

Лекция 3. Остаток про дискретные с.в. и введение в непрерывные

24 февраля 2021 г.

1 Случайный вектор

Несколько с.в. (случайный вектор)

Совместная функция вероятностей

$$p_{X,Y}(x, y) = \Pr(X = x \cap Y = y)$$

Маргинальные функции вероятностей

$$p_X(x) = \sum_y p_{X,Y}(x, y)$$

$$p_Y(y) = \sum_x p_{X,Y}(x, y)$$

Пример случайного вектора В клетках — $p_{X,Y}(x, y)$

Y	4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
	3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
	2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
	1		$\frac{1}{20}$		
		1	2	3	4

X

Blue arrow pointing down from row 2, column 2 to row 1, column 2: $p_X(2) = \frac{8}{20}$

Red arrow pointing right from row 2, column 2 to row 2, column 5: $p_Y(2) = \frac{5}{20}$

NB: все работает точно также и для большего числа с.в.

2 Линейность матожидания

Линейность матожидания (опять) Теперь у нас есть возможность смотреть $Z = g(X, Y)$

$$p_Z(z) = \Pr(g(X, Y) = z) = \sum_{(x,y): g(x,y)=z} p_{X,Y}(x, y)$$

$$E(Z) = \sum_x \sum_y g(x, y) p_{X,Y}(x, y)$$

Пусть $Z = X + Y$ как вычислить ее матожидание? Хорошо, что оно *линейно*

$$E(Z) = E(X) + E(Y)$$

Линейность матожидания: доказательство

$$\begin{aligned} E(Z) &= E(X + Y) = \sum_x \sum_y (x + y) p_{X,Y}(x, y) \\ &= \sum_x \sum_y x p_{X,Y}(x, y) + \sum_x \sum_y y p_{X,Y}(x, y) \\ &= \sum_x x \sum_y p_{X,Y}(x, y) + \sum_y y \sum_x p_{X,Y}(x, y) \\ &= \sum_x x p_X(x) + \sum_y y p_Y(y) \\ &= E(X) + E(Y) \end{aligned}$$

Окончательная линейность матожидания

$$E\left(\sum_i a_i X_i\right) = \sum_i a_i E(X_i)$$

NB: случайная величина может быть равна *константе*, поэтому мы не рассматриваем прибавление к сумме константы

Матожидание биномиального распределения *Напоминание:* $X \sim \text{Bin}(n, p)$, если X равен числу успехов в n независимых испытаниях Бернулли с вероятностью успеха p .

$X = \sum_{i=1}^n X_i$, где все $X_i \sim \text{Bern}(p)$

$$E(X) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np$$

Всегда можно посчитать *в лоб*

$$E(X) = \sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i}$$

3 С.в., условная на другой с.в.

Условная с.в. (на другой с.в.)

Было:

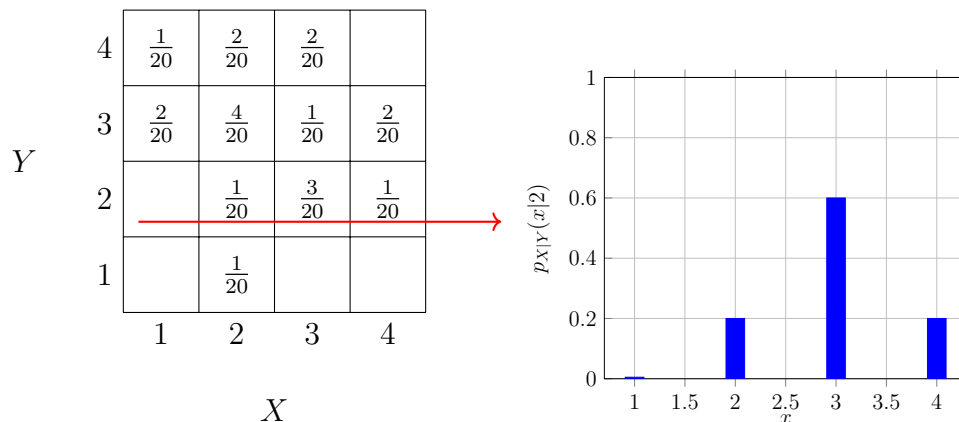
$$p_X(x | A) = \Pr(X = x | A)$$

Стало:

$$p_{X|Y}(x | y) = \Pr(X = x | Y = y)$$

$$p_{X|Y}(x | y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

NB: Условная вероятность определена только для y , у которых $p_Y(y) > 0$



Условные случайные векторы

$$\begin{aligned} & p_{X_1, \dots, X_n | Y_1, \dots, Y_m}(x_1, \dots, x_m | y_1, \dots, y_m) \\ &= \Pr(X_1 = x_1 \cap \dots \cap X_n = x_n | Y_1 = y_1 \cap \dots \cap Y_m = y_m) \end{aligned}$$

Чуть проще:

- $p_{X|Y,Z}(x | y, z) = \Pr(X = x | Y = y \cap Z = z)$
- $p_{X,Y|Z}(x, y | z) = \Pr(X = x \cap Y = y | Z = z)$

Работает правило умножения:

- Было: $\Pr(A \cap B \cap C) = \Pr(A) \Pr(B | A) \Pr(C | A \cap B)$
- Стало: $p_{X,Y,Z}(x, y, z) = p_X(x) p_{Y|X}(y | x) p_{Z|X,Y}(z | x, y)$

Условное матожидание

$$E(X | Y = y) = \sum_x x p_{X|Y}(x | y)$$

Условия просто *меняют вероятностную меру*, поэтому с условными с.в. все работает так же, как и с безусловными

Полные вероятность и матожидание Формула полной вероятности (оригинал): для разбиения Ω на A_i

$$p_X(x) = \sum_i \Pr(A_i) p_X(x | A_i)$$

Давайте теперь скажем, что $A_i = (Y = y)$, получим

$$p_X(x) = \sum_y p_Y(y) p_{X|Y}(x | y)$$

Полное матожидание — по аналогии

$$E(X) = \sum_y p_Y(y) E(X | Y = y)$$

NB: работает только когда ряд сходится *абсолютно*

4 Независимость

Независимость и условная независимость с.в.

- *Напоминание:* A и B независимы $\Leftrightarrow \Pr(A \cap B) = \Pr(A) \Pr(B)$
- Событие A и с.в. X независимы \Leftrightarrow

$$\forall x \quad \Pr(A \cap X = x) = \Pr(A) p_X(x)$$

- X и Y — независимые с.в. \Leftrightarrow

$$\forall x, y \quad p_{X,Y}(x, y) = p_X(x) p_Y(y)$$

NB: независимость по-прежнему означает, что одно событие (или с.в.) *не дает никакой информации* о другом событии (или с.в.)

Условная независимость с.в. Условие порождает новую меру \Rightarrow зависимость с.в. может меняться в зависимости от условия

Y	4	$\frac{1}{20}$	$\frac{2}{20}$	$\frac{2}{20}$	
	3	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{1}{20}$	$\frac{2}{20}$
	2		$\frac{1}{20}$	$\frac{3}{20}$	$\frac{1}{20}$
	1		$\frac{1}{20}$		
		1	2	3	4
		X			

- X и Y , очевидно, зависимы:

$$\begin{aligned} - p_X(1) &= \frac{3}{20} \\ - p_{X|Y}(1, 1) &= 0 \end{aligned}$$

- Но при условии $C = (X \leq 2 \cap Y \geq 3)$ становятся независимы:

$$\begin{aligned} - p_X(1 | C) &= \frac{1}{3}, p_X(2 | C) = \frac{2}{3} \\ - p_Y(3 | C) &= \frac{2}{3}, p_Y(4 | C) = \frac{1}{3} \end{aligned}$$

- Поэтому,

$$\begin{aligned}
- p_{X,Y}(1, 4 \mid C) &= \frac{1}{9} = p_X(1 \mid C)p_Y(4 \mid C) \\
- p_{X,Y}(1, 3 \mid C) &= \frac{2}{9} = p_X(1 \mid C)p_Y(3 \mid C) \\
- p_{X,Y}(2, 4 \mid C) &= \frac{2}{9} = p_X(2 \mid C)p_Y(4 \mid C) \\
- p_{X,Y}(2, 3 \mid C) &= \frac{4}{9} = p_X(2 \mid C)p_Y(3 \mid C)
\end{aligned}$$

Матожидание независимых с.в.

Пусть X и Y — независимые с.в., тогда

$$E(XY) = E(X)E(Y)$$

$$\begin{aligned}
E(XY) &= \sum_x \sum_y xyp_{X,Y}(x, y) = \sum_x \sum_y xyp_X(x)p_Y(y) \\
&= \sum_x xp_X(x) \sum_y yp_Y(y) = E(X)E(Y)
\end{aligned}$$

Дисперсия независимых с.в.

Пусть X и Y — независимые с.в., тогда

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Предположим, что $E(X) = E(Y) = 0$.

Тогда $E(XY) = E(X)E(Y) = 0$

$$\begin{aligned}
\text{Var}(X + Y) &= E((X + Y)^2) = E(X^2 + 2XY + Y^2) \\
&= E(X^2) + 2E(XY) + E(Y^2) = E(X^2) + E(Y^2) \\
&= \text{Var}(X) + \text{Var}(Y)
\end{aligned}$$

Дисперсия биномиалки Напоминание: $X \sim \text{Bin}(n, p) \Leftrightarrow X = \sum_{i=1}^n X_i$, где все $X_i \sim \text{Bern}(p)$ — независимы

$$\begin{aligned}
\text{Var}(X) &= \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) \\
&= \sum_{i=1}^n p(1 - p) = np(1 - p)
\end{aligned}$$

5 Более сложные распределения

Гипер-геометрическое распределение

Эксперимент: у нас в корзине есть N шаров, D из которых белые, $N - D$ — черные. Мы достаём случайные n шаров (не возвращая их в корзину). X — число белых шаров, которые мы достали. Тогда $X \sim HG(D, N, n)$

Матожидание: представим X_i как сумму индикаторных величин для всех белых шаров, что он был вытащен. Вероятность успеха $p = \frac{n}{N}$

$$E[X] = \sum_{i\text{- белый шар}} E[X_i] = D \cdot \frac{n}{N} = \frac{nD}{N}$$

Дисперсия:

$$\begin{aligned} E[X^2] &= \sum_{i\text{- белый шар}} E[X_i^2] + \sum_{i,j\text{- белые шары}} E[X_i X_j] = E[X] + D(D-1) \cdot \frac{n}{N} \cdot \frac{n-1}{N-1} \\ &= \frac{nD(N-n)(N-D)}{N^2(N-1)} \end{aligned}$$

При больших N очень близко к биномиальному распределению $\text{Bin}(n, \frac{D}{N})$.

Степенное распределение

$X \sim \text{row}(u, \beta)$ значит, что мы выбираем число из $[1..u]$, причем $\Pr(X = i) \sim i^{-\beta}$.
Более точно:

$$\Pr[X = i] = \begin{cases} C_{u,\beta} i^{-\beta}, & \text{если } i \in [1..u], \\ 0, & \text{иначе.} \end{cases}$$

$C_{u,\beta}$ — коэффициент нормализации

$$C_{u,\beta} = \left(\sum_{i=1}^u i^{-\beta} \right)^{-1} = H_{u,\beta}^{-1}$$

$H_{u,\beta}^{-1}$ — обобщенное гармоническое число

Замечание: если $\beta > 1$, то u может быть $+\infty$

Рассказать про приближение интегралами

6 Непрерывные случайные величины

Определение непрерывных с.в.

В случае с дискретными величинами:

$$X \sim \text{Geom}(p) \Rightarrow \Pr(X \geq t) = (1-p)^t$$

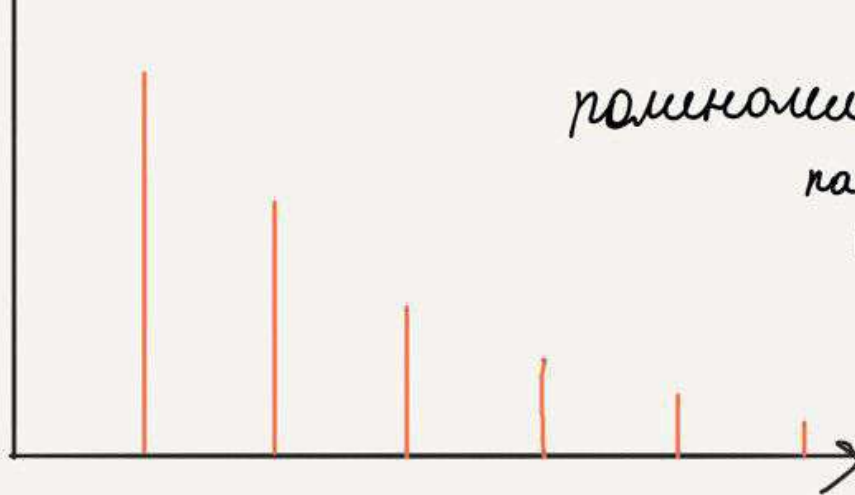
$$X \sim \text{pow}(+\infty, 2) \Rightarrow \Pr(X \geq t) =$$

$$= \frac{6}{\pi^2} \left(\sum_{k=t}^{+\infty} k^{-2} \right) \approx \frac{6}{\pi^2} \left(\int_t^{+\infty} x^{-2} dx \right) =$$

$$= \frac{6}{\pi^2} \cdot \frac{1}{t}$$

Степенное распределение:

$P_X(x)$



монотонно
падает