Лекция 6. Совместное распределение двух дискретных случайных величин

Курбацкий А. Н.

МШЭ МГУ

23 октября 2020

Содержание

Совместное распределение

Определение

Говорят, что задано совместное распределение двух дискретных случайных величин, измеряемых в одном и том же случайном эксперименте, если для каждой пары значений этих величин (x_i, y_j) задана вероятность $P(X = x_i, Y = y_j) = p_{ij}$, где x_i, $i = 1, \ldots, n$ - множество возможных значений X, а y_j , $j = 1, \ldots, m$ - множество всех возможных значений Y.

Совместное распределение двух дискретных случайных величин удобно записывать в виде таблицы:

$X \setminus Y$	y_1	y_2	$\ldots$	y_m
x_1	p_{11}	p_{12}	$\ldots$	p_{1m}
x_2	p_{21}	p_{22}	$\ldots$	p_{2m}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
x_n	p_{n1}	p_{n2}	$\ldots$	p_{nm}

Сумма всех значений p_{ij} : $\sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij} = 1$

Попробуйте самостоятельно!

Пример

Игральную кость бросают два раза. Определим случайную величину X *как число выпавших шестерок. Случайная величина* Y *будет принимать значение 0, если хотя бы на одной кости выпадет нечетное, и 1, если на обеих костях выпадет четное. Выпишем таблицу совместного распределения случайных величин* X *и* Y.

Попробуйте самостоятельно!

Пример

Игральную кость бросают два раза. Определим случайную величину X *как число выпавших шестерок. Случайная величина* Y *будет принимать значение 0, если хотя бы на одной кости выпадет нечетное, и 1, если на обеих костях выпадет четное. Выпишем таблицу совместного распределения случайных величин* X *и* Y*.*

Попробуйте самостоятельно!

Пример

Игральную кость бросают два раза. Определим случайную величину X как число выпавших шестерок. Случайная величина Y будет принимать значение 0, если хотя бы на одной кости выпадет нечетное, и 1, если на обеих костях выпадет четное. Выпишем таблицу совместного распределения случайных величин X и Y.

Случайная величина X может принимать три значения - 0, 1, 2, а случайная величина Y - два значения 0, 1. Следовательно, таблица совместного распределения будет иметь вид:

$X \setminus Y$	0	1
0	p_{11}	p_{12}
1	p_{21}	p_{22}
2	p_{31}	p_{32}

Осталось вычислить все вероятности $p_{ij} = P(X = x_i, Y = y_j)$.

Решение

- Легко видеть, что $p_{32} = P(X = 2, Y = 1) = 1/36$, так как ей соответствует только элементарный исход ω_e= $\{6, 6\}$.
- $p_{31} = P(X = 2, Y = 0) = 0$.
- $p_{22} = P(X = 1, Y = 1) = 4/36 = 1/8$, так как этому событию соответствует 4 элементарных исхода: $\{6,2\}$, $\{6,4\}$, $\{2,6\}$, $\{4,6\}$.
- $p_{21} = P(X = 1, Y = 0) = 6/36 = 1/6$, так как этому событию удовлетворяют исходы: $\{6,1\}$, $\{6,3\}$, $\{2,5\}$, $\{1,6\}$, $\{3,6\}$, $\{5,6\}$.
- $p_{12} = P(X = 0, Y = 1) = 4/36 = 1/8$, так как этому событию соответствуют исходы: $\{2,2\}$, $\{2,4\}$, $\{4,2\}$, $\{4,4\}$.
- $p_{11} = P(X = 0, Y = 0)$ может быть вычислено как $1 - p_{12} - p_{21} - p_{22} - p_{31} - p_{32} =$ $1 - 4/36 - 6/36 - 4/36 - 0 - 1/36 = 21/36$.

Итого, таблица совместного распределения X и Y имеет вид:

$X \setminus Y$	0	1
0	$21/36$	$4/36$
1	$6/36$	$4/36$
2	0	$1/36$

Содержание

Маргинальные распределения

Из таблицы совместного распределения двух дискретных случайных величин можно получить распределение каждой из случайных величин X и Y. Такие распределения называются маргинальными или частными.

Определение

Для того, чтобы получить маргинальное распределение X, то есть найти вероятность $P(X = x_i)$, надо просуммировать вероятности в i-ой строке таблицы совместного распределения:

$$P(X = x_i) = p_{i1} + p_{i2} + \ldots + p_{im}.$$

Для того, чтобы получить маргинальное распределение Y, то есть найти вероятности $P(Y = y_i)$, надо просуммировать вероятности в j-ом столбце таблицы совместного распределения:

$$P(Y = y_j) = p_{1j} + p_{2j} + \ldots + p_{nj}.$$

Пример

Задача

Рассмотрим совместное распределение X и Y из предыдущего примера

$X \setminus Y$	0	1
0	$21/36$	$4/36$
1	$6/36$	$4/36$
2	0	$1/36$

Найти маргинальные распределения случайных величин X и Y.

Пример

Задача

Рассмотрим совместное распределение X и Y из предыдущего примера

$X \setminus Y$	0	1
0	$21/36$	$4/36$
1	$6/36$	$4/36$
2	0	$1/36$

Найти маргинальные распределения случайных величин X и Y.

Решение

Маргинальные распределения X и Y задаются таблицами:

X	0	1	2
p	$25/36$	$10/36$	$1/36$

Y	0	1
p	$27/36$	$9/36$

Содержание

Понятие независимости

Теорема

Дискретные случайные величины X и Y независимы тогда и только тогда, когда для любых i и j, $i = 1, \ldots, n$, $j = 1, \ldots, m$
$p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j)$.

Эту формулу можно использовать для проверки независимости двух дискретных случайных величин.

Пример

Рассмотрим случайные величины X и Y, определенные в примере 1

$X \setminus Y$	0	1
0	21/36	4/36
1	6/36	4/36
2	0	1/36

Являются ли эти случайные величины независимыми?

Решение

Рассмотрим вероятность $p_{11} = P(X = 0, Y = 0) = 21/36$.
Мы уже находили маргинальные распределения X и Y

X	0	1	2
p	25/36	10/36	1/36

Y	0	1
p	27/36	9/36

В частности: $P(X = 0) = 25/36$ и $P(Y = 0) = 27/36$.
Проверим выполнение условий независимости:

$$P(X = 0, Y = 0) = P(X = 0) \cdot P(Y = 0)$$

$$21/36 \neq 25/36 \cdot 27/36.$$

Следовательно, X и Y зависят друг от друга. В этом примере нам повезло, так как проверка ограничилась лишь исследованием равенства для $i = 1$ и $j = 1$. Так бывает не всегда. Если бы условие независимости выполнялось бы для p_{11}, то нам пришлось бы проверять это условие для всех остальных p_{ij}.

Ковариация

Когда случайные величины X и Y зависимы, представляет интерес сила их взаимосвязи. Для этого используется понятие ковариации (то есть совместной вариации) двух случайных величин.

Определение

Ковариацией двух случайных величин X и Y называется $\mathrm{cov}(X,Y) = E[(X - E(X))(Y - E(Y))]$.
На практике для вычисления $\mathrm{cov}(X,Y)$ *чаще используется формула* $\mathrm{cov}(X,Y) = E(X \cdot Y) - E(X) \cdot E(Y)$.

Теорема

- *Если случайные величины X и Y независимы, то* $\mathrm{cov}(X,Y) = 0$;
- *Пусть $X_1 = a_1 + b_1 X$ и $Y_1 = a_2 + b_2 Y$, тогда* $\mathrm{cov}(X_1, Y_1) = b_1 b_2 \,\mathrm{cov}(X,Y)$.

Другими словами, величина ковариации между двумя случайными величинам зависит от их единиц измерения.

Корреляция

Более удобную меру связи двух с. в. даёт коэффициент корреляции.

Определение

Корреляцией двух случайных величин называется:

$$\operatorname{corr}(X, Y) = \frac{\operatorname{cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}.$$

Теорема

- *Если X и Y независимы, то* $\operatorname{corr}(X, Y) = 0$.
- $|\operatorname{corr}(X, Y)| \leq 1$.
- *Если* $\operatorname{corr}(X, Y) = 1$, *то* $Y = a + bX$, *где* $b > 0$.
  Если $\operatorname{corr}(X, Y) = -1$, *то* $Y = a + bX$, *где* $b < 0$.
- *Пусть* $X_1 = a_1 + b_1 X$ *и* $Y_1 = a_2 + b_2 Y$, *тогда*
  $\operatorname{corr}(X_1, Y_1) = \operatorname{corr}(X, Y)$, *если* $b_1 \cdot b_2 > 0$ *и*
  $\operatorname{corr}(X_1, Y_1) = -\operatorname{corr}(X, Y)$, *если* $b_1 \cdot b_2 < 0$.

Пример

Совместное распределение двух дискретных случайных величин X и Y задано таблицей:

$X \setminus Y$	0	1
0	0.1	0.3
1	0.4	

Найти $\operatorname{cov}(X, Y)$ и $\operatorname{corr}(X, Y)$.

Решение. Заполним до конца таблицу совместного распределения $P(X = 1, Y = 1) = 1 - 0.1 - 0.3 - 0.4 = 0.2$. Вычислим маргинальные распределения X и Y:

X	0	1
p	0.4	0.6

Y	0	1
p	0.5	0.5

При этом $E(X) = 0.6$, а $E(Y) = 0.5$.

Найдем распределение случайной величины $X \cdot Y$. Эта величина может принимать только два значения 0 и 1.
$P(X \cdot Y = 1) = P(X = 1, Y = 1) = 0.2$. Следовательно,
$P(X \cdot Y = 0) = 1 - 0.2 = 0.8$ и распределение $X \cdot Y$ задается таблицей:

$X \cdot Y$	0	1
p	0.8	0.2

$\mathrm{cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 0 \cdot 0.8 + 1 \cdot 0.2 - 0.6 \cdot 0.5 = 0.2 - 0.3 = -0.1$.
Для нахождения корреляции, необходимо вычислить дисперсии:

$$D(X) = E(X^2) - [E(X)]^2 = 0.4 \cdot 0.6 = 0.24,$$

$$D(Y) = E(Y^2) - [E(Y)]^2 = 0.5 \cdot 0.5 = 0.25.$$

Отсюда: $\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}} = \frac{-0.1}{\sqrt{0.24}\sqrt{0.25}} \approx -0.4$.

Корреляция и независимость

Важно!

Если случайные величины независимы, то их корреляция равна нулю. А вот обратное утверждать, как правило, нельзя.

Пример

Пусть совместное распределение X и Y задается таблицей:

$X \setminus Y$	-1	0	1
-1	0	0.2	0
0	0.2	0.2	0.2
1	0	0.2	0

Вычислим $\operatorname{cov}(X, Y)$ *и проверим величины на независимость.*

Корреляция и независимость

Важно!

Если случайные величины независимы, то их корреляция равна нулю. А вот обратное утверждать, как правило, нельзя.

Пример

Пусть совместное распределение X *и* Y *задается таблицей:*

$X \setminus Y$	-1	0	1
-1	0	0.2	0
0	0.2	0.2	0.2
1	0	0.2	0

Вычислим $\mathrm{cov}(X, Y)$ *и проверим величины на независимость.*

$$\mathrm{cov}(X, Y) = E(X \cdot Y) - E(X) \cdot E(Y) = 0$$

Убедитесь, что величины не являются независимыми!

Содержание

Условное распределение

Определение

Условным законом распределения с.в. X при условии Y называется любое соотношение, ставящее в соответствие значениям с.в. X условные вероятности их принятия при условии $Y = y$

$$P_{X|Y}(x_i|y_j) = P(X = x_i|Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

Определение

Условное мат. ожидание с.в. X при условии $Y = y$ называется математическое ожидание условного распределения X при условии $Y = y$

$$E(X|Y = y_j) = \sum_i x_i P_{X|Y}(x_i|y_j)$$

Функция регрессии

Определение

Функция регрессии с.в. X по Y называется функция, ставящее в соответствие числу y условное мат. ожидание X при условии y:

$$\varphi_{X|Y}(y) = E(X|Y = y)$$

Функция регрессии характеризует среднее значение одной с.в. при известном значении другой. Если разброс невелик, то это может быть информативно!

Определение

Условным математическим ожиданием X по Y называется называется случайная величина, равная $\varphi_{X|Y}(Y)$, которая обозначается $E(X|Y)$

Пример

Совместный закон распределения с. в. X и Y задан таблицей:

$X \setminus Y$	0	1	3
0	0.15	0.05	0.3
-1	0	0.15	0.1
-2	0.15	0	0.1

Найдите

а) законы распределения случайных величин X и Y;

б) EX, EY, DX, DY, $\mathrm{cov}(X, Y)$, $\mathrm{corr}(X, Y)$, а также математическое ожидание и дисперсию случайной величины $V = 6X - 4Y + 3$.

в) Найдите условное математическое ожидание $E(X|Y = 0)$ и выпишите функцию регрессии $\varphi_{X|Y}(y)$.

Пример

Совместный закон распределения с. в. X и Y задан таблицей:

$X \setminus Y$	0	1	3
0	0.15	0.05	0.3
-1	0	0.15	0.1
-2	0.15	0	0.1

Найдите

а) законы распределения случайных величин X и Y;

б) EX, EY, DX, DY, $\mathrm{cov}(X, Y)$, $\mathrm{corr}(X, Y)$, а также математическое ожидание и дисперсию случайной величины $V = 6X - 4Y + 3$.

в) Найдите условное математическое ожидание $E(X|Y = 0)$ и выпишите функцию регрессии $\varphi_{X|Y}(y)$.

Решение. Для того, чтобы найти законы распределений X и Y нужно просуммировать вероятности по строкам и столбцам соответственно:

X	0	-1	-2
P	0,5	0,25	0,25

Y	0	1	3
P	0,3	0,2	0,5

Зная законы распределений вычисляем математические ожидания, дисперсии и ковариацию:

$$EX = 0\cdot 0.5 - 1\cdot 0.25 - 2\cdot 0.25 = -0.75, \quad EY = 0\cdot 0.3 + 1\cdot 0.2 + 3\cdot 0.5 = 1.7,$$

$$DX = E(X^2) - (EX)^2 = 0^2\cdot 0.5 + (-1)^2\cdot 0.25 + (-2)^2\cdot 0.25 - (-0.75)^2 = \\ = 1.25 - 0.5625 = 0.6875,$$

$$DY = E(Y^2) - (EY)^2 = 0^2\cdot 0.3 + 1^2\cdot 0.2 + 3^2\cdot 0.5 - (1.7)^2 = \\ = 4.7 - 2.89 = 1.81,$$

$$\mathrm{cov}(X;Y) = E(XY) - EX\cdot EY = \\ = -1\cdot 1\cdot 0.15 - 1\cdot 3\cdot 0.1 - 2\cdot 3\cdot 0.1 + 0.75\cdot 1.7 = 0.225.$$

$$\mathrm{corr}(X;Y) = \frac{0.225}{\sqrt{0.6875}\sqrt{1.81}} \approx 0.2.$$

Мат. ожидание и дисперсию с.в. V вычислим по свойствам:

$$EV = E(6X - 4Y + 3) = 6EX - 4EY + 3 = -6 \cdot 0.75 - 4 \cdot 1.7 + 3 = -8.3,$$

$$DV = D(6X - 4Y + 3) = 36DX + 16DY + 2 \cdot 6 \cdot (-4)\,\mathrm{cov}(X; Y) =$$
$$= 36 \cdot 0.6875 + 16 \cdot 1.81 - 48 \cdot 0.225 = 53.71 - 10.8 = 42.91.$$

Составим условное распределение X от Y, пользуясь формулой $P(X|Y = 0) = \frac{P(X \cap Y = 0)}{P(Y = 0)}$:

$$P(X = 0|Y = 0) = \frac{P(X = 0, Y = 0)}{P(Y = 0)} = \frac{0.15}{0.3} = 1/2,$$

$$P(X = -1|Y = 0) = \frac{P(X = -1, Y = 0)}{P(Y = 0)} = 0,$$

$$P(X = -2|Y = 0) = \frac{P(X = -2, Y = 0)}{P(Y = 0)} = \frac{0.15}{0.3} = 1/2$$

Откуда $E(X|Y = 0) = 0 \cdot \frac{1}{2} - 1 \cdot 0 - 2 \cdot \frac{1}{2} = -1.$

Аналогично вычислим условные математические ожидания для $Y = 1$ и $Y = 3$
Для $Y = 1$:

$$P(X = 0|Y = 1) = \frac{P(X = 0, Y = 1)}{P(Y = 1)} = \frac{0.05}{0.2} = 0.25,$$
$$P(X = -1|Y = 1) = \frac{P(X = -1, Y = 1)}{P(Y = 1)} = \frac{0.15}{0.2} = 0.75,$$
$$P(X = -2|Y = 1) = \frac{P(X = -2, Y = 1)}{P(Y = 1)} = \frac{0}{0.2} = 0,$$

Откуда $E(X|Y = 1) = 0 \cdot \frac{1}{2} - 1 \cdot 0.75 - 2 \cdot 0 = -0.75$.

Для $Y = 3$:

$$P(X = 0|Y = 3) = \frac{P(X = 0, Y = 3)}{P(Y = 3)} = \frac{0.3}{0.5} = 0.6,$$

$$P(X = -1|Y = 3) = \frac{P(X = -1, Y = 3)}{P(Y = 3)} = \frac{0.1}{0.5} = 0.2,$$

$$P(X = -2|Y = 3) = \frac{P(X = -2, Y = 3)}{P(Y = 3)} = \frac{0.1}{0.5} = 0.2$$

Откуда $E(X|Y = 3) = 0 \cdot 0.6 - 1 \cdot 0.2 - 2 \cdot 0.2 = -0.6$.
Поэтому функция регрессии имеет вид

Y	0	1	3
$\varphi_{X\|Y}(y)$	-1	-0.75	-0.6

Свойства условного мат.ожидания

Теорема

- $E(C|Y) = C;$
- $E(E(X|Y)) = E(X);$
- $E(aX + b|Y) = aE(X|Y) + b$
- $E(X_1 + X_2|Y) = E(X_1|Y) + E(X_2|Y)$
- $E(h(Y)|Y) = h(Y);$
- $E(h(Y)X|Y) = h(Y)E(X|Y);$
- $E(X|Y) = E(X)$ *для независимых* X *и* Y.
- $\operatorname{cov}(X, g(Y)) = \operatorname{cov}(E(X|Y), g(Y)).$

Свойства условной дисперсии

Теорема

- $D(C|X) = 0$;
- $D(aX + b|Y) = a^2 D(X|Y)$;
- $D(X + h(Y)|Y) = D(X|Y)$;
- $D(h(Y)X|Y) = h^2(Y)D(X|Y)$;
- $D(X) = D(E(X|Y)) + E(D(X|Y))$.