Дополнение 1. Вероятностные понятия

Обычно в университетских учебниках теория вероятностей излагается как математическая дисциплина, основанная на аксиомах. Здесь используется неформальный подход. Вероятностные понятия рассматриваются как ступеньки, ведущие к статистическим терминам, которые, в свою очередь, нужны для понимания отчётов, создаваемых программами, предназначенными для статистической обработки данных.

В этом дополнении определяются функция распределения и плотность, математическое ожидание и дисперсия. Приводятся примеры важнейших вероятностных распределений. Объясняется, как моделировать случайность с помощью датчиков случайных чисел.

1.1. Повторный выбор наудачу

Изучается множество $\boldsymbol{\Omega}$,[1] состоящее из M объектов, обладающих признаками X, Y, …. Допустим, что исследователя интересует признак X. Поскольку количество объектов M обычно очень велико, то узнать значения признака X у всех объектов нереально. Поэтому для изучения выбираются n объектов случайным образом (с помощью компьютера, см. раздел 1.7), где размер выборки n намного меньше M.

Идею случайного выбора наглядно представляет следующий способ, пригодный только для небольших M. Пронумеруем объекты числами от 1 до M. Напишем номера 1, 2, …, M на бумажках и поместим бумажки в шляпу. Возьмём наугад одну бумажку. Обозначим через X_1 значение признака X у объекта с номером, указанным на вынутой бумажке. Вернём бумажку в шляпу и перемешаем бумажки. Повторим n раз выбор бумажек наудачу. В результате получим ряд значений X_1, …, X_n, который называется *выборкой*. Отметим, что X_i являются случайными величинами, так как они зависят от номеров выбранных наудачу бумажек.

Рассмотрим примеры.

В некотором регионе живут M человек. Социолог опрашивает n случайно встретившихся на пути жителей и узнаёт, в частности, их ежемесячный доход X. Если ответы были правдивыми, а число n достаточно велико, то в качестве оценки среднего дохода в регионе разумно использовать выборочное среднее $\overline{X} = (X_1 + \ldots + X_n) / n$.

Социолога также может интересовать вопрос, какую долю населения региона составляют мужчины. Пусть Y — это пол жителя: 0 — женский, 1 — мужской. Тогда выборочное среднее $\overline{Y}$ представляет собой частоту, с которой мужчины встречались в выборке. Эта частота служит естественной оценкой для доли мужского населения в регионе.

Обобщим последний пример. Рассмотрим множество $\boldsymbol{A} \subseteq \boldsymbol{\Omega}$ объектов, обладающих определённым свойством. Обозначим через $|\boldsymbol{A}|$ количество объектов во множестве $\boldsymbol{A}$. Тогда $|\boldsymbol{\Omega}| = M$. Представляет интерес доля $|\boldsymbol{A}|/M$. В качестве оценки для этой доли можно взять частоту объектов в выборке, обладающих данным свойством.

В следующей таблице приведены частоты, с которыми встречаются буквы русского алфавита, включая пробел между словами (—).

— 0,175	о 0,090	е, ё 0,072	а 0,062	и 0,062	т 0,053	н 0,053	с 0,045
р 0,040	в 0,038	л 0,035	к 0,028	м 0,026	д 0,025	п 0,023	у 0,021
я 0,018	ы 0,016	з 0,016	ь, ъ 0,014	б 0,014	г 0,013	ч 0,012	й 0,010
х 0,009	ж 0,007	ю 0,006	ш 0,006	ц 0,004	щ 0,003	э 0,003	ф 0,002

[1] Раньше $\boldsymbol{\Omega}$ красиво именовали *генеральной совокупностью*.

Например, в стихотворении М. Ю. Лермонтова «Бородино» содержится n = 2461 букв и пробелов между словами. Частота появления буквы «а» оказалась равной 162 / 2461 ≈ 0,066. Она лишь незначительно отличается от указанной в таблице частоты 0,062.

> Рассказывают, что, создавая свой код, Морзе отправился в ближайшую типографию и подсчитал число литер в наборных кассах. Буквам и знакам, для которых литер в этих кассах было припасено больше, он сопоставил более короткие кодовые обозначения.
>
> Аршинов М. Н., Садовский А. Е. «Коды и математика»

1.2. Функция распределения признака

Пусть x — произвольное действительное число. Рассмотрим для признака X множество

$$\boldsymbol{A}_{\le x} = \{X \le x\},$$

образованное объектами, для которых выполняется условие $X \le x$. Представляет интерес вопрос, как изменяется доля таких объектов при изменении границы x. Ответ на него выражается функцией

$$F(x) = |\boldsymbol{A}_{\le x}| \,/\, |\boldsymbol{\Omega}|,$$

которая называется *функцией распределения* признака X.

При возрастании x число объектов во множестве $\boldsymbol{A}_{\le x}$ может только увеличиваться. Поэтому функция $F(x)$ не убывает по x. Функция распределения $F(x)$ важна тем, что через неё для произвольных действительных чисел $a < b$ выражается доля объектов, принадлежащих множеству $\{a < X \le b\}$. Эта доля равна $F(b) - F(a)$. Действительно, из множества $\{X \le b\}$ надо исключить объекты, входящие во вложенное в него (ввиду условия $a < b$) множество $\{X \le a\}$.

Рассмотрим пример, относящийся к демографии.[2] На рис. 1 приведена столбиковая диаграмма (также называемая *гистограммой*)[3] продолжительности жизни T в России в 2010 г. Высота столбика гистограммы равна доле p_t тех, кто прожил t лет, среди всех умерших в России в течение 2010 г.

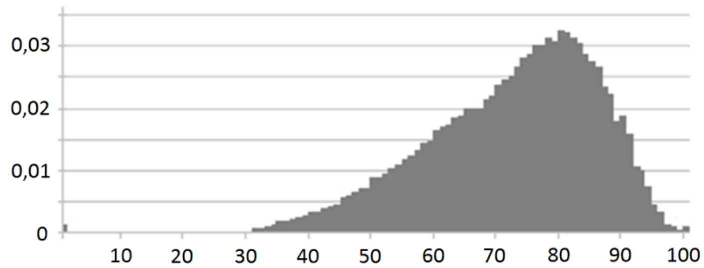

Рис. 1

Отметим, что «горка» заметно перекошена: правый склон намного круче. Наиболее высокие столбики, соответствующие наиболее вероятной продолжительности жизни, располагаются вблизи 80 лет. Небольшой «горбик» около 65 лет соответствует наиболее вероятной продолжительности жизни мужчин. Подобные диаграммы за разные годы представляют несомненный интерес для страховщиков и социальных служб.

---

[2] Пример основан на данных из статьи Горшкова В. А. «Распределение продолжительности жизни в России. Альтернативная оценка», размещённой на сайте demographia.ru в ноябре 2011 г.

[3] Термин, впервые использованный К. Пирсоном в 1895 году, происходит от ἱστός — (др.-греч.) *столб*.

Нетрудно понять, что для целых значений возраста t функцию распределения $F(t)$ продолжительности жизни T можно представить в виде суммы высот столбиков гистограммы:

$$F(t) = p_1 + \ldots + p_t.$$

Тем самым функция $F(t)$ выражает «накопленную» долю людей, проживших не более t лет.[4] Для произвольных возрастов $a < b$ доля людей, проживших больше a, но не более b лет, равна $F(b) – F(a)$ (рис. 2).

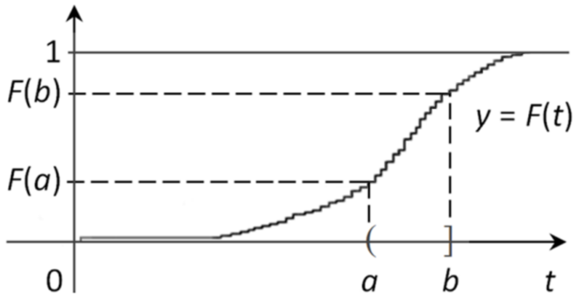

Рис. 2

Понятно, что полную информацию о продолжительности жизни во всей России получить затруднительно. Для оценки функции распределения $F(t)$ можно использовать выборочный метод. Скажем, изучить данные о смертности граждан, хранящиеся в нескольких случайно отобранных отделениях ЗАГС[5] из разных регионов страны. Предположим, что в результате запросов в отделения ЗАГС исследователь получил сведения о продолжительности жизни n человек. На основе полученной выборки подсчитываются частоты $\hat{p}_t$, по которым вычисляется *выборочная функция распределения*

$$\hat{F}_n(t) = \hat{p}_1 + \ldots + \hat{p}_t.$$

Для заданных возрастов $a < b$ в качестве оценки неизвестной доли $F(b) – F(a)$ можно использовать частоту $\hat{F}_n(b) - \hat{F}_n(a)$. Принципиальный вопрос: насколько большим должен быть размер выборки n для обеспечения требуемой точности оценки. Ответ будет дан в разделе 1.6.

1.3. Дискретные распределения

Пусть x — это одно из возможных значений признака X. Рассмотрим множество

$$\boldsymbol{A}_x = \{X = x\},$$

состоящее из объектов, для которых выполняется условие $X = x$. Долю множества $\boldsymbol{A}_x$ во множестве всех объектов $\boldsymbol{\Omega}$ обозначим через p:[6]

$$p = |\boldsymbol{A}_x| \,/\, |\boldsymbol{\Omega}|.$$

На практике доля p неизвестна. Для её оценки производится отбор n объектов из $\boldsymbol{\Omega}$ случайным образом. Значения признака X у этих объектов в 1.1 обозначались $X_1, \ldots, X_n$ и назывались элементами выборки. Разберём два «подводящих» примера и две простые вероятностные задачи, связанные с выборкой и множеством $\boldsymbol{A}_x$.

[4] Раньше $F(t)$ называлась *кумулятивной функцией распределения* (cumulative — (англ.) *накопленный*).

[5] Запись актов гражданского состояния.

[6] Более точное обозначение p_x не будем использовать во избежание громоздких формул.

Студент на экзамене берёт наудачу один из M билетов, среди которых L билетов — лёгкие. Понятно, что вероятность вытянуть лёгкий билет равна доле L/M лёгких билетов среди всех M билетов.

Задача 1. Каковы шансы, что X_1 окажется равным x? Другими словами: какова вероятность, что первый выбранный объект принадлежит множеству $\boldsymbol{A}_x$? Ответ: вероятность равна доле p.

Игральная кость бросается 3 раза. Результат трёх бросаний описывается набором из трёх цифр от 1 до 6. Всего таких наборов 6^3. Все наборы равновероятны. В частности, вероятность троекратного выпадения одного очка, т. е. вероятность появления набора (1, 1, 1), равна $1/6^3 = (1/6)^3$.

Задача 2. Чему равна вероятность, что все элементы выборки X_1, …, X_n окажутся равными x? Ответ: годятся $|\boldsymbol{A}_x|^n$ наборов из $|\boldsymbol{\Omega}|^n$, поэтому искомая вероятность равна $|\boldsymbol{A}_x|^n / |\boldsymbol{\Omega}|^n = p^n$.

Теперь решим более сложную задачу.

Задача 3. Найти вероятность, что ровно k элементов выборки ($0 \le k \le n$) окажутся равными x.

Сначала подсчитаем число различных вариантов размещения k единиц и $(n - k)$ нулей. Единицы соответствуют элементам выборки, оказавшимся равными x, нули — элементам выборки, принявшим другие значения. Например, пусть $k = 3$ и $n = 12$. Вот один из возможных вариантов размещения:

0, 0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 0

Сколько всего существует вариантов размещения? Покажем, что $12 \cdot 11 \cdot 10 / 6 = 220$.

Выполним подсчёт вариантов для произвольных k и n. Первую единицу можно поставить на любое из n мест, вторую единицу — на любое из $(n-1)$ оставшихся мест, …, k-ю единицу — на любое из $(n - k + 1)$ мест. Комбинируя возможности, получаем $n(n-1)\dots(n-k+1)$ вариантов расстановок. Однако расстановки, отличающиеся лишь порядком выбранных мест, дают одно и то же размещение. (Например, приведённое выше размещение могло получиться при последовательной расстановке единиц как на местах 3, 6, 8, так и на местах 8, 3, 6.) Поэтому не станем различать расстановки, отличающиеся лишь порядком номеров мест. Количество различных перестановок из k различных номеров равно $k! \equiv 1\cdot 2\cdot \ldots \cdot k$.[7] Действительно, первый номер можно поставить на любое из k мест в перестановке, второй номер — на любое из оставшихся $(k-1)$ мест, …, k-й номер — на единственное оставшееся место. В итоге получаем, что число вариантов размещения k единиц на n местах равно

$$\frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{k!\,(n-k)!}. \tag{1}$$

Для выражения (1) используется обозначение C_n^k, называемое *числом сочетаний*[8] из n по k или *биномиальным коэффициентом*. Последнее название связано с участием коэффициентов C_n^k в формуле *бинома Ньютона*:

$$(a+b)^n = \sum_{k=0}^{n} C_n^k a^k b^{n-k}. \qquad (\text{Здесь } C_n^0 = 1.) \tag{2}$$

Вернёмся к подсчёту вероятности, что ровно k элементов выборки окажутся равными x. Сначала выберем номера элементов со значением x — разместим единицы. Мы установили, что это можно сделать C_n^k способами. Аналогично задаче 2 нетрудно понять, что любой набор из k единиц и $(n - k)$ нулей появляется с вероятностью $p^k(1-p)^{n-k}$. Отсюда получаем искомую вероятность $C_n^k p^k (1-p)^{n-k}$.

---

[7] Обозначение $k!$ называется *факториалом* числа k (factorialis — (лат.) *производящий*, *умножающий*).

[8] Combination — (англ.) *сочетание*.

Сформулируем результаты решения задач на языке случайных величин.

Определение. *Случайными величинами* называются функции от элементов выборки.

Каждый элемент X_i сам является случайной величиной. Примерами случайных величин служат функции $X_1+...+X_n$, $X_1^2+...+X_n^2$, $\max\{X_1,...,X_n\}$.

Пожалуй, самые простые случайные величины — это *индикаторы* $\zeta_1, \dots, \zeta_n$, где

$$\zeta_i = I_{\{X_i=x\}} = \begin{cases} 1, & \text{если } X_i = x; \\ 0, & \text{если } X_i \neq x. \end{cases} \tag{3}$$

Обозначим через $\mathbf{P}(\xi=t)$ вероятность,[9] с которой случайная величина ξ принимает значение t. Тогда согласно задачам 1–3 имеем следующие представления:

$$\mathbf{P}(\zeta_1=1)=p, \quad \mathbf{P}(\zeta_1\cdot\ldots\cdot\zeta_n=1)=p^n, \quad \mathbf{P}(\zeta_1+\ldots+\zeta_n=k)=C_n^k p^k(1-p)^{n-k}.$$

Индикаторы $\zeta_1, \dots, \zeta_n$ также называют *испытаниями Бернулли*.[10] Если $\zeta_i = 1$, то говорят, что в i-м испытании случился «успех», если $\zeta_i = 0$ — «неудача». В этой терминологии случайная величина $S_n = \zeta_1+\ldots+\zeta_n$ представляет собой количество «успехов» в n испытаниях Бернулли.

Определение. *Распределением* случайной величины называется набор вероятностей, с которыми она принимает свои значения. Например, S_n имеет распределение

$$\mathbf{P}(S_n=k)=C_n^k p^k(1-p)^{n-k}, \quad k=0,1,\ldots,n. \tag{4}$$

Распределение (4) называется *биномиальным* с параметрами n и p.[11]

На рис. 3 изображена диаграмма биномиального распределения для $n = 100$ и $p = 0{,}5$. На ней над каждым значением $k=0,1,\ldots,100$ нарисована вертикальная линия длины $\mathbf{P}(S_n=k)$.

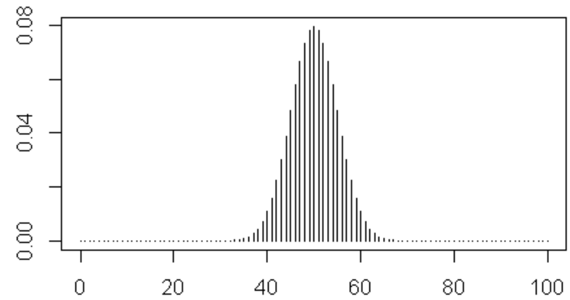

Рис. 3

Обратим внимание, что биномиальные вероятности практически не отличаются от нуля при $k < 35$ и $k > 65$. Вычислив сумму вероятностей от 35 до 65, получим $0{,}9982 \approx 1$. Таким образом, при 100 бросаниях монетки число выпавших «гербов» почти наверняка окажется между 35 и 65. Этот результат будет объяснён в разделе 1.6.

Рассмотрим ещё одну важную случайную величину и найдём её распределение. Пусть ν обозначает количество «неудач» в испытаниях Бернулли до появления первого «успеха». Перемножая вероятности наблюдения k «неудач» и одного «успеха» на $(k + 1)$-м месте, имеем

$$\mathbf{P}(\nu=k)=(1-p)^k p, \quad k=0,1,2,\ldots. \tag{5}$$

---

[9] Probabilitas — (лат.) *вероятность.*

[10] В честь Якоба Бернулли (1654 – 1705), впервые доказавшего закон больших чисел для этой модели.

[11] Binomial distribution — (англ.) *биномиальное распределение.*

Распределение (5) называется *геометрическим* ввиду того, что вероятности образуют геометрическую прогрессию, умноженную на p. Суммируя вероятности $\mathbf{P}(\nu = k)$, нетрудно получить, что

$$\mathbf{P}(\nu \le k) = 1 - (1-p)^{k+1}. \tag{6}$$

Отметим, что случайная величина ν принимает счётное[12] множество значений: $0, 1, 2, \ldots$.

Определение. Распределения случайных величин, имеющих конечное или счётное множество значений, называются *дискретными*.

1.4. Непрерывные распределения

На практике любой признак измеряется с определённой точностью. Например, температура больного — с точностью до 0,1 градуса, цена товара — с точностью до копейки, продолжительность жизни человека — с точностью до года (хотя при желании её можно узнать и с точностью до месяца, дня, часа). Всегда существует *достаточная точность* δ в том смысле, что практически нереально или неинтересно измерять признак с большей точностью. Поэтому можно считать, что все значения признака располагаются на действительной прямой в точках сетки с шагом δ.

Ранее в 1.2 было определено понятие функции распределения $F(x)$ признака X. Эта функция разрывна: она растёт скачками в точках сетки. Причём величина скачка в точке сетки с координатой x равна доле $p_x = |\boldsymbol{A}_x| \,/\, |\boldsymbol{\Omega}|$, где $\boldsymbol{A}_x = \{X = x\}$ — множество объектов, на которых признак X имеет значение x.

Таким образом, в реальности нет признаков с непрерывными функциями распределения. Зачем же тогда рассматриваются непрерывные распределения? На это есть несколько причин:

1. Непрерывное распределение представляет собой «идеал» распределения признака, измеряемого со сколько угодно высокой точностью и имеющего бесконечное множество значений (примером может служить определяемое ниже равномерное распределение на отрезке [0, 1]);
2. Непрерывные распределения появляются как пределы дискретных распределений (главный источник — предельные теоремы теории вероятностей, см. раздел 1.6);
3. Семейства непрерывных распределений, зависящие от нескольких параметров, используются в математической статистике в качестве моделей, аппроксимирующих распределения признаков.

Эксперимент. Представим, что вокруг оси вращается с малым трением стрелка, конец которой описывает окружность длины 1 (рис. 4). Стрелку раскрутили и дождались момента, когда она остановилась. Пусть η — координата точки окружности, на которую указала стрелка.

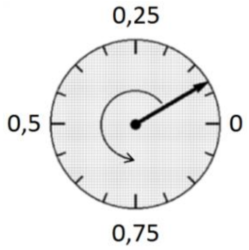

Рис. 4

[12] Бесконечное множество называется *счётным*, если его элементы можно занумеровать натуральными числами. В математическом анализе доказывается, что рациональные числа образуют счётное множество, а множество всех точек отрезка [0, 1] несчётно.

Теоретически эту координату можно измерить с любой точностью. Другими словами, η может принимать любые действительные значения из отрезка [0, 1]. Интуитивно понятно, что «в идеале» попадание конца стрелки в отрезок $[a, b]$, где $0 \le a < b \le 1$, происходит с вероятностью $b - a$, независимо от положения этого отрезка внутри отрезка [0, 1]:

$$\mathbf{P}(a \le \eta \le b) = b - a. \tag{7}$$

Говорят, что координата η имеет *равномерное* распределение на отрезке [0, 1]. Также случайную величину η называют координатой точки, *выбранной наудачу* из отрезка [0, 1]. Положив в формуле (7) $a = 0$ и $b = x$, находим функцию распределения случайной величины η:

$$F_\eta(x) = \mathbf{P}(\eta \le x) = \begin{cases} 0, & \text{если } x \le 0; \\ x, & \text{если } 0 < x < 1; \\ 1, & \text{если } x \ge 1. \end{cases} \tag{8}$$

Определение. *Экспоненциальной* с параметром $\lambda > 0$ называется случайная величина τ, функцией распределения которой задаётся формулой

$$F_\tau(x) = \mathbf{P}(\tau \le x) = \begin{cases} 0, & \text{если } x \le 0; \\ 1 - e^{-\lambda x}, & \text{если } x > 0. \end{cases} \tag{9}$$

На рис. 5 изображены графики функции распределения $F_\tau(x)$ при $\lambda = 2, 1, 1/2$.

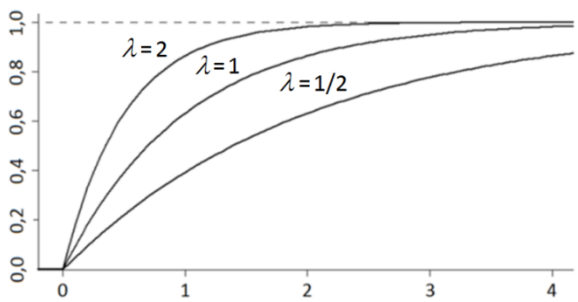

Рис. 5

Непрерывные распределения удобно задавать через их плотности.

Определение. Если существует такая неотрицательная функция $f_\xi(x)$, что для произвольных чисел $a < b$ выполнятся равенство

$$\mathbf{P}(a \le \xi \le b) = \int_a^b f_\xi(x)\,dx,$$

то $f_\xi(x)$ называется *плотностью распределения* случайной величины ξ (рис. 6).

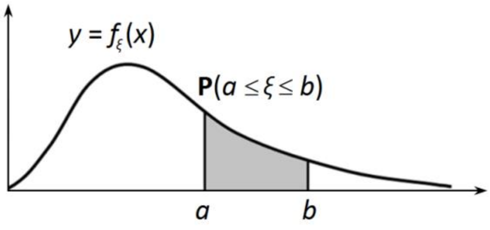

Рис. 6

Если плотность распределения $f_\xi(x)$ существует, то она является производной функции распределения $F_\xi(x)$:

$$f_\xi(x) = F_\xi'(x).$$

В частности, для экспоненциальной случайной величины τ плотность распределения $f_\tau(x)$ задаётся формулой

$$f_\tau(x) = F_\tau'(x) = \begin{cases} 0, & \text{если } x < 0; \\ \lambda e^{-\lambda x}, & \text{если } x \geq 0. \end{cases} \tag{10}$$

Случайные величины с дискретными распределениями плотностей не имеют.

1.5. Математическое ожидание и дисперсия

Математическое ожидание случайной величины является одним из основных понятий теории вероятностей. Давайте сначала рассмотрим его аналог из механики — центр масс.

Вопрос по механике. Штангист по ошибке поставил на штангу диски неодинаковой массы: слева диск массы m_1, а справа — диск массы m_2 (рис. 7). Если поднимать штангу одной рукой, то в какой точке грифа (который ради простоты считается невесомым) надо её держать?

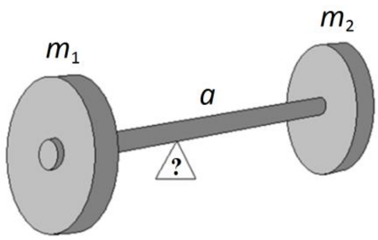

Рис. 7

Ответ. Пусть a — длина грифа. Тогда центр масс штанги (без учёта веса самого грифа) находится от левого диска на расстоянии $x_c = (0 \cdot m_1 + a \cdot m_2) / (m_1 + m_2) = am_2 / (m_1 + m_2)$.

В общем случае, когда на невесомой спице в точках с координатами x_1, x_2, … закреплены грузики с массами m_1, m_2, …, координата центра масс всей системы грузиков вычисляется согласно формуле

$$x_c = \sum x_i m_i \Big/ \sum m_i. \tag{11}$$

В теории вероятностей аналогами масс m_i служат вероятности $p_i = \mathbf{P}(\xi = x_i)$, причём $\sum p_i = 1$.

Для случайной величины ξ, принимающей значения x_1, x_2, … с вероятностями p_1, p_2, … соответственно, *математическое ожидание* определяется формулой

$$\mathbf{M}\xi = \sum x_i p_i. \tag{12}$$

Таким образом, математическое ожидание — это центр вероятностных масс. Для примера найдём математическое ожидание отдельного испытания Бернулли ζ_i (см. формулу (3)):

$$\mathbf{M}\zeta_i = 0 \cdot \mathbf{P}(\zeta_i = 0) + 1 \cdot \mathbf{P}(\zeta_i = 1) = 0 \cdot (1 - p) + 1 \cdot p = p.$$

Математическое ожидание случайной величины ξ с плотностью распределения $f_\xi(x)$ определяется формулой

$$\mathbf{M}\xi = \int_{-\infty}^{+\infty} x\, f_\xi(x)\, dx. \tag{13}$$

В качестве примера вычислим $\mathbf{M}\eta$ случайной величины η, имеющей равномерное распределение на отрезке [0, 1] (см. формулу (7)). Дифференцируя заданную формулой (8) функцию распределения $F_\eta(x)$, находим плотность распределения

$$f_\eta(x) = I_{[0,\,1]} = \begin{cases} 1, & \text{если } 0 \le x \le 1; \\ 0, & \text{в противном случае.} \end{cases}$$

Учитывая то, что первообразной для функции $f(x) = x$ служит функция $F(x) = x^2/2 + c$, получим:

$$\mathbf{M}\eta = \int_{-\infty}^{+\infty} x\, f_\eta(x)\, dx = \int_0^1 x \cdot 1\, dx = \frac{1^2}{2} - \frac{0^2}{2} = \frac{1}{2}.$$

Конечно, этот результат очевиден и без вычислений: центр масс однородного стержня находится в его середине. В конце дополнения приведены несколько задач на подсчёт математических ожиданий.

Отметим, что математическое ожидание определено не для любой случайной величины. Например, для случайной величины γ, имеющей *распределение Коши* с плотностью

$$f_\gamma(x) = \frac{1}{\pi}\frac{1}{1 + x^2},$$

интеграл из формулы (13) не определён, т. е. $\mathbf{M}\gamma$ не существует. Это вызвано тем, что плотность $f_\gamma(x)$ настолько медленно убывает при $x \to \pm\infty$, что интегралы

$$\int_{-\infty}^{0} x\, f_\gamma(x)\, dx \text{ и } \int_{0}^{+\infty} x\, f_\gamma(x)\, dx$$

оказываются бесконечными (*расходящимися*). Из-за этого возникает неопределённость вида $\infty - \infty$.

Математическое ожидание $\mathbf{M}\xi$ является естественной характеристикой *центра* распределения (центра вероятностной массы). Возникает вопрос, что взять в качестве меры типичного *разброса* (рассеяния вероятностной массы) вокруг математического ожидания. Такой мерой служит *стандартное отклонение* $\sqrt{\mathbf{D}\xi}$, где $\mathbf{D}\xi$ — определяемая ниже дисперсия случайной величины ξ.

Как и при знакомстве с математическим ожиданием, начнём с механической аналогии.

Вопрос по механике. На тонком стержне (числовой прямой) в точках с координатами x_i находятся массы m_i (рис. 8). Какаю координату a следует выбрать для точки крепления стержня к вертикальной оси, чтобы минимизировать *момент инерции относительно точки крепления*

$$I_a = \sum (x_i - a)^2 m_i \;?$$

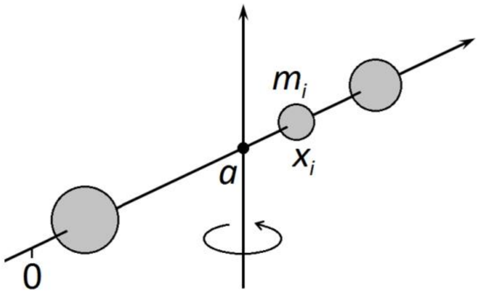

Рис. 8

Ответ. В задаче 4 в конце дополнения предлагается доказать, что минимум функции

$$f(a) = \sum (x_i - a)^2 m_i$$

достигается при $a = x_c$, где x_c — координата центра масс, заданная формулой (11).

В теории вероятностей аналогом момента инерции относительно центра масс $\sum (x_i - x_c)^2 m_i$ служит дисперсия случайной величины.

Для случайной величины ξ, принимающей значения x_1, x_2, … с вероятностями p_1, p_2, … соответственно, *дисперсия* определяется формулой

$$\mathbf{D}\xi = \sum (x_i - \mathbf{M}\xi)^2 p_i. \tag{14}$$

Для примера найдём дисперсию испытания Бернулли ζ_i. Выше было установлено, что $\mathbf{M}\zeta_i = p$. С учётом этого имеем:

$$\mathbf{D}\zeta_i = (0-p)^2 \cdot \mathbf{P}(\zeta_i = 0) + (1-p)^2 \cdot \mathbf{P}(\zeta_i = 1) = p^2(1-p) + (1-p)^2 p = p(1-p).$$

Для случайной величины ξ с плотностью распределения $f_\xi(x)$ непрерывным аналогом формулы (14) служит формула

$$\mathbf{D}\xi = \int_{-\infty}^{+\infty} (x - \mathbf{M}\xi)^2 f_\xi(x)\,dx. \tag{15}$$

В задачах, приведённых в конце дополнения, предлагается вычислить дисперсии некоторых случайных величин с дискретными и непрерывными распределениями.

Не всегда требуется полная количественная информация о случайной величине ξ, выражающаяся в её функции распределения $F_\xi(x)$. Иногда достаточно знать, где располагается область наиболее типичных значений случайной величины, под которой на практике обычно понимают интервал с границами $\mathbf{M}\xi \pm \sqrt{\mathbf{D}\xi}$ (см. рис. 9).

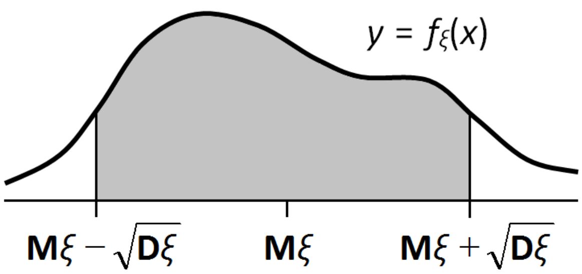

Рис. 9

1.6. Предельные теоремы

Прежде всего, познакомимся с теоремой Пуассона,[13] которую также называют *законом редких событий*. Эта теорема позволяет приближённо вычислять биномиальные вероятности $C_n^k p^k (1-p)^{n-k}$, когда параметр p имеет очень малое значение, а параметр n — очень большое. Проблема в том, что в таком случае при подсчёте C_n^k придётся перемножать большие числа, а p^k будет очень малым числом даже при умеренных значениях k.

Положим $\lambda = np$. Предположим, что λ не слишком велико (скажем, $\lambda < 10$). Покажем, что тогда для биномиальных вероятностей при небольших k можно использовать приближение

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda} \equiv \pi_k.$$

При $k = 0$ считается, что $\lambda^0/0! = 1$. Разумность такого приближения обосновывает

***Теорема Пуассона*.** Пусть k фиксировано, $n \to \infty$, $\lambda = np < \infty$. Тогда $C_n^k p^k (1-p)^{n-k} \to \pi_k$.

Д о к а з а т е л ь с т в о. Преобразуем биномиальную вероятность:

$$\frac{n(n-1)\dots(n-k+1)}{k!} p^k (1-p)^{n-k} = \frac{(np)^k}{k!}(1-p)^n \left[\frac{n}{n}\cdot\frac{n-1}{n}\dots\frac{(n-k+1)}{n}\cdot(1-p)^{-k}\right].$$

Здесь $(1-p)^n = (1-\lambda/n)^n \to e^{-\lambda}$, а выражение в квадратных скобках стремится к 1, так как число k фиксировано, и каждый сомножитель стремится к 1.

Определение. Дискретное распределение с вероятностями π_k в точках $k = 0, 1, 2, \dots$ называется *пуассоновским* с параметром λ.

Можно подсчитать, что λ служит математическим ожиданием случайной величины, имеющей пуассоновское распределение. В качестве примера использования пуассоновского приближения рассмотрим задачу из области страхования.

Вычисление страхового тарифа. Согласно статистике, накопленной страховой компанией, вероятность угона в течение года автомобиля определённой марки $p = 0{,}0123$. Из $n = 500$ страхуемых машин в следующем году предположительно будет угнано $\lambda = np \approx 6$ машин. Поскольку вероятность отдельного угона мала, а угоны предполагаются не связанными между собой, то можно использовать пуассоновское приближение для распределения числа K ожидаемых угонов. Задача заключается в определении величины страхового процента ρ, компенсирующего убытки в случае слишком большого числа угонов.

Решим эту задачу. Пусть C — цена автомобиля. Предположим, что в случае угона она выплачивается владельцу полностью. Тогда страховые выплаты составят сумму CK. Доход страховой компании равен ρCn. Зададим малое число α из соображения, что страховщик готов пренебречь возможностью осуществления события, происходящего с вероятностью α. Найдём величину ρ из условия $\mathbf{P}(CK \le \rho Cn) = \mathbf{P}(K \le \rho n) = 1-\alpha$. Для вероятности $\mathbf{P}(K \le \rho n)$ используем пуассоновское приближение:

$$\mathbf{P}(K \le \rho n) = \sum_{k \le \rho n} \mathbf{P}(K = k) \approx \sum_{k \le \rho n} \pi_k.$$

[13] Siméon Denis Poisson (1781 — 1840) — знаменитый французский математик, механик и физик.

Сумму вероятностей π_k в правой части этой формулы можно вычислять, например, с помощью программы Excel. В частности, при $\rho n = 10$ данная сумма равна 0,95087 ≈ 0,95. Отсюда

$$\alpha \approx 1 - 0{,}95 = 0{,}05 \ \text{ и } \ \rho = 10/n = 10/500 = 2\%.$$

Для иллюстрации на рис. 10 изображена диаграмма пуассоновского распределения с параметром $\lambda = 0{,}0123 \cdot 500$, аналогичная диаграмме биномиального распределения из 1.3:

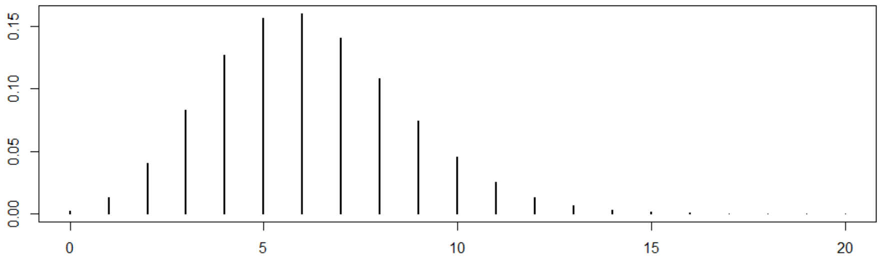

Рис. 10

Видим, что вероятности π_k становятся очень малыми при $k > 13$: $\pi_{14} \approx 0{,}0027$.

Теперь обсудим закон больших чисел. Ранее в 1.3 были определены испытания Бернулли $\zeta_1, \dots, \zeta_n$ и было установлено, что их сумма $S_n = \zeta_1 + \dots + \zeta_n$ имеет биномиальное распределение. В задачах 6 и 7 предлагается доказать, что $\mathbf{M}S_n = np$ и $\mathbf{D}S_n = np(1-p)$.

Пусть число испытаний n велико. Насколько вероятно тогда могут наблюдаться значения случайной величины S_n, значительно удалённые от $\mathbf{M}S_n$? Якоб Бернулли в 1713 году установил следующее неравенство: для произвольного числа $\varepsilon > 0$

$$\mathbf{P}(|S_n - \mathbf{M}S_n| \ge n\varepsilon) = \mathbf{P}(|S_n - np| \ge n\varepsilon) \le \frac{p(1-p)}{n\varepsilon^2}.$$

Заметим, что выражение, стоящее справа в формуле, стремится к нулю при $n \to \infty$. Обозначив через $\overline{X} = S_n/n$ частоту «успехов» в n испытаниях Бернулли, получаем сходимость

$$\mathbf{P}(|S_n - np| \ge n\varepsilon) = \mathbf{P}(|\overline{X} - p| \ge \varepsilon) \to 0 \quad \text{при } n \to \infty.$$

Другими словами, отклонение на произвольную величину ε частоты $\overline{X}$ от оцениваемой доли p становится сколь угодно маловероятным при увеличении размера выборки n. Это утверждение получило название ***закона больших чисел*** для схемы Бернулли.

Закон больших чисел допускает обобщение: пусть $\mu = (x_1 + \dots + x_M)/M$ — среднее значение признака X, подсчитанное по всем M объектам из $\boldsymbol{\Omega}$; $\overline{X} = (X_1 + \dots + X_n)/n$ — выборочное среднее. Тогда $\mathbf{P}(|\overline{X} - \mu| \ge \varepsilon) \to 0$ при $n \to \infty$.

Представляет интерес, насколько быстро убывает погрешность $\Delta_n = |\overline{X} - \mu|$ с увеличением размера выборки n. Оказывается, Δ_n имеет порядок малости $const/\sqrt{n}$. Более точный ответ даёт

Центральная предельная теорема. Положим $\sigma^2 = \frac{1}{M}\sum_{j=1}^{M}(x_j - \mu)^2$. Тогда имеет место сходимость

$$\mathbf{P}(\sqrt{n}\,(\overline{X} - \mu)/\sigma \le x) \to \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-y^2/2}\,dy \ \text{ при } n \to \infty. \qquad (16)$$

Здесь функция $\Phi(x)$ называется *функцией распределения стандартного нормального закона*. Известно, что её нельзя выразить в явном виде через элементарные функции. Производная

$$\varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$$

имеет форму «колокола». Величины некоторых площадей под её графиком показаны на рис. 11.

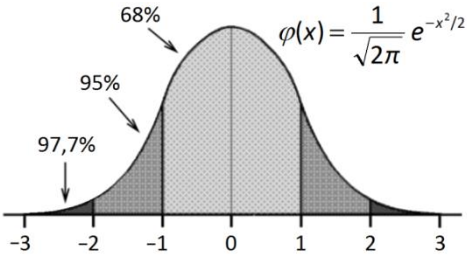

Рис. 11

Рассмотрим случайную величину Z с функцией распределения $\Phi(x)$. Она называется *стандартной нормальной.* Согласно рис. 11 приращение $\Phi(3) - \Phi(-3) \approx 0{,}997$. Следовательно, неравенство $|Z| \le 3$ выполняется с вероятностью 0,997. Иначе говоря, Z является практически ограниченной. В свою очередь, поскольку $\Phi(2) - \Phi(-2) \approx 0{,}95$, погрешность $\Delta_n = |\overline{X} - \mu|$ не будет превышать величины $2\sigma/\sqrt{n}$ с вероятностью 0,95 для больших n (**«*правило двух сигм*»**).

Однако обычно значение параметра σ неизвестно, поэтому на практике σ заменяют на зависящую от самих наблюдений оценку — *выборочное стандартное отклонение*

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}. \qquad (17)$$

Оценка погрешности измерений. Рассмотрим классическую вероятностную модель, которая применяется при многократных измерениях характеристики μ со случайными ошибками ε_i: $X_i = \mu + \varepsilon_i$, $i = 1, \ldots, n$. Предполагается, что измерения X_i проводятся в одних и тех же условиях и независимы между собой в том смысле, что результаты одних измерений не влияют на другие. Ошибки ε_i можно рассматривать как независимые и одинаково распределённые случайные величины с нулевым математическим ожиданием. Согласно центральной предельной теореме для большого количества измерений погрешность $\Delta_n = |\overline{X} - \mu|$ не должна превышать $2\sigma/\sqrt{n}$ с вероятностью около 0,95*.* Заменив параметр σ на оценку S из (17), получим неравенство $\overline{X} - 2S/\sqrt{n} < \mu < \overline{X} + 2S/\sqrt{n}$. Говорят, что $\overline{X} \pm 2S/\sqrt{n}$ — это границы *доверительного интервала* с коэффициентом доверия $\beta \approx 95\%$. Насколько на практике можно надеяться, что этот интервал накрывает истинное значение μ? В учебнике «Теория вероятностей и случайных процессов» В. Н. Тутубалин пишет:

«В «Философском очерке теории вероятностей» Лаплас, определив отношение массы Юпитера к массе Солнца, предлагает пари, что будущие поколения учёных не изменят найденное им число более чем на 1% (вероятность того, что это произойдёт, согласно Лапласу, ничтожно мала). Но современное значение отличается от найденного Лапласом на 2%».

Дело в том, что нередко оказывается, что наблюдения имеют вид $X_i = \mu + \Delta + \varepsilon_i$, где Δ — неизвестная *систематическая ошибка измерения*. Попытка оценить величину возможного отклонения $\overline{X}$ от μ лишь на основе разброса наблюдений X_i вокруг $\overline{X}$, вообще говоря, может приводить к ошибочным выводам.

Проиллюстрируем смысл центральной предельной теоремы для испытаний Бернулли.[14] Построим диаграммы биномиальных распределений для p = 1/5 и n = 10, 25, 50 (рис. 12).

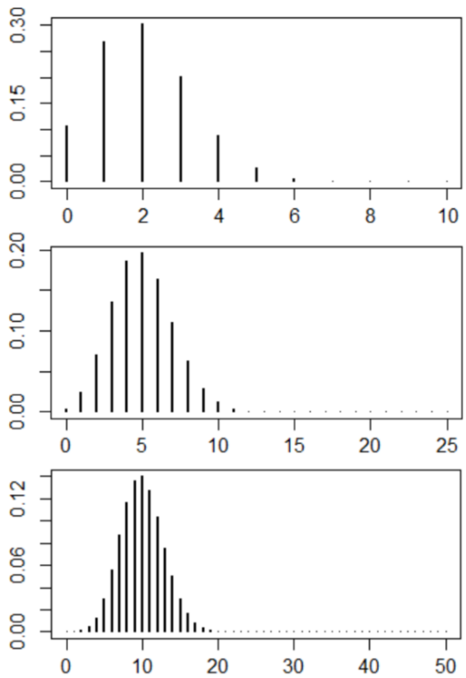

Рис. 12

Видим, что с увеличением n асимметрия распределения уменьшается, и огибающая столбиков (после переноса начала координат в точку np и соответствующего изменения масштаба) становится всё более похожей на плотность $\varphi(x)$ стандартного нормального распределения. Биномиальные вероятности при больших n оказываются практически нулевыми вне интервала

$$(np-3\sqrt{np(1-p)},\ np+3\sqrt{np(1-p)}).$$

Например, для n = 100 и p = 0,5 получаем интервал $(50-3\sqrt{25}, 50+3\sqrt{25})=(35, 65)$ (см. рис. 3).

1.7. Моделирование случайности

Каким образом можно осуществить случайный выбор n объектов из множества $\boldsymbol{\Omega}$, состоящего из M занумерованных объектов? На практике для этого используются *датчики псевдослучайных чисел* — алгоритмы, генерирующие серию номеров выбираемых объектов, причём очередной номер является функцией от предыдущего номера. Функция устроена сложным образом: весьма непросто предвидеть её значение, зная значение аргумента. Можно сказать, что при моделировании (имитации) выбора наудачу случайность подменяется алгоритмической сложностью.

В качестве примера сначала рассмотрим простой, но довольно «плохой» датчик — *метод середины квадрата*, предложенный Джоном фон Нейманом в 1946 году. Зададим некоторое четырёхзначное число n_0. Скажем, n_0 = 7777. Вычислим n_0^2 = 60**4817**29. Выделив четыре средние цифры, получим n_1 = 4817. Затем вычислим n_1^2 = 23**2034**89. Выделив четыре средние цифры, получим n_2 = 2034 и т. д. Поскольку всего четырёхзначных чисел существует 10^4, то процесс генерации номеров зациклится не позднее, чем через 10^4 шагов.[15]

---

[14] Этот частный случай, называемый теоремой Муавра — Лапласа, был доказан Лапласом в 1812 г.

[15] Например, число 3792 сразу воспроизводит само себя: 3792^2 = 14**3792**65.

Более сложный *мультипликативный датчик* работает по следующей схеме. Задаются три целых числа: стартовое число n_0, множитель m и делитель d. Далее последовательно вычисляются n_1, n_2, … согласно формуле

$$n_i = \text{остаток от деления произведения } (m \cdot n_{i-1}) \text{ на число } d.$$

Обычно m и d — очень большие числа. Множитель m выбирают так, чтобы последовательность $n_1, n_2, \ldots$, прежде чем зациклиться, пробегала все возможные значения от 1 до $d-1$. В результате изучения статистических свойств датчика для различных множителей Дж. Фишман и Л. Мур предложили использовать, в частности, $m = 630360016$ и $d = 2^{31} - 1 = 2147483647$.

В программном обеспечении иногда встречаются быстро работающие, но недостаточно качественные датчики. Известный пример — датчик RANDU с $m = 2^{16} + 3 = 65539$ и $d = 2^{31}$, вошедший в SSP — библиотеку научных программ для IBM-360. Чем он плох? Положим $y_i = n_i / d$. Для «хорошего» датчика точки с координатами y_i должны равномерно (всюду одинаково плотно) заполнять отрезок [0, 1]. Более того, если разбить $y_1, y_2, \ldots$ на последовательные тройки (y_{3i-2}, y_{3i-1}, y_{3i}), то точки с такими координатами должны равномерно заполнять трёхмерный единичный куб. Однако для последовательности, генерируемой датчиком RANDU, эти точки располагаются ***в точности*** на 15 плоскостях (рис. 13), заданных уравнениями

$$9y_{3i-2} - 6y_{3i-1} + y_{3i} = k, \quad \text{где } k = -5, \ldots, 9.$$

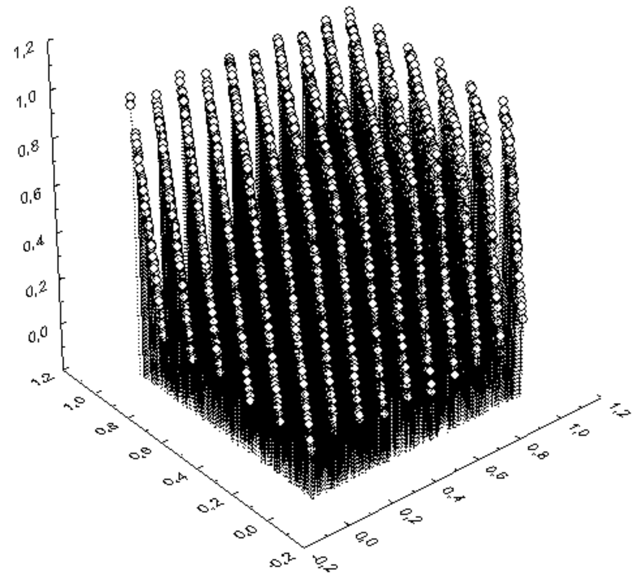

Рис. 13

В языке R по умолчанию используется датчик, называемый *вихрем Мерсенна*.[16] Он основывается на свойствах *простых чисел Мерсенна*[17] вида $2^n - 1$, где n — натуральное число, и обеспечивает быструю генерацию высококачественных псевдослучайных чисел y_i, равномерно распределённых на отрезке [0, 1]. Доказано, что если разбить y_1, y_2, … на последовательные наборы длины 623, то наборы будут равномерно заполнять 623-мерный единичный куб. Вихрь Мерсенна имеет огромный период, равный числу Мерсенна $2^{19937} - 1$. Существуют эффективные реализации вихря Мерсенна, превосходящие по скорости многие стандартные генераторы (в частности, работающие в 2-3 раза быстрее мультипликативных датчиков).

---

[16] Matsumoto M., Nishimura T. «Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator», ACM Transactions on Modeling and Computer Simulation, 1998, 8, 3–30.

[17] Марен Мерсенн (1588 – 1648) — французский математик, физик, философ и теоретик музыки.

Задачи

Попробуйте решить некоторые из них, запишите решения на листе бумаги и сдайте.

1. Пусть ξ — число очков, выпадающее на игральной кости. Вычислить $\mathbf{M}\xi$ и $\mathbf{D}\xi$.

2. Случайная величина η равномерно распределена на отрезке [0, 1]. Вычислить $\mathbf{D}\eta$.

3. Найти математическое ожидание экспоненциальной случайной величины τ (см. формулу (9)).

4. Доказать, что минимум функции $f(a)=\sum(x_i-a)^2 m_i$ достигается при $a=x_c$, где x_c — координата центра масс, заданная формулой (11).

5. Тест состоит из 20 вопросов, на каждый из которых предлагается 4 варианта ответа. Только один вариант правильный. С какой вероятностью студент даст 14 и более правильных ответов, если он выбирает варианты ответов наудачу? Запишите ответ в виде суммы и вычислите приближённое значение искомой вероятности.

6. Пусть S_n имеет биномиальное распределение с параметрами n и p (см. формулу (4)). Доказать, что $\mathbf{M}S_n = np$. (*Указание*. Для вычисления $\mathbf{M}S_n$ выразите числа сочетаний C_n^k через факториалы, сократите k и используйте бином Ньютона.)

7. Пусть S_n имеет биномиальное распределение с параметрами n и p (см. формулу (4)). Доказать, что $\mathbf{D}S_n = np(1-p)$. (*Указание*. При подсчёте $\mathbf{D}S_n$ представьте k^2 в виде $k(k-1)+k$.)

8*. Найти $\mathbf{M}\nu$, где ν имеет геометрическое распределение (см. формулу (5)).