

Лекция 7. Функции от с.в.

22 марта 2021 г.

1 Линейные преобразования

Как это в общем случае посчитать новое распределение линейной функции от с.в.? Рассмотрим $Y = aX + b$ (где $a \neq 0$, иначе это скучный случай, когда $Y = b$ с вероятностью 1). Пусть сначала X дискретная, и нам известна ее функция вероятностей. Тогда

$$p_Y(y) = \Pr(Y = y) = \Pr(aX + b = y) = \Pr\left(X = \frac{y-b}{a}\right) = p_X\left(\frac{y-b}{a}\right)$$

С непрерывными такое не работает, так как вероятность, что непрерывная с.в. равна конкретному числу, есть ноль. Но мы можем работать с функциями распределения! Допустим мы знаем $f_X(x)$ и $F_X(x)$. Рассмотрим случай $a > 0$

$$F_Y(y) = \Pr(Y \leq y) = \Pr(aX + b \leq y) = \Pr\left(X \leq \frac{y-b}{a}\right) = F_X\left(\frac{y-b}{a}\right),$$

$$f_Y(y) = F'_Y(y) = f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

И отдельно $a < 0$

$$F_Y(y) = \Pr(Y \leq y) = \Pr(aX + b \leq y) = \Pr\left(X \geq \frac{y-b}{a}\right) = 1 - F_X\left(\frac{y-b}{a}\right),$$

$$f_Y(y) = F'_Y(y) = -f_X\left(\frac{y-b}{a}\right) \cdot \frac{1}{a}.$$

Объединяя два случая, получаем:

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Докажем теперь, что линейное преобразование нормального распределения оставляет его нормальным. Пусть $X \sim N(\mu, \sigma^2)$ и $Y = aX + b$. Значит,

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\ &= \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left(-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}\right) \\ &= \frac{1}{(|a|\sigma)\sqrt{2\pi}} \exp\left(-\frac{(y - (b + a\mu))^2}{2(a\sigma)^2}\right), \end{aligned}$$

что есть функция плотности вероятности для $N(a\mu + b, (a\sigma)^2)$.

2 Нелинейные преобразования

Алгоритм поиска нового распределения $Y = g(X)$:

1. Найти функцию распределения $F_Y(y) = \Pr(Y \leq y) = \Pr(g(X) \leq y)$
2. Продифференцировать: $f_Y(y) = F'_Y(y)$.

Рассмотрим случай, когда $g(x)$ монотонно возрастает и дифференцируема.

$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(g(X) \leq y) = \Pr(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) \\ f_Y(y) &= F'_Y(y) = F'_X(g^{-1}(y)) = f_X(g^{-1}(y))(g^{-1}(y))' = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} \end{aligned}$$

Аналогично можно рассмотреть случай, когда $g(x)$ монотонно убывает и дифференцируема

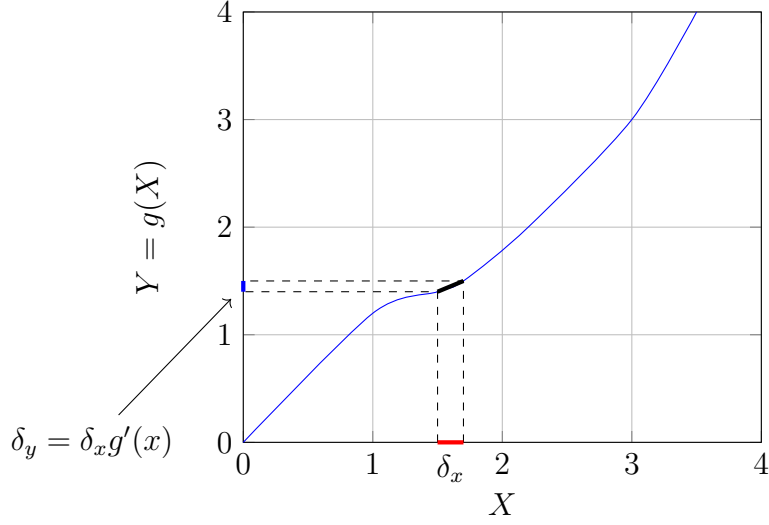
$$\begin{aligned} F_Y(y) &= \Pr(Y \leq y) = \Pr(g(X) \leq y) = \Pr(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y)) \\ f_Y(y) &= F'_Y(y) = -F'_X(g^{-1}(y)) = -f_X(g^{-1}(y))(g^{-1}(y))' = -\frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|} \end{aligned}$$

То есть в общем случае (когда $g(x)$ строго монотонна):

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|}$$

Интуитивное объяснение (которое использовалось на практике). Пусть $y = g(x)$

$$f_Y(y)\delta_y \approx \Pr(y \leq Y \leq y + \delta_y) = \Pr(x \leq X \leq x + \delta_x) = f_X(x)\delta_x = f_X(x)\frac{\delta_y}{g'(x)}$$



Для немонотонных $g(x)$ алгоритм остается прежний, но нет гарантий, что все пройдет легко. рассмотрим $g(x) = x^2$ и $Y = g(X)$ (при известной плотности X).

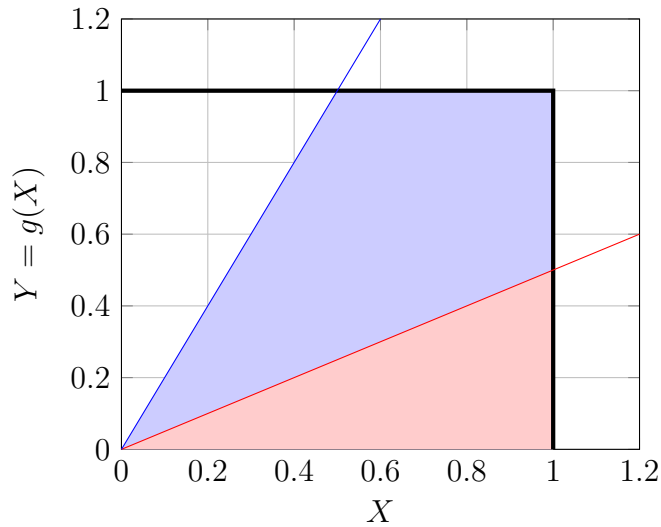
$$\begin{aligned}
F_Y(y) &= \Pr(Y \leq y) = \Pr(X^2 \leq y) = 1 - \Pr(X^2 > y) \\
&= 1 - \Pr(X > \sqrt{y}) - \Pr(X < -\sqrt{y}) \\
&= \Pr(X \leq \sqrt{y}) = \Pr(X < -\sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\
f_Y(y) &= F'_Y(y) = F'_X(\sqrt{y}) - F'_X(-\sqrt{y}) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} - f_X(-\sqrt{y}) \left(-\frac{1}{2\sqrt{y}}\right) \\
&= \frac{f_X(\sqrt{y}) + f_X(-\sqrt{y})}{2\sqrt{y}}
\end{aligned}$$

3 Функции нескольких с.в.

Если есть $Z = g(X, Y)$, то алгоритм нахождения ее функции распределения или плотности такой же. Рассмотрим на примере. X, Y – независимые, обе следуют $U(0, 1)$. Рассмотрим с.в. $Z = \frac{Y}{X}$.

Найдем функцию распределения $F_Z(z)$

$$F_Z(z) = \Pr\left(\frac{Y}{X} \leq z\right).$$



Если z меньше единицы, то это просто интеграл по красному треугольнику, то есть его площадь (так как совместная плотность $f_{X,Y}(x,y) = 1$ на всем квадрате).

$$F_Z(z) = \frac{z}{2}, \text{ если } z \in [0, 1].$$

Если $z > 1$, то это площадь всего квадрата минус площадь незакрашенного треугольника

$$F_Z(z) = 1 - \frac{1}{2z}.$$

Объединяя все вместе, получаем:

$$F_Z(z) = \begin{cases} 0, & \text{если } z < 0, \\ \frac{z}{2}, & \text{если } z \in [0, 1), \\ 1 - \frac{1}{2z}, & \text{если } z \geq 1. \end{cases}$$

3.1 Сумма независимых с.в.

В данном подразделе мы имеем две независимых с.в. X и Y , знаем их распределение и хотим узнать распределение с.в. $Z = X + Y$.

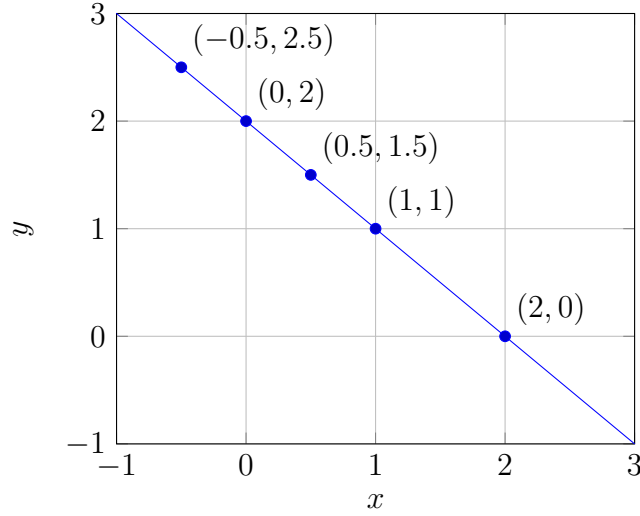
Дискретный случай

Чтобы посчитать вероятность того, что $Z = z$, необходимо, чтобы одновременно выполнялось:

- $X = x$ (где x — какое-то число из множества значений X)

- $Y = z - x$.

То есть мы берем все пары (x, y) с прямой $x + y = z$ и суммируем вероятности этих пар.



Заметим, что из-за независимости X и Y события $X = x$ и $Y = z - x$ независимы, поэтому

$$\Pr(X = x \cap Y = z - x) = p_X(x)p_Y(z - x).$$

Таким образом, формула выглядит так:

$$p_Z(z) = \sum_x p_X(x)p_Y(z - x)$$

Непрерывный случай

Аналогичная формула для непрерывных с.в.:

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_Y(z - x)dx$$

Чтобы ее доказать, обусловимся сначала на событии $X = x$. Тогда

$$f_{Z|X}(z | x) = f_{Y+x|X}(z | x) = f_{Y+x}(z) = f_Y(z - x).$$

Далее, по формуле полной вероятности имеем

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x)f_{Z|X}(z | x)dx = \int_{-\infty}^{+\infty} f_X(x)f_Y(z - x)dx /$$

Интересное следствие

Сумма независимых нормальных распределений есть нормальное распределение. Пусть $X \sim N(\mu_X, \sigma_X^2)$ и $Y \sim N(\mu_Y, \sigma_Y^2)$ — независимы. Посчитаем $f_Z(z)$, где $Z = X + Y$.

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sigma_X \sqrt{2\pi}} \exp\left(-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right) \cdot \frac{1}{\sigma_Y \sqrt{2\pi}} \exp\left(-\frac{(z-x-\mu_Y)^2}{2\sigma_Y^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi(\sigma_X^2 + \sigma_Y^2)}} \exp\left(-\frac{(z-\mu_X-\mu_Y)^2}{2(\sigma_X^2 + \sigma_Y^2)}\right) \end{aligned}$$

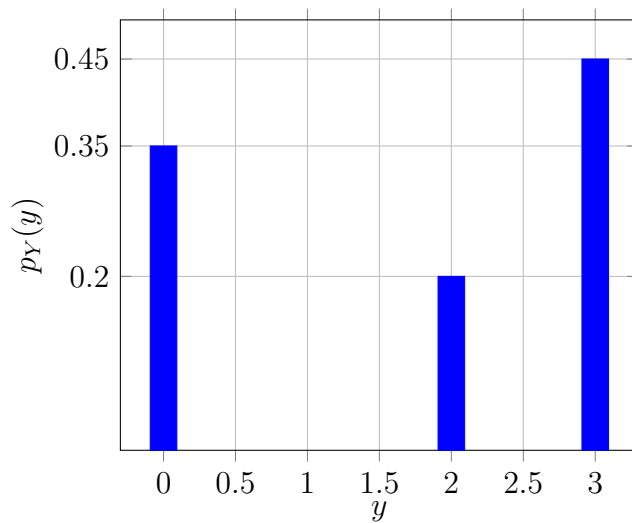
Тут было много несложной, но громоздкой математики, которую мы опустили. Таким образом, $Z \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$. Единственное удивительное здесь — то, что мы по-прежнему имеем нормальное распределение. Его параметры при этом очевидны из линейности матожиданий и из свойств дисперсии независимых с.в.

По индукции легко доказать, что для любого конечного множества независимых нормальных с.в. их сумма будет также нормальной с.в.

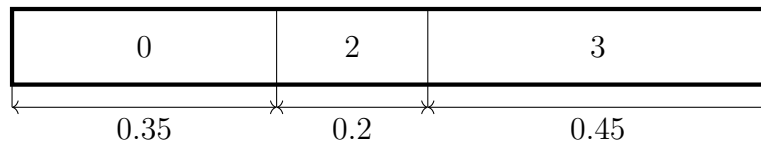
4 Эмитация одного распределения другим

Практически любое современное вычислительное устройство умеет генерировать (псевдо)случайные числа, но чаще всего только равномерное распределение. Этого на самом деле достаточно, чтобы получить случайную величину, следующую любому другому распределению. В данной части мы предполагаем, что мы умеем получать с.в. $X \sim U(0, 1)$ и хотим получить с.в. Y , для которой знаем функцию распределения $F_Y(y)$.

Рассмотрим простой случай. Y — дискретная с.в. с конечным набором значений. Пусть ее функция вероятности выглядит так:

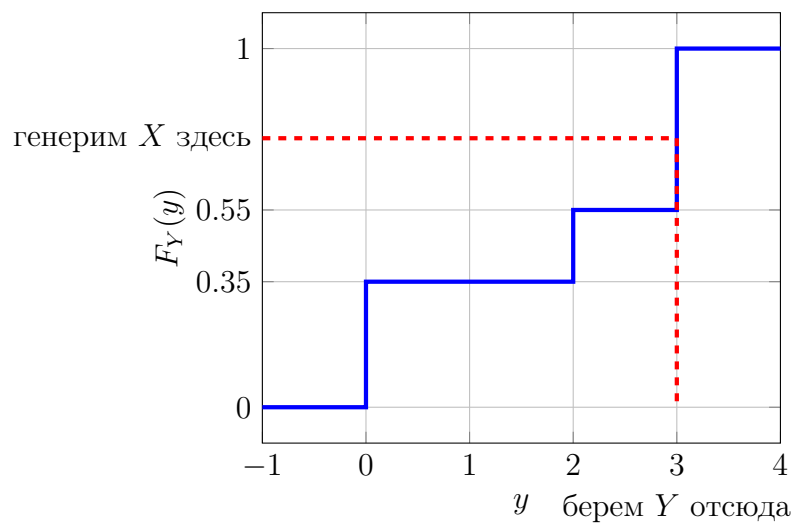


Как мы можем поступить. Возьмем отрезок единичной длины, разобьем его на отрезки длиной, соответствующей вероятности каждого значения с.в., сгенерим X и вернем значение, соответствующее тому значению, в которое попал X .

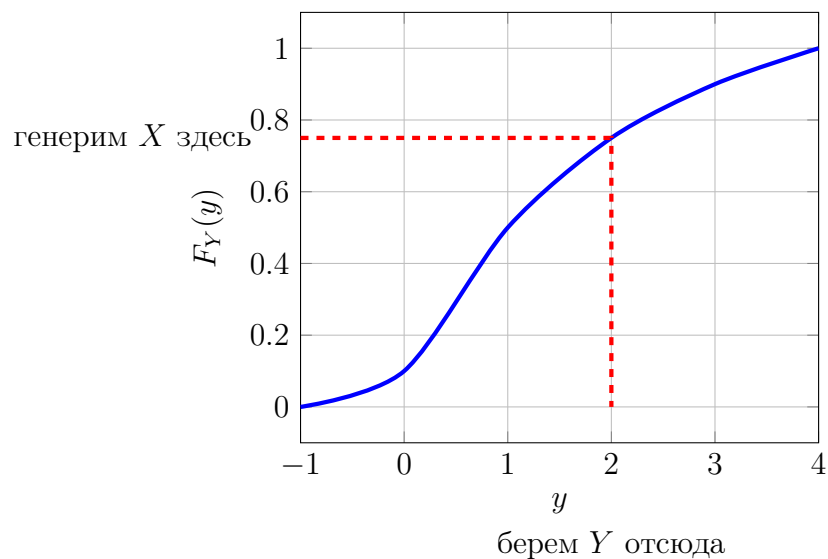


Таким методом можно смитировать любую с.в. с конечным множеством значений размера n за время запроса к $X \sim U(0, 1)$ плюс время $\Theta(\log(n))$, необходимое для нахождения отрезка, в который мы попали, двоичным поиском.

Интерпретация наших действий на основе функции распределения:



Давайте используем эту интерпретацию для непрерывной Y в случае с монотонной $F_Y(y)$.

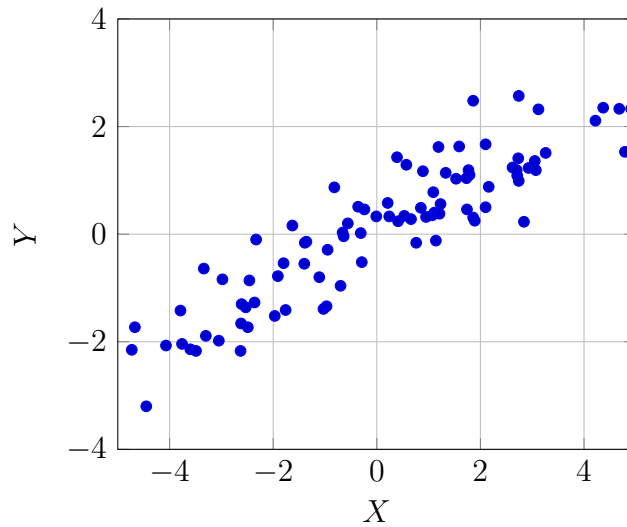


То есть после выбора X мы берем $Y = F_Y^{-1}(X)$. Почему это работает:

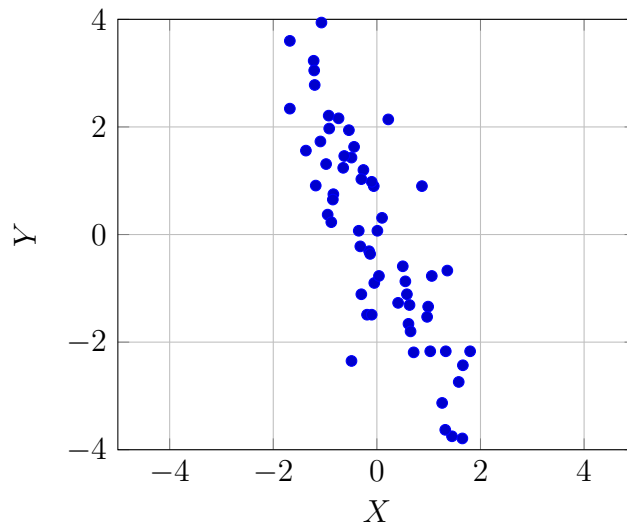
$$\Pr(Y \leq y) = \Pr(F_Y^{-1}(X) \leq y) = \Pr(X \leq F_Y(y)) = F_Y(y)$$

5 Ковариация

До сих пор мы про две с.в. могли сказать только то, зависимы ли они, или нет. Но иногда хочется определить степень этой зависимости, то есть понять, сколько информации одна с.в. дает про другую с.в. Рассмотрим пример. Пусть две с.в. X и Y имеют нулевое матожидание. Если они независимы, то $E[XY] = E[X]E[Y] = 0$. Но рассмотрим такие случаи: пусть каждая точка на графике равновероятна.



В таком случае велика вероятность, что либо X и Y оба меньше нуля, либо оба больше нуля, то есть $E[XY] > 0$. В следующем случае будет наоборот: X и Y с большой вероятностью имеют разные знаки, значит, $E[XY] < 0$.



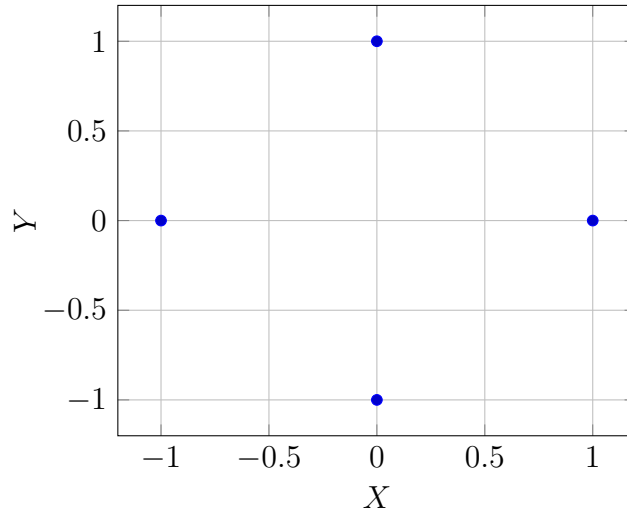
Чтобы измерить степень зависимости была введена величина, называемая *ковариацией* случайных величин X и Y .

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

Если две с.в. независимы, то для них будет верно

$$\text{Cov}(X, Y) = E(X - E(X))E(Y - E(Y)) = 0 \cdot 0 = 0$$

Правда обратное неверно: с.в. с нулевой ковариацией могут быть зависимы. Рассмотрим пример. Пусть четыре пары (X, Y) равновероятны: $(1, 0), (0, 1), (-1, 0), (0, -1)$



Заметим, что матожидания X и Y равны нулю, то есть $\text{Cov}(X, Y) = E[XY]$, но XY всегда равен нулю, поэтому ковариация нулевая. При этом X и Y явно зависимы: если $X = 1$, то Y однозначно равен нулю.

6 Свойства ковариации

Полезная формула.

Ковариация с.в. с самой собой равна ее дисперсии, причем у нас была очень милая формула:

$$\text{Cov}(X, X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

Есть ли что-то подобное для ковариации? Конечно, есть.

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] = E[XY - YE[X] - XE[Y] + E[X]E[Y]] \\ &= E[XY] - E[X]E[Y] - E[X]E[Y] + E[X]E[Y] = E[XY] - E[X]E[Y]. \end{aligned}$$

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$$

Линейные преобразования

Линейное преобразование одной с.в.:

$$\begin{aligned}
\text{Cov}(aX + b, Y) &= E[(aX + b)Y] - E[aX + b]E[Y] \\
&= aE[XY] + bE[Y] - aE[X]E[Y] - bE[Y] \\
&= a(E[XY] - E[X]E[Y]) = a \text{Cov}(X, Y)
\end{aligned}$$

Заметим, что прибавление константы к одной с.в. ничего не меняет, так как не меняется распределение $X - E[X]$, а умножение на константу как раз увеличивает этот множитель в ту же константу.

Одна с.в. – сумма двух других:

$$\begin{aligned}
\text{Cov}(X + Y, Z) &= E\left[(X + Y - E[X + Y])(Z - E[Z])\right] \\
&= E\left[(X - E[X])(Z - E[Z])\right] + E\left[(Y - E[Y])(Z - E[Z])\right] \\
&= \text{Cov}(X, Z) + \text{Cov}(Y, Z)
\end{aligned}$$

Ковариация и дисперсия суммы с.в.

Распишем дисперсию суммы двух с.в. по формуле

$$\begin{aligned}
\text{Var}(X + Y) &= E\left[(X + Y - E[X + Y])^2\right] \\
&= E\left[((X - E[X]) + (Y - E[Y]))^2\right] \\
&= E[(X - E[X])^2] + E[(Y - E[Y])^2] + 2E[(X - E[X])(Y - E[Y])] \\
&= \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)
\end{aligned}$$

Что насчет более двух с.в.? Предположим, что у нас есть набор с.в. X_1, \dots, X_n , и матожидание каждой с.в. равно нулю. Тогда имеем

$$\begin{aligned}
\text{Var}(X_1 + \dots + X_n) &= E((X_1 + \dots + X_n)^2) \\
&= E\left(\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j\right) \\
&= \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)
\end{aligned}$$

Если хотим показать то же самое для с.в. с ненулевым ожиданием, то заметим, что добавление константы к с.в. не меняет ни ее дисперсию, ни ковариацию с другой с.в., поэтому

$$\begin{aligned}
\text{Var}(X_1 + \dots + X_n) &= \text{Var}((X_1 - E[X_1]) + \dots + (X_n - E[X_n])) \\
&= \sum_{i=1}^n \text{Var}(X_i - E[X_i]) + \sum_{i \neq j} \text{Cov}((X_i - E[X_i]), (X_j - E[X_j])) \\
&= \sum_{i=1}^n \text{Var}(X_i) + \sum_{i \neq j} \text{Cov}(X_i, X_j)
\end{aligned}$$

7 Коэффициент корреляции

Ковариация – очень странная величина. Ее размер по сути зависит от разброса двух с.в., которые являются ее аргументом. То есть по фразе “ковариация X и Y равна единице” очень сложно сказать, много это, или мало, так как мы не знаем масштаба X и Y . Например, когда мы пытаемся посчитать ковариацию температуры воздуха и давления, то в зависимости от используемых величин измерения (цельсий или фаренгейты, паскали или атмосферы) будет разная ковариация. Причем у самой ковариации всегда есть единица измерения, равная единице измерения X , умноженная на Y (например, градус цельсия * паскаль), что еще больше усложняет осознание степени зависимости пары с.в. Поэтому была введена мера зависимости с.в., называемая коэффициентом корреляции $\rho(X, Y)$.

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

NB: он определен только для с.в. с ненулевой дисперсией.

Его основные свойства:

1. $\rho(X, Y) \in [-1, 1]$
2. Если X и Y независимы, то $\rho(X, Y) = 0$, как и ковариация (обратное не всегда верно, см. пример для нулевой ковариации)
3. Если $|\rho(X, Y)| = 1$, то X и Y линейно зависимы, то есть $(X - E[X]) = c(Y - E[Y])$.
Как следствие: $\rho(X, X) = \frac{\text{Cov}(X, X)}{\sigma_X^2} = 1$.
4. Линейные преобразования любой из двух с.в. не меняют модуль коэффициента корреляции: $\rho(aX + b, Y) = \frac{a \text{Cov}(X, Y)}{|a| \sigma_X \sigma_Y} = \text{sign}(a) \rho(X, Y)$

Для доказательства первого и третьего свойств рассмотрим с.в. X и Y с нулевыми матожиданиями и единичной дисперсией (для общего случая легко доказать все,

перейдя к с.в. $X' = \frac{X-E[X]}{\sigma_X}$ и $Y' = \frac{Y-E[Y]}{\sigma_Y}$, но мы опустим это для упрощения вычислений). Рассмотрим с.в. $Z = (X - \rho(X, Y)Y)^2$. Она неотрицательна, значит, и ее матожидание неотрицательно.

$$\begin{aligned} E[Z] &= E[(X - \rho Y)^2] = E[X^2] - 2\rho E[XY] + \rho^2 E[Y^2] \\ &= \text{Var}(X) - 2\rho \cdot \rho + \rho^2 \text{Var}(Y) = 1 - \rho^2 \geq 0 \end{aligned}$$

Отсюда следует, что во-первых, $\rho^2 \leq 1$, а во-вторых, если $|\rho| = 1$, то $E[Z] = 0$, что возможно только если Z всегда равен нулю, что в свою очередь подразумевает, что X всегда равен $\pm Y$ (в зависимости от знака ρ).

8 Интерпретация корреляции

Если две случайных величины зависимы, то чаще всего это не значит, что одна определяет другую и наоборот. Например, проведя исследование оценок в школе, можно обнаружить, что у учеников с хорошей оценкой по математике также хорошие оценки по литературе, однако это не значит, что знания по математике помогают в изучении литературы и наоборот. Это скорее означает, что у двух этих случайных величин есть какой-то общий фактор, который влияет на них обоих. В случае с оценками это может быть, например, степень усердия, которое тот или иной ученик прилагает к учебе.

Чуть более формально, давайте представим, что есть три случайных величины Z, U и V с нулевыми матожиданиями и единичными дисперсиями, которые все независимы друг от друга. И рассмотрим две других с.в. $X = Z + U$ и $Y = Z + V$. Посчитаем $\rho(X, Y)$

$$\begin{aligned} \text{Var}(X) &= \text{Var}(Y) = \text{Var}(Z) + \text{Var}(U) = 2 \\ \sigma_X &= \sigma_Y = \sqrt{2} \\ E[XY] &= E[Z^2 + UZ + VZ + UV] = \text{Var}(Z) = 1 \\ \rho(X, Y) &= \frac{E[XY]}{\sigma_X \sigma_Y} = \frac{1}{2} \end{aligned}$$

То есть коэффициент корреляции определяет вклад какой-то причины, которая влияет на обе с.в., в каждую с.в.

9 Пример важности корреляции

Допустим, вы решили инвестировать. И у вас есть 10\$, и вы вкладываете по доллару в какие-то 10 компаний. Средняя выручка с каждой компании равна 1\$ и среднеквадратичное отклонение выручки с каждой компании есть 1.3\$. Очевидно, ваша суммарная

выручка равна вашему вкладу, но каково среднеквадратичное отклонение? Если выручки от разных компаний независимы, То

$$\begin{aligned}\text{Var}(X_1 + \dots + X_{10}) &= \sum_{i=1}^{10} \text{Var}(X_i) = 10 \cdot 1.3^2 = 16.9 \\ \sigma_{X_1 + \dots + X_{10}} &= \sqrt{16.9} \approx 4.1\end{aligned}$$

То есть сумма довольно хорошо сконцентрирована, больше, чем при вкладе всего лишь в одну компанию. На большую прибыль надеяться не стоит, но и много проиграть шансов не так много. Теперь допустим, что у всех выручек есть какой-то общий фактор, который под ними лежит, и коэффициент корреляции между любыми X_i и X_j равен 0.9. Тогда

$$\begin{aligned}\text{Var}(X_1 + \dots + X_{10}) &= \sum_{i=1}^{10} \text{Var}(X_i) - \sum_{i \neq j} \text{Cov}(X_i, X_j) = 10 \cdot 1.3^2 + 90 \cdot 0.9 \cdot (1.3)^2 \approx 154 \\ \sigma_{X_1 + \dots + X_{10}} &= \sqrt{154} \approx 12.4\end{aligned}$$

То есть если падают акции хотя бы одной компании, то с большой вероятностью падают и акции других, что приводит к большим потерям. Примерно это и случилось в 2008 году, когда рухнул весь рынок сразу.