Л.Я. САВЕЛЬЕВ

ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ВЕРОЯТНОСТЕЙ

1

Министерство образования и науки РФ

Новосибирский государственный университет

Л.Я. Савельев

ЭЛЕМЕНТАРНАЯ ТЕОРИЯ ВЕРОЯТНОСТЕЙ

Часть 1

Новосибирск 2005

ББК 22.171
УДК 519.21

Савельев Л.Я. Элементарная теория вероятностей, 1.
Учебное пособие. Новосибирский государственный университет. Новосибирск, 2005. (158 с)

ISBN 5-94356-274-5

Первая часть книги посвящена теории. В ней сначала подробно описываются конечные вероятностные пространства. Чтобы читать ее, достаточно уметь оперировать с конечными суммами и произведениями. Переход к счетным пространствам требует знакомства с рядами. А для непрерывных пространств нужны производные и интегралы.

В книге много примеров. Задачи с решениями составляют вторую часть книги. Там повторяются основные
определения и формулы. Это позволяет при желании читать вторую часть независимо от первой.

Книга может быть полезна для изучающих теорию вероятностей и для имеющих дело с ее элементарными приложениями.

Рецензент канд. физ.-мат. наук Н. И. Чернова

Теорию вероятностей можно описательно определить как математическую теорию случайных явлений. С влиянием случая приходится сталкиваться при описании самых разных явлений. Часто это влияние настолько существенно, что им нельзя пренебречь. Поэтому теория вероятностей применяется во всех развитых областях науки. Прекрасным руководством по дискретной теории вероятностей является книга [11]. С различными приложениями можно познакомиться по книгам [9], [10]. Книга посвящена элементарной теории вероятностей. Вначале рассматриваются конечные вероятностные пространства. Для того, чтобы использовать эту простейшую модель, достаточно уметь оперировать с конечными суммами и произведениями. Переход к счетным пространствам требует знакомства с рядами. Для описываемых по аналогии пространств с плотностью нужны некоторые простые интегралы. Применяемый математический аппарат можно считать элементарным. Это и определило название книги. Все нужные сведения по алгебре и анализу можно найти в "Энциклопедии элементарной математики" [1] – [3] и "Лекциях по математическому анализу" [4], [5], а также в стандартных руководствах. Некоторые нужные сведения есть в приложении. В книге много примеров. Они компенсируют формальный стиль изложения и помогают лучше понимать содержание. Задачи с решениями составляют отдельную часть. Там повторяются определения нужных понятий, связанные с ними факты и формулы. Это позволяет при желании читать часть, посвященную задачам, независимо от остальных. Примеры и задачи дают некоторое представление о приложениях элементарной теории вероятностей в самых различных областях.

Благодарю С. В. Балакина и Н. И. Чернову за помощь, критику и советы.

Содержание

Теория

В этом разделе кратко описывается элементарная теория вероятностей.

1. Конечные вероятностные пространства

Эти пространства — самые простые. Они используются для исследования случайных событий, описываемых конечными множествами исходов.

1.1. Элементарная вероятность

Элементарная вероятность — основное понятие элементарной теории вероятностей. Все остальные понятия этой теории прямо или косвенно определяются через элементарную вероятность.

1.1.1. Пусть U — конечное множество и p — положительная нормированная функция на U:

$$p\,(u) \geqslant 0, \qquad \sum p\,(u) = 1 \qquad (u \in U)\,.$$

Вместе они составляют *конечное вероятностное пространство* (U, p). Элементы множества U называются *возможными исходами* или просто *исходами*, функция p — *элементарной вероятностью*. Значение $p\,(u)$ называется *вероятностью исхода u*. Элементарную вероятность p, не имеющую нулевых значений, условимся называть *невырожденной*, а пространство (U, p) с такой вероятностью — *невырожденным*. В этой главе рассматриваются только конечные вероятностные пространства, которые коротко называются *пространствами*. Исключив исходы с нулевыми вероятностями, можно любое пространство превратить в невырожденное. Иногда, правда, в соответствии с содержанием задачи удобно, наоборот, добавить исходы с нулевыми вероятностями.

1.1.2. Обозначим n число элементов множества U и положим $p(u) = 1/n$ для каждого $u \in U$. Ясно, что функция p с такими значениями есть элементарная вероятность на U. Вероятностное пространство (U, p) будем называть *пространством Лапласа* с n исходами и обозначать (L, n).

1.1.3. Пример. Обозначим $\mathbb{B}$ множество из элементов 0 и 1, а $\mathbb{B}^n$ — множество всех последовательностей из n элементов множества $\mathbb{B}$:

$$\mathbb{B} = \{0, 1\}, \qquad \mathbb{B}^n = \{u = (u_i) : u_i \in \mathbb{B}, i = 1, \ldots, n\}.$$

Будем называть последовательности из $\mathbb{B}^n$ *двоичными последовательностями* длины n.

Пусть $U = \mathbb{B}^n$, $u = (u_i) \in U$, $s(u) = \sum u_i$, $a \in [0, 1]$ и

$$p(u) = a^{s(u)} (1 - a)^{n - s(u)}$$

($s(u)$ равно числу единиц, а $n - s(u)$ — числу нулей в последовательности u). По индукции легко доказать (упражнение), что функция p с такими значениями есть элементарная вероятность на $U = \mathbb{B}^n$. Вероятностное пространство $(\mathbb{B}^n, p)$ с такой элементарной вероятностью будем называть *пространством Бернулли* с параметрами n, a и обозначать $B(n, a)$. Говорят также о *последовательности Бернулли* длины n с параметром a.

Заметим, что $B(n, 1/2) = (L, 2^n)$.

1.1.4. Пример. Пусть

$$U = \mathbb{B}^{1+n} = \{u = (u_i) : u_i \in \mathbb{B}, i = 0, 1, \ldots, n\},$$

$a \in [0, 1]$ и $Q = (q\,(x, y))$ — *стохастическая* 2×2*–матрица*:

$$q\,(x, y) \geqslant 0, \qquad q\,(x, 0) + q\,(x, 1) = 1 \qquad (x, y \in \mathbb{B})\,.$$

По индукции легко доказать (упражнение), что функция p со значениями

$$p\,(u) = a^{u_0}\,(1 - a)^{1-u_0}\,q\,(u_0, u_1) \ldots q\,(u_{n-1}, u_n)$$

является элементарной вероятностью на $U = \mathbb{B}^{1+n}$. Вероятностное пространство $\left(\mathbb{B}^{1+n}, p\right)$ с такой элементарной вероятностью будем называть *двоичным пространством Маркова* с параметрами n, a и оператором Q. Условимся обозначать такое пространство $BM\,(n, a, Q)$. Говорят также о *двоичной марковской последовательностью* длины $n{+}1$ с параметром a и оператором Q. Марковские последовательности часто называют *марковскими цепями*.

Заметим, что $BM\,(n, a, Q) = B\,(n + 1, a)$ при $q\,(1, 1) = a$, $q\,(0, 0) = 1 - a$.

1.2. Вероятность и среднее

Как правило задачи, решаемые с помощью теории вероятностей, сводятся к вычислению вероятностей случайных событий или средних значений случайных переменных.

1.2.1. Возьмем конечное вероятностное пространство (U, p). Каждое множество $X \subseteq U$ называется *случайным событием* или просто *событием* в пространстве (U, p). Число

$$P\,(X) = \sum_{u \in X} p\,(u)$$

называется *вероятностью события* X. Эта вероятность складывается из вероятностей исходов u, составляющих событие X (*благоприятных* для него). Выделяются пустое событие $X = \varnothing$ и полное событие $X = U$. Из определений следует, что

$$P(\varnothing) = 0, \qquad\qquad P(U) = 1.$$

Событие $X = \{u\}$, состоящее из одного единственного исхода u, называется *элементарным*. Оно часто отождествляется с составляющим его исходом: $\{u\} = u$. Из определений следует, что $P(\{u\}) = p(u)$.

Существенная разница между элементарной вероятностью p и вероятностью P состоит в том, что p определена на множестве U, а P — на классе $\mathcal{P} = \mathcal{P}(U)$ всех частей множества U. Поэтому оперировать с вероятностью P непосредственно гораздо сложнее, чем через элементарную вероятность p. Хотя интерес, как правило, представляют вероятности событий, а не вероятности отдельных исходов, из которых они складываются.

Элементарную вероятность p коротко тоже будем называть вероятностью, как P, сохраняя разницу в обозначениях.

1.2.2. Каждая функция на множестве U называется *случайной переменной* на вероятностном пространстве (U, p). Всюду, где нет специальных оговорок, имеются в виду *вещественные* случайные переменные. Их часто называют *случайными величинами*.

Множество $X \subseteq U$ удобно отождествлять с его *индикатором* $\operatorname{ind} X$ на U:

$$\operatorname{ind} X(u) = 1 \qquad (u \in X), \qquad \operatorname{ind} X(u) = 0 \qquad (u \notin X).$$

Условимся для простоты часто вместо *индикатор множества* говорить *множество* и писать X вместо $\operatorname{ind} X$. Отождествление множеств с их индикаторами позволяет считать события специальными случайными переменными и многие рассуждения проводить только для случайных переменных. Случайные переменные будем коротко называть *переменными*.

Будем часто для простоты вместо $\{u\}$ писать u, используя эту букву в разных смыслах.

Каждую переменную f на множестве U можно записать в виде линейной комбинации индикаторов $\operatorname{ind} u$ исходов $u \in U$:

$$f = \sum f(u) \operatorname{ind} u \qquad\qquad (u \in U)\,.$$

По определению $\operatorname{ind} u(v) = 1$ при $v = u$ и $\operatorname{ind} u(v) = 0$ при $v \neq u$. Поэтому сумма в равенстве имеет значение $f(u)$ для каждого $u \in U$. Группируя исходы, в которых переменная f имеет равные значения, получаем удобную *стандартную запись* для f:

$$f = \sum x \operatorname{ind} f^{-1}(x) \qquad\qquad (x \in X = f(U))\,.$$

Соглашение об отождествлении множеств с их индикаторами позволяет упростить эти записи:

$$f = \sum f(u)\, u \qquad (u \in U)\,, \qquad f = \sum x\, f^{-1}(x) \qquad (x \in X)\,.$$

Здесь u одновременно обозначает исход и его индикатор. Слагаемые в суммах являются функциями на U и на X.

1.2.3. Число

$$Ef = \sum f(u)\, p(u)$$

называется (*вероятностным* или *стохастическим*) *средним значением* переменной f на пространстве (U, p). Среднее значение называют также *математическим ожиданием*.

Из определений следует, что

$$E(\operatorname{ind} X) = \sum_{u \in X} p(u) = P(X)\,.$$

Вероятность события X равна среднему значению его индикатора. Это позволяет выводить формулы для вероятностей из формул для средних, используя их свойства.

1.2.4. Пример. Возьмем произвольную случайную переменную f на пространстве Лапласа с множеством U из n исходов. Среднее значение f равно среднему арифметическому:

$$Ef = \sum f(u)\,\frac{1}{n} = \frac{1}{n}\sum f(u).$$

Пусть X — событие, составленное из m исходов и $f = \operatorname{ind} X$. Тогда

$$P(X) = E(\operatorname{ind} X) = m/n.$$

В пространстве Лапласа вероятность события равна отношению числа благоприятных для него исходов к числу всех возможных исходов. Ее вычисление сводится к подсчету числа элементов в рассматриваемых множествах. Поэтому пространство Лапласа можно назвать *комбинаторным пространством*.

1.2.5. Задача д'Аламбера. Симметричная монета подбрасывается два раза. Игрок выигрывает, если оба раза монета падает гербом вверх. Какова вероятность выигрыша?

В статье «Герб и решетка» (1754) д'Аламбер рассуждал так. Возможны три исхода: монета при первом подбрасывании падает решкой вверх, и поэтому второй раз бросать ее нет смысла (0); монета в первый раз падает гербом вверх, а во второй — решкой (1); монета оба раза падает гербом вверх (2). Благоприятным является один исход из трех возможных. Следовательно, вероятность выигрыша равна $1/3$. Использовано комбинаторное пространство с тремя равновероятными исходами. Но их равновероятность никак не обоснована: исход 0 представляется более вероятным, чем исходы 1 и 2, так как в нем не учитывается результат второго подбрасывания.

Если этот результат учитывать, то пространство $(L, 3)$ нужно заменить на пространство $B(4, 1/2) = (L, 4)$ с множеством исходов $U = \{00, 01, 10, 11\}$ и вероятность выигрыша будет меньше:

$p(11) = (1/2)^2 = 1/4$. Такая формализация больше соответствует содержанию задачи, так как здесь равновозможность исходов отражает симметричность монеты (остальные неявные предположения одинаковы). Кроме того, в правильной модели вероятность выигрыша должна быть близка частоте появления пар гербов при многократном подбрасывании двух монет в одинаковых условиях. Статистика свидетельствует, что эта частота близка 1/4. (Упражнение: провести соответствующий компьютерный эксперимент.)

1.2.6. Пример. Случайная величина s_j со значениями

$$s_j(u) = u_j \qquad\qquad (u = (u_i) \in \mathbb{B}^n)$$

описывает число единиц на j-м месте ($1 \leqslant j \leqslant n$) в последовательности Бернулли $B(n, a)$. По индукции легко проверить (упражнение), что

$$\begin{aligned} &P(s_j = 1) = P\{u \in \mathbb{B}^n : u_j = 1\} = \\ &\qquad = \sum_{u_i, i \neq j} \Big(a^{u_1} (1-a)^{1-u_1} \dots a^{u_{j-1}} (1-a)^{1-u_{j-1}} \cdot a \times \\ &\times a^{u_{j+1}} (1-a)^{1-u_{j+1}} \dots a^{u_n} (1-a)^{1-u_n} \Big) = a(a + 1 - a)^{n-1} = a. \end{aligned}$$

Следовательно,

$$P(s_j = 0) = 1 - P(s_j = 1) = 1 - a.$$

Подчеркнем, что эти вероятности не зависят от номера места j в последовательности u.

Пусть $s = \sum s_j$. Тогда

$$P(s = m) = \binom{n}{m} a^m (1-a)^{n-m} = b(m, n, a)$$

для $m = 0, 1, \dots, n$. Сумма $s(u) = \sum s_j(u)$ равна числу единиц в двоичной строке $u = u_1 \dots u_n$. Говорят, что переменная s имеет *биномиальное распределение*.

1.2.7. Пример. Для двоичной марковской последовательности $BM(n, a, Q)$ аналогичные вероятности вычисляются сложнее. Они зависят от номера места k $(0 \leqslant k \leqslant n)$.

Используя определения в 1.1.4 и индукцию, нетрудно проверить (упражнение), что в марковском случае

$$P(s_k = 1) = P\{u \in \mathbb{B}^{n+1} : u_k = 1\} =$$
$$= \sum_{u_i, i \neq k} \Big(a^{u_0} (1-a)^{1-u_0} q(u_0, u_1) \dots q(u_{k-1}, 1) \times$$
$$\times q(1, u_{k+1}) \dots q(u_{n-1}, u_n)\Big) = b + (a-b)\, d^k,$$

где

$$d = q(1,1) + q(0,0) - 1, \qquad b = (1 - q(0,0)) \big/ (1-d)$$

(предполагается, что $d \neq 1$).

Следовательно,

$$P(s_k = 0) = 1 - b - (a-b)\, d^k.$$

В частности,

$$P(s_0 = 1) = a, \qquad P(s_0 = 0) = 1 - a.$$

Если $q(1,1) = a$, $q(0,0) = 1 - a$, то $d = 0$ и $b = a$. Марковская последовательность превращается в бернулевскую и

$$P(s_k = 1) = a, \qquad P(s_k = 0) = 1 - a$$

для каждого $k = 0, 1, \dots, n$.

1.3. Свойства среднего

Основные свойства среднего можно выразить одной фразой: среднее есть нормированный положительный линейный функционал на алгебре случайных переменных. Эти свойства среднего существенно помогают при вычислениях.

1.3.1. Рассмотрим множество $\mathcal{F} = \mathcal{F}(U)$ всех случайных переменных на вероятностном пространстве (U, p). Их можно складывать, перемножать и умножать на числа с соблюдением обычных правил действий с функциями. Поэтому $\mathcal{F}$ называют *алгеброй.* Среднее E каждой функции $f \in \mathcal{F}$ ставит в соответствие число Ef. Поэтому E называется *функционалом* на $\mathcal{F}$.

Линейность среднего E означает, что

$$E(f+g) = Ef + Eg, \qquad\qquad E(cf) = cEf$$

для каждых $f, g \in \mathcal{F}$ и числа c. *Положительность* и *нормированность* — что $Ef \geqslant 0$ для каждой $f \geqslant 0$ и $E(1) = 1$. Неравенство $f \geqslant 0$ эквивалентно $f(u) \geqslant 0$ для всех $u \in U$, а символ 1 обозначает число 1 и постоянную на U со значением 1. Формально постоянные тоже являются случайными величинами. Они обозначаются так же, как их значения. Из нормированности и линейности среднего следует равенство $Ec = c$ для каждой постоянной c.

Все указанные свойства среднего E легко проверяются:

$$\begin{aligned}
&E(f+g) = \sum (f(u) + g(u))\, p(u) = \\
&\qquad\qquad = \sum f(u)\, p(u) + \sum g(u)\, p(u) = Ef + Eg, \\
&E(cf) = \sum cf(u)\, p(u) = c \sum f(u)\, p(u) = cEf, \\
&Ef = \sum f(u)\, p(u) \geqslant 0 \qquad (f(u) \geqslant 0), \\
&E1 = \sum p(u) = 1.
\end{aligned}$$

Из линейности и положительности среднего следует его *монотонность*:

$$Ef \geqslant Eg \qquad\qquad (f \geqslant g).$$

В самом деле,

$$f \leqslant g \Leftrightarrow g - f \geqslant 0 \Rightarrow Eg - Ef = E\,(g - f) \geqslant 0 \Leftrightarrow Ef \leqslant Eg.$$

Заметим, что $E\,(-f) = -Ef$. Положительность функционала E не означает, что у него нет отрицательных значений. Неравенство $Ef \geqslant 0$ гарантируется только для $f \geqslant 0$.

1.3.2. Пример. Рассмотрим случайную переменную s на пространстве $B\,(n, a)$ — число единиц в последовательности Бернулли длины n с параметром a. Пусть $s_j\,(u) = u_j$ для $u = (u_i) \in \mathbb{B}^n$. Из определений следует, что

$$s\,(u) = \sum_{j=1}^{n} u_j = \sum_{j=1}^{n} s_j\,(u) \Leftrightarrow s = \sum_{j=1}^{n} s_j.$$

Поэтому

$$Es = E\left(\sum_{j=1}^{n} s_j\right) = \sum_{j=1}^{n} Es_j.$$

Но $Es_j = P\,(s_j = 1) = a$ (1.2.6). Следовательно,

$$Es = na.$$

В частности, $Es = n/2$ при $a = 1/2$. Приблизительно таким бывает число гербов при n-кратном подбрасывании симметричной монеты. Если $a = 1/4$, то $Es = n/4$. Приблизительно таким бывает число пар гербов при n-кратном подбрасывании двух симметричных монет. (Предполагается, что n достаточно большое.)

1.3.3. Пример. Для двоичной марковской последовательности $B\,(n, a, Q)$ формула для среднего числа единиц сложнее.

Используя равенства (1.2.7)

$$Es_k = P\,(s_k = 1) = b + (a - b)\,d^k$$

и формулу для геометрической прогрессии, получаем:

$$Es = \sum_{k=0}^{n} s_k = \sum_{k=0}^{n} \Big(b + (a-b)\, d^k \Big) =$$
$$= nb + (b+c) - c\, d^{n+1} = (n+1)\, b + c\left(1 - d^{n+1}\right),$$

где

$$c = (a-b) \Big/ (1-d)$$

(предполагается, что $d \neq 1$).

В частности, при $d = 0$ имеем:

$$Es = a + nb.$$

Если, кроме того, $b = a$, то $B\,(n, a, Q)$ превращается в $B\,(n+1, a)$ и $Es = (n+1)\, a$.

1.3.4. Пример. Для произведения $s_j s_k$ $(j \neq k)$ случайных переменных s_j, s_k на пространстве Бернулли $B\,(n, a)$ верно равенство

$$E\,(s_j s_k) = Es_j Es_k.$$

В самом деле, немного изменяя рассуждения 1.2.6, получаем (упражнение):

$$E\,(s_j s_k) = P\,(s_j = s_k = 1) = a^2\,(a + 1 - a)^{n-2} = a^2.$$

1.3.5. Для двоичной марковской последовательности в общем случае $E\,(s_j s_k) \neq Es_j Es_k$.

В самом деле, немного изменяя рассуждения 1.2.7, получаем (упражнение):

$$E\,(s_j s_k) = \left(b + (a-b)\, d^j\right) q\,(1,1)\,.$$

В то же время

$$Es_j Es_k = \left(b + (a-b)\, d^j\right) \left(b + (a-b)\, d^{j+1}\right).$$

В частности, при $j = 0$ имеем:

$$E\,(s_0 s_1) = aq\,(1,1)\,,$$
$$Es_0 = a, \qquad\qquad Es_1 = aq\,(1,1) + (1-a)\,q\,(0,1)\,.$$

Пусть $a = 1/2$, $q\,(1,1) = 1/4$, $q\,(0,1) = 3/4$. Тогда

$$E\,(s_0 s_1) = 1/8 \neq (1/2)\,(1/2) = Es_0 Es_1.$$

1.3.6. Соглашение об отождествлении событий с их индикаторами позволяет считать события специальными случайными переменными и отождествлять вероятности событий со средними значениями этих переменных. При определенной аккуратности использование одних и тех же обозначений для событий и их индикаторов не приводит к путанице и часто существенно упрощает запись формул.

Пусть A, B — события, отождествленные со своими индикаторами. Тогда $A \cap B = AB$ — пересечение событий A и B, $A \cup B = A + B - AB$ — их объединение, $A \setminus B = A - AB$ — дополнение B до A, отождествленные с соответствующими алгебраическими выражениями для их индикаторов. Если $AB = \varnothing$, то $A \cup B = A + B$. Если $B \subset A$, то $A \setminus B = A - B$. Так как среднее линейно, то:

$$E\,(A + B - AB) = E\,(A) + E\,(B) - E\,(AB)\,,$$
$$E\,(A - AB) = E\,(A) - E\,(AB)\,.$$

Эти равенства для средних эквивалентны следующим равенствам для вероятностей:

$$P\,(A \cup B) = P\,(A) + P\,(B) - P\,(A \cap B)\,,$$
$$P\,(A \setminus B) = P\,(A) - P\,(A \cap B)\,.$$

В частности,

$$P\,(A \cup B) = P\,(A) + P\,(B) \qquad\qquad (A \cap B = \varnothing)\,,$$
$$P\,(A \setminus B) = P\,(A) - P\,(B) \qquad\qquad (B \subset A)\,.$$

1.4. Дисперсия

Среднее значение очень грубо описывает случайную переменную: ее значения могут сильно отличаться от среднего. Дисперсия служит мерой этих отклонений.

1.4.1. Число
$$Df = E\left((f - Ef)^2\right)$$
называется *дисперсией* случайной переменной f на вероятностном пространстве (U, p). Из определений следует, что
$$Df = \sum (f(u) - Ef)^2 \, p(u).$$

Свойства дисперсии D определяются свойствами среднего E.

1.4.2. Так как $(f - Ef)^2 \geqslant 0$, то
$$Df \geqslant 0.$$
А так как среднее постоянной c равно ее значению, то
$$Dc = \sum (c - c)^2 p(u) = 0.$$
Если рассматриваемое вероятностное пространство (U, p) невырожденное ($p(u) > 0$ для всех $u \in U$), то нулевую дисперсию имеют только постоянные.

Прибавление постоянной к случайной переменной не изменяет ее дисперсии:
$$D(f + c) = E\left((f + c - E(f + c))^2\right) =$$
$$= E\left((f + c - Ef - Ec)^2\right) = E\left((f - Ef)^2\right) = Df.$$

1.4.3. Дисперсия не является линейным функционалом на $\mathcal{F}$:

$$D\,(cf) = E\left((cf - E\,(cf))^2\right) = E\left(c^2\,(f - Ef)^2\right) = c^2 Df,$$

$$D\,(f+g) = E\left((f + g - E\,(f+g))^2\right) = E\left((f - Ef)^2 + \right.$$

$$\left. + \,(g - Eg)^2 + 2E\,((f - Ef)\,(g - Eg))\right) = Df + Dg + 2C\,(f,g)\,,$$

где

$$C\,(f,g) = E\,((f - Ef)\,(g - Eg)) = E\,(fg) - EfEg.$$

Число $C\,(f,g)$ называется *корреляцией* случайных переменных f и g. Корреляция описывает линейную зависимость между ними. Если $C\,(f,g) = 0$, то

$$D\,(f+g) = Df + Dg.$$

Верно и обратное утверждение.

1.4.4. Число

$$Sf = \sqrt{Df} \geqslant 0$$

называется *стандартным отклонением* или, коротко, *стандартом* случайной переменной f. Извлечение корня уравновешивает квадрат $(f - Ef)^2$ в определении дисперсии. Стандарт обладает свойством положительной однородности:

$$S\,(cf) = |c|\,S\,(f)\,.$$

Если $c \geqslant 0$, то правая часть равна $cS\,(f)$.

Стандарт часто используется вместо дисперсии для оценки разброса значений случайной переменной вокруг среднего значения.

1.4.5. Для вычисления дисперсии удобно использовать формулу

$$Df = E\left(f^2\right) - (Ef)^2 .$$

Она доказывается так:

$$E\left((f - Ef)^2\right) = E\left(f^2 - 2f \cdot Ef + (Ef)^2\right) =$$
$$= E\left(f^2\right) - 2E\left(f \cdot Ef\right) + E\left((Ef)^2\right) =$$
$$= E\left(f^2\right) - 2\left(Ef\right)^2 + (Ef)^2 = E\left(f^2\right) - (Ef)^2 .$$

В частности, если $f = X$ — событие, то $X^2 = X$, $EX = P\left(X\right)$ и

$$DX = P\left(X\right) - \left(P\left(X\right)\right)^2 = P\left(X\right)\left(1 - P\left(X\right)\right) = P\left(X\right) P\left(X'\right) .$$

Дисперсия события X равна произведению его вероятности на вероятность его дополнения $X' = U - X$.

Для стандарта верно равенство

$$S\left(X\right) = \sqrt{P\left(X\right) P\left(X'\right)}.$$

1.4.6. **Пример.** Случайные переменные s_j на пространстве Бернулли $B\left(n, a\right)$ являются событиями и $Es_j = a$ $(j = 1, \ldots, n)$. Поэтому $Ds_j = a\left(1 - a\right)$.

Корреляции $C\left(s_j, s_k\right)$ при $j \neq k$ равны 0:

$$E\left(\left(s_j - a\right)\left(s_k - a\right)\right) = E\left(s_j s_k - s_j a - a s_k + a^2\right) =$$
$$E\left(s_j s_k\right) - aEs_j - aEs_k + a^2 = a^2 - a^2 - a^2 + a^2 = 0,$$

так как $E\left(s_j s_k\right) = a^2$ (1.3.4). Следовательно,

$$D\left(s_j + s_k\right) = Ds_j + Ds_k = 2a\left(1 - a\right) .$$

По индукции нетрудно обобщить эту формулу (упражнение) на n слагаемых $s_1, \ldots, s_n$. Дисперсия числа единиц в последовательности Бернулли длины n с параметром a выражается равенством

$$Ds = na\,(1-a)\,.$$

Стандартное отклонение числа единиц от среднего значения na имеет порядок $\sqrt{n}$.

1.4.7. Во многих прикладных задачах используется элементарная вероятность p со значениями

$$p\,(k) = \binom{l}{k}\binom{n-l}{m-k} \Big/ \binom{n}{m} = g\,(k, l, m, n)$$

на множестве $U = \{0, 1, \ldots, l\}$. Здесь l, m, n целые числа, удовлетворяющие неравенствам $0 \leqslant l \leqslant n$, $1 \leqslant m \leqslant n$. Вероятность p называется *гипергеометрическим распределением* с параметрами l, m, n.

Равенство $\sum p\,(k) = 1$ эквивалентно равенству

$$\sum_k \binom{l}{k}\binom{n-l}{m-k} = \binom{n}{m},$$

которое проверяется сравнением коэффициентов при x^m в равенстве

$$(1+x)^l\,(1+x)^{n-l} = (1+x)^n$$

после применения биномиальной формулы.

Вычислим среднее значение тождественной переменной f на U. Как легко проверить (упражнение), выразив биномиальные коэффициенты через факториалы,

$$k \cdot g\,(k, l, m, n) = (lm/n) \cdot g\,(k-1, l-1, m-1, n-1)\,.$$

Поэтому

$$E(f) = \sum_{k \geqslant 0} k \cdot g(k,l,m,n) =$$
$$= \frac{lm}{n} \sum_{k \geqslant 1} g(k-1,l-1,m-1,n-1) = \frac{lm}{n}.$$

Вычислим дисперсию f. Как легко проверить (упражнение),

$$k(k-1) \cdot g(k,l,m,n) =$$
$$= \frac{lm}{n} \frac{(l-1)(m-1)}{n-1} \cdot g(k-2,l-2,m-2,n-2).$$

Следовательно,

$$D(f) = \frac{lm}{n} \frac{(l-1)(m-1)}{n-1} + \frac{lm}{n} - \left(\frac{lm}{n}\right)^2 = \frac{n-m}{n-1} \frac{lm}{n} \left(1 - \frac{l}{n}\right).$$

Отметим связь гипергеометрического распределения с биномиальным, имеющим параметры m, l/n: переменные с этим распределением имеют одно и то же среднее значение, а их дисперсии отличаются коэффициентом $(n-m)/(n-1)$. Переменная f равна сумме $s_1 + \ldots + s_m$ двоичных переменных с одним и тем же распределением $P(s_i = 1) = l/n$, $P(s_i = 0) = 1 - l/n$. В биномиальном случае их естественно считать независимыми, а в гипергеометрическом — зависимыми.

1.4.8. Для описанной в 1.1.4 двоичной марковской последовательности $BM(n,a,Q)$ выражение для дисперсии числа единиц $s = \sum s_j$ гораздо сложнее, так как $C(s_j, s_k) \neq 0$ при $j \neq k$ и $Ds \neq \sum Ds_j$ при $d \neq 0$ (1.3.5).

Чтобы упростить формулы, введем для двоичных марковских последовательностей специальные обозначения:

$$p = q(1,1), \quad q = q(0,0), \quad 1-p = q(1,0), \quad 1-q = q(0,1);$$

$$
\begin{array}{lll}
d = p + q - 1, & b = (1-q)\,/\,(1-d)\,, & c = (a-b)\,/\,(1-d)\,; \\
\alpha = (1-b)\,/\,(1-d)\,, & \beta = b/\,(1-d)\,, & \gamma = (1+d)\,/\,(1-d)\,.
\end{array}
$$

В 1.2.7 и 1.3.3 были выписаны формулы для вероятностей и средних

$$
\begin{array}{ll}
P\,(s_j = 1) = b + (a-b)\,d^j, & P\,(s_j = 0) = 1 - b - (a-b)\,d^j; \\
Es_j = b + (a-b)\,d^j, & Es = (n+1)\,b + c\left(1 - d^{n+1}\right).
\end{array}
$$

Следующие несколько пунктов посвящены подробному вычислению дисперсии Ds для двоичной марковской последовательности.

1.4.9. Вычислим средние значения $E(s_j s_k) = P(s_j = s_k = 1)$. Докажем по индукции, что при $j \leqslant k$

$$
E\,(s_j s_k) = P\,(s_j = 1)\left(b + (1-b)\,d^{k-j}\right).
$$

Если $k = j$, то $s_j s_k = s_j^2$ и это равенство сводится к равенству для $P\,(s_j = 1)$, которое выписано в 1.2.7. Его легко доказать: если $j = 0$, то

$$
P\,(s_0 = 1) = a = b + (a-b) = b + (a-b)\,d^0.
$$

Если $j \geqslant 0$ и $P\,(s_j = 1) = b + (a-b)\,d^j$, то

$$
\begin{aligned}
P\,(s_{j+1} = 1) &= P\,(s_j = 1)\,q\,(1,1) + P\,(s_j = 0)\,q\,(0,1) = \\
&= P\,(s_j = 1)\,p + (1 - P\,(s_j = 1))\,(1-q) = P\,(s_j = 1)\,d + 1 - q = \\
&= \left(b + (a-b)\,d^j\right)d + b\,(1-d) = b + (a-b)\,d^{j+1}.
\end{aligned}
$$

Значит, доказываемое равенство для $P\,(s_j = 1)$ верно для всех номеров $j \geqslant 0$.

1.4.10. Докажем теперь равенство для $E\left(s_j s_k\right)$. Если $k=j$, то оно верно:

$$E\left(s_j s_k\right)=E\left(s_j^2\right)=P\left(s_j=1\right)=P\left(s_j=1\right)\left(b+(1-b)\,d^0\right).$$

Возьмем произвольный номер $k \geqslant j$ и предположим, что для него доказываемое равенство верно. Положим

$$p_j=P\left(s_j=1\right), \quad q_j=P\left(s_j=0\right), \quad p_{jk}=P\left(s_j=1, s_k=1\right).$$

Имеем:

$$\begin{gathered}
E\left(s_j s_{k+1}\right)=P\left(s_j s_{k+1}=1\right)=P\left(s_j=1, s_{k+1}=1\right)= \\
=P\left(s_j=1, s_k=1\right) q\,(1,1)+P\left(s_j=1, s_k=0\right) q\,(0,1)= \\
=p_{jk} p+\left(p_j-p_{jk}\right)(1-q)=p_{jk} d+p_j\,(1-q)= \\
=p_j\left(b+(1-b)\,d^{k-j}\right) d+p_j b\,(1-d)=p_j\left(b+(1-b)\,d^{k+1-j}\right).
\end{gathered}$$

Значит, доказываемое равенство для $E\left(s_j s_k\right)$ верно для всех $k \geqslant j$.

Заметим, что

$$C\left(s_j, s_k\right)=p_j q_j d^{k-j}.$$

В самом деле,

$$\begin{aligned}
C\left(s_j, s_k\right)=E\left(s_j s_k\right)-E s_j E s_k=p_j\left(b+(1-b)\,d^{k-j}\right)-p_j p_k= \\
=p_j\left(1-b-(a-b)\,d^j\right) d^{k-j}=p_j q_j d^{k-j}.
\end{aligned}$$

Если $d=0$, то $C\left(s_j, s_k\right)=0$.

1.4.11. Вычислим среднее значение переменной s^2. Заметим, что так как $s_j^2=s_j$, то

$$E\left(s^2\right) = E\left(\left(\sum_{j=0}^{n} s_j\right)^2\right) = E\left(\sum_{j=0}^{n} s_j^2\right) +$$
$$+ 2E\left(\sum_{j=0}^{n-1}\sum_{k=j+1}^{n} s_j s_k\right) = E\left(s\right) + 2\sum_{j=0}^{n-1}\sum_{k=j+1}^{n} E\left(s_j s_k\right).$$

Подставляя найденные в 1.3.3 и 1.4.8 значения $E\left(s\right)$ и $E\left(s_j s_k\right)$, используя формулы для сумм d^j и jd^j после несложных преобразований получаем (упражнение):

$$E\left(s^2\right) = n^2 b^2 + nb\left(1 + b + 2\left(c + \alpha d\right)\right) +$$
$$+ \left(b + c + 2\left(\left(\alpha - \beta\right)c - \alpha bd\right)\right) -$$
$$- \left(2\left(a - b\right)\alpha + \left(c + 2\left(c\left(\alpha - \beta\right) - \alpha\beta d\right)\right) \big/ n\right) nd^{n+1}.$$

Отметим промежуточные формулы:

$$\sum_{k=j+1}^{n} E\left(s_j s_k\right) = \left(n - j\right) b^2 + \left(a - b\right) b\left(n - j\right) d^j +$$
$$+ \alpha bd - \alpha bd^{n-j+1} + \alpha\left(a - b\right) d^{j+1} - \alpha\left(a - b\right) d^{n+1},$$
$$\sum_{j=0}^{n-1}\left(a - b\right) b\left(n - j\right) d^j = bcn - \beta cd + \beta cd^{n+1}.$$

1.4.12. Как легко проверить (упражнение),

$$\left(E\left(s\right)\right)^2 = n^2 b^2 + 2b\left(b + c\right) n + \left(b + c\right)^2 -$$
$$- \left(2b + \left(2\left(b + c\right) - cd^{n+1}\right) \big/ n\right) cnd^{n+1},$$

откуда, используя формулу из 1.4.5, получаем:

$$D\left(s\right) = E\left(s^2\right) - \left(E\left(s\right)\right)^2 = nb\left(1 - b\right)\gamma + B + C\left(n\right) nd^{n+1},$$

где

$$
\begin{aligned}
&B = (1-b-c)(b+c) + 2((\alpha-\beta)c - \alpha\beta d)\,d,\\
&C(n) = 2bc - 2(a-b)\,\alpha -\\
&\qquad\qquad - \left(c + 2(((\alpha-\beta)-(b+c))\,c - \alpha\beta d) + c^2 d^{n+1}\right) / \; n.
\end{aligned}
$$

В частности, если $a = p = 1 - q$ и марковская последовательность превращается в последовательность Бернулли, то $d = 0$, $b = p$, $c = 0$, $B = p(1-p)$, $d^{n+1} = 0$, $\gamma = 1$, и

$$D(s) = nb(1-b) + b(1-b) = (n+1)\,p(1-p)$$

в соответствии с формулой в 1.4.5. В марковском случае главная часть формулы умножается на коэффициент $\gamma = (1+d)/(1-d)$, характеризующий зависимость между переменными s (с индексом j).

Если $|d| < 1$, то при $n \to \infty$

$$D(s) \sim nb(1-b)\,\gamma.$$

Но когда $|d|$ близко к 1, то $D(s)/(nb(1-b)\,\gamma) \to 1$ довольно медленно.

1.5. Корреляционная теория

Эта теория используется при описании зависимости случайных переменных. Для ее изложения удобен геометрический язык.

1.5.1. Алгебра $\mathcal{F} = \mathcal{F}(U)$ переменных на вероятностном пространстве (U, p) является *векторным пространством*: переменные можно складывать и умножать на числа, соблюдая обычные для векторов правила. Это пространство можно сделать *евклидовым*, если с помощью среднего E определить *скалярное произведение* таких векторов:

$$\langle f, g\rangle = E(fg) \qquad\qquad (f, g \in \mathcal{F}).$$

Свойства скалярного произведения легко проверяются. Оно *симметрично*:

$$\langle f, g \rangle = E\,(fg) = E\,(gf) = \langle g, f \rangle\,;$$

линейно по каждой переменной:

$$\begin{aligned}\langle af + bg, h \rangle = E\,((af + bg)\,h) = E\,(afh + bgh) = \\ = aE\,(fh) + bE\,(gh) = a\,\langle f, h \rangle + b\,\langle g, h \rangle\end{aligned}$$

и по симметрии

$$\langle f, bg + ch \rangle = b\,\langle f, g \rangle + c\,\langle f, h \rangle$$

для каждых переменных $f, g, h \in \mathcal{F}$ и чисел a, b, c;
положительно на диагонали:

$$\langle f, f \rangle = E\left(f^2\right) \geqslant 0.$$

Если вероятность p невырожденная ($p\,(u) > 0$ для каждого исхода $u \in U$), то скалярное произведение $\langle \cdot, \cdot \rangle$ *невырожденное*:

$$\langle f, f \rangle = 0 \Leftrightarrow f = 0.$$

1.5.2. Каждую переменную $f \in \mathcal{F}$ можно центрировать с помощью ее среднего значения Ef и перейти к *центрированной* переменной

$$f_0 = f - Ef$$

со средним значением

$$Ef_0 = E\,(f - Ef) = Ef - Ef = 0.$$

Сумма и произведение на число центрированных переменных тоже центрированы. Значит, центрированные переменные составляют *подпространство* векторного пространства $\mathcal{F}$. Обозначим это подпространство $\mathcal{F}_0 = \mathcal{F}_0(U)$. Каждый вектор пространства $\mathcal{F}$ равен сумме некоторых вектора из $\mathcal{F}_0$ и постоянной:

$$f = f_0 + c, \qquad f_0 = f - Ef, \qquad c = Ef.$$

Основные определения корреляционной теории проще формулировать для центрированных переменных.

Заметим, что

$$Df_0 = E\left(f_0^2\right) = \langle f_0, f_0 \rangle,$$
$$Sf_0 = \sqrt{Df_0} = \sqrt{\langle f_0, f_0 \rangle}$$

для каждой центрированной f_0.

1.5.3. В соответствии с общим определением *нормы* вектора в евклидовом пространстве положим

$$\|f_0\| = \sqrt{\langle f_0, f_0 \rangle} = Sf_0.$$

Норма центрированной переменной равна ее стандарту.

Если $\|f_0\| \neq 0$, то центрированную переменную f_0 можно нормировать и перейти к нормированной переменной

$$\overline{f_0} = f_0 \, / \, \|f_0\|$$

с нормой

$$\left\|\overline{f_0}\right\| = \sqrt{\langle f_0 / \|f_0\|, f / \|f_0\| \rangle} = \sqrt{\langle f_0, f_0 \rangle} \, / \, \|f_0\| = \|f_0\| \, / \, \|f_0\| = 1.$$

Центрированные и нормированные переменные составляют сферу с центром 0 и радиусом 1 в евклидовом пространстве $\mathcal{F}_0$.

Если скалярное произведение для $\mathcal{F}$ невырожденное, то в $\mathcal{F}_0$ нет ненулевых векторов с нулевой нормой.

1.5.4. В соответствии с общим определением меры угла в евклидовом пространстве положим

$$K\left(f,g\right)=\left\langle \overline{f_0},\overline{g_0}\right\rangle ,$$

для каждых переменных f,g и

$$\overline{f_0}=\frac{f-Ef}{\|f-Ef\|},\qquad\qquad \overline{g_0}=\frac{g-Eg}{\|g-Eg\|}.$$

Предполагается, что знаменатели не равны 0. Это часто неявно делается и в дальнейшем. Число $K\left(f,g\right)$ называется *коэффициентом корреляции* между случайными переменными. Используя данные определения, получаем формулу:

$$K\left(f,g\right)=\frac{C\left(f,g\right)}{SfSg}=\frac{E\left(fg\right)-EfEg}{\sqrt{Df}\sqrt{Dg}}$$

(по предположению $Df>0$, $Dg>0$).

Из определений следует, что

$$K\left(f,g\right)=K\left(f_0,g_0\right).$$

Если $K\left(f,g\right)=0$, то случайные переменные f, g называются *некоррелированными*. Используя геометрическую интерпретацию ($\cos\left(\pi/2\right)=0$), их называют также *ортогональными*.

Из определений следует, что

$$K\left(f,g\right)=0\Leftrightarrow E\left(fg\right)=EfEg.$$

Коэффициент корреляции равен нормированной разности между средним значением произведения переменных и произведением их средних значений. Он описывает линейную зависимость между случайными переменными в среднем и служит хотя и грубой, но простой характеристикой этой зависимости.

1.5.5. Как для любого евклидова пространства, для $\mathcal{F}_0$ верно *неравенство Коши*:

$$|\langle f_0, g_0 \rangle| \leqslant \|f_0\| \cdot \|g_0\| \qquad\qquad (f_0, g_0 \in \mathcal{F}_0)\,.$$

Из него следует, что

$$|K\,(f, g)| \leqslant 1 \qquad\qquad (f, g \in \mathcal{F})\,.$$

О всяком свойстве случайной переменной условимся говорить, что оно верно *почти всюду*, если оно верно для всех исходов с ненулевыми вероятностями. В невырожденном пространстве, где исходов с нулевыми вероятностями нет, *почти всюду* означает *всюду*. Говорят, что переменные f, g почти всюду линейно зависимы, если существует число a и постоянная b такие, что $af - g = b$ почти всюду (то есть $af\,(u) - g\,(u) = b$ при $p\,(u) > 0$).

Исследуя дискриминант квадратного уравнения

$$E\left(f_0^2\right)x^2 - 2E\,(f_0 g_0)\,x + E\left(g_0^2\right) = 0,$$

нетрудно доказать (упражнение), что равенство

$$|K\,(f, g)| = 1$$

верно тогда и только тогда, когда переменные f, g почти всюду линейно зависимы.

Используя векторную интерпретацию, можно сказать, что равенства $K\,(f, g) = 1$ и $K\,(f, g) = -1$ определяют соответственно *одинаковую* и *противоположную направленность* f и g.

1.5.6. В 1.4.8 были вычислены корреляции для переменных s_j и s_k, описывающих появление единицы на j-м и k-м местах в двоичной марковской последовательности:

$$C\,(s_j, s_k) = p_j q_j d^{k-j} = D\,(s_j)\, d^{k-j}.$$

Следовательно, при $j < k$ верно равенство

$$K(s_j, s_k) = \sqrt{D(s_j) \;/\; D(s_k)}\, d^{k-j}.$$

Для последовательности Бернулли $d = 0$ и $K(s_j, s_k) = 0$. Переменные s_j и s_k некоррелированы.

1.5.7. Пример. Из формулы в 1.4.5 для Df следует, что

$$K(f, f) = \frac{E\left(f^2\right) - (Ef)^2}{\sqrt{Df}\sqrt{Df}} = \frac{Df}{Df} = 1$$
$$K(f, -f) = \frac{-E\left(f^2\right) + (Ef)^2}{\sqrt{Df}\sqrt{Df}} = -\frac{Df}{Df} = -1.$$

Вообще, для каждых переменных f, g и чисел a, b, c, d верны равенства

$$\begin{aligned} C(af+b, cg+d) &= E\left((af+b)(cg+d)\right) - E(af+b)\,E(cg+d) = \\ &= ac\left(E(fg) - EfEg\right), \\ S(af+b) &= |a|\,Sf, \qquad S(cg+d) = |c|\,Sg, \\ K(af+b, cg+d) &= \operatorname{sgn}(ac)\cdot K(f,g). \end{aligned}$$

В частности,

$$K(-f, g) = K(f, -g) = -K(f, g), \qquad K(-f, -g) = K(f, g).$$

1.5.8. Пример. На пространстве $B(n, a)$ рассмотрим случайные переменные s_j, $s = \sum s_j$, s/n $(1 \leqslant j \leqslant n)$.

Из 1.5.7 следует, что

$$K(s, s/n) = K(s, s) = 1.$$

Это соответствует линейной зависимости частоты появления единиц и их числа.

Так как (1.3.2, 1.3.4)

$$E\left(s_i s\right) = \sum_{j=1}^{n} E\left(s_i s_j\right) = a + (n-1)\, a^2,$$

$$Es_i = a, \qquad Es = na, \qquad Ds_j = a\left(1-a\right), \qquad Ds = na\left(1-a\right),$$

то

$$K\left(s_i, s\right) = \frac{a + (n-1)\, a^2 - na^2}{\sqrt{n} \cdot a\left(1-a\right)} = 1/n$$

для каждого $i = 1, \ldots, n$. При сделанных предположениях с увеличением числа слагаемых сумма меньше зависит от каждого из них.

1.5.9. Пример. Рассмотрим кроме суммы $s = \sum s_j$ еще произведение $t = \prod s_j$ переменных s_j на пространстве $B\left(n, a\right)$. Из определений следует, что

$$Et = P\left(s_1 = \ldots = s_n = 1\right) = a^n.$$

Так как $t^2 = t$, то

$$Dt = E\left(t^2\right) - \left(Et\right)^2 = Et - \left(Et\right)^2 = a^n\left(1-a^n\right).$$

А так как $st = nt$, то

$$E\left(st\right) = nEt = na^n.$$

Следовательно,

$$K\left(s, t\right) = \frac{na^n - na \cdot a^n}{\sqrt{na\left(1-a\right) a^n\left(1-a^n\right)}} = \sqrt{na^{n-1} \cdot \frac{1-a}{1-a^n}}.$$

Как и в примере 1.5.8, корреляция стремится к 0 при $n \to \infty$.

1.5.10. Пример. Рассмотрим пространство $B(2,a)$ и случайные переменные $f = s_1 + s_2$, $g = s_1 - s_2$ на нем.

Так как

$$Ef = 2a, \qquad Eg = 0,$$
$$E(fg) = E\left(s_1^2 - s_2^2\right) = E(s_1 - s_2) = 0,$$

то

$$C(f,g) = E(fg) - EfEg = 0$$

и, следовательно, $K(f,g) = 0$.

Вместе с тем переменные f и g явно зависимы: если $g \neq 0$, то $f = 1$. Коэффициент корреляции не отражает такую зависимость.

Рассмотрим комбинаторное пространство $(L,4)$ с множеством исходов $U = \{1,2,3,4\}$ и случайные переменные x, y со значениями $x(1) = -2$, $x(2) = -1$, $x(3) = 1$, $x(4) = 2$ и $y(u) = x^2(u)$.

Непосредственные вычисления дают:

$$Ex = 0, \qquad Ey = 10/4, \qquad E(xy) = 0.$$

Следовательно,

$$C(x,y) = E(xy) - ExEy = 0$$

и $K(x,y) = 0$.

Вместе с тем между переменными x и y существует функциональная зависимость: $y = x^2$. Коэффициент корреляции не отражает такую нелинейную зависимость.

Эти два примера показывают, что коэффициент корреляции может равняться 0 и тогда, когда случайные переменные явно зависимы. Поэтому нужны более адекватные меры зависимости между случайными переменными, описывающие не только линейную зависимость.

1.6. Распределения

Теория вероятностей как правило изучает не сами случайные переменные, а их распределения. Совместные распределения случайных переменных определяют их стохастическую зависимость.

1.6.1. Случайная переменная f на вероятностном пространстве (U, p) при замене исхода u на значение $x = f(u)$ естественным образом определяет новое вероятностное пространство (X, q) с множеством исходов

$$X = f(U) = \{f(u) : u \in U\}$$

и элементарной вероятностью $q = Pf^{-1}$, имеющей значения

$$q(x) = P\left(f^{-1}(x)\right) = \sum_{u : f(u) = x} p(u).$$

Ясно, что $q(x) \geqslant 0$. Кроме того,

$$\sum_x q(x) = \sum_x \left(\sum_{u : f(u) = x} p(u) \right) = \sum_u p(u) = 1.$$

Если переменная f числовая, что обычно предполагается, то множество X состоит из чисел. А множество U может состоять из любых элементов.

Элементарная вероятность q называется *элементарным распределением* случайной переменной f. Вероятность $Q = Pf^{-1}$, продолжающая q на класс событий, называется *распределением* f. Случайные переменные, имеющие одинаковые распределения, называются *одинаково распределенными*.

Элементарное распределение q коротко тоже будем называть распределением, как Q, сохраняя разницу в обозначениях. Распределение Q будет использоваться редко. Если пространство (U, p) невырожденное, то и (X, q) невырожденное.

Замена $u \in U$ на $x = f(u) \in X$ дает удобную формулу для среднего значения:

$$Ef = \sum xq(x).$$

Она получается группировкой исходов u, при которых $f(u) = x$:

$$\sum_u f(u)\, p(u) = \sum x \sum_{u: f(u)=x} p(u) = \sum_x q(x).$$

Точно так же получается формула

$$Df = \sum (x - Ef)^2\, q(x)\,.$$

1.6.2. Пример. Пусть (U_1, p_1) — комбинаторное пространство с множеством исходов $u_1 = \{1, 2, 3, 4, 5, 6\}$. Положим также $f_1(2) = f_1(4) = f_1(6) = 1$, $f_1(1) = f_1(3) = f_1(5) = -1$. Тогда $X_1 = \{-1, 1\}$ и $q_1(-1) = q_1(1) = 1/2$. Эта модель описывает выигрыш при выпадении четного числа очков и проигрыш при выпадении нечетного, когда игральную кость подбрасывают один раз.

Пусть теперь $U_2 = \{0, 1\}$, $p_2(0) = p_2(1) = 1/2$ и $f_2(0) = -1$, $f_2(1) = 1$. Тогда тоже $X_2 = \{-1, 1\}$ и $q_2(-1) = q_2(1) = 1/2$. Такая модель описывает выигрыш при выпадении герба и проигрыш при выпадении решки, когда симметричная монета подбрасывается один раз. Случайные переменные f_1 и f_2 имеют одинаковые распределения, хотя как функции различны и имеют разные области определения: $Pf_1^{-1} = Pf_2^{-1}$, но $f_1 \neq f_2$.

1.6.3. Рассмотрим пространство Бернулли $B(n, a)$ и случайную переменную $f = s$ (1.1.3). Ее множеством значений будет $X = \{0, 1, \ldots, n\}$. Она имеет *биномиальное распределение*:

$$q(x) = \sum_{u: s(u)=x} a^x (1-a)^{n-x} = \binom{n}{x} a^x (1-a)^{n-x}\,,$$

$$\binom{n}{x} = \frac{n!}{x!\,(n-x)!} \qquad (x = 0, 1, \ldots, n)\,.$$

Удобно считать, что этот биномиальный коэффициент равен 0 при $x < 0$ и при $x > n$.

Биномиальное распределение играет большую роль в элементарной теории вероятностей. Оно используется при решении многих задач.

1.6.4. Переходя от чисел к векторам, можно определить совместное распределение конечного семейства $f = (f_j)$ числовых случайных переменных f_j ($j \in K$) на одном и том же вероятностном пространстве (U, p). Семейство при замене исхода u на вектор $x = f(u)$ с координатами $x_j = f_j(u)$ определяет новое вероятностное пространство (X, q) с множеством исходов

$$X = f(U) = \{(f_j(u)) : u \in U\}$$

и элементарной вероятностью $q = Pf^{-1}$ со значениями

$$q(x) = P\left(f^{-1}(x)\right) = P\{u : f_j(u) = x_j, j \in K\}.$$

Эта элементарная вероятность называется *элементарным совместным распределением* случайных переменных f_j. Вероятность $Q = Pf^{-1}$, определяемая q, называется *совместным распределением* случайных переменных f_j. Этот термин часто будет употребляться и для q.

Вместе с тем, каждая из случайных переменных f_j имеет собственные элементарное распределение $q_j = Pf_j^{-1}$ и определяемое им распределение $Q_j = Pf_j^{-1}$. Эти распределения называются *маргинальными*. Каждая маргинальная вероятность q_j для семейства (f_j) определена на множестве $X_j = f_j(U)$ чисел, а совместная q — на множестве $X = f(U)$ векторов $|K|$-мерного пространства. Как правило множество индексов K составляется из номеров. Если $K = \{1, \ldots, m\}$, то $X \subseteq \mathbb{R}^m$.

1.6.5. Пример. Рассмотрим пространство Бернулли $B\,(2,a)$ и случайные переменные $f_1 = s_1$, $f_2 = s_2$ на нем. Из определений следует, что $U_1 = U_2 = \mathbb{B} = \{0,1\}$, $U = \mathbb{B}^2 = \{(0,0),(0,1),(1,0),(1,1)\}$ и $X_1 = X_2 = \mathbb{B}$, $X = \mathbb{B}^2$. Маргинальные и совместные распределения для $f = (s_1, s_2)$ описываются таблицами:

x_1	0	1
q_1	$1-a$	a

,

x_2	0	1
q_2	$1-a$	a

,

x	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
q	$(1-a)^2$	$a\,(1-a)$	$a\,(1-a)$	a^2

.

Для всех $x = (x_1, x_2)$ верно равенство $q\,(x) = q_1\,(x_1)\,q_2\,(x_2)$. Имея его в виду, вероятность q называют *произведением* вероятностей q_1, q_2 и пишут $q = q_1 \times q_2$.

Если интерпретировать эту модель как описание подбрасывания два раза симметричной монеты, то числа s_1 и s_2 появления гербов при первом и втором подбрасывании естественно считать независимыми. Это подводит к мысли связать независимость случайных переменных с равенством их совместного распределения произведению маргинальных.

1.6.6. Пример. Рассмотрим пространство $B\,(2, 1/2)$ и случайные переменные $f_1 = s_1 + s_2$, $f_2 = s_1 - s_2$. Очевидно, что $X_1 = \{0,1,2\}$, $X_2 = \{-1,0,1\}$, $X = \{(0,0),(1,-1),(1,1),(2,0)\}$. Маргинальное и совместное распределения для f_1, f_2 описываются таблицами:

x_1	0	1	2
q_1	$1/4$	$1/2$	$1/4$

,

x_2	-1	0	1
q_2	$1/4$	$1/2$	$1/4$

,

x	$(0,0)$	$(1,-1)$	$(1,1)$	$(2,0)$
q	$1/4$	$1/4$	$1/4$	$1/4$

.

Теперь $q \neq q_1 \times q_2$ $(q\,(0,0) = 1/4 \neq (1/4)\,(1/2) = q_1\,(0)\,q_2\,(0))$. Как отмечалось в 1.5.10, переменные f_1 и f_2 зависимы. Пример подкрепляет связь между независимостью случайных переменных и равенством их совместного распределения произведению маргинальных.

1.6.7. Рассмотрим конечное семейство $f = (f_j)$ переменных f_j $(j \in M)$ на пространстве (U, p), их маргинальные распределения q_j на множествах $X_j = f_j(U)$ и совместное распределение q. Если равенство

$$q(x) = \prod q_j(x_j) \qquad (*)$$

верно для всех $x = (x_j) \in X$, то q называют *произведением* q_j и пишут $q = \prod q_j$. Если $q = \prod q_j$, то говорят, что случайные переменные f_j (*стохастически*) *независимы*, а если $q \neq \prod q_j$, — что (*стохастически*) *зависимы*.

Примеры в 1.6.5 и 1.6.6 поясняют данное определение. Оно придает точный смысл описываемой там зависимости. Рассмотрим еще пространство Бернулли $B(n, a)$ и семейство переменных $s_j = u_j$ $(j = 1, \ldots, n)$ на нем (1.1.3). Из определений следует, что эти переменные независимы: $X_j = \mathbb{B}$, $X = \mathbb{B}^n$ и

$$q(x) = a^{\Sigma x_j} (1-a)^{n-\Sigma x_j} = \prod q_j(x_j).$$

Равенство (*) верно и для каждого множества индексов $K \subseteq M$ (конечного вместе с M). Для того, чтобы убедиться в этом, достаточно просуммировать обе части (*) по всем $x_j \in X_j$ для индексов $j \in M \setminus K$ и использовать равенства $\sum q_j(x_j) = 1$ для них.

Рассмотрим множества $B_j \subseteq X_j$, $j \in K$. Суммируя полученное из (*) равенство по всем $x_j \in B_j$, $j \in K$, убеждаемся в том, что

$$P(f_j \in B_j, j \in K) = \prod_{j \in K} P(f_j \in B_j). \qquad (**)$$

Здесь $f_j \in B_j$ — краткая запись события $\{u : f_j(u) \in B_j\}$. В левой части — пересечение этих событий при $j \in K$. При $B_j = \{x_j\}$ и

$K = M$ равенство (**) превращается в (*). Переход к $K \subset M$ эквивалентен условию $B_j = X_j$, $j \in M \setminus K$. Независимость случайных переменных f_j, $j \in K$ означает, что (**) верно при любых $B_j \subseteq X_j$. Обычно она так и определяется.

1.6.8. Рассмотрим конечное семейство случайных переменных f_j. Условимся говорить, что случайные переменные f_j *независимы в среднем*, если среднее значение произведения $\prod f_j$ равно произведению средних значений f_j:

$$E\left(\prod f_j\right) = \prod (Ef_j).$$

Для двух переменных независимость в среднем означает некоррелированность.

Для большого семейства переменных такая независимость может служить только очень грубой характеристикой. Хотя для некоторых специальных семейств она совпадает со стохастической независимостью (например, для семейств индикаторов).

Если случайные переменные независимы, то они независимы и в среднем. В самом деле, если $q = \prod q_j$, то

$$E\left(\prod f_j\right) = \sum_x \left(\prod_j x_j\right) q(x) = \sum_x \prod_j x_j q_j(x_j) = \\ = \prod_j \sum_{x_j} x_j q_j(x_j) = \prod_j (Ef_j).$$

В частности, если две переменные независимы, то они некоррелированы. Обратное не верно (1.5.10, 1.6.6). Но если переменные коррелируют, то они зависимы: независимость f, g влечет равенства $E(fg) = EfEg$, $C(f,g) = 0$, $K(f,g) = 0$.

1.6.9. Случайные переменные f_j называются *попарно независимыми*, если для каждых двух различных индексов j и k переменные f_j и f_k независимы. Ясно, что из независимости переменных f_j в совокупности следует их попарная независимость. Обратное не верно, как показывает классический *пример Бернштейна*.

Возьмем правильный тетраэдр с белой гранью (a), синей гранью (b), красной гранью (c) и трехцветной четвертой гранью (d), окрашенной в эти цвета. Благодаря своей симметричности тетраэдр при подбрасывании с одинаковой вероятностью падает на одну из этих граней. Поэтому для описания такого подбрасывания можно использовать комбинаторную модель с множеством исходов $\{a, b, c, d\}$. События $A = \{a, d\}$, $B = \{b, d\}$, $C = \{c, d\}$ описывают появление соответственно белого, синего, красного цвета в основании упавшего на плоскую поверхность тетраэдра. Событие $D = \{d\}$ — появление всех трех цветов. Заметим, что пересечения AB, AC, BC и ABC равны D — два цвета появляются в том и только в том случае, когда появляются все три.

Обозначим f_1, f_2, f_3 индикаторы соответственно событий A, B, C. Тогда каждая из случайных переменных f_j имеет распределение q_j со значениями $q_j\,(0) = q_j\,(1) = 1/2$. Совместные распределения q_{jk} переменных f_j, f_k $(j \neq k)$ имеют следующие значения: $q_{jk}\,(0,0) = q_{jk}\,(0,1) = q_{jk}\,(1,0) = q_{jk}\,(1,1) = 1/4$. А совместное распределение q переменных f_1, f_2, f_3 описывается таблицей

x	100	010	001	111	000	011	101	110
q	1/4	1/4	1/4	1/4	0	0	0	0

Таким образом, $q_{jk} = q_j \times q_k$, но $q \neq q_1 \times q_2 \times q_3$. Случайные переменные f_1, f_2, f_3 попарно независимы, а в совокупности зависимы.

1.6.10. События $A_j \subseteq U$ в пространстве (U, p) называются независимыми, если независимы их индикаторы $f_j = \operatorname{ind} A_j$. Проверять независимость событий удобнее непосредственно используя их вероятности.

Рассмотрим события $A, B \subseteq U$ и их индикаторы $f = \operatorname{ind} A$, $g = \operatorname{ind} B$. Такие случайные переменные f, g независимы тогда и только тогда, когда для пересечения $AB = A \cap B$ верно равенство

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

В самом деле, если f и g независимы, то

$$P(AB) = E(fg) = EfEg = P(A)\, P(B).$$

Проверим обратное утверждение. Элементарные распределения q, r для f, g и их совместное элементарное распределение s определяются таблицами

$$\begin{array}{c|c|c} & 1 & 0 \\ \hline q & P(A) & 1 - P(A) \end{array}, \qquad \begin{array}{c|c|c} & 1 & 0 \\ \hline r & P(B) & 1 - P(B) \end{array},$$

$$\begin{array}{c|c|c|c|c} & (1,1) & (1,0) & (0,1) & (0,0) \\ \hline s & P(AB) & P(AB') & P(A'B) & P(A'B') \end{array}.$$

Если $P(AB) = P(A)\, P(B)$, то легко проверить (упражнение), что $P(AB') = P(A)P(B')$, $P(A'B) = P(A')P(B)$, $P(A'B') = P(A')P(B')$. Значит, $s = q \times r$ и f, g независимы.

Независимость событий $A_j \subseteq U$ $(j \in M)$ в пространстве (U, p) выражается равенством

$$P\left(\bigcap_{j \in K} A_j\right) = \prod_{j \in K} P(A_j)$$

для каждого конечного $K \subseteq M$. Эти равенства обычно принимаются в качестве определения независимости событий A_j. Они сразу следуют из равенства (**).

1.6.11. Определение независимости событий через произведения их вероятностей можно в свою очередь использовать для определения независимости случайных переменных, которое эквивалентно данному в 1.6.7.

Рассмотрим, как в 1.6.7, семейство $f = (f_j)$ переменных f_j $(j \in M)$ на пространстве (U, p) с элементарными маргинальными и совместным распределениями q_j и q на множествах $X_j = f_j(U)$ и $X = f(U)$. По определению

$$q_j(x_j) = Pf_j^{-1}(x_j), \quad q(x) = Pf^{-1}(x) \quad (x = (x_j) \in X, x_j \in X_j).$$

Так как для конечного $K \subseteq M$

$$f^{-1}(x) = \{u : f_j(u) = x_j, j \in K\} = \\ = \bigcap \{u : f_j(u) = x_j\} = \bigcap f_j^{-1}(x_j),$$

то равенство $q(x) = \prod q_j(x_j)$ $(j \in K)$ означает, что

$$P\left(\bigcap f_j^{-1}(x_j)\right) = \prod P\left(f_j^{-1}(x_j)\right).$$

Выполнение этих равенств для всех значений x_j всех конечных семейств рассматриваемых переменных f_j эквивалентно их стохастической независимости. Используя определение независимости случайных переменных через произведения вероятностей их значений, можно упростить рассуждения в примере с тетраэдром в 1.6.9.

1.6.12. Реальная зависимость между случайными переменными f и g может быть несимметричной. Это отражают *коэффициенты регрессии* f на g и g на f:

$$R(f, g) = \frac{C(f, g)}{Dg} = \frac{E(fg) - EfEg}{Dg},$$
$$R(g, f) = \frac{C(f, g)}{Df} = \frac{E(fg) - EfEg}{Df}.$$

(Предполагаем, как обычно в корреляционной теории, что $Df > 0$ и $Dg > 0$.) Геометрически $R(f,g)$ и $R(g,f)$ интерпретируются как нормированные величины проекций вектора f на направление вектора g и наоборот:

$$R(f,g) = R(f_0, g_0) = \frac{\langle f_0, \overline{g_0} \rangle}{\|g_0\|},$$

$$R(f,g) = R(g_0, f_0) = \frac{\langle g_0, \overline{f_0} \rangle}{\|f_0\|}.$$

Здесь $\overline{f_0} = f_0 / \|f_0\|$, $\overline{g_0} = g_0 / \|g_0\|$ при $\|f_0\| > 0$, $\|g_0\| > 0$.

1.6.13. Так как

$$R(f,g) \cdot R(g,f) = (K(f,g))^2,$$

то $R(f,g)$ и $R(g,f)$ имеют один и тот же знак, а коэффициент корреляции равен их среднему геометрическому с общим знаком:

$$K(f,g) = \pm\sqrt{R(f,g)\, R(g,f)}.$$

В примерах 1.5.7 и 1.5.8

$$R(s_j, s) = 1, \qquad R(s, s_j) = 1/n;$$
$$R(s,t) = a^{n-1}, \qquad R(t,s) = (1-a) \big/ (1-a^n).$$

Если сумма $s = 0$, то каждое слагаемое $s_j = 0$ и произведение $t = 0$. Но если $s_j = 0$ для данного номера j или $t = 0$, то s может иметь любое значение из $\{0, 1, \ldots, n-1\}$.

1.6.14. Если отождествить события A, B с их индикаторами $\operatorname{ind} A$, $\operatorname{ind} B$ и воспользоваться равенствами $P(A) = E(\operatorname{ind} A)$, $P(B) = E(\operatorname{ind} B)$, то можно определить коэффициент корреляции для событий равенством

$$K(A,B) = \frac{P(AB) - P(A)\,P(B)}{\sqrt{P(A)\,(1-P(A))}\sqrt{P(B)\,(1-P(B))}}.$$

Вместо коэффициента корреляции для измерения зависимости между событиями можно использовать *коэффициент связи*

$$Q(A,B) = \frac{P(AB) - P(A)P(B)}{P(AB)P(A'B') + P(AB')P(A'B)}.$$

(В равенствах для коэффициентов корреляции и связи предполагается, что знаменатели не равны 0).

Равенство $Q(A,B) = 0$ эквивалентно $P(AB) = P(A)P(B)$ и означает независимость событий A, B.

Используя правила действий с вероятностями, легко проверить (упражнение), что

$$|K(A,B)| \leqslant |Q(A,B)|.$$

Равенство $Q(A,B) = 1$ означает, что либо $P(A) = P(AB)$, либо $P(B) = P(AB)$. Если рассматриваемое пространство невырожденное, то последние равенства эквивалентны включениям $A \subseteq B$ или $B \subseteq A$: одно из этих событий влечет другое.

1.6.15. Пример. Рассмотрим последовательность Бернулли $B(n,a)$ и события $A = \{u = (u_i) : u_j = 1\} = \{s_j = 1\}$ — появление единицы на j-м месте, $B = \{u = (u_i) : \prod u_i = 1\} = \{t - + = 1\}$ — появление единиц на всех местах. Из определений следует, что

$$\begin{aligned} &P(A) = a, && P(B) = a^n, && P(AB) = P(B) = a^n, \\ &P\left(A'B'\right) = 1 - a, && P\left(A'B\right) = 0, && P\left(AB'\right) = a\left(1 - a^{n-1}\right). \end{aligned}$$

Поэтому, в соответствии с включением $B \subseteq A$,

$$Q(A,B) = \frac{a^n - aa^n}{a^n(1-a)} = 1.$$

1.7. Информация и энтропия

Информация случайных переменных определяется как среднее значение некоторой специальной функции этих переменных. Она является удобной мерой зависимости между ними. Термин *информация* употребляется в теории вероятностей в специальном точно определенном смысле. Энтропия определяется через информацию. Элементарной теории информации посвящена книга [12].

1.7.1. Рассмотрим невырожденное вероятностное пространство (U, p) и случайные переменные f, g на нем. Пусть $X = f(U)$, $Y = g(U)$ и

$$Z = \{(x, y) = (f(u), g(u)) : u \in U\}.$$

(Подчеркнем, что часто $X \times Y \neq Z$, так как пары $(x, y) = (f(u), g(v))$ для некоторых исходов $u \neq v$ могут не принадлежать Z.) Обозначим q, r и s маргинальные и совместное распределения для f, g. По предположению p, а поэтому q, r и s строго положительны. Это позволяет определить случайную переменную $i(f, g)$, которая имеет значения

$$i(f, g, u) = \log \frac{s(f(u), g(u))}{q(f(u))\, r(g(u))} \qquad (u \in U).$$

Условимся называть ее *элементарной информацией* f, g. Основание логарифма обычно выбирается равным 2. Из определений следует, что тождественное равенство $i(f, g) = 0$ эквивалентно независимости f, g.

Число

$$I(f, g) = E(i(f, g))$$

называется *информацией* f, g. Из определений следует, что

$$I(f,g) = \sum p(u) \log \frac{s(f(u), g(u))}{q(f(u))\, r(g(u))} \qquad (u \in U),$$

$$I(f,g) = \sum s(x,y) \log \frac{s(x,y)}{q(x)\, r(y)} \qquad ((x,y) \in Z).$$

Ясно, что если f, g независимы, то $I(f,g) = E(0) = 0$.

Часто информацию переменных f и g называют *взаимной*.

1.7.2. Пусть $f = s_1$, $g = s_2$ из примера 1.6.5. Тогда имеем: $X = Y = \{0,1\}$ и $Z = X \times Y$, $s(x,y) = q(x)\, r(y)$ и $I(f,g) = 0$ в соответствии со стохастической независимостью переменных s_1,s_2.

Если $f = s_1 + s_2$, $g = s_1 - s_2$ (пример 1.6.6), то $X = \{0,1,2\}$, $Y = \{-1,0,1\}$ и

$$Z = \{(0,0), (1,-1), (1,1), (2,0)\} \neq X \times Y.$$

При $a = 1/2$ распределения q, r и s имеют следующие значения: $q = \{1/4, 1/2, 1/4\}$, $r = \{1/4, 1/2, 1/4\}$ и $s(x,y) = 1/4$ для всех $(x,y) \in Z$. Следовательно,

$$I(f,g) = 4\left(\frac{1}{4} \log \frac{1/4}{(1/2)(1/4)}\right) = \log 2 > 0$$

в соответствии со стохастической зависимостью случайных переменных $s_1 + s_2$ и $s_1 - s_2$.

1.7.3. Совсем не очевидно, что равенство $I(f,g) = 0$ эквивалентно стохастической независимости переменных f,g. Этот факт можно вывести из классической теоремы о мультипликативно и аддитивно взвешенных средних (геометрического и арифметического).

Рассмотрим натуральное число $n > 0$, семейство чисел $x_j \geqslant 0$ и семейство чисел $q_j > 0$ с суммой $\sum q_j = 1$ $(j = 1, \ldots, n)$.

Теорема о средних. *$\prod x_j^{q_j} \geqslant \sum q_j x_j$, причем равенство верно тогда и только тогда, когда $x_1 = \ldots = x_n$.*

Подробное элементарное доказательство этой теоремы есть в [8]. Там же выведены два следствия, которые нетрудно последовательно проверить (упражнение).

Пусть $s = (s_j)$, $t = (t_j)$; $s_j > 0$, $t_j > 0$, $\sum s_j = \sum t_j = 1$; $a = (a_j)$, $a_j = 1/n$ $(j = 1, \ldots, n)$;

$$\varphi(s,t) = \sum s_j \log t_j.$$

Следствие 1. *$\varphi(s,t) < \varphi(t,t)$ при $s \neq t$.*

Следствие 2. *$\varphi(t,t) > \varphi(a,a)$ при $t \neq a$.*

Таким образом, при каждом t функция $\varphi(\cdot,t) : s \to \varphi(s,t)$ имеет строгий максимум в точке $s = t$, а функция ψ со значениями

$$\psi(t) = -\varphi(t,t) = -\sum t_j \log t_j$$

имеет строгий максимум в точке $t = a$ (при $t_j = 1/n$ для $j = 1, \ldots, n$).

1.7.4. Используя следствие 1, легко доказать следующую теорему об информации.

Теорема об информации. *Информация случайных переменных f и g положительна. Она равна нулю, если и только если f и g независимы.*

□ Рассмотрим семейство s чисел $s(x,y)$ и семейство t чисел $t(x,y) = q(x)r(y)$, полученные произвольной нумерацией пар (x,y) значений $x = f(u)$, $y = g(u)$ случайных переменных f, g. Из определений вытекает, что

$$I(f,g) = \sum s(x,y) \log s(x,y) - \sum s(x,y) \log t(x,y).$$

По следствию 1 строгое неравенство $I(f,g) > 0$ верно при $s \neq t$ и $I(f,g) = 0$ при $s = t$. Это эквивалентно сформулированному утверждению. ■

Из теоремы об информации следует

Информационный критерий зависимости. *Случайные переменные f и g зависимы, если и только если их взаимная информация строго положительна.*

Равенство $I(f,g) = 0$ гарантирует стохастическую независимость в отличие от равенства $K(f,g) = 0$, обеспечивающего только независимость в среднем.

Например, в пространстве $B(2,1/2)$ для переменных $f = s_1 + s_2$, $g = s_1 - s_2$ (1.5.9) верно равенство $K(f,g) = 0$, но $I(f,g) = \log 2 > 0$. Переменные f, g некоррелированы, но стохастически зависимы. Точно так же легко проверить (упражнение), что для некоррелированных переменных x, y в 1.5.10 верно $I\,(f,g) = \log 2 > 0$, что доказывает их стохастическую зависимость.

1.7.5. Число

$$H\,(f) = I\,(f,f)$$

называется *энтропией* случайной переменной f на невырожденном вероятностном пространстве (U,p). Так как $s(x,x) = r(x) = q(x)$ для каждого $x = f\,(u)$, то

$$H\,(f) = \sum q\,(x) \log\,(1/q\,(x)) = -\sum q\,(x) \log q\,(x).$$

Эти равенства позволяют определить энтропию H независимо от информации J. Энтропия вместе с информацией положительна: $H\,(f) \geqslant 0$.

На пространстве (U,p) введем случайную переменную $h(f)$ со значениями

$$h\,(f) = i\,(f,f) = \log\,(1/q\,(f\,(u))) = -\log q\,(f\,(u)).$$

Из определений следует, что

$$H\,(f) = E\,(h\,(f)) = -\sum p\,(u) \log q\,(f\,(u)).$$

В примере 1.6.5 с пространством Бернулли $B(2, a)$ энтропия переменной $f = s_1$ выражается равенством

$$H(s_1) = -\left(a \log a + (1-a) \log(1-a)\right).$$

Такая же энтропия и у s_2. В примере 1.6.6

$$H(s_1 + s_2) = H(s_1 - s_2) = (3/2) \log 2.$$

Если $f = c$ — постоянная, то $q(c) = 1$ и

$$H(c) = 0.$$

А если случайная переменная f имеет *равномерное распределение* ($q(x) = 1/n$ для всех n значений x), то

$$H(f) = -n(1/n) \log 1/n = \log n.$$

1.7.6. Информация $I(f, g)$ случайных переменных f, g меньше каждой из энтропий $H(f)$, $H(g)$.

Действительно, так как выполнено $q(x) = \sum\limits_y s(x, y) \geqslant s(x, y)$ и $r(y) = \sum\limits_x s(x, y) \geqslant s(x, y)$, то

$$\frac{s(x, y)}{q(x)\, r(y)} \leqslant \frac{1}{q(x)}$$

для каждых пар $(x, y) = (f(u), g(u))$ значений x, y переменных f, g. Поэтому

$$I(f, g) = \sum_{(x,y)} s(x, y) \log \frac{s(x, y)}{q(x)\, r(y)} \leqslant \sum q(x) \log(1/q(x)) = H(f).$$

Точно так же $I(f, g) \leqslant H(g)$.

Пусть n — число значений случайной переменной f. Тогда из следствия 2 теоремы о средних вытекает, что

$$H(f) \leqslant -n(1/n) \log(1/n) = \log n.$$

Таким образом
$$0 \leqslant H(f) \leqslant \log n.$$
При этом $H(f) = 0$ эквивалентно $f = const$, а $H(f) = \log n$ тогда и только тогда, когда f имеет равномерное распределение.

Энтропия $H(f)$ служит средней *мерой неопределенности* случайной переменной f. Использование логарифмов позволяет произведения вероятностей превращать в суммы. Это упрощает выкладки.

1.7.7. Исследуем с информационной точки зрения процедуру угадывания. Предположим, что с помощью случайного выбора определена одна из $m = 32$ букв алфавита U, которую нужно угадать последовательным делением алфавита на непустые части. Будем считать, процедура описывается комбинаторным пространством.

Если разбить алфавит U на части A и B из k и $l = m - k$ букв, то информация о том, что выбранная буква $u \in A$ оценивается так. Рассмотрим какую-нибудь нумерацию f алфавита U и индикатор g множества A. Имеем:
$$\begin{array}{lll} X = \{1, \ldots, m\}, & Y = \{0, 1\}, & Z = X \times Y; \\ q(x) = 1/m, & r(1) = k/m, & r(0) = 1 - k/m; \end{array}$$
$$\begin{array}{llll} s(x,1) = 1/m & (x \in f(A)), & s(x,1) = 0 & (x \notin f(A)), \\ s(x,0) = 0 & (x \in f(A)), & s(x,0) = 1/m & (x \notin f(A)). \end{array}$$
Поэтому
$$I(f,g) = \frac{k}{m} \log \frac{1/m}{(1/m)(k/m)} + \frac{m-k}{m} \log \frac{1/m}{(1/m)(1-k/m)} =$$
$$= -\left(\frac{k}{m} \log \frac{k}{m} + \left(1 - \frac{k}{m}\right) \log \left(1 - \frac{k}{m}\right)\right) = H(g).$$

Максимум информации достигается, когда g имеет равномерное распределение: $k/m = 1/2$ и $k/m = m/2$. Оптимально делить алфавит на две равные части. Тогда в среднем можно угадывать выбранную букву за пять последовательных разбиений. Другие способы в среднем будут хуже, хотя случайно угадывание по одной букве может сразу привести к успеху. Но зато в противном случае неопределенность мало уменьшается.

1.7.8. Теория информации была разработана вначале для систем связи. Простейшую систему связи можно схематически описать с помощью следующей вероятностной модели.

Рассмотрим множество исходов $U = \{00, 01, 10, 11\}$, где величина $u = xy$ описывает передачу сигнала x и прием сигнала y. Событие $A_x = \{x0, x1\}$ означает передачу x, а $B_y = \{0y, 1y\}$ — прием y $(x, y \in \{0, 1\})$. Пусть $a_x > 0$, $a_0 + a_1 = 1$ и $t_{xy} > 0$, $t_{x0} + t_{x1} = 1$. Элементарная вероятность p имеет значения $p\,(xy) = a_x t_{xy}$. В частности, $p\,(xx) = a_x t_{xx}$.

Далее, заметим, что $A_x B_y = A_x \cap B_y = \{xy\}$ и $P\,(A_x) = a_x$, $P(A_x B_y) = p(xy) = a_x t_{xy}$. *Условная вероятность* принять y, когда передан x, по определению равна отношению $P(A_x B_y)/P(A_x)$:

$$P\,(B_y/A_x) = a_x t_{xy}/a_x = t_{xy}.$$

Это поясняет вероятностный смысл чисел t_{xy}. Наконец,

$$P\,(B_y) = p\,(0y) + p\,(1y) = a_0 t_{0y} + a_1 t_{1y}.$$

Случайные переменные f, g со значениями $f(xy) = x$, $g(xy) = y$ описывают переданный и принятый сигналы. Если в канале связи имеются помехи, изменяющие сигнал, то возможно, что $x \neq y$. Информация $I\,(f, g)$ оценивает качество системы связи. Имеем:

$$\begin{aligned}
&X = Y = \{0, 1\}, \qquad Z = X \times Y, \\
&q\,(x) = a_x, \qquad r\,(y) = a_0 t_{0y} + a_1 t_{1y}, \qquad s\,(x, y) = a_x t_{xy}.
\end{aligned}$$

Поэтому

$$I(f,g) = \sum_{x,y} a_x t_{xy} \log \frac{t_{xy}}{a_0 t_{0y} + a_1 t_{1y}}.$$

Если $t_{xy} = 1/2$ для всех x, y, то $I(f,g) = 0$: переданный и принятый сигналы независимы. Использовать такую систему связи нет смысла. Для любой системы рассматриваемого типа $I(f,g) \leqslant H(f) \leqslant \log 2$. Можно передать не больше одного *бита* информации за один раз.

1.7.9. По аналогии с коэффициентом корреляции $K(f,g)$ введем *коэффициент информации*

$$IK(f,g) = \frac{I(f,g)}{\sqrt{I(f,f)}\sqrt{I(g,g)}} = \frac{I(f,g)}{\sqrt{H(f)}\sqrt{H(g)}}$$

случайных переменных f и g, предполагая $I(f,f) = H(f) > 0$, $I(g,g) = H(g) > 0$. Так как

$$0 \leqslant I(f,g) \leqslant I(f,f) \wedge I(g,g) = H(f) \wedge H(g),$$

то

$$0 \leqslant IK(f,g) \leqslant 1.$$

Из теоремы об информации следует, что $IK(f,g) = 0$, если и только если случайные переменные f и g независимы. А из соотношений между информацией и энтропиями — что $IK(f,g) = 1$ тогда и только тогда, когда $I(f,g) = H(f) = H(g)$:

$$IK(f,g) = 1 \Leftrightarrow I(f,g) = I(f,f) = I(g,g) \Leftrightarrow \\ \Leftrightarrow I(f,g) = H(f) = H(g).$$

В самом деле, если $I(f,g) < H(f)$ или $I(f,g) < H(g)$, то вследствие $I(f,g) \leqslant H(f)$ и $I(f,g) \leqslant H(g)$ верны неравенства

$$(I(f,g))^2 < H(f)\,H(g), \qquad I(f,g) < \sqrt{H(f)}\sqrt{H(g)},$$
$$IK(f,g) < 1.$$

Коэффициент информации служит удобной мерой стохастической зависимости случайных переменных, не связанных линейно.

Условимся продолжать формально совместную элементарную вероятность s с множества Z пар возможных значений переменных f, g на весь прямоугольник $X \times Y$, приписав нулевые вероятности невозможным парам (x, y):

$$s\,(x, y) = 0 \qquad\qquad ((x, y) \in X \times Y \setminus Z)\,.$$

Кроме того, положим $0 \cdot \log 0 = 0$. Все это позволит в выкладках вместо суммирования по сложному множеству Z проводить суммирование по прямоугольнику $X \times Y$ и от двойных сумм переходить к простым суммам по множествам X и Y.

Удобно представлять маргинальные распределения q, r и совместное s для переменных f, g общей таблицей. В первом столбце записываются элементы множества X, а в первой строке — элементы множества Y. Во втором столбце помещаются вероятности $q\,(x)$, а во второй строке — вероятности $r\,(y)$. Остальную часть таблицы занимает s: в клетке (x, y) записывается вероятность $s\,(x, y)$. Если элементы каждого из множеств X и Y занумерованы, то таблицу можно упростить, не выписывая элементы.

Пример 1. Полная и упрощенная таблицы для распределений q, r, s переменных $f = s_1 + s_2$, $g = s_1 - s_2$, рассматривавшихся в 1.7.2 и 1.7.5, имеют вид:

$X \setminus Y$			0	1	-1
	$q \setminus r$		1/2	1/4	1/4
		s			
0	1/4		1/4	0	0
1	1/2		0	1/4	1/4
2	1/4		1/4	0	0

;

$q \setminus r$		1/2	1/4	1/4
	s			
1/4		1/4	0	0
1/2		0	1/4	1/4
1/4		1/4	0	0

.

Коэффициент информации выражается равенствами

$$IK\,(f,g)=\frac{\log 2}{\sqrt{(3/2)\log 2}\sqrt{(3/2)\log 2}}=\frac{2}{3}.$$

Пример 2. Полная и упрощенная таблицы для распределений q, r, s переменных x и y, рассматривавшихся в 1.5.10, имеют вид:

$X \setminus Y$			1	4
	$q \setminus r$		1/2	1/2
		s		
-2	1/4		0	1/4
-1	1/4		1/4	0
1	1/4		1/4	0
2	1/4		0	1/4

;

$q \setminus r$		1/2	1/4
	s		
1/4		0	1/4
1/4		1/4	0
1/4		1/4	0
1/4		0	1/4

.

Вычисления дают:

$$H\,(x)=\log 4=2\log 2,\qquad H\,(y)=\log 2,$$

$$\inf\,(x,y)=4\frac{1}{4}\log\left(\frac{1}{4}\Big/\left(\frac{1}{4}\frac{1}{2}\right)\right)=\log 2,$$

$$IK\,(x,y)=\frac{\log 2}{\sqrt{2\log 2}\sqrt{\log 2}}=\frac{1}{\sqrt{2}}.$$

Пример 3. Заменим $y = x^2$ на переменную $z = x^3$. Таблицы распределений тогда имеют вид:

$X \setminus Y$			-8	-1	1	8
	$q \setminus r$		1/4	1/4	1/4	1/4
		s				
-2	1/4		1/4	0	0	0
-1	1/4		0	1/4	0	0
1	1/4		0	0	1/4	0
2	1/4		0	0	0	1/4

;

$q \setminus r$		1/4	1/4	1/4	1/4
	s				
1/4		1/4	0	0	0
1/4		0	1/4	0	0
1/4		0	0	1/4	0
1/4		0	0	0	1/4

.

Соответственно:

$$H(x) = H(z) = 2\log 2,$$

$$\inf(x, y) = 4\frac{1}{4}\log\left(\frac{1}{4}\Big/\left(\frac{1}{4}\frac{1}{4}\right)\right) = 2\log 2,$$

$$IK(x, y) = 1.$$

В примерах 2 и 3 переменные $y = x^2$ и $z = x^3$ функционально зависят от переменной x. Но для z эта зависимость взаимно однозначна, а для y — нет. Разница хорошо видна на таблицах: в примере 3 в части для s каждая строка и каждый столбец имеют ровно по одному ненулевому элементу, а в примере 2 в каждом столбце есть два ненулевых элемента. Кроме того, в примере 3 таблицы квадратные, а в примере 2 — нет.

1.7.10. Связь между взаимной однозначностью и максимальным значением коэффициента информации имеет общий характер. Рассмотрим вероятностное пространство (U, p) и случайные переменные f, g на нем с множествами значений X, Y, распределениями q, r и совместным распределением s. Как обычно предполагается, что маргинальные распределения q, r невырожденные: $q\,(x) > 0$, $r\,(y) > 0$ при всех $x \in X$, $y \in Y$. Совместное распределение s может быть вырожденным: для некоторых пар значений (x, y) возможно равенство $s\,(x, y) = 0$.

Будем называть взаимно однозначное отображение $\chi : A \to B$ множества A на все множество B *подстановкой* A на B и обозначать композицию функций символом $\circ$. Условимся говорить, что переменная g *стохастически изоморфна* f, если существует подстановка $\varphi : X \to Y$ такая, что:

$$\begin{aligned} &g = \varphi \circ f, && r = q \circ \varphi^{-1}, \\ &s\,(x, y) = q\,(x) = r\,(y) && (y = \varphi\,(x)\,, x \in X)\,. \end{aligned}$$

Если переменная g стохастически изоморфна f, то и f стохастически изоморфна g, так как тогда для обратной подстановки $\psi = \varphi^{-1} : Y \to X$ верны равенства

$$\begin{aligned} &f = \psi \circ g, && q = r \circ \psi^{-1}, \\ &s\,(x, y) = r\,(y) = q\,(x) && (x = \psi\,(y)\,, y \in Y)\,. \end{aligned}$$

Поэтому в любом из этих случаев можно называть случайные переменные f и g *стохастически изоморфными*. Стохастически изоморфные переменные имеют одинаковое число значений и вероятности этих значений получаются друг из друга перестановками. Совместные ненулевые вероятности пар значений равны соответствующим маргинальным вероятностям.

Пример. В примере 3 пункта 1.7.9

$$U = \{00, 01, 10, 11\}, \quad X = \{-2, -1, 1, 2\}, \quad Y = \{-8, -1, 1, 8\},$$

переменные f, g и подстановки φ, ψ такие, что $g = \varphi \circ f$, $r = q \circ \varphi^{-1}$,

$f = \psi \circ g$, $q = r \circ \psi^{-1}$ представляются таблицами:

u	00	01	10	11
f	-2	-1	1	2

,

u	00	01	10	11
g	-8	-1	1	8

,

x	-2	-1	1	2
φ	-8	-1	1	8

,

y	-8	-1	1	8
ψ	-2	-1	1	2

.

Ясно, что

$$\varphi(x) = x^3, \qquad \psi(y) = y^{1/3}, \qquad \psi = \varphi^{-1}, \qquad \varphi = \psi^{-1},$$
$$g(u) = \varphi(f(u)) = (f(u))^3\,,\ f(u) = \psi(g(u)) = (g(u))^{1/3}\,.$$

Подстановки φ и ψ можно также представить таблицами, получающимися из таблицы для совместного распределения s заменой ненулевых элементов единицами:

$X \setminus Y$		-8	-1	1	8
	φ				
-2		1	0	0	0
-1		0	1	0	0
1		0	0	1	0
2		0	0	0	1

;

$X \setminus Y$		-1	8	-8	1
	φ				
-2		0	0	1	0
-1		1	0	0	0
1		0	0	0	1
2		0	1	0	0

;

$X \setminus Y$		-2	-1	1	2
	ψ				
-8		1	0	0	0
-1		0	1	0	0
1		0	0	1	0
8		0	0	0	1

;

$X \setminus Y$		-2	-1	1	2
	ψ				
-1		0	1	0	0
8		0	0	0	1
-8		1	0	0	0
1		0	0	1	0

.

Единицы выделяют пары соответствующих друг другу при этих подстановках элементов множеств X и Y. Таблицы для ψ получаются из соответствующих таблиц для φ транспонированием (поворотом вокруг главной диагонали). Таблицы показывают, что расположение единиц зависит от выбранных для записи элементов множеств X и Y порядков. Но всегда в каждой сроке и каждом столбце таблицы есть ровно одна единица. (Элементы множеств при этом не считаются: они в рассматриваемых таблицах играют роль индексов. При желании их можно заменить номерами.)

Теорема о коэффициенте информации. *1) Коэффициент информации $IK\left(f,g\right)$ случайных переменных f и g равен 0, если и только если f и g стохастически независимы.*

2) Коэффициент информации $IK\left(f,g\right)$ случайных переменных f и g равен 1, если и только если f и g стохастически изоморфны.

□ 1) Как уже отмечалось, утверждение о равенстве коэффициента информации нулю сразу следует из теоремы об информации.

2) Утверждение о равенстве коэффициента информации единице доказывается следующим образом. Как было показано

$$IK\left(f,g\right)=1\Leftrightarrow I\left(f,g\right)=H\left(f\right)=H\left(g\right).$$

Если переменные f, g стохастически изоморфны и связаны подстановками $\varphi: X\to Y$, $\psi=\varphi^{-1}: Y\to X$, то

$$s\left(\psi\left(y\right),y\right)=s\left(x,\varphi\left(x\right)\right)=q\left(\psi\left(y\right)\right)=r\left(y\right) \qquad \left(y=\varphi\left(x\right)\in Y\Leftrightarrow x=\psi\left(y\right)\in X\right).$$

Поэтому

$$I\left(f,g\right)=\sum_x s\left(x,\varphi\left(x\right)\right)\log\left(\frac{s\left(x,\varphi\left(x\right)\right)}{q\left(x\right)r\left(\varphi\left(x\right)\right)}\right)=$$
$$=\sum_x q\left(x\right)\log\left(\frac{q\left(x\right)}{q\left(x\right)r\left(\varphi\left(x\right)\right)}\right)=-\sum_x q\left(x\right)\log\left(r\left(\varphi\left(x\right)\right)\right)=$$
$$=-\sum_x r\left(\varphi\left(x\right)\right)\log\left(r\left(\varphi\left(x\right)\right)\right)=H\left(g\right).$$

Аналогично

$$I\left(f,g\right)=\sum_y s\left(\psi\left(y\right),y\right)\log\left(\frac{s\left(\psi\left(y\right),y\right)}{q\left(\psi\left(y\right),y\right)r\left(y\right)}\right)=H\left(f\right).$$

Следовательно, $I(f,g) = H(f) = H(g)$ и $IK(f,g) = 1$.

Предположим теперь, что эти равенства верны. Тогда, если $I(f,g) = H(f)$, то

$$\sum_{x,y} s(x,y) \log\left(\frac{s(x,y)}{q(x)\, r(y)}\right) = \\ = -\sum_{x} q(x) \log(q(x)) = -\sum_{x,y} s(x,y) \log(q(x)),$$

$$\sum_{x,y} s(x,y) \log\left(\frac{s(x,y)}{r(y)}\right) = \sum_{x,y} s(x,y) \log\left(\frac{s(x,y)}{q(x)\, r(y)} q(x)\right) = \\ = \sum_{x,y} s(x,y) \log\left(\frac{s(x,y)}{q(x)\, r(y)}\right) + \sum_{x,y} s(x,y) \log(q(x)) = \\ = I(f,g) - H(f) = 0,$$

$$\sum_{x,y} s(x,y) \log\left(\frac{s(x,y)}{q(x)}\right) = \sum_{x,y} s(x,y) \log\left(\frac{s(x,y)}{q(x)\, r(y)} r(y)\right) = \\ = \sum_{x,y} s(x,y) \log\left(\frac{s(x,y)}{q(x)\, r(y)}\right) + \sum_{x,y} s(x,y) \log(r(y)) = \\ = I(f,g) - H(g) = 0.$$

Пусть $x \in X$. Так как $s(x,y) \geqslant 0$, $\sum\limits_{y} s(x,y) = q(x)$, то выполнено $0 \leqslant s(x,y) \leqslant q(x)$, $s(x,y) \log(s(x,y)/q(x)) \leqslant 0$. Поэтому $\sum\limits_{x,y} s(x,y) \log(s(x,y)/q(x)) = 0 \Rightarrow s(x,y) \log(s(x,y)/q(x)) = 0$ для всех пар (x,y). А из $\sum\limits_{y} s(x,y) = q(x) > 0$ следует, что существует $y = \varphi(x) \in Y$, для которого $s(x,\varphi(x)) > 0$, $\log(s(x,\varphi(x))/q(x)) = 0$ и $s(x,\varphi(x)) = q(x)$. Такое значение $y = \varphi(x)$ единственное: $s(x,y) = 0$ для всех $y \neq \varphi(x)$, в противном случае было бы $\sum\limits_{y} s(x,y) > q(x)$. Значит $y_1 = \varphi(x_1) \neq \varphi(x_2) = y_2$ при $x_1 \neq x_2$ и в множестве Y элементов не меньше, чем в X. При этом $x = f(u)$, $g(u) = \varphi(f(u))$ для некоторого $u \in U$.

Пусть теперь $y \in Y$. Так как $s(x,y) \geqslant 0$ и $\sum_x s(x,y) = r(y)$, то $0 \leqslant s(x,y) \leqslant r(y)$, $s(x,y)\log(s(x,y)/r(y)) \leqslant 0$. Поэтому $\sum_{x,y} s(x,y)\log(s(x,y)/r(y)) = 0 \Rightarrow s(x,y)\log(s(x,y)/r(y)) = 0$ для всех пар (x,y). Из $\sum_x s(x,y) = r(y) > 0$ следует, что существует $x = \psi(y) \in X$, для которого $s(\psi(y),y) > 0$, $\log(s(\psi(y),y)/r(y)) = 0$ и $s(\psi(y),y) = r(y)$. Такое значение $x = \psi(y)$ единственное: $s(x,y) = 0$ для всех $x \neq \psi(y)$, в противном случае было бы $\sum_x s(x,y) > r(y)$. Значит $x_1 = \psi(y_1) \neq \psi(y_2) = x_2$ при $y_1 \neq y_2$ и в множестве X элементов не меньше, чем в Y. При этом $y = g(u)$, $f(u) = \psi(g(u))$ для некоторого $u \in U$.

Таким образом, если $IK(f,g) = 1$ или, что эквивалентно, $I(f,g) = H(f) = H(g)$, то для каждого $x \in X$ существует единственный элемент $y = \varphi(x) \in Y$, для которого $s(x,y) > 0$. При этом $s(x,\varphi(x)) = q(x)$. Точно так же для каждого $y \in Y$ существует единственный элемент $x = \psi(y) \in X$, для которого $s(x,y) > 0$. При этом $s(\psi(y),y) = r(y)$. Множества X и Y имеют одно и то же число элементов. Для всех $y \neq \varphi(x)$ и для всех $x \neq \psi(y)$ верно равенство $s(x,y) = 0$. Поэтому $(x,\varphi(x)) = (\psi(\varphi(x)),\varphi(x))$, $x = \psi(\varphi(x))$, $s(x,\varphi(x)) = q(x)$ для каждого $x \in X$ и $(\psi(y),y) = (\psi(y),\varphi(\psi(y)))$, $y = \varphi(\psi(y))$, $s(\psi(y),y) = r(y)$ для каждого $y \in Y$. Значит, $\varphi : X \to Y$ и $\psi : Y \to X$ являются взаимно обратными подстановками. Кроме того,

$$
\begin{aligned}
&g(u) = \varphi(f(u)), \qquad\qquad f(u) = \psi(g(u)), \\
&s(\psi(y),y) = s(x,\varphi(x)) = q(\psi(y)) = r(y) \quad (u \in U, x \in X, y \in Y).
\end{aligned}
$$

Переменные f и g стохастически изоморфны. ■

Таким образом, равенство $IK(f,g) = 1$ означает крайнюю степень зависимости между переменными f и g: эти переменные стохастически изоморфны. Они имеют одинаковое число значений и их распределения получаются друг из друга перестановками.

1.8. Условные средние и вероятности

Часто влияние случая ограничивается определенными условиями, которые описываются объединением исходов в группы, составляющие разбиение множества всех исходов. Тогда важную роль начинают играть переменные, имеющие одинаковые значения для объединенных исходов. Вводятся условные вероятности и средние значения, учитывающие сделанную группировку.

Как и в 1.5, в этом параграфе будет использоваться геометрический язык.

1.8.1. Рассмотрим невырожденное вероятностное пространство (U, p) и разбиение $\mathcal{B} = (B_j)$ множества U:

$$\sum B_j = U, \qquad B_j \neq \varnothing, \qquad B_j B_k = \varnothing \qquad (j \neq k)\,.$$

Случайные переменные $g = \sum y_j B_j$, имеющие при каждом j одно и то же значение $g\,(u) = y_j$ при всех $u \in B_j$, составляют подпространство $\mathcal{G} = \mathcal{G}\,(\mathcal{B})$ векторного пространства $\mathcal{F} = \mathcal{F}\,(U)$, составленного из переменных $f = \sum f\,(v)\,v$. (Здесь B_j одновременно обозначает множество и его индикатор, а v — элемент, элементарное множество и его индикатор: $v\,(u) = 1$ при $u = v$, и $v\,(u) = 0$ при $u \neq v$.) Нужно подобрать переменную $g \in \mathcal{G}$, имеющую то же среднее, что и данная переменная $f \in \mathcal{F}$. То есть, найти $g \in \mathcal{G}$, *близкую $f \in \mathcal{F}$ в среднем.*

Так как

$$f = \sum f B_j, \qquad Ef = \sum E\,(f B_j);$$
$$g = \sum y_j B_j, \qquad Eg = \sum y_j P\,(B_j);$$

то равенство $Ef = Eg$ получается при

$$y_j = E\,(f B_j) \,\big/\, P\,(B_j) = E\,(f \mid B_j)\,.$$

(По предположению пространство (U, p) невырождено, и поэтому $P(B_j) > 0$.) Переменная g с такими значениями называется *условным средним* переменной f при условии $\mathcal{B}$ и обозначается $E(f \mid \mathcal{B})$. Таким образом, по определению

$$E(f \mid \mathcal{B}) = \sum E(f \mid B_j) B_j.$$

Замечание. Более простую переменную $g = \sum y_j B_j \in \mathcal{G}$, близкую в среднем данной переменной $f \in \mathcal{F}$, можно найти и в вырожденном пространстве (U, p), где $P(B_j) = 0$ для некоторых индексов j. В этом случае значения y_j для этих j можно выбирать любые, так как для них $y_j P(B_j) = 0$ и по-прежнему $Eg = Ef$.

1.8.2. Если $f = A$ — событие (отождествленное со своим индикатором), то

$$E(f B_j) = E(A B_j) = P(A B_j)$$

и

$$E(A \mid \mathcal{B}) = \sum P(A \mid B_j) B_j,$$

где

$$P(A \mid B_j) = P(A B_j) \Big/ P(B_j).$$

Для каждого события B, имеющего вероятность $P(B) > 0$, число

$$P(A \mid B) = P(AB) \Big/ P(B)$$

называется *условной вероятностью A при условии B*. А случайная переменная

$$P(A \mid \mathcal{B}) = E(A \mid \mathcal{B})$$

называется *условной вероятностью A при условии $\mathcal{B}$*. Условные вероятности $P(A \mid B_j)$ служат значениями для $P(A \mid \mathcal{B})$.

Часто в качестве условия $\mathcal{B}$ выбирается разбиение, составленное из прообразов

$$B_z = h^{-1}(z) = (h = z) = \{u : h(u) = z\}$$

значений z переменной h на U. Тогда вместо $E\,(f \mid \mathcal{B})$ обычно пишут $E\,(f \mid h)$. По определению

$$E\,(f \mid h) = \sum E\,(f \mid h = z) \cdot (h = z),$$
$$E\,(f \mid h = z) = E\,(f \times (h = z)) \,\big/\, P\,(h = z)\,.$$

Если $f = A$ — событие, то

$$P\,(A \mid h) = \sum P\,(A \mid h = z) \cdot (h = z).$$

В 1.8.9 приведен простой пример.

Каждое разбиение $\mathcal{B} = (B_j)$ с перенумерованными множествами B_j порождается переменной $h = j \operatorname{ind}(B_j)$. Поэтому определения условных средних для конечных разбиений или переменных эквивалентны.

Если переменные f и h независимы, то

$$E\,(f \cdot (h = z)) = E\,(f) \cdot E\,(h = z) = E\,(f)\,P\,(h = z)\,,$$
$$E\,(f \mid h) = E\,(f)\,.$$

В частности, для независимых событий $f = A$ и $h = B$ верно равенство

$$P\,(A \mid B) = P\,(A)\,.$$

Равенство

$$E\,(f_1 f_2 \mid \mathcal{B}) = E\,(f_1 \mid \mathcal{B})\,E\,(f_2 \mid \mathcal{B})$$

определяет *условную независимость переменных* f_1, f_2 при условии $\mathcal{B}$. В частности, для событий $f_1 = A_1$, $f_2 = A_2$ равенство

$$P\,(A_1 A_2 \mid \mathcal{B}) = P\,(A_1 \mid \mathcal{B})\,P\,(A_2 \mid \mathcal{B})$$

определяет условную независимость событий при условии $\mathcal{B}$. Эти равенства обобщают независимость в среднем для переменных и обычную независимость для событий, получающихся при тривиальном разбиении $\mathcal{B} = \{U\}$.

Замечание. Понятие условного среднего $E\,(\cdot \mid \mathcal{B})$ обобщает понятие среднего E. Если разбиение $\mathcal{B} = \{U\}$ состоит из единственного множества U, то $E\,(\cdot \mid U) = E$: условное среднее равно постоянной $E\,(f)$. Точно так же $P\,(\cdot \mid U) = P$ и $P\,(A \mid U) = P\,(A)$.

1.8.3. Из определений сразу следует, что $E\,(f \mid \mathcal{B}) = f$ при $f = \sum x_j B_j \in \mathcal{G}$. В самом деле, в этом случае

$$E\,(fB_j) = E\,(x_j B_j) = x_j P\,(B_j)\,,$$
$$E\,(f \mid B_j) = E\,(fB_j) \Big/ P\,(B_j) = x_j$$

для каждого индекса j. Переменные из $\mathcal{G}$ сами служат своими условными средними.

Следовательно,

$$E\,(E\,(f \mid \mathcal{B}) \mid \mathcal{B}) = E\,(f \mid \mathcal{B})\,.$$

Повторное условное усреднение по $\mathcal{B}$ сохраняет результат. Это позволяет назвать условное усреднение по разбиению $\mathcal{B}$ *проектированием* на подпространство $\mathcal{G} = \mathcal{G}\,(\mathcal{B})$.

Рассмотрим еще разбиение $\mathcal{C} = (C_k)$ множества U более мелкое, чем $\mathcal{B} = (B_j)$: каждое B_j есть объединение некоторых C_k. Легко проверить (упражнение), что

$$E\,(E\,(f \mid \mathcal{C}) \mid \mathcal{B}) = E\,(f \mid \mathcal{B})\,.$$

Проектирование сначала на подпространство $\mathcal{H} = \mathcal{G}\,(\mathcal{C})$ пространства $\mathcal{F}$, а затем на подпространство $\mathcal{G} = \mathcal{G}\,(\mathcal{B})$ пространства $\mathcal{H}$ дает тот же результат, что и прямое проектирование сразу на $\mathcal{G}$. (Области постоянства переменных из $\mathcal{G}$ складываются из областей постоянства переменных из $\mathcal{H}$. Поэтому каждая переменная из $\mathcal{G}$ принадлежит $\mathcal{H}$ и $\mathcal{G} \subseteq \mathcal{H}$.)

1.8.4. Часто полезна *функциональная линейность* условного среднего, обобщающая обычную:

$$E\,(h_1 f_1 + h_2 f_2 \mid \mathcal{B}) = h_1 E\,(f_1 \mid \mathcal{B}) + h_2 E\,(f_2 \mid \mathcal{B})$$

для каждых $f_1, f_2 \in \mathcal{F}$ и $h_1, h_2 \in \mathcal{G} = \mathcal{G}\,(\mathcal{B})$. В частности, h_1, h_2 могут быть постоянными. Выписанное равенство для условных средних эквивалентно двум:

$$\begin{aligned} &E\,(f_1 + f_2 \mid \mathcal{B}) = E\,(f_1 \mid \mathcal{B}) + E\,(f_2 \mid \mathcal{B})\,, \\ &E\,(hf \mid \mathcal{B}) = hE\,(f \mid \mathcal{B}) \qquad\qquad (f, f_1, f_2 \in \mathcal{F}, h \in \mathcal{G})\,. \end{aligned}$$

Эти равенства легко проверяются:

$$\begin{aligned} &E\,(f_1 + f_2 \mid \mathcal{B}) = \sum E\,(f_1 + f_2 \mid B_j)\, B_j = \\ &= \sum E\,(f_1 \mid B_j)\, B_j + \sum E\,(f_2 \mid B_j)\, B_j = E\,(f_1 \mid \mathcal{B}) + E\,(f_2 \mid \mathcal{B}) \end{aligned}$$

и, при $h = \sum z_k B_k$, $hB_j = z_j$,

$$\begin{aligned} E\,(hf \mid \mathcal{B}) &= \sum_j E\,(hf \mid B_j)\, B_j = \sum_j z_j E\,(f \mid B_j)\, B_j = \\ &= \sum_k z_k B_k \cdot \sum_j E\,(f \mid B_j)\, B_j = hE\,(f \mid \mathcal{B})\,. \end{aligned}$$

Кроме того, так как $E\,(fb_j) \geqslant 0$ при $f \geqslant 0$, то

$$E\,(f \mid \mathcal{B}) \geqslant 0 \qquad\qquad (f \geqslant 0)\,.$$

И, как отмечалось,

$$E\,(E\,(f \mid \mathcal{B}) \mid \mathcal{B}) = E\,(f \mid \mathcal{B})\,.$$

Условное среднее $E\,(\cdot \mid \mathcal{B})$ каждой переменной $f \in \mathcal{F}$ ставит в соответствие переменную $g = E\,(f \mid \mathcal{B}) \in \mathcal{G} = \mathcal{G}\,(\mathcal{B})$. Это соответствие обладает свойством линейности, при нем положительным переменным f соответствуют положительные переменные g, и повторное усреднение при тех же условиях не изменяет результат. Имея в виду эти свойства, говорят, что условное среднее $E(\cdot \mid \mathcal{B})$ есть *положительный линейный оператор*, проектирующий $\mathcal{F}$ на $\mathcal{G} = \mathcal{G}(\mathcal{B})$.

1.8.5. Вероятностное пространство (U, p) можно интерпретировать как механическую систему точек $u \in U$ с положительными массами $p(u)$, в сумме равными единице. Вероятность $P(A)$ события $A \subseteq U$ равна массе тела A. Случайная переменная $f : U \to X$ ставит в соответствие системе (U, p) новую систему (X, q) точек $x \in X = f(U) \subseteq \mathbb{R}$ с массами $q(x) = Pf^{-1}(x) = P\{u : f(u) = x\}$. Сумма этих масс тоже равна единице.

Переменная f переносит массу с абстрактного множества U на вещественную прямую $\mathbb{R}$. Среднее значение Ef переменной f служит *центром масс* системы $(X, q) = \left(f(U), Pf^{-1}\right)$, а дисперсия — *моментом инерции* относительно этого центра.

Аналогия между вероятностным пространством и механической системой часто позволяет лучше понять основные определения. Представление о вероятности как о мере реализуемости события хорошо согласуется с интуитивным представлением о массе тела.

1.8.6. Рассмотрим множество $A \subseteq U$ и разбиение $\mathcal{B} = \{B, B'\}$, составленное из множества $B \subseteq U$ и его дополнения $B' = U - B$ ($P(B) > 0$, $P(B') > 0$). По определению

$$P(A \mid \mathcal{B}) = P(A \mid B)\, B + P\left(A \mid B'\right) B',$$

где

$$P(A \mid B) = P(AB) \;/\; P(B)\,, \quad P\left(A \mid B'\right) = P\left(AB'\right) \;/\; P\left(B'\right).$$

Условную вероятность $P(A \mid B)$ события A при условии B можно интерпретировать как долю массы части тела A, которая содержится в теле B в соответствующей механической системе.

Элементарная вероятность $p\,(\cdot \mid B)$ со значениями

$$p\,(u \mid B) = p\,(uB) \;/\; P\,(B) \qquad\qquad (u \in U)$$

определяет вероятностное пространство $(U, p\,(\cdot \mid B))$. В нем исходы $u \notin B$ $(u \in B')$ имеют нулевые вероятности. Исключение этих исходов приводит к невырожденному пространству $(B, p\,(\cdot \mid B))$ с множеством исходов B и элементарной вероятностью

$$p\,(u \mid B) = p\,(u) \;/\; P\,(B) \qquad\qquad (u \in B)\,.$$

Если $A \subseteq B$, то $AB = A$ и

$$P\,(A \mid B) = P\,(AB) \;/\; P\,(B) = \sum_{u \in A} p\,(u \mid B).$$

1.8.7. Из определения условной вероятности $P\,(A \mid B)$ следует *правило умножения* для вычисления вероятности пересечения $AB = A \cap B$ событий A, B:

$$P\,(AB) = P\,(A \mid B)\, P\,(B)\,.$$

Независимость событий A, B означает (1.6.10), что

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

Сравнивая выписанные равенства, получаем равенство

$$P\,(A \mid B) = P\,(A)$$

для независимых событий A, B. Вследствие симметрии для них верно также равенство

$$P\,(B \mid A) = P\,(B)\,.$$

По данному определению стохастическая независимость событий взаимна: если A не зависит от B, то и B не зависит от A. Определение не описывает односторонние зависимости.

Условные вероятности позволяют переписать формулы коэффициентов регрессии и связи для событий (1.6.12, 1.6.14). Так как

$$
\begin{aligned}
P(AB)-P(A)\,P(B) = P(AB)\left(P(B)+P\left(B'\right)\right)-P(A)\,P(B) = \\
= P(AB)\,P\left(B'\right) - \left(P(A)-P(AB)\right)P(B) = \\
= P(AB)\,P\left(B'\right) - P\left(AB'\right)P(B)\,,
\end{aligned}
$$

$$
\frac{P(AB)-P(A)\,P(B)}{P(B)\,P(B')} = \frac{P(AB)}{P(B)} - \frac{P(AB')}{P(B')}.
$$

Эта формула показывает, что коэффициент регрессии $R(A,B)$ учитывает не только зависимость события A от события B, но и зависимость A от дополнения B' события B.

Легко проверить (упражнение), что для коэффициента связи $Q(A,B)$ между событиями верна формула

$$
Q(A,B) = \frac{P(A \mid B) - P(A \mid B')}{P(A \mid B)\,P(A' \mid B') + P(A' \mid B)\,P(A \mid B')}.
$$

Этот коэффициент учитывает зависимость A, A' и B, B'.

1.8.8. Поясним сказанное об условных вероятностях абстрактным примером с простыми числами.

Рассмотрим комбинаторное пространство с множеством исходов $U = \{1, 2, \ldots, 100\}$. Пусть $A = \{$ 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97 $\}$ — множество всех простых чисел в U, а B — множество всех нечетных. Тогда A' — множество всех составных чисел в U, а B' — множество всех четных. Из определений следует, что $|A| = 25$, $|A'| = 75$, $|B| = |B'| = 50$ и $|AB| = 24$, $|AB'| = 1$, $|A'B| = 26$, $|A'B'| = 49$. Так как в комбинаторном пространстве вычисление вероятностей событий сводится к подсчету чисел элементов в этих событиях, то вычисления упрощаются.

Имеем:

$$
\begin{aligned}
&P(A \mid B) = |AB| \,/\, |B| = 24/50 = 0.48,\\
&P(B \mid A) = |AB| \,/\, |A| = 24/25 = 0.96,\\
&P\left(A \mid B'\right) = \left|AB'\right| \,/\, \left|B'\right| = 1/50 = 0.02,\\
&P\left(B \mid A'\right) = \left|A'B\right| \,/\, \left|A'\right| = 26/75 \approx 0.35,
\end{aligned}
$$

откуда

$$
\begin{aligned}
&R(A,B) = P(A \mid B) - P\left(A \mid B'\right) = 0.48 - 0.02 = 0.46,\\
&R(B,A) = P(B \mid A) - P\left(B \mid A'\right) \approx 0.96 - 0.35 = 0.61,\\
&K(A,B) = \sqrt{R(A,B)\,R(B,A)} \approx \sqrt{0.46 \cdot 0.61} \approx 0.53,\\
&Q(A,B) = \frac{24 \cdot 100 - 25 \cdot 50}{24 \cdot 49 + 26 \cdot 1} \approx 0.96.
\end{aligned}
$$

При выборе наугад числа из множества $\{1, 2, \ldots, 100\}$ вероятность выбрать простое равна 0.25. Если выбирать только из нечетных чисел, то она увеличивается до 0.48. Если же выбирать только из четных, то вероятность выбрать простое число уменьшается до 0.02. Равенство $Q(A,B) \approx 0.96$ подтверждает тесную связь между простотой и нечетностью натуральных чисел из первой сотни. Заметим, что $Q(A,B') = -Q(A,B) \approx -0.96$. Между простотой и четностью тоже есть тесная связь, но отрицательная: выбранное наугад четное число как правило не будет простым.

1.8.9. Продолжим пример с простыми числами.

Рассмотрим переменную h на комбинаторном пространстве с $U = \{1, 2, \ldots, 100\}$, описывающую количество $h(u)$ простых чисел в десятке U_i, которому принадлежит число u:

i	1	2	3	4	5	6	7	8	9	10
$h(u)$	4	4	2	2	3	2	2	3	2	1

.

Переменная h порождает разбиение $\mathcal{B} = \{B_1, B_2, B_3, B_4\}$, составленное из множеств $B_1 = U_{10}$, $B_2 = U_3 + U_4 + U_6 + U_7 + U_9$, $B_3 = U_5 + U_8$, $B_4 = U_1 + U_2$.

Пусть $f = C$ — множество всех простых чисел в U. Тогда

$$
\begin{aligned}
&P\,(C \mid h) = P\,(C \mid B_1)\,B_1 + \cdots + P\,(C \mid B_4)\,B_4,\\
&P\,(C) = E\,(P\,(C \mid h)) =\\
&\qquad = P\,(C \mid B_1)\,P\,(B_1) + \cdots + P\,(C \mid B_4)\,P\,(B_4) =\\
&\qquad\qquad = \frac{1}{10}\cdot\frac{1}{10} + \frac{2}{10}\cdot\frac{5}{10} + \frac{3}{10}\cdot\frac{2}{10} + \frac{4}{10}\cdot\frac{2}{10} = \frac{25}{100},
\end{aligned}
$$

как и должно быть.

Заменим h на Ah — переменную, описывающую количество $A(u)h(u)$ четных простых чисел в десятке U_i, которому принадлежит число u. Переменная Ah порождает разбиение $\mathcal{D} = \{\mathcal{D}_0, \mathcal{D}_1\}$, составленное из множеств $D_0 = U_2 + \cdots + U_{10}$, $D_1 = U_1$. В этом случае

$$
\begin{aligned}
&P\,(C \mid Ah) = P\,(C \mid D_0)\,D_0 + P\,(C \mid D_1)\,D_1,\\
&P\,(C) = E\,(P\,(C \mid Ah)) = P\,(C \mid D_0)\,P\,(D_0) + P\,(C \mid D_1)\,P\,(D_1) =\\
&\qquad\qquad = \frac{24}{90}\cdot\frac{90}{100} + \frac{1}{10}\cdot\frac{1}{10} = \frac{25}{100}.
\end{aligned}
$$

Результаты те же, что и при непосредственном вычислении отношения числа благоприятных исходов к числу всех.

1.8.10. Определение условного среднего $g = E\,(f \mid \mathcal{B})$, данное в 1.8.1, можно пояснить, используя геометрический язык.

Переменные $g = \sum y_j B_j$ образуют подпространство $\mathcal{G} = \mathcal{G}\,(\mathcal{B})$ евклидова пространства $\mathcal{F} = \mathcal{F}\,(\mathcal{U})$ со скалярным произведением $\langle f_1, f_2\rangle = E\,(f_1, f_2)$, описанным в 1.5.1. Разбиение $\mathcal{B} = B_j$ является ортогональной базой для $\mathcal{G}$. Ее можно нормировать, заменив B_j

на $b_j = B_j / \sqrt{P(B_j)}$:

$$\|B_j\|^2 = E\left(B_j^2\right) = E(B_j) = P(B_j)\,,$$
$$\|b_j\|^2 = E\left(b_j^2\right) = E(B_j / P(B_j)) = E(B_j) / P(B_j) = 1.$$

По определению

$$\begin{aligned} E(f \mid \mathcal{B}) &= \sum (E(fB_j) / P(B_j))\, B_j = \\ &= \sum E\left(fB_j \,/\, \sqrt{P(B_j)}\right)\left(B_j \,/\, \sqrt{P(B_j)}\right) = \\ &= \sum E(fb_j)\, b_j = \sum \langle f, b_j \rangle\, b_j. \end{aligned}$$

Это значит, что условное среднее $E(f \mid \mathcal{B})$ является ортогональной проекцией вектора f на подпространство $\mathcal{G}(\mathcal{B})$. И поэтому $E(f \mid \mathcal{B})$ есть *ближайшая в среднем квадратичном* к f переменная из $\mathcal{G}(\mathcal{B})$.

В самом деле, пусть $g = E(f \mid \mathcal{B})$ и $h \in \mathcal{G}$. Тогда из свойств условного среднего следует, что

$$\begin{aligned} \langle f - g, h \rangle &= E((f-g)\,h) = E(E((f-g)\,h \mid \mathcal{B})) = \\ &= E(hE(f-g \mid \mathcal{B})) = E(h \cdot 0) = 0. \end{aligned}$$

То есть разность $f-g$ ортогональна подпространству $\mathcal{G}$: $f - g \perp \mathcal{G}$. Следовательно, по теореме Пифагора, для всякого $h \in G$

$$\|f-h\|^2 = \|f-g+g-h\|^2 = \|f-g\|^2 + \|g-h\|^2 \geqslant \|f-g\|^2\,.$$

(Так как $\mathcal{G}$ — подпространство, то $g - h \in \mathcal{G}$ при $g \in \mathcal{G}$, $h \in \mathcal{G}$.)

Замечание. Если пространство (U,p) вырожденное и $P(B_j)$=0 для некоторых множеств B_j из разбиения $\mathcal{B}$ множества U, то из-за вырожденности скалярного произведения $\langle f_1, f_2 \rangle = E(f_1 f_2)$ для пространства $\mathcal{F} = \mathcal{F}(U,p)$ переменных на U нельзя взять ортонормированную базу $b_j = B_j / \sqrt{P(B_j)}$.

Но по-прежнему можно рассматривать ортогональную базу $\mathcal{B} = (B_j)$ с некоторыми векторами, имеющими нулевую длину:

$$\|B_j\| = \sqrt{P\,(B_j)} = 0 \ \text{ при } \ P\,(B_j) = 0.$$

Условное среднее $E\,(f \mid \mathcal{B})$ и в вырожденном случае можно определить как ортогональную проекцию f на $\mathcal{G}\,(\mathcal{B})$. Но она уже не будет единственной, как в невырожденном случае.

1.9. Формулы полной вероятности и Байеса

Описываемые в этом параграфе формула полной вероятности и следующая из нее формула Байеса очень содержательны и часто применяются.

1.9.1. Как в 1.8.1 рассмотрим невырожденное вероятностное пространство (U, p) и разбиение $\mathcal{B} = (B_j)$ множества U. По определению условной вероятности для каждого события $A \subseteq U$ верно равенство

$$P\,(A \mid \mathcal{B}) = \sum P\,(A \mid B_j)\, B_j.$$

Так как (1.8.3)

$$E\,(P\,(A \mid \mathcal{B})) = E\,(E\,(A \mid \mathcal{B})) = E\,(A) = P\,(A)\,, \quad E\,(B_j) = P\,(B_j)\,,$$

то, взяв средние значения в обеих частях определяющего $P\,(A \mid \mathcal{B})$ равенства, получаем

$$P\,(A) = \sum P\,(A \mid B_j)\, P\,(B_j).$$

Это и есть *формула полной вероятности*.

Задачи, решаемые с помощью этой формулы, схематически можно описать так.

Известны:

(1) вероятности $P\,(B_j)$ нескольких исключающих друг друга условий B_j, одно из которых обязательно выполняется;

(2) условные вероятности $P\,(A \mid B_j)$ события A при условии B_j.

Какова вероятность $P\,(A)$ события A?

По формуле полная вероятность $P\,(A)$ складывается из его частичных вероятностей $P\,(A \mid B_j)$ с весами $P\,(B_j)$.

Формула полной вероятности часто используется и в теоретических рассуждениях.

1.9.2. Поясним применение формулы полной вероятности на примере с простыми числами.

Как в 1.8.8 возьмем множество $U = \{1, 2, \ldots, 100\}$ и множество A простых чисел в U. Разобьем U на группы десятков B_1, B_2, B_3, B_4, в которых содержится соответственно 1, 2, 3, 4 простых числа. Группа B_1 состоит из одного десятка $\{91, 92, \ldots, 100\}$, группа B_2 — из пяти десятков, B_3 — из двух, B_4 — из двух. При выборе наугад числа из десятка при условии, что в нем $j = 1, 2, 3, 4$ простых чисел, вероятность выбрать простое число будет соответственно $P\,(A \mid B_j) = 1/10, 2/10, 3/10, 4/10$.

Рассмотрим три различные способа выбора групп B_j: (1) группа выбирается наугад и $P\,(B_j) = 1/4$; (2) группа выбирается пропорционально количеству простых чисел в десятке и $P\,(B_j) = j/10$; (3) группа выбирается пропорционально числу десятков в ней и $P\,(B_j) = 1/10, 5/10, 2/10, 2/10$. Соответственно по формуле полной вероятности получаем:

$$P\,(A) = \frac{1}{4}\left(\frac{1}{10} + \frac{2}{10} + \frac{3}{10} + \frac{4}{10}\right) = \frac{1}{4},$$

$$P\,(A) = \frac{1}{10}\frac{1}{10} + \frac{2}{10}\frac{2}{10} + \frac{3}{10}\frac{3}{10} + \frac{4}{10}\frac{4}{10} = \frac{3}{10},$$

$$P\,(A) = \frac{1}{10}\frac{1}{10} + \frac{2}{10}\frac{5}{10} + \frac{3}{10}\frac{2}{10} + \frac{4}{10}\frac{2}{10} = \frac{1}{4}.$$

Первый и третий способы дают тот же результат, что и простой выбор наугад числа из рассматриваемых ста. Второй способ увеличивает вероятность выбрать простое число, что понятно.

1.9.3. В связи с задачей о вычислении вероятности события по известным условным вероятностям и вероятностям условий возникает естественная обратная задача: вычислить вероятность выполнения условия при совершившемся событии.

Для решения такой задачи применяется *формула Байеса*:

$$P(B_j \mid A) = \frac{P(A \mid B_j)\, P(B_j)}{\sum P(A \mid B_k)\, P(B_k)}.$$

Эта формула следует из равенств

$$P(AB_j) = P(A \mid B_j)\, P(B_j) = P(B_j \mid A)\, P(A)$$

и формулы полной вероятности. Простота вывода не умаляет важности этой формулы. Ее часто называют также *формулой вероятностей гипотез*.

Задачи, решаемые с помощью формулы Байеса, можно схематически описать так.

Известны:

(1) вероятности $P(B_j)$ нескольких исключающих друг друга предположений B_j, одно из которых верно;

(2) условные вероятности $P(A \mid B_j)$ события A при условии B_j.

Какова условная вероятность $P(B_j \mid A)$ верности предположения B_j, если событие A произошло?

Формула Байеса часто используется при решении статистических задач. Если вероятность $P(A)$ известна, то вместо формулы Байеса можно использовать равенство

$$P(B_j \mid A) = P(A \mid B_j)\, P(B_j) / P(A).$$

1.9.4. Поясним применение формулы Байеса на примере с простыми числами.

Возьмем те же U, A и B_j, что и в 1.9.2. Пусть наугад выбранное из U число оказалось простым. Какова вероятность того, что оно выбрано из десятка с j простыми числами? При выборе наугад вероятности групп B_j пропорциональны числам десятков в них: $P(B_j) = 1/10, 5/10, 2/10, 2/10$. Условные вероятности выбрать простое число $P(A \mid B_j) = 1/10, 2/10, 3/10, 4/10$. По формуле Байеса получаем:

$$
\begin{aligned}
P(B_1 \mid A) &= 0.1 \cdot 0.1/0.25 = 0.04,\\
P(B_2 \mid A) &= 0.2 \cdot 0.5/0.25 = 0.40,\\
P(B_3 \mid A) &= 0.3 \cdot 0.2/0.25 = 0.24,\\
P(B_4 \mid A) &= 0.4 \cdot 0.2/0.25 = 0.32.
\end{aligned}
$$

Наиболее вероятно, что в десятке, из которого выбрано простое число, их было два (а не четыре, как можно было ожидать). Это объясняется тем, что таких десятков больше, чем других. Вероятность того, что в этом десятке было четыре простых числа меньше (хотя и ненамного).

Более надежно предположить, что в десятке было четное число простых:

$$P(B_2 + B_4 \mid A) = P(B_2 \mid A) + P(B_4 \mid A) = 0.72.$$

Наименее вероятно, что в десятке, из которого выбрано простое число, оно было единственное. Это естественно, потому что такой десяток тоже единственный.

1.10. Закон больших чисел

Этот фундаментальный закон связывает стохастические средние со статистическими. Связь между ними выражается в самых разных формах. Закон больших чисел используется в теоретических рассуждениях и при решении прикладных задач.

1.10.1. Рассмотрим вероятностное пространство (U, p), случайную переменную f на U и произвольное число $\varepsilon > 0$. Для них верно

Неравенство Чебышева

$$P\left(u : |f\left(u\right) - Ef| \geqslant \varepsilon\right) \leqslant \varepsilon^{-2} Df.$$

□ Пусть $g = |f - Ef|^2$, $B = \{u : |f\left(u\right) - Ef| \geqslant \varepsilon\} = \{u : g \geqslant \varepsilon^2\}$. Тогда, отождествляя событие B со своим индикатором, имеем: $g \geqslant Bg \geqslant \varepsilon^2 B$ и

$$Df = Eg \geqslant \varepsilon^2 E\left(B\right) = \varepsilon^2 P\left(B\right).$$

Неравенство $Df \geqslant \varepsilon^2 P\left(B\right)$ эквивалентно доказываемому. ■

Неравенство Чебышева часто записывают в другой форме:

$$P\{u : |f\left(u\right) - Ef| < \varepsilon\} \geqslant 1 - \varepsilon^{-2} Df.$$

Заметим, что при слишком малых числах ε неравенство Чебышева становится тривиальным: его правая часть может быть больше единицы (а в другой форме — меньше нуля).

1.10.2. Рассмотрим последовательность попарно некоррелированных случайных переменных f_j $(j = 1, \ldots, n)$, имеющих одинаковые распределения. Пусть

$$Ef_j = a, \qquad Df_j = b^2.$$

Рассмотрим еще среднее арифметическое $f = n^{-1} \sum f_j$ этих переменных. Благодаря одинаковой распределенности и некоррелированности переменных f_j верны равенства:

$$Ef = n^{-1} na = a, \qquad Df = n^{-2} nb^2 = n^{-1} b^2.$$

Неравенство Чебышева для этого случая переписывается так:

$$P\{u : \left|n^{-1} \sum f_j\left(u\right) - a\right| \geqslant \varepsilon\} \leqslant b^2 / \left(n\varepsilon^2\right).$$

Или, в эквивалентной форме,

$$P\{u : \left|n^{-1} \sum f_j(u) - a\right| < \varepsilon\} \geqslant 1 - b^2/\left(n\varepsilon^2\right).$$

Для каждого $\varepsilon > 0$ при всех достаточно больших n правая часть первого из выписанных неравенств произвольно близка к нулю, а второго — к единице. Оба этих неравенства означают, что вероятность отклонения среднего арифметического переменных f_j от их стохастического среднего a может быть произвольно мала при достаточно большом числе этих переменных.

Выписанные неравенства выражают *закон больших чисел Чебышева*.

Замечание. Из закона больших чисел Чебышева *не следует*, что неравенство $\left|n^{-1} \sum f_j(u) - a\right| < \varepsilon$ выполняется при всех исходах u для каждого n, большего некоторого $n(\varepsilon)$. Это неравенство только имеет вероятность, близкую к единице.

Неравенство Чебышева позволяет оценить погрешность при замене стохастического среднего арифметическим с данной надежностью. Возьмем $\alpha \in]0, 1/2[$ и условимся считать каждое событие B с вероятностью $P(B) \leqslant \alpha$ *практически невозможным*, а каждое событие A с вероятностью $P(A) \geqslant 1 - \alpha$ — *практически достоверным*. Неравенство $b^2/\left(n\varepsilon^2\right) \leqslant \alpha$ эквивалентно неравенству $\varepsilon \geqslant b/\sqrt{n\alpha}$. Значит, отклонение среднего арифметического от среднего стохастического при сделанных предположениях больше, чем на $\varepsilon = b/\sqrt{n\alpha}$ практически невозможно. А для достижения данной точности ε с практической достоверностью (или *надежностью* $1 - \alpha$) достаточно взять $n \geqslant b^2/\left(\alpha\varepsilon^2\right)$.

В качестве верного значения измеряемой величины часто используется среднее арифметическое результатов большого числа независимых измерений, произведенных в одинаковых условиях. Это основано на законе больших чисел. Из-за случайных отклонений такая оценка бывает иногда довольно грубой и увеличение числа измерений не всегда учитывает ее точность.

1.10.3. Рассмотрим последовательность переменных $f_j = s_j$ $(j = 1, \ldots, n)$ в пространстве Бернулли $B\,(n, a)$ (1.1.3). Их среднее арифметическое $f = n^{-1} \sum s_j$ описывает частоту появления единицы, а среднее стохастическое $Ef = a$ — вероятность появления единицы на каждом данном месте. Переменные s_j независимы и одинаково распределены, $Es_j = a$ и $Ds_j = a\,(1-a) = b^2$.

Из неравенства Чебышева следует *неравенство Бернулли*:

$$P\{u : \left| n^{-1} \sum s_j\,(u) - a \right| \geqslant \varepsilon\} \leqslant a\,(1-a) \,/\, \left(n\varepsilon^2\right).$$

Эквивалентная форма:

$$P\{u : \left| n^{-1} \sum s_j\,(u) - a \right| < \varepsilon\} \geqslant 1 - a\,(1-a) \,/\, \left(n\varepsilon^2\right).$$

Эти неравенства выражают *закон больших чисел Бернулли*.

Закон больших чисел Бернулли выражает связь между частотой появления данного события в последовательности независимых испытаний, произведенных в одинаковых условиях, и вероятностью этого события. Частота служит оценкой для вероятности. Неравенство Бернулли позволяет с нужной надежностью указать точность этой оценки и необходимое число испытаний.

1.10.4. Пример. Один из основоположников математической статистики К. Пирсон подбросил монету $n = 24000$ раз. Герб выпал $m = 12012$ раз. Частота его появления равна $m/n = 0.5005$.

Если считать, что монета была симметричной, а подбрасывания производились независимо и одним и тем же способом, то опыт К. Пирсона можно описать последовательностью Бернулли $B\,(n, a)$ при $n = 24000$ и $a = 1/2$, отождествляя появление герба с появлением единицы. Тогда среднее арифметическое $n^{-1} \sum s_j$ будет описывать частоту появления герба. Вероятность появления герба при j-м подбрасывании $(j = 1, 2, \ldots, 24000)$ равна $a = 1/2$. В опыте отклонение $|m/n - a| = 0.0005$. Условимся считать практически невозможными события, вероятность которых меньше $\alpha = 0.001$. Из неравенства Бернулли следует, что практически невозможны отклонения $|m/n - a| \geqslant 0.103$. В опыте К. Пирсона оно значительно меньше. Этот опыт не противоречит предположению о симметричности монеты.

Для того, чтобы отклонение $\varepsilon = 0.0005$ противоречило предположению о симметричности монеты, нужно было бы получить такое отклонение, подбросив монету $n \geqslant 10^9$ раз. По статистике отклонение в таких пределах получается при значительно меньшем числе подбрасываний. Полученное для n неравенство объясняется грубостью оценки в неравенстве Чебышева, неизбежной из-за его универсальности.

Упражнение. Провести соответствующий компьютерный эксперимент, моделируя подбрасывание симметричной монеты.

1.11. Экспоненциальные полиномы

Так здесь называются начальные отрезки ряда, в который разлагается экспоненциальная функция. С их помощью в 1.12 будут получены уточнения неравенств Чебышева и Бернулли. Предварительно описываются свойства экспоненциальных полиномов. Все эти свойства подробно доказаны в [8].

1.11.1. Будем называть l-м *экспоненциальным полиномом* (или, короче, *экспоненциалом*) полином l-й степени e_l с коэффициентами $1/k!$ $(0 \leqslant k \leqslant l)$:

$$e_l\,(x) = \sum_{0 \leqslant k \leqslant l} x^k/k! \qquad\qquad (x \in \mathbb{R})\,.$$

В частности,

$$e_0\,(x) \equiv 1, \qquad e_1\,(x) \equiv 1 + x, \qquad e_2\,(x) \equiv 1 + x + x^2/2,$$
$$e_3\,(x) \equiv 1 + x + x^2 + x^3/6.$$

Главное свойство экспоненциалов состоит в том, что они преобразуют суммы в произведения с небольшим искажением:

$$e_l(x_1)\cdots e_l(x_n) = e_l(x_1+\cdots+x_n) + r_l(x_1,\ldots,x_n)\,, \qquad (1)$$

где

$$r_l(x_1,\ldots,x_n) = \sum_{l<k\leqslant nl}\ \sum_{k(1),\ldots,k(n)} x_1^k(1)\cdots x_n^k(n)\ /\ (k(1)!\cdots k(n)!)$$

(внутреннее суммирование производится по всем $k(1),\ldots,k(n)\geqslant 0$ с суммой $k(1)+\cdots+k(n)=k$). Равенство (1) верно для всех вещественных $x_1,\ldots,x_n$ (комплексных тоже, но они не понадобятся). Его легко доказать по индукции (упражнение) или используя *полиномиальную формулу* ([8]).

1.11.2. В экспоненциальном полиноме сумма членов с достаточно большими номерами произвольно мала:

$$|e_m(x)-e_l(x)| \leqslant 2\,|x|^{l+1}\ /\ (l+1)! \qquad (2)$$

для всех $m\geqslant l\geqslant 2\,|x|$. В самом деле, при этих условиях

$$|e_m(x)-e_l(x)| \leqslant \sum_{l<k\leqslant m} |x|^k\ /\ k!,$$

$$(l+1)!\,|x|^{k-l-1}\ /\ k! \leqslant |x|^{k-l-1}\ /\ ((l+2)\cdots k) \leqslant$$
$$\leqslant |x|^{k-l-1}\ /\ (l+2)^{k-l-1} \leqslant (1/2)^{k-l-1}$$

для каждого $k>l+1$. Поэтому

$$\sum_{l<k\leqslant m} |x|^k\ /\ k! \leqslant \Big(|x|^{l+1}\ /\ (l+1)!\Big)\sum_{l<k\leqslant m}(1/2)^{k-l-1} \leqslant$$
$$\leqslant 2\,|x|^{l+1}\ /\ (l+1)!.$$

Так как $|x|/k \leqslant 1/2$ при $k \geqslant 2|x|$, то

$$2|x|^{l+1}/(l+1)! \leqslant c(x)/2^{l-2|x|-1},$$

где $c(x)$ равно произведению дробей $|x|/j$ при $j < 2|x|$. Следовательно, $2|x|^{l+1}/(l+1)! \to 0$, когда $l \to \infty$. Следовательно, $|e_m(x) - e_l(x)| \to 0$, когда $m \geqslant l$ и $l \to \infty$. Сумма членов с номерами $l+1, \ldots, m$ экспоненциала e_m произвольно мала, когда l достаточно велико. Это используется в различных оценках с экспоненциалами.

1.11.3. Неравенство (2) позволяет оценить разность

$$r_l(x_1, \ldots, x_n) = e_l(x_1) \cdots e_l(x_n) - e_l(x_1 + \cdots + x_n).$$

При $l \geqslant 2(|x_1| + \cdots + |x_n|)$ верно неравенство

$$|r_l(x_1, \ldots, x_n)| \leqslant 2(|x_1| + \cdots + |x_n|)^{l+1} / (l+1)! \qquad (3)$$

В самом деле, по индукции легко доказать, что

$$|r_l(x_1, \ldots, x_n)| \leqslant e_{nl}(|x_1| + \cdots + |x_n|) - e_l(|x_1| + \cdots + |x_n|).$$

Отсюда и из (2) следует (3).

Из неравенства (3) в частности следует, что

$$|r_2(x_1, x_2)| \leqslant (|x_1| + |x_2|)^3/3 \qquad (|x_1| + |x_2| \leqslant 1).$$

1.11.4. Для каждого вещественного числа x экспоненциалы с достаточно большими номерами имеют строго положительные значения. Если $x \geqslant 0$, то $e_l(x) \geqslant 0$ для каждого номера $l \geqslant 0$. Это следует из определения экспоненциалов. Точнее:

$$e_l(0) = 1, \qquad e_l(x) \geqslant 1 \qquad (l \geqslant 0, \quad x \geqslant 0).$$

Если же $x < 0$, то неравенство $e_l(x) > 0$ становится верным при $l > (1 + 4|x|)(1 + \log_2|x|)$.

В самом деле, из (1) следует, что

$$e_l(x)\,e_l(-x) = e_l(x-x) + r_l(x,-x) = 1 + r_l(x,-x)\,.$$

Так как $-x = |x| > 0$ при $x < 0$, то отсюда вытекает, что $e_l(x) > 0$ вместе с $e_l(-x) > 0$ при $|r_l(x,-x)| < 1$. Из (2) следует, что

$$|r_l(x,-x)| \leqslant 2\,(2\,|x|)^{l+1} \,/\, (l+1)!$$

при $l \geqslant 4\,|x|$. Пусть $4\,|x| \leqslant k < 4\,|x| + 1$. Тогда при $l > k$

$$2\,(2\,|x|)^{l+1} \,/\, (l+1)! \leqslant 4\,|x|\,(2\,|x| \,/\, (l_0+1))^{l-k+1} < 4\,|x|^{l_0}\,2^{-(l-k+1)}.$$

Следовательно, $|r_l(x,-x)| < 1$ при $l > k$ и $4\,|x|^k\,2^{-(l-k+1)} \leqslant 1$. Из последнего неравенства следует $2 + k\log_2|x| - (l-k+1) \leqslant 0$, что эквивалентно $l \geqslant k\,(1+\log_2|x|) + 1$. Так как $k < 4\,|x| + 1$, то это неравенство верно при $l > (1+4\,|x|)\,(1+\log_2|x|)$.

1.11.5. Рассмотрим конечную последовательность независимых случайных переменных f_j $(j = 1, \ldots, n)$ на вероятностном пространстве (U, p), не предполагая, что они имеют одинаковое распределение.

Положим:

$$Ef_j = a_j, \qquad Df_j = b_j, \qquad a = \sum a_j, \qquad b^2 = \sum b_j^2.$$

Будем предполагать, что $b > 0$ (то есть хотя бы одна из дисперсий $Df_j = b_j > 0$). Вместе с f_j рассмотрим еще переменные

$$g_j = (f_j - a_j)\,/b, \qquad\qquad g = \sum g_j.$$

Переменные g_j тоже независимы и $Eg_j = 0$. Их сумма центрирована и нормирована:

$$Eg = 0 \qquad\qquad Dg = 1.$$

Для каждых чисел $\varepsilon > 0$, $t > 0$ и экспоненциала e_l рассмотрим переменные tg_j, $t\,(g-\varepsilon)$ и их композиции с e_l:

$$h_j = e_l\,(tg_j) = \sum_{0\leqslant k\leqslant l} t^k g_j^k \;/\; k!,$$
$$h = e_l\,(t\,(g-\varepsilon)) = \sum_{0\leqslant k\leqslant l} t^k\,(g-\varepsilon)^k \;/\; k!.$$

Из независимости g_j следует независимость случайных переменных h_j.

1.11.6. Для переменных g и h нетрудно получить неравенство чебышевского типа.

Рассмотрим число $c > 0$, ограничивающее абсолютные значения всех переменных g_j:

$$|g_j\,(u)| \leqslant c \qquad\qquad (u \in U, \quad 1 \leqslant j \leqslant n)\,.$$

Заметим, что

$$|g\,(u)| \leqslant \sum |g_j\,(u)| \leqslant nc \qquad\qquad (u \in U)\,.$$

Будем предполагать, что

$$t \leqslant 1/c, \tag{T}$$

$$l \geqslant l\,(n, c, \varepsilon) = 1/c^2 + (1 + 4\,(n + \varepsilon/c))\,(1 + \log_2\,(n + \varepsilon/c))\,. \tag{L}$$

Лемма. *При сделанных предположениях*

$$P\{u : g\,(u) \geqslant \varepsilon\} \leqslant E\,(h)\,.$$

Это неравенство оценивает вероятность значительных отклонений переменной g от ее среднего значения $Eg = 0$ *в положительную сторону*. Докажем его.

Пусть

$$A = \{u : g(u) \geqslant \varepsilon\} = \{u : g(u) - \varepsilon \geqslant 0\}.$$

Тогда

$$h(u) = e_l(t(g(u) - \varepsilon)) \geqslant 1 \qquad (u \in A).$$

Из условий следует, что

$$l > (1 + 4t(nc + \varepsilon))(1 + \log_2(t(nc + \varepsilon))) \geqslant$$
$$\geqslant (1 + 4t|g(u) - \varepsilon|)(1 + \log_2(t|g(u) - \varepsilon|))$$

для каждого $u \in U$. Откуда (1.11.4)

$$h(u) = e_l(t(g(u) - \varepsilon)) > 0 \qquad (u \in U).$$

Значит, $A \leqslant h \quad (h(u) \geqslant 1$ влечет $h(u) > 0)$ и

$$P(A) = E(A) \leqslant E(h),$$

что и требовалось доказать. ■

1.11.7. Для того, чтобы получить оценку для среднего Eh, оценим сначала средние Eh_j. Будем использовать те же обозначения и предположения, что и в 1.11.6.

Лемма.

$$Eh_j \leqslant e_l\left(\frac{b_j^2}{b^2}\frac{t^2}{2}\left(1 + \frac{tc}{2}\right)\right).$$

Так как $E\left(g_j^0\right) = E(1) = 1$, $E\left(g_j^1\right) = E(g) = 0$, то

$$Eh_j = 1 + \sum_{2 \leqslant k \leqslant l} t^k E\left(g_j^k\right) / k!.$$

Пусть $k \geqslant 2$. Тогда $g_j^k \leqslant c^{k-2} g_j^2$. А так как $E\left(g_j^2\right) = D\left(g_j\right) = b_j^2/b^2$, то

$$E\left(g_j^k\right) \leqslant c^{k-2} b_j^2/b^2.$$

Кроме того,

$$c^{k-2}\frac{t^k}{k!} \leqslant c^{k-2}\frac{t^k}{2\cdot 3^{k-2}} \leqslant \frac{t^2}{2}\left(\frac{tc}{3}\right)^{k-2}.$$

Откуда

$$\sum_{2\leqslant k\leqslant l}\frac{t^k}{k!}E\left(g_j^k\right) \leqslant \frac{b_j^2}{b^2}\frac{t^2}{2}\frac{1}{1-tc/3}.$$

Если $tc \leqslant 1$ (T), то

$$\frac{1}{1-tc/3} = 1+\frac{tc/3}{1-tc/3} \leqslant 1+\frac{tc/3}{1-1/3} = 1+\frac{tc}{2}.$$

Поэтому

$$\sum_{2\leqslant k\leqslant l}\frac{t^k}{k!}E\left(g_j^k\right) \leqslant \frac{b_j^2}{b^2}\frac{t^2}{2}\left(1+\frac{tc}{2}\right)$$

и

$$Eh_j \leqslant 1+\frac{b_j^2}{b^2}\frac{t^2}{2}\left(1+\frac{tc}{2}\right).$$

Остается использовать неравенство

$$e_l\left(x\right) \geqslant 1+x,$$

верное для каждых $l > 0$ и $x \geqslant 0$, заметив, что $l > 0$ по предположению $l\left(n, c, \varepsilon\right) > 0$ (L).

1.11.8. Оценим теперь среднее $E(h)$, выбрав $\varepsilon > 0$. Сохранив обозначение $l(n, c, \varepsilon)$ из 1.11.6 положим еще

$$R_l(n, c, \varepsilon) = \frac{2}{(l+1)!}\left(\left(n + \frac{\varepsilon}{c}\right)^{l+1} + \left(\frac{1}{c^2} + \frac{\varepsilon}{c}\right)^{l+1}\right).$$

Лемма. Если $0 < t \leqslant 1/c$ и $l \geqslant l(n, c, \varepsilon)$, то

$$E(h) = e_l\left(-t\varepsilon + \frac{t^2}{2}\left(1 + \frac{tc}{2}\right)\right) + R_l, \qquad R_l \leqslant R_l(n, c, \varepsilon).$$

Заметим, что при $t \leqslant 1/c$ (T)

$$\sum |c_j| + |-t\varepsilon| = \frac{t^2}{2}\left(1 + \frac{tc}{2}\right) + t\varepsilon \leqslant \frac{1}{c^2} + \frac{\varepsilon}{c},$$

а так как $l(n, c, \varepsilon) \geqslant 1/c^2 + \varepsilon/c$, то из $l \geqslant l(n, c, \varepsilon)$ следует, что

$$|r_l(c_1, \ldots, c_n, -t\varepsilon)| \leqslant \beta = \frac{2}{(l+1)!}\left(\frac{1}{c^2} + \frac{\varepsilon}{c}\right)^{l+1}.$$

Значит,

$$E(h) \leqslant e_l\left(-t\varepsilon + \frac{t^2}{2}\left(1 + \frac{tc}{2}\right)\right) + (\alpha + \beta).$$

Так как $\alpha + \beta = R_l(n, c, \varepsilon)$, то неравенство леммы верно. ■

1.12. Экспоненциальные неравенства

С помощью описанных в 1.11 экспоненциальных полиномов можно уточнить оценки вероятностей в неравенствах Чебышева и Бернулли.

1.12.1. Рассмотрим последовательность независимых случайных переменных f_j $(j = 1, \ldots, n)$ на вероятностном пространстве (U, p). Будем использовать те же обозначения и предположения, что и в 1.11.5 и 1.11.6.

Пусть $\varepsilon > 0$, $0 < t \leqslant 1/c$ и по-прежнему

$$l\,(n, c, \varepsilon) = 1/c^2 + (1 + 4\,(n + \varepsilon/c))\,(1 + \log_2(n + \varepsilon/c))\,,$$
$$R_l\,(n, c, \varepsilon) = \frac{2}{(l+1)!}\left(\left(n + \frac{\varepsilon}{c}\right)^{l+1} + \left(\frac{1}{c^2} + \frac{\varepsilon}{c}\right)^{l+1}\right).$$

Тогда из лемм 1.11.6 и 1.11.8 следует

Теорема. *Если $l \geqslant l\,(n, c, \varepsilon)$, то*

$$P\{u : g\,(u) \geqslant \varepsilon\} \leqslant e_l\left(-t\varepsilon + \frac{t^2}{2}\left(1 + \frac{tc}{2}\right)\right) + R_l,$$
$$R_l \leqslant R_l\,(n, c, \varepsilon)\,.$$

Целесообразно упростить полученную оценку. Если $\varepsilon \leqslant 1/c$, то $t = c$ удовлетворяет условиям и в этом случае.

$$-t\varepsilon + \frac{t^2}{2}\left(1 + \frac{tc}{2}\right) = -\varepsilon^2 + \frac{\varepsilon^2}{2}\left(1 + \frac{\varepsilon c}{2}\right) = -\frac{\varepsilon^2}{4}\,(2 - \varepsilon c) \leqslant -\frac{\varepsilon^2}{4}.$$

Если $\varepsilon > 1/c$, то при $t = 1/c$ и $c \geqslant 1$

$$-t\varepsilon + \frac{t^2}{2}\left(1 + \frac{tc}{2}\right) = -\frac{\varepsilon}{c} + \frac{3}{4}\frac{1}{c^2} \leqslant -\frac{1}{4}\frac{\varepsilon}{c} = -\frac{\varepsilon^2}{4}\frac{1}{\varepsilon c}.$$

Положим

$$\varphi\,(\varepsilon c) = \min\{1, 1/\,(\varepsilon c)\}.$$

Из полученных неравенств следует, что

$$-t\varepsilon + \frac{t^2}{2}\left(1 + \frac{tc}{2}\right) \leqslant -\frac{\varepsilon^2}{4}\varphi\,(\varepsilon c)$$

при $t = \varepsilon \leqslant 1/c$ или $\varepsilon > 1/c = t$.

1.12.2. Сказанное в 1.12.1 позволяет сформулировать сразу вытекающее из теоремы

Следствие 1. *Если* $l \geqslant l(n,c,\varepsilon)$, *то*

$$P\{u : g(u) \geqslant \varepsilon\} \leqslant e_l\left(-\varepsilon^2\varphi(\varepsilon c)/4\right) + R_l(n,c,\varepsilon).$$

Полученное неравенство оценивает вероятность отклонения переменной g от нуля в положительную сторону. Сохранив сделанные предположения, из этого неравенства легко получить аналогичные оценки для отрицательных и абсолютных отклонений.

Следствие 2. *Если* $l \geqslant l(n,c,\varepsilon)$, *то*

$$P\{u : g(u) \leqslant -\varepsilon\} \leqslant e_l\left(-\varepsilon^2\varphi(\varepsilon c)/4\right) + R_l(n,c,\varepsilon).$$

Это неравенство получается переходом от переменных f_j, g_j, g к переменным $-f_j$, $-g_j$, $-g$. Переменные $-f_j$ независимы вместе с f_j. Кроме того, $E(-f_j) = -E(f_j) = -a_j$, $D(-f_j) = D(f_j) = b^2$, $-g_j = -(f_j - a_j)/b = ((-f_j) - (-a_j))/b$. Поэтому для суммы $-g = \sum(-g_j)$ верно следствие 1, а для вероятности события $\{u : g(u) \leqslant -\varepsilon\} = \{u : -g(u) \geqslant \varepsilon\}$ верно доказываемое неравенство. Следствие 2 эквивалентно следствию 1.

Так как

$$\{u : |g(u)| \geqslant \varepsilon\} = \{u : g(u) \geqslant \varepsilon\} \cup \{u : g(u) \leqslant -\varepsilon\},$$

то из следствий 1 и 2 вытекает

Следствие 3. *Если* $l \geqslant l(n,c,\varepsilon)$, *то*

$$P\{u : |g| \geqslant \varepsilon\} \leqslant 2\left(e_l\left(-\varepsilon^2\varphi(\varepsilon c)/4\right) + R_l(n,c,\varepsilon)\right).$$

Условимся для определенности *экспоненциальным неравенством* называть неравенство следствия 3.

1.12.3. Предположим теперь, что независимые переменные f_j $(j = 1, \ldots, n)$ имеют одно и то же распределение, $Ef_j = a$, $Df_j = b^2 > 0$. Как в 1.11.5 положим

$$g_j = (f_j - a) / \left(\sqrt{n} b\right), \qquad g = \sum g_j = \sqrt{n} \left(\sum f_j / n - a\right) / b.$$

Сумма g пропорциональна отклонению среднего арифметического $\sum f_j / n$ случайных переменных f_j от их общего среднего a. Для оценки вероятности значительного отклонения можно использовать экспоненциальное неравенство.

Пусть $\varepsilon > 0$, $c > 0$ и $|g_j| \leqslant c$ для $j = 1, \ldots, n$. Верна

Теорема. *Если* $l \geqslant l(n, c, \varepsilon)$, *то*

$$P\{u : \left|\sum f_j(u) / n - a\right| \geqslant \varepsilon b / \sqrt{n}\} \leqslant \\ \leqslant 2 \left(e_l \left(-\varepsilon^2 \varphi(\varepsilon c) / 4\right) + R_l(n, c, \varepsilon)\right).$$

Так как

$$\{u : \left|\sum f_j(u) / n - a\right| \geqslant \varepsilon b / \sqrt{n}\} = \{u : |g(u)| \geqslant \varepsilon\},$$

то эта теорема сразу следует из экспоненциального неравенства в 1.12.2.

Заметим, что из неравенства Чебышева (1.10.1) следует, что

$$P\{u : \left|\sum f_j(u) / n - a\right| \geqslant \varepsilon b / \sqrt{n}\} \leqslant 1 / \varepsilon^2.$$

Неравенство теоремы часто оказывается точнее. Можно доказать (упражнение), что для каждых чисел $\alpha > 0$, $\gamma > 0$ существуют число $x_0 > 0$ и номер l_0 такие, что

$$\alpha e_l(-\gamma x) < 1 / x^2 \qquad (x > x_0, \;\; l \geqslant l_0).$$

1.12.4. Рассмотрим пространство Бернулли $B(n,a)$, случайные переменные $f_j = s_j$, их сумму $s = \sum s_j$ и среднее арифметическое s/n. Верны равенства

$$E(s_j) = a, \qquad D(s_j) = b^2 = a(1-a), \qquad (1 \leqslant j \leqslant n).$$

Так как $|s_j| \leqslant c = 1$, то можно конкретизировать выражения для $l(n,c,\varepsilon)$ и $R_l(n,c,\varepsilon)$:

$$l(n,1,\varepsilon) = 1 + (1 + 4(n+\varepsilon))(1 + \log_2(n+\varepsilon)),$$
$$R_l(n,1,\varepsilon) = \frac{2}{(l+1)!}\left((n+\varepsilon)^{l+1} + (1+\varepsilon)^{l+1}\right).$$

Кроме того, $\varphi(\varepsilon) = \min\{1, 1/\varepsilon\}$.

Из теоремы 1.12.3 вытекает

Следствие. *Если $l \geqslant l(n,1,\varepsilon)$, то*

$$P\{u : |s/n - a| \geqslant \varepsilon\sqrt{a(1-a)/n}\} \leqslant$$
$$\leqslant 2\left(e_l\left(-\varepsilon^2\varphi(\varepsilon)/4\right) + R_l(n,1,\varepsilon)\right).$$

Это неравенство уточняет неравенство Бернулли в 1.10.3.

Легко проверить (упражнение), что $a(1-a) \leqslant 1/4$, когда $0 \leqslant a \leqslant 1$. Поэтому в неравенстве следствия можно $\varepsilon\sqrt{a(1-a)/n}$ заменить на $\varepsilon/(2\sqrt{n})$.

1.12.5. Предположим, что последовательность n измерений некоторой величины описывается последовательностью независимых и одинаково распределенных случайных переменных f_j, общее среднее значение которых $E(f_j) = a$ равно значению измеряемой величины, а дисперсия $D(f_j) = b^2$ характеризует точность каждого измерения.

Выберем $\alpha \in]0, 1/2[$ и условимся считать событие B, вероятность которого $P(B) \leqslant \alpha$, *практически невозможным*, а каждое событие A, вероятность которого $P(A) \geqslant 1 - \alpha$ — *практически* *достоверным*. Неравенство $b^2/\left(n\varepsilon^2\right) \leqslant \alpha$ верно, если $\varepsilon = b/\sqrt{n\alpha}$. Из неравенства Чебышева (1.10.2) следует, что при такой точности измерений отклонение среднего арифметического результатов измерений от значения измеряемой случайной величины практически невозможно. В частности, если $b^2 = 1$, $n = 1000$ и $\alpha = 0.001$, то измерения могут иметь точность $\varepsilon = 1$.

Неравенство Чебышева позволяет также оценить число измерений n, практически достоверно обеспечивающее нужную точность. Неравенство $1-b^2/(n\varepsilon^2) \geqslant 1-\alpha$ эквивалентно $n \geqslant b^2/(\alpha\varepsilon^2)$. При таком числе измерений точность ε практически достоверна. В частности, если $b^2 = 0.01$, $\alpha = 0.01$, то $n \geqslant 100$.

Теорема 1.12.3 позволяет улучшить эти результаты. Пусть снова $b = 1$, $n = 1000$ и $\alpha = 0.001$. Кроме того, возможные отклонения при каждом измерении меньше пяти единиц и поэтому $|g_j j\,(u)| \leqslant 10^{-3/2} \cdot 1 \cdot 5 \leqslant 1/6 = c$. Положим $x = \varepsilon\sqrt{\varphi\,(\varepsilon c)}/2$. При $l \geqslant 30$ неравенство $2\left(e_l\left(-x^2\right) + R_l\,(n, c, \varepsilon)\right) \leqslant \alpha$ верно, если $x = 3$, как нетрудно проверить (упражнение). Следовательно, $\varepsilon = 2x/\sqrt{\varphi\,(\varepsilon c)} = 6$ $(\varphi\,(6/6) = \varphi\,(1) = 1)$, $\varepsilon b/\sqrt{n} = 6/\sqrt{1000} < 0.2$. Отклонение $|\sum f_j/n - a| \geqslant 0.2$ практически невозможно. Неравенство Чебышева гарантировало практическую невозможность только отклонения $|\sum f_j/n - a| \geqslant 1$.

Если $b^2 = 0.01$, $\alpha = 0.01$ и $\Delta = \varepsilon b/\sqrt{n} = 0.1$, то при $l \geqslant 30$ и $c = 1/6$ неравенство $2(e_l(-x^2) + R_l(n, c, \varepsilon)) \leqslant \alpha$ верно, если $x \geqslant 2.5$. Следовательно, допустимая граница отклонений $|\sum f_j(u)/n - a|$ в теореме 1.12.3 будет определяться числом $\varepsilon = 2x/\sqrt{\varphi(\varepsilon c)} = 5$ $(\varphi(\varepsilon c) = \varphi(5/6) = 1)$. Число измерений, обеспечивающее практическую невозможность отклонения $|\sum f_j/n - a| \geqslant \Delta$, находится из неравенства $\Delta \geqslant \varepsilon b/\sqrt{n}$, эквивалентного $n \geqslant \varepsilon^2 b^2/\Delta^2$. В рассматриваемом примере $n \geqslant 25 \cdot 0.01/0.01 = 25$. Неравенство Чебышева требовало $n \geqslant 100$.

Точно так же, используя следствие 1.12.4, можно улучшить оценку для числа подбрасываний, полученную в 1.10.4 (упражнение).

2. Дискретные вероятностные пространства

Дискретным называется вероятностное пространство со счетным множеством исходов. Конечные множества включаются в класс счетных, и конечные вероятностные пространства являются дискретными. Бесконечные дискретные пространства не намного сложнее конечных.

2.1. Элементарная вероятность

Элементарная вероятность является основным понятием для дискретных пространств.

2.1.1. Пусть U — счетное множество и p — положительная нормированная функция на U:

$$p\,(u) \geqslant 0, \qquad \sum p\,(u) = 1 \qquad (u \in U)\,.$$

Вместе они составляют *дискретное вероятностное пространство* (U, p). Если множество U бесконечно, то сумма $\sum p\,(u)$ имеет бесконечно много слагаемых и определяется как предел семейства конечных сумм ([4]).

Элементы $u \in U$ по-прежнему называются *исходами*, функция p — *элементарной вероятностью*, число $p\,(u)$ — *вероятностью исхода* u. Если $p\,(u) > 0$ для всех $u \in U$, то вероятность p и пространство (U, p) называют *невырожденными*.

Так как элементы каждого счетного множества можно снабдить номерами, то удобно в качестве множеств исходов выбирать конечные или бесконечные множества номеров. Особенно часто выбирают множества $\mathbb{N} = \{1, 2, \ldots\}$ или $\mathbb{P} = \{0, 1, 2, \ldots\}$.

2.1.2. Рассмотрим несколько примеров. Во всех примерах $p\,(u) > 0$ следует из определения и проверяется только $\sum p\,(u) = 1$.

Пример 1. *Геометрическое распределение* с параметром $a \in]0,1[$ описывается элементарной вероятностью p на множестве $\mathbb{P}$, имеющей значения

$$p(n) = a(1-a)^n.$$

По формуле для суммы геометрической прогрессии

$$\sum p(n) = a \sum (1-a)^n = a(1-(1-a))^{-1} = aa^{-1} = 1.$$

Пример 2. *Распределение Паскаля* с параметрами $a \in]0,1[$, $m \in \mathbb{N}$ описывается элементарной вероятностью p на множестве $\mathbb{P}$, имеющей значения

$$p(n) = \binom{n+m-1}{n} a^m (1-a)^n = p(n,m,a).$$

Заметим, что

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

По общей биномиальной формуле ([8])

$$\sum p(n) = a^m \sum_n \binom{-m}{n} (-(1-a))^n = a^m (1-(1-a))^{-m} = 1.$$

В связи с биномиальными коэффициентами для $-m$ распределение Паскаля называют также *отрицательным биномиальным распределением*. В отличие от обычного (*положительного*) биномиального распределения отрицательное определено на бесконечном множестве исходов.

Распределение Паскаля можно описывать с помощью биномиальных коэффициентов. Так как

$$\binom{n+m-1}{n} = \binom{n+m-1}{m-1},$$

то

$$p\,(n, m, a) = ab\,(n, n+m-1, 1-a) = ab\,(m-1, n+m-1, a)\,.$$

Заметим еще, что $p\,(n, 1, a) = a\,(1-a)^n$: при $m = 1$ распределение Паскаля превращается в геометрическое.

Пример 3. *Логарифмическое распределение* с параметром $a \in]0, 1[$ описывается элементарной вероятностью p на множестве $\mathbb{N}$, имеющей значения

$$p\,(n) = -\,(1-a)^n \;/\; (n \log a)\,.$$

Используя разложение функции $\log\,(1+x)$ в степенной ряд, при $x = -\,(1-a)$ получаем

$$\sum p\,(n) = -\sum \left((1-a)^n\,/n\right) / \log a = \log a / \log a = 1.$$

Пример 4. *Распределение Пуассона* с параметром $a > 0$ описывается элементарной вероятностью p на множестве $\mathbb{P}$, имеющей значения

$$p\,(n) = (a^n/n!)\,e^{-a} = p\,(n, a)\,.$$

Используя разложение функции e^x в степенной ряд, при $x = a$ получаем

$$\sum p\,(n) = e^{-a} \sum (a^n/n!) = e^{-a} e^a = 1.$$

Распределение Пуассона играет выдающуюся роль в теории вероятностей. Оно применяется при решении самых разных теоретических и прикладных задач.

Пример 5. *Распределение Гольдбаха* описывается элементарной вероятностью p на множестве натуральных чисел

$$M = \{m = l^k : l > 1, k > 1\},$$

имеющей значения

$$p\,(m) = 1/\,(m-1)\,.$$

Можно доказать (упражнение; см. [4]), что $\sum p\,(m) = 1$. Суммы по специальным множествам, подобные этой, рассматриваются в теории чисел.

2.2. Вероятность и среднее

Для дискретного вероятностного пространства вероятности и средние определяются так же, как для конечного. Существенная разница между конечными и бесконечными пространствами заключается в том, что на бесконечных пространствах не все случайные переменные имеют средние значения и дисперсии.

2.2.1. Пусть (U, p) — дискретное вероятностное пространство и $A \subseteq U$ — *событие*. Так как p положительна и нормирована, то

$$0 \leqslant P\,(A) = \sum_{u \in A} p\,(u) \leqslant \sum_{u \in U} p\,(u) = 1.$$

Число $P\,(A)$ называется *вероятностью* события A. Вероятность P определена на классе $\mathcal{P} = \mathcal{P}\,(U)$ всех частей множества U.

Для любого множества индексов M независимость событий $A_j \subseteq U$, $j \in M$ означает, что

$$P\left(\bigcap A_j\right) = \prod P(A_j) \qquad\qquad (j \in K)$$

для каждого конечного $K \subseteq M$.

2.2.2. Каждая (вещественная) функция на множестве U называется *случайной переменной* или, коротко, просто *переменной*.

Событие A часто отождествляется с переменной $\operatorname{ind} A$. Это отождествление позволяет использовать удобные записи для переменной f на U:

$$f = \sum f(u) \cdot u \quad (u \in U), \quad f = \sum x \cdot f^{-1}(x) \quad (x \in X = f(U)).$$

Рассмотрим семейство $f = (f_j)$ переменных f_j $(j \in M)$ на пространстве (U, p), их маргинальные распределения q_j на множествах $X_j = f_j(U)$ и совместные распределения q_K для $j \in K$ и каждого конечного $K \subseteq M$. Если равенство

$$q_K(x) = \prod q_j(x_j)$$

верно для всех $x = (x_j) \in X$ $(j \in K)$, то q_K называют *произведением* q_j и пишут $q_K = \prod q_j$. Если $q_K = \prod q_j$ $(j \in K)$ для всех конечных $K \subseteq M$, то говорят, что случайные переменные f_j (*стохастически*) *независимы*, а если $q_K \neq \prod q_j$, — что (*стохастически*) *зависимы*.

Независимость событий $A_j \subseteq U$ эквивалентна независимости их индикаторов $f_j = \operatorname{ind} A_j$ $(j \in M)$.

2.2.3. Для каждой переменной f на пространстве (U, p) и каждого конечного множества $K \subseteq U$ определена конечная сумма

$$E(f, K) = \sum_{u \in K} f(u)\, p(u).$$

Если семейство этих конечных сумм ограничено (и только в этом случае), существует сумма

$$E(f) = \sum f(u)\, p(u),$$

равная пределу $E(f, K)$ при $K \to U$. Это значит ([4]), что для каждого $\varepsilon > 0$ существует конечное множество $K(\varepsilon) \subseteq U$ такое, что неравенство

$$|E(f, K) - E(f)| < \varepsilon$$

верно для всех конечных множеств $K \subseteq U$, содержащих $K(\varepsilon)$. В этом случае переменную f называют *суммируемой* по p.

Число $E(f)$ называется *средним значением* суммируемой по p переменной f на дискретном пространстве (U, p). Из определений следует, что

$$E(\operatorname{ind} A) = P(A)$$

для каждого $A \subseteq U$.

Каждая ограниченная переменная f суммируема по p: если $|f(u)| \leqslant c$ при всех $u \in U$, то

$$\left|\sum_{u \in K} f(u)\, p(u)\right| \leqslant \sum_{u \in K} |f(u)|\, p(u) \leqslant cP(K) \leqslant c$$

для каждого конечного множества $K \subseteq U$ и семейство $(f(u)\, p(u))$ суммируемо. В частности, каждая постоянная на U суммируема по p.

2.2.4. Рассмотрим несколько примеров вычисления средних. Всюду будет рассматриваться *тождественная переменная* f со значениями $f(n) = n$ для $n \in U = \mathbb{P}$ или $n \in U = \mathbb{N}$. Среднее $E(f)$ тождественной переменной f является центром для распределения масс $p(n)$ в точках n вещественной прямой $\mathbb{R}$.

Пример 1. При геометрическом распределении с параметром a $(0 < a < 1)$

$$E(f) = \sum_{n \geqslant 1} na(1-a)^n = (1-a)/a.$$

В самом деле, как легко проверить по индукции (упражнение), для конечных сумм $S(n) = S(K(n))$, $K(n) = \{m : 0 \leqslant m < n\}$ верны равенства

$$\begin{aligned} S(n) = \sum_{0\leqslant m<n} ma(1-a)^m = \\ = a(1-a)\left(\frac{1-(1-a)^n}{(1-(1-a))^2} - \frac{n(1-a)^{n-1}}{1-(1-a)}\right) = \\ = \frac{1-a}{a}\left(1-(1-a)^n - na(1-a)^{n-1}\right). \end{aligned}$$

Поэтому

$$E(f) = \lim_{n\to\infty} S(n) = \frac{1-a}{a}.$$

Пример 2. При распределении Паскаля с параметрами m, a ($m \in \mathbb{N}$, $0 < a < 1$)

$$E(f) = \sum_{n\geqslant 1} np(n, m, a) = m(1-a)/a.$$

В самом деле, из определения $p(n, m, a)$ следует, что

$$\begin{aligned} & np(n, m, a) = m((1-a)/a) \cdot p(n-1, m+1, a), \\ & \sum_{n\geqslant 1} p(n-1, m+1, a) = 1. \end{aligned}$$

Пример 3. При логарифмическом распределении с параметром a $(0 < a < 1)$

$$E(f) = -\sum_{n\geqslant 1} (1-a)^n / \log a = (1-a)/(a\log a).$$

Пример 4. При распределении Пуассона с параметром $a > 0$

$$E(f) = \sum_{n\geqslant 1} np(n, a) = a\sum_{n\geqslant 1} p(n-1, a) = a,$$

так как

$$np\,(n,a) = n\frac{a^n}{n!}e^{-a} = a\frac{a^{n-1}}{(n-1)!}e^{-a} = ap\,(n-1,a)\,,$$
$$\sum_{n\geqslant 1} p\,(n-1,a) = \sum_{n\geqslant 0} p\,(n,a) = 1.$$

Пример 5. При распределении Гольдбаха тождественная переменная не имеет конечного среднего значения:

$$\sum_{m\in M} m/\,(m-1) = \infty,$$

так как $m/\,(m-1) \geqslant 1$. Другими словами, распределение Гольдбаха не имеет конечного центра.

2.3. Свойства среднего

Суммы суммируемых семейств обладают теми же свойствами, что и конечные суммы. Поэтому средние для дискретных пространств обладают теми же свойствами, что и для конечных. Нужно только проверять суммируемость рассматриваемых переменных.

2.3.1. Рассмотрим множество $\mathcal{F} = \mathcal{F}\,(U)$ всех переменных на дискретном пространстве (U,p) и выделим множество $\mathcal{L} = \mathcal{L}\,(U,p)$ суммируемых по p. Сумма и произведение на число суммируемых переменных тоже суммируемы. Произведение суммируемых переменных может быть несуммируемо.

Среднее E является *функционалом* на $\mathcal{L}$: каждой суммируемой переменной ставится в соответствие число Ef. Про суммируемые переменные говорят также, что они *имеют среднее значение.*

В общем случае суммируемые переменные на (U,p) образуют векторное пространство, но не образуют алгебры.

Контрпример. Возьмем $s = 1 + \sigma$, $\sigma > 0$ и рассмотрим последовательность $(1/n^s)$, $n \geqslant 1$. Она суммируема. В самом деле,

$$\frac{1}{(n+1)^s} + \cdots + \frac{1}{(2n)^s} \leqslant n \cdot \frac{1}{n^s} = \frac{1}{n^\sigma},$$
$$\frac{1}{(2^{k-1}+1)^s} + \cdots + \frac{1}{(2^k)^s} \leqslant n \cdot \frac{1}{(2^{k-1})^\sigma} = \frac{1}{(2^\sigma)^{k-1}},$$

откуда следует, что конечные суммы последовательности $(1/n^s)$ ограничены числом

$$c = 1 + \frac{1}{2^s} + \frac{1/2^\sigma}{1 - 1/2^\sigma}.$$

Равенство

$$\zeta(s) = \sum \frac{1}{n^s}$$

определяет знаменитую *ζ-функцию Римана*, которая играет важную роль в теории чисел. Известно, что $\zeta(2) = \pi^2/6$.

Пусть $U = \mathbb{N}$, $p(n) = 6/(\pi n)^2$ и $f(n) = \sqrt{n}$. Тогда

$$E(f) = \left(6/\pi^2\right) \sum \frac{1}{n^{3/2}} = \frac{6}{\pi^2}\,\zeta(\frac{3}{2}),$$
$$E\left(f^2\right) = (6/\pi)^2 \sum \frac{1}{n} = \infty.$$

Неограниченность семейства конечных сумм гармонического ряда $(1/n)$ нетрудно проверить (упражнение), разбивая их на двоичные отрезки и оценивая получаемые частичные суммы числом $1/2$ снизу каждую.

Рассмотрим еще переменную $g = f + c$, где c — постоянная. Произведение $fg = f^2 + cf$ несуммируемо вместе с f^2.

2.3.2. Основные свойства среднего E можно выразить фразой: E является *нормированным положительным линейным фун-*

кционалом на пространстве $\mathcal{L}$ суммируемых переменных. Это значит, что для каждых $f, g \in \mathcal{L}$ и $c \in \mathbb{R}$ верны равенства:

$$\begin{aligned} &E(f+g) = E(f) + E(g), && E(cf) = cE(f), \\ &E(f) \geqslant 0 \qquad (f \geqslant 0), && E(1) = 1. \end{aligned}$$

Эти равенства проверяются так же, как в конечном случае (1.3.1). Из линейности и положительности среднего следует его монотонность:

$$E(f) \leqslant E(g) \qquad\qquad (f \leqslant g).$$

Добавим еще одно важное свойство среднего E: его *счетную аддитивность.*

Рассмотрим последовательность суммируемых по p переменных f_n на пространстве (U, p). Предположим, что последовательность значений $f_n(u)$ для каждого $u \in U$ имеет сумму $f(u)$ и что последовательность средних значений $E(f_n)$ тоже имеет некоторую сумму. Счетная аддитивность среднего E означает, что переменная $f = \sum f_n$ суммируема по p и верно равенство

$$E(f) = \sum E(f_n).$$

Это равенство легко проверить, используя возможность изменять порядок суммирования для суммируемых семейств ([4]). Ограниченность семейства конечных сумм для $(E(f_n))$ влечет суммируемость семейств $(f_n(u)\, p(u))$, $(f(u)\, p(u))$ и обеспечивает равенства

$$\begin{aligned} Ef = \sum_u f(u)\, p(u) = \sum_u \sum_n f_n(u)\, p(u) = \\ = \sum_n \sum_u f_n(u)\, p(u) = \sum_n E(f_n). \end{aligned}$$

Из счетной аддитивности среднего E следует счетная аддитивность вероятности P. Рассмотрим последовательность попарно непересекающихся событий $A_n \subseteq U$ и их объединение A. Индикаторы $f_n = \operatorname{ind} A_n$ и $f = \operatorname{ind} A$ удовлетворяют условиям счетной аддитивности: $Ef_n = P(A_n)$, $\sum\limits_{n \in K} P(A_n) = P\left(\bigcup\limits_{n \in K} A_n\right) \leqslant 1$ для каждого конечного $K \subseteq \mathbb{N}$, $f(u) = \sum f_n(u) = 1$ при $u \in A$ и $f(u) = 0$ при $u \notin A$. Следовательно

$$P(A) = E(\operatorname{ind} A) = \sum E(\operatorname{ind} A_n) = \sum P(A_n).$$

Счетная аддитивность среднего E и вероятности P для бесконечных пространств — свойство существенно более сильное, чем конечная аддитивность. При использовании счетной аддитивности вероятности P надо помнить об условии $A_j A_k = \varnothing$ $(j \neq k)$. В общем случае нужно переходить к среднему E.

2.3.3. Линейность среднего часто позволяет упростить его вычисление.

Рассмотрим множество $U = \mathbb{P}^m$ *мультиномеров* $u = (u_i)$, составленных из номеров $u_i \in \mathbb{P}$ $(i = 1, \dots, m)$. Назовем число $|u| = u_1 + \dots + u_m$ *нормой* мультиномера u. Обозначим U_n множество мультиномеров с нормой n. Множества U_n конечны, и поэтому их объединение $U = \cup U_n$ счетно ([4]). Ясно, что U_k, U_n при $k \neq n$ не имеют общих элементов. Пусть $0 < a < 1$. Равенство

$$p(u) = a^m (1 - a)^{|u|}$$

определяет элементарную вероятность p на множестве U, причем $p(U_n) = b(n, m, a)$.

Каждая координата $f_i(u) = u_i$ мультиномера u имеет геометрическое распределение

$$q_i(n) = P\{u : u_i = n\} = \sum_{u_j : j \neq i} a^{m-1}(1-a)^{\sum u_j} \cdot a(1-a)^n = \\ = a(1-a)^n.$$

Элементарная вероятность p равна произведению q_i и является элементарным совместным распределением переменных f_i.

Сумма $|f| = \sum f_i$ имеет распределение Паскаля: так как множество $u_n = \{u : |u| = n\}$ состоит из $\binom{n+m-1}{m}$ элементов ([8]), то $P(|f| = n) = P(U_n) = b(n, m, a)$. Зная это, можно сразу выписать среднее значение

$$E(|f|) = \sum_{1 \leqslant i \leqslant m} E(f_i) = m(1-a)/a$$

в соответствии с 2.2.4.

Замечание. Распределение Паскаля применяют для описания числа нулей, появившихся до m-й по порядку единицы в бесконечной последовательности Бернулли: $b(n, m, a)$ есть вероятность того, что $s_{j(1)} = \ldots = s_{j(m)} = 1$ при некоторых номерах $j(1) < \ldots < j(m) = n + m$ и $s_j = 0$ при остальных $j < n + m$. Так как число n случайно и может принимать любые целые положительные значения, то нельзя для описания взять какое-нибудь пространство $B(l, a)$ заранее. А множество бесконечных двоичных последовательностей несчетно, и дискретная модель непосредственно неприменима. Переход к распределениям позволяет обойти эту трудность.

2.4. Дисперсия

Дисперсия служит мерой отклонений значений случайной переменной от ее среднего значения.

2.4.1. Выделим в алгебре $\mathcal{L} = \mathcal{L}(U, p)$ суммируемых по p переменных на дискретном пространстве (U, p) подалгебру $\mathcal{L}^2 = \mathcal{L}^2(U, p)$ квадратично суммируемых по p. По определению, переменная f на U *квадратично суммируема* по p, если переменная f^2 суммируема по p. Другими словами, $f \in \mathcal{L}^2$ означает, что f^2 имеет среднее значение. Так как $f^2 \geqslant 0$, то это записывают неравенством $E\left(f^2\right) < \infty$.

На бесконечных дискретных пространствах существуют переменные, которые суммируемы, но не квадратично суммируемы (контрпример в 2.3.1).

Если f^2 суммируема по p, то и f суммируема по p. В само деле, $|f(u)\, p(u)| \leqslant p(u)$ при $f(u) \leqslant 1$ и $|f(u)\, p(u)| \leqslant f^2(u)\, p(u)$ при $f(u) > 1$. Следовательно

$$|f(u)\, p(u)| \leqslant f^2(u)\, p(u) + p(u)$$

для каждого $u \in U$. Так как $\sum p(u) = 1$, то из ограниченности конечных сумм для f^2 следует их ограниченность для f.

2.4.2. Число

$$D(f) = E\left((f - Ef)^2\right)$$

называется *дисперсией* квадратично суммируемой по p переменной f на пространстве (U, p). Для проверки суммируемости выражения $(f - Ef)^2$ используется равенство

$$(f - Ef)^2 = f^2 - 2f \cdot Ef + (Ef)^2,$$

суммируемость по p постоянных и суммируемость f, обеспечиваемая предполагаемой суммируемостью f^2.

Свойства дисперсии D определяются свойствами среднего E:

$$D(f) \geqslant 0, \qquad D(c) = 0;$$
$$D(f + c) = D(f), \qquad D(cf) = c^2 D(f);$$
$$D(f + g) = D(f) + D(g) - 2C(f, g),$$

где

$$C(f,g) = E((f - Ef)(g - Eg)) = E(fg) - Ef \cdot Eg.$$

Кроме того

$$D(f) = E\left(f^2\right) - (Ef)^2.$$

Эти равенства проверяются так же, как в конечном случае (1.4.2 – 1.4.3). Число $C(f,g)$ называется *корреляцией* переменных f, g.

Равенство

$$D(f+g) = D(f) + D(g)$$

эквивалентно $C(f,g) = 0$.

Число

$$S(f) = \sqrt{D(f)} \geqslant 0$$

называется *стандартным отклонением* переменной f. Или, коротко, ее *стандартом*. Его используют в качестве единицы масштаба для измерения отклонений значений переменной f от ее среднего значения $E(f)$.

2.4.3. Рассмотрим несколько примеров вычисления дисперсий. Будем использовать тождественную переменную f и распределения p на $\mathbb{P}$ или $\mathbb{N}$, взятые из примеров в 2.2.4.

Пример 1. При геометрическом распределении с параметром a $(0 < a < 1)$

$$D(f) = (1-a)/a^2.$$

В самом деле, используя формулу суммирования по частям

([4]) и равенство $E(f) = (1-a)/a$ из 2.2.4, получаем:

$$E\left(f^2\right) = \sum_{n\geqslant 0} n^2 a (1-a)^n = a \lim_{n\to\infty}\left(\sum_{0\leqslant m<n} m^2 (1-a)^m\right) =$$
$$= a\lim_{n\to\infty}\left(-\frac{n^2(1-a)^n}{a} + \sum_{0\leqslant m<n}\frac{(1-a)^{m+1}}{a}(2m+1)\right) =$$
$$= 2(1-a)\sum_{m\geqslant 0} m(1-a)^m + (1-a)\sum_{m\geqslant 0}(1-a)^m =$$
$$= 2(1-a)^2/a^2 + (1-a)/a,$$
$$(Ef)^2 = (1-a)^2/a^2,$$
$$D(f) = E\left(f^2\right) - (Ef)^2 = (1-a)/a^2.$$

Пример 2. Для распределения Паскаля с параметрами m, a ($m > 0$ целое, $0 < a < 1$) верны равенства

$$n(n-1)\,p(n,m,a) = (m+1)\,m\frac{(1-a)^2}{a^2}p(n-2,m+2,a)\,,$$
$$\sum_{n\geqslant 0} n(n-1)\,p(n,m,a) = (m+1)\,m\frac{(1-a)^2}{a^2}\times$$
$$\times\sum_{n\geqslant 2} p(n-2,m+2,a) = (m+1)\,m\left((1-a)^2/a^2\right),$$

откуда
$$E\left(f^2\right) = (m+1)\,m\left((1-a)^2/a^2\right) + Ef.$$

А так как $Ef = m(1-a)/a$, то

$$D(f) = E\left(f^2\right) - (Ef)^2 = m(1-a)/a^2.$$

Пример 3. При логарифмическом распределении с параметром a $(0 < a < 1)$

$$D(f) = -\frac{1-a}{a^2 \log a}\left(1 + \frac{1-a}{\log a}\right).$$

В самом деле,

$$E\left(f^2\right) = -\sum_{n\geqslant 1} n^2 (1-a)^n / (n \log a) =$$
$$= -\sum_{n\geqslant 1} n (1-a)^n / \log a = -(1-a) / \left(a^2 \log a\right),$$
$$(Ef)^2 = (1-a)^2 / (a \log a)^2,$$

и формула $D(f) = E\left(f^2\right) - (Ef)^2$ дает нужный результат.

Пример 4. При распределении Пуассона с параметром $a > 0$

$$D(f) = a.$$

В самом деле,

$$E\left(f^2\right) = \sum_{n\geqslant 1} n^2 \frac{a^n}{n!} e^{-a} = \sum_{n\geqslant 1} n(n-1)\frac{a^n}{n!} e^{-a} + \sum_{n\geqslant 1} n \frac{a^n}{n!} e^{-a} =$$
$$= a^2 e^{-a} \sum_{n\geqslant 2} \frac{a^{n-2}}{(n-2)!} + Ef = a^2 + a,$$
$$D(f) = E\left(f^2\right) - (Ef)^2 = a^2 + a - a^2 = a.$$

При распределении Пуассона дисперсия и среднее значение равны параметру a.

2.5. Корреляционная теория

Геометрический язык позволяет изложить эту теорию для дискретных пространств так же, как для конечных, используя квадратичную суммируемость рассматриваемых переменных.

2.5.1. Рассмотрим векторное пространство $\mathcal{L}^2 = \mathcal{L}^2(U, p)$ квадратично суммируемых по p переменных на невырожденном дискретном пространстве (U, p). Каждая квадратично суммируемая переменная суммируема. Поэтому $\mathcal{L}^2$ является подпространством пространства $\mathcal{L} = \mathcal{L}\,(U, p)$ суммируемых переменных на (U, p).

Если переменные f, g на U квадратично суммируемы по p, то их произведение суммируемо по p. Это следует из неравенства

$$|fg| \leqslant \left(f^2 + g^2\right)/2,$$

которое обеспечивает ограниченность конечных сумм семейства $f\,(u)\,g\,(u)\,p\,(u)\ \ (u \in U)$. Суммируемость произведения fg позволяет определить скалярное произведение

$$\langle f, g\rangle = E\,(fg)$$

для каждых квадратично суммируемых переменных f, g и сделать $\mathcal{L}^2$ евклидовым пространством. Свойства скалярного произведения проверяются так же, как для конечного пространства (1.5.1). Скалярное произведение симметрично, линейно по каждой переменной, положительно на диагонали и невырождено при невырожденной элементарной вероятности p (не имеющей нулевых значений, что предполагается).

2.5.2. Квадратично суммируемые переменные с нулевым средним значением образуют подпространство $\mathcal{L}^2_0 = \mathcal{L}^2_0(U, p)$ евклидова пространства $\mathcal{L}^2 = \mathcal{L}^2(U, p)$. Каждой переменной $f \in \mathcal{L}^2$ соответствует центрированная переменная $f_0 = f - Ef \in \mathcal{L}^2_0$. Из определений следует, что

$$D\left(f_0\right)=E\left(f_0^2\right)=\left\langle f_0, f_0\right\rangle=\left\|f_0\right\|^2,$$
$$S\left(f_0\right)=\sqrt{D\left(f_0\right)}=\sqrt{\left\langle f_0, f_0\right\rangle}=\left\|f_0\right\|$$

для каждой $f \in \mathcal{L}^2$. Дисперсия центрированной переменной равна квадрату ее евклидовой длины, а стандартное отклонение — ее длине в пространстве $\mathcal{L}_0^2$.

Скалярное произведение позволяет измерять зависимость между квадратично суммируемыми переменными как величину угла между векторами в евклидовом пространстве. Число

$$K(f, g)=\frac{C(f, g)}{S(f)\, S(g)}=\frac{E(f g)-E(f)\, E(g)}{\sqrt{D(f)} \sqrt{D(g)}}$$

называется *коэффициентом корреляции* между квадратично суммируемыми переменными $f \neq 0$, $g \neq 0$. (Так как по предположению элементарная вероятность p невырожденная, то $Df \neq 0$, $Dg \neq 0$) Из определений следует, что

$$K(f, g)=\left\langle\overline{f_0}, \overline{g_0}\right\rangle,$$

где

$$\overline{f_0}=(f-E f) /\|f-E f\|, \qquad \overline{g_0}=(g-E g) /\|g-E g\| .$$

Для каждых векторов f, g евклидова пространства $\mathcal{L}^2$ верно неравенство Коши

$$|\langle f, g\rangle| \leqslant\|f\|\|g\|$$

и, следовательно,

$$K(f, g) \leqslant\left\|\overline{f_0}\right\|\left\|\overline{g_0}\right\|=1.$$

Коэффициент корреляции симметричен:

$$K(f, g)=K(g, f) .$$

Если $K(f,g) = 0$, то переменные f, g называются *некоррелированными* или *ортогональными*. Равенство $K(f,g) = 1$ означает линейную зависимость между f и g (1.5.5). В общем случае дискретных пространств для коэффициента корреляции верны все равенства, выписанные в 1.5.7.

2.6. Распределения

Переход к распределению позволяет отвлечься от специфики рассматриваемой переменной и сосредоточиться на вероятностной стороне дела. Все сказанное по этому поводу для конечных пространств в 1.6 сохраняет силу для дискретных.

2.6.1. Переменная f на дискретном пространстве (U,p) определяет новое пространство (X,q) с множеством исходов $X = f(U)$ и элементарной вероятностью $q = Pf^{-1}$, имеющей значения

$$q(x) = P\left(f^{-1}(x)\right) = \sum_{u:f(u)=x} p(u).$$

Элементарная вероятность $q = Pf^{-1}$ называется *элементарным распределением* переменной f, а определяемая q вероятность $Q = Pf^{-1}$ — ее *распределением*:

$$Q(B) = P(A) = P\left(f^{-1}(B)\right) \qquad \left(A = f^{-1}(B) \subseteq U, B \subseteq X\right).$$

Замена $u \in U$ на $x = f(u) \in X$ дает удобные формулы для среднего значения и дисперсии (когда они существуют):

$$E(f) = \sum xq(x), \qquad D(f) = \sum (x - Ef)^2 q(x).$$

Они получаются группировкой исходов u, для которых $f(u) = x$.

В примерах 1 – 5 пункта 2.2.4 элементарные вероятности p являлись распределениями для тождественной переменной f: всюду $U = X$, $u = f(u) = f(x) = x$, $q(x) = P\left(f^{-1}(x)\right) = P(\{x\}) = p(x)$. В примерах 1 – 4 пункта 2.4.3 для переменных $g = f^2$ всюду $X = \{x = u^2 : u \in U\}$, $g^{-1}(x) = \sqrt{x}$, $q(x) = p(\sqrt{x})$.

2.6.2. Совместные распределения переменных на дискретных пространствах определяются так же как для переменных на конечных пространствах (1.6.4). Не умаляя общности (элементы конечного множества всегда можно снабдить номерами), вместо произвольных индексов при обозначении семейства $f = (f_j)$ переменных на дискретном пространстве (U, p) будем рассматривать номера $j = 1, \ldots, m$ и писать также $f = (f_1, \ldots, f_m)$.

Значение $f(u) = (f_1(u), \ldots, f_m(u))$ для каждого $u \in U$ есть вектор в m-мерном вещественном пространстве $\mathbb{R}^m$. Семейство f определяет множество

$$X = f(U) = \{x = f(u) : u \in U\}$$

векторов $x = (x_j)$ с координатами $x_j = f_j(u)$. Множество $X \subseteq \mathbb{R}^m$ счетно вместе с U ([4]).

Равенство

$$q(x) = P\left(f^{-1}(x)\right) = P\{u : f_j(u) = x_j, 1 \leqslant j \leqslant m\}$$

Определяет элементарную вероятность $q = P^{-1}$ на множестве X, которая называется *элементарным совместным распределением* переменных f_j. Определяемая q вероятность $Q = Pf^{-1}$ для событий $B \subseteq X$ называется *совместным распределением* переменных f_j.

Элементарные распределения $q_j = Pf_j^{-1}$, а также распределения $Q_j = Pf_j^{-1}$ для переменных f_j называются *маргинальными*. Они определены для множеств исходов $X_j = f_j(U) \subseteq \mathbb{R}$.

В 2.3.3 описана связь между геометрическим распределением и распределением Паскаля. Там $U = X = \mathbb{P}^m$, $p(u) = a^m (1-a)^{|u|}$ и $f_i(u) = u_i$ для $u = (u_i)$ $(i = 1, \ldots, m)$, откуда следует, что $f(u) = (f_1(u), \ldots, f_m(u)) = (u_1, \ldots, u_m) = u$ и f является тождественной переменной. Элементарная вероятность p является одновременно совместным распределением для $f : q = p$. Маргинальные распределения q_i геометрические и определены на $X_i = \mathbb{P}$:

$$q_i(n) = P(u : u_i = n) = a(1-a)^n$$

для каждого $n \in \mathbb{P}$.

2.6.3. Рассмотрим векторную переменную $f = (f_1, \ldots, f_n)$, координаты которой f_j $(j = 1, \ldots, n)$ являются числовыми переменными на пространстве (U, p), их маргинальные распределения q_j на множествах $X_j = f_j(U) \subseteq \mathbb{R}$ и совместное распределение q на множестве $X = f(U) \subseteq \mathbb{R}^n$. Если $q = q_1 \cdots q_n$, то говорят, что переменные $f_1, \ldots, f_n$ *(стохастически) независимы*, а если $q \neq q_1 \cdots q_n$, — что *(стохастически) зависимы*.

Для независимых переменных $f_1, \ldots, f_n$, имеющих средние и дисперсии, верны равенства

$$E(f_1 \cdots f_n) = E(f_1) \cdots E(f_1),$$
$$D(f_1 + \cdots + f_n) = D(f_1) + \cdots + D(f_n).$$

Все, сказанное в 1.6.11 – 1.6.14 о независимых событиях, коэффициентах связи и регрессии для конечных вероятностных пространств верно и для дискретных. Нужно только всюду проверять суммируемость.

2.6.4. Рассмотрим пространство (U, p) с множеством исходов $U = \mathbb{P}^n$ и элементарной вероятностью $p(u) = a^m (1-a)^{|u|}$, описанное в 2.3.3. Из определений следует, что переменные $f_i(u) = u_i$ независимы:

$$p(u) = a^m (1-a)^{\sum u_i} = \prod a(1-a)^{u_i} = \prod q_i.$$

Поэтому кроме повторенного в 2.3.3 равенства для среднего значения суммы $|f| = \sum f_i$ можно сразу выписать и ее дисперсию:

$$D(|f|) = \sum_{1 \leqslant i \leqslant m} D(f_i) = m(1-a)/a^2$$

в соответствии с 2.4.3.

2.7. Информация и энтропия

Для дискретного пространства (U, p) элементарную информацию $i(f, g)$ переменных f, g можно определить так же, как и для конечных пространств в 1.7.1, равенством

$$i(f, g) = \log \frac{s(f(u), g(u))}{q(f(u))\, r(g(u))} \qquad (u \in U).$$

Здесь q, r, s маргинальные и совместное распределения переменных f, g. Но так как среднее значение

$$I(f, g) = E(i(f, g)),$$

выражающее взаимную информацию f и g, существует в дискретных пространствах не всегда, то использовать теперь информацию в качестве меры зависимости можно только с оговорками.

Кроме того, на бесконечном счетном множестве X не существует равномерного распределения, обеспечивающего максимальную энтропию, как в случае конечного X.

Преодоление всех этих трудностей выходит за рамки элементарной теории. Отметим только, что вместо $I(f, g)$ можно рассматривать семейство информаций $I_k(f_k, g_k)$ для сужений переменных f, g на конечные множества $K \subseteq U$ и различные характеристики этого семейства.

Упражнение. Уточнить и проверить информационный критерий независимости для дискретных пространств (1.7.4).

2.8. Условные средние и вероятности

В дискретных пространствах условия могут описываться бесконечными разбиениями. Нужно только всюду проверять суммируемость. Счетная аддитивность позволяет сказанное об условных средних в 1.8 для конечных пространств обобщить на дискретные.

2.8.1. Рассмотрим невырожденное дискретное пространство (U, p) и разбиение $\mathcal{B} = (B_j)$ множества U:

$$\sum B_j = U, \qquad B_j \neq \varnothing, \qquad B_j B_k = \varnothing \quad (j \neq k)\,.$$

Разбиение $\mathcal{B}$ может быть конечным или бесконечным. Пусть $\mathcal{G} = \mathcal{GL}(\mathcal{B})$ — подпространство векторного пространства $\mathcal{L} = \mathcal{L}(U, p)$ суммируемых по p переменных на U, составленное из суммируемых по p переменных $g = \sum y_j B_j$.

Условным средним переменной f при условии $\mathcal{B}$ называется переменная

$$E\,(f \mid G) = \sum E\,(f \mid B_j)\, B_j.$$

Переменная $g = E\,(f \mid G)$ принадлежит $\mathcal{G}$ и имеет то же среднее значение, что и f. В самом деле, благодаря счетной аддитивности и линейности среднего E,

$$\begin{aligned} E\,(g) &= \sum E\,(E\,(f \mid B_j)\, B_j) = \\ &= \sum E\,(f \mid B_j)\, E\,(B_j) = \sum \big(E\,(f B_j) \;/\; P\,(B_j)\big)\, P\,(B_j) = \\ &= \sum E\,(f B_j) = E\left(f \sum B_j\right) = E\,(f)\,. \end{aligned}$$

Каждая ограниченная переменная f на U имеет условное среднее при любом $\mathcal{B}$.

Переменная

$$P\,(A \mid \mathcal{B}) = E\,(A \mid \mathcal{B})$$

называется *условной вероятностью* события $A \subseteq U$.

Если разбиение $\mathcal{B}$ составлено из прообразов $B_z = h^{-1}\,(z)$ значений z переменной h на U, то вместо $E\,(f \mid \mathcal{B})$ пишут также $E\,(f \mid h)$:

$$\begin{aligned} &E\,(f \mid h) = \sum E\,(f \mid h = z)\, h^{-1}\,(z), \\ &E\,(f \mid h = z) = E\,\big(f \cdot h^{-1}\,(z)\big) \;/\; P\,\big(h^{-1}\,(z)\big)\,. \end{aligned}$$

Для события $f = A$ это равенство превращается в

$$P(f \mid h) = \sum P(A \mid h = z)\,(h^{-1}(z).$$

Из определений следует, что

$$E(f \mid h) = f$$

для независимых переменных f, h.

Условная независимость переменных и событий определяется так же, как в 1.8.1.

2.8.2. Условное среднее для суммируемых переменных на дискретных пространствах имеет те же свойства, что и для переменных на конечных пространствах:

$$E(f \mid \mathcal{B}) = f \qquad (f \in \mathcal{G}); \qquad E(E(f \mid \mathcal{B}) \mid \mathcal{B}) = E(f \mid \mathcal{B});$$
$$E(E(f \mid \mathcal{C}) \mid \mathcal{B}) = E(f \mid \mathcal{B}),$$

если разбиение $\mathcal{C}$ мельче разбиения $\mathcal{B}$;

$$E(f_1 + f_2 \mid \mathcal{B}) = E(f_1 \mid \mathcal{B}) + E(f_2 \mid \mathcal{B}), \quad E(hf \mid \mathcal{B}) = hE(f \mid \mathcal{B})$$

при $f, f_1, f_2 \in \mathcal{L}$ и $h \in \mathcal{G}$;

$$E(f \mid \mathcal{B}) \geqslant 0 \qquad\qquad (f \in \mathcal{L}, \quad f \geqslant 0).$$

Условное среднее $E(\cdot \mid \mathcal{B})$ есть положительный линейный оператор, проектирующий $\mathcal{L}$ на $\mathcal{GL}(\mathcal{B})$.

Сказанное в 1.8.5 – 1.8.7 о механических системах, коэффициентах связи и регрессии, правиле умножения для вероятностей верно и для дискретных пространств.

2.8.3. Рассмотрим вероятностное пространство (U, p) с множеством $U = \sum (\mathbb{B}^n \times \{n\})$, элементами которого являются пары (u, n), составленные из двоичных строк $u = u_1 \dots u_n \in \mathbb{B}^n$ и чисел $n \in \mathbb{P}$, а элементарная вероятность p имеет значения

$$p(u, n) = a^{|u|} (1-a)^{n-|u|} p(n, \lambda),$$

где

$$|u| = u_1 + \dots + u_n, \qquad p(n, \lambda) = e^{-\lambda} \lambda^n / n!.$$

Пусть

$$f(u, n) = |u|, \qquad h(u, n) = n.$$

Тогда

$$P(h = n) = p(n, \lambda), \qquad P(f = m \mid h = n) = b(m, n, a).$$

Найдем распределение переменной f и ее среднее. Имеем:

$$P(f = m) = \sum_{n \geqslant 0} P(f = m \mid h = n) P(h = n) =$$
$$= \sum_{n \geqslant 0} b(m, n, a) p(n, \lambda) = \frac{(\lambda a)^m}{m!} e^{-\lambda} \sum_{n \geqslant m} \frac{(\lambda (1-a))^{n-m}}{(n-m)!} =$$
$$= \frac{(\lambda a)^m}{m!} e^{-\lambda} e^{\lambda(1-a)} = e^{-\lambda a} (\lambda a)^m / m! = p(m, \lambda a).$$

Переменная f имеет распределение Пуассона с параметром λa.

Так как

$$E(f \cdot (h = n)) = \sum_m m\, b(m, n, a) p(n, \lambda) = na \cdot p(n, \lambda),$$

то

$$
\begin{aligned}
&E\,(f \mid h=n) = E\,(f \cdot (h=n))\,/P\,(h=n) = \\
&\qquad\qquad\qquad\qquad = na \cdot p\,(n,\lambda)\,/p\,(n,\lambda) = na, \\
&E\,(f \mid h) = \sum E\,(f \mid h=n)\,(h=n) = \sum na \cdot (h=n), \\
&E\,(f) = E\,(E\,(f \mid h)) = E\left(\sum na \cdot (h=n)\right) = \\
&\qquad\qquad = \sum na \cdot P\,(h=n) = a \sum np\,(n,\lambda) = \lambda a
\end{aligned}
$$

в соответствии с пуассоновским распределением переменной f.

Переменная f равна сумме, составленной из случайного числа двоичных слагаемых, которое имеет пуассоновское распределение. Такие суммы встречаются в самых разных задачах.

2.8.4. Геометрическое распределение обладает важным свойством: у него нет *последействия*.

Пусть переменная f имеет геометрическое распределение с параметром a и m, n — целые положительные числа. Тогда

$$
P\,(f \geqslant m+n \mid f \geqslant m) = P\,(f \geqslant n)\,.
$$

В само деле, так как $f \geqslant m+n \Rightarrow f \geqslant m$, то

$$
\begin{aligned}
&P\,(f \geqslant m+n, f \geqslant m) = P\,(f \geqslant m+n) = \\
&\qquad = \sum_{l \leqslant m+n} a\,(1-a)^l = (1-a)^{m+n} = P\,(f \geqslant m) \cdot P\,(f \geqslant n)\,.
\end{aligned}
$$

Поэтому

$$
P\,(f \geqslant m+n \mid f \geqslant m) = P\,(f \geqslant m+n) \Big/ P\,(f \geqslant m) = P\,(f \geqslant n)\,.
$$

2.8.5. Определение условного среднего, данное в 2.8.1, можно пояснить, используя геометрический язык. Для этого рассмотрим невырожденное евклидово пространство $\mathcal{L}^2 = \mathcal{L}^2(U,p)$ квадратично суммируемых переменных на U и его подпространство $\mathcal{G}^2 = \mathcal{X}\mathcal{L}^2(\mathcal{B})$, составленное из квадратично суммируемых переменных $g = \sum y_j B_j$. Квадратичная суммируемость g означает, что

$$E\left(g^2\right) = E\left(\sum y_j^2 B_j\right) = \sum E\left(y_j^2 B_j\right) = \sum y_j^2 P\left(B_j\right) < \infty.$$

Переменные

$$b_j = B_j / \sqrt{P\left(B_j\right)}$$

образуют ортонормированную базу подпространства $\mathcal{G}^2$, которая конечна или бесконечна вместе с разбиением $\mathcal{B}$. Условное среднее $g = E(f \mid \mathcal{B})$ является ортогональной проекцией переменной $f \in \mathcal{L}^2$ на подпространство $\mathcal{G}^2$ и определяется равенством

$$g = \sum \langle f, b_j \rangle b_j,$$

которое эквивалентно

$$g = \sum y_j B_j, \qquad y_j = E\left(f B_j\right) / P\left(B_j\right).$$

Но обобщение на дискретные пространства сказанного в 1.8.10 для конечных пространств выходит за рамки элементарной теории: геометрия бесконечномерных пространств сложнее геометрии конечномерных. Вместе с тем пользоваться общими свойствами евклидова пространства $\mathcal{L}^2$ часто бывает удобно.

2.9. Формулы полной вероятности и Байеса

Формулы полной вероятности и Байеса, выписанные в 1.9, без изменений обобщаются на дискретные пространства. Нужно только иметь в виду, что суммы в этих формулах могут состоять из бесконечного числа слагаемых и надо проверять суммируемость.

2.9.1. Рассмотрим невырожденное дискретное пространство (U, p) и разбиение $\mathcal{B} = (B_j)$ множества U. Для каждого события $A \subseteq U$ верна *формула полной вероятности*

$$P(A) = \sum P(A \mid B_j)\, P(B_j).$$

Чтобы ее получить, достаточно взять среднее E от обеих частей равенства

$$E(A \mid B) = \sum E(A \mid B_j)\, B_j.$$

В самом деле, используя счетную аддитивность E и определения условных вероятностей, получаем:

$$P(A) = E(A) = X(E(A \mid \mathcal{B})) = E\left(\sum E(A \mid B_j)\, B_j\right) = \\ = \sum E(A \mid B_j)\, E(B_j) = \sum P(A \mid B_j)\, P(B_j).$$

С помощью формулы полной вероятности решаются задачи такого же типа, который был описан в 1.9.1. Для бесконечных разбиений эта формула применялась в 2.8.3 и 2.8.4.

2.9.2. Формула Байеса сразу следует из определения условной вероятности для событий и формулы полной вероятности. Запись сумм без явного указания области суммирования позволяет формулу Байеса вместе с формулой полной вероятности без изменений перенести с конечного случая на дискретный:

$$P(B_j \mid A) = P(A \mid B_j)\, P(B_j) \Big/ \sum P(A \mid B_k)\, P(B_k).$$

Задачи, решаемые с помощью формулы Байеса, схематически описаны в 1.9.3. Если вероятность $P(A)$ известна, то вместо формулы Байеса можно использовать равенство

$$P(B_j \mid A) = P(A \mid B_j)\, P(B_j) \Big/ P(A).$$

2.9.3. Для примера рассмотрим композицию биномиального и пуассоновского распределений, описанную в 2.8.3, и вычислим вероятность $P(h = n \mid f = m)$ того, что вторым элементом пары (u, n) является n, если известно, что в строке u ровно m единиц.

В 2.8.3 было доказано, что

$$P(f = m \mid h = n) = b(m, n, a), \qquad P(h = n) = p(n, \lambda),$$
$$P(f = m) = p(m, \lambda a).$$

Следовательно,

$$P(h = n \mid f = m) = P(f = m \mid h = n) P(h = n) \big/ P(f = m) =$$
$$= \frac{n!}{m!\,(n-m)!} a^m (1-a)^{n-m} e^{-\lambda} \frac{\lambda^n}{n!} e^{\lambda a} \frac{m!}{(\lambda a)^m} =$$
$$= e^{-\lambda(1-a)} \frac{(\lambda(1-a))^{n-m}}{(n-m)!} = p(n-m, \lambda(1-a)).$$

Заметим, что распределение Пуассона $p(n, \lambda)$ с параметром λ имеет наибольшее значение при $n = [\lambda]$. В самом деле, неравенства $p(n, \lambda) \geqslant p(n-1, \lambda)$, $p(n, \lambda) > p(n+1, \lambda)$ эквивалентны неравенствам $\lambda - 1 < n \leqslant \lambda$. Таким образом, наиболее вероятно, что вторым элементом пары (u, n), где в строке u ровно m единиц, является $n = m + [\lambda(1-a)]$.

2.10. Усиленный закон больших чисел

Неравенство Чебышева и следующий из него закон больших чисел, описанные в 1.10 для конечных пространств, верны и для дискретных. Более сильные результаты можно получить с помощью неравенства Колмогорова. Для бесконечных дискретных пространств можно сформулировать законы больших чисел Чебышева и Бернулли в предельной форме.

Из неравенства Чебышева следует, что

$$P\{u : \left| n^{-1} \sum f_j(u) - a \right| \geqslant \varepsilon\} \to 0 \qquad (n \to \infty).$$

Это — закон больших чисел Чебышева в предельной форме. Эквивалентная форма:

$$P\{u : \left|n^{-1} \sum f_j(u) - a\right| < \varepsilon\} \to 1 \qquad (n \to \infty).$$

Из неравенства Бернулли следует, что при каждом $\varepsilon > 0$

$$P\{u : \left|n^{-1} \sum s_j(u) - a\right| \geqslant \varepsilon\} \to 0 \qquad (n \to \infty).$$

Это — закон больших чисел Бернулли в предельной форме. Эквивалентная форма:

$$P\{u : \left|n^{-1} \sum s_j(u) - a\right| < \varepsilon\} \to 1 \qquad (n \to \infty).$$

2.10.1. Рассмотрим евклидово пространство $\mathcal{L}_0^2 = \mathcal{L}_0^2(U, p)$ квадратично суммируемых переменных с нулевыми средними. По определению

$$\langle f, g\rangle = E(fg), \qquad \|f\|^2 = E\left(f^2\right) = D(f) \qquad \left(f, g \in \mathcal{L}_0^2\right).$$

Пусть f_j $(1 \leqslant j \leqslant n)$ — попарно ортогональные (некоррелированные) переменные из $\mathcal{L}_0^2$ и

$$g_k = \sum_{1 \leqslant j \leqslant k} f_j, \qquad h_k = \sum_{k < j \leqslant n} f_j \qquad (h_n = 0),$$

$$g = \max_{1 \leqslant k \leqslant n} |g_k|, \qquad \sigma^2 = D(g_n) = \sum_{1 \leqslant j \leqslant n} D(f_j).$$

Возьмем еще число $\varepsilon > 0$ и рассмотрим событие

$$A = (g \geqslant \varepsilon) = \{u : g(u) \geqslant \varepsilon\}.$$

Докажем неравенство

$$P(A) \leqslant \sigma^0/\varepsilon^2.$$

Для каждого $k = 1, \ldots, n$ положим:

$$A_k = \{u : |g_k(u)| \geqslant \varepsilon, |g_j(u)| < \varepsilon \ \ (1 \leqslant j < k)\}.$$

Ясно, что

$$A_k A_l = \varnothing \quad (k \neq l), \qquad A = \sum A_k;$$
$$g_n = g_k + h_k, \qquad A_k g_n = A_k g_k + A_k h_k.$$

Из попарной ортогональности f_j следует ортогональность $A_k g_k$, $A_k h_k$:

$$E(A_k g_k \cdot A_k h_k) = 0, \qquad A_k g_k \perp A_k h_k.$$

Используя теорему Пифагора, линейность и монотонность среднего E, получаем:

$$\|A_k g_n\|^2 = \|A_k g_k\|^2 + \|A_k h_k\|^2 \geqslant \|A_k g_k\|^2 = \\ = E\left(A_k g_k^2\right) \geqslant E\left(\varepsilon^1 A_k\right) = \varepsilon^2 P(A_k).$$

А так как f_j попарно ортогональны, то благодаря этому неравенству

$$\sum_j D(f_j) = \sum_j \|f_j\|^2 = \|g_n\|^2 = E\left(g_n^2\right) \geqslant E\left(A g_n^2\right) = \\ = E\left(\sum_k A_k g_n^2\right) = \sum_k E\left(A_k g_n^2\right) \geqslant \varepsilon^2 \sum_k P(A_k) = \varepsilon^2 P(A).$$

Отсюда сразу следует доказываемое неравенство.

Перепишем его в исходных обозначениях:

$$P\left(\max_{1 \leqslant k \leqslant n} \left|\sum_{1 \leqslant j \leqslant k} f_j\right| \geqslant \varepsilon\right) \leqslant \sigma^2/\varepsilon^2. \tag{K1}$$

Это и есть *неравенство Колмогорова*. При $n = 1$ и $f_1 = f - Ef$ из него получается неравенство Чебышева.

Можно рассматривать попарно некоррелированные переменные f_j^0 с произвольными средними значениями $E\left(f_j^0\right) = a_j$. Тогда неравенство Колмогорова нужно применять к центрированным переменным $f_j = f_j^0 - a_j$, имеющим те же дисперсии:

$$P\left(\max_{1\leqslant k\leqslant n}\left|\sum_{1\leqslant j\leqslant k}\left(f_j^0 - a_j\right)\right| \geqslant \varepsilon\right) \leqslant \sigma^2/\varepsilon^2.$$

2.10.2. Доказанное неравенство (K1) оценивает вероятность $P(A)$ сверху. Другое неравенство Колмогорова оценивает $P(A)$ снизу. Предположим дополнительно, что рассматриваемые переменные f_j равномерно ограничены постоянной $c > 0$: $|f_j(u)| \leqslant c$ для всех $j = 1, \ldots, n$ и $u \in U$. Тогда верно неравенство

$$P\left(\max_{1\leqslant k\leqslant n}\left|\sum_{1\leqslant j\leqslant k} f_j\right| \geqslant \varepsilon\right) \geqslant 1 - (c+\varepsilon)^2/\sigma^2. \qquad \text{(K2)}$$

В самом деле, так как $A + A' = U$, $AA' = \varnothing$, то $Ag_n + A'g_n = g_n$ и $Ag_n \perp A'g_n$. Поэтому

$$\begin{aligned}
\|g_n\|^2 &= \left\|Ag_n + A'g_n\right\|^2 = \|Ag_n\|^2 + \left\|A'g_n\right\|^2, \\
\|Ag_n\|^2 &= \|g_n\|^2 - \left\|A'g_n\right\|^2 \geqslant \|g_n\|^2 - \varepsilon^2 P\left(A'\right) = \\
&= \|g_n\|^2 - \varepsilon^2 + \varepsilon^2 P(A).
\end{aligned}$$

Вместе с тем по определению множества A_k верно неравенство $|A_k g_{k-1}| < \varepsilon$ и поэтому

$$|A_k g_k| = |A_k(g_{k-1} + f_k)| \leqslant |A_k g_{k-1}| + |A_k f_k| \leqslant \varepsilon + c.$$

А так как A_k определяется переменными f_j с номерами $j \leqslant k$, а h_k — переменными f_j с номерами $j > k$, то $A_k \perp h_k$ и

$$\|A_k h_k\|^2 = E\left(a_k h_k^2\right) = E\left(A_k\right) E\left(h_k^2\right) = P\left(A\right) \left\|h_k^2\right\|.$$

Кроме того, $A_k \perp A_l$ при $k \neq l$ и

$$A = \sum A_k, \qquad \|h_k\|^2 = \|g_n\|^2 - \|g_k\|^2 \leqslant \|g_n\|^2.$$

Используя все это, получаем:

$$\begin{aligned}\|Ag_n\|^2 &= \left\|\sum_k A_k g_n\right\|^2 = \sum_k \|A_k g_n\|^2 = \sum_k \|A_k g_k + A_k h_k\|^2 = \\ &= \sum_k \|A_k g_k\|^2 + \sum_k \|A_k h_k\|^2 = \sum_k E\left(A_k g_k^2\right) + \sum_k E\left(A_k h_k^2\right) \leqslant \\ &\leqslant (\varepsilon + c)^2 \sum_k P\left(A_k\right) + \|g_n\|^2 \sum_k P\left(A_k\right) = \\ &= \left((\varepsilon + c)^2 + \|g_n\|^2\right) P\left(A\right).\end{aligned}$$

Сравнивая полученные для $\|Ag_n\|^2$ неравенства, находим:

$$\|g_n\|^2 - \varepsilon^2 + \varepsilon^2 P\left(A\right) \leqslant \left((\varepsilon + c)^2 + \|g_n\|^2\right) P\left(A\right),$$

откуда

$$\begin{aligned}P\left(A\right) &\geqslant \frac{\|g_n\|^2 - \varepsilon^2}{\|g_n\|^2 - \varepsilon^2 + (\varepsilon + c)^2} = 1 - \frac{(\varepsilon + c)^2}{\|g_n\|^2 - \varepsilon^2 + (\varepsilon + c)^2} \geqslant \\ &\geqslant 1 - (\varepsilon + c)^2 \big/ \|g_n\|^2.\end{aligned}$$

Неравенство (К2) доказано $\left(\|g_n\|^2 = \sigma^2\right)$.

Это неравенство успешно применяется в теоретических рассуждениях.

2.10.3. По определению

$$\sigma = \sqrt{D\left(g_n\right)} = \sqrt{\sum_{1\leqslant j\leqslant n} D\left(f_j\right)}$$

есть стандарт для суммы $g_n = \sum\limits_{1\leqslant j\leqslant n} f_j$, измеряющий ее отклонение от среднего значения $E\left(g_n\right) = 0$. Его удобно использовать в качестве единицы масштаба для измерения этого отклонения.

Пусть $t > 0$ и $\varepsilon = t\sigma$. Тогда неравенство Колмогорова (K1) имеет вид:

$$P\left(\max_{1\leqslant k\leqslant n}|g_k| \geqslant t\sigma\right) \leqslant 1/t^2.$$

Его можно переписать в эквивалентной форме:

$$P\left(\max_{1\leqslant k\leqslant n}|g_k| < t\sigma\right) \geqslant 1 - 1/t^2.$$

В частности

$$P\left(\max_{1\leqslant k\leqslant n}|g_k| < 3\sigma\right) \geqslant 0.88.$$

Отклонения, меньшие 3σ, можно считать удовлетворительными, а отклонения, большие 3σ, — маловероятными.

Подчеркнем, что неравенство $\max\limits_{1\leqslant k\leqslant n}|g_k| < \varepsilon$ означает выполнение всех неравенств $|g_1|<\varepsilon, \ldots, |g_k|<\varepsilon$. А неравенство $\max\limits_{1\leqslant k\leqslant n}|g_k|\geqslant\varepsilon$ — выполнение хотя бы одного из неравенств $|g_1| \geqslant \varepsilon, \ldots, |g_k| \geqslant \varepsilon$.

2.10.4. Рассмотрим последовательность Бернулли с параметрами n, a. Пусть как и прежде переменная s_j описывает появление единицы на j-м месте: $s_j\left(u\right) = u_j$ для $u = u_1 \ldots u_n$ и $j = 1, \ldots, n$. Тогда $E\left(s_j\right) = a$ и $D\left(s_j\right) = a\left(1 - a\right)$. Введем кроме обычной частоты еще *взвешенные частоты* появления единиц:

$$\nu\left(k, n\right) = \left(kn^{-1}\right) k^{-1} \sum_{1\leqslant j\leqslant k} s_j = n^{-1} \sum_{1\leqslant j\leqslant k} s_j \quad \left(k = 1, \ldots, n\right).$$

Из неравенства Колмогорова следует *усиленное неравенство Бернулли*

$$P\left(\max_{1\leqslant k\leqslant n}|\nu\,(k,n)-a|\geqslant\varepsilon\right)\leqslant a\,(1-a)\,/\,\left(n\varepsilon^2\right).$$

Оно оценивает вероятность того, что хотя бы одно отклонение $|\nu(k,n)-a|\ \geqslant\ \varepsilon$ взвешенной частоты $\nu\,(k,n)$ от вероятности a больше ε. (В частности, и отклонения $|\nu\,(k,n)-a|\geqslant\varepsilon$ для частоты $\nu\,(n,n)$, вероятность которого оценивает обычное неравенство Бернулли из 1.10.3.)

Пусть $\alpha>0$ и $n\,(\varepsilon)=\left[a\,(1-a)\,/\,\left(\alpha\varepsilon^2\right)\right]$ — целая часть числа $a\,(1-a)\,/\,\left(\alpha\varepsilon^2\right)$. Из усиленного неравенства Бернулли следует, что при всех $n>n\,(\varepsilon)$ вероятность отклонения $|\nu\,(k,n)-a|\geqslant\varepsilon$ для хотя бы одного k меньше α.

Из усиленного неравенства Бернулли следует, что при каждом $\varepsilon>0$

$$P\left(\max_{1\leqslant k\leqslant n}|\nu\,(k,n)-a|\geqslant\varepsilon\right)\to 0\qquad\qquad (n\to\infty)\,.$$

Это — *усиленный закон больших чисел* в форме Бернулли. Он уточняет связь между частотой появления данного события в последовательности независимых испытаний, произведенных в одинаковых условиях, и вероятностью этого события. Вероятность события служит мерой его реализуемости.

3. Непрерывные вероятностные пространства

В этих пространствах множеством исходов служит множество $U = \mathbb{R}$ вещественных чисел, а роль элементарной вероятности играет положительная нормированная интегрируемая функция $p: U \to \mathbb{R}$, называемая плотностью. Элементарным такое пространство предлагается называть в связи с использованием элементарного математического аппарата.

Основная цель главы — показать аналогию между дискретным и непрерывным случаями. Поэтому она написана бегло. Общий случай подробно рассматривается в [7]. Он опирается на общую теорию меры и интеграла.

3.1. Плотность

Плотность является основным понятием элементарного вероятностного пространства.

3.1.1. Будем называть *плотностью* каждую положительную интегрируемую функцию $p : U \to \mathbb{R}$ такую, что

$$\int p(u)\, du = 1 \qquad\qquad (u \in U)\,.$$

Множество $U = \mathbb{R}$ вещественных чисел и плотность $p : U \to \mathbb{R}$ составляют *элементарное вероятностное пространство* (U, p). Если $p(u) > 0$ для всех $u \in U$, то плотность p и пространство (U, p) называют *невырожденными*.

Выделяются ступенчатые плотности, тесно связанные с элементарными вероятностями.

Рассмотрим последовательность чисел $a_n \in \mathbb{R}$ и последовательность попарно непересекающихся ограниченных интервалов

$X_n \subseteq \mathbb{R}$. Они определяют на $\mathbb{R}$ *ступенчатую* функцию

$$f = \sum a_n X_n, \quad f(u) = a_n \ (u \in X_n), \quad f(u) = 0 \ \left(u \notin \sum X_n\right).$$

Ступенчатая функция не имеет сложных разрывов и может быть интегрируема.

Обозначим $\Delta(X_n)$ длину интервала X_n. Если ряд $(a_n \Delta(X_n))$ суммируем, то функция f интегрируема и

$$\int f = \sum a_n \Delta(X_n).$$

Если

$$a_n \geqslant 0, \qquad \sum a_n \Delta(X_n) = 1,$$

то функция f является плотностью.

Выпуклую комбинацию $\alpha p + \beta q$ ($\alpha \geqslant 0$, $\beta \geqslant 0$, $\alpha + \beta = 1$) плотности p и плотности q условимся называть их *смесью*.

3.1.2. Рассмотрим примеры. Для ступенчатых плотностей они являются модификациями примеров 1 – 4 дискретных распределений из 2.1.2.

Пример 1. *Ступенчатое геометрическое распределение* с параметром $a \in]0,1[$ описывается плотностью p, имеющей значения

$$p(u) = a(1-a)^n \quad (u \in [n, n+1[), \qquad p(u) = 0 \quad (u < 0),$$

и интеграл

$$\int p = \sum p(n) = a \sum (1-a)^n = 1.$$

Пример 2. *Ступенчатое распределение Паскаля* с параметрами $a \in]0,1[$, $m \in \mathbb{N}$ описывается плотностью p, имеющей значения

$$p(u) = \binom{n+m-1}{n} a^m (1-a)^n \ (u \in [n, n+1[), \quad p(u) = 0 \ (u < 0),$$

и интеграл

$$\int p = \sum p(n) = a^m \sum_n \binom{-m}{n} (-(1-a))^n = 1.$$

Пример 3. *Ступенчатое логарифмическое распределение* с параметром $a \in]0,1[$ описывается плотностью p, имеющей значения

$$p(u) = -(1-a)^n \big/ (n \log a) \quad (u \in [n, n+1[), \quad p(u) = 0 \quad (u < 1),$$

и интеграл

$$\int p = \sum p(n) = -\sum ((1-a)^n / n) / \log a = \log a / \log a = 1.$$

Пример 4. *Ступенчатое распределение Пуассона* с параметром $a > 0$ описывается плотностью p, имеющей значения

$$p(u) = (a^n / n!)\, e^{-a} \quad (u \in [n, n+1[)\,, \qquad p(u) = 0 \quad (u < 0)\,,$$

и интеграл

$$\int p = \sum p(n) = e^{-a} \sum (a^n / n!) = e^{-a} e^{a} = 1.$$

Пример 5. *Равномерное распределение* на отрезке $[a,b] \subseteq \mathbb{R}$ $(a < b)$ задается плотностью p, имеющей значения

$$p(u) = \frac{1}{(b-a)} \quad (u \in [a,b])\,, \qquad p(u) = 0 \quad (u \notin [a,b])\,.$$

Данное распределение также ступенчатое. Описывается равномерное распределение единицы массы на отрезке $[a,b]$.

Пример 6. *Треугольное распределение* на отрезке $[a,b] \subseteq \mathbb{R}$ $(a < b)$ задается плотностью p, имеющей значения

$$\begin{aligned} p(u) &= \frac{4}{(b-a)} \left(1 - \frac{1}{(b-a)} \left| \frac{a+b}{2} - u \right| \right) && (u \in [a,b])\,, \\ p(u) &= 0 && (u \notin [a,b])\,. \end{aligned}$$

Описывается симметричное относительно точки 0 распределение единицы массы на отрезке $[a, b]$, уплотняющееся к середине. Название объясняется видом графика плотности. Треугольное распределение называется также *распределением Симпсона*.

Пример 7. *Экспоненциальное* (*показательное*) *распределение* с параметром α задается плотностью p, имеющей значения

$$p\,(u) = \alpha e^{-\alpha u} \quad (u \geqslant 0)\,, \qquad\qquad p\,(u) = 0 \quad (u < 0)\,.$$

Из свойств экспоненты и правил интегрирования следует, что $p\,(u) > 0$ $(u \geqslant 0)$ и

$$\alpha \int\limits_0^\infty e^{-\alpha u} du = \int\limits_0^\infty e^{-x} dx = -\,(0-1) = 1.$$

Таким образом описывается распределение единицы массы на интервале $[0, \infty[$. Его плотность быстро убывает при удалении от начала и тем быстрее, чем больше параметр α. Показательное распределение является непрерывным аналогом геометрического распределения.

Пример 8. *Нормальное распределение* с параметрами $a \in \mathbb{R}$, $\sigma > 0$ задается плотностью $p = g$, имеющей значения

$$g\,(u) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-a)^2}{2\sigma^2}} \qquad\qquad (-\infty < u < \infty)\,.$$

В приложении доказано, что

$$g\,(u) > 0 \quad (-\infty < u < \infty)\,, \qquad\qquad \int\limits_{-\infty}^{\infty} g\,(u) du = 1$$

и исследованы свойства функции g.

Описывается симметричное относительно точки a распределение единицы массы на всей вещественной прямой, уклоняющееся к a тем быстрее, чем меньше параметр σ.

Среди нормальных распределений выделяется *стандартное*, с параметрами $a = 0$, $\sigma = 1$. Оно имеет плотность

$$g(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \qquad (-\infty < x < \infty).$$

Нормальное распределение играет исключительную роль в теории вероятностей. Его называют также *распределением Гаусса*. Нормальная плотность g является гладкой на $\mathbb{R}$ функцией.

Пример 9. Пусть $\alpha \geqslant 0$, $\beta \geqslant 0$, $\alpha + \beta = 1$ и $\lambda > 0$, $\mu > 0$. Смесь экспоненциальных распределений с параметрами λ, μ задается плотностью p, имеющей значения

$$p(u) = \alpha\lambda e^{-\lambda u} + \beta\mu e^{-\mu u} \quad (u \geqslant 0), \qquad p(u) = 0 \quad (u < 0).$$

Рассматриваются и смеси нескольких экспоненциальных распределений.

Пример 9. *Распределение Коши* задается плотностью $p = g$, имеющей значения

$$p(u) = \frac{1}{\pi} \frac{1}{1 + u^2} \qquad (-\infty < u < \infty).$$

Применяя формулу Ньютона-Лейбница, получаем

$$\frac{1}{\pi} \int\limits_{-\infty}^{\infty} \frac{dx}{1 + x^2} = \frac{1}{\pi} (\arctan(\infty) - \arctan(-\infty)) = 1.$$

Рассматривается также распределение Коши с параметрами $a \in \mathbb{R}$ и $\lambda > 0$, задаваемое плотностью p со значениями

$$p(u) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (u - a)^2} \qquad (-\infty < u < \infty).$$

Это распределение симметрично относительно точки a, и плотность p имеет в ней наибольшее значение $p(a) = 1/(\pi\lambda)$.

3.2. Вероятность и среднее

Для непрерывного вероятностного пространства вероятности и средние определяются так же, как для дискретного. Нужно только заменить элементарную вероятность плотностью, а суммы — интегралами.

3.2.1. Рассмотрим непрерывное вероятностное пространство (U, p). Каждое множество $A \subseteq U$ для индикатора которого существует интеграл $\int Ap$, назовем *случайным событием* или просто *событием*. Пусть $A \subseteq U$ — событие. Так как p положительна и нормирована, то

$$0 \leqslant P(A) = \int Ap \leqslant \int p = 1.$$

Число $P(A)$ называется *вероятностью* события A.

Каждую функцию $f : U \to \mathbb{R}$, не имеющую сложных разрывов, будем называть *случайной переменной* или просто *переменной*. Условимся в дальнейшем рассматривать только те события, индикаторы которых являются случайными переменными. Примерами таких событий служат интервалы вещественной прямой и их конечные объединения.

3.2.2. Будем говорить, что переменная $f : U \to \mathbb{R}$ интегрируема по плотности $p : U \to \mathbb{R}$, если существует интеграл

$$E(f) = \int fp.$$

Каждая ограниченная переменная f интегрируема по p.

Число $E(f)$ называется *средним значением* интегрируемой переменной f на элементарном пространстве (U, p). Из определений следует, что

$$E(A) = E(\operatorname{ind} A) = P(A)$$

для каждого события $A \subseteq U$, отождествленного с $\operatorname{ind} A$.

3.2.3. Рассмотрим несколько примеров вычисления средних. Всюду будет рассматриваться *тождественная переменная* f со значениями $f(u) = u$ для $u \in U = \mathbb{R}$. Среднее $E(f)$ тождественной переменной f является центром для распределения на прямой $\mathbb{R}$ единицы массы с плотностью p.

Примеры 1 – 4. Средние для ступенчатых геометрического распределения, распределения Паскаля, логарифмического распределения и распределения Пуассона вычисляются с помощью результатов в 2.2.4 для дискретных распределений. Нужно только добавлять $1/2$, получающуюся за счет интегрирования по интервалам $[n, n+1[$:

$$\int\limits_{-\infty}^{\infty} p(u)\,du = \sum p(n) \int\limits_{n}^{n+1} udu = \sum p(n)\, \frac{(n+1)^2 - n^2}{2} =$$

$$= \sum \left(np(n) + \frac{1}{2} p(n) \right) = \sum np(n) + 1/2.$$

Это — эффект равномерного распределения массы $p(n)$ в точке n по интервалу $[n, n+1[$.

Подчеркнем, что по условию всюду $f(u) = u$ $(u \in \mathbb{R})$.

Пример 5. Для равномерного распределения на отрезке $[a, b]$ получаем

$$E(f) = \frac{1}{b-a} \int\limits_{a}^{b} udu = \frac{a+b}{2}.$$

Пример 6. Для треугольного распределения на отрезке $[a, b]$ имеем

$$E(f) = \frac{1}{3(b-a)^2} \left(a^3 + b^3 - 2\left(\frac{a+b}{2} \right)^3 \right).$$

Пример 7. Для показательного распределения с параметром α интегрирование по частям дает

$$E(f) = 1/\alpha.$$

Пример 7. Для нормального распределения с параметрами α, σ среднее есть $E(f) = \alpha$.

Среднее для смеси равно смеси средних. А для распределения Коши среднего не существует.

3.3. Свойства среднего

Так как средние значения переменных определены интегралами, то их свойства определяются свойствами интегралов.

3.3.1. Рассмотрим множество $\mathcal{F} = \mathcal{F}(U)$ всех переменных на непрерывном пространстве (U, p) и выделим из них множество $\mathcal{L} = \mathcal{L}(U, p)$ интегрируемых по p. Сумма и произведение на число интегрируемых переменных тоже интегрируемы. Произведение интегрируемых переменных может быть неинтегрируемо.

Среднее E является *функционалом* на $\mathcal{L}$: каждой интегрируемой переменной ставится в соответствие число Ef. Про интегрируемые переменные говорят также, что они *имеют среднее значение.*

В общем случае интегрируемые переменные на (U, p) образуют векторное пространство, но не образуют алгебры.

Основные свойства среднего E можно выразить фразой: E является *нормированным положительным линейным функционалом* на пространстве $\mathcal{L}$ интегрируемых переменных. Это значит, что для каждых $f, g \in \mathcal{L}$ и $c \in \mathbb{R}$ верны равенства:

$$
\begin{aligned}
&E(f+g) = E(f) + E(g), && E(cf) = cE(f), \\
&E(f) \geqslant 0 \qquad (f \geqslant 0), && E(1) = 1.
\end{aligned}
$$

Эти равенства проверяются так же, как в дискретном случае. Из линейности и положительности среднего следует его монотонность:

$$
E(f) \leqslant E(g) \qquad\qquad (f \leqslant g).
$$

3.3.2. Добавим еще одно важное свойство среднего E: его *секвенциальную непрерывность* при равномерной сходимости. Из сказанного в приложении об интегралах и перестановке пределов следует, что для каждой последовательности интегрируемых переменных $f_n \in \mathcal{L}$, равномерно сходящейся к переменной $f \in \mathcal{L}$, верно равенство

$$E(f) = \lim_{n \to \infty} E(f_n).$$

Секвенциальная непрерывность среднего E связана со счетной аддитивностью вероятности P, но для нее нужно использовать другую сходимость.

3.4. Дисперсия

Дисперсия служит мерой отклонений значений случайной переменной от ее среднего значения.

3.4.1. Выделим в алгебре $\mathcal{L} = \mathcal{L}(U, p)$ интегрируемых по p переменных на непрерывном пространстве (U, p) подалгебру $\mathcal{L}^2 = \mathcal{L}^2(U, p)$ квадратично интегрируемых по p. По определению, переменная f на U *квадратично интегрируема* по p, то есть $f \in \mathcal{L}^2$, если $f \in \mathcal{L}$ и $f^2 \in \mathcal{L}$.

Число

$$D(f) = E\left((f - Ef)^2\right)$$

называется *дисперсией* квадратично интегрируемой по p переменной f на пространстве (U, p). Для проверки интегрируемости $(f - Ef)^2$ используется равенство

$$(f - Ef)^2 = f^2 - 2f \cdot Ef + (Ef)^2,$$

интегрируемость по p постоянных и интегрируемость f и f^2.

Свойства дисперсии D определяются свойствами среднего E:

$$\begin{aligned} &D(f) \geqslant 0, && D(c) = 0; \\ &D(f+c) = D(f), && D(cf) = c^2 D(f); \\ &D(f+g) = D(f) + D(g) - 2C(f,g), && \end{aligned}$$

где

$$C(f,g) = E((f - Ef)(g - Eg)) = E(fg) - Ef \cdot Eg.$$

Кроме того
$$D(f) = E\left(f^2\right) - (Ef)^2.$$

Эти равенства проверяются так же, как в дискретном случае. Число $C(f,g)$ называется *ковариацией* переменных f, g. Интегрируемость произведения fg обеспечивается интегрируемостью f^2, g^2 и равенством

$$fg = \frac{1}{2}\left((f+g)^2 - f^2 - g^2\right).$$

Равенство
$$D(f+g) = D(f) + D(g)$$
эквивалентно $C(f,g) = 0$.

Число
$$S(f) = \sqrt{D(f)} \geqslant 0$$
называется *стандартным отклонением* переменной f или, коротко, ее *стандартом*. Его используют в качестве единицы масштаба для измерения отклонений значений переменной f от ее среднего значения $E(f)$.

3.4.2. Рассмотрим несколько примеров вычисления дисперсий. Будем использовать тождественную переменную f и результаты 3.2.2.

Примеры 1 – 4. Дисперсии для рассматривавшихся в 3.2.2 ступенчатых распределений вычисляются с помощью результатов в 2.4.3 для дискретных распределений. Нужно только сделать поправки, связанные с интегралами

$$\int_{n}^{n+1} (u - E(f))^2 \, p(u) \, du.$$

Пример 5. Для равномерного распределения на отрезке $[a, b]$ имеем:

$$D(f) = \frac{1}{12} (b - a)^2 .$$

Пример 6. Для треугольного распределения на отрезке $[a, b]$ имеем:

$$D(f) = \frac{1}{24} (b - a)^2 .$$

Пример 7. Для показательного распределения с параметром α имеем:

$$D(f) = 1/\alpha^2.$$

Пример 8. Для нормального распределения с параметрами α, σ, как показано в приложении,

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

3.4.3. Возьмем случайную переменную $f \in \mathcal{L}^2$, ее среднее $E(f) = a$, дисперсию $D(f) = \sigma^2$ и число $\varepsilon > 0$. Предположим, что множество $\{x : |f(x) - a| \geqslant \varepsilon\}$ — событие и его индикатор A — случайная переменная. В этом случае, как и в дискретном, верно

Неравенство Чебышева.

$$P(\{x : |f(x) - a| \geqslant \varepsilon\}) \leqslant \frac{D(f)}{\varepsilon^2}.$$

□ Доказательство аналогично данному в дискретном случае, так как среднее E обладает теми же свойствами. Интегрируя и переходя к средним значениям, из неравенств

$$(f(u) - a)^2 \geqslant (f(u) - a)^2 A(u) \geqslant \varepsilon^2 A(u)$$

получаем соотношения

$$D(f) = E\left((f-a)^2\right) \geqslant \varepsilon^2 E(A) = \varepsilon^2 P(A),$$

из которых следует доказываемое неравенство. ■

Как и в дискретном случае, неравенство Чебышева показывает, что дисперсия может служить для оценки вероятности отклонения случайной величины от ее среднего значения.

Замечание. Дополнительное предположение о том, что множество $\{x : |f(x) - a| \geqslant \varepsilon\}$ является событием, связано с данным определением элементарного вероятностного пространства.

3.5. Корреляционная теория

Геометрический язык позволяет изложить эту теорию для непрерывных пространств так же, как для дискретных, используя квадратичную интегрируемость рассматриваемых переменных.

3.5.1. Рассмотрим векторное пространство $\mathcal{L}^2 = \mathcal{L}^2(U,p)$ квадратично интегрируемых по p переменных на непрерывном пространстве (U,p). Пространство $\mathcal{L}^2$ является подпространством пространства $\mathcal{L} = \mathcal{L}(U,p)$ интегрируемых переменных на (U,p).

Если переменные f, g на U квадратично интегрируемы по p, то их произведение интегрируемо по p. Интегрируемость произведения fg позволяет определить скалярное произведение

$$\langle f, g\rangle = E(fg)$$

для каждых квадратично интегрируемых переменных f, g и сделать $\mathcal{L}^2$ евклидовым пространством. Свойства скалярного произведения проверяются так же, как для дискретного пространства. Скалярное произведение симметрично, линейно по каждой переменной, положительно на диагонали, но может быть вырожденным.

3.5.2. Квадратично интегрируемые переменные с нулевым средним значением составляют подпространство $\mathcal{L}_0^2 = \mathcal{L}_0^2(U,p)$ евклидова пространства $\mathcal{L}^2 = \mathcal{L}^2(U,p)$. Каждой переменной $f \in \mathcal{L}^2$ соответствует полученная центрированием переменная $f_0 = f - Ef \in \mathcal{L}_0^2$. Из определений следует, что

$$D(f_0) = E\left(f_0^2\right) = \langle f_0, f_0 \rangle = \|f_0\|^2,$$
$$S(f_0) = \sqrt{D(f_0)} = \sqrt{\langle f_0, f_0 \rangle} = \|f_0\|$$

для каждой $f \in \mathcal{L}^2$. Дисперсия центрированной переменной равна квадрату ее евклидовой нормы, а стандартное отклонение — ее норме в пространстве $\mathcal{L}_0^2$.

Норма для евклидова пространства $\mathcal{L}^2$ определяет сходимость последовательности случайных переменных $f_n \in \mathcal{L}^2$ к переменной $f \in \mathcal{L}^2$. По определению

$$f_n \to f \Leftrightarrow \|f_n - f\| \to 0.$$

В частности, можно определить среднеквадратическую сумму A ряда попарно непересекающихся событий $A_i \subseteq A$:

$$A = \sum A_i \Leftrightarrow \left\| A - \sum_{i=1}^{n} A_i \right\| \to 0 \qquad (n \to \infty),$$

а вместе с этим определить *счетную аддитивность* вероятности P для такой суммы:

$$P(A) = \sum P(A_i) \qquad \left(A = \sum A_i\right).$$

Такая счетная аддитивность следует из равенств

$$P(A) - \sum_{i=1}^{n} P(A_i) = \int Ap - \sum_{i=1}^{n} \int A_i p = \int \left(A - \sum_{i=1}^{n} A_i\right) p =$$
$$= \int \left(A - \sum_{i=1}^{n} A_i\right)^2 p = \left\| A - \sum_{i=1}^{n} A_i \right\|^2 .$$

Здесь множества отождествлены со своими индикаторами и использована их идемпотентность.

3.5.3. Скалярное произведение позволяет измерять зависимость между квадратично интегрируемыми переменными как величину угла между векторами в евклидовом пространстве. Число

$$K(f,g) = \frac{C(f,g)}{S(f)\,S(g)} = \frac{E(fg) - E(f)\,E(g)}{\sqrt{D(f)}\sqrt{D(g)}}$$

называется *коэффициентом корреляции* между квадратично интегрируемыми переменными f, g, для которых $Df > 0$, $Dg > 0$. Из определений следует, что

$$K(f,g) = \left\langle \overline{f_0}, \overline{g_0} \right\rangle ,$$

где

$$\overline{f_0} = (f - Ef) \,/\, \|f - Ef\| , \qquad \overline{g_0} = (g - Eg) \,/\, \|g - Eg\| .$$

Для каждых векторов f, g евклидова пространства $\mathcal{L}^2$ верно неравенство Коши

$$|\langle f, g\rangle| \leqslant \|f\| \, \|g\|$$

и, следовательно,

$$K(f,g) \leqslant \left\|\overline{f_0}\right\| \left\|\overline{g_0}\right\| = 1.$$

Коэффициент корреляции симметричен:

$$K(f,g) = K(g,f).$$

Если $K(f,g) = 0$, то переменные f, g называются *некоррелированными* или *ортогональными*. Равенство $K(f,g) = 1$ означает *стохастическую* линейную зависимость между f и g.

3.6. Функция распределения

Функция $F : \mathbb{R} \to \mathbb{R}$ со значениями

$$F(x) = \int\limits_{-\infty}^{x} p(u)\,du \qquad (x \in \mathbb{R})$$

называется *функцией распределения* для плотности p.

С помощью функции распределения удобно вычислять вероятности интервалов:

$$P([a,b[) = \int\limits_{a}^{b} p(u)\,du = F(b) - F(a) \qquad (-\infty \leqslant a < b \leqslant \infty).$$

Пример. Для равномерного распределения функция F имеет значения

$$F(x) = \begin{cases} 0, & (x < a), \\ \dfrac{x-a}{b-a} & (a \leqslant x \leqslant b), \\ 1 & (x > b), \end{cases}$$

откуда

$$P([u,v[) = \frac{v-u}{b-a} \qquad (a \leqslant u < v \leqslant b).$$

Функция распределения $F : \mathbb{R} \to \mathbb{R}$ непрерывна и дифференцируема в каждой точке $x \in \mathbb{R}$, в которой непрерывна плотность

p. Для ступенчатых плотностей функции распределения кусочно линейны.

Замечание. Понятие распределения случайной переменной на вероятностном пространстве (U, p), аналогичное описанному в 2.6 для дискретного пространства, входит за элементарные рамки.

3.7. Случайные векторы

Случайный вектор $f = (f_1, \ldots, f_m) : U \to \mathbb{R}$ составляется из случайных переменных $f_i : U \to \mathbb{R}$ $(i = 1, \ldots, m)$. Для эффективной теории в качестве множества исходов U нужно прямую $\mathbb{R}$ тоже заменить многомерным пространством $\mathbb{R}^n$. Чтобы определить плотность $p : \mathbb{R}^n \to \mathbb{R}$ и среднее значение $E\,(f)$ случайного вектора $f : \mathbb{R}^n \to \mathbb{R}^m$, надо использовать кратные интегралы. Вместе с понятием распределения случайного вектора это выходит за элементарные рамки.

Отсутствие совместного распределения не позволяет ввести коэффициент информации для случайных переменных на непрерывных пространствах по аналогии с дискретными.

3.8. Условные средние и вероятности

В непрерывных пространствах, как и в дискретных, условия могут описываться бесконечными разбиениями. Нужно только всюду проверять интегрируемость.

3.8.1. Рассмотрим непрерывное пространство (U, p) и разбиение $\mathcal{B} = (B_j)$ множества U на события B_j:

$$\sum B_j = U, \qquad P\,(B_j) \neq 0, \qquad B_j B_k = \varnothing \qquad (j \neq k)\,.$$

Разбиение $\mathcal{B}$ может быть конечным или бесконечным. Пусть $\mathcal{G} = \mathcal{GL}(\mathcal{B})$ — подпространство векторного пространства $\mathcal{L} = \mathcal{L}(U, p)$ интегрируемых по p переменных на U, составленное из интегрируемых по p переменных $g = \sum y_j B_j$.

Условным средним переменной f при условии $\mathcal{B}$ называется переменная

$$E(f \mid \mathcal{B}) = \sum E(f \mid B_j)\, B_j.$$

Переменная $g = E(f \mid \mathcal{B})$ принадлежит $\mathcal{G}$ и имеет то же среднее значение, что и f. В самом деле, благодаря счетной аддитивности и линейности среднего E,

$$\begin{aligned} E(g) = \sum E(E(f \mid B_j)\, B_j) = \\ = \sum E(f \mid B_j)\, E(B_j) = \sum \big(E(fB_j) \mathbin{/} P(B_j)\big)\, P(B_j) = \\ = \sum E(fB_j) = E\Big(f \sum B_j\Big) = E(f)\,. \end{aligned}$$

Каждая ограниченная переменная f на U имеет условное среднее при любом $\mathcal{B}$.

Переменная

$$P(A \mid \mathcal{B}) = E(A \mid \mathcal{B})$$

называется *условной вероятностью* события $A \subseteq U$.

Рассмотрим переменную h со счетным множеством значений z. Если разбиение $\mathcal{B}$ составлено из прообразов $B_z = h^{-1}(z)$ значений z переменной h на U, которые предполагаются событиями, то вместо $E(f \mid \mathcal{B})$ пишут также $E(f \mid h)$:

$$\begin{aligned} &E(f \mid h) = \sum E(f \mid h = z)\, h^{-1}(z), \\ &E(f \mid h = z) = E\big(f \cdot h^{-1}(z)\big) \mathbin{/} P\big(h^{-1}(z)\big)\,. \end{aligned}$$

Для события $f = A$ это равенство превращается в

$$P(f \mid h) = \sum P(A \mid h = z)\, h^{-1}(z).$$

3.8.2. Условное среднее для интегрируемых переменных на непрерывных пространствах имеет те же свойства, что и для переменных на дискретных пространствах:

$$E\,(f \mid \mathcal{B}) = f \qquad (f \in \mathcal{G})\,; \qquad E\,(E\,(f \mid \mathcal{B}) \mid \mathcal{B}) = E\,(f \mid \mathcal{B})\,;$$
$$E\,(E\,(f \mid \mathcal{C}) \mid \mathcal{B}) = E\,(f \mid \mathcal{B})\,,$$

если разбиение $\mathcal{C}$ мельче разбиения $\mathcal{B}$;

$$E\,(f_1 + f_2 \mid \mathcal{B}) = E\,(f_1 \mid \mathcal{B}) + E\,(f_2 \mid \mathcal{B})\,, \quad E\,(hf \mid \mathcal{B}) = hE\,(f \mid \mathcal{B})$$

при $f, f_1, f_2 \in \mathcal{L}$ и $h \in \mathcal{G}$;

$$E\,(f \mid \mathcal{B}) \geqslant 0 \qquad\qquad (\,f \in \mathcal{L}, \quad f \geqslant 0\,)\,.$$

Условное среднее $E\,(\cdot \mid \mathcal{B})$ есть положительный линейный оператор, проектирующий $\mathcal{L}$ на $\mathcal{GL}\,(\mathcal{B})$.

3.8.3. Для непрерывных пространств верны формула полной вероятности и формула Байеса, аналогичные описанным в 2.9 для дискретных пространств. Доказательства тоже идентичны.

Рассмотрим непрерывное вероятностное пространство (U, p) и разбиение $\mathcal{B} = (B_j)$ множества U на события B_j. Для каждого события $A \subseteq U$ верна *формула полной вероятности*

$$P\,(A) = \sum P\,(A \mid B_j)\,P\,(B_j)\,.$$

Верна также *формула Байеса*

$$P\,(B_j \mid A) = P\,(A \mid B_j)\,P\,(B_j)\,/\sum P\,(A \mid B_k)\,P\,(B_k)\,.$$

С помощью этих формул решаются задачи, аналогичные описанным в 1.9.1 и 1.9.3.

3.9. Классические предельные теоремы

Доказываются несколько классических предельных теорем теории вероятностей. Формально они описывают пределы некоторых последовательностей функций и их интегралов. В доказательствах используются формула Тейлора для гладких функций и формула Стирлинга о приближении факториала ([3], [11]).

3.9.1. Пусть: $\alpha \in]0,1[$, $\beta = 1-\alpha$; $j = 0,1,2,\ldots,k$; $k = 1,2,\ldots$; $l = k-j$; $\mu = k\alpha$, $\sigma = \sqrt{k\alpha\beta}$; $u(j) = \sigma^{-1}(j-\mu)$;

$$b(u(j),k) = \begin{cases} \binom{k}{j}\alpha^j\beta^{k-j} & \text{при } j \leqslant k, \\ 0 & \text{при } j < k; \end{cases}$$

$$g(u) = \frac{1}{\sqrt{2\pi}}e^{-u^2/2} \quad (u \in \mathbb{R});$$

$$-\infty < t < v < \infty.$$

Для простоты записей зависимость рассматриваемых переменных от k не всегда указывается явно. В частности, $\sigma = \sigma(k)$.

Лемма о локальном приближении.

$$b(u,k) = g(u)\cdot\sigma^{-1}\left(1 + c(u)\cdot\sigma^{-1}\right)$$

при всех $u = u(j) \in [t,v]$ и некоторых $M > 0$, $c(u) : |c(u)| \leqslant M$.

□ Заметим, что $j = u\sigma + \mu$ и $t \leqslant u \leqslant v$,

$$t\sigma + \mu \leqslant j \leqslant v\sigma + \mu \Leftrightarrow k - (v\sigma + \mu) \leqslant l \leqslant k - (t\sigma + \mu).$$

Применяя формулу Стирлинга

$$n! = \sqrt{2\pi n}n^n e^{-n}e^{\theta(n)} \qquad \left(0 < \theta(n) < \frac{1}{12n}\right),$$

получаем ($|\theta(j,k,l)| \leqslant \frac{1}{12}\left(j^{-1}+k^{-1}+l^{-1}\right)$):

$$b(u,k) = \frac{k!}{j!l!}\alpha^j\beta^l = \frac{1}{\sqrt{2\pi}}\sqrt{\frac{k}{jl}}\left(\frac{k\alpha}{j}\right)^j\left(\frac{k\beta}{l}\right)^l \cdot e^{\theta(j,k,l)}.$$

Как легко проверить, равенство $j\,l/k = \sigma^2\left(1 + c(u)\cdot\sigma^{-1}\right)$ выполняется при всех $u = u(j) \in [t,v]$ и некоторых $M_1 > 0$, $c_1(u)$ таких, что $|c_1(u)| \leqslant M_1$. по формуле Тейлора для $1/\sqrt{1+x}$ отсюда вытекает, что

$$\sqrt{\frac{k}{jl}} = \frac{1}{\sigma}\left(1 + c_2(u)\cdot\frac{1}{\sigma}\right)$$

при всех $u = u(j) \in [t,v]$ и некоторых $M_2 > 0$, $c_2(u)$ таких, что $|c_2(u)| \leqslant M_2$.

Применяя формулу Тейлора для $\ln(1+x)$, получаем:

$$\ln\left(\frac{k\alpha}{j}\left(\frac{k\beta}{l}\right)^l\right) = -(k\alpha + \sigma u)\left(\beta u\frac{1}{\sigma} - \frac{1}{2}\beta^2u^2\frac{1}{\sigma^2} + c_3(u)\frac{1}{\sigma^3}\right) -$$
$$-(k\beta - \sigma u)\left(-\alpha u\frac{1}{\sigma} - \frac{1}{2}\alpha^2u^2\frac{1}{\sigma^2} + c_4(u)\frac{1}{\sigma^3}\right) = -\frac{1}{2}u^2 + c_5(u)\,\sigma$$

при всех $u = u(j) \in [t,v]$ и некоторых $M_i > 0$, $c_i(u)$ таких, что $|c_i(u)| \leqslant M_i$ ($i = 3,4,5$). Используя формулу Тейлора для e^x, находим:

$$\left(\frac{k\alpha}{j}\right)^j\left(\frac{k\beta}{l}\right)^l = e^{-u^2/2}\cdot\left(1 + c_6(u)\frac{1}{\sigma}\right)$$

при всех $u = u(j) \in [t,v]$ и некоторых $M_6 > 0$, $c_6(u)$ таких, что $|c_6(u)| \leqslant M_6$.

Снова применяя формулу Тейлора, получаем:

$$e^{\theta(j,k,l)} = 1 + c_7(u)\frac{1}{\sigma}$$

при всех $u = u(j) \in [t, v]$ и некоторых $M_7 > 0$, $c_7(u)$ таких, что $|c_7(u)| \leqslant M_7$.

Из полученных соотношений вытекает, что

$$b(u,k) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \frac{1}{\sigma} \cdot \left(1 + \frac{c_2(u)}{\sigma}\right) \left(1 + \frac{c_6(u)}{\sigma}\right) \left(1 + \frac{c_7(u)}{\sigma}\right) = \\ = g(u) \left(1 + c(u) \frac{1}{\sigma}\right)$$

при всех $u = u(j) \in [t, v]$ и некоторых $M > 0$, $c(u)$ таких, что $|c(u)| \leqslant M$. ■

По определению $x(k) \sim y(k) \Leftrightarrow x(k)/y(k) \to 1$ $(k \to \infty)$. Из полученных результатов следует

Теорема Муавра (1732).

$$b(u,k) \sim \frac{1}{\sigma(k)} g(u) \qquad (n \to \infty)$$

равномерно по $u = u(j) \in [t, v]$.

□ Используя лемму о локальном приближении, получаем, что $b(u,k) / \left(\sigma^{-1} g(u)\right) = 1 + c(u)\sigma^{-1}$ при всех $u = u(j) \in [t, v]$ и некоторых $M > 0$, $c(u)$ таких, что $|c(u)| \leqslant M$. Вместе с тем $\sigma = \sigma(k) = (k\alpha\beta)^{-1/2} \to 0$ $(k \to \infty)$. ■

Теорему Муавра называют *локальной предельной теоремой*.

3.9.2. Пусть:

$$B(v,k) = \sum_{u(j)<v} b(u(j), k), \qquad G(v) = \int\limits_{-\infty}^{v} g(u)\, du,$$

$$g_k(u) = \begin{cases} g(u(j)) & (u \in [u(j), u(j+1)[\,), \\ 0 & (u \in \mathbb{R} \setminus [u(0), u(k+1)[\,), \end{cases}$$

$$\|f\| = \sup\{|f(u)| : u \in \mathbb{R}\}.$$

Докажем несколько вспомогательных утверждений.

Лемма 1. $\int\limits_{\sqrt{\ln k}}^{\infty} e^{-u^2/2} du \leqslant \frac{1}{\sqrt{k}} \quad (k \geqslant 3).$

□ Заметим, что $t = \sqrt{\ln k} \geqslant 1$ при $k \geqslant 3$. Поэтому

$$\frac{e^{-t^2/2}}{t} \leqslant e^{-t^2/2} = e^{-\left(\sqrt{\ln k}\right)^2/2} = \left(e^{l \ln k}\right)^{-1/2} = \frac{1}{k}.$$

Отсюда следует доказываемое неравенство. ■

Лемма 2. $\|g - g_k\| \leqslant \frac{1}{2\pi\sigma(k)} \quad \left(k \geqslant k(0) \geqslant \frac{\left(1+\sqrt{\beta}\right)^2}{\alpha^2}\right).$

□ Заметим, что

$$|g(u) - g_k(u)| =$$
$$= |g(u) - g(u(j))| \leqslant |g(u(j+1)) - g(u(j))| \leqslant \frac{e^{-1/2}}{2\pi\sigma} \leqslant \frac{1}{2\pi\sigma}$$

при $u \in [u(j), u(j+1)[$. Пусть $k \geqslant k(0) \geqslant \left(1+\sqrt{\beta}\right)^2/\alpha^2$. Тогда

$$u(0) = \frac{-k\alpha}{\sigma} \leqslant -\sqrt{\ln k}, \qquad \sqrt{\ln k} \leqslant u(k) = \frac{(k-1)\alpha}{\sigma}$$

и, следовательно,

$$|g(u) - g_k(u)| \leqslant g(u) \leqslant g(u(0)) \leqslant g(-\sqrt{\ln k}) = \frac{1}{\sqrt{2\pi k}} \leqslant \frac{1}{\sqrt{2\pi}\sigma}$$
$$(u \in]-\infty, u(0)[).$$

Вместе с тем

$$|g(u) - g_k(u)| \leqslant g\left(u - \frac{1}{\sigma}\right) \leqslant g(u(k)) \leqslant g(\sqrt{\ln k}) =$$
$$= \frac{1}{\sqrt{2\pi k}} \leqslant \frac{1}{\sqrt{2\pi}\sigma} \qquad (u \in [u(k+1), \infty[). \quad ■$$

Пусть M — число, фигурирующее в лемме о локальном приближении (3.9.1). Тогда верна

Лемма 3.

$$B(v,k) = \int\limits_{-\infty}^{v} g_k(u)\,du + R(v), \qquad |R(v)| \leqslant (M+1)\,\sigma^{-1}(k).$$

□ Определим $j(v)$ условием $v \in]u(j(v)), u(j(v)+1)]$ и пусть $r(u) = c(u)\,g(u)\,\sigma^{-2}$, $|c(u)| \leqslant M$. Из леммы о локальном приближении следует, что

$$B(v,k) = \sum b(u(j), k) = \sum g(u(j))\,\sigma^{-1} + \sum r(u(j)) =$$

$$= \int\limits_{-\infty}^{u(j(v)+1)} g_k(u)\,du + \sum r(u(j)) = \int\limits_{-\infty}^{v} g_k(u)\,du + R(v),$$

$$R(v) = \int\limits_{-\infty}^{u(j(v)+1)} g_k(u)\,du + \sum r(u(j)) \qquad (j \leqslant j(v)).$$

Заметим, что $|g_k| \leqslant |g| \leqslant 1/\sqrt{2\pi} \leqslant 1$ и поэтому

$$\left| \int\limits_{-\infty}^{u(j(v)+1)} g_k(u)\,du \right| \leqslant \|g_k\|\,\sigma^{-1} \leqslant \sigma^{-1},$$

$$\left| \sum r(u(j)) \right| \leqslant M\sigma^{-1} \sum \int\limits_{u(j)}^{u(j+1)} g_k(u)\,du \leqslant$$

$$\leqslant M\sigma^{-1} \int\limits_{-\infty}^{u(j(v)+1)} g_k(u)\,du \leqslant M\sigma^{-1}\frac{2}{\sqrt{2\pi}} \leqslant M\sigma^{-1}.$$

Следовательно, $|r(v)| \leqslant \sigma^{-1} + M\sigma^{-1} = (M+1)\,\sigma^{-1}$. ■

Лемма 4. $B(v,k) = G(v) + c(v)\cdot\sqrt{\ln k}\cdot\sigma^{-1}(k)$ *при всех* $v \in \mathbb{R}$, $k \geqslant k(0) \geqslant \left(1+\sqrt{\beta}\right)\alpha^{-2}$ *и некоторых* $L > 0$, $c(v)$ *таких, что* $|c(v)| < L$.

□ Из леммы 3 следует, что

$$B(v,k) - G(v) = \int\limits_{-\infty}^{v} \Delta_k(u)\,du,$$

$$\Delta_k(u) = g_k(u) - g(u)\,, \qquad |r(v)| \leqslant (M+1)\,\sigma^{-1}(k)\,.$$

Вместе с тем

$$\left|\int\limits_{-\infty}^{v} \Delta_k(u)\,du\right| \leqslant \left|\int\limits_{-\infty}^{-\sqrt{\ln k}} \Delta_k(u)\,du\right| + \left|\int\limits_{-\sqrt{\ln k}}^{\sqrt{\ln k}} \Delta_k(u)\,du\right| + \left|\int\limits_{\sqrt{\ln k}}^{\infty} \Delta_k(u)\,du\right|.$$

Используя неравенства лемм 1 и 3, при $k \geqslant k(0) \geqslant \left(1+\sqrt{\beta}\right)\alpha^{-2}$ получаем:

$$\left|\int\limits_{-\infty}^{-\sqrt{\ln k}} \Delta_k(u)\,du\right| \leqslant \int\limits_{\sqrt{\ln k}}^{\infty} g(u)\,du \leqslant \frac{1}{\sqrt{k}},$$

$$\left|\int\limits_{\sqrt{\ln k}}^{-\infty} \Delta_k(u)\,du\right| \leqslant \int\limits_{\sqrt{\ln k}}^{\infty} g\left(u-\sigma^{-1}\right)du \leqslant \frac{1}{\sqrt{k}},$$

$$\left|\int\limits_{-\sqrt{\ln k}}^{\sqrt{\ln k}} \Delta_k(u)\,du\right| \leqslant \|g_k - g\| \cdot 2\sqrt{\ln k} \leqslant \sqrt{\ln k}\,\sigma^{-1}(k)\,.$$

Следовательно, при $k \geqslant k(0) \geqslant \left(1+\sqrt{\beta}\right)\alpha^{-2}$ верно неравенство

$$|B(v,k) - G(v)| \leqslant \left(\sqrt{\alpha\beta} + \sqrt{\ln k} + \sqrt{\alpha\beta} + (M+1)\right)\sigma^{-1}(k)\,. \;■$$

Рассмотрим приращения биномиальной и нормальной функций распределения ($-\infty \leqslant t < v \leqslant \infty$):

$$B(t, v, k) = B(v, k) - B(t, k), \qquad G(t, v) = G(v) - G(t).$$

Из полученных результатов следует

Теорема Лапласа (1801). $B(t, v, k) \to G(t, v)$ $(k \to \infty)$ *равномерно по* t, v.

□ используя лемму 4, получаем:

$$\begin{aligned}|B(t, v, k) - G(t, v)| &= |B(v, k) - G(v) - (B(t, k) - G(t))| = \\ &= |c(v) - c(t)| \sqrt{\ln k}\,\sigma^{-1}(k) \leqslant L\sqrt{\ln k}\,\sigma^{-1}(k) = \\ &= L\sqrt{\ln k}/\sqrt{k\alpha\beta} \to 0 \qquad (k \to \infty). \ \blacksquare\end{aligned}$$

Теорему Лапласа называют *интегральной предельной теоремой.*

3.9.3. Рассмотрим распределение Бернулли с параметрами k, α/k и распределение Пуассона с параметром $\alpha > 0$:

$$b(\alpha/k, j, k) = \binom{k}{j}\left(\frac{\alpha}{k}\right)^j \left(1 - \frac{\alpha}{k}\right)^{k-j} \qquad (0 \leqslant j \leqslant k),$$

$$p(\alpha, j) = \frac{\alpha^j}{j!}e^{-j} \qquad (j \geqslant 0).$$

Верна

Теорема Пуассона (1832). $b(\alpha/k, j, k) \to p(\alpha, j)$ $(k \to \infty)$.

□ В самом деле

$$\begin{aligned}b(\alpha/k, j, k) &= \frac{k(k-1)\cdots(k-j+1)}{j!}\left(\frac{\alpha}{k}\right)^j\left(1 - \frac{\alpha}{k}\right)^{k-j} = \\ &= \frac{\alpha^j}{j}\left(1 - \frac{\alpha}{k}\right)^k \left(1 - \frac{1}{k}\right)\left(1 - \frac{2}{k}\right)\cdots\left(1 - \frac{j-1}{k}\right)\left(1 - \frac{\alpha}{k}\right)^{-j} \to \\ &\underset{k\to\infty}{\longrightarrow} \frac{\alpha^j}{j!}e^{-j} = p(\alpha, j). \qquad \blacksquare\end{aligned}$$

Это еще одна классическая локальная предельная теорема. теоремы Муавра, Лапласа и Пуассона часто применяются при решении теоретических и прикладных задач.

Литература

[1] *Энциклопедия элементарной математики. Т.1. Арифметика.* Москва, Гостехиздат, 1951.

[2] *Энциклопедия элементарной математики. Т.2. Алгебра.* Москва, Гостехиздат, 1951.

[3] *Энциклопедия элементарной математики. Т.3. Функции и пределы (основы анализа).* Москва, Гостехиздат, 1952.

[4] Савельев Л. Я. *Лекции по математическому анализу. Часть 1.* Новосибирск, НГУ, 1969.

[5] Савельев Л. Я. *Лекции по математическому анализу. Части 2.1, 2.2–3.* Новосибирск, НГУ, 1973.

[6] Савельев Л. Я. *Интегрирование равномерно измеримых функций.* Новосибирск, НГУ, 1984.

[7] Савельев Л. Я. *Приложения к теории вероятностей.* Новосибирск, НГУ, 1989.

[8] Савельев Л. Я. *Комбинаторика и вероятность.* Новосибирск, Наука, 1975.

[9] Грэхем Р., Кнут Д., Поташник О. *Конкретная математика. Основание информатики.* Москва, Мир, 1998.

[10] Робертс Ф. С. *Дискретные математические модели с приложениями к социальным, биологическим и экологическим задачам.* Москва, Наука, 1986.

[11] Феллер В. *Введение в теорию вероятностей и ее приложения. Том 1.* Москва, Мир, 1984.

[12] Яглом А. М., Яглом И. М. *Вероятность и информация.* Москва, Наука, 1973.

Алфавитный указатель