

— Что мы знаем о лисе?  
— Ничего. И то не все.  
Борис Заходер

Цель этой заметки — доказать теорему Иссерлиса о подсчёте ожиданий для многомерного нормального распределения. По дороге выведем функцию производящую моменты для одномерного нормального  $\mathcal{N}(\mu, \sigma^2)$  и многомерного нормального  $\mathcal{N}(\mu, C)$ .

Если  $Y \sim \mathcal{N}(\mu, C)$ , то

$$\mathbb{E}(Y_1) = \mu_1$$

$$\mathbb{E}(Y_1 Y_2) = \mu_1 \mu_2 + c_{12}$$

$$\mathbb{E}(Y_1 Y_2 Y_3) = \mu_1 \mu_2 \mu_3 + \mu_1 c_{23} + \mu_2 c_{13} + \mu_3 c_{12}.$$

$$\mathbb{E}(Y_1 Y_2 Y_3 Y_4) = \mu_1 \mu_2 \mu_3 \mu_4 + \mu_1 \mu_4 c_{23} + \mu_2 \mu_4 c_{13} + \mu_3 \mu_4 c_{12} + \mu_1 \mu_3 c_{24} + \mu_1 \mu_2 c_{34} + \mu_2 \mu_3 c_{14} + c_{12} c_{34} + c_{13} c_{24} + c_{14} c_{23}.$$

Каждое слагаемое идёт с единичным весом. Каждое слагаемое содержит все индексы от единицы до максимального ровно по одному разу. Все варианты перемножения  $\mu_i$  и  $c_{jk}$  присутствуют. Множитель  $\mu_i$  «съедает» один индекс, а  $c_{jk}$  «съедает» два индекса.

## Почти доказательство

Если вектор  $Y$  имеет нормальное распределение  $\mathcal{N}(\mu, C)$ , то  $\mathbb{E}(Y_1) = \mu_1$  и  $\mathbb{E}(Y_1 Y_2) = \mu_1 \mu_2 + c_{12}$ .

А как выглядят ожидания  $\mathbb{E}(Y_1 Y_2 Y_3)$ ,  $\mathbb{E}(Y_1 Y_2 Y_3 Y_4)$  и так далее?

Конечно, они должны быть функциями от ожиданий  $\mu_i$  и ковариаций  $c_{jk}$ , так как эти параметры полностью описывают многомерное нормальное распределение.

Поглядев на  $\mathbb{E}(Y_1)$  и  $\mathbb{E}(Y_1 Y_2)$  мы видим, что эти функции являются многочленами от  $\mu_i$  и  $c_{jk}$ . Давайте предположим, что и дальнейшие ожидания тоже будут многочленами.

Конечно, произвольный многочлен может содежать слагаемые в духе  $\mu_1^2 \mu_3^4 c_{15} c_{27}^3 c_{29}^9$ . Однако  $\mathbb{E}(Y_1 Y_2 Y_3)$  должно быть линейно по  $Y_1$ . Например, при увеличении  $Y_1$  в два раза двойка должна выноситься из каждого слагаемого многочлена! При увеличении  $Y_1$  в два раза во столько же раз растут ожидания  $\mu_1$  и ковариации  $c_{12}$ ,  $c_{13}$ . Значит каждое слагаемое многочлена должно содержать ровно одну из этих величин в качестве сомножителя. Например, слагаемое  $\mu_1 c_{12} \mu_3$  невозможно, так как растёт в 4 раза при увеличении  $Y_1$  в два раза.

Аналогично рассуждая про  $Y_2$  и  $Y_3$  мы понимаем, что в каждом слагаемом каждый индекс от 1 до 3 должен быть упомянут ровно один раз.

$$\mathbb{E}(Y_1 Y_2 Y_3) = ? \mu_1 \mu_2 \mu_3 + ? \mu_1 c_{23} + ? \mu_2 c_{13} + ? \mu_3 c_{12}.$$

Более того, формула должна быть симметричной, а именно, должна сохраняться при смене индексов. Коэффициенты при  $\mu_1 c_{23}$ ,  $\mu_2 c_{13}$  и  $\mu_3 c_{12}$  должны совпадать.

Также можно было рассуждать по размерностям, если  $Y_1$  измеряется в пудах,  $Y_2$  — в вершках, а  $Y_3$  — в саженьях, то каждое слагаемое многочлена должно иметь такие же единицы измерения, [пуд  $\times$  вершок  $\times$  сажень]. А слагаемые, где какой-то индекс повторяется имеют неподходящие единицы измерения. Например,  $\mu_2 c_{23}$  измеряется в [вершок<sup>2</sup>  $\times$  сажень].

Начнём охоту за коэффициентами многочлена! Занулим все слагаемые кроме  $\mu_1 \mu_2 \mu_3$ . Для этого возьмём независимые  $Y_1 \sim \mathcal{N}(1, 1)$ ,  $Y_2 \sim \mathcal{N}(1, 1)$  и  $Y_3 \sim \mathcal{N}(1, 1)$ . С одной стороны для них  $\mathbb{E}(Y_1 Y_2 Y_3) = 1$ . В многочлене при этом остаётся лишь слагаемое  $\mu_1 \mu_2 \mu_3$ , следовательно, коэффициент при нём равен 1.

Теперь занулим все слагаемые кроме  $\mu_1 c_{23}$ . Для этого возьмём  $Y_1 \sim \mathcal{N}(1, 1)$ ,  $Y_2 = Y_3 \sim \mathcal{N}(0, 1)$  и независимы от  $Y_1$ . В этом случае  $\mathbb{E}(Y_1 Y_2 Y_3) = 1 \cdot 1$  и коэффициент при слагаемом  $\mu_1 c_{23}$  также равен 1.

Итого мы получили, что все возможные слагаемые присутствуют с весом 1,

$$\mathbb{E}(Y_1 Y_2 Y_3) = \mu_1 \mu_2 \mu_3 + \mu_1 c_{23} + \mu_2 c_{13} + \mu_3 c_{12}.$$

Для следующего ожидания  $\mathbb{E}(Y_1 Y_2 Y_3 Y_4)$  сработают аналогичные рассуждения. Во-первых, чтобы сохранялась линейность по отдельным  $Y_i$  ни одно слагаемое не может иметь повторяющихся индексов. Во-вторых, все коэффициенты многочлена равны единице. Например, чтобы посмотреть в новом многочлене на коэффициент при слагаемом  $c_{12} c_{34}$  нужно взять  $Y_1 = Y_2 \sim \mathcal{N}(0, 1)$  и независимую величину  $Y_3 = Y_4 \sim \mathcal{N}(0, 1)$ . А чтобы глянуть на коэффициент при  $\mu_1 \mu_2 c_{34}$  нужно взять независимые  $Y_1 \sim \mathcal{N}(1, 1)$ ,  $Y_2 \sim \mathcal{N}(1, 1)$  и  $Y_3 = Y_4 \sim \mathcal{N}(0, 1)$ . Единственно возможный вариант,

$$\mathbb{E}(Y_1 Y_2 Y_3 Y_4) = \mu_1 \mu_2 \mu_3 \mu_4 + \mu_1 \mu_4 c_{23} + \mu_2 \mu_4 c_{13} + \mu_3 \mu_4 c_{12} + \mu_1 \mu_3 c_{24} + \mu_1 \mu_2 c_{34} + \mu_2 \mu_3 c_{14} + c_{12} c_{34} + c_{13} c_{24} + c_{14} c_{23}.$$

Пример.

Вектор  $Y$  имеет совместное нормальное распределение,

$$\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 5 & -1 & -2 \\ -1 & 6 & -3 \\ -2 & -3 & 7 \end{pmatrix} \right).$$

Найдите  $\mathbb{E}(Y_1 Y_2 Y_3)$  и  $\mathbb{E}(Y_1 Y_2 Y_2)$ .

Решение:

$$\mathbb{E}(Y_1 Y_2 Y_3) = \mu_1 \mu_2 \mu_3 + \mu_1 c_{23} + \mu_2 c_{13} + \mu_3 c_{12} = 1 \cdot 2 \cdot 3 + 1 \cdot (-3) + 2 \cdot (-2) + 3 \cdot (-1) = -2.$$

Если какой-то индекс повторяется, то его надо повторить :)

$$\mathbb{E}(Y_1 Y_2 Y_2) = \mu_1 \mu_2 \mu_2 + \mu_1 c_{22} + \mu_2 c_{12} + \mu_2 c_{12}.$$

Всё! Доказательство опирается на допущение, что  $\mathbb{E}(Y_1 Y_2 \dots Y_n)$  является многочленом от  $\mu_i$  и  $c_{jk}$  для нормального распределения. Для других распределений это не верно. В оставшейся части мы залатаем это предположение доказав теорему Иссерлиса через производящие функции.

## Одномерная функция производящая моменты

Первая задача. Найдите функцию производящую моменты  $\text{mgf}(u) = \mathbb{E}(\exp(uX))$  для нормальной  $X \sim \mathcal{N}(0, 1)$ .

Перейдём к интегралам!

$$\mathbb{E}(\exp(uX)) = \int_{\mathbb{R}} \exp(ux) f(x) dx = \int_{\mathbb{R}} \exp(ux) \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx.$$

Для взятия интеграла выделим полный квадрат внутри экспоненты:

$$ux - x^2/2 = -\frac{1}{2}(x^2 - 2ux + u^2 - u^2) = -\frac{1}{2}(x - u)^2 + \frac{1}{2}u^2.$$

Возвращаемся к интегралу:

$$\mathbb{E}(\exp(uX)) = \dots = \int_{\mathbb{R}} \exp(u^2/2) \frac{1}{\sqrt{2\pi}} \exp(-(x - u)^2/2) dx = \exp(u^2/2) \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} \exp(-(x - \sigma)^2/2) dx.$$

Замечаем, что последний интеграл — это площадь под нормальной функцией плотности, смещённой на  $u$  вправо. И эта площадь равна единице.

Задача раз решена,

$$\mathbb{E}(\exp(uX)) = \exp(u^2/2)$$

Вторая задача. Найдите функцию производящую моменты  $\text{mgf}_Y(u) = \mathbb{E}(\exp(uY))$  для нормальной  $Y \sim \mathcal{N}(\mu; \sigma^2)$ .

Сначала стандартизируем  $Y$ ,  $Y = \mu + \sigma X$ , где  $X \sim \mathcal{N}(0; 1)$ :

$$\mathbb{E}(\exp(uY)) = \mathbb{E}(\exp(u\mu) \exp(u\sigma X)) = \exp(u\mu) \mathbb{E}(\exp(u\sigma X)).$$

Ожидание в конце мы уже де-факто считали,  $\mathbb{E}(\exp(u\sigma X)) = \exp(u^2\sigma^2/2)$ .

Отсюда

$$\text{mgf}_Y(u) = \mathbb{E}(\exp(uY)) = \exp(\mu u + \sigma^2 u^2/2), \text{ если } Y \sim \mathcal{N}(\mu, \sigma^2).$$

## Многомерная функция производящая моменты

Третья задача. Найдите функцию производящую моменты для случайного вектора  $Y \sim \mathcal{N}(\mu, C)$ .

По-определению,  $\text{mgf}_Y(u) = \mathbb{E}(\exp(u^T Y))$ .

Заметим, что  $u^T Y$  это скалярная случайная величина с нормальным распределением  $\mathcal{N}(u^T \mu, u^T C u)$ .

Снова быстро получаем функцию производящую моменты,

$$\text{mgf}_Y(u) = \exp(u^T \mu + u^T C u/2).$$

Производящая функция выглядит как экспоненты от квадратичной функции,

$$\text{mgf}_Y(u) = \exp(q(u)), \quad q(u) = u^T \mu + u^T C u/2.$$

## Доказательство через производящие функции

Четвёртая задача. Для случайного вектора  $Y \sim \mathcal{N}(\mu, C)$  последовательно найдите  $\mathbb{E}(Y_1)$ ,  $\mathbb{E}(Y_1 Y_2)$ ,  $\mathbb{E}(Y_1 Y_2 Y_3)$  и  $\mathbb{E}(Y_1 Y_2 Y_3 Y_4)$ .

Вспомним, что  $\mathbb{E}(Y_1) = \text{mgf}'_1(0)$ ,  $\mathbb{E}(Y_1 Y_2) = \text{mgf}''_{12}(0)$  и так далее.

Немного заранее подготовимся! Во-первых, в нуле  $q(0) = 0$ ,  $\exp(q(0)) = 1$ .

Найдём первую производную  $q'_1(u) = \mu_1 + c_{11}^T u$ , где  $c_1$  — первый столбец матрицы  $C$ . Для наглядности перепишем её в скалярном виде

$$q'_1(u_1, u_2, \dots, u_n) = \mu_1 + c_{11}u_1 + c_{12}u_2 + \dots + c_{1n}u_n.$$

В нуле первая производная равна  $q'_1(0) = \mu_1$  соответствующему ожиданию.

Вторая производная  $q''_{12}(u) = c_{12}$  тождественно равна соответствующей ковариации.

Третья производная  $q'''_{123} = 0$  тождественно равна нулю, ведь  $q(u)$  — квадратичная функция.

А теперь считаем ожидания по очереди,

$$\mathbb{E}(Y_1) = \text{mgf}'_1 = \exp(q)q'_1 = 1 \cdot \mu_1 = \mu_1.$$

Пока что ничего неожиданного, мы же сами обозначили  $\mathbb{E}(Y_1)$  как  $\mu_1$ .

Пойдём дальше!

$$\mathbb{E}(Y_1 Y_2) = \text{mgf}''_{12} = \exp(q)(q'_1 q'_2 + q''_{12}) = \mu_1 \mu_2 + c_{12}.$$


---

Это тождество, верное для любых случайных величин, не только для нормальных,  $\mathbb{E}(Y_1 Y_2) = \mathbb{E}(Y_1) \mathbb{E}(Y_2) + \text{Cov}(Y_1, Y_2)$ .

Продолжаем,

$$\mathbb{E}(Y_1 Y_2 Y_3) = \text{mgf}_{123}'' = \exp(q)(q_1' q_2' q_3' + q_3' q_1'' + q_1' q_2'' + q_2' q_1'') = \mu_1 \mu_2 \mu_3 + \mu_1 c_{23} + \mu_2 c_{13} + \mu_3 c_{12}.$$

Дифференцируя дальше убеждаемся в сохранении закономерности,

$$\mathbb{E}(Y_1 Y_2 Y_3 Y_4) = \mu_1 \mu_2 \mu_3 \mu_4 + \mu_1 \mu_4 c_{23} + \mu_2 \mu_4 c_{13} + \mu_3 \mu_4 c_{12} + \mu_1 \mu_3 c_{24} + \mu_1 \mu_2 c_{34} + \mu_2 \mu_3 c_{14} + c_{12} c_{34} + c_{13} c_{24} + c_{14} c_{23}.$$

## Стайн

TODO: добавить про лемму Стайна и <https://math.stackexchange.com/questions/3507099>

$$\mathbb{E}(X g(X)) = \mu \mathbb{E}(g(X)) + \sigma^2 \mathbb{E}(g'(X)).$$

<https://yedizhang.github.io/Isserlis.html>

TODO: добавить первым элегантно доказательство

Заметим, что формулы Иссерлиса в общем случае не верны для других случайных величин. Например, подбросим правильную монетку два раза. Величина  $X_1$  — индикатор того, что в первый раз выпал орёл,  $X_2$  — индикатор того, что во второй раз выпал орёл, а  $X_3$  — индикатор того, что результаты двух бросков совпали. Все ожидания равны  $\mu_1 = \mu_2 = \mu_3 = 1/2$ . Величины  $X_1$ ,  $X_2$  и  $X_3$  попарно независимы, поэтому все ковариации равны нулю,  $c_{12} = c_{23} = c_{13} = 0$ . Однако зависимы в совокупности,  $\mathbb{E}(X_1 X_2 X_3) = \mathbb{E}(X_1 X_2) = 1/4$ . И формула Иссерлиса не верна,

$$1/4 = \mathbb{E}(X_1 X_2 X_3) \neq \mu_1 \mu_2 \mu_3 + \mu_1 c_{23} + \mu_2 c_{13} + \mu_3 c_{12} = 1/8.$$

TODO: добавить, что верно обобщение на кумулянты

<https://stats.stackexchange.com/questions/616234>

---