

Многомерное нормальное распределение

Борис Демешев, boris.demeshev@gmail.com

14 сентября 2012 г.

Векторные случайные величины

Векторная случайная величина - это просто вектор-столбец из случайных величин:

$$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix} \quad (1.1)$$

В принципе можно говорить о векторе-строке или о случайной матрице, но нам понадобятся только вектор-столбцы. Две самых распространенных характеристики одномерной случайной величины X - это среднее $\mathbb{E}(X)$ и дисперсия $\text{Var}(X)$. У векторных случайных величин также есть среднее:

$$\mathbb{E}(\vec{X}) = \begin{pmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \\ \dots \\ \mathbb{E}(X_n) \end{pmatrix} \quad (1.2)$$

И ковариационная матрица:

$$\text{Var}(\vec{X}) = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \dots & \text{Var}(X_n) \end{pmatrix} \quad (1.3)$$

В ковариационной матрице в i -ой строке в j -ом столбце находится $\text{Cov}(X_i, X_j)$. На диагонали i -ый элемент, это $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$. Из-за этого сразу следует, что ковариационная матрица симметрична.

Для простоты мы будем обозначать транспонирование с помощью штриха, вот так: A' . Путаницы с производной не возникнет.

С помощью транспонирования легко сказать, что ковариационная матрица симметрична: $\text{Var}(X)' = \text{Var}(X)$.

Конечно же, дисперсию можно записать через математическое ожидание: $\text{Var}(X) = E[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

Для ковариационной матрицы это выглядит так:

$$\text{Var}(\vec{X}) = E[(X - \mathbb{E}(X)) \cdot (X - \mathbb{E}(X))'] = \mathbb{E}(XX') - \mathbb{E}(X) \cdot \mathbb{E}(X)' \quad (1.4)$$

Упр. Убедитесь в этом.

Мы быстренько пройдемся по свойствам этих вещей:

Если A - неслучайная матрица и b - неслучайный вектор подходящих размеров, то:

1. $\mathbb{E}(A\vec{X} + b) = A\mathbb{E}(\vec{X}) + b$
2. $\mathbb{E}(\vec{X}A + b) = \mathbb{E}(\vec{X})A + b$

3. $\text{Var}(AX + b) = A \text{Var}(X) A'$.
 4. $\text{Var}(X)$ - неотрицательно определена, имеет n неотрицательных собственных чисел (с учетом кратности), где n размерность вектора X .
- Проверка этих свойств - еще одно упражнение по линейной алгебре.

Краткая схема разных определений

Есть три с половиной подхода определить многомерное нормальное распределение.

Вектор \vec{X} имеет многомерное нормальное распределение, если выполнено одно из трех равносильных условий:

1. Вектор \vec{X} представим в виде $\vec{X} = A\vec{Z} + \mu$, где \vec{Z} — это вектор независимых стандартных нормальных случайных величин
2. Любая линейная комбинация $c_1X_1 + c_2X_2 + \dots + c_nX_n$ имеет одномерное нормальное распределение
3. Характеристическая функция вектора \vec{X} имеет вид:

$$\varphi(\vec{u}) = \exp \left(i \cdot \mu' \cdot u - \frac{1}{2} u' V u \right) \quad (2.1)$$

В случае $\det(V) \neq 0$ первые три формулировки становятся эквивалентны четвертой:

Вектор \vec{X} имеет невырожденное многомерное нормальное распределение, если

1. Функция плотности вектора \vec{X} имеет вид:

$$p(\vec{x}) = (2\pi)^{-n/2} (\det(V))^{-1/2} \cdot \exp \left(-\frac{1}{2} (\vec{x} - \mu) V^{-1} (\vec{x} - \mu)' \right) \quad (2.2)$$

Мы выбираем первое свойство как определение, а остальные будут доказывать как теоремы.

Определение и функция плотности

Очень часто бывает удобно говорить о константе как о нормально распределенной случайной величине с нулевой дисперсией. Это просто соглашение. Никакой функции плотности у константы конечно же нет! Просто когда мы говорим $N(\mu, \sigma^2)$, $\sigma > 0$, мы имеем ввиду «честное» нормальное распределение с функцией плотности

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x - \mu)^2}{2\sigma^2} \right) \quad (2.3)$$

, а когда говорим $N(\mu, 0)$ имеем ввиду просто константу μ .

Т.е. говоря «нормальное распределение» мы теперь можем иметь ввиду и константу. Если же нам надо подчеркнуть, что речь идет о «честном» нормальном распределении с функцией плотности, то мы будем говорить «невырожденное нормальное распределение».

Теперь мы можем дать несколько необычное определение одномерной нормальной величины... Пусть a и b - это константы.

Определение 2.4. Если случайная величина X представима в виде $X = aZ + b$, где $Z \sim N(0, 1)$, то мы говорим, что X имеет **нормальное распределение** $N(b, a^2)$. Если, кроме того, $a \neq 0$, то мы говорим, что X имеет **невырожденное нормальное распределение**.

Теперь мы готовы сказать, что такое многомерное нормальное распределение. Пусть матрица A - неслучайная, вектор b - неслучайный. Размерности подходящие.

Определение 2.5. Если $X = AZ + b$, где вектор Z состоит из независимых $Z_i \sim N(0; 1)$, то мы говорим, что вектор X имеет **нормальное распределение** $N(b, AA')$. Если, кроме того, A обратимая матрица, $\det(A) \neq 0$, то мы говорим, что X имеет **невырожденное нормальное распределение**.

В этой главе и в книге в целом мы **постараемся** уточнять, но обычно из контекста понятно, включает ли автор в понятие нормального распределения вырожденные случаи. Как правило, да, включает.

Пример 3.1. Вектор X имеет вырожденное нормальное распределение, у него нет двумерной функции плотности:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} Z_1 \\ Z_1 \end{pmatrix} \quad (3.2)$$

Невырожденный нормальный вектор имеет функцию плотности...

Теорема 3.3. Вектор X имеет невырожденное нормальное распределение $N(b, AA')$, если и только если его функция плотности имеет вид

$$p(\vec{x}) = (2\pi)^{-n/2} (\det(AA'))^{-1/2} \cdot \exp \left(-\frac{1}{2} (\vec{x} - b)(AA')^{-1} (\vec{x} - b)' \right) \quad (3.4)$$

Если не обращать внимания на константу, которая нужна чтобы интеграл под функцией плотности равнялся единице, то:

$$p(\vec{x}) \sim \exp \left(-\frac{1}{2} (\vec{x} - b)(AA')^{-1} (\vec{x} - b)' \right) \quad (3.5)$$

Доказательство. □

Напоминалка про характеристические функции

Опытный охотник увидев уши, торчащие из кустов, скажет: «Ага, это заяц!» Для этих же целей нужна и характеристическая функция. Она позволяет опознать закон распределения случайной величины.

Узнав например, что характеристическая функция случайной величины W равна $\varphi(u) = \frac{\exp(iu) - 1}{iu}$, опытный охотник скажет: «Ага, это равномерная на $[0; 1]$!». Увидев характеристическую функцию $\varphi(u) = (p \cos(u) + ip \sin(u) + 1 - p)^n$, опытный охотник узнает биномиальное распределение $Bin(n, p)$.

Настала пора для формального определения:

Определение 3.6. Характеристической функцией случайной величины W называется функция $\varphi : \mathbb{R} \rightarrow \mathbb{C}$:

$$\varphi_W(u) := \mathbb{E}(\cos(uW)) + i\mathbb{E}(\sin(uW)) \quad (3.7)$$

Или более кратко:

$$\varphi_W(u) := \mathbb{E}(\exp(iuW)) \quad (3.8)$$

Косинус и синус - ограниченные функции, поэтому для любой случайной величины W существуют $\mathbb{E}(\sin(uW))$ и $\mathbb{E}(\cos(uW))$. А значит и характеристическая функция всегда существует. Тот факт, что по характеристической функции можно узнать закон распределения строго формулируется в виде теоремы:

Теорема 3.9. Пусть X_1 и X_2 — две случайные величины. Функции распределения $F_1(t)$ и $F_2(t)$ совпадают если и только если совпадают характеристические функции $\varphi_1(u)$ и $\varphi_2(u)$

Для описания многомерных нормальных величин характеристическая функция оказывается более удобна, чем функция плотности. Дело в том, что у вырожденных нормальных векторов функции плотности нет, а характеристическая функция существует всегда.

Еще два эквивалентных определения

Теорема 3.10. Вектор X имеет нормальное распределение $N(b, AA')$, возможно вырожденное, если и только если его характеристическая функция имеет вид

$$\varphi(\vec{u}) = \exp\left(i \cdot \mu' \cdot u - \frac{1}{2} u' V u\right) \quad (3.11)$$

Важная особенность многомерного нормального распределения.

Теорема 4.1. Вектор $X = (X_1, \dots, X_n)$ имеет многомерное нормальное распределение (возможно вырожденное) если и только если любая линейная комбинация $Y = c_1 X_1 + c_2 X_2 + \dots + c_n X_n$ имеет одномерное нормальное распределение (возможно вырожденное).

Доказательство. Если X - многомерное нормальное, то $X = AZ + b$. Следовательно, $Y = c_1 X_1 + c_2 X_2 + \dots + c_n X_n = \vec{c}X = \vec{c}AZ + \vec{c}b$. Значит Y по определению 2.5 имеет нормальное распределение.

Если любая линейная комбинация нормальна...

□

Стандартизация

В одномерном случае, если $\text{Var}(X) \neq 0$, то случайную величину X можно «стандартизировать», т.е. превратить в случайную величину с нулевым средним и единичной дисперсией:

$$Z := \frac{X - \mathbb{E}(X)}{\sqrt{\text{Var}(X)}} \quad (4.2)$$

У случайной величины Z : $\mathbb{E}(Z) = 0$, $\text{Var}(Z) = 1$.

В многомерном случае, если матрица $\text{Var}(X)$ обратима, то случайный вектор \vec{X} можно «стандартизовать», т.е. превратить в вектор некоррелированных случайных величин с нулевым средним и единичной дисперсией.

$$\vec{Z} := (\text{Var}(X))^{-1/2}(\vec{X} - \mathbb{E}(\vec{X})) \quad (4.3)$$

Для этой операции нужно понимать, что такое $A^{\frac{1}{2}}$. Для положительно определенной матрицы эта операция корректна. Если A положительно определена, то у нее есть разложение $A = PDP^{-1}$ причем D - диагональная матрица, где на главной диагонали стоят положительные собственные значения A . И, стало быть, $A^{\frac{1}{2}} = PD^{\frac{1}{2}}P^{-1}$.

Если нужно решить какую-то задачу, связанную с многомерным нормальным распределением, то стандартизация может здорово облегчить вычисления. Даже в вырожденном случае имеет смысл перейти к рассмотрению независимых нормальных $N(0; 1)$ случайных величин.

Пример 4.4. Пусть рост любой женщины имеет распределение $N(165, 25)$. А корреляция между ростом матери и ростом дочери равна ρ . Какова вероятность того, что дочь высокая (выше среднего роста), если мама - высокая?

Начинаем решать:

Обозначим X_1 - рост матери, X_2 - рост дочери. $\mathbb{P}(X_2 > 165 | X_1 > 165) = \frac{\mathbb{P}(X_1 > 165 \cap X_2 > 165)}{\mathbb{P}(X_1 > 165)} = 2\mathbb{P}(X_1 > 165 \cap X_2 > 165)$.

Если решать без стандартизации, «в лоб». Вылезает интеграл, который берется усердным трудом...

Если перейти к стандартным Z_1 и Z_2 ...

Многомерное — это больше чем несколько одномерных

Из теоремы 4.1 следует в частности, что:

Теорема 4.5. Если вектор \vec{X} имеет многомерное нормальное распределение, то и любая его компонента X_i имеет нормальное распределение. Кроме того, если распределение \vec{X} — невырожденное, то и распределение каждой компоненты X_i невырожденное.

Обратное утверждение неверно. Мы приведем пример, в котором X_1 и X_2 имеют нормальное распределение по отдельности, но не имеют совместного нормального распределения.

Пример 5.1. Пусть $X_1 \sim N(0; 1)$, а K — это случайная величина равновероятно принимающая значения 1 и -1 , причем K и X_1 независимы. Определим

$$X_2 = K \cdot |X_1| \quad (5.2)$$

Проверим, что X_2 имеет нормальное распределение. Для положительных x :

$$\begin{aligned} \mathbb{P}(X_2 \leq x) &= \mathbb{P}(K = -1) + \mathbb{P}(K = 1) \cdot \mathbb{P}(|X_1| \leq x) = \\ &= 0.5 + 0.5\mathbb{P}(|X_1| \leq x) = 0.5 + \mathbb{P}(X_1 \in [0; x]) = \mathbb{P}(X_1 \leq x) = F(x) \end{aligned} \quad (5.3)$$

Для отрицательных x проверьте сами!

Можно заметить, что $X_1 + X_2$ не является нормально распределенной случайной величиной. Более того, распределение $X_1 + X_2$ не является непрерывным. Действительно,

$$\mathbb{P}(X_1 + X_2 = 0) = \mathbb{P}(X_1 + K|X_1| = 0) = \mathbb{P}(K \neq \text{sign}(X_1)) = 0.5 \quad (5.4)$$

Если бы вектор (X_1, X_2) имел совместное нормальное распределение, то сумма $X_1 + X_2$ была бы нормально распределенной. Кстати, найдем ковариацию между X_1 и X_2 :

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2) = \mathbb{E}(X_1 X_2) = \mathbb{E}(X_1 \cdot K \cdot |X_1|) = \\ &= \mathbb{E}(K)\mathbb{E}(X_1 \cdot |X_1|) = 0 \end{aligned} \quad (5.5)$$

Значит наши X_1 и X_2 были некоррелированы, нормальны по отдельности. Вместе с тем они не были нормальны в совокупности. И конечно, они зависимы, т.к. $|X_1| = |X_2|$.

Некоррелированность и независимость

Для многомерного нормального распределения некоррелированность равносильна независимости.

Доказательство для двумерного...

Условное распределение

Предположим, что вектор (X, Y) имеет совместное нормальное распределение. Часто возникает задача прогнозирования случайной величины Y , если значение случайной величины X известно.

Для удобства вычислений всегда используем стандартизацию! Вместо исходных X и Y рассмотрим:

$$Z_1 = \frac{X - \mathbb{E}(X)}{\sigma_X}, \quad Z_2 = \frac{Y - \mathbb{E}(Y)}{\sigma_Y} \quad (5.6)$$

Естественно, $Z_1 \sim N(0; 1)$ и $Z_2 \sim N(0; 1)$ и $\text{Corr}(Z_1, Z_2) = \text{Corr}(X, Y)$.

Найдем условную функцию плотности Z_2 при известном Z_1 :

$$p(z_2|z_1) = \frac{p(z_1, z_2)}{p(z_1)} \quad (5.7)$$

Поскольку нас интересует только зависимость от z_2 , то на $p(z_1)$ можно не обращать внимания. Значком \sim мы будем обозначать равенство с точностью до константы не зависящей от z_2 :

$$\begin{aligned} p(z_2|z_1) &\sim p(z_1, z_2) \sim \exp \left(-\frac{1}{2}(z_1, z_2) \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right) = \\ &= \exp \left(-\frac{1}{2(1-\rho^2)}(z_1, z_2) \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right) = \\ &= \exp \left(-\frac{1}{2(1-\rho^2)}(z_2^2 - 2\rho z_1 z_2 + z_1^2) \right) \sim \exp \left(-\frac{1}{2(1-\rho^2)}(z_2 - \rho z_1)^2 \right) \end{aligned} \quad (5.8)$$

Сравним полученный результат с функцией плотности одномерного нормального распределения:

$$p(x) \sim \exp \left(-\frac{(x - \mu)^2}{2\sigma^2} \right) \quad (6.1)$$

Замечаем, что $\sigma^2 = 1 - \rho^2$ и $\mu = \rho z_1$. И оформляем результат вычислений в виде теоремы:

Теорема 6.2. Если Z_1 и Z_2 имеют совместное нормальное распределение, $\mathbb{E}(Z_i) = 0$, $\text{Var}(Z_i) = 1$ и $\text{Corr}(Z_1, Z_2) = \rho$, то условное распределение Z_1 при известном Z_2 является нормальным и:

$$\mathbb{E}(Z_2|Z_1 = z) = \rho \cdot z \quad (6.3)$$

$$\text{Var}(Z_2|Z_1 = z) = 1 - \rho^2 \quad (6.4)$$

Формулы дают нам смысл:

1. Корреляция стандартизированных нормальных величин показывает на сколько в среднем растет одна случайная величина при росте другой на единицу.
2. Условная дисперсия Z_2 не зависит от конкретного значения Z_1 . Это несколько неожиданно, но интуитивного объяснения я не знаю.

Для возврата к исходным X и Y замечаем, что условие $X = x$ равносильно тому, что $Z_1 = \frac{x - \mu_x}{\sigma_x}$ и, кроме того, $Y = \mu_y + \sigma_y Z_2$.

Сделав обратную замену, получаем:

Теорема 6.5. Если X и Y имеют совместное нормальное распределение, $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$ и $\text{Corr}(X, Y) = \rho$, то условное распределение Y при известном X является нормальным и:

$$\mathbb{E}(Y|X = x) = \mu_y + \rho \sigma_y \frac{x - \mu_x}{\sigma_x} \quad (6.6)$$

$$\text{Var}(Y|X = x) = (1 - \rho^2) \sigma_y^2 \quad (6.7)$$

Получаем трактовку корреляции: Если X и Y имеют совместное нормальное распределение, то корреляция показывает на сколько своих стандартных отклонений в среднем растет Y при росте X на одно свое стандартное отклонение.

Геометрический смысл ковариационной матрицы

Определение 6.8. Матрица R называется **матрицей поворота** если одновременно выполнены два условия: $R'R = I$ и $\det(R) = 1$

Почему определение поворота именно такое?

Условие $RR' = I$ означает две вещи:

1. Вектор $R\vec{x}$ имеет такую же длину, как и вектор \vec{x} :

$$|\vec{x}|^2 = \vec{x}'\vec{x} = \vec{x}'R'R\vec{x} = (R\vec{x})'(R\vec{x}) = |R\vec{x}|^2 \quad (6.9)$$

2. Угол между \vec{x} и \vec{y} равен углу между $R\vec{x}$ и $R\vec{y}$

$$\cos(\vec{x}, \vec{y}) = \frac{\vec{x}'\vec{y}}{|\vec{x}||\vec{y}|} = \frac{\vec{x}'R'R\vec{y}}{|R\vec{x}||R\vec{y}|} = \frac{(R\vec{x})'(R\vec{y})}{|R\vec{x}||R\vec{y}|} = \cos(R\vec{x}, R\vec{y}) \quad (6.10)$$

Условию $R'R = I$ подходят матрицы с определителем $\det(A) = \pm 1$. Дополнительное условие $\det(R) = 1$ исключает «отражения».

Упражнение. Мы доказали, что матрицы вида $R'R = I$ сохраняют углы и длины. Докажите, что никакие другие матрицы не сохраняют одновременно углы и длины.

Решение. Рассмотрим вектор $e_k = (0, 0, \dots, 0, 1, 0, \dots, 0)'$.

Из линейной алгебры:

Теорема 7.1. Если $A_{n \times n}$ действительная симметричная положительно полу-определенная матрица, то

1. У A имеется ровно n действительных собственных чисел
2. A представима в виде

$$A = RDR^{-1} = RDR' \quad (7.2)$$

, где D — диагональная матрица из собственных чисел матрицы A , а R — матрица поворота из собственных векторов матрицы A .

Для начала сформулируем геометрические факты:

Определение 7.3. Множество точек называется **эллипсоидом**, если оно задается системой уравнений

$$(\vec{x} - \vec{x}_0)'A(\vec{x} - \vec{x}_0) = 1 \quad (7.4)$$

, где A — положительно определенная матрица. Точка \vec{x}_0 — центр эллипсоида. В двумерном случае эллипсоид называют **эллипсом**.

Почему определение эллипса именно такое?

Матрицу A можно представить в виде $A = R'DR$, где R — матрица поворота, а D — диагональная матрица собственных чисел матрицы A . Отсюда получаем, что уравнение эллипса можно записать в виде:

$$(R(\vec{x} - \vec{x}_0))'D(R(\vec{x} - \vec{x}_0)) = 1 \quad (7.5)$$

Если ввести обозначения $\vec{z} = R(\vec{x} - \vec{x}_0)$, то уравнение примет вид:

$$\sum_{i=1}^n d_{ii} z_i^2 = 1 \quad (7.6)$$

Именно в силу этого представления:

Определение 7.7. Для эллипса $(\vec{x} - \vec{x}_0)'A(\vec{x} - \vec{x}_0) = 1$ собственные векторы матрицы A называют **направлениями полуосей**. Если λ_i — это собственное число матрицы A , то величины $1/\sqrt{\lambda_i}$ называют **длинами полуосей**.

Теперь мы готовы нарисовать линии уровня многомерного нормального распределения!

Теорема 7.8. Для невырожденного нормального распределения линии уровня функции плотности являются эллипсоидами. Направления главных осей задаются собственными векторами ковариационной матрицы. Соотношение длин полуосей равно соотношению корней из собственных чисел ковариационной матрицы.

Доказательство. Для доказательства нам потребуется технический факт из линейной алгебры:

Теорема 7.9. Если \vec{a} собственный вектор для матрицы V с собственным числом λ , то \vec{a} собственный вектор для матрицы V^{-1} с собственным числом $1/\lambda$.

Доказательство. Если \vec{a} собственный вектор матрицы V , то с одной стороны:

$$V^{-1} \cdot (V \cdot \vec{a}) = V^{-1}(\lambda \vec{a}) = \lambda V^{-1} \vec{a} \quad (7.10)$$

С другой стороны:

$$(V^{-1} \cdot V) \cdot \vec{a} = \vec{a} \quad (7.11)$$

Т.е. $\lambda V^{-1} \vec{a} = \vec{a}$. Или:

$$V^{-1} \vec{a} = \frac{1}{\lambda} \vec{a} \quad (8.1)$$

□

Итак, пусть $X \sim N(\vec{\mu}; V)$.

Тогда условие

$$p(\vec{x}) = const \quad (8.2)$$

после преобразований равносильно тому, что:

$$(\vec{x} - \vec{\mu}) \cdot V^{-1} \cdot (\vec{x} - \vec{\mu})' = const \quad (8.3)$$

Т.е. мы получили наше определение эллипсоида с $A = \frac{1}{const} V^{-1}$.

Направления полуосей задаются собственными векторами A , значит они совпадают с собственными векторами V .

Длины полуосей обратно пропорциональны корням из собственных чисел A , значит они прямо пропорциональны корням из собственных чисел V . □

Пример 8.4. Пусть $X \sim N(0; V)$ и

$$V = \begin{pmatrix} 5 & 1 \\ 1 & 9 \end{pmatrix} \quad (8.5)$$

Нарисуйте линии уровня функции плотности $p(x_1, x_2)$

Решение.

Тут обязательно картинки. Эллипс. Главные оси. С кодом R и Sage!

В случае независимых нормальных случайных величин с одинаковой дисперсией линиями уровня будут окружности (сферы при более высоких размерностях). Поскольку поворот никак не влияет на линию уровня мы бесплатно получаем следующую теорему:

Теорема 8.6. Если матрица R — это матрица поворота и вектор $\vec{Z} \sim N(\vec{0}, I)$ то вектор $R\vec{Z} \sim N(\vec{0}, I)$

Хи-квадрат распределение

Как известно,

Определение 8.7. Если случайная величина W представима в виде $W = \sum_{i=1}^k Z_i^2$, где Z_i — независимые стандартные нормальные случайные величины, то говорят, что W имеет **хи-квадрат распределение с k степенями свободы**.

К сожалению, проверять, что что-то имеет хи-квадрат распределение напрямую очень неудобно.

Пример 8.8. Давайте попробуем по определению показать, что если $X_i \sim N(\mu; \sigma^2)$ и независимы, то

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \sim \chi_{n-1}^2 \quad (8.9)$$

Как всегда, сначала стандартизируем наши X_i .

...

Гораздо более удобным оказывается следующий способ:

Теорема 8.10. Если вектор \vec{Z} состоит из независимых стандартных нормальных случайных величин и все собственные числа симметричной матрицы A равны либо нулю, либо единице, то

$$\vec{Z}' A \vec{Z} \sim \chi_r^2 \quad (8.11)$$

где r — количество собственных чисел матрицы A равных единице.

Доказательство. Матрица A представима в виде $A = R'DR$. Поэтому:

$$\vec{Z}'A\vec{Z} = \vec{Z}'R'DR\vec{Z} = (R\vec{Z})'D(R\vec{Z}) \quad (9.1)$$

Остается заметить, что $R\vec{Z}$ — это вектор независимых стандартных нормальных случайных величин в силу теоремы 8.6. Если обозначить этот вектор буквой $\vec{W} = R\vec{Z}$, то

$$\vec{Z}'A\vec{Z} = \vec{W}'D\vec{W} = \sum_{i=1}^n d_{ii}W_i^2 \quad (9.2)$$

Матрица D — это матрица собственных чисел матрицы A , т.е. d_{ii} равны либо нулю, либо единице. Получается, что $\vec{Z}'A\vec{Z}$ — это сумма r стандартных независимых нормальных случайных величин. \square

Теорема 9.3. Собственные числа симметричной матрицы A равны либо нулю, либо единице, если и только если $A^2 = A$

Доказательство. С одной стороны:

$$A^2 = (R'DR)(R'DR) = R'D(RR')DR = R'DDR = R'D^2R \quad (9.4)$$

С другой стороны $A = R'DR$. Значит $A^2 = A$ если и только если $D^2 = D$. Но D — диагональная матрица, поэтому условие $D^2 = D$ равносильно тому, что на диагонали стоят либо нули, либо единицы. \square

Пример 9.5. Тот же пример только быстрее... ..

Границы на хвостовые вероятности

При изучении броуновского движения полезны два неравенства.

Теорема 9.6. Если $Z \sim N(0; 1)$, то:

$$\frac{1}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x + x^{-1}} \leq \mathbb{P}(Z \geq x) \leq \frac{1}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x}, \quad x > 0 \quad (9.7)$$

Доказательство. Заметим, что $\exp(-t^2/2)' = -t \exp(-t^2/2)$. Получаем верхнюю границу:

$$\begin{aligned} \mathbb{P}(Z \geq x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt \leq \\ &\leq \int_x^\infty \left(\frac{t}{x} \right) \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt = \frac{1}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x} \end{aligned} \quad (9.8)$$

Заметим, что $(t^{-1} \exp(-t^2/2))' = (1 + t^{-2}) \exp(-t^2/2)$. Получаем нижнюю границу:

$$\begin{aligned} \mathbb{P}(Z \geq x) &= \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt \geq \int_x^\infty \left(\frac{1 + t^{-2}}{1 + x^{-2}} \right) \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt = \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{1 + x^{-2}} \frac{\exp(-x^2/2)}{x} = \frac{1}{\sqrt{2\pi}} \frac{\exp(-x^2/2)}{x + x^{-1}} \end{aligned} \quad (9.9)$$

\square

Теорема 9.10. Если $Z \sim N(0; 1)$, то:

$$\frac{2x}{\sqrt{2\pi e}} \leq \mathbb{P}(|Z| \leq x) \leq \frac{2x}{\sqrt{2\pi}}, \quad 0 < x \leq 1 \quad (9.11)$$

Доказательство. Если $|t| \leq 1$, то

$$\frac{1}{\sqrt{2\pi e}} \leq \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) \leq \frac{1}{\sqrt{2\pi}} \quad (10.1)$$

Интегрируя это неравенство от $-x$ до x получаем требуемое. \square

Еще кусок из блога... (ссылка?) ...

Откуда взялось π в формуле?

Есть много способов объяснить, откуда берется π в формуле функции плотности. Вот то, которое нравится мне.

Рассмотрим пару независимых стандартных нормальных случайных величин X и Y . Их функция плотности имеет вид:

$$p(x, y) = c \cdot \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \quad (10.2)$$

Рассмотрим более подробно функцию

$$f(x, y) = \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \quad (10.3)$$

Значение $f(x, y)$ зависит только от расстояния от точки (x, y) до начала координат. Значит объем под «шляпой» является фигурой вращения.

Картинка (слева и справа повернутая):

...

Уже понятно, что π — в деле. Остается вспомнить, что объем фигуры вращения определяется по формуле $Vol = \int_a^b \pi r^2(t) dt$.

В нашем случае: $a = 0$, $b = f(0, 0) = 1$, а $r(t)$ находится из условия:

$$\exp\left(-\frac{1}{2}r^2(t)\right) = t \quad (10.4)$$

Находим $r^2(t)$ и получаем:

$$r^2(t) = -2\ln(t) \quad (10.5)$$

Находим объем фигуры вращения:

$$Vol = \int_0^1 \pi r^2(t) dt = \int_0^1 \pi(-2\ln(t)) dt = -2\pi \int_0^1 \ln(t) dt = 2\pi \quad (10.6)$$

Интеграл от $\ln(t)$ можно взять либо по частям, либо заметив, что:

$$\int_0^1 \ln(t) dt = -\int_0^\infty \exp(-t) dt \quad (10.7)$$

Картинка:

Настоящие знатоки берут интеграл $\int_0^\infty \exp(-t)dt$ по методу Мамикона Мнацаканяна [1] в уме глядя на картинку:

Пояснение к картинке:

Производная функции $\exp(-x)$ равна ей самой умноженной на минус один. Поэтому тень от касательной всегда имеет длину один. Интересующая нас площадь равна площади треугольника плюс оставшаяся площадь. Касательные заметающие оставшуюся площадь можно перенести так, чтобы они замели треугольник. Значит интеграл равен удвоенной площади треугольника.

Мы доказали, что объем под функцией $f(x, y)$ равен 2π . Следовательно, функция плотности $p(x, y)$ должна иметь вид:

$$p(x, y) = \frac{1}{2\pi} \cdot \exp\left(-\frac{1}{2}(x^2 + y^2)\right) \quad (11.1)$$

Для независимых величин $p(x, y) = p(x)p(y)$, следовательно:

$$p(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{1}{2}x^2\right) \quad (11.2)$$

ЦПТ — доказательство через характеристические функции

Теорема 11.3. Пусть X_n — последовательность случайных величин с характеристическими функциями $\varphi_n(u)$. Пусть кроме того, X — случайная величина с характеристической функцией $\varphi(u)$. Последовательность X_n сходится по распределению к случайной величине X если и только если последовательность функций $\varphi_n(u)$ сходится поточечно к функции $\varphi(u)$.

Список литературы

[1] Tom Apostol. A visual approach to calculus problems. *Engineering and science*, 2000.

Как находить площади фигур по методу Мамикона Мнацаканяна! Площадь под экспонентой или площадь между следом заднего и переднего колеса велосипеда... Без интегрирования!

<http://www.cco.caltech.edu/~mamikon/>.