

Теория вероятностей и математическая статистика

Линейные комбинации случайных величин.

Глеб Карпов

ФКН ВШЭ

Линейная комбинация случайных величин

Математическое ожидание комбинации

- Часто важно анализировать поведение суммы двух или более случайных величин
- В основном мы хотим понять, как ведет себя переменная $T = aX \pm bY$, каковы ее характеристики.
- Линейное свойство математического ожидания: $E[aX \pm bY] = aE[X] \pm bE[Y]$.

$$\begin{aligned} E[T] &= \sum_{i=1}^n \sum_{j=1}^m (ax_i \pm by_j) P(X = x_i, Y = y_j) = \\ &= \sum_{i=1}^n \sum_{j=1}^m ax_i P(X = x_i, Y = y_j) \pm \sum_{i=1}^n \sum_{j=1}^m by_j P(X = x_i, Y = y_j) = \\ &= a \sum_{i=1}^n x_i \underbrace{\sum_{j=1}^m P(X = x_i, Y = y_j)}_{\text{marginal for } X} \pm b \sum_{j=1}^m y_j \underbrace{\sum_{i=1}^n P(X = x_i, Y = y_j)}_{\text{marginal for } Y} = \\ &= a \sum_{i=1}^n x_i P(X = x_i) \pm b \sum_{j=1}^m y_j P(Y = y_j) = aE[X] \pm bE[Y] \end{aligned}$$

Линейная комбинация случайных величин

Но дисперсия - это уже совсем другая история...

- Нас интересует $Var[T] = Var[aX \pm bY]$.
- Давайте применим сокращенную формулу для дисперсии:

$$\begin{aligned}Var[T] &= E[T^2] - (E[T])^2 = E[(aX \pm bY)^2] - (aE[X] \pm bE[Y])^2 \\&= E[a^2X^2 \pm 2abXY + b^2Y^2] - ((aE[X])^2 \pm 2abE[X]E[Y] + (bE[Y])^2) \\&= a^2E[X^2] \pm 2abE[XY] + b^2E[Y^2] - a^2(E[X])^2 \mp 2abE[X]E[Y] - b^2(E[Y])^2 \\&= a^2Var[X] + b^2Var[Y] \pm 2ab\left(E[XY] - E[X]E[Y]\right)\end{aligned}$$

- В процессе использовали свойство линейности математического ожидания.
- Слагаемое $E[XY] - E[X]E[Y]$ называется **ковариацией** X и Y .

Ковариация и независимость

Ков(x,y)=0
Независим.

Утверждение: Если X и Y - независимые случайные величины, то $E[XY] = E[X]E[Y]$

i Proof

Используем определение независимости величин:

$$\forall x_i, \forall y_j : P_X(X = x_i) P_Y(Y = y_j) = P(\{X = x_i\} \cap \{Y = y_j\}) = P(X = x_i, Y = y_j)$$

$$\begin{aligned} E[XY] &= \sum_{i=1}^n \sum_{j=1}^m x_i y_j P(X = x_i, Y = y_j) = \sum_{i=1}^n \sum_{j=1}^m x_i y_j P_X(x_i) P_Y(y_j) \\ &= \sum_{i=1}^n x_i P_X(x_i) \sum_{j=1}^m y_j P_Y(y_j) = E[X]E[Y]. \end{aligned}$$

Будьте внимательны: обратное утверждение неверно!

Корреляция

- Ковариация не является хорошей мерой зависимости, поскольку она зависит от масштабов X и Y .
- Идея: нормализовать ковариацию и избавиться от масштабов.
- Корреляция:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}}$$

- Она строго ограничена от -1 до 1 и:

$$\text{Corr}(X, Y) = 1 \leftrightarrow Y = kX + b, k > 0$$

$$\text{Corr}(X, Y) = -1 \leftrightarrow Y = -kX + b, k > 0$$

- Важно: Если переменные независимы, то их ковариация (и корреляция) равна нулю. Обратное утверждение НЕВЕРНО. Очень может быть, что ковариация равна нулю, но при этом величины являются зависимыми

$$E(X \cdot Y) = -1 \cdot 3 \cdot \frac{3}{4} - \frac{6}{4} = -\frac{15}{4}$$

Пример:

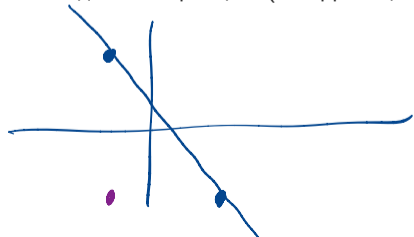
$$a \cdot (-1) + b = 3 \quad -2X + 1$$

$$a \cdot 2 + b = -3$$

$$3a = -6$$

$$a = -2$$

1. Найдите маргинальные функции вероятности для X и Y .
2. Независимы ли X и Y ?
3. Найдите ковариацию (и корреляцию) между X и Y .



x	-1	2
P_X	$\frac{3}{4}$	$\frac{1}{4}$

	$Y = 3$	$Y = -3$
$X = -1$	$\frac{3}{4}$	0
$X = 2$	0	$\frac{1}{4}$

y	-3	3
P_Y	$\frac{1}{4}$	$\frac{3}{4}$

$$P(Y=3) \cdot P(X=2) = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$$

Решение задачи

1. Маргинальные функции вероятности

Для X :

$$P(X = -1) = \sum_j P(X = -1, Y = y_j) = \frac{3}{4} + 0 = \frac{3}{4}$$

$$P(X = 2) = \sum_j P(X = 2, Y = y_j) = 0 + \frac{1}{4} = \frac{1}{4}$$

Для Y :

$$P(Y = 3) = \sum_i P(X = x_i, Y = 3) = \frac{3}{4} + 0 = \frac{3}{4}$$

$$P(Y = -3) = \sum_i P(X = x_i, Y = -3) = 0 + \frac{1}{4} = \frac{1}{4}$$

Пример

Проверка независимости

Проверим условие независимости для пары:

$$P(X = -1, Y = 3) = \frac{3}{4} \quad \text{и} \quad P(X = -1)P(Y = 3) = \frac{3}{4} \cdot \frac{3}{4} = \frac{9}{16} \neq \frac{3}{4}$$

Ответ: X и Y не независимы.

Пример

Ковариация и корреляция

Сначала найдем математические ожидания:

$$E[X] = (-1) \cdot \frac{3}{4} + 2 \cdot \frac{1}{4} = -\frac{3}{4} + \frac{1}{2} = -\frac{1}{4}$$

$$E[Y] = 3 \cdot \frac{3}{4} + (-3) \cdot \frac{1}{4} = \frac{9}{4} - \frac{3}{4} = \frac{3}{2}$$

Теперь найдем $E[XY]$:

$$\begin{aligned} E[XY] &= (-1) \cdot 3 \cdot \frac{3}{4} + (-1) \cdot (-3) \cdot 0 + 2 \cdot 3 \cdot 0 + 2 \cdot (-3) \cdot \frac{1}{4} \\ &= -\frac{9}{4} + 0 + 0 - \frac{6}{4} = -\frac{15}{4} \end{aligned}$$

Пример

Ковариация:

$$\begin{aligned} Cov(X, Y) &= E[XY] - E[X]E[Y] = -\frac{15}{4} - \left(-\frac{1}{4}\right) \cdot \frac{3}{2} \\ &= -\frac{15}{4} + \frac{3}{8} = -\frac{30}{8} + \frac{3}{8} = -\frac{27}{8} \end{aligned}$$

Для корреляции найдем дисперсии:

$$\begin{aligned} E[X^2] &= (-1)^2 \cdot \frac{3}{4} + 2^2 \cdot \frac{1}{4} = \frac{3}{4} + 1 = \frac{7}{4} \\ Var[X] &= E[X^2] - (E[X])^2 = \frac{7}{4} - \left(-\frac{1}{4}\right)^2 = \frac{7}{4} - \frac{1}{16} = \frac{27}{16} \end{aligned}$$

$$\begin{aligned} E[Y^2] &= 3^2 \cdot \frac{3}{4} + (-3)^2 \cdot \frac{1}{4} = \frac{27}{4} + \frac{9}{4} = 9 \\ Var[Y] &= E[Y^2] - (E[Y])^2 = 9 - \left(\frac{3}{2}\right)^2 = 9 - \frac{9}{4} = \frac{27}{4} \end{aligned}$$

Пример

Корреляция:

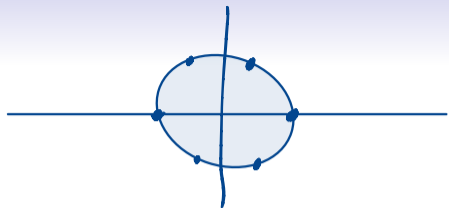
$$\begin{aligned} \text{Corr}(X, Y) &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}} = \frac{-\frac{27}{8}}{\sqrt{\frac{27}{16} \cdot \frac{27}{4}}} \\ &= \frac{-\frac{27}{8}}{\sqrt{\frac{729}{64}}} = \frac{-\frac{27}{8}}{\frac{27}{8}} = -1 \end{aligned}$$

Корреляция равна -1 , что означает полную отрицательную линейную зависимость между X и Y .

$$E[XY] = \frac{1}{2} \frac{\sqrt{3}}{2} \cdot \frac{1}{6} + \frac{1}{2} \left(-\frac{\sqrt{3}}{2}\right) \cdot \frac{1}{6} +$$

$$+ \frac{1}{2} \frac{\sqrt{3}}{2} \cdot \frac{1}{6} + \frac{1}{2} \frac{\sqrt{3}}{2} \cdot \frac{1}{6} = 0$$

Пример



Пусть известно распределение случайной величины U :

U	$u_1 = 0$	$u_2 = \pi/3$...	$u_6 = 5\pi/3$
P_U	$\frac{1}{6}$	$\frac{1}{6}$

1. Введём переменные $X = \cos U$ и $Y = \sin U$. Постройте таблицу их совместного распределения.
2. Являются ли они зависимыми? Рассмотрите, например, $P(X = \frac{1}{2})$ и $P(X = \frac{1}{2} | Y = \frac{\sqrt{3}}{2})$.
3. Найдите математические ожидания $E[X]$ и $E[Y]$.
4. Найдите ковариацию и корреляцию между X и Y .

x	-1	-1/2	1/2	1
P_X	1/6	2/6	1/6	1/6

y	$-\sqrt{3}/2$	0	$\sqrt{3}/2$
P_Y	2/6	2/6	2/6

$X \setminus Y$	0	$\sqrt{3}/2$	$-\sqrt{3}/2$
1	1/6	0	0
-1	1/6	0	0
1/2	0	1/6	1/6
-1/2	0	1/6	1/6

Слайд для записей

x	-1	$-\frac{1}{2}$	$\frac{1}{2}$	1
P_x	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

y	$-\frac{\sqrt{3}}{2}$	0	$\frac{\sqrt{3}}{2}$
P_y	$\frac{2}{6}$	$\frac{2}{6}$	$\frac{2}{6}$

$$E[X] = 1 \cdot \frac{1}{6} - 1 \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{2}{6} - \frac{1}{2} \cdot \frac{2}{6} = 0$$

$$E[Y] = \frac{\sqrt{3}}{2} \cdot \frac{2}{6} - \frac{\sqrt{3}}{2} \cdot \frac{2}{6} = 0$$

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