambda$

Используя разложение экспоненты в ряд Маклорена:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!},$$

покажем, что распределение Пуассона является нормированным, т.е. сумма вероятностей равна единице:

$$\sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1. \quad (3.7)$$

Математическое ожидание (среднее значение) распределения (при проведении выкладок введем замену  $m = n - 1$ )

$$\begin{aligned} E(n) = \bar{n} &= \sum_{n=0}^{\infty} n P_n = \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \sum_{n=1}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1} \cdot \lambda}{(n-1)!} = e^{-\lambda} \cdot \lambda \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} = \\ &= e^{-\lambda} \cdot \lambda \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda} = \lambda. \end{aligned} \quad (3.8)$$

Дисперсию распределения Пуассона будем искать на основе среднего и среднеквадратичного значений:

$$D(n) = \overline{n^2} - \bar{n}^2. \quad (3.9)$$

Для этого посчитаем среднеквадратичное значение:

$$\begin{aligned} \overline{n^2} &= \sum_{n=0}^{\infty} n^2 P_n = \sum_{n=0}^{\infty} n^2 \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= e^{-\lambda} \sum_{n=0}^{\infty} n \frac{\lambda^n}{(n-1)!} = e^{-\lambda} \sum_{n=1}^{\infty} n \frac{\lambda^n}{(n-1)!} = \end{aligned} \quad (3.10)$$

Проведем замену  $m = n - 1$ :

$$= e^{-\lambda} \sum_{m=0}^{\infty} (m+1) \frac{\lambda^{m+1}}{m!} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} (m+1) \frac{\lambda^m}{m!} = \quad (3.11)$$

Раскроем скобки и разделим на две суммы:

$$= \lambda e^{-\lambda} \left( \sum_{m=0}^{\infty} m \frac{\lambda^m}{m!} + \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \right) = \quad (3.12)$$

Первая сумма нам уже известна по вычислению среднего значения, вторая сумма – это разложенная в ряд экспонента.

$$= \lambda e^{-\lambda}(\lambda e^{\lambda} + e^{\lambda}) = \lambda(\lambda + 1). \quad (3.13)$$

Подставляя (3.13) и (3.8) в (3.9), получаем:

$$D(n) = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Соответственно, среднеквадратичное отклонение для распределения Пуассона равно

$$\sigma = \sqrt{\lambda}. \quad (3.14)$$

Оценим наиболее вероятное значение  $n$ . Для этого нужно решить уравнение

$$\frac{dP_n}{dn} = \frac{d}{dn} \left( \frac{\lambda^n}{n!} e^{-\lambda} \right) = 0.$$

При такой постановке задачи возникает проблема взятия производной от дискретной функции  $n!$ . Эту проблему можно обойти, если использовать приближенное выражение для  $n!$  (формулу Стирлинга):

$$n! = \sqrt{2\pi n} \left( \frac{n}{e} \right)^n.$$

Таким образом, получаем

$$\frac{dP_n}{dn} = \frac{d}{dn} \left( \frac{(e\lambda)^n}{n^{n+0.5}} \frac{e^{-\lambda}}{\sqrt{2\pi}} \right) = \frac{e^{-\lambda}}{\sqrt{2\pi}} \frac{d}{dn} \left( \frac{(e\lambda)^n}{n^{n+0.5}} \right) = 0.$$

Для того чтобы взять производную от  $n^{n+0.5}$ , вспомним, как можно взять производную от функции  $f(x) = x^x$ :

$$\frac{df}{dx} = \frac{d}{dx}(x^x) = \frac{d}{dx}(e^{x \ln x}) = e^{x \ln x}(\ln x + 1) = x^x(\ln x + 1).$$

Таким образом, производная от  $n^{n+0.5}$ :

$$\begin{aligned} \frac{d}{dn}(n^{n+0.5}) &= \frac{d}{dn}(e^{(n+0.5)\ln n}) = \\ &= e^{(n+0.5)\ln n} \left( \ln n + \frac{n+0.5}{n} \right) = n^{n+0.5} \left( \ln n + 1 + \frac{0.5}{n} \right). \end{aligned}$$

Зная выражение для производной от  $n^{n+0.5}$ , можем вернуться к поиску наиболее вероятного  $n$ :

$$\begin{aligned} \frac{d}{dn} \left( \frac{(e\lambda)^n}{n^{n+0.5}} \right) &= \\ &= \frac{(e\lambda)^n \ln(e\lambda)n^{n+0.5} - (e\lambda)^n n^{n+0.5} \left( \ln n + 1 + \frac{0.5}{n} \right)}{n^{2n+1}} = 0. \end{aligned}$$

Сократив члены в числителе, получаем

$$\ln(e\lambda) - \left( \ln n + 1 + \frac{0.5}{n} \right) = 0,$$

откуда следует

$$\ln \lambda = \ln n + \frac{0.5}{n}.$$

После потенцирования получаем

$$\lambda = ne^{0.5/n}.$$

При достаточно больших  $n$  можно разложить экспоненту в ряд:

$$\lambda = n \left( 1 + \frac{0.5}{n} + \dots \right),$$

откуда получаем оценку на наиболее вероятное  $n$ :

$$n_{\text{вер.}} \approx \lambda - 0.5.$$

Данный вывод не претендует на абсолютную математическую строгость, но демонстрирует, что наиболее вероятным является ближайшее целое слева от среднего значения. Если же среднее значение само является целым числом  $\lambda = k$ , где  $k$  – целое, то максимальными являются вероятности при  $n = k$  и  $n = k - 1$ , которые равны друг другу. Действительно, в случае  $\lambda = k$ :

$$P_{n=k} = \frac{k^k}{k!} e^{-k} = \frac{k^{k-1}}{(k-1)!} e^{-k} = P_{n=k-1}.$$

Таким образом, можно утверждать, что если, например,  $\lambda = 10$ , то наиболее вероятными являются  $n = 9$  и  $n = 10$  и  $P_9 = P_{10}$ .

Оценим коэффициент асимметрии распределения Пуассона, который определяется так:

$$\gamma = \frac{E[(x - \bar{x})^3]}{\sigma^3}. \quad (3.15)$$

$$E[(x - \bar{x})^3] = \overline{(x - \bar{x})^3} = \bar{x}^3 - 3\bar{x}^2 \cdot \bar{x} + 3\bar{x} \cdot \bar{x}^2 - \bar{x}^3 = \bar{x}^3 - 3\bar{x}^2 \cdot \bar{x} + 2\bar{x}^3. \quad (3.16)$$

Для определения асимметрии распределения Пуассона необходимо вычислить  $\bar{n}^3$ :

$$\begin{aligned} \bar{n}^3 &= E[n^3] = \sum_{n=0}^{\infty} n^3 P_n = \sum_{n=0}^{\infty} n^3 \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= \lambda \cdot e^{-\lambda} \sum_{n=1}^{\infty} n^2 \frac{\lambda^{n-1}}{(n-1)!} = \end{aligned}$$

Проведем замену  $m = n - 1$ :

$$\begin{aligned} &= \lambda \cdot e^{-\lambda} \sum_{m=0}^{\infty} (m+1)^2 \frac{\lambda^m}{m!} = \\ &= \lambda \cdot e^{-\lambda} \left( \sum_{m=0}^{\infty} m^2 \frac{\lambda^m}{m!} + 2 \sum_{m=0}^{\infty} m \frac{\lambda^m}{m!} + \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \right). \end{aligned}$$

Такие суммы мы уже считали – (3.7), (3.8) и (3.10):

$$= \lambda \cdot e^{-\lambda} \cdot e^{\lambda}(\lambda(\lambda + 1) + 2\lambda + 1) = \lambda^3 + 3\lambda^2 + \lambda. \quad (3.17)$$

Подставляем (3.17) в (3.16):

$$E[(x - \bar{x})^3] = \lambda^3 + 3\lambda^2 + \lambda - 3\lambda(\lambda^2 + \lambda) + 2\lambda^3 = \lambda. \quad (3.18)$$

Подставляем (3.18) и (3.14) в (3.15), получаем коэффициент асимметрии для распределения Пуассона:

$$\gamma = 1/\sqrt{\lambda}.$$

При больших средних значениях коэффициент асимметрии распределения Пуассона стремится к нулю, распределение становится все более симметричным и по форме стремится к распределению Гаусса с  $\sigma = \sqrt{\lambda}$  (см. раздел 5).

Распределение Пуассона имеет положительный коэффициент асимметрии, т.е. распределение более вытянуто вправо. Такая асимметрия приводит к тому, что правая и левая погрешности распределения Пуассона при малых статистиках перестают быть равными друг другу и требуют отдельных вычислений.

Для распределения Гаусса в пределах  $\pm\sigma$  от среднего лежит 68.3 % событий, т.е. на каждый из хвостов распределения приходится примерно по 15.8 % событий. Можно воспользоваться этой вероятностью при оценке погрешностей при малых средних в распределении Пуассона.

Допустим, в эксперименте мы измеряли дискретную величину, отвечающую требованиям пуассоновского потока, и получили малое целое число  $k$ .

Для оценки правой погрешности представим, что истинное среднее значение измеряемой величины лежит правее измеренной величины  $k$  и равно  $\lambda_r$  (в общем случае истинное среднее значение – нецелое). Мы не знаем, чему равна величина  $\lambda_r$ , но по аналогии с распределением

Гаусса предполагаем, что суммарная вероятность наблюдения  $k, k - 1, k - 2, \dots 0$  событий должна составлять не более 15.8 %. Таким образом, получаем условие на правую границу погрешности:

$$\sum_{n=0}^{n=k} P_{n,\lambda_r} = 0.158.$$

По аналогии можно оценить левую границу  $- \lambda_l$ . В этом случае получаем следующее уравнение:

$$\sum_{n=k}^{\infty} P_{n,\lambda_l} = 0.158, \text{ либо } \sum_{n=0}^{k-1} P_{n,\lambda_l} = 0.842.$$

Оценим погрешность при измеренном  $k = 0$ . Левая граница диапазона при этом не вычисляется, а для  $\lambda_r$  получаем

$$\begin{aligned} \sum_{n=0}^{n=k} P_{n,\lambda_r} &= \sum_{n=0}^0 P_{n,\lambda_r} = 0.158, \\ e^{-\lambda_r} &= 0.158, \\ \lambda_r &= -\ln 0.158 = 1.845. \end{aligned}$$

Тогда

$$\sigma_r = \lambda_r - k = 1.845.$$

Правая погрешность при нулевой статистике имеет наиболее частое приложение при постановке пределов различных экзотических событий. Допустим, мы в течение года искали монополю Дирака, работая с детектором, площадь которого равна  $1 \text{ м}^2$ . Экспериментальный результат, который мы получили,  $N = 0$  (т.е. не зарегистрировали монополей). Тогда можно сказать, что поток частиц  $I = N/(St) < 1.845$  частицы/(год·м<sup>2</sup>). Это значение получено в доверительном интервале 68.3 %.

В большинстве случаев при постановке пределов для 0 зарегистрированных событий принято рассматривать 95 % доверительный интервал ( $\pm 2\sigma$ ):

$$e^{-\lambda_r} = \frac{100 \% - 95 \%}{2 \cdot 100 \%} = 0.025,$$

$$\lambda_r = -\ln 0.025 = 3.68,$$

$$\sigma_r = \lambda_r - k = 3.68.$$

Таким образом, для данного примера в доверительном интервале 95 % поток монополей составляет  $I < 3.68$  частицы/(год·м<sup>2</sup>).

Рассмотрим случай определения погрешностей при  $k = 1$ :

$$\sum_{n=0}^{n=k} P_{n,\lambda_r} = \sum_{n=0}^1 P_{n,\lambda_r} = 0.158,$$

$$e^{-\lambda_r} + \lambda_r \cdot e^{-\lambda_r} = 0.158,$$

$e^{-\lambda_r}(1 + \lambda_r) = 0.158$  – трансцендентное уравнение, которое может быть решено численно. Решая это уравнение, получаем  $\lambda_r \approx 3.305$ . Тогда  $\sigma_r = \lambda_r - k = 3.305 - 1 = 2.205$ .

Определим левую границу доверительного интервала:

$$\sum_{n=0}^{k-1} P_{n,\lambda_l} = \sum_{n=0}^{1-1=0} P_{n,\lambda_l} = 0.842,$$

$$e^{-\lambda_l} = 0.842,$$

$$\lambda_l = -\ln 0.842 \approx 0.172,$$

$$\sigma_l = k - \lambda_l = 1 - 0.172 = 0.828,$$

т.е. при  $k = 1$ :  $\sigma_r = 2.31$ ,  $\sigma_l = 0.83$ .

Теперь рассмотрим для  $k = 2$ :

$$\begin{aligned} \sum_{n=0}^{k-1} P_{n,\lambda_l} &= \sum_{n=0}^{2-1=1} P_{n,\lambda_l} = 0.842, \\ e^{-\lambda_l} + \lambda_l \cdot e^{-\lambda_l} &= 0.842, \\ \lambda_l &= 0.706, \\ \sigma_l = k - \lambda_l &= 2 - 0.706 = 1.294, \\ \sum_{n=0}^{n=k} P_{n,\lambda_r} &= \sum_{n=0}^2 P_{n,\lambda_r} = 0.158, \\ e^{-\lambda_r} + \lambda_r \cdot e^{-\lambda_r} + \frac{\lambda_r^2}{2} e^{-\lambda_r} &= 0.158, \\ \lambda_r &\approx 4.644, \\ \sigma_r = \lambda_r - k &= 4.644 - 2 = 2.644; \end{aligned}$$

т.е. при  $k = 2$ :  $\sigma_r = 2.64$ ,  $\sigma_l = 1.29$ .

Точные расчетные значения в доверительном интервале 68.3 % для  $k \leq 30$  представлены в табл. 3.1.

Таблица 3.1. Погрешности распределений Пуассона при малых статистиках для доверительного интервала 68.3 %

$k$	$\lambda_l$	$\lambda_r$	$\sigma_l$	$\sigma_r$
0	0	1.845	0	1.845
1	0.172	3.305	0.828	2.305
2	0.706	4.644	1.294	2.644
3	1.364	5.925	1.636	2.925
4	2.082	7.170	1.918	3.170
5	2.836	8.391	2.164	3.391
6	3.616	9.592	2.384	3.592
7	4.413	10.780	2.587	3.780
8	5.226	11.955	2.774	3.955
9	6.050	13.120	2.950	4.120
10	6.885	14.278	3.115	4.278
11	7.727	15.428	3.273	4.428
12	8.577	16.571	3.423	4.571
13	9.433	17.709	3.567	4.709
14	10.295	18.842	3.705	4.842
15	11.162	19.971	3.838	4.971
16	12.033	21.096	3.967	5.096
17	12.909	22.217	4.091	5.217
18	13.787	23.334	4.213	5.334
19	14.670	24.449	4.33	5.449
20	15.555	25.561	4.448	5.561
21	16.444	26.670	4.556	5.670
22	17.335	27.776	4.665	5.776
23	18.228	28.880	4.772	5.880
24	19.124	29.982	4.876	5.982
25	20.022	31.082	4.978	6.082
26	20.922	32.180	5.078	6.180
27	21.824	33.276	5.174	6.276
28	22.728	34.371	5.272	6.371
29	23.633	35.463	5.367	6.463
30	24.540	36.555	5.460	6.555

### Пример 3.1. Поиск распада протонов

В черенковском водном детекторе массой  $m = 10^6$  кг за один год экспозиции не было зарегистрировано ни одного распада протона. Оценить время жизни протона.

Если протон – распадающаяся частица, то число протонов со временем должно уменьшаться по закону радиоактивного распада:

$$N = N_0 \cdot e^{-t/\tau}.$$

В каждой молекуле воды содержится  $N_p = 10$  протонов. Количество протонов в 1000 т воды:

$$N_0 = \frac{N_A N_p m}{\mu} = 6 \cdot 10^{23} \frac{1}{\text{моль}} \cdot 10 \cdot \frac{10^9 \text{ г}}{18 \frac{\text{г}}{\text{моль}}} = 3.3 \cdot 10^{32} \text{ шт.}$$

Поскольку время жизни протона  $\tau$  во много раз превышает время экспозиции, экспонента может быть разложена в ряд:

$$e^{-t/\tau} = 1 - t/\tau.$$

Таким образом, из закона радиоактивного распада получаем уравнение

$$N/N_0 = 1 - t/\tau.$$

Поскольку в эксперименте не было зарегистрировано ни одного распада, при подстановке  $N = N_0$  получаем, что  $\tau \rightarrow \infty$ .

Однако из свойств распределения Пуассона известна правая граница 68 % доверительного интервала при  $k = 0$ ,  $\lambda_r = 1.845$ . Можем записать:

$$N = N_0 - \lambda_r = N_0 - 1.8.$$

Тогда  $(N_0 - 1.8)/N_0 = 1 - t/\tau$ , откуда получаем

$$\tau \geq tN_0/1.8 = 1.8 \cdot 10^{32} \text{ лет.}$$

### Пример 3.2. Расчет погрешности скорости счета детектора

Детектор за  $\tau = 10$  с зарегистрировал  $N = 141$  событие, удовлетворяющее условию пуассоновского потока. Определить скорость счета детектора и ее погрешность, пренебрегая погрешностью измерения времени.

Скорость счета детектора равна

$$\nu = N/\tau = 141/10 = 14.1 \text{ с}^{-1}.$$

Оценим погрешность, учитывая, что для величины, подчиняющейся распределению Пуассона,  $D(N) = N$ :

$$\begin{aligned}\sigma_\nu &= \sqrt{D(\nu)} = \sqrt{D(N/\tau)} = \sqrt{D(N)/\tau^2} = \sqrt{N/\tau^2} = \\ &= \sqrt{N}/\tau = 11.87/10 = 1.19 \text{ с}^{-1}.\end{aligned}$$

Таким образом, скорость счета детектора составляет  $14.1 \pm 1.2 \text{ с}^{-1}$ .

### Пример 3.3. Мертвое время детектора

Регистрирующая система комплекса установок НЕВОД-ДЕКОР-СКТ имеет мертвое время  $\tau = 3.3$  мс (данное время обусловлено протоколом передачи данных Ethernet). Определить, какая доля событий попадает в мертвое время, если суммарная скорость счета установок комплекса составляет  $\nu = 20 \text{ с}^{-1}$ .

Поток вторичных космических лучей, регистрируемых комплексом, является классическим примером пуассоновского потока событий.

События не будут зарегистрированы комплексом, если они попадают в мертвое время после предшествующего события. Таким образом, доля «потерянных» событий равна

$$P_{\text{потери}} = \sum_{n=1}^{\infty} P_n = 1 - P_0 = 1 - e^{-\nu\tau} = 1 - e^{-0.0033 \cdot 20} = 0.0639.$$

Таким образом, 6.4 % событий попадают в мертвое время.

### **Пример 3.4. Мертвое время системы «пинг-понг»**

Для снижения доли потерь систем с высокой нагрузкой используют так называемые «пинг-понг» системы, когда обрабатывающая электроника имеет второй канал, который полностью дублирует первый и работает вместо него, если первый занят обработкой. Первый канал системы, в свою очередь, может начать обработку события, если занят второй канал. Каналы попеременно переключаются и обрабатывают большее количество событий.

Например, спектрометрический тракт имеет мертвое время  $\tau = 50$  мкс и работает при нагрузке  $\nu = 2000$  с<sup>-1</sup>.

Согласно примеру 3.3, доля событий, попадающих в мертвое время канала, равна

$$P_{\text{потери}} = 1 - e^{-\nu\tau} = 1 - e^{-0.00005 \cdot 2000} = 0.095.$$

Однако, если поставить два канала, работающих в «пинг-понг» режиме, то система будет терять только второе и более событий, пришедших в мертвое время первого из сработавших каналов. Доля потерь в этом случае составит

$$\begin{aligned} P_{\text{потери}} &= \sum_{n=2}^{\infty} P_n = 1 - P_0 - P_1 = 1 - e^{-\nu\tau} - \nu\tau e^{-\nu\tau} = \\ &= 1 - e^{-\nu\tau}(1 + \nu\tau) = 1 - e^{-0.1}(1 + 0.1) = 0.0047. \end{aligned}$$

Таким образом, в режиме «пинг-понг» вместо 9.5 % событий будет теряться менее 0.5 %.

### **Пример 3.5. Учет мертвого времени**

Показания нейтронного детектора при размещении рядом с нейтронным генератором составляют  $\nu = 237$  нейтронов в секунду. Определить нагрузку электроники детектора, если мертвое время электроники детектора составляет  $\tau = 500$  мкс.

Поток нейтронов от генератора имеет пуассоновский характер, поэтому нейтроны могут приходить в детектор с малой разницей по времени, что приводит к просчетам. Скорость счета событий, которые электроника детектора успеваеет зарегистрировать, описывается формулой

$$v = v_0 e^{-v_0 \tau}.$$

Таким образом, задача восстановления «истинной» скорости счета сводится к решению этого уравнения. Уравнение может быть решено разложением экспоненты до первой или второй степени малости.

В первом приближении имеем

$$\begin{aligned} v e^{v_0 \tau} &= v_0, \\ v(1 + v_0 \tau + \dots) &= v_0, \\ v + v v_0 \tau &= v_0, \\ v_0 &= \frac{v}{1 - v \tau} = \frac{237 \text{ с}^{-1}}{1 - 237 \text{ с}^{-1} \cdot 0.0005 \text{ с}} = \\ &= \frac{237 \text{ с}^{-1}}{1 - 0.1185} = 269 \text{ с}^{-1}. \end{aligned}$$

Во втором приближении

$$\begin{aligned} v e^{v_0 \tau} &= v_0, \\ v \left( 1 + v_0 \tau + \frac{(v_0 \tau)^2}{2!} + \dots \right) &= v_0, \\ v + v v_0 \tau + 0.5 v v_0^2 \tau^2 &= v_0, \\ 0.5 v v_0^2 \tau^2 - v_0(1 - v \tau) + v &= 0, \\ v_0 &= \frac{(1 - v \tau) \pm \sqrt{(1 - v \tau)^2 - 2v^2 \tau^2}}{v \tau^2} = \end{aligned}$$

$$\begin{aligned}
&= \frac{(1 - 0.1185) \pm \sqrt{(1 - 0.1185)^2 - 2 \cdot 0.1185^2}}{0.1185 \cdot 0.0005 \text{ с}} = \\
&= \frac{0.8815 \pm 0.865423}{0.1185 \cdot 0.0005 \text{ с}}, \\
&\quad v_0^- = 271 \text{ с}^{-1}, \\
&\quad v_0^+ = 29484 \text{ с}^{-1}.
\end{aligned}$$

В качестве ответа принимается результат  $v_0 = 271 \text{ с}^{-1}$  как наиболее близкий к измеренной скорости счета нейтронов.

### **Пример 3.6. Режим одноэлектронной подсветки фотоумножителя**

Одним из наиболее надежных способов измерения коэффициента усиления диодной системы фотоумножителя является режим одноэлектронной подсветки. В данном режиме уровень подсветки фотокатода подбирается таким образом, чтобы вероятность выбивания двух и более электронов с фотокатода по отношению к вероятности выбивания одного электрона не превышала 5 %. Определить условия данного режима.

Если фотокатод подсвечивается стабильными вспышками светодиода или лазера, число выбитых фотоэлектронов подчиняется распределению Пуассона (формула 3.6). Основное условие одноэлектронного режима может быть записано как

$$\frac{P_{n>1}}{P_{n=1}} < 0.05.$$

Вероятность выбивания двух и более электронов может быть записана так:

$$P_{n>1} = \sum_{n=2}^{\infty} P_n = 1 - P_0 - P_1.$$

Таким образом, можно провести преобразования:

$$\frac{P_{n>1}}{P_{n=1}} = \frac{1 - P_0 - P_1}{P_1} = \frac{1 - e^{-\lambda} - \lambda e^{-\lambda}}{\lambda e^{-\lambda}} = \frac{e^{\lambda} - 1 - \lambda}{\lambda} < 0.05.$$

В режиме одноэлектронной подсветки  $\lambda$  много меньше единицы, и экспонента может быть разложена в ряд до третьего порядка:

$$e^{\lambda} = 1 + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots$$

Таким образом получаем квадратичное неравенство

$$\lambda^2 + 3\lambda - 0.3 < 0,$$

положительным решением которого является

$$\lambda < 0.097.$$

Однако эта величина не может быть измерена экспериментально, поскольку для этого необходимо знать коэффициент усиления диодной системы ФЭУ, который и должен определяться в режиме одноэлектронной подсветки. Для того чтобы выйти из замкнутого круга, используют эффективность срабатывания ФЭУ, которая равна

$$\varepsilon = 1 - P_0 = 1 - e^{-\lambda} = 1 - e^{-0.097} = 0.092.$$

Таким образом, основываясь на свойствах распределения Пуассона, можно утверждать, что если эффективность срабатывания ФЭУ на стабильную подсветку составляет менее 9.2 %, то примесь многоэлектронных событий по отношению к одноэлектронным не будет превышать 5 %.

## 4. Равномерное распределение

Простейшим распределением непрерывной случайной величины является равномерное распределение. Случайная величина, подчиняющаяся равномерному распределению, находится в интервале  $a \leq x \leq b$ , а функция плотности вероятности равна константе:

$$f(x) = C.$$

Значение константы легко получить из условия нормировки

$$\int_a^b C dx = C(b - a) = 1,$$

откуда получаем

$$f(x) = \frac{1}{b - a}.$$

Найдем среднее значение:

$$\begin{aligned} E(x) = \bar{x} &= \int_a^b x f(x) dx = \int_a^b \frac{x}{b - a} dx = \frac{1}{2} \frac{x^2}{b - a} \Big|_a^b = \\ &= \frac{1}{2} \frac{b^2 - a^2}{b - a} = \frac{b + a}{2}. \end{aligned}$$

Для того чтобы найти дисперсию, вычислим средне-квадратичное значение:

$$\begin{aligned} E(x^2) = \overline{x^2} &= \int_a^b x^2 f(x) dx = \int_a^b \frac{x^2}{b - a} dx = \frac{1}{3} \frac{x^3}{b - a} \Big|_a^b = \\ &= \frac{1}{3} \frac{b^3 - a^3}{b - a} = \frac{b^2 + ab + a^2}{3}. \end{aligned}$$

Теперь найдем дисперсию:

$$\begin{aligned} D &= \overline{x^2} - \bar{x}^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \\ &= \frac{4(b^2 + ab + a^2) - 3(b^2 + 2ab + a^2)}{12} = \frac{b^2 - 2ab + a^2}{12} = \\ &= \frac{(b-a)^2}{12}. \end{aligned}$$

Интегрируя плотность вероятности, получаем функцию вероятности:

$$F(x) = \int_a^x f(x)dx = \int_a^b \frac{1}{b-a} dx = \frac{x-a}{b-a}.$$

Вид функции вероятности представлен на рис. 4.1.

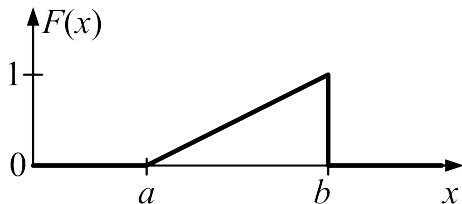


Рис. 4.1. Вид функции вероятности равномерного распределения

**Стандартным равномерным распределением** называют распределение при  $a = 0$  и  $b = 1$ , в этом случае функция плотности вероятности  $f(x) = 1$ . Стандартное равномерное распределение имеет огромное значение для методов математического моделирования, а случайную величину, подчиняющуюся этому распределению, часто называют **базовой случайной величиной**.

## 5. Распределение Гаусса

Распределение Гаусса, которое также называют *нормальным распределением* или *распределением Гаусса-Лапласа*, – это непрерывное распределение случайной величины, плотность вероятности которого в одномерном случае задается функцией Гаусса:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (5.1)$$

Одним из важных свойств функции Гаусса является невозможность выражения неопределенного интеграла от этой функции с помощью аналитических функций. Однако это не мешает проверке нормированности функции, поскольку при проверке интеграл берется в бесконечных пределах.

Поскольку параметр  $\mu$  не влияет на площадь под кривой, а лишь смещает кривую относительно оси ординат, проверку нормировки будем проводить при  $\mu = 0$  и в этом случае можем воспользоваться четностью функции ( $f(-x) = f(x)$ ):

$$\int_{-\infty}^{\infty} f(x) dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} dx = \frac{2}{\sigma\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx.$$

Чтобы взять интеграл

$$A = \int_0^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx,$$

домножим его на такой же, но по другой переменной:

$$\begin{aligned}
A^2 &= \int_0^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \int_0^{\infty} e^{-\frac{y^2}{2\sigma^2}} dy = \int_0^{\infty} \int_0^{\infty} e^{-\frac{x^2+y^2}{2\sigma^2}} dx dy = \\
&= | \text{замена } x = r \cos \varphi, y = r \sin \varphi, dx dy = r dr d\varphi | = \\
&= \int_0^{2\pi} d\varphi \int_0^{\infty} e^{-\frac{r^2}{2\sigma^2}} r dr = 2\pi \int_0^{\infty} e^{-\frac{r^2}{2\sigma^2}} r dr = \\
&= 2\pi\sigma^2 \int_0^{\infty} e^{-\frac{r^2}{2\sigma^2}} d\left(\frac{r^2}{2\sigma^2}\right) = -2\pi\sigma^2 e^{-\frac{r^2}{2\sigma^2}} \Big|_0^{\infty} = 2\pi\sigma^2.
\end{aligned}$$

Таким образом, получаем

$$A = \int_0^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx = \sigma\sqrt{2\pi},$$

откуда следует, что распределение Гаусса нормированно.

Определим математическое ожидание для распределения Гаусса:

$$\begin{aligned}
E(x) = \bar{x} &= \int_{-\infty}^{\infty} x f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\
&= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.
\end{aligned}$$

В первом интеграле сделаем замену  $t = x - \mu$ , в итоге под интегралом выходит нечетная функция, поэтому при симметричных пределах интегрирования в результате получаем нуль. Второй интеграл равен  $\mu$ , так как подынтегральная функция нормированна. Следовательно,

$$E(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} t e^{-\frac{t^2}{2\sigma^2}} dt + \mu = 0 + \mu = \mu.$$

Для вычисления дисперсии распределения Гаусса определим средний квадрат:

$$\begin{aligned}
 E(x^2) &= \overline{x^2} = \int_{-\infty}^{\infty} x^2 f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \\
 &+ \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} 2\mu x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx - \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mu^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \\
 &+ 2\mu \int_{-\infty}^{\infty} x f(x) dx - \mu^2 \int_{-\infty}^{\infty} f(x) dx.
 \end{aligned}$$

В первом интеграле сделаем замену  $t = x - \mu$ , подынтегральная функция – четная, поэтому можно заменить нижний предел интегрирования на 0 и домножить интеграл на 2. Второй интеграл – математическое ожидание, третий – нормировка.

$$\begin{aligned}
 E(x^2) &= \overline{x^2} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{t^2}{2\sigma^2}} dt + 2\mu^2 - \mu^2 = \\
 &= \frac{2}{\sigma\sqrt{2\pi}} \int_0^{\infty} t^2 e^{-\frac{t^2}{2\sigma^2}} dt + \mu^2.
 \end{aligned}$$

Теперь сделаем замену

$$y = \frac{t^2}{2\sigma^2}, \quad t = \sigma\sqrt{2y}, \quad dt = \frac{\sigma\sqrt{2}dy}{2\sqrt{y}}.$$

Получаем

$$\begin{aligned}
 E(x^2) &= \overline{x^2} = \frac{2}{\sigma\sqrt{2\pi}} \int_0^{\infty} 2\sigma^2 y e^{-y} \frac{\sigma\sqrt{2}dy}{2\sqrt{y}} + \mu^2 = \\
 &= \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} \sqrt{y} e^{-y} dy + \mu^2 = \frac{2\sigma^2}{\sqrt{\pi}} \int_0^{\infty} y^{3/2-1} e^{-y} dy + \mu^2 = \\
 &= \frac{2\sigma^2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) + \mu^2,
 \end{aligned}$$

где  $\Gamma(z)$  – гамма-функция,  $\Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}$ .

Таким образом, получаем значение для среднего квадрата распределения Гаусса:

$$E(x^2) = \overline{x^2} = \sigma^2 + \mu^2.$$

Теперь найдем дисперсию распределения Гаусса:

$$D = \overline{x^2} - \bar{x}^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2.$$

Таким образом, два главных параметра распределения Гаусса являются:  $\mu$  – средним значением,  $\sigma^2$  – дисперсией.

На рис. 5.1 в качестве примера приведены варианты распределения Гаусса с одинаковым средним значением  $\mu = 3$  и разными значениями дисперсии.

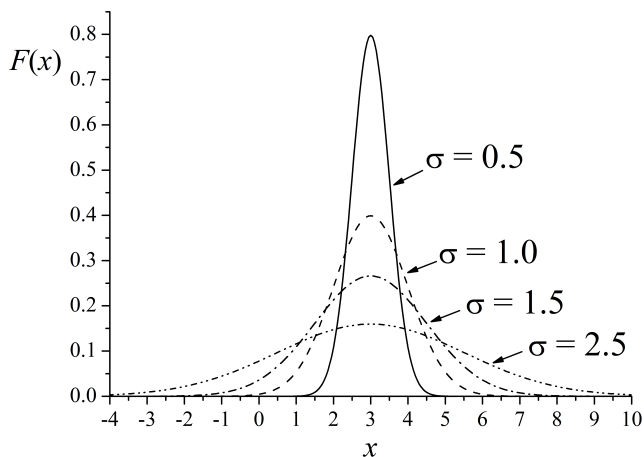


Рис. 5.1. Распределения Гаусса с одинаковым средним значением  $\mu = 3$  и разными значениями дисперсии

Определим ширину распределения Гаусса на полувысоте. Максимум распределения всегда наблюдается при  $x = \mu$ , этот максимум равен

$$f_{\text{макс.}} = f(x = \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{0}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}}.$$

Необходимо определить координаты, соответствующие левому и правому значению полувысоты:

$$\begin{aligned} f(x) &= \frac{f_{\text{макс.}}}{2}, \\ \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} &= \frac{1}{2} \frac{1}{\sigma\sqrt{2\pi}}, \\ e^{-\frac{(x-\mu)^2}{2\sigma^2}} &= \frac{1}{2}, \\ -\frac{(x-\mu)^2}{2\sigma^2} &= \ln \frac{1}{2}, \\ \frac{(x-\mu)^2}{2\sigma^2} &= \ln 2, \\ (x-\mu)^2 &= 2\sigma^2 \ln 2, \\ x &= \mu \pm \sigma\sqrt{2 \ln 2}. \end{aligned}$$

Таким образом, ширина распределения Гаусса на полувысоте

$$FWHM = x_{\text{пр.}} - x_{\text{л.}} = 2\sigma\sqrt{2 \ln 2} = 2.35\sigma.$$

Поскольку интеграл от функции Гаусса не может быть выражен через аналитические функции, функция вероятности распределения Гаусса

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

может быть рассчитана только численными методами. На рис. 5.2 приведены функции вероятности для плотностей вероятностей, представленных на рис. 5.1.

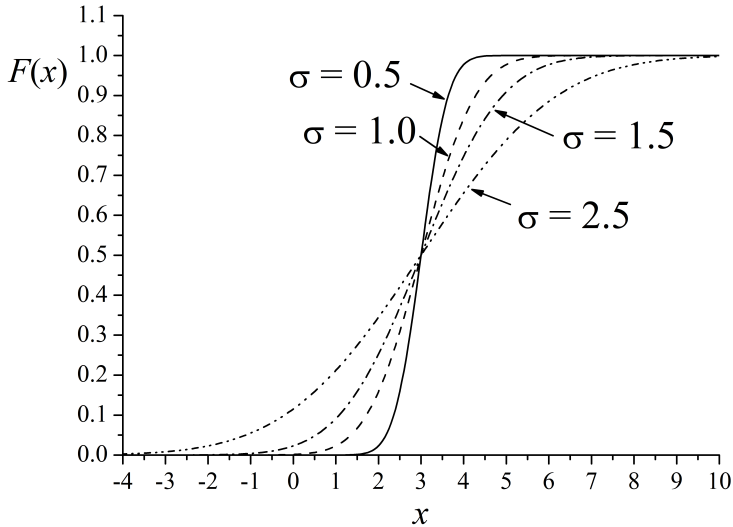


Рис. 5.2. Функции вероятности распределений Гаусса с одинаковым средним значением  $\mu = 3$  и разными значениями дисперсии

При анализе результатов измерений часто используют вероятность того, что нормальная случайная величина со средним значением  $\mu$  отклонится от него на величину  $n\sigma$ .

$$\begin{aligned}
 P(\mu - n\sigma < x < \mu + n\sigma) &= \int_{-n\sigma}^{n\sigma} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\
 &= F(\mu + n\sigma) - F(\mu - n\sigma).
 \end{aligned}$$

Значения вероятности для целых  $n$  от 1 до 6 приведены в табл. 5.1.

Таблица 5.1. Вероятность того, что распределенная по нормальному закону случайная величина со средним значением  $\mu$  отклонится от него на величину  $n\sigma$

$n$	$F(\mu + n\sigma) - F(\mu - n\sigma)$
1	0.682689492137
2	0.954499736104
3	0.997300203937
4	0.999936657516
5	0.999999426697
6	0.999999998027

### Пример 5.1

Предполагается, что отклик детектора при регистрации  $\alpha$ -частиц подчиняется распределению Гаусса, отношение дисперсии к среднему заряду постоянно и равно  $D(Q)/E(Q) = 0.28$  пКл. Определить, какой процент сигналов будет иметь заряд более 11 пКл, если средний сигнал равен  $E(Q) = 8$  пКл.

По условию мы имеем дело с распределением Гаусса со средним значением  $\mu = E(x) = 8$  пКл. Дисперсия распределения откликов детектора при таком среднем равна  $\sigma^2 = D(Q) = E(Q) \cdot 0.28$  пКл =  $8 \cdot 0.28$  пКл<sup>2</sup> = 2.24 пКл<sup>2</sup>. Откуда получаем, что среднеквадратичное отклонение равно  $\sigma = 1.497$  пКл.

Таким образом, величина заряда 11 пКл при  $\mu = 8$  пКл и  $\sigma = 1.497$  пКл является правой границей диапазона  $\mu \pm 2\sigma$ . Согласно табл. 5.1 за пределами этого диапазона лежит 4.56 % событий. Поскольку распределение Гаусса симметрично относительно среднего значения, правее границы 11 пКл ( $\mu + 2\sigma$ ) должно находиться 2.28 % сигналов.

## Пример 5.2. Определение эффективного динамического диапазона детектора

Спектрометрические тракты 12-х динодов фотоумножителей оптических модулей черенковского водного калориметра НЕВОД имеют динамический диапазон от 0 до 1000 фотоэлектронов (ф.э.). Оптические модули снабжены светодиодной системой мониторинга, спектр откликов ФЭУ на светодиодную подсветку имеет форму распределения Гаусса. Фотоумножители настроены таким образом, что отношение дисперсии к среднему заряду постоянно и равно  $D(Q)/E(Q) = 1.5$  ф.э. Определить максимальный средний сигнал, который может регистрироваться таким фотоумножителем без опасения выйти за динамический диапазон спектрометрического тракта.

Если средний сигнал на выходе ФЭУ будет иметь заряд 1000 ф.э., то половина сигналов будет превышать эту величину, таким образом, нужно подобрать величину среднего сигнала, не приводящую к «зашкаливанию».

С формальной точки зрения распределение Гаусса не имеет границ. Однако при проведении оценок принято в качестве пределов использовать границы диапазонов  $\mu \pm 3\sigma$  или  $\mu \pm 5\sigma$ . В первом случае за пределами диапазона расположено 0.27 % событий, а во втором – только 0.000006 % событий (см. табл. 5.1).

Среднеквадратичное отклонение зависит от среднего значения:

$$\begin{aligned}\sigma^2 &= D(Q) = E(Q) \cdot 1.5 \text{ ф.э.}, \\ \sigma &= \sqrt{1.5\mu}.\end{aligned}$$

Таким образом, для определения максимального среднего значения необходимо решить уравнение

$$\begin{aligned}
 \mu + n\sigma &= Q_{\text{макс.}}, \\
 \mu + n\sqrt{1.5\mu} &= Q_{\text{макс.}}, \\
 n\sqrt{1.5\mu} &= Q_{\text{макс.}} - \mu, \\
 1.5\mu n^2 &= (Q_{\text{макс.}} - \mu)^2, \\
 1.5\mu n^2 &= Q_{\text{макс.}}^2 - 2Q_{\text{макс.}}\mu + \mu^2, \\
 \mu^2 - (2Q_{\text{макс.}} + 1.5n^2)\mu + Q_{\text{макс.}}^2 &= 0, \\
 \mu &= \frac{(2Q_{\text{макс.}} + 1.5n^2) \pm \sqrt{6Q_{\text{макс.}}n^2 + 2.25n^4}}{2}.
 \end{aligned}$$

Ответ со знаком «+» перед квадратным корнем соответствует среднему значению, расположенному правее  $Q_{\text{макс.}}$ . Такое распределение, пересекающее предельную границу своим левым хвостом, неинтересно для нас. Будем использовать ответ со знаком «-» перед квадратным корнем.

Для диапазона « $3\sigma$ » получаем

$$\begin{aligned}
 \mu &= \frac{(2 \cdot 1000 + 1.5 \cdot 3^2) - \sqrt{6 \cdot 1000 \cdot 3^2 + 2.25 \cdot 3^4}}{2} = \\
 &= \frac{2013.5 - \sqrt{54182}}{2} = 890.4 \text{ ф.э.}
 \end{aligned}$$

При таком среднем значении среднеквадратичное отклонение будет равно

$$\sigma = \sqrt{1.5\mu} = \sqrt{1.5 \text{ ф.э.} \cdot 890.4 \text{ ф.э.}} = 36.5 \text{ ф.э.},$$

и граница  $\mu + 3\sigma$  пройдет по правой границе диапазона линейности спектрометрического тракта.

Для диапазона « $5\sigma$ » имеем

$$\begin{aligned}
 \mu &= \frac{(2 \cdot 1000 + 1.5 \cdot 5^2) - \sqrt{6 \cdot 1000 \cdot 5^2 + 2.25 \cdot 5^4}}{2} = \\
 &= \frac{2037.5 - \sqrt{151406}}{2} = 824.2 \text{ ф.э.}
 \end{aligned}$$

В этом случае среднеквадратичное отклонение равно

$$\sigma = \sqrt{1.5\mu} = \sqrt{1.5 \text{ ф.э.} \cdot 824.2 \text{ ф.э.}} = 35.2 \text{ ф.э.},$$

и граница  $\mu + 5\sigma$  пройдет по правой границе диапазона линейности спектрометрического тракта.

Таким образом, правая граница эффективного диапазона по разным оценкам находится в пределах 820–890 ф.э.

### **Пример 5.3. Оценка среднеквадратичного отклонения распределения**

Дано распределение фотоэлектронных умножителей ФЭУ-200 по коэффициенту сшивки диапазонов 12-го и 9-го динодов (табл. 5.2 и рис. 5.3). Оценить среднеквадратичное отклонение этого распределения.

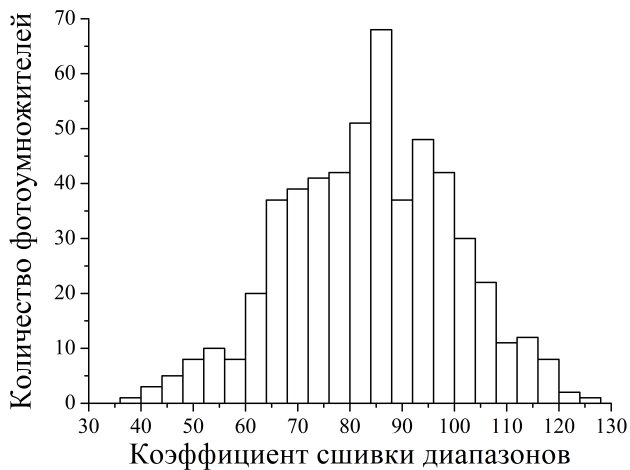


Рис. 5.3. Распределение фотоэлектронных умножителей ФЭУ-200 по коэффициенту сшивки диапазонов 12-го и 9-го динодов

Таблица 5.2. Распределение фотоэлектронных умножителей ФЭУ-200 по коэффициенту сшивки диапазонов 12-го и 9-го диодов

Коэффициент сшивки диапазонов	Количество фотоумножителей
34	0
38	1
42	3
46	5
50	8
54	10
58	8
62	20
66	37
70	39
74	41
78	42
82	51
86	68
90	37
94	48
98	42
102	30
106	22
110	11
114	12
118	8
122	2
126	1
130	0

Распределение по форме напоминает распределение Гаусса, для которого ширина на полувысоте связана со среднеквадратичным отклонением как

$$FWHM = 2.35\sigma.$$

Таким образом, можно быстро оценить среднеквадратичное отклонение. Как видно из таблицы, распределение имеет максимум при  $x_{\text{вер.}} = 86$ , при котором наблюдается  $N_{\text{макс.}} = 68$  ФЭУ. Полувысота распределения равна 34 ФЭУ. Левая граница полувысоты находится между значениями  $x_{\text{л.}} = 62 - 66$ , правая граница  $x_{\text{пр.}} = 98 - 102$ , т.е.  $FWHM = x_{\text{пр.}} - x_{\text{л.}} = 100 - 64 = 36$ .

При грубой оценке достаточно разделить ширину на полувысоте пополам:

$$\sigma = \frac{FWHM}{2} = \frac{36}{2} = 18,$$

или, более точно,

$$\sigma = \frac{FWHM}{2.35} = \frac{36}{2.35} = 15.3.$$

Отметим, что такая прикидочная оценка требует несколько очень простых арифметических действий и занимает очень мало времени.

Теперь проверим справедливость проведенной оценки. Определим среднее значение распределения:

$$\begin{aligned} \bar{x} &= \frac{\sum x_i N_i}{\sum N_i} = \\ &= \frac{38 \cdot 1 + 42 \cdot 3 + 46 \cdot 5 + \dots + 118 \cdot 8 + 122 \cdot 2 + 126 \cdot 1}{1 + 3 + 5 + \dots + 8 + 2 + 1} = \\ &= \frac{45792}{546} = 83.9. \end{aligned}$$

Среднеквадратичное значение:

$$\begin{aligned}\overline{x^2} &= \frac{\sum x_i^2 N_i}{\sum N_i} = \\ &= \frac{38^2 \cdot 1 + 42^2 \cdot 3 + 46^2 \cdot 5 + \dots + 122^2 \cdot 2 + 126^2 \cdot 1}{1 + 3 + 5 + \dots + 2 + 1} = 7291.3.\end{aligned}$$

Теперь можем рассчитать среднеквадратичное отклонение:

$$\sigma = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{7291.3 - 83.9^2} = 15.9.$$

Как видно, полученная оценка среднеквадратичного отклонения близка к расчетной. Этот метод часто используют для качественной проверки результатов расчета с помощью программных пакетов обработки данных, а также при оценке правильности данных, представляемых в докладах и отчетах. Однако стоит помнить, что если анализируемое распределение имеет отличную от нормального форму (имеется подложка, асимметрия, длинные хвосты и др.), то оцененное по ширине на полувысоте значение среднеквадратичного отклонения будет значительно меньше, чем значение, рассчитанное напрямую.

#### **Пример 5.4. Распределение Рэлея**

Ось пучка частиц проходит перпендикулярно плоскости детектора через начало координат. Распределение частиц относительно каждой из осей координат подчиняется распределению Гаусса со среднеквадратичным отклонением  $\sigma$ . Рассмотрим распределение расстояний от точки прохождения частицы до начала координат.

Поскольку ось пучка проходит через начало координат, распределения частиц по координатам  $x$  и  $y$  имеют вид

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, \quad -\infty < x < \infty,$$

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{y^2}{2\sigma^2}}, \quad -\infty < y < \infty.$$

Поскольку отклонение частиц пучка относительно каждой из координат независимы друг от друга, плотности вероятностей можно перемножить:

$$f(x, y) = f(x)f(y) = \frac{1}{\sigma^2 2\pi} e^{-\frac{x^2+y^2}{2\sigma^2}},$$

$$-\infty < x < \infty, \quad -\infty < y < \infty.$$

Точка на координатной плоскости может задаваться в полярных координатах  $(r, \varphi)$ :

$$x = r \cos \varphi,$$

$$y = r \sin \varphi,$$

$$r = \sqrt{x^2 + y^2}.$$

При замене системы координат в процессе интегрирования используется якобиан  $J_{r,\varphi} = r$ :

$$dx dy = J_{r,\varphi} dr d\varphi = r dr d\varphi,$$

$$f(r, \varphi) = \frac{dF(r, \varphi)}{dr d\varphi} = \frac{dF(x, y)}{dx dy} J_{r,\varphi} = f(x, y) J_{r,\varphi} =$$

$$= \frac{1}{\sigma^2 2\pi} e^{-\frac{x^2+y^2}{2\sigma^2}} r = \frac{r}{\sigma^2 2\pi} e^{-\frac{r^2}{2\sigma^2}}.$$

Поскольку пучок частиц аксиально симметричен, т.е. плотность вероятности по углу  $\varphi$  постоянна и равна  $f(\varphi) = \frac{1}{2\pi}$ , а координаты  $r$  и  $\varphi$  независимы друг от друга, то получаем

$$f(r, \varphi) = f(r)f(\varphi) = \frac{1}{2\pi} f(r).$$

Таким образом, получаем распределение частиц в пучке по расстояниям до начала координат, которое называется *распределением Рэля*:

$$f(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}, \quad 0 < r < \infty. \quad (5.2)$$

Параметр  $\sigma$  в распределении Рэля называют *параметром масштаба*.

Вид распределения Рэля для разных значений  $\sigma$  представлен на рис. 5.4.

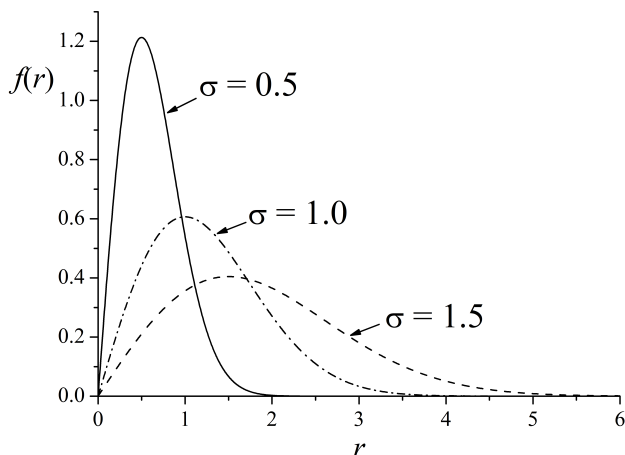


Рис. 5.4. Распределение Рэля для разных значений  $\sigma$

В отличие от распределения Гаусса, функция вероятности распределения Рэля выражается через аналитические функции:

$$\begin{aligned} F(r) &= \int_0^r f(r) dr = \int_0^r \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr = \int_0^r e^{-\frac{r^2}{2\sigma^2}} d\left(\frac{r^2}{2\sigma^2}\right) = \\ &= -e^{-\frac{r^2}{2\sigma^2}} \Big|_0^r = 1 - e^{-\frac{r^2}{2\sigma^2}}. \end{aligned}$$

По функции вероятности видно, что распределение Рэлея нормированно:  $F(r \rightarrow \infty) \rightarrow 1$ .

Также функция вероятности позволяет определить медиану распределения:

$$\begin{aligned} F(r_{\text{мед.}}) &= 0.5, \\ 1 - e^{-\frac{r_{\text{мед.}}^2}{2\sigma^2}} &= 0.5, \\ e^{-\frac{r_{\text{мед.}}^2}{2\sigma^2}} &= 0.5, \\ -\frac{r_{\text{мед.}}^2}{2\sigma^2} &= \ln \frac{1}{2}, \\ r_{\text{мед.}}^2 &= 2\sigma^2 \ln 2 = \sigma^2 \ln 4, \\ r_{\text{мед.}} &= \sigma \sqrt{\ln 4}. \end{aligned}$$

Найдем наиболее вероятное значение для распределения Рэлея:

$$\frac{df(r)}{dr} = 0,$$

$$\frac{d}{dr} \left( \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \right) = \frac{1}{\sigma^2} \left( e^{-\frac{r^2}{2\sigma^2}} + r e^{-\frac{r^2}{2\sigma^2}} \left( -\frac{r}{\sigma^2} \right) \right) = 0,$$

$$1 - \frac{r^2}{\sigma^2} = 0,$$

$$r_{\text{вер.}} = \sigma.$$

Рассчитаем среднее значение:

$$E(r) = \bar{r} = \int_0^{\infty} r f(r) dr = \int_0^{\infty} \frac{r^2}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr.$$

Делаем замену переменной

$$t = \frac{r^2}{2\sigma^2}, \quad r = \sigma \sqrt{2t}, \quad dr = \frac{\sigma}{\sqrt{2t}} dt.$$

Тогда

$$\begin{aligned} E(r) = \bar{r} &= \int_0^{\infty} \frac{2t\sigma^2}{\sigma^2} e^{-t} \frac{\sigma}{\sqrt{2t}} dt = \sigma\sqrt{2} \int_0^{\infty} \sqrt{t} e^{-t} dt = \\ &= \sigma\sqrt{2} \int_0^{\infty} t^{3/2-1} e^{-t} dt = \sigma\sqrt{2} \Gamma\left(\frac{3}{2}\right) = \sigma\sqrt{2} \frac{\sqrt{\pi}}{2} = \sigma\sqrt{\frac{\pi}{2}}. \end{aligned}$$

Среднеквадратичное значение:

$$E(r^2) = \overline{r^2} = \int_0^{\infty} r^2 f(r) dr = \int_0^{\infty} \frac{r^3}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} dr.$$

Используем ту же замену  $t = \frac{r^2}{2\sigma^2}$ :

$$E(r^2) = \overline{r^2} = \int_0^{\infty} \frac{\sigma^3(2t)^{3/2}}{\sigma^2} e^{-t} \frac{\sigma}{\sqrt{2t}} dt = 2\sigma^2 \int_0^{\infty} t e^{-t} dt = 2\sigma^2.$$

Таким образом, дисперсия распределения Рэлея

$$D(r) = \overline{r^2} - \bar{r}^2 = 2\sigma^2 - \sigma^2 \frac{\pi}{2} = \left(2 - \frac{\pi}{2}\right) \sigma^2.$$

## 6. Распределение Коши

В предыдущем разделе было показано, что ширина на полувысоте распределения Гаусса пропорциональна среднеквадратичному отклонению  $\sigma$ . Вследствие этого иногда формируется ошибочное представление, что среднеквадратичное отклонение способно охарактеризовать любое распределение. Продемонстрируем ошибочность этого представления.

Представим себе поливочную систему в виде равномерно вращающегося вокруг своей оси сопла, из которого по горизонтали бьет струя воды. Пусть на единичном расстоянии от поливочной системы находится бесконечно длинный забор. Найдем плотность вероятности попадания воды на различные участки забора.

Распределение воды по углам поворота сопла имеет равномерный характер. Будем откладывать угол поворота от перпендикуляра, соединяющего забор и поливочную систему. В этом случае вода попадает на различные участки забора при углах поворота сопла:

$$-\frac{\pi}{2} < \theta < \frac{\pi}{2}.$$

Будем учитывать только ту воду, которая бьет в сторону забора. В этом случае плотность вероятности равна

$$f(\theta) = \frac{dF}{d\theta} = \frac{1}{\pi}.$$

Плотность вероятности распределения воды по различным участкам забора может быть найдена как

$$f(x) = \frac{dF}{d\theta} \frac{d\theta}{dx}.$$

Поскольку поливочная система находится на единичном расстоянии от забора, координата участка забора  $x$  связана с углом поворота:

$$\begin{aligned}x &= \operatorname{tg} \theta, \\ \theta &= \operatorname{arctg} x, \quad -\infty < x < \infty, \\ \frac{d\theta}{dx} &= \frac{1}{1+x^2}.\end{aligned}$$

Таким образом, получаем распределение

$$f(x) = \frac{dF}{d\theta} \frac{d\theta}{dx} = \frac{1}{\pi} \frac{1}{1+x^2}. \quad (6.1)$$

Данное распределение называется *распределением Коши*, его вид представлен на рис. 6.1.

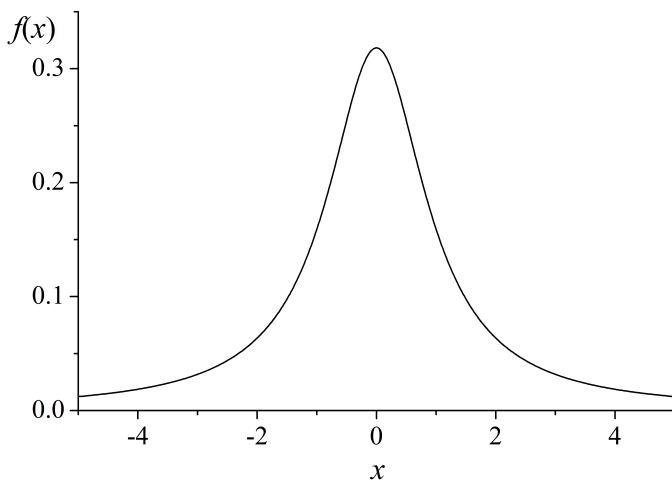


Рис. 6.1. Распределение Коши

Полученное распределение номированно:

$$\begin{aligned}\int_{-\infty}^{\infty} f(x)dx &= \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{1}{1+x^2} dx = \\ &= \frac{1}{\pi} \operatorname{arctg} x \Big|_{-\infty}^{\infty} = \frac{1}{\pi} \left( \frac{\pi}{2} + \frac{\pi}{2} \right) = 1.\end{aligned}$$

Поскольку распределение Коши является четной функцией, при подсчете среднего значения под интегралом появляется нечетная функция, и интегрирование в симметричных пределах дает нулевое среднее значение:

$$E(x) = \bar{x} = \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{x}{1+x^2} dx = 0.$$

Рассмотрим среднеквадратичное значение:

$$\begin{aligned}E(x^2) = \overline{x^2} &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_{-\infty}^{\infty} \frac{1}{\pi} \frac{x^2}{1+x^2} dx = \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1+x^2-1}{1+x^2} dx = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1+x^2}{1+x^2} dx - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{1+x^2} dx = \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} dx - 1 \rightarrow \infty.\end{aligned}$$

Среднеквадратичное значение для распределения Коши стремится к бесконечности.

Таким образом, распределение Коши является представителем распределений с бесконечной дисперсией.

## 7. Гамма-распределение

Ближайшим «родственником» гамма-распределения является гамма-функция. Вспомним некоторые основные ее свойства.

Гамма-функция имеет вид

$$\Gamma(z) = \int_0^{\infty} x^{z-1} e^{-x} dx. \quad (7.1)$$

Свойства гамма-функции:

1.  $\Gamma(z = 1) = 1$ .

$$\Gamma(1) = \int_0^{\infty} x^{1-1} e^{-x} dx = \int_0^{\infty} e^{-x} dx = -e^{-x} \Big|_0^{\infty} = e^0 - e^{-\infty} = 1.$$

2.  $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ .

$$\begin{aligned} \Gamma\left(\frac{1}{2}\right) &= \int_0^{\infty} x^{\frac{1}{2}-1} e^{-x} dx = \int_0^{\infty} \frac{e^{-x}}{\sqrt{x}} dx = \left| x = \frac{y^2}{2}, dx = y dy \right| = \\ &= \int_0^{\infty} \frac{e^{-y^2/2} y dy}{\sqrt{y^2/2}} = \sqrt{2} \int_0^{\infty} e^{-y^2/2} dy = \end{aligned}$$

(воспользуемся тем, что подынтегральная функция – четная)

$$= \frac{\sqrt{2}}{2} \int_{-\infty}^{\infty} e^{-y^2/2} dy = \frac{\sqrt{2}}{2} \sqrt{2\pi} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy =$$

(теперь подынтегральная функция – распределение Гаусса)

$$= \frac{\sqrt{2}}{2} \sqrt{2\pi} = \sqrt{\pi}.$$

$$3. \Gamma(z + 1) = z\Gamma(z).$$

$$\begin{aligned} \Gamma(z + 1) &= \int_0^{\infty} x^{z+1-1} e^{-x} dx = \int_0^{\infty} x^z e^{-x} dx = \\ |u = x^z, du = zx^{z-1} dx, dv = e^{-x} dx, v = -e^{-x}| &= \\ = uv|_0^{\infty} - \int_0^{\infty} v du = -x^z e^{-x}|_0^{\infty} + \int_0^{\infty} zx^{z-1} e^{-x} dx &= \\ = z \int_0^{\infty} x^{z-1} e^{-x} dx = z\Gamma(z). \end{aligned}$$

Таким образом, можно подсчитать гамма-функцию от любого полуцелого числа:

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2},$$

$$\Gamma\left(\frac{5}{2}\right) = \frac{3}{2} \cdot \Gamma\left(\frac{3}{2}\right) = \frac{3}{4}\sqrt{\pi},$$

$$\Gamma\left(\frac{1}{2} + n\right) = \frac{(2n)!}{4^n n!} \sqrt{\pi}.$$

$$4. \Gamma(n + 1) = n!, \text{ где } n - \text{ целое число.}$$

Действительно, как следует из предыдущего свойства:

$$\begin{aligned} \Gamma(n + 1) &= n\Gamma(n) = n(n - 1)\Gamma(n - 1) = \\ &= n(n - 1)(n - 2)\dots 1\Gamma(1) = n! \end{aligned}$$

Гамма-распределение имеет следующую плотность вероятности, зависящую от двух параметров  $\alpha$  и  $\beta$ :

$$f(x) = f(x, \alpha, \beta) = \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} x^{\alpha} e^{-x/\beta},$$

$$0 \leq x < \infty, \alpha > -1, \beta > 0.$$

Проверим нормировку гамма-распределения:

$$\begin{aligned}
 \int_0^{\infty} f(x, \alpha, \beta) dx &= \int_0^{\infty} \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} x^{\alpha} e^{-x/\beta} dx = \\
 &= \left| y = \frac{x}{\beta}, dx = \beta dy \right| = \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} \int_0^{\infty} (\beta y)^{\alpha} e^{-y} \beta dy = \\
 &= \frac{1}{\Gamma(\alpha + 1)} \int_0^{\infty} y^{\alpha} e^{-y} dy = \frac{1}{\Gamma(\alpha + 1)} \Gamma(\alpha + 1) = 1.
 \end{aligned}$$

Вычислим математическое ожидание:

$$\begin{aligned}
 E(x) = \bar{x} &= \int_0^{\infty} x f(x, \alpha, \beta) dx = \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} \int_0^{\infty} x x^{\alpha} e^{-x/\beta} dx = \\
 &= \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} \int_0^{\infty} x^{\alpha+1} e^{-x/\beta} dx = \\
 &= \frac{\Gamma(\alpha + 2)\beta^{\alpha+2}}{\Gamma(\alpha + 1)\beta^{\alpha+1}} \int_0^{\infty} \frac{1}{\Gamma(\alpha + 2)\beta^{\alpha+2}} x^{\alpha+1} e^{-x/\beta} dx = \\
 &= \frac{\Gamma(\alpha + 2)\beta}{\Gamma(\alpha + 1)} \int_0^{\infty} f(x, \alpha + 1, \beta) dx = \frac{\Gamma(\alpha + 2)\beta}{\Gamma(\alpha + 1)} = \\
 &= \frac{\Gamma(\alpha + 1)(\alpha + 1)\beta}{\Gamma(\alpha + 1)} = (\alpha + 1)\beta.
 \end{aligned}$$

Для того чтобы найти дисперсию, вычислим средне-квадратичное значение:

$$\begin{aligned}
 E(x^2) = \overline{x^2} &= \int_0^{\infty} x^2 f(x, \alpha, \beta) dx = \\
 &= \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} \int_0^{\infty} x^2 x^{\alpha} e^{-x/\beta} dx = \\
 &= \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} \int_0^{\infty} x^{\alpha+2} e^{-x/\beta} dx = \\
 &= \frac{\Gamma(\alpha + 3)\beta^{\alpha+3}}{\Gamma(\alpha + 1)\beta^{\alpha+1}} \int_0^{\infty} \frac{1}{\Gamma(\alpha + 3)\beta^{\alpha+3}} x^{\alpha+2} e^{-x/\beta} dx = \\
 &= \frac{\Gamma(\alpha + 3)\beta^2}{\Gamma(\alpha + 1)} \int_0^{\infty} f(x, \alpha + 2, \beta) dx = \frac{\Gamma(\alpha + 3)\beta^2}{\Gamma(\alpha + 1)} = \\
 &= \frac{\Gamma(\alpha + 1)(\alpha + 2)(\alpha + 1)\beta^2}{\Gamma(\alpha + 1)} = (\alpha + 2)(\alpha + 1)\beta^2.
 \end{aligned}$$

Теперь найдем дисперсию:

$$D = \overline{x^2} - \bar{x}^2 = (\alpha + 2)(\alpha + 1)\beta^2 - [(\alpha + 1)\beta]^2 = (\alpha + 1)\beta^2.$$

Вычислим наиболее вероятное значение для гамма-распределения, для этого приравняем к нулю производную:

$$\begin{aligned} \frac{df(x, \alpha, \beta)}{dx} &= \frac{d}{dx} \left( \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} x^\alpha e^{-x/\beta} \right) = \\ &= \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} \frac{d}{dx} (x^\alpha e^{-x/\beta}) = 0, \end{aligned}$$

$$\frac{d}{dx} (x^\alpha e^{-x/\beta}) = \alpha x^{\alpha-1} e^{-x/\beta} - \frac{1}{\beta} x^\alpha e^{-x/\beta} = 0.$$

Сократив, получаем

$$\alpha - \frac{1}{\beta} x = 0,$$

$$x_{\text{вер.}} = \alpha\beta.$$

Таким образом, наиболее вероятное значение для гамма-распределения на величину одной  $\beta$  меньше среднего значения этого распределения.

## 8. Производящая функция моментов

Знания среднего значения и среднеквадратичного отклонения недостаточно для полного описания распределения. Распределение может быть описано с помощью моментов, только если известен их полный набор  $\{\mu_i\}$  и они конечны. Такой набор моментов удобно получать и выражать через производящие функции моментов.

*Производящей функцией начальных моментов (ПФМ) непрерывной случайной величины  $x$*  называют математическое ожидание случайной величины  $e^{tx}$ , где  $t$  – действительный параметр:

$$M_x(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx. \quad (8.1)$$

*Производящей функцией начальных моментов (ПФМ) дискретной случайной величины  $n$*  называют математическое ожидание случайной величины  $e^{tn}$ :

$$M_n(t) = E(e^{tn}) = \sum_{n=0}^{\infty} e^{tn} P_n. \quad (8.2)$$

Производящая функция моментов позволяет находить все начальные моменты случайной величины  $x$ .

Разложим экспоненту из формулы (8.1) в ряд Маклорена:

$$\begin{aligned} M_x(t) &= E(e^{tx}) = E \left[ 1 + xt + \frac{(xt)^2}{2!} + \frac{(xt)^3}{3!} + \dots \right] = \\ &= \int_{-\infty}^{\infty} \left( 1 + xt + \frac{(xt)^2}{2!} + \frac{(xt)^3}{3!} + \dots \right) f(x) dx = \\ &= \int_{-\infty}^{\infty} f(x) dx + t \int_{-\infty}^{\infty} x f(x) dx + \frac{t^2}{2!} \int_{-\infty}^{\infty} x^2 f(x) dx + \end{aligned}$$

$$\begin{aligned}
& + \frac{t^3}{3!} \int_{-\infty}^{\infty} x^3 f(x) dx + \dots = \\
& = 1 + E(x)t + E(x^2) \frac{t^2}{2!} + E(x^3) \frac{t^3}{3!} + \dots = \\
& = 1 + \mu_1 t + \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \dots = \sum_{k=0}^{\infty} \mu_k \frac{t^k}{k!},
\end{aligned}$$

где  $\mu_k$  – начальный момент  $k$ -го порядка.

Таким образом, зная ПФМ, начальный момент  $k$ -го порядка может быть найден путем разложения ПФМ в ряд по степеням  $t$  и поиска коэффициента перед членом  $t^k/k!$  или путем дифференцирования:

$$\mu_k = E(x^k) = \left. \frac{\partial^k M_x}{\partial t^k} \right|_{t=0} = M_x^{(k)}(0). \quad (8.3)$$

Помимо производящей функции начальных моментов можно ввести производящую функцию центральных моментов, определив ее как математическое ожидание  $e^{t(x-\mu)}$ , где  $\mu$  – среднее значение для данного распределения:

$$N_x(t) = E(e^{t(x-\mu)}) = \int_{-\infty}^{\infty} e^{(x-\mu)t} f(x) dx,$$

$$N_n(t) = E(e^{t(n-\mu)}) = \sum_{n=0}^{\infty} e^{t(n-\mu)} P_n.$$

Производящая функция центральных моментов однозначно связана с производящей функцией начальных моментов:

$$N_x(t) = M_x(t) e^{-\mu t},$$

$$M_x(t) = N_x(t) e^{\mu t}.$$

Производящая функция центральных моментов позволяет вычислять центральные моменты:

$$\begin{aligned}
 N_x(t) &= E(e^{t(x-\mu)}) = \\
 &= E \left[ 1 + (x-\mu)t + \frac{((x-\mu)t)^2}{2!} + \frac{((x-\mu)t)^3}{3!} + \dots \right] = \\
 &= \int_{-\infty}^{\infty} \left( 1 + (x-\mu)t + \frac{((x-\mu)t)^2}{2!} + \frac{((x-\mu)t)^3}{3!} + \dots \right) f(x) dx = \\
 &= \int_{-\infty}^{\infty} f(x) dx + t \int_{-\infty}^{\infty} (x-\mu) f(x) dx + \frac{t^2}{2!} \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx + \\
 &\quad + \frac{t^3}{3!} \int_{-\infty}^{\infty} (x-\mu)^3 f(x) dx + \dots = 1 + E((x-\mu))t + \\
 &\quad + E((x-\mu)^2) \frac{t^2}{2!} + E((x-\mu)^3) \frac{t^3}{3!} + \dots = \\
 &= 1 + \nu_1 t + \nu_2 \frac{t^2}{2!} + \nu_3 \frac{t^3}{3!} + \dots = \sum_{k=0}^{\infty} \nu_k \frac{t^k}{k!}.
 \end{aligned}$$

Стоит обратить внимание на тот факт, что первый центральный момент всегда равен нулю:

$$\nu_1 = E(x - \mu) = E(x) - E(\mu) = \mu - \mu = 0.$$

Второй центральный момент всегда равен дисперсии распределения (если она конечна):

$$\nu_2 = E((x - \mu)^2) = D(x).$$

**Пример 8.1. ПФМ биномиального распределения**

$$\begin{aligned}
 M_n(t) &= E(e^{tn}) = \sum_{n=0}^{\infty} e^{tn} P_n = \sum_{n=0}^N e^{nt} C_N^n p^n (1-p)^{N-n} = \\
 &= \sum_{n=0}^N C_N^n (pe^t)^n (1-p)^{N-n} = \\
 &\text{(воспользуемся биномом Ньютона)} = (pe^t + 1 - p)^N.
 \end{aligned}$$

Зная ПФМ, найдем среднее значение и среднеквадратичное отклонение:

$$E(n) = \mu_1 = \left. \frac{\partial M_n}{\partial t} \right|_{t=0} = N(pe^t + 1 - p)^{N-1} pe^t \Big|_{t=0} = Np,$$

$$\begin{aligned} E(n^2) = \mu_2 &= \left. \frac{\partial^2 M_x}{\partial t^2} \right|_{t=0} = \left. \frac{\partial}{\partial t} \frac{\partial M_n}{\partial t} \right|_{t=0} = \\ &= \left. \frac{\partial}{\partial t} (N(pe^t + 1 - p)^{N-1} pe^t) \right|_{t=0} = \\ &= N(N-1)(pe^t + 1 - p)^{N-2} (pe^t)^2 + \\ &+ N(pe^t + 1 - p)^{N-1} pe^t \Big|_{t=0} = N(N-1)p^2 + Np. \end{aligned}$$

$$D(n) = \mu_2 - \mu_1^2 = Np(1-p).$$

### Пример 8.2. ПФМ распределения Пуассона

$$\begin{aligned} M_n(t) = E(e^{tn}) &= \sum_{n=0}^{\infty} e^{tn} P_n = \sum_{n=0}^{\infty} e^{nt} \frac{\lambda^n}{n!} e^{-\lambda} = \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)}. \end{aligned}$$

Получим среднее значение и дисперсию распределения Пуассона через ПФМ. При этом  $\mu_1$  – как коэффициент при слагаемом  $t$ , а  $\mu_2$  – как коэффициент при слагаемом  $t^2/2$ . Для этого сначала разложим в ряд экспоненту, которая входит в показатель степени, а затем уже разложим экспоненту, которая является основанием всей функции:

$$\begin{aligned} M_n(t) = e^{\lambda(e^t - 1)} &= e^{\lambda \left( \sum_{k=0}^{\infty} \frac{t^k}{k!} - 1 \right)} = e^{\lambda \left( t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots \right)} = \\ &= 1 + \left[ \lambda \left( t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots \right) \right] + \frac{1}{2!} \left[ \lambda \left( t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots \right) \right]^2 + \\ &+ \frac{1}{3!} \left[ \lambda \left( t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots \right) \right]^3 + \dots \end{aligned}$$

Несмотря на сложность полученного выражения, нетрудно заметить, что член с первой степенью  $t$  может быть получен только из раскрытия первой скобки, и коэффициент равен  $\mu_1 = \lambda$ .

Для получения  $\mu_2$  проанализируем, в каких скобках после возведения их в степень останется слагаемое  $t^2$ . Очевидно, что оно есть в первой скобке, и множитель перед  $t^2/2$  равен  $\lambda$ . В третьей скобке после возведения ее в куб не останется членов со степенями, меньшими трех, поэтому третья и более старшие скобки нас не интересуют. Таким образом, вклад во второй начальный момент может дать только вторая скобка, возведем ее в квадрат:

$$\frac{1}{2!} \left[ \lambda \left( t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots \right) \right]^2 = \frac{1}{2!} \left[ \lambda^2 \left( t^2 + \frac{t^4}{4} + \frac{t^6}{36} + \dots \right. \right. \\ \left. \left. + 2t \frac{t^2}{2!} + 2t \frac{t^3}{3!} + 2 \frac{t^2 t^3}{2! 3!} + \dots \right) \right].$$

После возведения скобки в квадрат видно, что единственным коэффициентом перед членом  $t^2/2$  является  $\lambda^2$ .

Складывая коэффициенты из первой и второй скобок, получаем

$$\mu_2 = \lambda + \lambda^2.$$

После чего можно получить уже знакомое выражение для дисперсии распределения Пуассона:

$$D(n) = \mu_2 - \mu_1^2 = \lambda.$$

### Пример 8.3. ПФМ равномерного распределения

$$M_x(t) = E(e^{xt}) = \int_a^b e^{xt} f(x) dx = \frac{1}{b-a} \int_a^b e^{xt} dx = \\ = \frac{1}{b-a} \frac{1}{t} e^{xt} \Big|_a^b = \frac{e^{bt} - e^{at}}{t(b-a)}.$$

Разложив экспоненты в ряд, получаем:

$$\begin{aligned}
 M_x(t) &= \frac{\sum_{n=0}^{\infty} \frac{(bt)^n}{n!} - \sum_{n=0}^{\infty} \frac{(at)^n}{n!}}{t(b-a)} = \frac{\sum_{n=1}^{\infty} \frac{(b^n - a^n)t^n}{n!}}{t(b-a)} = \\
 &= \sum_{n=1}^{\infty} \frac{(b^n - a^n)t^{n-1}}{(b-a)n(n-1)!} = (\text{введем замену } k = n - 1) = \\
 &= \sum_{k=0}^{\infty} \frac{(b^{k+1} - a^{k+1})t^k}{(b-a)(k+1)k!}.
 \end{aligned}$$

Коэффициент, стоящий перед членом  $t^k/k!$ , является начальным моментом  $k$ -го порядка.

Таким образом, для равномерного распределения имеем

$$E(x^k) = \mu_k = \frac{b^{k+1} - a^{k+1}}{(b-a)(k+1)}.$$

#### Пример 8.4. ПФМ распределения Гаусса

$$\begin{aligned}
 M_x(t) = E(e^{tx}) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{xt} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{xt - \frac{(x-\mu)^2}{2\sigma^2}} dx.
 \end{aligned}$$

Для того чтобы взять этот интеграл, необходимо выделить полный квадрат в степени экспоненты:

$$\begin{aligned}
 xt - \frac{(x-\mu)^2}{2\sigma^2} &= \frac{2xt\sigma^2 - (x-\mu)^2}{2\sigma^2} = \\
 &= \frac{2xt\sigma^2 - x^2 + 2x\mu - \mu^2}{2\sigma^2} = \frac{-x^2 + 2x(\mu + t\sigma^2) - \mu^2}{2\sigma^2} = \\
 &= \frac{[-x^2 + 2x(\mu + t\sigma^2) - (\mu + t\sigma^2)^2] + (\mu + t\sigma^2)^2 - \mu^2}{2\sigma^2} = \\
 &= \frac{-(x - \mu - t\sigma^2)^2 + (\mu + t\sigma^2)^2 - \mu^2}{2\sigma^2} = \\
 &= -\frac{(x - \mu - t\sigma^2)^2}{2\sigma^2} + \mu t + \frac{t^2\sigma^2}{2}.
 \end{aligned}$$

Таким образом, с учетом проведенных преобразований, можем записать:

$$\begin{aligned}
 M_x(t) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu-t\sigma^2)^2}{2\sigma^2} + \mu t + \frac{(t\sigma)^2}{2}} dx = \\
 &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu-t\sigma^2)^2}{2\sigma^2}} e^{\mu t + \frac{(t\sigma)^2}{2}} dx = \\
 &= e^{\mu t + \frac{(t\sigma)^2}{2}} \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu-t\sigma^2)^2}{2\sigma^2}} dx = e^{\mu t + \frac{(t\sigma)^2}{2}}.
 \end{aligned}$$

Зная ПФМ, найдем среднее значение и среднеквадратичное отклонение:

$$\begin{aligned}
 E(x) = \mu_1 &= \left. \frac{\partial M_x}{\partial t} \right|_{t=0} = e^{\mu t + \frac{(t\sigma)^2}{2}} (\mu + t\sigma^2) \Big|_{t=0} = \mu, \\
 E(x^2) = \mu_2 &= \left. \frac{\partial^2 M_x}{\partial t^2} \right|_{t=0} = \left. \frac{\partial}{\partial t} \frac{\partial M_x}{\partial t} \right|_{t=0} = \\
 &= \left. \frac{\partial}{\partial t} \left( e^{\mu t + \frac{(t\sigma)^2}{2}} (\mu + t\sigma^2) \right) \right|_{t=0} = \\
 &= e^{\mu t + \frac{(t\sigma)^2}{2}} (\mu + t\sigma^2)^2 + e^{\mu t + \frac{(t\sigma)^2}{2}} \sigma^2 \Big|_{t=0} = \mu^2 + \sigma^2, \\
 D(x) &= \mu_2 - \mu_1^2 = \sigma^2.
 \end{aligned}$$

На основе производящей функции начальных моментов можно записать производящую функцию центральных моментов распределения Гаусса:

$$\begin{aligned}
 N_x(t) &= M_x(t) e^{-\mu t} = e^{\mu t + \frac{(t\sigma)^2}{2}} e^{-\mu t} = e^{\frac{(t\sigma)^2}{2}} = \sum_{k=0}^{\infty} \frac{\left(\frac{(t\sigma)^2}{2}\right)^k}{k!} = \\
 &= 1 + \sigma^2 \frac{t^2}{2} + \sigma^4 \frac{t^4}{2^2 2!} + \sigma^6 \frac{t^6}{2^3 3!} + \dots + \sigma^{2k} \frac{t^{2k}}{2^k k!}.
 \end{aligned}$$

Обратим внимание на то, что в составе ряда отсутствуют члены с нечетными степенями  $t$ , т.е. все нечетные центральные моменты распределения Гаусса равны нулю:

$$\nu_{2n+1} = 0.$$

В свою очередь, любой четный центральный момент распределения Гаусса может быть вычислен по формуле:

$$\nu_{2n} = \frac{(2n)!}{2^n n!} \sigma^{2n}.$$

### Пример 8.5. ПФМ гамма-распределения

$$\begin{aligned} M_x(t) &= E(e^{xt}) = \int_0^{\infty} e^{xt} f(x, \alpha, \beta) dx = \\ &= \frac{1}{\Gamma(\alpha + 1) \beta^{\alpha+1}} \int_0^{\infty} e^{xt} x^{\alpha} e^{-x/\beta} dx = \\ &= \frac{1}{\Gamma(\alpha + 1) \beta^{\alpha+1}} \int_0^{\infty} x^{\alpha} e^{-x/\beta + xt} dx. \end{aligned}$$

Преобразуем показатель у экспоненты:

$$\begin{aligned} -\frac{x}{\beta} + xt &= -x \left( \frac{1}{\beta} - t \right) = -x \frac{1 - \beta t}{\beta} = -\frac{x}{\beta_2}, \\ \text{где } \beta_2 &= \frac{\beta}{1 - \beta t}. \end{aligned}$$

Теперь подставляем  $\beta_2$  в выражение для ПФМ и получаем

$$\begin{aligned} M_x(t) &= \frac{1}{\Gamma(\alpha + 1) \beta^{\alpha+1}} \int_0^{\infty} x^{\alpha} e^{-x/\beta + xt} dx = \\ &= \frac{1}{\Gamma(\alpha + 1) \beta^{\alpha+1}} \int_0^{\infty} x^{\alpha} e^{-x/\beta_2} dx = \\ &= \frac{\Gamma(\alpha + 1) \beta_2^{\alpha+1}}{\Gamma(\alpha + 1) \beta^{\alpha+1}} \int_0^{\infty} \frac{1}{\Gamma(\alpha + 1) \beta_2^{\alpha+1}} x^{\alpha} e^{-x/\beta_2} dx = \\ &= \frac{\beta_2^{\alpha+1}}{\beta^{\alpha+1}} \int_0^{\infty} f(x, \alpha, \beta_2) dx = \frac{\left( \frac{\beta}{1 - \beta t} \right)^{\alpha+1}}{\beta^{\alpha+1}} = (1 - \beta t)^{-\alpha-1}. \end{aligned}$$

Зная ПФМ, найдем среднее значение и среднеквадратичное отклонение:

$$E(x) = \mu_1 = \left. \frac{\partial M_x}{\partial t} \right|_{t=0} = (-\alpha - 1)(1 - \beta t)^{-\alpha-2}(-\beta)|_{t=0} = (\alpha + 1)\beta,$$

$$\begin{aligned} E(x^2) = \mu_2 &= \left. \frac{\partial^2 M_x}{\partial t^2} \right|_{t=0} = \left. \frac{\partial}{\partial t} \frac{\partial M_x}{\partial t} \right|_{t=0} = \\ &= \left. \frac{\partial}{\partial t} (\beta(\alpha + 1)(1 - \beta t)^{-\alpha-2}) \right|_{t=0} = \\ &= \beta^2(\alpha + 1)(\alpha + 2)(1 - \beta t)^{-\alpha-3}|_{t=0} = \beta^2(\alpha + 1)(\alpha + 2), \end{aligned}$$

$$D(x) = \mu_2 - \mu_1^2 = (\alpha + 1)\beta^2.$$

## 9. Распределение $\chi^2$

При фитировании функциональных зависимостей методом наименьших квадратов возникает вопрос, какому распределению подчиняется случайная величина

$$u = \sum_{i=1}^k x_i^2,$$

если  $x_i$  подчиняются распределению Гаусса. Рассмотрим этот вопрос.

Для начала выясним, какому распределению подчиняется квадрат случайной величины, распределенной по Гауссу. Для решения этой задачи удобно использовать обратную функцию, т.е. введем функцию  $h = x^2$ , тогда обратная функция будет иметь вид  $x = \pm\sqrt{h}$ . Тогда функция распределения величины  $h$ :

$$g(h)dh = f(x(h)) \left| \frac{dx}{dh} \right| dh.$$

Производная  $\frac{dx}{dh} = \frac{1}{2\sqrt{h}}$ . Таким образом, функция распределения величины  $h$  примет следующий вид:

$$g(h)dh = f(\sqrt{h})\frac{1}{2\sqrt{h}}dh + f(-\sqrt{h})\frac{1}{2\sqrt{h}}dh. \quad (9.1)$$

Здесь функция  $f$  – это функция распределения Гаусса, куда вместо величины  $x$  подставляется величина  $\sqrt{h}$ . Так как в распределении Гаусса переменная  $x$  возводится в квадрат, получаем следующее соотношение:

$$g(h) = \frac{1}{\sqrt{2\pi}}e^{-h/2}\frac{1}{2\sqrt{h}}2 = \frac{1}{\sqrt{2\pi}}h^{-1/2}e^{-h/2}. \quad (9.2)$$

Теперь соотнесем функцию  $g(h)$  и гамма-функцию и увидим, что функция  $g(h)$  относится к семейству гамма-функций с  $\alpha = -1/2$ ,  $\beta = 2$ :

$$\begin{aligned} g(h) &= \frac{1}{\sqrt{2\pi}} h^{-1/2} e^{-h/2} = \frac{1}{\sqrt{\pi} \cdot 2^{1/2}} h^{-1/2} e^{-h/2} = \\ &= \frac{1}{\Gamma\left(-\frac{1}{2} + 1\right) \cdot 2^{1/2}} h^{-1/2} e^{-h/2} = f\left(x, \alpha = -\frac{1}{2}, \beta = 2\right). \end{aligned}$$

Теперь рассмотрим производящую функцию моментов для суммы случайных величин  $\sum_{i=1}^k h_i$ . Она выглядит как  $k$ -мерный интеграл:

$$\begin{aligned} M_{\sum h_i} &= \iiint e^{t \sum h_i} g(h_1) g(h_2) \dots g(h_k) dh_1 dh_2 \dots dh_k = \\ &= \iiint e^{t \sum h_i} \prod_{i=1}^k g(h_i) dh_1 dh_2 \dots dh_k = \\ &= \iiint \prod_{i=1}^k e^{th_i} g(h_i) dh_1 dh_2 \dots dh_k = \prod_{i=1}^k \int e^{th_i} g(h_i) dh_i. \end{aligned}$$

Если все случайные величины в сумме подчиняются одному и тому же распределению, то производящая функция моментов суммы случайных величин равна возведенной в степень  $k$  производящей функции моментов случайной величины:

$$M_{\sum h_i} = |M_h(t)|^k.$$

Из примера 8.4 известна ПФМ для случайной величины  $h$ , запишем ее с учетом  $\alpha = -1/2$ ,  $\beta = 2$ :

$$M_h(t) = (1 - \beta t)^{-\alpha-1} = (1 - 2t)^{-1/2}.$$

Таким образом, можно записать ПФМ для распределения случайной величины, равной сумме квадратов случайных величин, распределенных по Гауссу:

$$M_{\sum h_i} = |M_h(t)|^k = (1 - 2t)^{-k/2}. \quad (9.3)$$

Соотношение (9.3) является ПФМ для распределения  $\chi^2$ . Причем нетрудно видеть, что данное распределение также относится к семейству гамма-функций с  $\alpha = k/2 - 1$ ,  $\beta = 2$ .

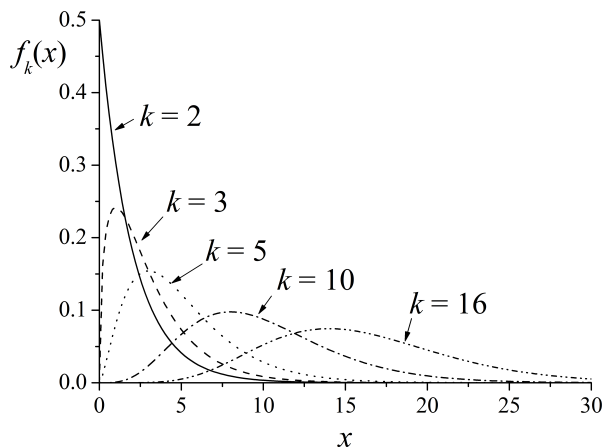


Рис. 9.1. Функция плотности вероятности распределения  $\chi^2$  для разных значений  $k$

Наконец, можно записать саму функцию плотности вероятности распределения  $\chi^2$ . Она будет выглядеть следующим образом (примеры показаны на рис. 9.1):

$$f_k(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right) \cdot 2^{k/2}} x^{k/2-1} e^{-x/2}. \quad (9.4)$$

Если  $k$  – четное, т.е.  $k = 2n$ , то

$$\Gamma\left(\frac{k}{2}\right) = \Gamma\left(\frac{2n}{2}\right) = (n-1)!$$

(см. свойства гамма-функции).

Если же  $k$  – нечетное, т.е.  $k = 2n + 1$ , то

$$\Gamma\left(\frac{k}{2}\right) = \Gamma\left(\frac{2n+1}{2}\right) = \frac{(2n)!}{4^n n!} \sqrt{\pi}$$

(см. свойства гамма-функции).

Найдем среднее значение и дисперсию для распределения  $\chi^2$ .

Согласно примеру 8.4,

$$E(x) = \mu_1 = \beta(\alpha + 1) = 2(k/2 - 1 + 1) = k,$$

$$D(x) = \sigma^2 = \beta^2(\alpha + 1) = 4(k/2 - 1 + 1) = 2k.$$

Наиболее вероятное значение для распределения  $\chi^2$ :

$$x_{\text{вер.}} = \alpha\beta = 2\left(\frac{k}{2} - 1\right) = k - 2.$$

Эти свойства распределения  $\chi^2$  понадобятся в дальнейшем при проверке гипотез.

При проверке гипотез в качестве предельного значения остаточной суммы квадратов используют 95-й или 99-й процентиль распределения  $\chi^2$ . Значения процентилей для  $k \leq 30$  приведены в табл. 9.1.

Таблица 9.1. Значения 95-го и 99-го процентилей  
распределения  $\chi^2$

k	$\chi_{0.95}^2$	$\chi_{0.99}^2$
1	3.84	6.63
2	5.99	9.21
3	7.81	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21
11	19.68	24.72
12	21.03	26.22
13	22.36	27.69
14	23.68	29.14
15	25.00	30.58
16	26.30	32.00
17	27.59	33.41
18	28.87	34.81
19	30.14	36.19
20	31.41	37.57
21	32.67	38.93
22	33.92	40.29
23	35.17	41.64
24	36.42	42.98
25	37.65	44.31
26	38.89	45.64
27	40.11	46.96
28	41.34	48.28
29	42.56	49.59
30	43.77	50.89

В большинстве программных пакетов для анализа данных квантили распределения  $\chi^2$  вычисляются по приближительным формулам. При этом одной из наиболее распространенных является аппроксимация Корниша-Фишера:

$$\chi_{k,\eta}^2 = k + A\sqrt{k} + B + \frac{C}{\sqrt{k}} + \frac{D}{k} + \frac{E}{k\sqrt{k}},$$

где

$$\begin{aligned} A &= d\sqrt{2}, \\ B &= \frac{2}{3}(d^2 - 1), \\ C &= d \cdot \frac{d^2 - 7}{9\sqrt{2}}, \\ D &= \frac{6d^4 + 14d^2 - 32}{405}, \\ E &= d \cdot \frac{9d^4 + 256d^2 - 433}{4860\sqrt{2}}, \\ d &= 2.0637 \left( \ln \frac{1}{1-\eta} - 0.16 \right)^{0.4274} - 1.5774 \\ &\quad \text{при } 0.5 \leq \eta \leq 0.999, \\ d &= -2.0637 \left( \ln \frac{1}{\eta} - 0.16 \right)^{0.4274} + 1.5774 \\ &\quad \text{при } 0.001 \leq \eta < 0.5. \end{aligned}$$

### **Пример 9.1. Расчет 95-го процентиля распределения $\chi^2$**

Получить аппроксимирующую формулу для 95-го процентиля и рассчитать по ней значение для  $k = 150$ .

Для  $\eta = 0.95$  имеем

$$\begin{aligned}
 d_{0.95} &= 2.0637 \left( \ln \frac{1}{1-\eta} - 0.16 \right)^{0.4274} - 1.5774 = \\
 &= 2.0637 \left( \ln \frac{1}{1-0.95} - 0.16 \right)^{0.4274} - 1.5774 = 1.6445, \\
 A_{0.95} &= d_{0.95} \sqrt{2} = 1.6445 \sqrt{2} = 2.3257, \\
 B_{0.95} &= \frac{2}{3} (d_{0.95}^2 - 1) = \frac{2}{3} (1.6445^2 - 1) = 1.1363, \\
 C_{0.95} &= d_{0.95} \frac{d_{0.95}^2 - 7}{9\sqrt{2}} = \\
 &= 1.6445 \frac{1.6445^2 - 7}{9\sqrt{2}} = -0.5550, \\
 D_{0.95} &= \frac{6d_{0.95}^4 + 14d_{0.95}^2 - 32}{405} = \\
 &= \frac{6 \cdot 1.6445^4 + 14 \cdot 1.6445^2 - 32}{405} = \\
 &= \frac{43.8820 + 37.8613 - 32}{405} = 0.1228, \\
 E_{0.95} &= d_{0.95} \frac{9d_{0.95}^4 + 256d_{0.95}^2 - 433}{4860\sqrt{2}} = \\
 &= 1.6445 \frac{9 \cdot 1.6445^4 + 256 \cdot 1.6445^2 - 433}{4860\sqrt{2}} = \\
 &= 1.6445 \frac{65.8231 + 692.3213 - 433}{4860\sqrt{2}} = 0.0780.
 \end{aligned}$$

Таким образом, получаем формулу для 95-го процентиля:

$$\begin{aligned}
 \chi_{k,0.95}^2 &= k + A_{0.95} \sqrt{k} + B_{0.95} + \\
 &\quad + \frac{C_{0.95}}{\sqrt{k}} + \frac{D_{0.95}}{k} + \frac{E_{0.95}}{k\sqrt{k}} = \\
 &= k + 2.3257\sqrt{k} + 1.1363 - \frac{0.555}{\sqrt{k}} + \\
 &\quad + \frac{0.1228}{k} + \frac{0.078}{k\sqrt{k}}.
 \end{aligned} \tag{9.5}$$

Для  $k = 150$  получаем

$$\begin{aligned} \chi_{150,0.95}^2 &= 150 + 2.3257\sqrt{150} + \\ &+ 1.1363 - \frac{0.555}{\sqrt{150}} + \frac{0.1228}{150} + \frac{0.078}{150\sqrt{150}} = \\ &= 150 + 28.4839 + 1.1367 - 0.453 + 0.008 + 0 = 179.18. \end{aligned}$$

### **Пример 9.2. Расчет 99-го процентиля распределения $\chi^2$**

Получить аппроксимирующую формулу для 99-го процентиля и рассчитать по ней значение для  $k = 150$ .

Рассчитаем параметр  $d$  при  $\eta = 0.99$ :

$$\begin{aligned} d_{0.99} &= 2.0637 \left( \ln \frac{1}{1-\eta} - 0.16 \right)^{0.4274} - 1.5774 = \\ &= 2.0637 \left( \ln \frac{1}{1-0.99} - 0.16 \right)^{0.4274} - 1.5774 = 2.327, \\ A_{0.99} &= d_{0.99}\sqrt{2} = 2.327\sqrt{2} = 3.2908, \\ B_{0.99} &= \frac{2}{3}(d_{0.99}^2 - 1) = \frac{2}{3}(2.327^2 - 1) = 2.9433, \\ C_{0.99} &= d_{0.99} \frac{d_{0.99}^2 - 7}{9\sqrt{2}} = 2.327 \frac{2.327^2 - 7}{9\sqrt{2}} = -0.2898, \\ D_{0.99} &= \frac{6d_{0.99}^4 + 14d_{0.99}^2 - 32}{405} = \\ &= \frac{6 \cdot 2.327^4 + 14 \cdot 2.327^2 - 32}{405} = \\ &= \frac{175.9287 + 75.8090 - 32}{405} = 0.5426, \\ E_{0.99} &= d_{0.99} \frac{9d_{0.99}^4 + 256d_{0.99}^2 - 433}{4860\sqrt{2}} = \\ &= 2.327 \frac{9 \cdot 2.327^4 + 256 \cdot 2.327^2 - 433}{4860\sqrt{2}} = \\ &= 2.327 \frac{263.8931 + 1386.2218 - 433}{4860\sqrt{2}} = 0.4121. \end{aligned}$$

Таким образом, получаем приближительную формулу для 99-го перцентиля:

$$\begin{aligned}
 \chi_{k,0.99}^2 &= k + A_{0.99}\sqrt{k} + B_{0.99} + \\
 &+ \frac{C_{0.99}}{\sqrt{k}} + \frac{D_{0.99}}{k} + \frac{E_{0.99}}{k\sqrt{k}} = \\
 &= k + 3.2908\sqrt{k} + 2.9433 - \frac{0.2898}{\sqrt{k}} + \\
 &+ \frac{0.5426}{k} + \frac{0.4121}{k\sqrt{k}}.
 \end{aligned} \tag{9.6}$$

Для  $k = 150$  получаем

$$\begin{aligned}
 \chi_{150,0.99}^2 &= 150 + 3.2908\sqrt{150} + \\
 &+ 2.9433 - \frac{0.2898}{\sqrt{150}} + \frac{0.5426}{150} + \frac{0.4121}{150\sqrt{150}} = \\
 &= 150 + 40.3039 + 2.9433 - 0.0237 + 0.0036 + 0.0002 = \\
 &= 193.23.
 \end{aligned}$$

### **Пример 9.3. Оценка 98-го перцентиля распределения $\chi^2$**

Оценить 98-й перцентиль распределения для  $k = 150$ .

При больших значениях  $k$  распределение  $\chi^2$  стремится к нормальному распределению. Из табл. 5.1 известно, что в диапазоне  $\mu \pm 2\sigma$  лежит 95.4 % распределения Гаусса, а правее границы  $\mu + 2\sigma$  лежит 2.3 %. Таким образом, примерное положение 98-го квантиля распределения  $\chi^2$  может быть оценено как

$$\chi_{k,0.98}^2 = k + 2\sqrt{2k}.$$

Для  $k = 150$  имеем

$$\chi_{k,0.98}^2 = 150 + 2\sqrt{300} = 150 + 2 \cdot 17.3 = 184.6.$$

Легко убедиться, что полученная оценка лежит между значениями 95-го и 99-го перцентилей из примеров 9.1 и 9.2 (179.18 и 193.23), но в отличие от них не требует громоздких расчетов.

# Заключение

Перечислим основные характеристики рассмотренных распределений.

## Биномиальное распределение

Тип распределения: дискретное.

Параметры распределения: вероятность благоприятного исхода  $p$  и число испытаний  $N$ .

Выражение для вероятности:

$$P_n = \frac{N!}{n!(N-n)!} \cdot p^n \cdot (1-p)^{N-n}, \quad n = 0, 1, 2, \dots, N.$$

Математическое ожидание (среднее):

$$E(n) = \bar{n} = Np.$$

Дисперсия:

$$D_n = E((n - E(n))^2) = \overline{n^2} - \bar{n}^2 = Np(1-p).$$

Производящая функция начальных моментов:

$$M_n(t) = E(e^{tn}) = (pe^t + 1 - p)^N.$$

## Распределение Пуассона

Тип распределения: дискретное.

Параметры распределения: среднее число событий за единицу времени  $\mu$  (при фиксированном значении интервала времени также используется параметр  $\lambda = \mu t$ ).

Выражение для вероятности:

$$P_n = \frac{(\mu t)^n}{n!} \cdot e^{-\mu t} = \frac{\lambda^n}{n!} \cdot e^{-\lambda}, \quad n = 0, 1, 2, \dots, \infty.$$

Математическое ожидание:

$$E(n) = \bar{n} = \mu t = \lambda.$$

Дисперсия:

$$D(n) = \mu t = \lambda.$$

Производящая функция начальных моментов:

$$M_n(t) = e^{\lambda(e^t - 1)}.$$

### **Равномерное распределение**

Тип распределения: непрерывное.

Параметры распределения: границы распределения  $a$  и  $b$ .

Плотность вероятности:

$$f(x) = \frac{1}{b - a}, \quad a < x < b.$$

Математическое ожидание:

$$E(x) = \bar{x} = \frac{a + b}{2}.$$

Дисперсия:

$$D(x) = \frac{(b - a)^2}{12}.$$

Производящая функция начальных моментов:

$$M_x(t) = \frac{e^{bt} - e^{at}}{t(b - a)}.$$

### **Распределение Гаусса**

Тип распределения: непрерывное.

Параметры распределения: среднее значение  $\mu$  и среднеквадратичное отклонение  $\sigma$ .

Плотность вероятности:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Математическое ожидание:

$$E(x) = \bar{x} = \mu.$$

Дисперсия:

$$D(x) = \sigma^2.$$

Производящая функция начальных моментов:

$$M_x(t) = e^{\mu t + (t\sigma)^2/2}.$$

### **Распределение Рэлея**

Тип распределения: непрерывное.

Параметры распределения: параметр масштаба  $\sigma$ .

Плотность вероятности:

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \quad 0 < x < \infty.$$

Математическое ожидание:

$$E(x) = \bar{x} = \sigma\sqrt{\frac{\pi}{2}}.$$

Дисперсия:

$$D(x) = \left(2 - \frac{\pi}{2}\right) \sigma^2.$$

### **Распределение Коши**

Тип распределения: непрерывное.

Плотность вероятности:

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad -\infty < x < \infty.$$

Математическое ожидание:

$$E(x) = \bar{x} = 0.$$

Дисперсия:

$$D(x) \rightarrow \infty.$$

### **Гамма-распределение**

Тип распределения: непрерывное.

Параметры распределения:  $\alpha$  и  $\beta$ .

Плотность вероятности:

$$f(x) = \frac{1}{\Gamma(\alpha + 1)\beta^{\alpha+1}} x^\alpha e^{-x/\beta}, \\ 0 < x < \infty, \alpha > -1, \beta > 0.$$

Математическое ожидание:

$$E(x) = \bar{x} = (\alpha + 1)\beta.$$

Дисперсия:

$$D(x) = (\alpha + 1)\beta^2.$$

Производящая функция начальных моментов:

$$M_x(t) = (1 - \beta t)^{-\alpha-1}.$$

### **Распределение $\chi^2$**

Тип распределения: непрерывное.

Параметры распределения: распределение  $\chi^2$  является частным случаем гамма-распределения с  $\alpha = k/2 - 1$ ,  $\beta = 2$ .

Плотность вероятности:

$$f_k(x) = \frac{1}{\Gamma\left(\frac{k}{2}\right) \cdot 2^{k/2}} x^{k/2-1} e^{-x/2}, \quad 0 < x < \infty, \quad k = 1, 2, 3, \dots, \infty.$$

Математическое ожидание:

$$E(x) = \bar{x} = k.$$

Дисперсия:

$$D(x) = 2k.$$

Производящая функция начальных моментов:

$$M_x(t) = (1 - 2t)^{-k/2}.$$

## Рекомендуемый список литературы

1. Худсон Д. Статистика для физиков. Лекции по теории вероятностей и элементарной статистике. 2-е изд. М.: Мир, 1970.
2. Кендалл М., Стьюарт А. Статистические выводы и связи. М.: Наука, 1973.
3. Корн Г., Корн Т. Справочник по математике (для научных работников и инженеров). М.: Наука, 1973.
4. Канеман Д. Думай медленно... решай быстро. М.: АСТ, 2021.
5. Талеб Н. Черный лебедь. Под знаком непредсказуемости. 2-е изд., доп. М.: Изд-во «КоЛибри», 2021.

## Рекомендуемый список интернет-источников

1. [www.library.mephi.ru](http://www.library.mephi.ru)
2. <http://elibrary.ru/>
3. <https://www.sciencedirect.com/browse/journals-and-books>
4. <http://www.nature.com/>
5. <http://link.springer.com/>
6. <http://pdg.lbl.gov/>
7. <https://www.nndc.bnl.gov/>
8. <https://www-nds.iaea.org/>
9. <https://www.nist.gov/pml/productsservices/physical-reference-data>
10. <https://trc.nist.gov/>

Семён Сергеевич Хохлов  
Анна Николаевна Дмитриева

## **Свойства распределений случайных величин**

Учебное пособие

Редактор Е.Е. Шумакова

Подписано в печать 05.09.2022. Формат 60×84 1/16.

Печ. л. 7,5. Уч.-изд. л. 7,5. Тираж 100 экз.

Изд. № 013-1. Заказ № .

Национальный исследовательский ядерный  
университет «МИФИ».

Типография НИЯУ МИФИ.

115409, Москва, Каширское ш., д. 31.