

Предисловие

Настоящее издание представляет собой продолжение пособия «Теория вероятностей» того же автора и сохраняет ту же структуру: каждый параграф содержит теоретическую часть и задачи. Однако ввиду повышенной по сравнению с первой книгой сложности материала теоретическая часть значительно расширена, содержит не только формулы и теоремы, а во многих случаях – и их доказательства, то есть представляет собой по существу краткий конспект лекций по разделам «Двумерные случайные величины», «Энтропия и информация». Задачи, предлагаемые студентам для решения в каждом параграфе, также часто носят теоретический характер.

Ссылки на указанное пособие (I и II части) имеют формат, например, (I,3), то есть часть I, параграф 3.

Часть III. Двумерные случайные величины

Введение

На практике нередко приходится *одновременно* учитывать значения нескольких величин, неслучайных и случайных. Это приводит к понятию **системы случайных величин**. Математически такую систему можно также называть **многомерной** случайной величиной или **случайным вектором**.

Пример. Состояние погоды в данной географической точке и в данный момент времени описывается некоторым набором параметров: температура воздуха, атмосферное давление, направление и скорость ветра, наличие, вид и интенсивность атмосферных осадков. Очевидно, что эти параметры можно изучать по отдельности, но для синоптиков имеет огромное значение изучение параметров погоды в их совокупности и взаимодействии.

Случайный вектор (n -мерная случайная величина) представляет собой упорядоченную совокупность обычных (непрерывных или дискретных) случайных величин: $\mathbf{X} = (X_1, X_2, \dots, X_n)$. Каждая координата (**компонента**) X_i имеет определенные математические характеристики, изученные нами ранее (функции распределения, математическое ожидание, дисперсию), однако изучение взаимодействия компонент случайного вектора приводит, очевидно, к появлению новых математических понятий.

В общем случае изучение взаимодействия многих случайных факторов – очень сложная математическая задача. Поэтому в дальнейшем мы ограничимся *двумерными* случайными величинами $n = 2$ (и в этом случае *пару* случайных величин будем обозначать (X, Y)).

1. Дискретная двумерная величина

По аналогии с рядом распределения обычной ДСВ (II,1) для задания всех вероятностей, связанных с парой (X, Y) дискретных случайных величин, используется **таблица совместного распределения**, которая имеет следующий вид:

$X \backslash Y$	x_1	x_2	$\dots$	x_i	$\dots$	P_Y
y_1	p_{11}	p_{21}	$\dots$	p_{i1}	$\dots$	p_{1y}
y_2	p_{12}	p_{22}	$\dots$	p_{i2}	$\dots$	p_{2y}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
y_j	p_{1j}	p_{2j}	$\dots$	p_{ij}	$\dots$	p_{jy}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
P_X	p_{1x}	p_{2x}	$\dots$	p_{ix}	$\dots$	1

Число строк и столбцов в таблице не указано. Это значит, что, как и в таблице для одномерного распределения, каждое из них может быть и бесконечным.

Каждое число p_{ij} во внутренней части таблицы обозначает вероятность того, что *в одном испытании* две случайные величины примут соответствующие значения:

$$p_{ij} = P(X = x_i, Y = y_j).$$

Строка P_X и столбец P_Y представляют собой *одномерные* распределения величин X и Y по отдельности и получаются суммированием вероятностей p_{ij} соответственно по столбцам и строкам (здесь используется формула разложения

по полной группе событий (I, 6)): $p_{ix} = \sum_j p_{ij}$, $p_{jy} = \sum_i p_{ij}$. Эти распределения удовлетворяют основному свойству ряда распределения (II, 1):

$$\sum_i p_{ix} = \sum_j p_{jy} = \sum_{i,j} p_{ij} = 1.$$

Они позволяют вычислить раздельно числовые характеристики компонент X и Y , например:

$$M(X) = \sum_i x_i p_{ix},$$

$$D(X) = \sum_i x_i^2 p_{ix} - (M(X))^2.$$

Если величины X , Y *независимы*, то есть принимают все свои значения независимо друг от друга, то при любых i, j будет выполняться равенство

$$P_{ij} = P_{ix}P_{jy}$$

(его легко проверить *вычислением*). Это условие можно сформулировать в других формах:

- а) для любых i, j, k, l $P_{ij}P_{kl} = P_{il}P_{kj}$;
- б) все столбцы и все строки матрицы $P = (p_{ij})$ пропорциональны;
- в) ранг этой матрицы равен 1.

Задачи

1.1. Монету бросили 3 раза. Составить таблицу совместного распределения двух случайных величин: X – количество выпавших «орлов», Y – число бросаний до выпадения первого «орла» (если «орел» так и не выпал, считаем $Y=3$). Найти одномерные распределения компонент, их числовые характеристики.

1.2. Независимые дискретные случайные величины X, Y заданы рядами распределения

X	8	12	16
P	0,4	0,5	0,1

Y	1	2
P	0,3	0,7

а) Составить ряды распределения величин

$U = X + Y$, $V = X + 4Y$ (Указание. Одинаковые значения величины объединяются, а соответствующие вероятности складываются), $S = XY$, $T = X/Y$. Проверить во всех случаях выполнение теорем сложения и умножения для математических ожиданий и дисперсий.

б) Составить таблицы совместного распределения для пар величин (X, U) , (U, V) , (S, T) . Являются ли эти пары независимыми? Проверить, что одномерные распределения величин, полученные по этим таблицам, совпадают с их распределениями, полученными в п.а).

1.3. Заполнить до конца таблицу совместного распределения

$X \backslash Y$	1	3
0	0,1	0,2
5	0,3	

Убедиться, что величины X, Y не являются независимыми. Найти их индивидуальные числовые характеристики.

2. Регрессия

Пусть известно, что величина X в испытании приняла значение x_i . Тогда для каждого из возможных значений y_j величины Y можно вычислить со-

ответствующую условную вероятность (применяем формулу условной вероятности (I, 4))

$$P_{X=x_i}(Y=y_j) = \frac{P(X=x_i, Y=y_j)}{P(X=x_i)} = \frac{p_{ij}}{p_{ix}}.$$

Эти вероятности составляют *условное распределение* величины Y , соответствующее данному значению x_i величины X . Они удовлетворяют основному условию,

накладываемому на ряд распределения: $\sum_j \frac{p_{ij}}{p_{ix}} = 1$. Это позволяет вычислить

условное математическое ожидание

$$M_{X=x_i}(Y) = \frac{\sum_j y_j p_{ij}}{p_{ix}}.$$

Определение. Зависимость условного математического ожидания величины Y от значения величины X называется *регрессией* Y на X .

Графически закон регрессии отображается в виде *ломаной регрессии*, соединяющей последовательно точки с координатами $(x_i; M_{X=x_i}(Y))$.

Аналогично рассматривается регрессия X на Y .

Задачи

2.1. В условиях задач 1.1, 1.2б) найти все условные математические ожидания, построить ломаные регрессии.

3. Ковариация и корреляция

Как приближение истинного закона регрессии рассматривают *прямую регрессии*. При этом в качестве критерия наилучшего приближения используется *правило наименьших квадратов* (ПНК).

Сформулируем и решим задачу о наилучшем приближении. Пусть искомая прямая имеет уравнение вида $y = kx + b$, а *отклонение* от нее точки $(x_i; y_j)$ равно $\delta_{ij} = y_j - (kx_i + b)$ (эту величину считаем значением случайной величины Δ). Согласно ПНК, наилучшей прямой является та, для которой $M(\Delta^2)$ достигает минимума. Задача сводится к минимизации функции двух переменных

$$\Phi(k, b) = \sum_{i,j} \delta_{ij}^2 p_{ij} = \sum_{i,j} (y_j - kx_i - b)^2 p_{ij}.$$

Вычислив производные $\frac{\partial \Phi}{\partial k}$ и $\frac{\partial \Phi}{\partial b}$ и приравняв их к нулю, после перегруппировки членов получим:

$$\begin{cases} k \sum_{i,j} x_i^2 p_{ij} + b \sum_{i,j} x_i p_{ij} = \sum_{i,j} x_i y_j p_{ij}, \\ k \sum_{i,j} x_i p_{ij} + b \sum_{i,j} p_{ij} = \sum_{i,j} y_j p_{ij}. \end{cases}$$

Учтем равенства из параграфа 1:

$$\sum_{i,j} p_{ij} = 1, \quad \sum_{i,j} x_i p_{ij} = \sum_i x_i \sum_j p_{ij} = \sum_i x_i p_{i\cdot} = M(X) \quad \text{и т.д. и получим:}$$

$$\begin{cases} kM(X^2) + bM(X) = M(XY), \\ kM(X) + b = M(Y), \end{cases}$$

откуда легко получить равенство

$$k(M(X^2) - (M(X))^2) = M(XY) - M(X)M(Y).$$

В левой части имеем, очевидно, $kD(X)$, а для правой введем специальный термин.

Определение. Величина $K(X, Y) = M(XY) - M(X)M(Y)$ называется *ковариацией* пары случайных величин X, Y .

Вычислив также величину b и подставив в уравнение прямой, получим, что оно имеет вид

$$y - M(Y) = \frac{K(X, Y)}{D(X)}(x - M(X)).$$

Введем еще величину

$$r(X, Y) = \frac{K(X, Y)}{\sqrt{D(X)D(Y)}} = \frac{K(X, Y)}{\sigma_x \sigma_y},$$

называемую *корреляцией* пары случайных величин X, Y . Тогда уравнение прямой регрессии Y на X принимает окончательный вид:

$$y - M(Y) = r(X, Y) \frac{\sigma_y}{\sigma_x} (x - M(X)).$$

Очевидно, что парное к нему уравнение прямой регрессии X на Y имеет вид

$$x - M(X) = r(X, Y) \frac{\sigma_x}{\sigma_y} (y - M(Y)).$$

Ковариация и корреляция характеризуют степень связи двух величин. Сформулируем их основные свойства, вытекающие из соответствующих свойств математического ожидания и дисперсии (II, 2).

1. Симметрия: $K(X, Y) = K(Y, X)$; $r(X, Y) = r(Y, X)$.

2. Линейность ковариации по каждому из аргументов:

$$K(CX, Y) = CK(Y, X); \quad K(X, Y \pm Z) = K(X, Y) \pm K(X, Z).$$

Следствие. $r(CX, Y) = \pm r(Y, X)$ в зависимости от знака константы.

3. Если величины X, Y независимы, то $K(X, Y) = r(X, Y) = 0$. В частности, $K(X, C) = r(X, C) = 0$.

Следствие. $K(X + C, Y) = K(X, Y)$; $r(X + C, Y) = r(Y, X)$ (при смещении величин на константу ковариация и корреляция не меняются).

Замечание. Утверждение, обратное свойству 3, вообще говоря, неверно: случайные величины могут быть *некоррелированными*, но зависимыми. В качестве примера можете рассмотреть следующую таблицу совместного распределения (если в клетке стоит 0, то для наглядности таблицы его можно опустить):

$X \backslash Y$	-2	-1	1	2
-2				0,1
-1		0,4		
1			0,4	
2	0,1			

4. $K(X, X) = D(X)$; $r(X, X) = 1$.

5. Как и для дисперсии, для ковариации существует второй способ вычисления — как *совместного центрального момента*:

$$K(X, Y) = M((X - M(X))(Y - M(Y))) = M(X_C Y_C)$$

6. Общая формула сложения дисперсий:

$$D(X \pm Y) = D(X) + D(Y) \pm 2K(X, Y).$$

7. $|K(X, Y)| \leq \sigma_x \sigma_y$; $|r(X, Y)| \leq 1$.

Доказательство. Рассмотрим при любом действительном значении t величину $Z_t = X_C + tY_C$. Так как $M(X_C) = M(Y_C) = M(Z_t) = 0$, то

$$D(Z_t) = M(Z_t^2) = M(X_C^2) + 2tM(X_C Y_C) + t^2M(Y_C^2) = \sigma_x^2 + 2tK(X, Y) + t^2\sigma_y^2.$$

Последнее выражение относительно t представляет собой полный квадрат, который ни при каком t не принимает отрицательных значений (это дисперсия!). Поэтому соответствующий дискриминант не положителен. Легко убедиться, что отсюда и вытекает неравенство 7.

Более того, известно, что дисперсия обращается в ноль только для величины, тождественно равной нулю. Таким образом, трехчлен имеет корень (то есть его дискриминант равен нулю, и только в этом случае неравенство 7 может обратиться в равенство) только если при некотором значении t_0 $Z_{t_0} = 0$, то есть величины X и Y *линейно связаны*.

Из перечисленных свойств вытекает, что корреляция характеризует степень зависимости (*тесноту связи*) между двумя случайными величинами. Она:

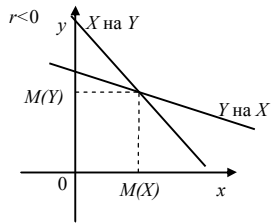
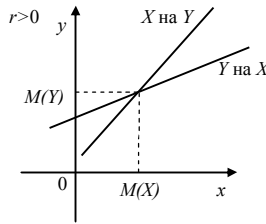
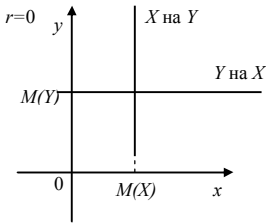
обращается в 0 для пары независимых величин;

равна ± 1 тогда и только тогда, когда две величины линейно связаны;

в остальных случаях $0 < |r(X, Y)| < 1$;

не меняется (с точностью до знака) при линейном изменении каждой из величин.

Эти свойства отражаются и на прямых регрессии. Они:



пересекаются в точке $(M(X); M(Y))$;

перпендикулярны (т.е. параллельны осям координат) для независимых величин;

сливаются при $r = \pm 1$ и только в этом случае.

При этом знак корреляции совпадает со знаком угловых коэффициентов обеих прямых (положительная и отрицательная корреляция).

Как частный случай корреляции можно рассмотреть *корреляцию случайных событий*. Для любого события A рассмотрим *индикаторную* случайную величину X_A , которая принимает значение 1, если в результате испытания A событие произошло, и значение 0 в противном случае. Для двух связанных событий A и B можно рассмотреть совместное распределение величин X_A и X_B и вычислить соответствующий коэффициент корреляции, который и назовем коэффициентом корреляции событий A и B : $r(A, B) = r(X_A, X_B)$.

Пример. Ключ у меня лежит либо в правом, либо в левом кармане. Искать его я начинаю с кармана, выбранного наудачу с равной вероятностью для каждого, но в 80% случаев он почему-то находится в другом кармане. Найти коэффициент корреляции между событиями A : «я начал поиск с правого кармана» и B : «ключ лежит в правом кармане».

Решение. Составим таблицу совместного распределения индикаторных величин:

По ней легко вычислить, что $r(A, B) = -0,6$.

Заметим, что значения 0 и 1 для индикаторных множеств несущественны. Можно перейти к любым другим (различным) значениям, что соответствует *линейному* преобразованию случайных величин, т.е. не влияет (с точностью до знака) на значение $r(A, B)$.

$X_A \backslash X_B$	0	1
0	0,1	0,4
1	0,4	0,1

Задачи

3.1. Каждый четвертый мужчина — блондин, у каждого десятого голубые глаза, четверо из ста обладают обоими признаками. Найти коэффициент корреляции двух признаков.

3.2. а) В условиях задачи 1.1 вычислить ковариацию и коэффициент корреляции величин X , Y . Составить уравнения прямых регрессии, сравнить их гра-

фически с ломаными регрессии.

б) То же – для пар величин в задаче 1.2б).

3.3. В условиях задачи 1.3 убедиться, что ломанные регрессии и прямые регрессии совпадают.

3.4. Пусть каждая из независимых случайных величин U, V принимает только значения 1 и 2 с равной вероятностью. Показать, что величины $X = U + V, Y = UV$ связаны однозначной, но не линейной зависимостью.

Убедиться, что ломанные регрессии совпадают, а прямые регрессии – нет.

Рассмотреть также случаи:

$$X = U + V, Y = U^2 + V^2;$$

$$X = U + V, Y = U - V.$$

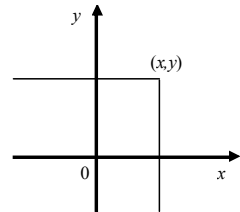
4. Функции распределения

По аналогии со случаем одной случайной величины (II, 4) введем для пары случайных величин (X, Y) описывающую ее функцию двух переменных (*интегральную функцию распределения*): для любой пары действительных чисел (x, y)

$$F(x, y) = P(X < x, Y < y).$$

Геометрически эта функция выражает вероятность попадания значения пары (X, Y) в *квадрант* плоскости (см. рис.):

Свойства этой функции аналогичны одномерному случаю.



$$1. \quad 0 \leq F(x, y) \leq 1.$$

2. Если одна из переменных неограниченно убывает, то значение функции убывает к нулю:

$$\lim_{x \rightarrow -\infty} F(x, y) = 0, \quad \lim_{y \rightarrow -\infty} F(x, y) = 0.$$

Действительно, события $(X < -\infty, Y < y)$ и $(X < x, Y < -\infty)$ являются невозможными.

3. Если одна из переменных неограниченно возрастает, то функция двух переменных приближается к значению одномерной функции другой переменной:

$$\lim_{x \rightarrow +\infty} F(x, y) = F_Y(y), \quad \lim_{y \rightarrow +\infty} F(x, y) = F_X(x).$$

Действительно, событие $(X < +\infty, Y < y)$ совпадает с событием $(Y < y)$.

$$4. \quad \text{Наконец, } \lim_{\substack{x \rightarrow +\infty \\ y \rightarrow +\infty}} F(x, y) = 1.$$

5. Пусть $x_1 < x_2$. Тогда разность

$$F(x_2, y) - F(x_1, y) = P(x_1 \leq X < x_2, Y < y) \geq 0$$

ния пары (X, Y) в *полуплоскости*. Таким образом, интегральная функция *не убывает* по каждой из переменных.

6. Теперь можно найти вероятность попадания значения пары (X, Y) в прямоугольник:

$$P(x_1 \leq X < x_2, y_1 \leq Y < y_2) = P(x_1 \leq X < x_2, Y < y_2) - P(x_1 \leq X < x_2, Y < y_1) = \\ = F(x_2, y_2) + F(x_1, y_1) - F(x_2, y_1) - F(x_1, y_2).$$

Замечание. Напомним, что для *непрерывных* случайных величин употребление знаков $\leq$ и $<$ безразлично.

Последнее выражение можно записать в иной форме, если функция $F(x, y)$ *дважды дифференцируема*.

Определение. Дифференциальной функцией распределения (двумерной плотностью вероятности) называется функция

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}.$$

Двойной интеграл по любому прямоугольнику от этой функции, очевидно, выражается той же комбинацией $F(x_2, y_2) + F(x_1, y_1) - F(x_2, y_1) - F(x_1, y_2)$, поэтому для любого прямоугольника

$$P(x_1 \leq X < x_2, y_1 \leq Y < y_2) = \iint_{\substack{x_1 \leq x < x_2 \\ y_1 \leq y < y_2}} f(x, y) dx dy.$$

Но, как известно из теории двойного интеграла, любую область можно приблизить областями, состоящими из прямоугольников, поэтому последнее равенство переносится на любую область D в плоскости переменных (x, y) :

$$7. P((X, Y) \in D) = \iint_D f(x, y) dx dy.$$

Следствие. Интегрируя по всей плоскости, получим 1 как вероятность достоверного события: $\iint_{R^2} f(x, y) dx dy = 1$.

Далее получим формулы, связывающие одномерные и двумерную функции плотности распределения.

$$8. f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy, \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx.$$

Доказательство. Рассмотрим сначала интегральную функцию (применяем св.7):

$$F_X(x) = P(X < x) = P(X < x, -\infty < Y < +\infty) = \int_{-\infty}^x dx \int_{-\infty}^{+\infty} f(x, y) dy.$$

Для получения первой из формул 8 теперь достаточно взять производную по x (используем теорему о производной интеграла по верхнему пределу).

Замечание. Для формально верной записи следовало бы внутри инте-

грала изменить обозначение переменной, т.к. x — это обозначение предела интегрирования. Однако на практике такая «путаница» не приводит к недоразумениям, мы будем применять такие же обозначения и далее.

Теперь легко получить формулы для вычисления *индивидуальных* числовых характеристик величин X, Y .

$$9. M(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \iint_{R^2} x f(x, y) dx dy,$$

$$D(X) = \iint_{R^2} (x - M(X))^2 f(x, y) dx dy = \iint_{R^2} x^2 f(x, y) dx dy - (M(X))^2$$

(аналогично для $M(Y), D(Y)$).

Вопрос об условиях распределений, условных математических ожиданиях и о законе регрессии для непрерывного совместного распределения более сложен, т.к. условие $(X = x)$ само по себе имеет нулевую вероятность. Поэтому следует сначала рассмотреть для величины X малый интервал значений, а затем перейти к пределу.

Итак, рассмотрим *условное распределение* величины Y при условии $(x \leq X < x + \Delta x)$ (используем формулу условной вероятности и свойство 7):

$$F_{x \leq X < x + \Delta x}(y) = P_{x \leq X < x + \Delta x}(Y < y) = \frac{P(x \leq X < x + \Delta x, Y < y)}{P(x \leq X < x + \Delta x)} = \frac{\int_x^{x+\Delta x} dx \int_{-\infty}^y f(x, y) dy}{\int_x^{x+\Delta x} dx \int_{-\infty}^{+\infty} f(x, y) dy}.$$

Если Δx мало, то во внутренних интегралах (в числителе и знаменателе) можно считать, что аргумент x принимает только начальное значение:

$$F_{x \leq X < x + \Delta x}(y) \approx \frac{\Delta x \int_{-\infty}^y f(x, y) dy}{\Delta x \int_{-\infty}^{+\infty} f(x, y) dy}.$$

При $\Delta x \rightarrow 0$ равенство в пределе становится точным. Учтем еще свойство 8:

$$F_{X=x}(y) = \frac{\int_{-\infty}^y f(x, y) dy}{\int_{-\infty}^{+\infty} f(x, y) dy} = \frac{\int_{-\infty}^y f(x, y) dy}{f_X(x)}.$$

Дифференцируя по y , получим:

$$f_{X=x}(y) = \frac{f(x, y)}{\int_{-\infty}^{+\infty} f(x, y) dy} = \frac{f(x, y)}{f_X(x)}.$$

Следствия. 1 (аналог формулы умножения вероятностей).

$$f(x, y) = f_X(x) \cdot f_{Y=x}(y).$$

2. Если компоненты X, Y независимы, то $f_{Y=x}(y)$ не зависит от x и совпадает с $f_Y(y)$.

Заметим, что все эти формулы имеют смысл лишь для тех значений x , для которых $f_X(x) > 0$ (в тех интервалах оси абсцисс, где $f_X(x) = 0$, величина X значений не принимает). Для *этих* значений x теперь можно вычислить и *условные математические ожидания*.

$$M_{Y=x}(Y) = \frac{\int_{-\infty}^{+\infty} y f(x, y) dy}{\int_{-\infty}^{+\infty} f(x, y) dy} = \frac{\int_{-\infty}^{+\infty} y f(x, y) dy}{f_X(x)}$$

и построить *линию регрессии* Y на X , уравнение которой имеет вид $y = M_{Y=x}(Y)$.

Условные математические ожидания обладают следующим свойством (правило повторного математического ожидания):

$$M(M_{Y=x}(Y)) = \int_{-\infty}^{+\infty} M_{Y=x}(Y) f_X(x) dx = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} y f(x, y) dy = \iint_{R^2} y f(x, y) dx dy = M(Y)$$

(среднее значение условного математического ожидания величины равно ее безусловному математическому ожиданию).

Все полученные формулы можно переписать и для условных распределений величины X , ее условных математических ожиданий, регрессии X на Y .

Формулы для ковариации, корреляции, свойства этих величин, уравнения прямых регрессии переносятся на случай непрерывных случайных величин без изменений с единственным уточнением

$$M(XY) = \iint_{R^2} xy f(x, y) dx dy.$$

Задачи

4.1. Дана интегральная функция распределения системы двух случайных ве-

$$\text{личин: } F(x, y) = F_1(x)F_1(y), \text{ где } F_1(x) = \begin{cases} 0, & x \leq 0, \\ \sin x, & 0 < x \leq \frac{\pi}{2}, \\ 1 & \frac{\pi}{2} < x. \end{cases}$$

Записать выражение для функции двумерной плотности распределения, вы-

числить следующие вероятности: $P\left(X < \frac{\pi}{3}\right)$, $P\left(Y > \frac{\pi}{4}\right)$, $P\left(X < \frac{\pi}{3}, Y > \frac{\pi}{4}\right)$.

4.2. На всей плоскости задана дифференциальная функция распределения системы двух случайных величин: $f(x, y) = \frac{C}{(x^2 + y^2 + 1)^2}$. Найти: величину C ,

вероятность $P(X^2 + Y^2 < r^2)$ в зависимости от r , указать значения r , соответствующие значениям этой вероятности 0,95 и 0,999.

4.3. Дана дифференциальная функция распределения системы двух случайных величин: $f(x, y) = C \sin(x + y)$ при $x, y \in \left[0; \frac{\pi}{2}\right]$ и $f(x, y) = 0$ вне этого

квадрата. Найти дифференциальные функции одномерных распределений, условные математические ожидания и уравнения прямых регрессии для этой пары случайных величин. Являются ли они независимыми?

4.4. Найти одномерные распределения, условные математические ожидания и уравнения прямых регрессии для пары случайных величин, распределенных равномерно:

а) в четверти круга $x^2 + y^2 \leq R^2$, находящейся в первом квадранте;

б) в треугольнике с вершинами (0,0), (0,2), (5,0).

4.5. Пара случайных величин равномерно распределена в круге. Показать, что эти величины зависимы, но не коррелированы.

4.6. На всей плоскости дана функция двумерной плотности распределения

$$f(x, y) = Ae^{-4x^2 - 6xy - 9y^2}.$$

а) Найти одномерные функции плотностей распределения величин X и Y . Убедиться, что они соответствуют нормальным распределениям с нулевым средним и различными среднеквадратичными отклонениями (найти их значения). Отсюда определить величину A .

б) Найти функции плотности условных распределений $f_{Y=y}(x)$ и $f_{X=x}(y)$ для любого значения переменных y и x соответственно.

в) Найти законы регрессии, убедиться, что теоретические линии регрессии являются прямыми. По их уравнениям установить значение коэффициента корреляции величин X и Y .

5. Функции случайных величин. Свертка

Пусть $g(x, y)$ — функция двух переменных, определенная на всей плоскости. Тогда можно определить новую случайную величину

$$Z = g(X, Y).$$

Ее числовые характеристики, очевидно, вычисляются по формулам

$$M(Z) = \iint_{R^2} g(x, y) f(x, y) dx dy, \quad D(Z) = \iint_{R^2} (g(x, y))^2 f(x, y) dx dy - (M(Z))^2$$

(если эти интегралы существуют).

Чтобы определить интегральную функцию величины Z , рассмотрим множества уровня функции g , определяемые для любого действительного числа z следующим образом: $G_z = \{(x, y) : g(x, y) < z\}$. Тогда (используем св.7 из предыдущего параграфа)

$$F_Z(z) = P(Z < z) = P((X, Y) \in G_z) = \iint_{G_z} f(x, y) dx dy$$

(записать выражение для производной этой функции $f_Z(z)$ в общем случае затруднительно).

Пример. Пусть $Z = X^2 + Y^2$. Тогда $G_z = \begin{cases} \emptyset, & z \leq 0, \\ \{(x, y) : x^2 + y^2 < z\}, & z > 0. \end{cases}$

Рассмотрим простейший случай $g(x, y) = x + y$ (т.е. $Z = X + Y$ — сумма двух компонент случайного вектора). Тогда для любого значения z множество $G_z = \{(x, y) : x + y < z\}$ представляет собой полуплоскость, и

$$F_Z(z) = F_{X+Y}(z) = \iint_{G_z} f(x, y) dx dy = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-x} f(x, y) dy.$$

Отсюда

$$f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx.$$

Отдельно рассмотрим случай, когда компоненты X, Y независимы. Тогда

$$F(x, y) = P(X < x, Y < y) = F_X(x)F_Y(y), \quad f(x, y) = \frac{\partial^2 F}{\partial x \partial y} = f_X(x)f_Y(y)$$

и

$$f_{X+Y}(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z-x) dx.$$

Такое преобразование двух дифференциальных функций называется их *свёрткой*.

Пример. Рассмотрим свертку двух функций показательного распределения. Пусть $f_X(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x > 0, \end{cases} \quad f_Y(y) = \begin{cases} 0, & y < 0, \\ \mu e^{-\mu y}, & y > 0. \end{cases}$ Применим формулу свертки. Очевидно, при $z < 0$ из равенства $x + y = z$ вытекает, что либо $x < 0$, либо $y < 0$, откуда $f_{X+Y}(z) = 0$. Пусть теперь $z > 0$. На прямой $x + y = z$ достаточно взять только отрезок между осями, где обе переменные положительны. Тогда

$$f_{X+Y}(z) = \int_0^z \lambda e^{-\lambda x} \mu e^{-\mu(z-x)} dx = \lambda \mu e^{-\mu z} \int_0^z e^{(\mu-\lambda)x} dx. \text{ Здесь надо рассмотреть случаи}$$

$\lambda = \mu$ и $\lambda \neq \mu$. Окончательный результат:

$$\text{при } \lambda = \mu \quad f_{X+Y}(z) = \begin{cases} 0, & z < 0, \\ \lambda^2 z e^{-\lambda z}, & z > 0, \end{cases}$$

$$\text{при } \lambda \neq \mu \quad f_{X+Y}(z) = \begin{cases} 0, & z < 0, \\ \frac{\lambda \mu}{\mu - \lambda} (e^{-\lambda z} - e^{-\mu z}), & z > 0. \end{cases}$$

Задачи

5.1. Для величины, являющейся сверткой двух независимых случайных величин, распределенных (каждая) равномерно на отрезке $[0; 1]$, найти: функции распределения, числовые характеристики.

5.2. В условиях задачи 4.1а) найти функции распределения и числовые характеристики:

- а) свертки двух данных случайных величин;
- б) суммы их квадратов.

5.3. Проверить, что свертка двух независимых стандартных нормальных величин также распределена по нормальному закону. Найти его параметры.

6. Двумерное нормальное распределение

А. Рассмотрим сначала простейший случай двумерного распределения с *независимыми* компонентами, имеющими *нормальное* распределение с нулевыми средними. Итак, если $X \in N(0, \sigma_x)$, $Y \in N(0, \sigma_y)$, то перемножим соответствующие функции плотности распределения и получим

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-L(x, y)},$$

где

$$L(x, y) = \frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right).$$

Рассмотрим линии уровня функций $L(x, y)$ и $f(x, y)$. Они представляют собой эллипсы и задаются уравнениями $\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = t^2$ (полуоси эллипса равны $t\sigma_x$ и $t\sigma_y$). Обозначив через $\mathcal{E}_t$ часть плоскости, ограниченную этим

ЭЛЛИПСОМ, ВЫЧИСЛИМ

$$P((X, Y) \in \mathfrak{I}_t) = \frac{1}{2\pi\sigma_x\sigma_y} \iint_{\mathfrak{I}_t} e^{-L(x,y)} dx dy.$$

После замены переменных $x = \sigma_x u$, $y = \sigma_y v$ будет $L(u, v) = \frac{u^2 + v^2}{2}$. Это позволяет перейти к полярным координатам:

$$P((X, Y) \in \mathfrak{I}_t) = \frac{1}{2\pi} \iint_{u^2 + v^2 \leq t^2} e^{-\frac{\rho^2}{2}} \rho d\rho d\varphi = \frac{1}{2\pi} \int_0^{2\pi} d\varphi \int_0^t e^{-\frac{\rho^2}{2}} \rho d\rho.$$

Последний интеграл легко вычисляется. Получим окончательный ответ:

$$P((X, Y) \in \mathfrak{I}_t) = 1 - e^{-\frac{t^2}{2}}.$$

С ростом t эта вероятность быстро приближается к единице. Легко также найти *доверительные эллипсы* (аналог доверительных интервалов в одномерном случае): если задать доверительную вероятность γ и приравнять $P((X, Y) \in \mathfrak{I}_t) = \gamma$, то, очевидно, $t(\gamma) = \sqrt{-2\ln(1-\gamma)}$, а полуоси эллипса равны $t(\gamma)\sigma_x$ и $t(\gamma)\sigma_y$.

Замечание. Эллипс, соответствующий $t = 1$, назовем *основным*.

Б. Чтобы перейти к общему случаю, требуется выполнить сначала поворот осей координат.

Пусть система координат $OX'Y'$ повернута на острый угол α (в положительном направлении) относительно системы $OX Y$. Из курса линейной алгебры известно, что в этом случае «старые» и «новые» координаты любой точки

связаны равенствами
$$\begin{cases} x = x' \cos \alpha - y' \sin \alpha, \\ y = x' \sin \alpha + y' \cos \alpha. \end{cases}$$
 Так как координаты точки одновременно являются значениями пары случайных величин, то таким же будет соотношение пар (X, Y) и (X', Y') :

$$\begin{cases} X = X' \cos \alpha - Y' \sin \alpha, \\ Y = X' \sin \alpha + Y' \cos \alpha. \end{cases}$$

Предположим, что пара (X', Y') удовлетворяет описанным выше условиям: $X' \in N(0, \sigma_{x'})$, $Y' \in N(0, \sigma_{y'})$, причем эти величины независимы. Это условие позволяет применить теоремы сложения как для математических ожиданий, так и для дисперсий (II, 2): $M(X) = M(Y) = 0$, $\sigma_x^2 = \sigma_{x'}^2 \cos^2 \alpha + \sigma_{y'}^2 \sin^2 \alpha$, $\sigma_y^2 = \sigma_{x'}^2 \sin^2 \alpha + \sigma_{y'}^2 \cos^2 \alpha$. Далее, учитывая, что $M(X'Y') = M(X')M(Y') = 0$, получим: $K(X, Y) = M(XY) = M(\sin \alpha \cos \alpha (X'^2 - Y'^2)) = \sin \alpha \cos \alpha (\sigma_{x'}^2 - \sigma_{y'}^2)$,

$$r = r(X, Y) = \frac{\sin \alpha \cos \alpha (\sigma_x^2 - \sigma_y^2)}{\sqrt{(\sigma_x^2 \cos^2 \alpha + \sigma_y^2 \sin^2 \alpha)(\sigma_x^2 \sin^2 \alpha + \sigma_y^2 \cos^2 \alpha)}}.$$

Оказывается (убедитесь в этом самостоятельно), из последнего равенства легко получить простое соотношение между дисперсиями всех четырех величин:

$$1 - r^2 = \frac{\sigma_x^2 \sigma_y^2}{\sigma_x^2 \sigma_y^2}.$$

Приведем к координатам (x, y) функции $f(x', y') = \frac{1}{2\pi\sigma_{x'}\sigma_{y'}} e^{-L(x', y')}$,

$L(x', y') = \frac{1}{2} \left(\frac{x'^2}{\sigma_{x'}^2} + \frac{y'^2}{\sigma_{y'}^2} \right)$. Из только что доказанного равенства вытекает, что

$\frac{1}{2\pi\sigma_{x'}\sigma_{y'}} = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}$. В выражение для L подставим формулы *обратного*

поворота: $L(x, y) = \frac{1}{2} \left(\frac{(x \cos \alpha + y \sin \alpha)^2}{\sigma_{x'}^2} + \frac{(-x \sin \alpha + y \cos \alpha)^2}{\sigma_{y'}^2} \right)$. Используя

выведенные выше формулы, можно получить:

$$L(x, y) = \frac{1}{2(1-r^2)} \left(\frac{x^2}{\sigma_x^2} - 2r \frac{x}{\sigma_x} \frac{y}{\sigma_y} + \frac{y^2}{\sigma_y^2} \right).$$

Рассмотрим вопрос об одномерных распределениях и прямых регрессии.

Выделим полный квадрат: $L(x, y) = \frac{1}{2(1-r^2)} \left(\frac{x^2}{\sigma_x^2} (1-r^2) + \left(\frac{y}{\sigma_y} - r \frac{x}{\sigma_x} \right)^2 \right)$. От-

сюда $f(x, y) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-r^2}} e^{-\frac{\left(y - r \frac{\sigma_y}{\sigma_x} x \right)^2}{2(\sigma_y\sqrt{1-r^2})^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{x^2}{2\sigma_x^2}}$, что полностью со-

ответствует формуле $f(x, y) = f_{X=x}(y) \cdot f_X(x)$ из параграфа 4. Делаем выводы:

1. Одномерное распределение величины X есть обычное нормальное распределение: $X \in N(0, \sigma_x)$ разумеется, то же верно и для Y .

2. Условное распределение величины Y при фиксированном значении

$X = x$ есть нормальное распределение с параметрами $M_{X=x}(Y) = r \frac{\sigma_y}{\sigma_x} x$

(то есть линия регрессии Y на X есть прямая $y = r \frac{\sigma_y}{\sigma_x} x$, а линия регрессии X на Y , очевидно, имеет уравнение $x = r \frac{\sigma_x}{\sigma_y} y$) и среднеквадратичным отклонением $\sigma_y \sqrt{1-r^2}$.

Заметим, что эта величина не зависит от значения x . Свойство постоянства дисперсии всех условных распределений называется *гомоскедастичностью*. Заметим, что это свойство сохраняется для любых прямых в плоскости, параллельных одному направлению, т.к. с помощью поворота осей на нужный угол это направление можно сделать координатным.

В. Наконец, с помощью переноса осей (заменяем x на $x - a_x$, y на $y - a_y$) переходим к общему нормальному распределению, характеризующемуся *пятью* параметрами $a_x, a_y, \sigma_x, \sigma_y, r$ и описываемому функциями

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}} e^{-L(x, y)},$$

$$L(x, y) = \frac{1}{2(1-r^2)} \left(\frac{(x-a_x)^2}{\sigma_x^2} - 2r \frac{x-a_x}{\sigma_x} \frac{y-a_y}{\sigma_y} + \frac{(y-a_y)^2}{\sigma_y^2} \right).$$

В курсе линейной алгебры доказывается, что к такому виду можно привести любую линейно-квадратичную форму

$$L(x, y) = Ax^2 + 2Bxy + Cy^2 + Dx + Ey + F,$$

если она *положительно определена*, т.е. если $A > 0, AC > B^2$. После этого можно

найти и значение константы $\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}$.

Задачи

6.1. Убедитесь, что функция двумерной плотности распределения, заданная в задаче 4.6, соответствует случаю Б. На основании этого определить значения величин σ_x, σ_y, r , а также величину угла поворота осей координат, приводящего пару величин (X, Y) к паре независимых нормальных величин (X', Y') . Построить основной эллипс рассеяния.

6.2. Та же задача для функций плотности распределения:

а) $f(x, y) = \frac{1}{\pi} e^{-0,5(5x^2 + 2xy + y^2)}$; б) $f(x, y) = Ce^{-x^2 - 2xy - 4y^2}$ (найти величину C).

Часть IV. Энтропия и информация

Введение

В этой части учебника кратко изложены основы современной математической теории информации.

Информация нужна тому, кто находится в состоянии *неопределенности*. Тому, кто знает все, что ему может понадобиться, информация не нужна.

Получение информации приводит к уменьшению неопределенности. Поэтому информации и неопределенность являются разновидностями (с противоположным знаком) единой величины, которую называют *энтропией*.

Для вывода точной математической меры энтропии следует построить более точную математическую модель явления (то есть системы, обладающей определенной мерой неопределенности). Теория вероятностей как раз и представляет для этой цели необходимые инструменты.

1. Энтропия случайного события

Рассмотрим сначала простейшую модель, описываемую как схема *классической вероятности*.

А. Пусть имеется система S , которую можно обнаружить в одном из N равновозможных состояний (система рассматривается в момент, когда состояние не определялось или просто неизвестно нам). Эту же ситуацию в терминах теории вероятностей можно описать как испытание, имеющее N равновозможных исходов. Введем функцию *энтропии* системы как функцию от N (обозначим ее $H(S) = \varphi(N)$), которая должна удовлетворять следующим аксиомам.

1. $\varphi(1) = 0$ (система, заведомо находящаяся в определенном состоянии, не имеет энтропии).
2. С ростом N $\varphi(N)$ возрастает.
3. $\varphi(MN) = \varphi(M) + \varphi(N)$.

Здесь нужны пояснения, сначала – на примере. Информация о карте, случайно извлеченной из колоды (36 равновозможных вариантов) состоит из наименования масти (4 варианта) и достоинства карты (9 вариантов). Естественно считать, что та и другая информация *суммируются*. В общем виде ситуацию можно понимать как сложное испытание, состоящее из двух независимых простых испытаний.

Очевидно, что свойствам 1 – 3 удовлетворяет функция $\varphi(N) = \log_a N$ при любом основании $a > 1$. Для уточнения определим, что

4. $\varphi(2) = 1$.

Это значит, что за единицу энтропии (эту единицу называют *битом*) выбирается энтропия простейшего испытания с двумя равновероятными исходами (подбрасывание монеты). Таким образом,

$$H(S) = \varphi(N) = \log_2 N.$$

Поскольку основание логарифма в дальнейшем меняться не будет, то будем (как это и принято в теории информации) применять сокращенное обозначение $\log$ вместо $\log_2$.

Б. Теперь рассмотрим случай *двух неравновероятных* исходов испытания. Пусть, согласно схеме классической вероятности, событию A благоприятствуют M исходов из N возможных, т.е. $p = P(A) = \frac{M}{N}$. Это значит: *если* в результате испытания событие A произошло, то оставшаяся энтропия равна $\log M$, то есть она *уменьшилась* на величину $\Delta H = \log N - \log M = -\log \frac{M}{N} = -\log p$ (знак «минус» здесь и далее не должен нас беспокоить: вероятности меньше единицы, то есть их логарифмы отрицательны).

Пример. Вы ничего не знаете относительно дня рождения девушки, с которой Вас знакомят (энтропия $\log 365$), но, не успев задать вопрос, замечаете у нее кулон с изображением, скажем, символа созвездия Льва. Это сразу уменьшает энтропию до $\log 31$, то есть Вы получили информацию, равную $-\log \frac{31}{365}$. Не удивительно, что это число практически совпадает с $\log 12$ (выбор одного из 12 практически равновероятных вариантов).

Однако могло произойти и противоположное событие $\bar{A}$ — этому случаю соответствует изменение энтропии $\Delta H = -\log(1-p)$. Таким образом, *изменение энтропии* между состояниями «до испытания» и «после испытания» можно рассматривать как случайную величину, заданную рядом распределения

ΔH	$-\log p$	$-\log(1-p)$
P	p	$1-p$

Определение. Энтропией испытания с двумя возможными исходами (иначе говоря, энтропией случайного события, вероятность которого известна) называется *среднее* значение изменения энтропии, т.е.

$$H(S) = H(A) = M(\Delta H) = -p \log p - (1-p) \log(1-p).$$

Замечания. 1. Естественно распространить это определение на случай *любого*, а не только рационального значения вероятности p события A .

2. Эта величина есть *уменьшение* энтропии в результате проведенного испытания с известным результатом (A или $\bar{A}$). Поэтому она же называется сред-

ней информацией $I(S)$.

3. Удобно ввести специальную функцию $\eta(p) = -p \log p$, определенную и положительную на интервале $(0; 1)$. Тогда $H(A) = \eta(p) + \eta(1-p)$. Функция η :

достигает максимального значения $\frac{1}{e}$ при $p = \frac{1}{e}$;

имеет предел 0 на обоих концах интервала (легко проверить, применяя правило Лопиталя), что позволяет доопределить $\eta(0) = \eta(1) = 0$.

Для вычисления энтропии и информации в конкретных случаях можно использовать специальную таблицу для функции $\eta(p)$.

4. Из предыдущего замечания вытекает, что при $p=0$ и $p=1$ будет $H(A)=0$ (испытание с одним возможным исходом энтропии не имеет). В остальных случаях $H(A) > 0$. Найдем точку максимума, вычислив производную

по переменной p : $\frac{d}{dp} H(A) = -\log p - 1 + \log(1-p) + 1 = \log \frac{1-p}{p}$. Приравняв

эту производную к нулю, получим, очевидно, $p = \frac{1}{2}$ и $H(A) = 1$. Таким образом, наибольшая энтропия случайного события достигается при его вероятности, равной $\frac{1}{2}$, и эта наибольшая энтропия равна 1, как и предполагалось выше.

5. Из определения очевидно, что H как функция от p симметрична: $H(A) = H(\bar{A})$.

Задачи

1.1. Определите максимально возможную энтропию системы, состоящей из четырех элементов, каждый из которых может быть в четырех состояниях равновероятно.

1.2. В урне два белых и три черных шара. Из урны вынимают подряд два шара. Найдите энтропию появления двух белых шаров.

1.3. Какое количество информации несут сообщения:

а) «бубны – козыри»;

б) «на двух костях, брошенных одновременно, выпало 8 очков»;

в) «монета выпала орлом в трех бросаниях из 8»?

2. Энтропия дискретной случайной величины

Очевидно, что для дискретной случайной величины, принимающей только два значения (то есть имеющей ряд распределения такого вида),

X	x_1	x_2
P	p	$1-p$

энтропия вычисляется так же, как и в предыдущем параграфе. Действительно, в результате испытания происходит либо событие $A = (X = x_1)$, либо событие $\bar{A} = (X = x_2)$. Без подробного обоснования (оно по существу не отличается от этого случая) дадим

Определение. Пусть дискретная случайная величина X задана рядом распределения

X	x_1	x_2	$\dots$	x_i	$\dots$
P	p_1	p_2	$\dots$	p_i	$\dots$

($\sum_i p_i = 1$). Ее энтропией называется величина

$$H(X) = -\sum_i p_i \log p_i = \sum_i \eta(p_i).$$

Замечания. 1. Энтропия всегда положительна, за исключением вырожденного случая, когда величина X принимает только одно значение, т.е. не является случайной.

2. Как видим, сами значения x_i при вычислении энтропии не используются. Действительно, информация о результате «розыгрыша» случайной величины состоит в том, какое из *заранее известных* возможных значений она приняла в результате испытания, то есть само значение информации не несет. Следствие: если две случайные величины имеют один и тот же набор вероятностей p_i (даже в разном порядке), то их энтропии совпадают.

3. Если первую строку ряда распределения X заменить на значения $-\log p_i$, то можно записать $H(X) = -M(\log p)$. Впрочем, эта запись не вполне корректна, так как среди значений p_i могут встретиться одинаковые, но объединять «значения» величины $\log p$ с суммированием их вероятностей в данном случае нельзя — это приводит к уменьшению энтропии (докажите!).

4. Как и для случая двух значений, докажем, что при данном числе n значений случайной величины наибольшее значение энтропии достигается, если все вероятности равны: $p_i = \frac{1}{n}$ (это значение равно $\log n$).

Пусть $i > 1$. Для вычисления частной производной по выбранной переменной $p_i = t$ фиксируем остальные переменные, *кроме* одной (пусть это будет

p_1), т.к. сумма всех вероятностей должна сохраняться. Поэтому

$p_1 = 1 - t - \sum_{\substack{j=2 \\ j \neq i}}^n p_j = b - t$. Теперь можно записать $H(X) = a + \eta(t) + \eta(b - t)$ (a и

b – константы). Приравняв нулю производную по t , как и в предыдущем параграфе, получим равенство $t = p_i = p_1$. То есть все p_i равны, что и требовалось доказать.

Задачи

- 2.1. Алфавит состоит из букв a, b, c, d. Вероятности появления букв равны соответственно 0,25; 0,25; 0,34; 0,16. Определите количество информации, приходящееся на символ сообщения, составленного с помощью такого алфавита
- 2.2. Найти энтропию случайной величины, распределенной по биномиальному закону (см. II, 3) при $n = 5$, $p = \frac{1}{3}$ (см. I, 7). Почему она не совпадает с суммой энтропий для каждого испытания?
- 2.3. Найти энтропию случайной величины – числа очков на двух костях, брошенных одновременно.
- 2.4. В двух корзинах имеется по 15 яблок, причем в первой урне – 5 красных, 7 белых и 3 черных, а во второй соответственно 4,4 и 7. Из каждой корзины вынимается по одному яблоку. Определите, для какой из корзин исход опыта является более определенным.
- 2.5. Для случайной величины, распределенной по геометрическому закону (см. II, 3) доказать равенство $H(X) = -\log q - \frac{q}{p} \log p$.
- 2.6. Доказать, что при объединении значений дискретной случайной величины ее энтропия уменьшается.

3. Условная энтропия. Взаимная информация

Пусть пара дискретных случайных величин (X, Y) задана таблицей совместного распределения, то есть набором вероятностей p_{ij} (см. (III,1)). По аналогии с предыдущим параграфом введем понятие *совместной энтропии* пары (X, Y) :

$$H(X, Y) = - \sum_{i,j} p_{ij} \log p_{ij} = \sum_{i,j} \eta(p_{ij}).$$

Сумма фактически распространяется только на те пары значений (x_i, y_j) , которые могут реализоваться, то есть для которых $p_{ij} > 0$ (напомним: мы приняли, что $\eta(0) = 0$).

В (III,1) обосновано, что при фиксированном значении одной из величин (например, при $X = x_i$) условное распределение величины Y задается значениями y_j и соответствующими им вероятностями $\frac{p_{ij}}{p_{ix}}$ (i фиксировано). Это позволяет ввести для этого значения x_i условную энтропию

$$\begin{aligned} H_{X=x_i}(Y) &= \sum_j \eta\left(\frac{p_{ij}}{p_{ix}}\right) = \sum_j \frac{p_{ij}}{p_{ix}} \log \frac{p_{ij}}{p_{ix}} = \frac{1}{p_{ix}} \left(\sum_j p_{ij} (\log p_{ij} - \log p_{ix}) \right) = \\ &= \frac{1}{p_{ix}} \left(\sum_j \eta(p_{ij}) - \log p_{ix} \sum_j p_{ij} \right) = \frac{1}{p_{ix}} \sum_j \eta(p_{ij}) - \log p_{ix}. \end{aligned}$$

Затем, как и ранее, находим *среднее* значение этой энтропии по всем значениям x_i , которую и назовем *полной условной энтропией*:

$$H_X(Y) = \sum_i p_{ix} H_{X=x_i}(Y) = \sum_{i,j} \eta(p_{ij}) - \sum_i \eta(p_{ix}).$$

Эту величину можно трактовать как остаточную неопределенность знаний о величине Y , если значение X известно. Например, если X, Y независимы, то $p_{ij} = p_{ix} p_{jy}$, отсюда очевидно, что $H_{X=x_i}(Y) = H_X(Y) = H(Y)$, то есть знание значения одной величины не меняет неопределенности второй.

Формулы, написанные выше, дают равенство

$$H_X(Y) = H(X, Y) - H(X).$$

Очевидно, что по аналогии будет

$$H_Y(X) = H(X, Y) - H(Y),$$

что в целом дает теорему *сложения* энтропий:

$$H(X, Y) = H(X) + H_Y(X) = H(Y) + H_X(Y).$$

(энтропия системы величин равна сумме энтропии одной из них и условной энтропии второй при известной первой). Это правило можно распространить на любое количество связанных случайных величин.

Из этих равенств вытекает также равенство

$$H(X) - H_Y(X) = H(Y) - H_X(Y).$$

Такая разность выражает уменьшение неопределенности одной из связанных величин в связи со знанием значения второй и поэтому называется *взаимной информацией* $I(X \leftrightarrow Y)$ или просто $I(X, Y)$. Ее можно также записать как

$$I(X, Y) = H(X) + H(Y) - H(X, Y).$$

Очевидно, что для независимых X, Y $I(X, Y) = 0$. Верна и обратная

Теорема. $I(X, Y) \geq 0$, причем равенство достигается *только* для независимых величин X, Y .

Прежде чем доказывать теорему, сформулируем

Следствие. $H(X, Y) \leq H(X) + H(Y)$ (неопределенность совокупности случайных величин никогда не превышает сумму их неопределенностей).

Доказательство теоремы.

$$\begin{aligned} I(X, Y) &= H(X) + H(Y) - H(X, Y) = -\sum_i p_{ix} \log p_{ix} - \sum_j p_{jy} \log p_{jy} + \sum_{i,j} p_{ij} \log p_{ij} = \\ &= -\sum_i \sum_j p_{ij} \log p_{ix} - \sum_j \sum_i p_{ij} \log p_{jy} + \sum_{i,j} p_{ij} \log p_{ij} = \sum_{i,j} p_{ij} \log \frac{p_{ij}}{p_{ix} p_{jy}} = \\ &= -\log e \sum_{i,j} p_{ij} \ln \frac{p_{ix} p_{jy}}{p_{ij}}. \end{aligned}$$

Функция $\psi(x) = x - 1 - \ln x$ при $x > 0$ имеет единственный нулевой минимум при $x = 1$ (легко проверить, что $\psi(1) = \psi'(1) = 0$ и что $\psi''(x) > 0$). Отсюда $-\ln x \geq x - 1$, причем равенство достигается только при $x = 1$. Таким образом, если выполняется условие независимости: $p_{ij} = p_{ix} p_{jy}$ *при всех* i, j , то $I(X, Y) = 0$, в противном случае

$$I(X, Y) > \log e \sum_{i,j} p_{ij} \left(\frac{p_{ix} p_{jy}}{p_{ij}} - 1 \right) = \log e \left(\sum_{i,j} p_{ix} p_{jy} - \sum_{i,j} p_{ij} \right) = \log e (1 \cdot 1 - 1) = 0.$$

Простейшим частным случаем условной и взаимной информации является информационная связь между случайными событиями. Как указано в конце параграфа III, 3, ее можно свести к связи между индикаторными случайными величинами. Так, для примера, приведенного в этом параграфе, вычисляем последовательно:

$$\begin{aligned} H(A) &= H(X_A) = H(B) = H(X_B) = 2\eta(0,5) = 1, \\ H(A, B) &= H(X_A, X_B) = 2\eta(0,1) + 2\eta(0,4) = 1,722, \\ H_A(B) &= H_B(A) = 0,722, \\ I(A, B) &= 0,278. \end{aligned}$$

Задачи

- 3.1. Каждый четвертый мужчина — блондин, у каждого десятого голубые глаза, четверо из ста обладают обоими признаками. Найти условную энтропию каждого из двух признаков и их взаимную информацию.
- 3.2. Найти условные энтропии и величину взаимной информации двух случайных величин по данной таблице их совместного распределения:

$X \backslash Y$	-1	1
0	0	0,1
1	0,4	0,2
2	0,3	0

- 3.3. Можно ли найти значения $H_X(Y)$ и $I(X, Y)$, если известны $H(X)$, $H(Y)$ и $H_Y(X)$?
- 3.4. В условиях задачи 1.1 части III найти взаимную информацию двух случайных величин.
- 3.5. Пусть величина Y является функцией величины X (то есть значение Y полностью определяется значением X ; также говорят, что X обуславливает Y). Доказать для этого случая равенства $H(X, Y) = H(X)$, $H_X(Y) = 0$, $I(X, Y) = H(Y)$ и объяснить их смысл.
- 3.6. В условиях предыдущей задачи рассмотреть случай *взаимно-однозначного соответствия* двух величин.
- 3.7. Пусть величина X принимает с равной вероятностью значения от 1 до $2n$, а величина Y равна 0, если значение X нечетно, и 1, если оно четно. Найти значения величин $H(X)$, $H(Y)$, $H(X, Y)$, $H_X(Y)$, $H_Y(X)$, $I(X, Y)$.

4. Энтропия непрерывных величин. Интервал неразличимости

Перейдем к рассмотрению случая *непрерывной* случайной величины X , заданной функцией плотности вероятностей $f(x)$, причем эта функция отлична от 0 только на отрезке $[a; b]$. Разобьем этот отрезок на *равные* части длиной Δx и рассмотрим *дискретную* случайную величину $\bar{X}$, которая в пределах каждого отрезка разбиения принимает определенное значение $\bar{x}_i$, выбранное на этом интервале (напомним, что сами значения ДСВ при вычислении энтропии вообще не используются). Вероятности же принятия каждого из этих значений приближенно равны $f(\bar{x}_i)\Delta x$. Поэтому

$$\begin{aligned} H(X) &\approx H(\bar{X}) \approx \sum_i \eta(f(\bar{x}_i)\Delta x) = -\sum_i f(\bar{x}_i)\Delta x \log(f(\bar{x}_i)\Delta x) = \\ &= -\sum_i f(\bar{x}_i) \log(f(\bar{x}_i)) \Delta x - \log \Delta x \sum_i f(\bar{x}_i) \Delta x. \end{aligned}$$

При достаточно малом Δx обе интегральные суммы можно заменить на соот-

ветствующие определенные интегралы. Учитывая, что $\int_a^b f(x)dx = 1$, получим:

$$H(X) \approx -\int_a^b f(x) \log f(x) dx - \log \Delta x.$$

Результат выглядит парадоксальным: известно, что для получения окончательной точной формулы требуется устремить Δx к нулю, но тогда последнее слагаемое неограниченно растет, т.е. $\lim_{\Delta x \rightarrow 0} H(\bar{X}) = +\infty$, и первое слагаемое вообще смысла не имеет.

Суть дела состоит в том, что *полная информация* о непрерывной случайной величине действительно неограниченно велика. Она подразумевает неограниченно точное измерение значений величины, полученных в результате испытания (или, с точки зрения функций распределения, абсолютно точное знание любых значений $f(x)$). На практике любые *непрерывные* величины известны с определенной точностью, то есть существует разумный нижний предел для Δx . Этот предел и называют *интервалом неразличимости*. Обозначив его просто Δ , получим практически применимую формулу $H(X) = H^*(X) - \log \Delta$, где

$$H^*(X) = -\int_a^b f(x) \log f(x) dx = \int_a^b \eta(f(x)) dx$$

(назовем эту величину *приведенной энтропией*).

Заметим еще, что в реальных приложениях случайная величина X является величиной именованной, то есть выраженной в некоторых единицах. Тогда под интервалом неразличимости следует понимать наименьший интервал, который может быть измерен. Если X и Δ измерять именно в таких единицах, то $\Delta \approx 1$, и разница между H и H^* исчезает. Впрочем, единица измерения может быть и другой.

Как и для случая вычисления математического ожидания и дисперсии, интеграл можно распространить на всю действительную ось:

$$H^*(X) = \int_{-\infty}^{+\infty} \eta(f(x)) dx$$

(напомним: мы считаем, что $\eta(0) = 0$).

Пример. Рассмотрим случай величины, распределенной равномерно на отрезке $[a; b]$ (см. II, 6). Очевидно, в этом случае

$$H^*(X) = -\int_a^b \frac{1}{b-a} \log \frac{1}{b-a} dx = \log(b-a).$$

Для случая двумерного непрерывного распределения аналогичные рассуждения приводят к формуле

$$H(X, Y) = H^*(X, Y) - \log \Delta_x - \log \Delta_y,$$

где

$$H^*(X, Y) = \iint_{R^2} \eta(f(x, y)) dx dy.$$

Интервалы неразличимости Δ_x и Δ_y могут не совпадать. Более того, X и Y могут быть величинами, измеряемыми в разных единицах (например, температура воздуха и атмосферное давление).

Рассмотрим для случая непрерывных величин формулы для условной энтропии и взаимной информации:

$$H_X(Y) = H(X, Y) - H(X) = H^*(X, Y) - H^*(X) - \log \Delta_y.$$

Отбросив последнее слагаемое, получим

$$H_X^*(Y) = H^*(X, Y) - H^*(X).$$

Так как

$$\begin{aligned} H^*(X) &= - \int_{-\infty}^{+\infty} f_X(x) \log f_X(x) dx = - \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(x, y) dy \right) \log f_X(x) dx = \\ &= - \iint_{R^2} f(x, y) \log f_X(x) dx dy, \end{aligned}$$

то

$$H_X^*(Y) = - \iint_{R^2} f(x, y) \log \frac{f(x, y)}{f_X(x)} dx dy.$$

Для взаимной информации $I(X, Y)$ использование величин H или H^* безразлично! Получаем окончательно

$$\begin{aligned} I(X, Y) &= H(X) + H(Y) - H(X, Y) = H^*(X) + H^*(Y) - H^*(X, Y) = \\ &= \iint_{R^2} f(x, y) \log \frac{f(x, y)}{f_X(x) f_Y(y)} dx dy. \end{aligned}$$

Аналогично дискретному случаю доказывается, что эта величина положительна, за исключением того случая, когда величины X, Y независимы.

Задачи

4.1. Доказать, что для величины, распределенной по показательному закону с параметром λ (см. II, 6), $H^*(X) = (1 - \ln \lambda) \log e = \log \frac{e}{\lambda}$.

4.2. Доказать, что для величины, распределенной по нормальному закону с параметрами a, σ (см. II, 6), $H^*(X) = \left(\frac{1}{2} + \ln(\sigma \sqrt{2\pi}) \right) \log e$.

4.3. Дана функция «треугольного» распределения

$$f(x) = \begin{cases} \frac{2}{a^2}(a-x), & x \in (0; a), \\ 0, & x \notin (0; a). \end{cases}$$

Показать, что для случайной величины X , имеющей эту плотность распределения, $H^*(X) = \left(\frac{1}{2} - \ln \frac{2}{a}\right) \log e$.

4.4. Показать, что для пары случайных величин, равномерно распределенной в ограниченной области D , $H^*(X, Y) = \log S(D)$, где $S(D)$ – площадь области.

4.5. Опираясь на результаты двух последних задач, найти взаимную информацию пары случайных величин, заданных в задаче III, 4.4б).

Приложения

I. Типовой расчет «Двумерные случайные величины»

Задача 1

Независимые случайные величины U, V даны своими рядами распределения:

U	-1	0	1	2
P	0,4	0,3	0,2	0,1

V	0	2
P	0,4	0,6

Составить таблицу совместного распределения случайных величин X и Y , которые выражаются через U, V по указанным формулам (для выбора варианта см. таблицу 1). Найти ковариацию, коэффициент корреляции. Построить ломаные регрессии и прямые регрессии.

1. U .
2. V .
3. $|U|$.
4. U^2 .
5. $U + V$.
6. $U - V$.
7. $\min(U, V)$.
8. $\max(U, V)$.
9. $U^2 + V$.
10. $U^2 - V$.
11. $|U| + V$.
12. $|U| - V$.

Задача 2

Векторная случайная величина (X, Y) распределена равномерно в пределах заданного треугольника ABC . Найти одномерные функции плотности распределения. Построить прямые регрессии Y на X и X на Y , найти коэффициент корреляции.

Таблица 1

ЛН	X	Y	A	B	C
1	1	7	0,0	0,-4	-3,-1
2	1	8	0,0	-4,0	0,-1
3	1	9	0,0	0,-3	-4,-3
4	1	10	0,0	-2,-3	-2,2
5	1	11	0,0	0,1	4,1
6	1	12	-1,0	0,6	-1,6
7	2	7	0,0	1,-3	0,-4
8	2	8	3,0	3,-2	0,-2
9	2	9	0,0	2,0	0,-3
10	2	10	3,0	3,4	0,4
11	2	11	0,0	1,-2	-3,-2
12	2	12	0,0	0,3	2,3
13	3	5	0,0	2,0	2,5
14	3	6	0,0	0,5	-2,3
15	3	7	-3,0	0,-2	-3,-2
16	3	8	0,0	0,3	-1,3
17	3	9	0,0	2,1	0,3
18	3	10	0,0	1,1	1,-3
19	4	5	0,0	-4,0	-4,1
20	4	6	0,0	2,0	0,-5
21	4	7	0,0	-5,0	0,-2
22	4	8	0,0	-4,0	0,1
23	4	11	0,0	2,0	-1,2
24	4	12	0,0	2,0	0,3

В задачах 1, 2: вычислить энтропии

$H(X), H(Y), H(X, Y), H_Y(X), H_X(Y)$, проверить выполнение формулы сложения энтропий.

II. Таблица значений функции $\eta(p) = -p \log_2 p$

0,01	0,0664		0,51	0,4954
0,02	0,1129		0,52	0,4906
0,03	0,1518		0,53	0,4854
0,04	0,1858		0,54	0,4800
0,05	0,2161		0,55	0,4744
0,06	0,2435		0,56	0,4684
0,07	0,2686		0,57	0,4623
0,08	0,2915		0,58	0,4558
0,09	0,3127		0,59	0,4491
0,10	0,3322		0,60	0,4422
0,11	0,3503		0,61	0,4350
0,12	0,3671		0,62	0,4276
0,13	0,3826		0,63	0,4199
0,14	0,3971		0,64	0,4121
0,15	0,4105		0,65	0,4040
0,16	0,4230		0,66	0,3956
0,17	0,4346		0,67	0,3871
0,18	0,4453		0,68	0,3783
0,19	0,4552		0,69	0,3694
0,20	0,4644		0,70	0,3602
0,21	0,4728		0,71	0,3508
0,22	0,4806		0,72	0,3412
0,23	0,4877		0,73	0,3314
0,24	0,4941		0,74	0,3215
0,25	0,5000		0,75	0,3113
0,26	0,5053		0,76	0,3009
0,27	0,5100		0,77	0,2903
0,28	0,5142		0,78	0,2796
0,29	0,5179		0,79	0,2687
0,30	0,5211		0,80	0,2575
0,31	0,5238		0,81	0,2462
0,32	0,5260		0,82	0,2348
0,33	0,5278		0,83	0,2231
0,34	0,5292		0,84	0,2113
0,35	0,5301		0,85	0,1993
0,36	0,5306		0,86	0,1871
0,37	0,5307		0,87	0,1748
0,38	0,5305		0,88	0,1623
0,39	0,5298		0,89	0,1496
0,40	0,5288		0,90	0,1368
0,41	0,5274		0,91	0,1238
0,42	0,5256		0,92	0,1107
0,43	0,5236		0,93	0,0974
0,44	0,5211		0,94	0,0839
0,45	0,5184		0,95	0,0703
0,46	0,5153		0,96	0,0565
0,47	0,5120		0,97	0,0426
0,48	0,5083		0,98	0,0286
0,49	0,5043		0,99	0,0144
0,50	0,5000		1,00	0,0000

Оглавление

Часть III. Двумерные случайные величины

Введение	2
1. Дискретная двумерная величина	3
2. Регрессия	4
3. Ковариация и корреляция	5
4. Функции распределения	9
5. Функции случайных величин. Свертка	13
6. Двумерное нормальное распределение	15

Часть IV. Энтропия и информация

Введение	19
1. Энтропия случайного события	19
2. Энтропия дискретной случайной величины	22
3. Условная энтропия. Взаимная информация	23
4. Энтропия непрерывных величин. Интервал неразличимости	26
Приложения	30