Учебник по Эконометрике

Лекция 1: Векторные Пространства

Джон Стачурски

Лекции: Акшай Шенкер

Перевел: Алексей Кедо

1 октября 2020 г.

Обзор

Линейная алгебра является основой математики и, в частности, эконометрики:

- проведение базовых вычислений с данными
- решение линейных уравнений используя данные
- продвинутые операции, такие как квадратичная минимизация

В центре внимания данной главы:

1. векторные пространства: линейные операции, нормы, линейные подпространства, линейная независимость, базисы и т.д.
2. теорема ортогональной проекции

Векторное пространство

Символ $\mathbb{R}^N$ показывает набор любых векторов длины N, или N векторов

N-вектор $\mathbf{x}$ – это список из N действительных чисел:

$$\mathbf{x} = (x_1, \ldots, x_N) \text{ ,где } \quad x_n \in \mathbb{R} \text{ для любого } n$$

Также мы можем записать $\mathbf{x}$ вертикально, вот так:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

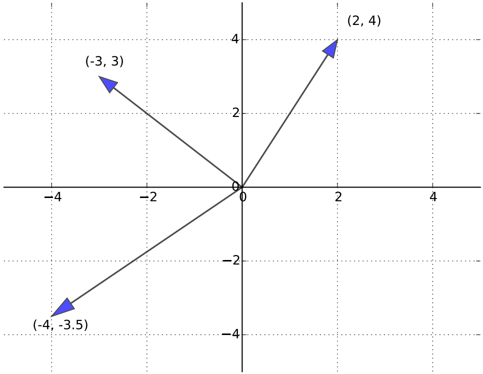


Рис.: Три вектора в $\mathbb{R}^2$

Вектор из единиц будет обозначен $\mathbf{1}$

$$\mathbf{1} := \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

Вектор из нулей будет обозначен $\mathbf{0}$

$$\mathbf{0} := \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Линейные операции

Две базовые алгебраические операции:

1. Сложение векторов
2. Умножение на скаляр

1. **Сумма** $\mathbf{x} \in \mathbb{R}^N$ и $\mathbf{y} \in \mathbb{R}^N$ обозначается

$$\mathbf{x} + \mathbf{y} =: \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_N + y_N \end{pmatrix}$$

Пример 1:

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + \begin{pmatrix} 2 \\ 4 \\ 6 \\ 8 \end{pmatrix} := \begin{pmatrix} 3 \\ 6 \\ 9 \\ 12 \end{pmatrix}$$

Пример 2:

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} := \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$$

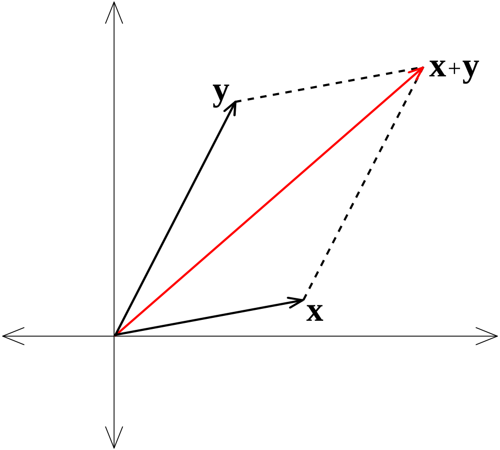


Рис.: Сложение векторов

2. **Умножение на скаляр** $\alpha \in \mathbb{R}$ и $\mathbf{x} \in \mathbb{R}^N$ обозначается

$$\alpha \mathbf{x} = \alpha \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} := \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_N \end{pmatrix}$$

Пример 1:

$$0.5 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} := \begin{pmatrix} 0.5 \\ 1.0 \\ 1.5 \\ 2.0 \end{pmatrix}$$

Пример 2:

$$-1 \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} := \begin{pmatrix} -1 \\ -2 \\ -3 \\ -4 \end{pmatrix}$$

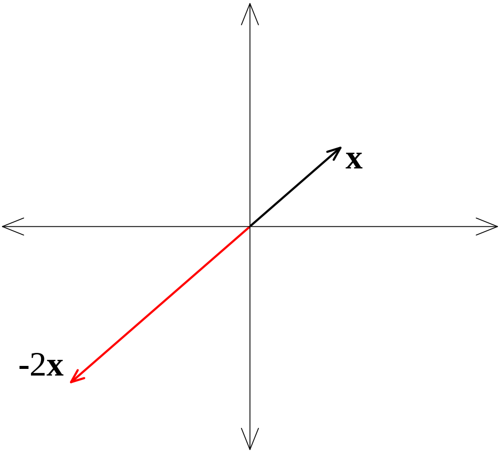


Рис.: Уножение на скаляр

Вычитание выполняется поэлементно, аналогично умножению

$$\mathbf{x} - \mathbf{y} := \begin{pmatrix} x_1 - y_1 \\ x_2 - y_2 \\ \vdots \\ x_N - y_N \end{pmatrix}$$

Определение можно дать в терминах сложения и умножения на скаляр

$$\mathbf{x} - \mathbf{y} := \mathbf{x} + (-1)\mathbf{y}$$

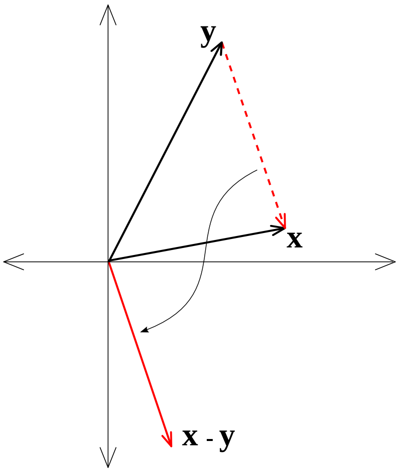


Рис.: Разница векторов

Скалярное произведение

Скалярное произведение двух векторов $\mathbf{x}$ и $\mathbf{y}$ в $\mathbb{R}^N$ обозначается $\langle \mathbf{x}, \mathbf{y} \rangle$, и является суммой произведения их элементов:

$$\langle \mathbf{x}, \mathbf{y} \rangle := \sum_{n=1}^{N} x_n y_n$$

Факт. (??)

Для любых $\alpha, \beta \in \mathbb{R}$ и любых $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, верны следующие утверждения:

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$,
2. $\langle \alpha \mathbf{x}, \beta \mathbf{y} \rangle = \alpha\beta \langle \mathbf{x}, \mathbf{y} \rangle$, и
3. $\langle \mathbf{x}, \alpha \mathbf{y} + \beta \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle + \beta \langle \mathbf{x}, \mathbf{z} \rangle$.

Свойства можно легко проверить с помощью определений умножения на скаляр и скалярного произведения.

Для 2., например, возьмите любые $\alpha, \beta \in \mathbb{R}$ и любые $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$:

$$\langle \alpha \mathbf{x}, \beta \mathbf{y} \rangle = \sum_{n=1}^{N} \alpha x_n \beta y_n = \alpha\beta \sum_{n=1}^{N} x_n y_n = \alpha\beta \langle \mathbf{x}, \mathbf{y} \rangle$$

Нормы и расстояние

Норма (Эвклида) $\mathbf{x} \in \mathbb{R}^N$ обозначается

$$\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$$

Интерпретация:

- $\|\mathbf{x}\|$ показывает "длину" вектора $\mathbf{x}$
- $\|\mathbf{x} - \mathbf{y}\|$ показывает расстояние между векторами $\mathbf{x}$ и $\mathbf{y}$

Факт. (??) Для любого $\alpha \in \mathbb{R}$ и любых $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$, верны следующие утверждения:

1. $\|\mathbf{x}\| \geq 0$ и $\|\mathbf{x}\| = 0$ тогда и только тогда, когда $\mathbf{x} = \mathbf{0}$
2. $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (неравенство треугольника)
4. $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$ (неравенство Коши — Буняковского)

Свойства 1. и 2. легко доказать (упражнение)

Свойство 4. рассматривается в ЕТ упражнении ??

Покажем доказательство свойства 3. с помощью свойств скалярного произведения (факт ??)

$$\begin{aligned}
\|\mathbf{x}+\mathbf{y}\|^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y} \rangle \\
&= \langle \mathbf{x}, \mathbf{x} \rangle + 2\langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \\
&\leq \langle \mathbf{x}, \mathbf{x} \rangle + 2|\langle \mathbf{x}, \mathbf{y} \rangle| + \langle \mathbf{y}, \mathbf{y} \rangle
\end{aligned}$$

Применим неравенство Коши — Буняковского

$$\|\mathbf{x}+\mathbf{y}\|^2 \leq (\|\mathbf{x}\|+\|\mathbf{y}\|)^2$$

Убираем квадраты и получаем неравенство треугольника

Линейная комбинация векторов $\mathbf{x}_1, \ldots, \mathbf{x}_K$ in $\mathbb{R}^N$ — это вектор

$$\mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_K \mathbf{x}_K$$

где $\alpha_1, \ldots, \alpha_K$ скаляры

Пример.

$$0.5 \begin{pmatrix} 6.0 \\ 2.0 \\ 8.0 \end{pmatrix} + 3.0 \begin{pmatrix} 0 \\ 1.0 \\ -1.0 \end{pmatrix} = \begin{pmatrix} 3.0 \\ 4.0 \\ 1.0 \end{pmatrix}$$

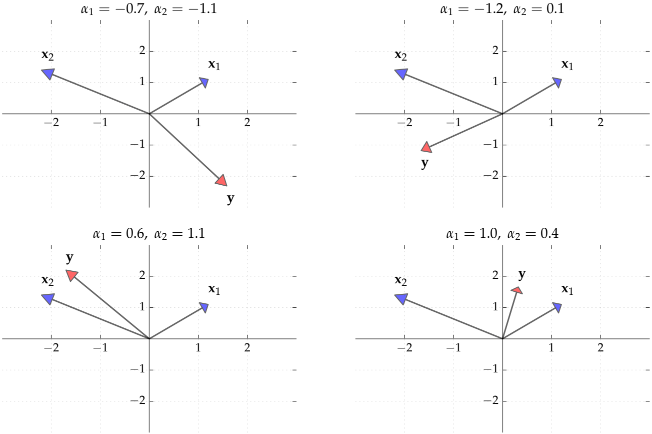


Рис.: Линейные комбинации $\mathbf{x}_1, \mathbf{x}_2$

Линейная оболочка

Пусть $X \subset \mathbb{R}^N$ некое непустое множество

Множество всех возможных линейных комбинаций X называют **линейной оболочкой** X, обозначается $\operatorname{span}(X)$

Для конечного $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ линейную оболочку можно записать так

$$\operatorname{span}(X) := \left\{ \text{все } \sum_{k=1}^{K} \alpha_k \mathbf{x}_k, \text{ где } (\alpha_1, \ldots, \alpha_K) \in \mathbb{R}^K \right\}$$

Пример. Четыре вектора, обозначенные $\mathbf{y}$ на предыдущем рисунке, лежат в линейной оболочке $X = \{\mathbf{x}_1, \mathbf{x}_2\}$

Может ли *любой* вектор в $\mathbb{R}^2$ быть создан линейной комбинацией $\mathbf{x}_1, \mathbf{x}_2$?

Ответ да. Мы докажем это в §**??**

Пример. Пусть $X = \{\mathbf{1}\} \subset \mathbb{R}^2$, где $\mathbf{1} := (1,1)$

Линейная оболочка X состоит из всех векторов вида

$$\alpha\mathbf{1} = \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} \quad , \text{ где } \quad \alpha \in \mathbb{R}$$

Представляет собой прямую на плоскости, которая проходит через

- вектор $\mathbf{1}$ (при $\alpha = 1$)
- начало координат $\mathbf{0}$ (при $\alpha = 0$)

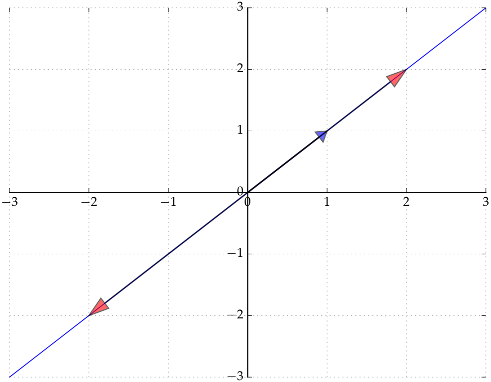


Рис.: Линейная оболочка $\mathbf{1} := (1, 1)$ в $\mathbb{R}^2$

Пример. Набор канонически базисных векторов $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ линейно независим в $\mathbb{R}^N$

Доказательство. Пусть $\alpha_1, \ldots, \alpha_N$ коэффициенты, такие что $\sum_{k=1}^{N} \alpha_k \mathbf{e}_k = \mathbf{0}$

Эквивалентно,

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \sum_{k=1}^{N} \alpha_k \mathbf{e}_k = \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

В частности, $\alpha_k = 0$ для всех k

Пример. Пусть $\mathbf{x}_1 = (3, 4, 2)$ и $\mathbf{x}_2 = (3, -4, 0.4)$

По определению, линейная оболочка — все возможные вектора в виде

$$\mathbf{y} = \alpha \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} + \beta \begin{pmatrix} 3 \\ -4 \\ 0.4 \end{pmatrix} \quad \text{, где} \quad \alpha, \beta \in \mathbb{R}$$

Это плоскость, проходящая через

- вектор $\mathbf{x}_1$
- вектор $\mathbf{x}_2$
- начало координат $\mathbf{0}$

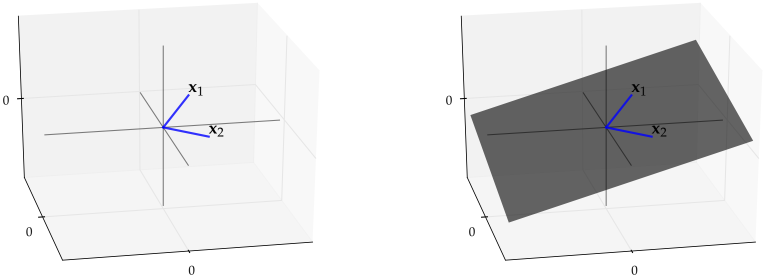


Рис.: Линейная оболочка $\mathbf{x}_1, \mathbf{x}_2$

Пример. Рассмотрим векторы $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\} \subset \mathbb{R}^N$, где

$$\mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \mathbf{e}_N := \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}$$

$\mathbf{e}_n$ состоит из нулей, кроме n-ого элемента равного 1

Вектора $\mathbf{e}_1, \ldots, \mathbf{e}_N$ назвают **канонически базисными векторами** of $\mathbb{R}^N$

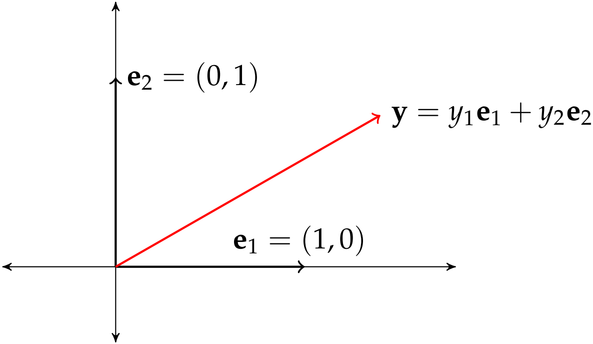


Рис.: Базисные векторы в $\mathbb{R}^2$

Пример. (прод.)

Линейная оболочка $\{\mathbf{e}_1, \ldots, \mathbf{e}_N\}$ эквивалентна всему $\mathbb{R}^N$

Доказательство для $N = 2$:

Пусть $\mathbf{y} \in \mathbb{R}^2$, тогда

$$\mathbf{y} := \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y_2 \end{pmatrix}$$

$$= y_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} = y_1 \mathbf{e}_1 + y_2 \mathbf{e}_2$$

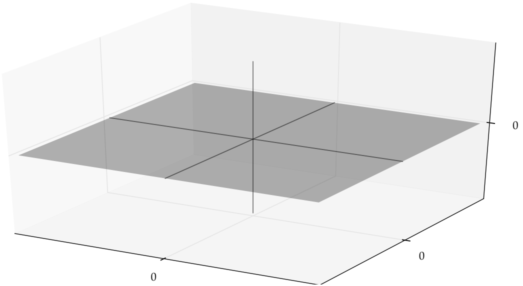
Таким образом, $\mathbf{y} \in \operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\}$

Так как $\mathbf{y}$ произвольный, мы показали, что $\operatorname{span}\{\mathbf{e}_1, \mathbf{e}_2\} = \mathbb{R}^2$

Пример. Рассмотрим множество

$$P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$$

Графически, $P =$ плоскость в $\mathbb{R}^3$с координатой высоты $= 0$



Пример. (прод.)

Если $\mathbf{e}_1 = (1,0,0)$ и $\mathbf{e}_2 = (0,1,0)$, тогда $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} = P$

Чтобы подтвердить утверждение, возьмем $\mathbf{x} = (x_1, x_2, 0)$, любой элемент P. Можно записать $\mathbf{x}$ в следующем виде

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ 0 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2$$

Другими словами, $P \subset \text{span}\{\mathbf{e}_1, \mathbf{e}_2\}$

И наоборот, у нас имеется $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} \subset P$ (почему?)

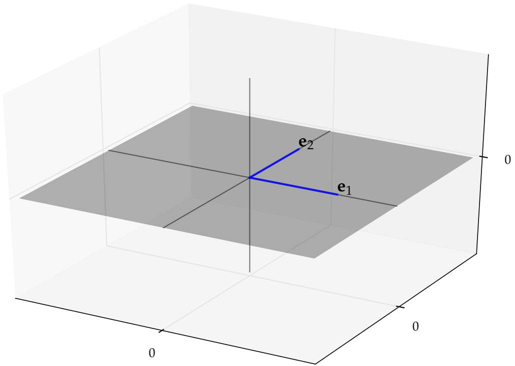


Рис.: $\text{span}\{\mathbf{e}_1, \mathbf{e}_2\} = P$

Факт. (??) Если X и Y непустые подмножества $\mathbb{R}^N$ и $X \subset Y$, тода $\operatorname{span}(X) \subset \operatorname{span}(Y)$

Доказательство. Возьмем любой непустой $X \subset Y \subset \mathbb{R}^N$

Пусть $\mathbf{z} \in \operatorname{span}(X)$, тогда имеется

$$\mathbf{z} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k \text{ для некоторых } \mathbf{x}_1, \ldots, \mathbf{x}_K \in X,\ \alpha_1, \ldots, \alpha_K \in \mathbb{R}$$

Доказательство. (прод.) Так как $X \subset Y$, каждый $\mathbf{x}_k$ также находится в Y, получается

$$\mathbf{z} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k \text{ для некоторых } \mathbf{x}_1, \dots, \mathbf{x}_K \in Y,\ \alpha_1, \dots, \alpha_K \in \mathbb{R}$$

Значит, $\mathbf{z} \in \operatorname{span}(Y)$

Линейная независимость

Важные вопросы:

- Когда матрица обратима?
- Когда аргументы регрессии страдают от коллинеарности?
- Когда система линейных уравнений имеет решение?

Все эти вопросы тесно связаны с линейной независимостью

Непустое множество векторов $X := \{\mathbf{x}_1, \dots, \mathbf{x}_K\} \subset \mathbb{R}^N$ называется **линейно независимым**, если

$$\sum_{k=1}^{K} \alpha_k \mathbf{x}_k = \mathbf{0} \implies \alpha_1 = \cdots = \alpha_K = 0$$

Пример. Рассмотрим $\mathbf{x}_1 = (1.2, 1.1)$ и $\mathbf{x}_2 = (-2.2, 1.4)$

Пусть α_1 и α_2 скаляры, такие что

$$\alpha_1 \begin{pmatrix} 1.2 \\ 1.1 \end{pmatrix} + \alpha_2 \begin{pmatrix} -2.2 \\ 1.4 \end{pmatrix} = \mathbf{0}$$

Это является линейной системой из двух уравнений α_1 и α_2

Единственное решение $\alpha_1 = \alpha_2 = 0$

$\{\mathbf{x}_1, \mathbf{x}_2\}$ линейно независимы

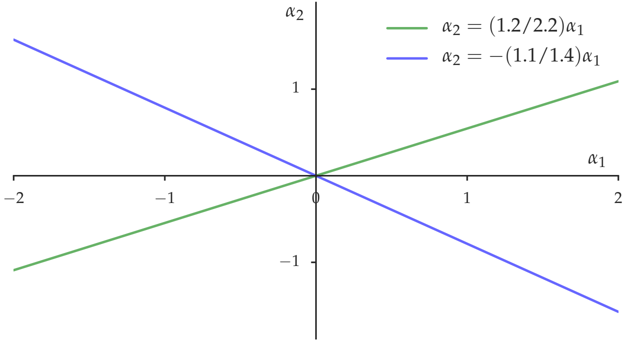


Рис.: Единственное решение $\alpha_1 = \alpha_2 = 0$

Пример. Базисные векторы $\{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ линейно независимы в $\mathbb{R}^N$

Проверим это. Пусть $\alpha_1, \dots, \alpha_N$ коэффициенты, такие что $\sum_{k=1}^{N} \alpha_k \mathbf{e}_k = \mathbf{0}$, тогда

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{pmatrix} = \sum_{k=1}^{N} \alpha_k \mathbf{e}_k = \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

То есть $\alpha_k = 0$ для всех k

Таким образом, $\{\mathbf{e}_1, \dots, \mathbf{e}_N\}$ линейно независимы

Теорема. (??) Пусть $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \subset \mathbb{R}^N$. Для $K > 1$, следующие утверждения эквивалентны:

1. X линейно независим
2. X_0 — подходящее подмножество $X \implies \operatorname{span} X_0$ — подходящее подмножество $\operatorname{span} X$
3. Ни один из векторов X не может быть записан как линейная комбинация оставшихся

Доказательство есть в упражнениях. См. ET упр. **??** и решение

Пример. Если убрать любой базисный вектор, линейная оболочка уменьшится

Рассмотрим случай с $N = 2$

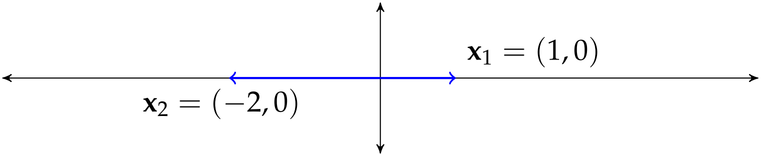
Мы знаем, что $\mathrm{span}\{\mathbf{e}_1, \mathbf{e}_2\} = \mathbb{R}^2$

- Если убрать любой элемент из $\mathrm{span}\{\mathbf{e}_1, \mathbf{e}_2\}$, линейная оболочка превратится в прямую.

Тем не менее, пусть $\mathbf{x}_1 = (1, 0)$ и $\mathbf{x}_2 = (-2, 0)$

Векторы не являются линейно независимыми, так как $\mathbf{x}_2 = -2\mathbf{x}_1$

- Если убрать любой из векторов, линейная оболочка не изменится — останется горизонтальной прямой
- имеется $\mathbf{x}_2 = -2\mathbf{x}_1$, значит любой вектор может быть записан как линейная комбинация оставшегося



Факт. (**??**) Если $X := \{\mathbf{x}_1, \dots, \mathbf{x}_K\}$ линейно независимы, тогда

1. любое подмножество X линейно независимо,
2. X не содержит $\mathbf{0}$, и
3. $X \cup \{\mathbf{x}\}$ линейно независимы для всех $\mathbf{x} \in \mathbb{R}^N$, таких что $\mathbf{x} \notin \operatorname{span} X$.

Доказательство показано в упражнении (упр. **??** в ЕТ)

Линейная независимость и единственность

Линейная независимость - ключевое условие для того, чтобы решение системы линейных уравнений существовало *и* было единственным

Теорема. (??) Пусть $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ некоторое множество векторов в $\mathbb{R}^N$. Следующие утверждения эквивалентны:

1. X линейно независим
2. Для каждого $\mathbf{y} \in \mathbb{R}^N$ существует не более одного множества скаляров $\alpha_1, \ldots, \alpha_K$, такого что

$$\mathbf{y} = \alpha_1 \mathbf{x}_1 + \cdots + \alpha_K \mathbf{x}_K \tag{1}$$

Доказательство. (1. $\implies$ 2.)

Пусть X линейно независим, возьмем любой $\mathbf{y}$

Предположим противоположное — (1) выполняется для нескольких наборов скаляров, получается

$$\exists\ \alpha_1, \ldots, \alpha_K \text{ и } \beta_1, \ldots, \beta_K \ \text{ s.t. } \ \mathbf{y} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k = \sum_{k=1}^{K} \beta_k \mathbf{x}_k$$

$$\therefore \quad \sum_{k=1}^{K} (\alpha_k - \beta_k)\mathbf{x}_k = \mathbf{0}$$

$$\therefore \quad \alpha_k = \beta_k \quad \text{для всех} \quad k$$

Доказательство. (2. $\implies$ 1.)

Если 2. выполняется, то существует не более одного набора скаляров, такого что

$$\mathbf{0} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k$$

Так как при $\alpha_1 = \cdots = \alpha_k = 0$ это равенство выполняется, больше не существует скаляров, при которых $\mathbf{0} = \sum_{k=1}^{K} \alpha_k \mathbf{x}_k$

Значит, X линейно независим по определению

Линейные подпространства

Непустое подмножество S множества $\mathbb{R}^N$ называется **линейным подпространством** (или просто **подпространством**) множества $\mathbb{R}^N$, если

$$\mathbf{x}, \mathbf{y} \in S \text{ и } \alpha, \beta \in \mathbb{R} \implies \alpha\mathbf{x} + \beta\mathbf{y} \in S$$

Другими словами, $S \subset \mathbb{R}^N$ 'заперт' с векторным сложением и умножением на скаляр

Пример. Если X — некое непустое подмножество $\mathbb{R}^N$, тогда $\operatorname{span} X$ — линейное подпространство $\mathbb{R}^N$

Пример. $\mathbb{R}^N$ — линейное подпространство $\mathbb{R}^N$

Пример. Возьмем любой вектор $\mathbf{a} \in \mathbb{R}^N$, множество $A := \{\mathbf{x} \in \mathbb{R}^N : \langle \mathbf{a}, \mathbf{x} \rangle = 0\}$ является линейным подпространством $\mathbb{R}^N$

Чтобы показать это, пусть $\mathbf{x}, \mathbf{y} \in A$, $\alpha, \beta \in \mathbb{R}$ и $\mathbf{z} := \alpha\mathbf{x} + \beta\mathbf{y} \in A$

Получается

$$\langle \mathbf{a}, \mathbf{z} \rangle = \langle \mathbf{a}, \alpha\mathbf{x} + \beta\mathbf{y} \rangle = \alpha \langle \mathbf{a}, \mathbf{x} \rangle + \beta \langle \mathbf{a}, \mathbf{y} \rangle = 0 + 0 = 0$$

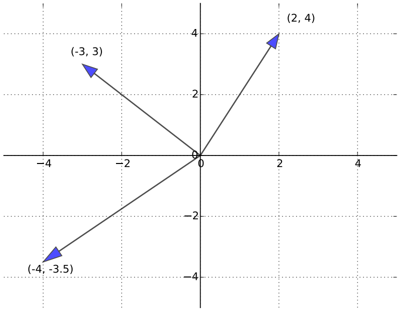
Значит, $\mathbf{z} \in A$

Факт. (??) Если S — линейное подпространство $\mathbb{R}^N$, тогда

1. $\mathbf{0} \in S$
2. $X \subset S \implies \operatorname{span} X \subset S$, и
3. $\operatorname{span} S = S$

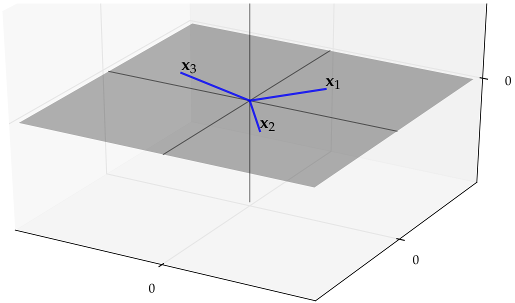
Теорема. (??) Пусть S — линейное подпространство $\mathbb{R}^N$. Если S охватывает K векторов, тогда любое линейно независимое подмножество S имеет не более K векторов

Пример. Вспомним базисные векторы $\{\mathbf{e}_1, \mathbf{e}_2\}$, охватывающие $\mathbb{R}^2$. Из Теоремы **??** следует, что три вектора ниже линейно зависимы



Пример. Плоскость $P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$ из примера **??** в ЕТ может быть образована двумя векторами

По теореме **??**, три вектора на рисунке ниже линейно независимы



Базисы и размерность

Теорема. (??) Пусть $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ — некие N векторов в $\mathbb{R}^N$. Следующие утверждения эквивалентны:

1. $\operatorname{span} X = \mathbb{R}^N$
2. X линейно независим

Для доказательства смотрите **??** в ET

Пусть S — линейное подпространство $\mathbb{R}^N$ и $B \subset S$

Множество B называется **базисом** S, если

1. B охватывает S и
2. B линейно независимо

Из Теоремы **??**, когда B является базисом S, каждая точка S имеет ровно представление как линейной комбинации элементов B

Из Теоремы **??**, любые N линейно независимых векторов в $\mathbb{R}^N$ формируют базис в $\mathbb{R}^N$

Пример. Вспомним плоскость из примера выше

$$P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$$

Мы показали, что $\mathrm{span}\{\mathbf{e}_1, \mathbf{e}_2\} = P$ для

$$\mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Более того, $\{\mathbf{e}_1, \mathbf{e}_2\}$ линейно независимы (почему?)

Значит, $\{\mathbf{e}_1, \mathbf{e}_2\}$ являются базисом для P

Теорема. (**??**) Если S — линейное подпространство $\mathbb{R}^N$ отличное от $\{\mathbf{0}\}$, тогда

1. S имеет по меньшей мере один базис и
2. каждый базис S имеет одинаковое количество элементов.

Если S — линейное подпространство $\mathbb{R}^N$, то обычное число, определенное в Теореме **??** называется **размерностью** S, и обозначается $\dim S$

Пример. Для $P := \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1, x_2 \in \mathbb{R}\}$, $\dim P = 2$, потому что

$$\mathbf{e}_1 := \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 := \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

базис имеет два элемента

Пример. Прямая $\{\alpha \mathbf{x} \in \mathbb{R}^N : \alpha \in \mathbb{R}\}$, проходящая через начало координат, имеет размерность 1

Говорят, что в $\mathbb{R}^N$ одноэлементное подпространство $\{\mathbf{0}\}$ имеет нулевую размерность

Возьмем множество из K векторов, насколько большой будет его линейная оболочка с точки зрения размерности?

Теорема. (??) Если $X := \{\mathbf{x}_1, \ldots, \mathbf{x}_K\} \subset \mathbb{R}^N$, то

1. $\dim \operatorname{span} X \leq K$ и
2. $\dim \operatorname{span} X = K$ тогда и только тогда, когда X линейно независим

Для доказательства смотрите упражнение **??** в ET

Факт. (??) Следующие утверждения верны:

1. Пусть S и S' являются линейными подпространствами $\mathbb{R}^N$ размерности K. Если $S \subset S'$, то $S = S'$
2. Если S — линейное подпространство $\mathbb{R}^N$ размерности M и $M < N$, то $S \neq \mathbb{R}^N$

Линейные отображения

Функция $T\colon \mathbb{R}^K \to \mathbb{R}^N$ называется **линейной**, если

$$T(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha T\mathbf{x} + \beta T\mathbf{y} \qquad \forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^K,\ \forall\, \alpha, \beta \in \mathbb{R}$$

Примечание:

- Линейные функции обычно записываются с большой буквы
- Обычно, когда это удобно, аргументы в скобках опускаются

Пример. Функция $T\colon \mathbb{R} \to \mathbb{R}$, определяемая как $Tx = 2x$, линейна

Чтобы увидеть это, возьмем любые α, β, x, y в $\mathbb{R}$, тогда

$$T(\alpha x + \beta y) = 2(\alpha x + \beta y) = \alpha 2x + \beta 2y = \alpha Tx + \beta Ty$$

Пример. Функция $f\colon \mathbb{R} \to \mathbb{R}$, определяемая как $f(x) = x^2$, нелинейна

Чтобы увидеть это, возьмем $\alpha = \beta = x = y = 1$, получается

$$f(\alpha x + \beta y) = f(2) = 4$$

Тем не менее, $\alpha f(x) + \beta f(y) = 1 + 1 = 2$

Примечание: Неправильно думать о линейных функциях как о тех, чей график является прямой

Пример. Функция $f: \mathbb{R} \to \mathbb{R}$, определяемая $f(x) = 1 + 2x$, нелинейна

Возьмем $\alpha = \beta = x = y = 1$. Имеется

$$f(\alpha x + \beta y) = f(2) = 5$$

Тем не менее, $\alpha f(x) + \beta f(y) = 3 + 3 = 6$

Такой вид функции называется **афинной** функцией

По определению, если T линейна, то изменение порядка в

$$T[\sum_{k=1}^{K} \alpha_k \mathbf{x}_k] = \sum_{k=1}^{K} \alpha_k T \mathbf{x}_k$$

будет действительным пока $K = 2$

Индукция расширяет это до произвольных K

Факт. (??) Если $T\colon \mathbb{R}^K \to \mathbb{R}^N$ является линейным отображением, то

$$\operatorname{rng}(T) = \operatorname{span}(V) \quad \text{, где} \quad V := \{T\mathbf{e}_1, \dots, T\mathbf{e}_K\}$$

где $\mathbf{e}_k$ является k-ым базисным вектором в $\mathbb{R}^K$

Доказательство. Любой $\mathbf{x} \in \mathbb{R}^K$ может быть выражен как $\sum_{k=1}^K \alpha_k \mathbf{e}_k$. Значит, $\operatorname{rng}(T)$ является множеством всех точек следующей формы

$$T\mathbf{x} = T\left[\sum_{k=1}^K \alpha_k \mathbf{e}_k\right] = \sum_{k=1}^K \alpha_k T\mathbf{e}_k$$

так как $\alpha_1, \dots, \alpha_K$ меняется для всех комбинаций. Это совпадает с определением $\operatorname{span}(V)$

Ядром линейного отображения $T\colon \mathbb{R}^K \to \mathbb{R}^N$ называют

$$\ker(T) := \{\mathbf{x} \in \mathbb{R}^K : T\mathbf{x} = \mathbf{0}\}$$

Факт. (??) Если $T\colon \mathbb{R}^K \to \mathbb{R}^N$ линейное отображение, то

$$\operatorname{rng} T = \operatorname{span} V, \quad \text{где } V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$$

Доказательства простые (выполните в качестве упражнения)

Линейная независимость и биекция

Много научных и практических задач являются задачами 'обратными'

- мы наблюдаем результаты, но не их причины
- как можно работать в обратном порядке, от результатов к причинам?

Примеры

- какие предпочтения потребителей привели к наблюдаемому рыночному поведению?
- какие ожидания привели к данному сдвигу обменных курсов?

В общем, обратную задачу можно выразить как

$$\overset{\text{model}}{F}(x) = \overset{\text{outcome}}{y}$$

what x led to outcome y?

- имеет ли эта задача решение?
- является ли оно едиственным?

Ответы зависят от того, является ли F сюръекцией, инъекцией и т.д.

Лучший вариант — биекция

Но возникают и другие ситуации

Теорема. (**??**) Если T — линейная функция из $\mathbb{R}^N$ в $\mathbb{R}^N$, то все следующие утверждения эквивалентны:

1. T является биекцией.
2. T является сюръекцией.
3. T является инъекцией.
4. $\ker T = \{\mathbf{0}\}$.
5. $V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_N\}$ линейно независимо.
6. $V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_N\}$ формирует базис $\mathbb{R}^N$.

Смотрите упражнение **??** в ЕТ для доказательства

Если любое из этих условий выполняется, то T называют **несингулярной**. В ином случае T назвают **сингулярной**

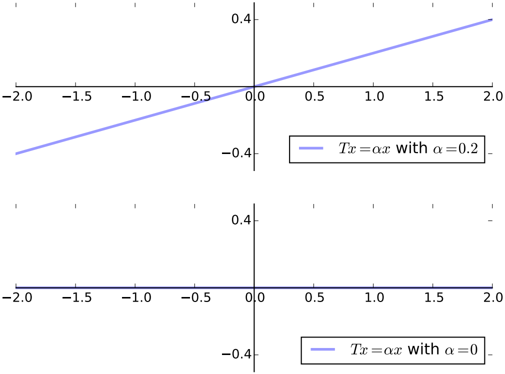


Рис.: Случай с $N = 1$, несингулярная и сингулярная функции

Если T несингулярна, то, будучи биекцией, она должна иметь обратную функцию T^{-1}, которая так же является биекцией (факт **??** на странице **??**)

Факт. (**??**) Если $T\colon \mathbb{R}^N \to \mathbb{R}^N$ несингулярна, то и T^{-1} несингулярна.

Для доказательства, смотрите упр. **??**

Отображения при различных размерностях

Помните, что результаты выше применимы к отображениям из $\mathbb{R}^N$ в $\mathbb{R}^N$

Все меняется, когда мы смотрим на линейные отображения при различных размерностях

Общие правили для линейных отображений:

- отображения из меньших в большие размерности не могут быть сюръекцией
- отображения из больших в меньшие размерности не могут быть инъекцией

Ни один из случаев не может быть биекцией

Теорема. (??) Для линейного отображения T из $\mathbb{R}^K \to \mathbb{R}^N$, следующие утверждения верны:

1. Если $K < N$, то T не сюръекция.
2. Если $K > N$, то T не инъекция.

Доказательство. (часть 1)

Пусть $K < N$ и отображение $T \colon \mathbb{R}^K \to \mathbb{R}^N$ линейное

Пусть $V := \{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$, имеется

$$\dim(\mathrm{rng}(T)) = \dim(\mathrm{span}(V)) \leq K < N$$

$$\therefore \quad \mathrm{rng}(T) \neq \mathbb{R}^N$$

Значит, T не сюръекция

Доказательство. (часть 2)

Предположим обратное, что T является инъекцией

Пусть $\alpha_1, \ldots, \alpha_K$ набор векторов, такой что

$$\alpha_1 T\mathbf{e}_1 + \cdots + \alpha_K T\mathbf{e}_K = \mathbf{0}$$

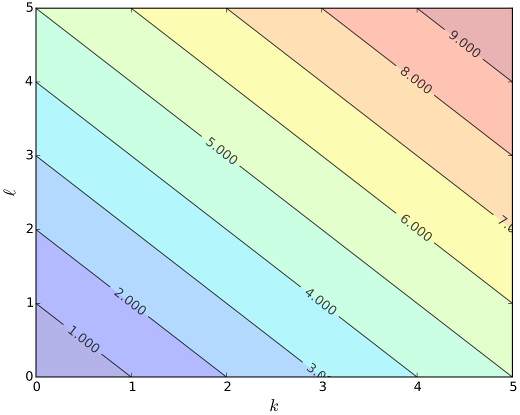
$$\therefore \quad T(\alpha_1 \mathbf{e}_1 + \cdots + \alpha_K \mathbf{e}_K) = \mathbf{0} \qquad \text{(по линейности)}$$

$$\therefore \quad \alpha_1 \mathbf{e}_1 + \cdots + \alpha_K \mathbf{e}_K = \mathbf{0} \qquad \text{(так как } \ker(T) = \{\mathbf{0}\}\text{)}$$

$$\therefore \quad \alpha_1 = \cdots = \alpha_K = 0 \qquad \text{(по независимости } \{\mathbf{e}_1, \ldots \mathbf{e}_K\}\text{)}$$

Мы показали, что $\{T\mathbf{e}_1, \ldots, T\mathbf{e}_K\}$ линейно независимы

Но тогда $\mathbb{R}^N$ содержит линейно независимое множество с $K > N$ векторами — противоречие



Пример. Функция издержек $c(k,\ell) = rk + w\ell$ не может быть инъекцией

Ортогональные векторы и проекции

Ключевой концепцией курса является ортогональность – не только векторов, но и случайных величин

Пусть $\mathbf{x}$ и $\mathbf{z}$ векторы в $\mathbb{R}^N$

Если $\langle \mathbf{x}, \mathbf{z} \rangle = 0$, то мы называем $\mathbf{x}$ и $\mathbf{z}$ **ортогональными**

Записывается $\mathbf{x} \perp \mathbf{z}$

В $\mathbb{R}^2$ ортогональный значит перпендикулярный

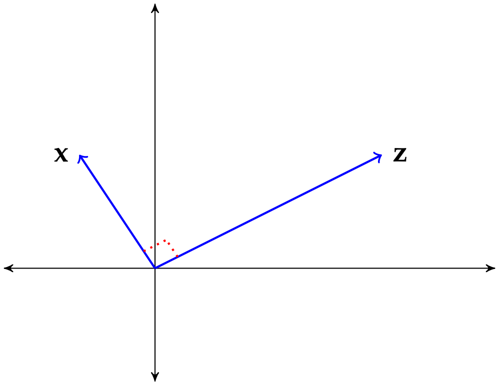


Рис.: $\mathbf{x} \perp \mathbf{z}$

Пусть S линейное подпространство

Говорят, что $\mathbf{x}$ **ортогонален** S, если $\mathbf{x} \perp \mathbf{z}$ для всех $\mathbf{z} \in S$

Записывается $\mathbf{x} \perp S$

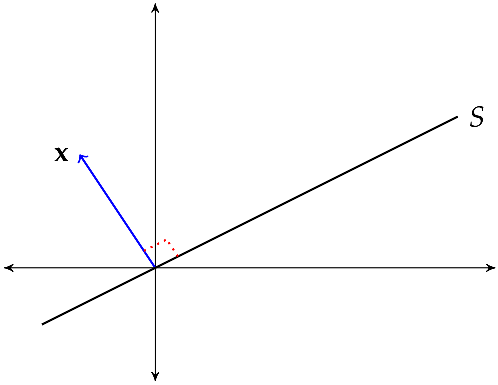


Рис.: $\mathbf{x} \perp S$

Факт. (??) (Теорема Пифагора)

Если $\{\mathbf{z}_1, \ldots, \mathbf{z}_K\}$ является ортогональным множеством, то

$$\|\mathbf{z}_1 + \cdots + \mathbf{z}_K\|^2 = \|\mathbf{z}_1\|^2 + \cdots + \|\mathbf{z}_K\|^2$$

Докажите в качестве упражнения

Факт. (??) Если $O \subset \mathbb{R}^N$ является ортогональным множеством и $\mathbf{0} \notin O$, то O линейно независимо

Ортогональной множество $O \subset \mathbb{R}^N$ называется **ортонормальное множество** Если $\|\mathbf{u}\| = 1$ для всех $\mathbf{u} \in O$

Ортонормальное множество, охватывающее линейное подпространство S в $\mathbb{R}^N$ называют **ортонормальным базисом** S

- примером ортонормального базиса для всего в $\mathbb{R}^N$ является канонический базис $\{\mathbf{e}_1, \dots, \mathbf{e}_N\}$

Факт. (??) Если $\{\mathbf{u}_1, \dots, \mathbf{u}_K\}$ является ортонормальным множеством и $\mathbf{x} \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_K\}$, то

$$\mathbf{x} = \sum_{k=1}^{K} \langle \mathbf{x}, \mathbf{u}_k \rangle \mathbf{u}_k$$

Возьмем $S \subset \mathbb{R}^N$, **ортогональное дополнение** S будет

$$S^\perp := \{\mathbf{x} \in \mathbb{R}^N : \mathbf{x} \perp S\}$$

Факт. (??) Для любого непустого $S \subset \mathbb{R}^N$, пространство $S^\perp$ является линейным подпространством $\mathbb{R}^N$

Доказательство. Если $\mathbf{x}, \mathbf{y} \in S^\perp$ и $\alpha, \beta \in \mathbb{R}$, то $\alpha\mathbf{x} + \beta\mathbf{y} \in S^\perp$, так как для любых $\mathbf{z} \in S$

$$\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle = \alpha \times 0 + \beta \times 0 = 0$$

Факт. (??) Для $S \subset \mathbb{R}^N$, выполняется $S \cap S^\perp = \{\mathbf{0}\}$

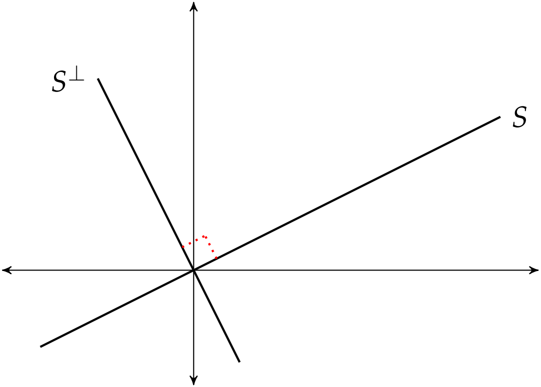


Рис.: Ортогональное дополнение S в $\mathbb{R}^2$

Теорема ортогональной проекции

Задача:

При данном $\mathbf{y} \in \mathbb{R}^N$ и подпространстве S, найти ближайший элемент S к $\mathbf{y}$

Формально: Решаем

$$\hat{\mathbf{y}} := \operatorname*{argmin}_{\mathbf{z} \in S} \|\mathbf{y} - \mathbf{z}\| \tag{2}$$

Существование, единственность решения не очевидны

Теорема ортогональной проекции: $\hat{\mathbf{y}}$ всегда существует, причем в единственном числе

Также дает полезную характеристику

Теорема. (??) [Теорема ортогональной проекции I]

Пусть $\mathbf{y} \in \mathbb{R}^N$ и S является непустым линейным подпространством в $\mathbb{R}^N$.

Следующие утверждения верны:

1. Задача оптимизации (2) имеет ровно одно решение
2. $\hat{\mathbf{y}} \in \mathbb{R}^N$ является решением (2) тогда и только тогда, когда $\hat{\mathbf{y}} \in S$ и $\mathbf{y} - \hat{\mathbf{y}} \perp S$

Единственное решение $\hat{\mathbf{y}}$ называется **ортогональной проекцией $\mathbf{y}$ на S**

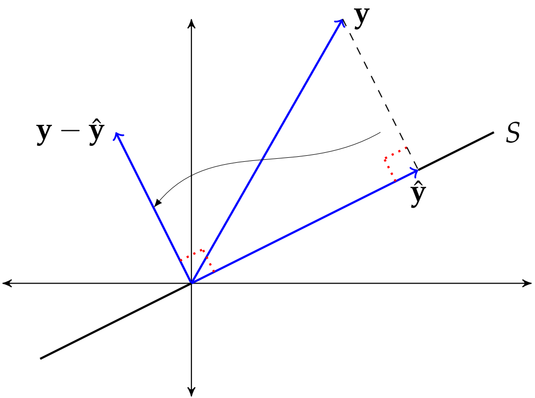


Рис.: Ортогональная проекция

Доказательство. (достаточности 2.)

Пусть $\mathbf{y} \in \mathbb{R}^N$ и S является линейным подпространством в $\mathbb{R}^N$

Пусть $\hat{\mathbf{y}}$ — вектор в S, удовлетворяющий условию $\mathbf{y} - \hat{\mathbf{y}} \perp S$

Пусть $\mathbf{z}$ — некая точка в S. Получается

$$\|\mathbf{y} - \mathbf{z}\|^2 = \|(\mathbf{y} - \hat{\mathbf{y}}) + (\hat{\mathbf{y}} - \mathbf{z})\|^2 = \|\mathbf{y} - \hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{y}} - \mathbf{z}\|^2$$

Второе равенство следует из $\mathbf{y} - \hat{\mathbf{y}} \perp S$ и Теоремы Пифагора

Так как $\mathbf{z}$ был произвольной точкой в S, получается
$\|\mathbf{y} - \mathbf{z}\| \geq \|\mathbf{y} - \hat{\mathbf{y}}\|$ для всех $\mathbf{z} \in S$

Пример. Пусть $\mathbf{y} \in \mathbb{R}^N$ и $\mathbf{1} \in \mathbb{R}^N$ является вектором из единиц

Пусть S является множеством постоянных векторов в $\mathbb{R}^N$ — S является линейной оболочкой $\{\mathbf{1}\}$

Ортогональная проекция $\mathbf{y}$ на S — это $\hat{\mathbf{y}} := \bar{y}\mathbf{1}$, где $\bar{y} := \frac{1}{N}\sum_{n=1}^N y_n$

Ясно, что $\hat{\mathbf{y}} \in S$

Чтобы показать, что $\mathbf{y} - \hat{\mathbf{y}}$ ортогонален к S, нужно проверить $\langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{1} \rangle = 0$ (смотрите упр. **??** на странице **??**). Это верно, так как

$$\langle \mathbf{y} - \hat{\mathbf{y}}, \mathbf{1} \rangle = \langle \mathbf{y}, \mathbf{1} \rangle - \langle \hat{\mathbf{y}}, \mathbf{1} \rangle = \sum_{n=1}^N y_n - \bar{y} \langle \mathbf{1}, \mathbf{1} \rangle = 0$$

Если зафиксировать подпространство S, мы получим функциональную связь

$$\mathbf{y} \;\mapsto\; \text{его ортогональная проекция } \hat{\mathbf{y}} \in S$$

Это четко определенная функция из $\mathbb{R}^N$ в $\mathbb{R}^N$

Функция обычно обозначается как $\mathbf{P}$

- $\mathbf{P}(\mathbf{y})$ или $\mathbf{P}\mathbf{y}$ представляет $\hat{\mathbf{y}}$

$\mathbf{P}$ называется **ортогональным проекционным отображением на S**, записывается как

$$\mathbf{P} = \operatorname{proj} S$$

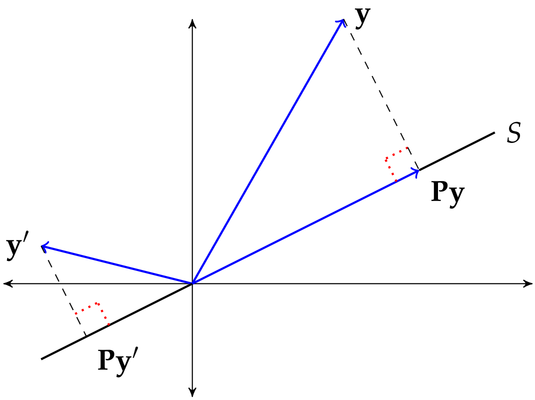


Рис.: Ортогональная проекция под $\mathbf{P}$

Теорема. (??) [Теорема ортогональной проекции II]

Пусть S является неким линейным подпространством $\mathbb{R}^N$, и $\mathbf{P} = \operatorname{proj} S$. Следующие утверждения верны:

1. $\mathbf{P}$ — линейная функция

Более того, для любого $\mathbf{y} \in \mathbb{R}^N$, соблюдается

2. $\mathbf{P}\mathbf{y} \in S$,
3. $\mathbf{y} - \mathbf{P}\mathbf{y} \perp S$,
4. $\|\mathbf{y}\|^2 = \|\mathbf{P}\mathbf{y}\|^2 + \|\mathbf{y} - \mathbf{P}\mathbf{y}\|^2$,
5. $\|\mathbf{P}\mathbf{y}\| \leq \|\mathbf{y}\|$,
6. $\mathbf{P}\mathbf{y} = \mathbf{y}$ тогда и только тогда, когда $\mathbf{y} \in S$, и
7. $\mathbf{P}\mathbf{y} = \mathbf{0}$ тогда и только тогда, когда $\mathbf{y} \in S^\perp$.

Для доказательства смотрите страницу ?? и упражнение ??

Ниже приводится основополагающий результат

Факт. (??) Если $\{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$ является ортонормальным базисом для S, то для каждого $\mathbf{y} \in \mathbb{R}^N$,

$$\mathbf{P}\mathbf{y} = \sum_{k=1}^{K} \langle \mathbf{y}, \mathbf{u}_k \rangle \, \mathbf{u}_k \tag{3}$$

Доказательство. Для начала, правая сторона (3) находится в S, так как это линейная комбинация векторов, охватывающих S

Далее, мы знаем, что $\mathbf{y} - \mathbf{P}\mathbf{y} \perp S$ тогда и только тогда, когда $\mathbf{y} - \mathbf{P}\mathbf{y} \perp \mathbf{u}_j$ для каждого $\mathbf{u}_j$ из множества базисных векторов (упражнение упр. **??**)

Дл любого $\mathbf{y} - \mathbf{P}\mathbf{y} \perp \mathbf{u}_j$, выполняется слудующее

$$\langle \mathbf{y} - \mathbf{P}\mathbf{y}, \mathbf{u}_j \rangle = \langle \mathbf{y}, \mathbf{u}_j \rangle - \sum_{k=1}^{K} \langle \mathbf{y}, \mathbf{u}_k \rangle \langle \mathbf{u}_k, \mathbf{u}_j \rangle$$

$$= \langle \mathbf{y}, \mathbf{u}_j \rangle - \langle \mathbf{y}, \mathbf{u}_j \rangle = 0$$

Это подтверждает, что $\mathbf{y} - \mathbf{P}\mathbf{y} \perp S$

Факт. (??) Пусть S_i является линейным подпространством $\mathbb{R}^N$ для $i = 1, 2$ и $\mathbf{P}_i = \operatorname{proj} S_i$. Если $S_1 \subset S_2$, то

$$\mathbf{P}_1\mathbf{P}_2\mathbf{y} = \mathbf{P}_2\mathbf{P}_1\mathbf{y} = \mathbf{P}_1\mathbf{y} \quad \text{для всех } \mathbf{y} \in \mathbb{R}^N$$

Остаточная проекция

Спроецируем $\mathbf{y}$ на S, где S является линейным подпространством $\mathbb{R}^N$

- Ближайшая точка к $\mathbf{y}$ на S — это $\hat{\mathbf{y}} := \mathbf{P}\mathbf{y}$, здесь $\mathbf{P} = \operatorname{proj} S$
- Если $\mathbf{y}$ не находится в S, ошибка $\mathbf{y} - \mathbf{P}\mathbf{y}$ существует

Введем оператор $\mathbf{M}$, который берет $\mathbf{y} \in \mathbb{R}^N$ и возвращает остаток

$$\mathbf{M} := \mathbf{I} - \mathbf{P} \tag{4}$$

где $\mathbf{I}$ является тождественным отображением $\mathbb{R}^N$

Для любых $\mathbf{y}$ выполняется $\mathbf{M}\mathbf{y} = \mathbf{I}\mathbf{y} - \mathbf{P}\mathbf{y} = \mathbf{y} - \mathbf{P}\mathbf{y}$

В регрессионном анализе $\mathbf{M}$ проявляется как матрица, называемая 'аннигилятором'

Мы говорим о $\mathbf{M}$ как об остаточной проекции

Пример. Вспомним, что проекция $\mathbf{y} \in \mathbb{R}^N$ на $\mathrm{span}\{\mathbf{1}\}$ — это $\bar{y}\mathbf{1}$

Остаточная проекция $\mathbf{M}_c\mathbf{y} := \mathbf{y} - \bar{y}\mathbf{1}$

- вектор ошибок, полученный, когда элементы вектора предсказываются его выборочным средним

Факт. (??) Пусть S является подпространством $\mathbb{R}^N$, $\mathbf{P} = \operatorname{proj} S$, и $\mathbf{M}$ является остаточной проекцией, определенной в (4). Верны следующие утверждения:

1. $\mathbf{M} = \operatorname{proj} S^{\perp}$
2. $\mathbf{y} = \mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y}$ для любых $\mathbf{y} \in \mathbb{R}^N$
3. $\mathbf{P}\mathbf{y} \perp \mathbf{M}\mathbf{y}$ для любых $\mathbf{y} \in \mathbb{R}^N$
4. $\mathbf{M}\mathbf{y} = \mathbf{0}$ тогда и только тогда, когда $\mathbf{y} \in S$
5. $\mathbf{P} \circ \mathbf{M} = \mathbf{M} \circ \mathbf{P} = \mathbf{0}$

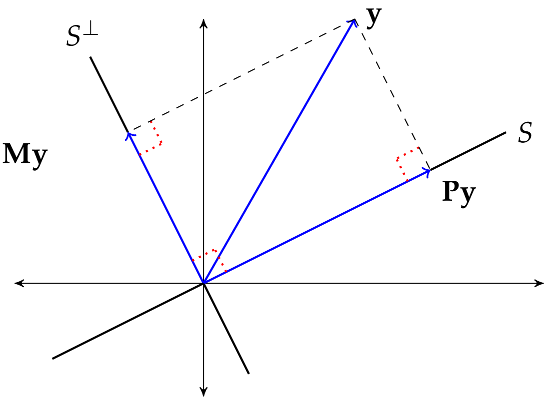


Рис.: Остаточная проекция

Если S_1 и S_2 — два подпространства $\mathbb{R}^N$ и $S_1 \subset S_2$, то $S_2^\perp \subset S_1^\perp$

Результат факта **??** обратный для $\mathbf{M}$

Факт. (**??**) Пусть S_1 и S_2 — два подпространства $\mathbb{R}^N$ и $\mathbf{y} \in \mathbb{R}^N$. Пусть $\mathbf{M}_1$ и $\mathbf{M}_2$ являются проекциями на $S_1^\perp$ и $S_2^\perp$ соответственно. Если $S_1 \subset S_2$, то

$$\mathbf{M}_1\mathbf{M}_2\mathbf{y} = \mathbf{M}_2\mathbf{M}_1\mathbf{y} = \mathbf{M}_2\mathbf{y}$$

Ортогонализация Грама-Шмидта

Вспомним, что любое ортогональное множество $\mathbb{R}^N$ не содержащее $\mathbf{0}$ линейно независимо – факт **??**

Вот (важное) частично обратное этому утверждению

Теорема. (**??**) Для каждого линейно независимого множества $\{\mathbf{b}_1, \ldots, \mathbf{b}_K\} \subset \mathbb{R}^N$, существует ортонормальное множество $\{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$, такое что

$$\operatorname{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\} = \operatorname{span}\{\mathbf{u}_1, \ldots, \mathbf{u}_k\} \quad \text{для} \quad k = 1, \ldots, K$$

Формальные доказательства решаются как упражнения **??** to **??**

Доказательство дает важный алгоритм построения ортонормированного множества $\{\mathbf{u}_1, \ldots, \mathbf{u}_K\}$

Первый шаг — построить ортогональные множества $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ с линейной оболочкой, идентичной $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ для каджого k

Построение $\{\mathbf{v}_1, \ldots, \mathbf{v}_K\}$ использует **ортогонализацию Грама-Шмидта**:

Для каждого $k = 1, \ldots, K$, пусть

1. $B_k := \operatorname{span}\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$,
2. $\mathbf{P}_k := \operatorname{proj} B_k$ и $\mathbf{M}_k := \operatorname{proj} B_k^\perp$,
3. $\mathbf{v}_k := \mathbf{M}_{k-1}\mathbf{b}_k$, где $\mathbf{M}_0$ является тождественным отображением, и
4. $V_k := \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$.

В шаге 3. мы сопоставляем каждый последующий элемент $\mathbf{b}_k$ в подпространство, ортогональное подпространству, созданному с помощью $\mathbf{b}_1, \ldots, \mathbf{b}_{k-1}$

В завершение, определим $\mathbf{u}_k$ с помощью $\mathbf{u}_k := \mathbf{v}_k / \|\mathbf{v}_k\|$

Множество векторов $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ является ортонормальным с линейной оболочкой, равной V_k

Учебник по Эконометрике

Лекция 2: Линейная Алгебра и Матрицы

Джон Стачурски

Лекции: Акшай Шенкер

Перевел: Алексей Кедо

1 октября 2020 г.

Матрицы

Обычная $N \times K$ **матрица**:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix}$$

Символ a_{nk} означает элемент, стоящий в n-ой строке k-ого столбца

$N \times K$ матрица также называется

- **вектором строки**, если $N = 1$
- **вектором столбца**, если $K = 1$

Если $N = K$, то матрицу $\mathbf{A}$ называют **квадратной**

Если $\mathbf{A}$ квадратная и $a_{nk} = a_{kn}$ для любых k и n, то $\mathbf{A}$ называют **симметричной**

Часто элементы матрицы $\mathbf{A}$ представляют собой коэффициенты в системе линейных уравнений

$$\begin{gathered} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1K}x_K = b_1 \\ \vdots \\ a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{NK}x_K = b_N \end{gathered}$$

Для матрицы $\mathbf{A}$ применяются обозначения:

- $\mathrm{row}_n\,\mathbf{A}$ означает n-ую строку $\mathbf{A}$
- $\mathrm{col}_k\,\mathbf{A}$ означает k-ый столбец $\mathbf{A}$

Символы $\mathbf{0}$ и $\mathbf{1}$ представляют собой матрицы, все элементы которых равны нулю и единице соответственно

Для квадратной матрицы $\mathbf{A}$, элементы a_{nn} называют **главной диагональю**:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix}$$

Единичная матрица:

$$\mathbf{I} := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Заметим, что $\operatorname{col}_n \mathbf{I} = \mathbf{e}_n$ — n-ый канонический базис $\mathbb{R}^N$

Алгебраические операции над матрицами

Операции над матрицами:

- Умножение на скаляр
- Сложение
- Умножение матриц

Умножение на скаляр выполняется поэлементно, как с векторами:

$$\gamma \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} := \begin{pmatrix} \gamma a_{11} & \gamma a_{12} & \cdots & \gamma a_{1K} \\ \gamma a_{21} & \gamma a_{22} & \cdots & \gamma a_{2K} \\ \vdots & \vdots & & \vdots \\ \gamma a_{N1} & \gamma a_{N2} & \cdots & \gamma a_{NK} \end{pmatrix}$$

Сложение тоже выполняется поэлементно:

$$\begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1K} \\ b_{21} & \cdots & b_{2K} \\ \vdots & \vdots & \vdots \\ b_{N1} & \cdots & b_{NK} \end{pmatrix}$$

$$:= \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1K}+b_{1K} \\ a_{21}+b_{21} & \cdots & a_{2K}+b_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1}+b_{N1} & \cdots & a_{NK}+b_{NK} \end{pmatrix}$$

Заметим, что матрицы должны быть одинакового размера

Умножение матриц:

Произведение $\mathbf{AB}$: i, j-ый элемент — скалярное умножение i-ой строки $\mathbf{A}$ и j-ого столбца $\mathbf{B}$

$$c_{ij} = \langle \mathrm{row}_i \, \mathbf{A}, \, \mathrm{col}_j \, \mathbf{B} \rangle = \sum_{k=1}^{K} a_{ik} b_{kj}$$

Для $i = j = 1$:

$$\begin{pmatrix} a_{11} & \cdots & a_{1K} \\ a_{21} & \cdots & a_{2K} \\ \vdots & \vdots & \vdots \\ a_{N1} & \cdots & a_{NK} \end{pmatrix} \begin{pmatrix} b_{11} & \cdots & b_{1J} \\ b_{21} & \cdots & b_{2J} \\ \vdots & \vdots & \vdots \\ b_{K1} & \cdots & b_{KJ} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1J} \\ c_{21} & \cdots & c_{2J} \\ \vdots & \vdots & \vdots \\ c_{N1} & \cdots & c_{NJ} \end{pmatrix}$$

Здесь

$$c_{11} = \langle \mathrm{row}_1(\mathbf{A}), \mathrm{col}_1(\mathbf{B}) \rangle = \sum_{k=1}^{K} a_{1k} b_{k1}$$

Предположим, что размер $\mathbf{A}$ — $N \times K$, $\mathbf{B}$ — $J \times M$

- $\mathbf{AB}$ определена, только если $K = J$
- Размер $\mathbf{AB}$ — $N \times M$

Запомните правило:

Произведение $N \times K$ и $K \times M$ равно $N \times M$

Умножение не коммутативно: $\mathbf{AB} \neq \mathbf{BA}$

Заметим, что произведение $\mathbf{BA}$ определено, только если $N = M$ соблюдается

Факт. (??) Для согласованных матриц $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ и скаляра α, верно

1. $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$,
2. $\mathbf{A}(\mathbf{B}+\mathbf{C}) = \mathbf{AB}+\mathbf{AC}$,
3. $(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$,
4. $\mathbf{A}\alpha\mathbf{B} = \alpha\mathbf{AB}$, и
5. $\mathbf{AI} = \mathbf{A}$ и $\mathbf{IA} = \mathbf{A}$, где $\mathbf{I}$ — единичная матрица.

Здесь "согласованость" значит, что операция определена при заданных размерностях матриц *k*-**ая мощность** квадратной матрицы $\mathbf{A}$ определяется как

$$\mathbf{A}^k := \underbrace{\mathbf{A} \cdots \mathbf{A}}_{k \text{ terms}}$$

Если матрица $\mathbf{B}$ такая, что $\mathbf{B}^2 = \mathbf{A}$, то $\mathbf{B}$ называется **квадратным корнем** матрицы $\mathbf{A}$ и записывается как $\sqrt{\mathbf{A}}$

Возьмем матрицу $\mathbf{A}$ размерности $N \times K$ и $K \times 1$ вектор столбца $\mathbf{x}$, перемножение $\mathbf{Ax}$:

$$\mathbf{Ax} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NK} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{pmatrix}$$

$$= x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{N1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{N2} \end{pmatrix} + \cdots + x_K \begin{pmatrix} a_{1K} \\ a_{2K} \\ \vdots \\ a_{NK} \end{pmatrix}$$

$$= \sum_{k=1}^{K} x_k \operatorname{col}_k \mathbf{A}$$

Матрицы как отображения

Можно размышлять о матрице $\mathbf{A}$ размерна $N \times K$ как об отображении из $\mathbb{R}^K$ в $\mathbb{R}^N$:

$$\mathbf{x} \mapsto \mathbf{A}\mathbf{x}$$

Такое отображение линейно

Как насчет примеров линейных функций, не использующих матрицы?

...на самом деле, таких не существует!

Множеству линейных функций из $\mathbb{R}^K$ в $\mathbb{R}^N$ инъективно соответствует множество матриц $N \times K$:

Теорема. (??) Пусть T является функцией из $\mathbb{R}^K$ в $\mathbb{R}^N$. Следующее эквивалентно:

1. T линейна.
2. Существует матрица $\mathbf{A}$ размера $N \times K$, такая что $T\mathbf{x} = \mathbf{A}\mathbf{x}$ для всех $\mathbf{x} \in \mathbb{R}^K$.

Доказательство. (1. $\implies$ 2.)

Пусть $T\colon \mathbb{R}^K \to \mathbb{R}^N$ линейна

Мы собираемся построить матрицу $\mathbf{A}$ размерности $N \times K$, такую что

$$T\mathbf{x} = \mathbf{A}\mathbf{x}, \qquad \forall\, \mathbf{x} \in \mathbb{R}^K$$

Как обычно, пусть $\mathbf{e}_k$ — k-ый канонический базисный вектор в $\mathbb{R}^K$

Определим матрицу $\mathbf{A}$ как $\operatorname{col}_k(\mathbf{A}) = T\mathbf{e}_k$. Возьмем любой $\mathbf{x} = (x_1, \ldots, x_K) \in \mathbb{R}^K$. По линейности, получается

$$T\mathbf{x} = T\left[\sum_{k=1}^{K} x_k \mathbf{e}_k\right] = \sum_{k=1}^{K} x_k T\mathbf{e}_k = \sum_{k=1}^{K} x_k \operatorname{col}_k(\mathbf{A}) = \mathbf{A}\mathbf{x}$$

Доказательство. (2 $\implies$ 1) Возьмем матрицу $\mathbf{A}$ размера $N \times K$ и пусть T определяется

$$T\colon \mathbb{R}^K \to \mathbb{R}^N, \qquad T\mathbf{x} = \mathbf{A}\mathbf{x}$$

Возьмем любые $\mathbf{x}$, $\mathbf{y}$ в $\mathbb{R}^K$, и любые скаляры α и β

Правила матричной арифметики говорят нам, что

$$T(\alpha\mathbf{x} + \beta\mathbf{y}) := \mathbf{A}(\alpha\mathbf{x} + \beta\mathbf{y}) = \alpha\mathbf{A}\mathbf{x} + \beta\mathbf{A}\mathbf{y} =: \alpha T\mathbf{x} + \beta T\mathbf{y}$$

Рассмотрим возможность решения системы линейных уравнений, например, $\mathbf{Ax} = \mathbf{b}$

Существование: можем ли мы найти $\mathbf{x}$, удовлетворяющий уравнению при любых заданных $\mathbf{b}$?

- является ли линейное отображение $T\mathbf{x} = \mathbf{Ax}$ сюръекцией?
- эквивалентно, равен ли $\operatorname{rng} T$ всему $\mathbb{R}^N$?

Пространство столбцов

Диапазон T — все вектора вида $T\mathbf{x} = \mathbf{A}\mathbf{x}$, где $\mathbf{x}$ варьируется в $\mathbb{R}^K$

Для $\mathbf{x} \in \mathbb{R}^K$ получается

$$\mathbf{A}\mathbf{x} = \sum_{k=1}^{K} x_k \operatorname{col}_k \mathbf{A}$$

Таким образом, $\operatorname{rng} T$ равен **пространству столбцов** $\mathbf{A}$ — линейная оболочка столбцов $\mathbf{A}$

$$\operatorname{colspace} \mathbf{A} := \operatorname{span}\{\operatorname{col}_1 \mathbf{A}, \ldots, \operatorname{col}_K \mathbf{A}\}$$

В итоге,

$$\operatorname{colspace} \mathbf{A} = \operatorname{rng} T = \{\mathbf{A}\mathbf{x} : \mathbf{x} \in \mathbb{R}^K\}$$

Ранг

Эквивалентные вопросы

- Насколько велик диапазон линейного отображения $T\mathbf{x} = \mathbf{A}\mathbf{x}$?
- Насколько велико пространство столбцов $\mathbf{A}$?

Очевидной мерой размера линейного подпространства является его размерность

Размерность $\operatorname{colspace} \mathbf{A}$ известна как **ранг** $\mathbf{A}$

$$\operatorname{rank} \mathbf{A} := \dim \operatorname{colspace} \mathbf{A}$$

Так как $\operatorname{colspace} \mathbf{A}$ — линейная оболочка K векторов, получается

$$\operatorname{rank} \mathbf{A} = \dim \operatorname{colspace} \mathbf{A} \leq K$$

$\mathbf{A}$ имеет **полный ранг системы столбцов**, если

$$\operatorname{rank} \mathbf{A} = \text{ количество столбцов } \mathbf{A}$$

Факт. (??) Для любой матрицы $\mathbf{A}$, следующие утверждения эквивалентны:

1. $\mathbf{A}$ является полным рангом системы столбцов
2. Столбцы $\mathbf{A}$ линейно независимы
3. Если $\mathbf{Ax} = \mathbf{0}$, то $\mathbf{x} = \mathbf{0}$

Квадратные матрицы и обратимость

Рассмотрим случай с матрицей $\mathbf{A}$ размера $N \times N$

Мы ищем условия, при которых для каждого $\mathbf{b} \in \mathbb{R}^N$, существует только один $\mathbf{x} \in \mathbb{R}^N$, такой что $\mathbf{A}\mathbf{x} = \mathbf{b}$

Пусть T является линейным отображением $T\mathbf{x} = \mathbf{A}\mathbf{x}$

- Когда каждая точка $\mathbf{b} \in \mathbb{R}^N$ имеет только один прообраз в T?
- Эквивалентно, когда T — биекция?

Напомним, линейные биекции называются несингулярными функциями.

Факт. (??) Для матрицы $\mathbf{A}$ размера $N \times N$ следующее эквивалентно:

1. Столбцы $\mathbf{A}$ линейно независимы.
2. Столбцы $\mathbf{A}$ формируют базис $\mathbb{R}^N$.
3. $\operatorname{rank} \mathbf{A} = N$.
4. $\operatorname{colspace} \mathbf{A} = \mathbb{R}^N$.
5. $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{y} \implies \mathbf{x} = \mathbf{y}$.
6. $\mathbf{A}\mathbf{x} = \mathbf{0} \implies \mathbf{x} = \mathbf{0}$.
7. Для каждого $\mathbf{b} \in \mathbb{R}^N$, уравнение $\mathbf{A}\mathbf{x} = \mathbf{b}$ имеется решение.
8. Для каждого $\mathbf{b} \in \mathbb{R}^N$, уравнение $\mathbf{A}\mathbf{x} = \mathbf{b}$ имеет единственное решение.

Если любое из эквивалентных условий факта **??** верное, мы назовем **несингулярной** не только отображение T, но и основную матрицу $\mathbf{A}$

Если хоть одно — и, следовательно, все — из этих условий не выполняются, то $\mathbf{A}$ называется **сингулярной**

Любая биекция имеет обратную функцию (смотрите §?? в ET)

Любое несингулярное отображение T имеет несингулярное обратное T^{-1} (факт **??** на странице **??**)

- если T создано матрицей $\mathbf{A}$, обратное T^{-1} также связано с матрицей, называемой обратной $\mathbf{A}$

Теорема. (??) Для несингулярной $\mathbf{A}$ следующие утверждения верны:

1. Существует квадратная матрица $\mathbf{B}$, такая что $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, где $\mathbf{I}$ — единичная матрица. Матрица $\mathbf{B}$ называется **обратной** $\mathbf{A}$, и записывается как $\mathbf{A}^{-1}$.
2. Для каждого $\mathbf{b} \in \mathbb{R}^N$, единственное решение задачи $\mathbf{Ax} = \mathbf{b}$ записывается как

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Пример. Расстмотрим линейный спрос для N товаров

$$q_n = \sum_{k=1}^{N} a_{nk} p_k + b_n, \quad n = 1, \ldots, N$$

где q_n и p_n — количество и цена n-ого товара

Мы хотим вычислить обратную функцию спроса, которая показывает зависимость цены от количества

Записшем систему в матричной форме: $\mathbf{q} = \mathbf{A}\mathbf{p} + \mathbf{b}$

Если столбцы $\mathbf{A}$ линейно независимы, то система обратима — единственное решение существует для любых фиксированных $\mathbf{q}$ и $\mathbf{b}$:

$$\mathbf{p} = \mathbf{A}^{-1}(\mathbf{q} - \mathbf{b})$$

Для матриц $\mathbf{A}$ и $\mathbf{B}$ размера $N \times N$, если

- $\mathbf{B}$ является **левой обратной**, в том смысле, что $\mathbf{BA} = \mathbf{I}$
- или $\mathbf{B}$ является **правой обратной**, в том смысле, что $\mathbf{AB} = \mathbf{I}$

Тогда $\mathbf{A}$ обратима и $\mathbf{B}$ является обратной матрицей $\mathbf{A}$

Факт. (??) Пусть $\mathbf{A}$ и $\mathbf{B}$ — квадратные матрицы размера $N \times N$. Если $\mathbf{B}$ является левой или правой обратной матрицы $\mathbf{A}$, то $\mathbf{A}$ несингулярна и $\mathbf{B}$ — ее обратная матрица.

Факт. (??) Если $\mathbf{A}$ и $\mathbf{B}$ несингулярны и $\alpha \neq 0$, то

1. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$,
2. $(\alpha\mathbf{A})^{-1} = \alpha^{-1}\mathbf{A}^{-1}$, и
3. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

Определитель

Определитель — уникальное число для любой квадратной матрицы $\mathbf{A}$

Пусть $S(N)$ является множеством всех биекций из $\{1, \dots, N\}$ в саму себя

Для $\pi \in S(N)$ определим **функцию знака** π

$$\operatorname{sgn}(\pi) := \prod_{m<n} \frac{\pi(m) - \pi(n)}{m - n}$$

Определитель матрицы $\mathbf{A}$ размера $N \times N$ — это

$$\det \mathbf{A} := \sum_{\pi \in S(N)} \operatorname{sgn}(\pi) \prod_{n=1}^{N} a_{\pi(n)n}$$

Факт. (??) Если $\mathbf{I}$ — единичная матрица размера $N \times N$, $\mathbf{A}$ и $\mathbf{B}$ являются марицами размера $N \times N$ и $\alpha \in \mathbb{R}$, то

1. $\det \mathbf{I} = 1$,
2. $\mathbf{A}$ несингулярна тогда и только тогда, когда $\det \mathbf{A} \neq 0$,
3. $\det(\mathbf{AB}) = \det(\mathbf{A}) \det(\mathbf{B})$,
4. $\det(\alpha \mathbf{A}) = \alpha^N \det(\mathbf{A})$, и
5. $\det(\mathbf{A}^{-1}) = (\det(\mathbf{A}))^{-1}$.

В случае матрицы размера 2×2 можно показать, что

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

Доказательство. (Факт ??)

Возьмем кважратную матрицу $\mathbf{A}$ и предположим, что правая обратная матрица $\mathbf{B}$ существует:

$$\mathbf{AB} = \mathbf{I}$$

Тогда обе матрицы $\mathbf{A}$ и $\mathbf{B}$ несингулярны, так как по правилам факта ??:

$$\det(\mathbf{A})\det(\mathbf{B}) = 1$$

Оба определителя $\det \mathbf{A}$ и $\det \mathbf{B}$ ненулевые и, следовательно, обе матрицы несингулярные. Далее, $\mathbf{AB} = \mathbf{I}$, значит

$$\mathbf{BAB} = \mathbf{B}$$

Умножим на $\mathbf{B}^{-1}$, получается $\mathbf{BA} = \mathbf{I}$

Диагональные и треугольные матрицы

Квадратная матрица назвается **нижней треугольной**, если каждый элемент строго выше главной диагонали равен нулю

Пример.

$$\mathbf{L} := \begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{pmatrix}$$

Квадратная матрица назвается **верхней треугольной**, если каждый элемент строго ниже главной диагонали равен нулю

Пример.

$$\mathbf{U} := \begin{pmatrix} 1 & 2 & 3 \\ 0 & 5 & 6 \\ 0 & 0 & 1 \end{pmatrix}$$

Квадратная матрица назвается **треугольной**, если она верхняя или нижняя треугольная

Факт. (??) Если $\mathbf{A} = (a_{mn})$ треугольная, то $\det \mathbf{A} = \prod_{n=1}^{N} a_{nn}$.

Связанные линейные уравнения просто решить

Пример.

$$\begin{pmatrix} 1 & 0 & 0 \\ 2 & 5 & 0 \\ 3 & 6 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

становится

$$\begin{aligned} x_1 &= b_1 \\ 2x_1 + 5x_2 &= b_2 \\ 3x_1 + 6x_2 + x_3 &= b_3 \end{aligned}$$

Верхнее уравнение включает только x_1, так что его можно решить напрямую

Подставьте это значение во второе уравнение и решите для x_2 и т.д.

Рассмотрим квадратную матрицу $\mathbf{A}$ размера $N \times N$

N элементов a_{nn} называются **главной диагональю**

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix}$$

Квадратная матрица $\mathbf{D}$ называется **диагональной**, если все значения вне главной диагонали равны нулю

$$\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_N \end{pmatrix}$$

Часто записывается как

$$\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_N)$$

Диагональные системы очень легко решить

Пример.

$$\begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$$

эквивалентно

$$\begin{aligned} d_1 x_1 &= b_1 \\ d_2 x_2 &= b_2 \\ d_3 x_3 &= b_3 \end{aligned}$$

Факт. (??) Если $\mathbf{C} = \operatorname{diag}(c_1, \ldots, c_N)$ и $\mathbf{D} = \operatorname{diag}(d_1, \ldots, d_N)$, то

1. $\mathbf{C} + \mathbf{D} = \operatorname{diag}(c_1 + d_1, \ldots, c_N + d_N)$
2. $\mathbf{C}\mathbf{D} = \operatorname{diag}(c_1 d_1, \ldots, c_N d_N)$
3. $\mathbf{D}^k = \operatorname{diag}(d_1^k, \ldots, d_N^k)$ для любых $k \in \mathbb{N}$
4. $d_n \geq 0$ для всех $n \implies \mathbf{D}^{1/2}$ существует и равняется
$$\operatorname{diag}(\sqrt{d_1}, \ldots, \sqrt{d_N})$$
5. $d_n \neq 0$ для всех $n \implies \mathbf{D}$ несингулярна и
$$\mathbf{D}^{-1} = \operatorname{diag}(d_1^{-1}, \ldots, d_N^{-1})$$

Доказательства: Проверьте 1 и 2 напрямую. Части 3-5 следуют из 1 и 2.

След, транспонирование, симметрия

След матрицы определяется как

$$\operatorname{trace}\begin{pmatrix} a_{11} & \cdots & a_{1N} \\ \vdots & & \vdots \\ a_{N1} & \cdots & a_{NN} \end{pmatrix} = \sum_{n=1}^{N} a_{nn}$$

Факт. (??) Если $\mathbf{A}$ и $\mathbf{B}$ являются квадратными матрицами и $\alpha, \beta \in \mathbb{R}$, то

$$\operatorname{trace}(\alpha\mathbf{A} + \beta\mathbf{B}) = \alpha \operatorname{trace}(\mathbf{A}) + \beta \operatorname{trace}(\mathbf{B})$$

Если $\mathbf{A}$ размера $N \times M$ и $\mathbf{B}$ размера $M \times N$, то $\operatorname{trace}(\mathbf{AB}) = \operatorname{trace}(\mathbf{BA})$

Транспонирование превращает матрицу $\mathbf{A}$ размера $N \times K$ в матрицу $\mathbf{A}^\mathsf{T}$ размера $K \times N$, определяемую как

$$\mathrm{col}_n(\mathbf{A}') = \mathrm{row}_n(\mathbf{A})$$

Пример. Если

$$\mathbf{A} := \begin{pmatrix} 10 & 40 \\ 20 & 50 \\ 30 & 60 \end{pmatrix} \quad \text{, то} \quad \mathbf{A}' = \begin{pmatrix} 10 & 20 & 30 \\ 40 & 50 & 60 \end{pmatrix}$$

Если

$$\mathbf{B} := \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \quad \text{, то} \quad \mathbf{B}' := \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}$$

Квадратная матрица $\mathbf{A}$ называется **симметричной**, если $\mathbf{A}^\top = \mathbf{A}$

- эквивалентно, $a_{nk} = a_{kn}$ для всех k и n

Матрицы $\mathbf{A}^\top\mathbf{A}$ и $\mathbf{A}\mathbf{A}^\top$ всегда корректно определены и симметричны

Факт. (??) Для согласующихся матриц $\mathbf{A}$ и $\mathbf{B}$ транспонирование удовлетворяет

1. $(\mathbf{A}^\top)^\top = \mathbf{A}$,
2. $(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top\mathbf{A}^\top$,
3. $(\mathbf{A}+\mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$, и
4. $(c\mathbf{A})^\top = c\mathbf{A}^\top$ для любых постоянных c.

Факт. (??) Для любой квадратной матрицы $\mathbf{A}$, верно

1. $\mathrm{trace}(\mathbf{A}) = \mathrm{trace}(\mathbf{A}^\intercal)$ и
2. $\det(\mathbf{A}^\intercal) = \det(\mathbf{A})$.
3. Если $\mathbf{A}$ несингулярна, то $\mathbf{A}^\intercal$ тоже несингулярна, и $(\mathbf{A}^\intercal)^{-1} = (\mathbf{A}^{-1})^\intercal$.

Если $\mathbf{a}$ и $\mathbf{b}$ — векторы размера $N \times 1$, умножение матриц $\mathbf{a}^\intercal \mathbf{b} = \mathbf{b}^\intercal \mathbf{a}$ равняется $\sum_{n=1}^{N} a_n b_n$,

- так же, как скалярное произведение $\langle \mathbf{a}, \mathbf{b} \rangle$

Собственные значения и собственные векторы

Возьмем матрицу $\mathbf{A}$ размера $N \times N$

Обычно $\mathbf{A}$ отображает $\mathbf{x}$ в какое-то произвольное новое место $\mathbf{Ax}$

Но иногда $\mathbf{x}$ будет только масштабироваться:

$$\mathbf{Ax} = \lambda \mathbf{x} \quad \text{для некоторого скаляра } \lambda \tag{1}$$

Если (1) выполняется и $\mathbf{x}$ is ненулевой, то

1. $\mathbf{x}$ называется **собственным вектором** $\mathbf{A}$ и λ называется **собственным значением**
2. $(\mathbf{x}, \lambda)$ называется **собственной парой**

Ясно, что $(\mathbf{x}, \lambda)$ — собственная пара $\mathbf{A} \implies (\alpha\mathbf{x}, \lambda)$ — собственная пара $\mathbf{A}$ для любых ненулевых α

Пример. Пусть

$$\mathbf{A} := \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix}$$

Тогда

$$\lambda = 2 \quad \text{и} \quad \mathbf{x} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

формируют собственную пару, потому что $\mathbf{x} \neq \mathbf{0}$ и

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} 1 & -1 \\ 3 & 5 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \lambda \mathbf{x}$$

Пример. Рассмотрим матрицу

$$\mathbf{R} := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Вызывает вращение против часовой стрелки в любой точке на 90°

Значит ни одна точка $\mathbf{x}$ не масштабируется

Значит не существует пары $\lambda \in \mathbb{R}$ и $\mathbf{x} \neq \mathbf{0}$, такой что

$$\mathbf{R}\mathbf{x} = \lambda\mathbf{x}$$

- Другими словами, не существует вещественных собственных пар

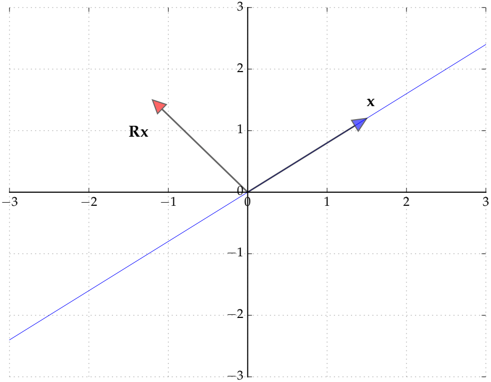


Рис.: Матрица $\mathbf{R}$ поворачивает точки на 90°

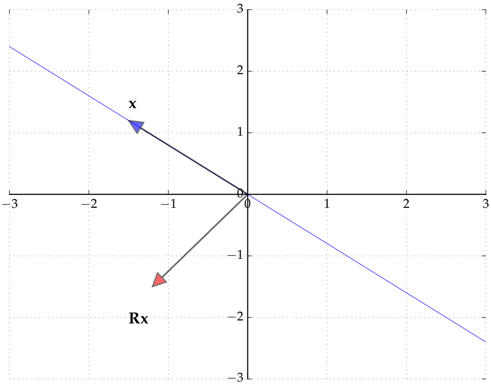


Рис.: Матрица $\mathbf{R}$ поворачивает точки на 90°

Но $\mathbf{Rx} = \lambda \mathbf{x}$ может выполняться, <u>если</u> мы допускаем комплексные числа

Пример.

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} i \\ 1 \end{pmatrix} = i \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

То есть

$$\mathbf{Rx} = \lambda \mathbf{x} \quad \text{для} \quad \lambda := i \quad \text{и} \quad \mathbf{x} := \begin{pmatrix} 1 \\ -i \end{pmatrix}$$

Тогда $(\mathbf{x}, \lambda)$ является собственной парой при условии, что мы допускаем комплексные числа

Факт. (??) для любой квадратной матрицы $\mathbf{A}$

$$\lambda \text{ является собственным значением } \mathbf{A} \iff \det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

Доказательство. Пусть $\mathbf{A}$ — матрица размера $N \times N$ и $\mathbf{I}$ — единичная матрица размера $N \times N$

Получается

$$\begin{aligned}
\det(\mathbf{A} - \lambda \mathbf{I}) = 0 &\iff \mathbf{A} - \lambda \mathbf{I} \text{ сингулярно} \\
&\iff \exists \mathbf{x} \neq \mathbf{0} \text{ s.t. } (\mathbf{A} - \lambda \mathbf{I})\mathbf{x} = \mathbf{0} \\
&\iff \exists \mathbf{x} \neq \mathbf{0} \text{ s.t. } \mathbf{A}\mathbf{x} = \lambda \mathbf{x} \\
&\iff \lambda \text{ является собственным значением } \mathbf{A}
\end{aligned}$$

Пример. В случае матрицы 2×2,

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Longrightarrow \quad \mathbf{A} - \lambda \mathbf{I} = \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix}$$

$$\begin{aligned} \therefore \quad \det(\mathbf{A} - \lambda \mathbf{I}) &= (a-\lambda)(d-\lambda) - bc \\ &= \lambda^2 - (a+d)\lambda + (ad-bc) \end{aligned}$$

Значит собственные значения $\mathbf{A}$ являются двумя корнями уравнения

$$\lambda^2 - (a+d)\lambda + (ad-bc) = 0$$

Эквивалентно,

$$\lambda^2 - \mathrm{trace}(\mathbf{A})\lambda + \det(\mathbf{A}) = 0$$

Существование собственных значений

Возьмем матрицу $\mathbf{A}$ размера $N \times N$

Факт. Существуют комплексные числа $\lambda_1, \ldots, \lambda_N$, такие что

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \prod_{n=1}^{N} (\lambda_n - \lambda)$$

Каждое такое λ_i является собственным значением $\mathbf{A}$, потому что

$$\det(\mathbf{A} - \lambda_i \mathbf{I}) = \prod_{n=1}^{N} (\lambda_n - \lambda_i) = 0$$

Важно: не все собственные значения обязательно различны — могут быть повторы

Факт. (??) Возьмем матрицу $\mathbf{A}$ размера $N \times N$ с собственными значениями $\lambda_1, \dots, \lambda_N$. Получается

1. $\det(\mathbf{A}) = \prod_{n=1}^{N} \lambda_n$
2. $\mathrm{trace}(\mathbf{A}) = \sum_{n=1}^{N} \lambda_n$
3. если $\mathbf{A}$ симметрична, то $\lambda_n \in \mathbb{R}$ для всех n
4. если $\mathbf{A}$ несингулярна, то eigenvalues of $\mathbf{A}^{-1} = 1/\lambda_1, \dots, 1/\lambda_N$
5. если $\mathbf{A} = \mathrm{diag}(d_1, \dots, d_N)$, то $\lambda_n = d_n$ для всех n

Значит матрица $\mathbf{A}$ несингулярна $\iff$ все ее собственные значения ненулевые

Квадратичные формы

Возьмем матрицу $\mathbf{A}$ размера $N \times N$

Квадратичная функция или **квадратичная форма** в $\mathbb{R}^N$, связанная с матрицей $\mathbf{A}$, — это отображение Q, определяемое как

$$Q(\mathbf{x}) := \mathbf{x}^\mathsf{T}\mathbf{A}\mathbf{x} = \sum_{j=1}^{N}\sum_{i=1}^{N} a_{ij}x_i x_j$$

Пример. Пусть $N = 2$ и $\mathbf{A}$ — единичная матрица $\mathbf{I}$. Тогда

$$Q(\mathbf{x}) = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$$

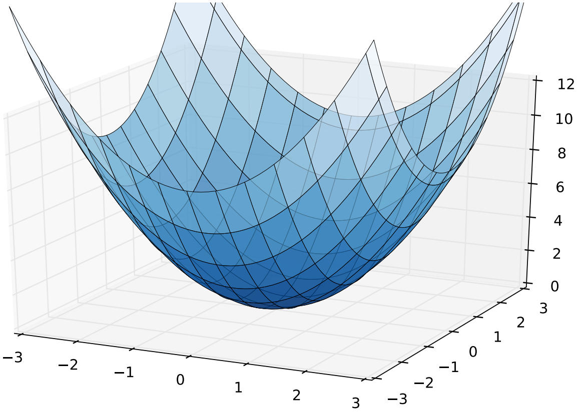


Рис.: Квадратичная функция $Q(\mathbf{x}) = x_1^2 + x_2^2$

Внимание:

- График для $Q(\mathbf{x}) = \|\mathbf{x}\|^2 = x_1^2 + x_2^2$ лежит всюду выше нуля

Матрица $\mathbf{A}$ с квадратичной формой с указанным выше свойством $Q(\mathbf{x}) \geq 0$ называется *положительно определенной*

В более общем смысле, симметричная матрица $\mathbf{A}$ размера $N \times N$ называется

- **неотрицательно определенной**, если $\mathbf{x}^\top \mathbf{A}\mathbf{x} \geq 0$ для всех $\mathbf{x} \in \mathbb{R}^N$,
- **положительно определенной**, если $\mathbf{x}^\top \mathbf{A}\mathbf{x} > 0$ для всех $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$,
- **неположительно определенной**, если $\mathbf{x}^\top \mathbf{A}\mathbf{x} \leq 0$ для всех $\mathbf{x} \in \mathbb{R}^N$, и
- **отрицательно определенной**, если $\mathbf{x}^\top \mathbf{A}\mathbf{x} < 0$ для всех $\mathbf{x} \in \mathbb{R}^N$ with $\mathbf{x} \neq \mathbf{0}$.

Если $\mathbf{A}$ не подходит ни к одной из этих категорий, то $\mathbf{A}$ называется **неопределенной**

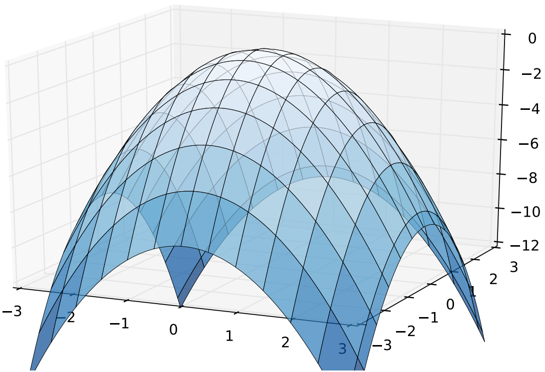


Рис.: Квадратичная функция $Q(\mathbf{x}) = -x_1^2 - x_2^2$

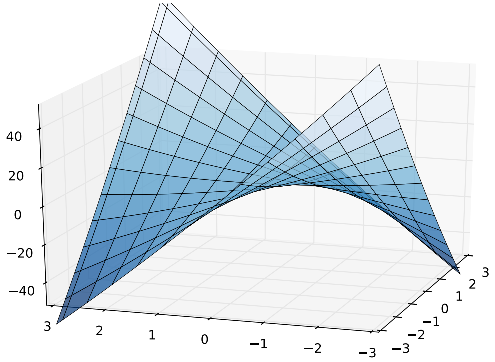


Рис.: Квадратичная функция $Q(\mathbf{x}) = x_1^2/2 + 8x_1x_2 + x_2^2/2$

Когда матрица $\mathbf{A}$ диагональная:

$$\mathbf{A} = \operatorname{diag}(d_1, \ldots, d_N) \quad \text{подразумевает} \quad Q(\mathbf{x}) = d_1 x_1^2 + \cdots + d_N x_N^2$$

Диагольная матрица положительно определена, тогда и только тогда, когда все диагональные элементы положительны

Факт. (??) Пусть $\mathbf{A}$ — некоторая симметричная матрица. Матрица $\mathbf{A}$

1. положительна определена тогда и только тогда, когда все ее собственные значения положительны
2. отрицательно определена тогда и только тогда, когда все ее собственные значения отрицательны

...аналогично для неположительно и неотрицательно определенных

Факт. (??) Если $\mathbf{A}$ положительно определена, то $\mathbf{A}$ несингулярна, $\det \mathbf{A} > 0$

Необходимое (но не достаточное) условие для каждого вида определенности:

Факт. (??) Если $\mathbf{A}$ положительна определена, то каждый элемент a_{nn} на главной диагонали положительный, то же самое для неотрицательной, неположительной и отрицательной.

Матрицы проекции

Напомним, что для любого подпространства $\mathbb{R}^N$, S, соответствующая проекция $\mathbf{P} = \operatorname{proj} S$ является линейным отображением из $\mathbb{R}^N$ в $\mathbb{R}^N$

Вспомним теорему **??**: существует матрица $\hat{\mathbf{P}}$ размера $N \times N$, такая что $\mathbf{P}\mathbf{x} = \hat{\mathbf{P}}\mathbf{x}$ для всех $\mathbf{x} \in \mathbb{R}^N$

- с этого момента $\mathbf{P}$ также будет означать соответствующую матрицу

Как выглядит эта матрица?

Теорема. (??) Пусть S — подпространство в $\mathbb{R}^N$. Если $\mathbf{P} = \operatorname{proj} S$, то

$$\mathbf{P} = \mathbf{B}(\mathbf{B}^\intercal \mathbf{B})^{-1} \mathbf{B}^\intercal \tag{2}$$

для каждой матрицы $\mathbf{B}$, такой что столбцы $\mathbf{B}$ формируют базис S

Смотрите упражнение **??** для доказательства

- Матрица $\mathbf{M} = \mathbf{I} - \mathbf{P}$ обозначает остаточную проекцию (смотрите страницу **??**)

Пример. Вспомним пример **??** на странице **??**

Мы выяснили, что проекция $\mathbf{y} \in \mathbb{R}^N$ на $\operatorname{span}\{\mathbf{1}\}$ — это $\bar{y}\mathbf{1}$

Тот же результат с помощью теоремы (**??**):

- Так как $\mathbf{1}$ является базисом $\operatorname{span}\{\mathbf{1}\}$:

$$\mathbf{P} = \operatorname{proj}\ \operatorname{span}\{\mathbf{1}\} \implies \mathbf{P} = \mathbf{1}(\mathbf{1}^\top \mathbf{1})^{-1}\mathbf{1}^\top = \frac{1}{N}\mathbf{1}\mathbf{1}^\top$$

- Значит, $\mathbf{P}\mathbf{y} = \bar{y}\mathbf{1}$, как и ожидалось
- Соответствующая остаточная проекция

$$\mathbf{M}_c = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^\top$$

Факт. (??) По постановке теоремы **??**, получается

1. $\mathbf{MB} = \mathbf{0}$
2. $\mathbf{PB} = \mathbf{B}$

Докажите в качестве упражнения (упр. **??** в ET)

Легко заметить, что $\mathbf{M}_c$ в предыдущем примере отображает $\mathbf{1}$ в $\mathbf{0}$

Квадратная матрица $\mathbf{A}$ является **идемпотентной**, если $\mathbf{AA} = \mathbf{A}$

Факт. (??) И $\mathbf{P}$, и $\mathbf{M}$ симметричны и идемпотентны

(Упражнение: проверьте прямыми вычислениями)

Интуиция: проецирование на подпространство дважды — это то же самое, что проецировать один раз – вспомните факт **??** на странице **??**

Факт. (**??**) Пусть S является линейным подпространством в $\mathbb{R}^N$. Если $\mathbf{P} = \operatorname{proj} S$ и $\mathbf{M}$ — остаточная проекция, то

1. $\operatorname{rank} \mathbf{P} = \operatorname{trace} \mathbf{P} = \dim S$ и
2. $\operatorname{rank} \mathbf{M} = \operatorname{trace} \mathbf{M} = N - \dim S$

Доказательство.

- Ранг линейного отображения — это размерность его диапазона. Когда $\mathbf{P} = \operatorname{proj} S$, Диапазон $\mathbf{P}$ равняется S
- Чтобы показать, что $\operatorname{trace} \mathbf{P} = \dim S$ также соблюдается, используем факт **??**– $\operatorname{trace} \mathbf{P} = \dim S$,
- Это следует из того, что $\operatorname{trace} \mathbf{M} = N - \dim S$, потому что

$$\operatorname{trace} \mathbf{M} = \operatorname{trace}(\mathbf{I} - \mathbf{P}) = \operatorname{trace} \mathbf{I} - \operatorname{trace} \mathbf{P} = N - \dim S$$

Переопределенные системы уравнений

Рассмотрим систему уравнений в виде $\mathbf{Ax} = \mathbf{b}$, когда:

- Матрица $\mathbf{A}$ размера $N \times K$ имеет полный ранг столбцов
- Вектор $\mathbf{x}$ имеет размер $K \times 1$
- Вектор $\mathbf{b}$ имеет размер $N \times 1$
- $K \leq N$

Принимая как данность $\mathbf{A}$ и $\mathbf{b}$, мы ищем $\mathbf{x} \in \mathbb{R}^K$, такой что $\mathbf{Ax} = \mathbf{b}$

Если $K = N$, то система имеет ровно одно решение

Когда $N > K$, система уравнений считается **переопределенной**:

- количество уравнений $>$ количества неизвестных
- количество ограничений $>$ количества степеней свободы

Возможно, не удастся найти $\mathbf{b}$, который удовлетворяет всем N уравнениям

Вспомним линейное отображение $T \colon \mathbb{R}^K \to \mathbb{R}^N$, где $T\mathbf{x} = \mathbf{A}\mathbf{x}$

Следующие утверждения эквивалентны:

1. существует $\mathbf{x} \in \mathbb{R}^K$, такой что $\mathbf{A}\mathbf{x} = \mathbf{b}$
2. вектор $\mathbf{b} \in \operatorname{colspace} \mathbf{A}$
3. вектор $\mathbf{b} \in \operatorname{rng} T$

Теорема **??** на странице **??**: когда $K < N$, функция T не может быть сюръекцией – возможные $\mathbf{b}$ находятся вне диапазона T

Когда $K < N$, случай с $\mathbf{b} \in \operatorname{colspace} \mathbf{A}$ является “очень редким“, потму что:

- точка $\mathbf{b}$ является произвольной точкой в $\mathbb{R}^N$
- пространство $\operatorname{colspace} \mathbf{A}$ имеет размерность K
- подпространста $\mathbb{R}^N$ с размерностями K имеют “нулевую меру Лебега” – “шанс” того, что $\mathbf{b}$ лежит в этом подпространстве, крошечный

Стандартный подход: признать, что точного решения может не существовать

Следует сосредоточиться на поиске $\mathbf{x} \in \mathbb{R}^K$, чтобы $\mathbf{A}\mathbf{x}$ оказалось настолько близко к $\mathbf{b}$, насколько возможно

- близки по обычной Евклидовой норме

Задача минимизации называется **методом наименьших квадратов**

$$\hat{\mathbf{x}} := \underset{\mathbf{x} \in \mathbb{R}^K}{\operatorname{argmin}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\| \tag{3}$$

При условии, что $\mathbf{A}$ размера $N \times K$ с $K \leq N$ и $\mathbf{b}$ размера $N \times 1$, мы можем использовать теорему ортогональной проекции для решения (3)

Теорема. (**??**) Если $\mathbf{A}$ имеет полный ранг столбцов, то (3) имеет единственное решение

$$\hat{\mathbf{x}} := (\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top \mathbf{b} \tag{4}$$

Доказательство.

Пусть:

- $\mathbf{A}$ и $\mathbf{b}$ будут как в формулировке теоремы
- $\hat{\mathbf{x}}$ как в (4) и
- $S := \operatorname{colspace} \mathbf{A}$

При условии полного ранга столбцов, столбцы $\mathbf{A}$ формируют базис для S. Применяем теорему **??**, ортогональная проекция $\mathbf{b}$ на S:

$$\mathbf{P}\mathbf{b} := \mathbf{A}(\mathbf{A}^\intercal \mathbf{A})^{-1}\mathbf{A}^\intercal \mathbf{b} = \mathbf{A}\hat{\mathbf{x}} \tag{5}$$

Поскольку теорема об ортогональной проекции дает единственный минимизатор в терминах ближайшей точки в S к $\mathbf{b}$,

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\| < \|\mathbf{b} - \mathbf{y}\| \quad \text{для всех } \mathbf{y} \in S,\ \mathbf{y} \neq \mathbf{A}\hat{\mathbf{x}} \tag{6}$$

Доказательство. (прод.) Возьмем любой $\mathbf{x} \in \mathbb{R}^K$, такой что $\mathbf{x} \neq \hat{\mathbf{x}}$

Получается $\mathbf{Ax} \in S$

К тому же, так как $\mathbf{x} \neq \hat{\mathbf{x}}$ и $\mathbf{A}$ имеют полный ранг столбцов, должно быть $\mathbf{Ax} \neq \mathbf{A}\hat{\mathbf{x}}$ (упр. **??**)

Значит

$$\|\mathbf{b} - \mathbf{A}\hat{\mathbf{x}}\| < \|\mathbf{b} - \mathbf{Ax}\| \quad \text{для всех } \mathbf{x} \in \mathbb{R}^K,\ \mathbf{x} \neq \hat{\mathbf{x}}$$

Другими словами, $\hat{\mathbf{x}}$ является единственным решением (3)

В (4) матрица $(\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}$ называется **псевдообратной** матрицы $\mathbf{A}$

Если $K = N$, то решение методом наименьших квадратов $\hat{\mathbf{x}}$ в (4) сокращается до:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

Что произойдет, если столбцы $boldA$ не будут линейно независимыми?

- множество $\operatorname{colspace} \mathbf{A}$ все еще линейное подпространство и теорема ортогональной проекции все еще дает нам ближайшую точку $\mathbf{Pb}$ к $\mathbf{b}$ в $\operatorname{colspace} \mathbf{A}$
- так как $\mathbf{Pb} \in \operatorname{colspace} \mathbf{A}$, все еще существует вектор $\hat{\mathbf{x}}$, такой что $\mathbf{Pb} = \mathbf{A}\hat{\mathbf{x}}$
- но таких векторов существует бесконечное множество

Смотрите упражнение **??**

QR-разложение

QR-разложение данной матрицы $\mathbf{A}$ — это произведение вида$\mathbf{QR}$

- первая матрица имеет ортонормированные столбцы и
- вторая является верхней треугольной

Приложения включают задачи наименьших квадратов и вычисление собственных значений

Теорема. (??) Если матрица $\mathbf{A}$ размера $N \times K$ имеет полный ранг столбцов, то существует факторизация $\mathbf{A} = \mathbf{QR}$, где

1. $\mathbf{R}$ размера $K \times K$ является верхней треугольной и несингулярной, и
2. $\mathbf{Q}$ размера $N \times K$ имеет ортонормированные столбцы

Смотрите страницу **??** в ET для доказательства

Возьмем разложение $\mathbf{A} = \mathbf{QR}$, рещение методом наименьших квадратов $\hat{\mathbf{x}}$, определенное в (4), может также записываться как:

$$\hat{\mathbf{x}} = \mathbf{R}^{-1}\mathbf{Q}^{\mathsf{T}}\mathbf{b}$$

Смотрите упр. ??

Умножение на $\mathbf{R}$:

$$\mathbf{R}\hat{\mathbf{x}} = \mathbf{Q}^{\mathsf{T}}\mathbf{b}$$

Диагонализация и спектральная теория

Если $f\colon A \to A$ и $g\colon B \to B$, то говорят, что g **топологически сопряжен** с f везде, где существует непрерывная биекция $\tau\colon B \to A$, такая что

$$f = \tau \circ g \circ \tau^{-1}$$

Может быть полезно, если g проще f

Говорят, что квадратная матрица $\mathbf{A}$ **подобна** матрице $\mathbf{B}$, если существует обратимая матрица $\mathbf{P}$, такая что $\mathbf{A} = \mathbf{PBP}^{-1}$

$$
\begin{tikzcd}
\bullet \arrow[r, "\mathbf{B}"] & \bullet \arrow[d, "\mathbf{P}"] \\
\underset{\mathbf{x}}{\bullet} \arrow[u, "\mathbf{P}^{-1}"] \arrow[r, "\mathbf{A}"] & \underset{\mathbf{Ax}}{\bullet}
\end{tikzcd}
$$

Рис.: $\mathbf{A}$ подобна $\mathbf{B}$

Если $\mathbf{A}$ подобна диагональной матрице, то $\mathbf{A}$ называется **диагонализируемой**

Нас интересует подобие простым матрицам, диагональные матрицы — самый простой вид

Факт. (??) Если $\mathbf{A}$ подобна $\mathbf{B}$, то $\mathbf{A}^t$ подобна $\mathbf{B}^t$ для всех $t \in \mathbb{N}$

Пример. Мы хотим вычислить $\mathbf{A}^t$ для некоторых данных $t \in \mathbb{N}$

Если $\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$ для некоторых $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$, то по факту **??** и факту **??**, получается

$$\mathbf{A}^t = \mathbf{P} \operatorname{diag}(\lambda_1^t, \ldots, \lambda_N^t)\mathbf{P}^{-1}$$

Диагонализация и собственные значения

Факт. (??) Если $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ для некоторых $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \dots, \lambda_N)$, то $(\mathrm{col}_n \mathbf{P}, \lambda_n)$ является собственной парой $\mathbf{A}$ для любого n

Доказательство. Заметим, что $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ значит, что $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{\Lambda}$

Equating the nth column on each side gives Рассмотрим столбец n с каждой стороны:

$$\mathbf{A}\mathbf{p}_n = \lambda_n \mathbf{p}_n$$

Где $\mathbf{p}_n := \mathrm{col}_n \mathbf{P}$

Заметим, что $\mathbf{p}_n$ — ненулевой вектор, так как $\mathbf{P}$ обратима

Но когда $\mathbf{A}$ диагонализируема?

Факт. (3.3.7) Матрица $\mathbf{A}$ размера $N \times N$ диагонализируема тогда и только тогда, когда она имеет N линейно независимых собственных векторов

В некоторых случаях, мы можем получить еще более простое разложение матрицы, если матрица $\mathbf{P}$ имеет ортогональные столбцы

Такие матрицы называются **ортогональными матрицами**

Факт. (??) Если $\mathbf{Q}$ и $\mathbf{P}$ — ортогональные матрицы размера $N \times N$, то

1. $\mathbf{Q}^\mathsf{T}$ ортогональна и $\mathbf{Q}^{-1} = \mathbf{Q}^\mathsf{T}$,
2. $\mathbf{QP}$ ортогонально, и
3. $\det \mathbf{Q} \in \{-1, 1\}$.

Если $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$ и $\mathbf{Q}$ имеют ортогональные столбцы, то

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathsf{T}}$$

Ясно, что $\mathbf{A}$ должна быть симметричной. Следующая теорема показывает, что симметрия $\mathbf{A}$ также достаточна

Теорема. (??) Если $\mathbf{A}$ симметрична, то $\mathbf{A}$ может быть диагонализована как $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathsf{T}}$, где $\mathbf{Q}$ — ортогональная матрица и $\mathbf{\Lambda}$ — диагональная матрица, сформированная из собственных значений $\mathbf{A}$

Вышеупомянутая теорема была версией **теоремы о спектральном разложении**

$\mathbf{Q\Lambda Q}^\intercal$ называется **симметричным разложением на собственные значения** матрицы $\mathbf{A}$ – воздействие $\mathbf{A}$ на вектор $\mathbf{x}$ размера $N \times 1$:

$$\mathbf{Ax} = \sum_{n=1}^{N} \lambda_n (\mathbf{u}_n^\intercal \mathbf{x}) \mathbf{u}_n$$

где λ_n является n-ым собственным значением $\mathbf{A}$ и $\mathbf{u}_n = \mathrm{col}_n\, \mathbf{Q}$

Сравните с $\mathbf{x} = \sum_{n=1}^{N} (\mathbf{u}_n^\intercal \mathbf{x}) \mathbf{u}_n$

Факт. (??) Если $\mathbf{A}$ неотрицательно определена, то $\sqrt{\mathbf{A}}$ существует и равняется $\mathbf{Q}\sqrt{\mathbf{\Lambda}}\mathbf{Q}^\mathsf{T}$. Матрица $\sqrt{\mathbf{\Lambda}}$ равняется $\mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_N})$

Факт. (??) Если $\mathbf{A}$ положительно определена, то существует несингулярная, верхняя треугольная матрица $\mathbf{R}$, такая что $\mathbf{A} = \mathbf{R}^\mathsf{T}\mathbf{R}$

Такое разложение называется разложением Холецкого

Доказательство. (Разложение Холецкого) Мы можем написать:

$$\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^\intercal = \mathbf{Q}\sqrt{\boldsymbol{\Lambda}}\sqrt{\boldsymbol{\Lambda}}\mathbf{Q}^\intercal = (\sqrt{\boldsymbol{\Lambda}}\mathbf{Q}^\intercal)^\intercal\sqrt{\boldsymbol{\Lambda}}\mathbf{Q}^\intercal$$

Затем применим QR-разложение к $\sqrt{\boldsymbol{\Lambda}}\mathbf{Q}^\intercal$:

$$\sqrt{\boldsymbol{\Lambda}}\mathbf{Q}^\intercal = \tilde{\mathbf{Q}}\mathbf{R}$$

где $\mathbf{R}$ несингулярная и верхняя треугольная, и $\tilde{\mathbf{Q}}$ имеет ортонормированные столбцы

Так как столбцы $\tilde{\mathbf{Q}}$ ортонормированные,

$$\mathbf{A} = (\tilde{\mathbf{Q}}\mathbf{R})^\intercal\tilde{\mathbf{Q}}\mathbf{R} = \mathbf{R}^\intercal\tilde{\mathbf{Q}}^\intercal\tilde{\mathbf{Q}}\mathbf{R} = \mathbf{R}^\intercal\mathbf{R}$$

Нормы и непрерывность

Возьмем векторную последовательность $\{\mathbf{x}_n\}$ в $\mathbb{R}^K$ и любую точку $\mathbf{x} \in \mathbb{R}^K$. Говорят, что $\{\mathbf{x}_n\}$ **сходится** к $\mathbf{x}$ и записывается $\mathbf{x}_n \to \mathbf{x}$, если для любых $\epsilon > 0$, существует $N \in \mathbb{N}$, такое что $\|\mathbf{x}_n - \mathbf{x}\| < \epsilon$ при $n \geq N$

Эквивалентно, последовательность действительных чисел $z_n := \|\mathbf{x}_n - \mathbf{x}\|$ сходится к нулю ы $\mathbb{R}$ при $n \to \infty$

Факт. (3.3.11) Имеют место следующие результаты:

1. если $\mathbf{x}_n \to \mathbf{x}$ и $\mathbf{y}_n \to \mathbf{y}$, то $\mathbf{x}_n + \mathbf{y}_n \to \mathbf{x} + \mathbf{y}$.
2. если $\mathbf{x}_n \to \mathbf{x}$ и $\alpha \in \mathbb{R}$, то $\alpha \mathbf{x}_n \to \alpha \mathbf{x}$.
3. $\mathbf{x}_n \to \mathbf{x}$ тогда и только тогда, когда $\mathbf{a}^\mathsf{T} \mathbf{x}_n \to \mathbf{a}^\mathsf{T} \mathbf{x}$ для всех $\mathbf{a} \in \mathbb{R}^K$.

Мы хотим распространить понятие сходимости на матрицы

Норма матрицы $\mathbf{A}$ размера $N \times K$:

$$\|\mathbf{A}\| := \max\left\{\|\mathbf{A}\mathbf{x}\| \,:\, \mathbf{x} \in \mathbb{R}^K,\ \|\mathbf{x}\| = 1\right\} \tag{7}$$

Значение нормы матрицы в общем случае найти непросто

Однако, матричная норма ведет себя как векторная норма

Факт. (3.3.12) Для любых согласующихся матриц $\mathbf{A}$ и $\mathbf{B}$, матричная норма удовлетворяет

1. $\|\mathbf{A}\| \geq 0$ и $\|\mathbf{A}\| = 0$ тогда и только тогда, когда все записи $\mathbf{A}$ нулевые,
2. $\|\alpha \mathbf{A}\| = |\alpha| \|\mathbf{A}\|$ для любых скаляров α,
3. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$, и
4. $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$.

Факт. (??) Для любых матриц $\mathbf{A}$ размера $J \times K$ с элементами a_{jk}, выходит

$$\|\mathbf{A}\| \leq \sqrt{JK} \max_{jk} |a_{jk}|$$

если каждый элекмент $\mathbf{A}$ близок к нулю, то $\|\mathbf{A}\|$ тоже близка к нулю

Ряд Неймана

В дальнейшем мы изучаем динамические системы вида

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{b}$$

Существует ли “стационарный“ вектор $\mathbf{x} \in \mathbb{R}^N$ в том смысле, что $\mathbf{x}_t = \mathbf{x}$ значит, что $\mathbf{x}_{t+1} = \mathbf{x}$?

Мы ищем $\mathbf{x} \in \mathbb{R}^N$, являющийся решением системы уравнений

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{b} \qquad (\mathbf{A} \text{ is } N \times N \text{ и } \mathbf{b} \text{ is } N \times 1) \tag{8}$$

Рассмотрим случай со скалярами $x = ax + b$

Если $|a| < 1$, то существует единственное решение

$$\bar{x} = \frac{b}{1-a} = b\sum_{k=0}^{\infty} a^k$$

Лемма о рядах Неймана помогает обобщить это до $\mathbb{R}^N$

Теорема. (**??**) Если матрица $\mathbf{A}$ квадратная и $\|\mathbf{A}^j\| < 1$ для некоторых $j \in \mathbb{N}$, то $\mathbf{I} - \mathbf{A}$ обратима, и

$$(\mathbf{I} - \mathbf{A})^{-1} = \sum_{i=0}^{\infty} \mathbf{A}^i$$

Когда выполняется условие леммы о рядах Неймана, (8) имеет единственное решение

$$\bar{\mathbf{x}} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{b}$$

Чтобы проверить это решение, используем **спектральный радиус** $\mathbf{A}$:

$$\varrho(\mathbf{A}) := \max\{|\lambda| : \lambda \text{ является собственным значением } \mathbf{A}\}$$

$|\lambda|$ — это **модуль** возможно комплексного числа λ

Факт. Если $\varrho(\mathbf{A}) < 1$, то $\|\mathbf{A}^j\| < 1$ для некоторых $j \in \mathbb{N}$

Почему достаточно $\varrho(\mathbf{A}) < 1$?

Нам нужно $\sum_{i=0}^{t} \mathbf{A}^i(\mathbf{I} - \mathbf{A})$ было близко к $\mathbf{I}$ для больших t

Получается:

$$\sum_{i=0}^{t} \mathbf{A}^i(\mathbf{I} - \mathbf{A}) = \sum_{i=0}^{t} \mathbf{A}^i - \sum_{i=0}^{t} \mathbf{A}^{i+1} = \mathbf{I} - \mathbf{A}^{t+1}$$

Упростим до случая, где $\mathbf{A}$ диагонализируется:

$$\mathbf{A} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^{-1}$$

где $\boldsymbol{\Lambda}$ — диагональная матрица, содержащая собственные значения $\lambda_1, \ldots, \lambda_N$ матрицы $\mathbf{A}$ на главной диагонали.

Теперь используем факт **??**,

$$\mathbf{A}^t = \mathbf{P} \begin{pmatrix} \lambda_1^t & 0 & \cdots & 0 \\ 0 & \lambda_2^t & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_N^t \end{pmatrix} \mathbf{P}^{-1}$$

Если $\varrho(\mathbf{A}) < 1$, то $|\lambda_n| < 1$ для всех n, и значит $\lambda_n^t \to 0$ при $t \to \infty$. Из этого следует, что $\mathbf{A}^t \to \mathbf{0}$

Учебник по Эконометрике

Лекция 3: Основы теории вероятностей

Джон Стачурски

Лекции: Акшай Шенкер

Перевел: Алексей Кедо

1 октября 2020 г.

Вероятность фундаментальна для статистики и эконометрики, но технически сложна:

- множество событий, которым мы хотим присвоить вероятности, может быть очень большим
- нам нужны способы справиться со сложностью

Прежде чем мы начнем:

- множество S исчеслимо, если оно конечно или может быть представлено как последовательность
- в противном случае, множество S неисчеслимо

Пространство элементарных событий

Пространство элементарных событий можно представить себе как "список" всех возможных результатов в данном случайном эксперименте:

- пространство элементарных событий обычно обозначается как Ω
- пространством элементарных событий может быть только непустое множество
- типичный элемент Ω обозначается как ω

Реализация неопределенности приведет к выбору конкретной $\omega \in \Omega$

Пример. В случайном эксперименте, заключающемся в бросании кости один раз, множество возможных результатов представлено как $\Omega := \{1, \dots, 6\}$

Пример. Обезьяна Бертона Малкиела с завязанными глазами метает дротики в мишень радиусом 1

Введем обычные декартовы координаты с началом в центре мишени

Пусть (h, v) — типичные координаты, измеренные по горизонтали и вертикали соответственно

Естественное пространство элементарных событий — это $\Omega := \{(h, v) \in \mathbb{R}^2 : \|(h, v)\| \leq 1\}$ – также называется **единичным кругом** в $\mathbb{R}^2$

Неформально, **событие** — это подмножество Ω (мы скоро рассмотрим некоторые предостережения)

Событие A происходит всякий раз, когда отдельный $\omega \in \Omega$, выбранный в случайном эксперименте, оказывается в A

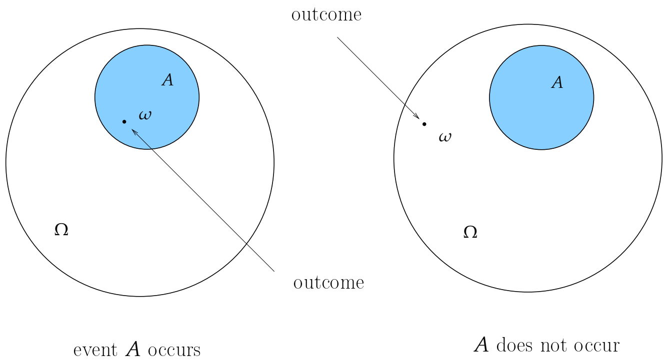


Рис.: Результаты и события

Вероятности и события

Можем ли мы присвоить вероятности каждому $\omega \in \Omega$?

Рассмотрим модель метания дротика, где Ω is $\mathbb{R}$:

- для $A \subset \Omega$, вероятность того, что дротик попадет в A пропорционален площади A
- вероятность точки $\omega \in \Omega$ будет меньше любой области A, содержащей ω
- для любых $\epsilon > 0$, мы можем найти A, содержащее ω, с площадью меньше ϵ

вероятность попадания в ω меньше ϵ для любых $\epsilon > 0$, значит вероятность попадания в ω должна быть равна нулю!

Итог: когда пространство элементарных событий исчеслимо, присваивать вероятности событиям (подмножествам Ω), не каждому $\omega \in \Omega$

Но можем ли мы присваивать вероятности *каждому* попространству Ω?

В модели дротика:

$$\mathbb{P}(A) = \frac{\lambda(A)}{\pi}$$

где $\lambda(A) :=$ площадь множества A

Определение площади A для всех $A \subset \Omega$ проблематично:

- пространство Ω, наш круг для дартс в $\mathbb{R}^2$, содержит много подмножеств, что производит странные явления
- Парадокс Банаха — Тарского

Решение: не принимайте множество событий за все подмножества Ω

Возьмите множество событий как определенные "хорошие" подмножества Ω, обозначенные $\mathscr{F}$

Присваивайте вероятности только подмножествам Ω в $\mathscr{F}$

Сигма-алгебра

Как мы можем гарантировать, что $\mathscr{F}$ достаточно большой? В разумной вероятностной модели мы в идеале хотим:

- событие "не A" принадлежит $\mathscr{F}$, если $A \in \mathscr{F}$
- событие "A или B" принадлежит $\mathscr{F}$, если $A \in \mathscr{F}$ и $A \in \mathscr{F}$

Формально, $\mathscr{F}$ — это **σ-алгебра** множества Ω, если

1. $A \in \mathscr{F} \implies A^c \in \mathscr{F}$,
2. $A_1, A_2, \ldots \in \mathscr{F} \implies \cup_{n=1}^{\infty} A_n \in \mathscr{F}$ и
3. $\Omega \in \mathscr{F}$

1. - 3. подразумевают, что $\varnothing \in \mathscr{F}$, если $\mathscr{F}$ — σ-алгебра

Событие $\varnothing$ называется **невозможным событием**

Событие Ω называется **достоверным событием**

Пример. Множество $\{\Omega, \varnothing\}$ — σ - алгебра, называемая **тривиальной σ-алгеброй**

Борелевская σ -алгебра

σ-алгебра событий меняется от задачи к задаче

В $\mathbb{R}^N$ мы используем Борелевские множества, обозначенные $\mathscr{B}(\mathbb{R}^N)$

- наименьшая σ-алгебра, содержащая все прямоугольники в $\mathbb{R}^N$

Почему Борелевская σ-алгебра?

- исключает "странные" множества
- включает повседневные полезные множества (включая плоскости и гиперплоскости, круги, сферы, многоугольники, конечные множества и последовательности точек)

Вероятности

Для данного события $B \in \mathscr{F}$, символ $\mathbb{P}(B)$ показывает "вероятность, что событие B случится"

> *$\mathbb{P}(B)$ показывает вероятность, что когда неопределенность решена и некоторые $\omega \in \Omega$ выбраны "естественно", то утверждение $\omega \in B$ является верным*

Нам нужно установить ограничения, чтобы сделать вероятности правильными

Например, мы хотим исключить $\mathbb{P}(B) = -93$ для некоторых B

Пусть Ω — непустое множество и $\mathscr{F}$ — σ-алгебра подпространств Ω. **Вероятность** $\mathbb{P}$ на $(\Omega, \mathscr{F})$ — функция из $\mathscr{F}$ в $[0, 1]$, которая удовлетворяет

1. $\mathbb{P}(\Omega) = 1$ и
2. $\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ для любой непересекающейся последовательности множеств $A_1, A_2, \ldots \in \mathscr{F}$

$\mathbb{P}$ также называется **вероятностной мерой**; втроем $(\Omega, \mathscr{F}, \mathbb{P})$ называются **вероятностным пространством**

Аксиома 1.: мы требуем $\mathbb{P}(\Omega) = 1$, так как, по построению, любая возможная ω лежит в множестве Ω.

Аксиома 2. называется **счетной аддитивностью**

В формулировке аксиомы (ii) несовместность попарна: любая различная пара A_i, A_j не имеет общих точек

Счетная аддитивность подразумевает конечную **аддитивность**:

$$\mathbb{P}(A_1 \cup \cdots \cup A_k) = \mathbb{P}(A_1) + \cdots + \mathbb{P}(A_k) \tag{1}$$

всякий раз, когда $A_1, \ldots, A_k$ несовместные

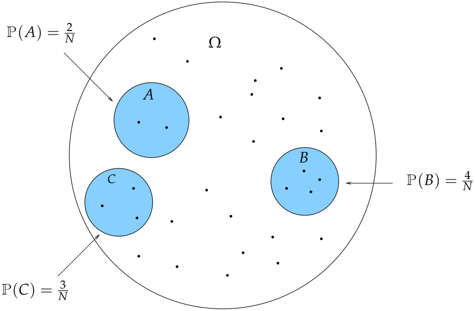


Рис.: Каждая из N точек случается с вероятностью $1/N$

$$\mathbb{P}(A \cup B \cup C) = \frac{9}{N} = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C)$$

Пример. Пусть $\Omega := \{1, \dots, 6\}$ представляют шесть разных граней кубика, как в примере **??**

Так как Ω конечна, пусть $\mathscr{F}$ — множество всех подмножеств Ω

Определим вероятность $\mathbb{P} \colon \mathscr{F} \to [0, 1]$

$$\mathbb{P}(A) := \frac{|A|}{6} \quad \text{, где } |A| := \text{ количество элементов в множестве } A \tag{2}$$

Легко заметить, что $0 \leq \mathbb{P}(A) \leq 1$ для любых $A \in \mathscr{F}$ и что $\mathbb{P}(\Omega) = 1$

Пример. (прод.) Касательно аддитивности, предположим, что A и B — два несовместных подмножества $\{1, \ldots, 6\}$

Тогда $|A \cup B| = |A| + |B|$, значит

$$\mathbb{P}(A \cup B) = \frac{|A \cup B|}{6} = \frac{|A| + |B|}{6} = \frac{|A|}{6} + \frac{|B|}{6} = \mathbb{P}(A) + \mathbb{P}(B)$$

Это доказывает аддитивность для пар множеств. Аналогичный аргумент подтверждает аддитивность для любого конечного набора

Конечная аддитивность в этом случае эквивалентна счетной аддитивности, поскольку общее количество различных событий конечно.

Пример.

Чип памяти состоит из миллиардов переключателей/битов

- Переключатели могут быть выключены или включены (0 или 1)

Генератор случайных чисел обращается к N битам, включая или выключая каждый из них

Получается

- $\Omega := \{(b_1, \ldots, b_N)$: где b_n равняется 0 или 1 для каждого $n\}$
- $\mathbb{P}(A) := 2^{-N}(\#A)$

Упражнение: Покажите, что $\mathbb{P}$ — вероятность

Пример. Рассмотрим снова модель мишени, где Ω — это единичный круг в $\mathbb{R}^2$

Для пространства событий, возьмем $\mathscr{F}$ как множество Борелевских подмножеств в $\mathbb{R}^2$, лежащее в Ω

Для $\mathbb{P}$ мы следуем "равномерному"распределению вероятностей, заданному

То есть, $\mathbb{P}(B) = \lambda(B)/\pi$ для каждого $B \in \mathscr{F}$

Функция λ, которая назначает область для Борелевских множеств, известна как счетно-аддитивная, то есть $\lambda(\cup_n A_n) = \sum_{n=1}^{\infty} \lambda(A_n)$ при условии, что эти множества несовместные

Очевидно, $\mathbb{P}(\Omega) = 1$

Мера Лебега

Функция λ, отображающая Борелевские множества на свою "территорию", формально известна как **Мера Лебега**

§?? в ET дает краткое введение в эту концепцию

Свойства вероятностной меры

Факт. (??) Пусть $(\Omega, \mathscr{F}, \mathbb{P})$ — вероятностное пространство и $A, B \in \mathscr{F}$

Если $A \subset B$, то

1. $\mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$,
2. $\mathbb{P}(A) \leq \mathbb{P}(B)$ (**монотонность**)
3. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$, и
4. $\mathbb{P}(\varnothing) = 0$.

Доказательство. Когда $A \subset B$, имеется $B = (B \setminus A) \cup A$, значит

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A)$$

Все результаты следуют из этого (почему?)

Факт. (??) Если A и B — какие-нибудь (не обязательно несовместные) события, то

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Докажите в качестве упражнения **??** в ЕТ

Факт подразумевает, что **полуаддитивность**: для любых $A, B \in \mathscr{F}$, имеется

$$\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$$

Условная вероятность и независимость

Условная вероятность A при данном B:

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{3}$$

Вероятность A, при данной информации, что B случилось

События A и B называются **независимыми**, если $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

- Если A и B независимы, то

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A)$$

Пример. Эксперимент: бросим кубик дважды

$$\Omega := \{(i,j) : i,j \in \{1,\dots,6\}\} \quad \text{и} \quad \mathbb{P}(E) := \#E/36$$

Теперь рассмотрим события

$$A := \{(i,j) \in \Omega : i \text{ четное}\} \quad \text{и} \quad B := \{(i,j) \in \Omega : j \text{ четное}\}$$

В этом случае мы имеем

$$A \cap B = \{(i,j) \in \Omega : i \text{ и } j \text{ четные}\}$$

Упражнение: убедитесь, что $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Следовательно, A и B независимы с вероятностью $\mathbb{P}$

Закон полной вероятности

Закон полной вероятности утверждает, что:

Факт. (??) Если $A \in \mathscr{F}$ и $B_1, \dots, B_M$ — части Ω с $\mathbb{P}(B_m) > 0$ для всех m, то

$$\mathbb{P}(A) = \sum_{m=1}^{M} \mathbb{P}(A \mid B_m) \cdot \mathbb{P}(B_m)$$

Доказательство. Возьмем $A \in \mathscr{F}$ и части $B_1, \dots, B_M$:

$$\mathbb{P}(A) = \mathbb{P}[A \cap (\cup_{m=1}^{M} B_m)] = \mathbb{P}[\cup_{m=1}^{M}(A \cap B_m)]$$

$$= \sum_{m=1}^{M} \mathbb{P}(A \cap B_m) = \sum_{m=1}^{M} \mathbb{P}(A \mid B_m) \cdot \mathbb{P}(B_m)$$

Теорема Байеса

Теорема Байеса: для любых событий A и B с положительной вероятностью, выполняется

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{4}$$

Доказательство. Из определения условной вероятности:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \quad \text{и} \quad \mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Следовательно $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\,\mathbb{P}(B) = \mathbb{P}(B \mid A)\,\mathbb{P}(A)$

Перегруппировка уравнения (4)

Пример. Банки используют автоматизированные системы для обнаружения мошеннических или незаконных транзакций

Рассмотрим тест, который отвечает на каждую транзакцию с P или N:

- P значит "положительно" (транзакция отмечена как мошенническая)
- N значит "отрицательно" (транзакция отмечена как нормальная)

Пусть F значит мошенническая, предположим

- $\mathbb{P}(P \mid F) = 0.99$ (тест отмечает 99% мошеннических транзакций),
- $\mathbb{P}(P \mid F^c) = 0.01$ (процент ложных срабатываний), и
- $\mathbb{P}(F) = 0.001$ (распространенность мошенничества)

Какова вероятность мошенничества при положительном тесте?

Обратите внимание на закон Байеса

$$\mathbb{P}(F \mid P) = \frac{\mathbb{P}(P \mid F)\mathbb{P}(F)}{\mathbb{P}(P)}$$

и обратите внимание на закон полной вероятности

$$\mathbb{P}(P) = \mathbb{P}(P \mid F)\mathbb{P}(F) + \mathbb{P}(P \mid F^c)\mathbb{P}(F^c)$$

Следовательно

$$\mathbb{P}(F \mid P) = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.01 \times 0.999} = \frac{11}{122} \approx \frac{1}{11}$$

Случайные переменные

Неформально: "значение, которое изменяется случайным образом"

Формально: **случайная переменная** x — функция из Ω в $\mathbb{R}$

Интерпретация: случайные величины преобразуют исходы в пространстве элементарных событий в числовые исходы.

Главная идея:

- "природа" выбирает ω в Ω
- случайная переменная сообщает результат как $x(\omega) \in \mathbb{R}$

Пример. Предположим, что Ω — множество бесконечных двоичных последовательностей

$$\Omega := \{(b_1, b_2, \ldots) : b_n \in \{0,1\} \text{ для каждого } n\}$$

Мы можем создавать различные отображения случайных переменных $\Omega \to \mathbb{R}$:

- количество "подбрасываний" до первого "орла":

$$x(\omega) = x(b_1, b_2, \ldots) = \min\{n : b_n = 1\}$$

- количество "орлов" в первых 10 "подбрасываниях":

$$x(\omega) = x(b_1, b_2, \ldots) = \sum_{n=1}^{10} b_n$$

- количество подбрасываний до первого орла:

$$x(\omega) = x(b_1, b_2, \dots) = \min\{n \in \mathbb{N} : b_n = 1\}$$

- **Двоичная случайная величина или случайная величина Бернулли** говорит нам, возникают ли какие-либо орлы в первые 10 подбрасываний:

$$x(\omega) = y(b_1, b_2, \dots) := \min\left\{\sum_{n=1}^{10} b_n, 1\right\} \tag{5}$$

Случайная величина Бернулли

Бернулли или **двоичная случайная величина** с.в. x принимает значения $\{0, 1\}$

Теперь мы рассмотрим общий способ создания с.в. Бернулли

Пусть Q — утверждение, например, "a больше 3"

Определение: $\mathbb{1}\{Q\}$ равняется единице, если Q верно, в ином случае нулю.

Определим

$$x(\omega) = \mathbb{1}\{\omega \in A\} \text{ , где } A \in \mathscr{F}$$

С.в. показывает, случается ли событие C

Распространенный вариант обозначений: для произвольного $A \in \mathscr{F}$:

$$\mathbb{1}_A(\omega) := \mathbb{1}\{\omega \in A\} := \begin{cases} 1 & \text{, если } \omega \in A \\ 0 & \text{, иначе} \end{cases}$$

Факт. (??) Если $A_1, \ldots, A_N$ — подмножества Ω, то

1. $\mathbb{1}_{\bigcap_{n=1}^N A_n} = \prod_{n=1}^N \mathbb{1}_{A_n}$ и
2. $\mathbb{1}_{\bigcup_{n=1}^N A_n} = \sum_{n=1}^N \mathbb{1}_{A_n}$ всякий раз, когда множества несовместные

Смотрите упражнение ?? для доказательства

Здесь равенство означает оценку при любом $\omega \in \Omega$

Условные обозначения

Общие условные обозначения для с.в.:

$$\{x \text{ обладает некоторым свойством}\} :=$$
$$\{\omega \in \Omega : x(\omega) \text{ обладает некоторым свойством}\}$$

Пример.

$$\{x \leq 2\} := \{\omega \in \Omega : x(\omega) \leq 2\}$$

$$\therefore \quad \mathbb{P}\{x \leq 2\} := \mathbb{P}\{\omega \in \Omega : x(\omega) \leq 2\}$$

Пример. Возьмем случайную величину x и $a \leq b$, мы утверждаем, что

$$\mathbb{P}\{x \leq a\} \leq \mathbb{P}\{x \leq b\}$$

Это выполняется, так как

$$\{x \leq a\} := \{\omega \in \Omega : x(\omega) \leq a\}$$

$$\subset \{\omega \in \Omega : x(\omega) \leq b\} := \{x \leq b\}$$

Теперь применим монотонность: $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

Равенства, неравенства и арифметические операции следует интерпретировать *точечно*:

- $x \leq y \iff x(\omega) \leq y(\omega)$ для всех $\omega \in \Omega$,
- $x = y \iff x(\omega) = y(\omega)$ для всех $\omega \in \Omega$, и
- $z = \alpha x + \beta y \iff z(\omega) = \alpha x(\omega) + \beta y(\omega)$ для всех $\omega \in \Omega$

Случайные переменные — измеримые функции

Пусть $(\Omega, \mathscr{F}, \mathbb{P})$ — некоторое вероятностное пространство

Пусть B — некоторое подпространство $\mathbb{R}$

Рассмотрим вероятность

$$\mathbb{P}\{x \in B\} := \mathbb{P}\{\omega \in \Omega : x(\omega) \in B\}$$

где x — некоторая функция из Ω в $\mathbb{R}$

Нет возможности быть уверенным, что $\{\omega \in \Omega : x(\omega) \in B\}$ является элементом $\mathscr{F}$

- $\mathbb{P}\{x \in B\}$ может быть не определена

Мы должны ввести ограничения:

- для B мы естественно ограничиваем внимание до $\mathscr{B}(\mathbb{R})$, Борелевское подмножество $\mathbb{R}$
- для x мы требуем $\{x \in B\} \in \mathscr{F}$ всюду, где B — Борелевская множество

Формальное определение случайной величины:

Случайная величина в $(\Omega, \mathscr{F})$ — функция $x \colon \Omega \to \mathbb{R}$, удовлетворяющая

$$\{\omega \in \Omega : x(\omega) \in B\} \in \mathscr{F} \quad \text{для всех } B \in \mathscr{B}(\mathbb{R}) \tag{6}$$

Эти виды функций также называются $\mathscr{F}$-измеримыми функциями

Обозначение прообраза: $x^{-1}(B)$ — это все $\omega \in \Omega$, такие что $x(\omega) \in B$

Перепишем (6) как

$$x^{-1}(B) \in \mathscr{F} \quad \text{для всех } B \in \mathscr{B}(\mathbb{R})$$

Таким образом, x "откатывает" Борелевские множества до событий

Измеримые преобразования

Мы хотим обсудить некоторые преобразования x

Например, $y := e^x$. y тоже случайная величина?

Да, при условии, что преобразование удовлетворяет Борелевской измеримости

Формально, $f\colon \mathbb{R} \to \mathbb{R}$ называется **Борелевски измеримой** или **$\mathscr{B}$-измеримой**, если

$$f^{-1}(B) \in \mathscr{B}(\mathbb{R}) \quad \text{для всех } B \in \mathscr{B}(\mathbb{R}) \tag{7}$$

Класс $\mathscr{B}$-измеримых функций огромен: любая непрерывная функция, любая возрастающая функция и т.д.

Предположим, что f $\mathscr{B}$-измерима и x — случайная величина

Мы имеем $\{y \in B\} \in \mathscr{F}$ для всех $B \in \mathscr{B}(\mathbb{R})$, так как

$$\{y \in B\} = \{f(x) \in B\} = \{x \in f^{-1}(B)\} \tag{8}$$

Таким образом, $y = f(x)$ — случайная величина

Ожидания

Мы хотим определить ожидания для произвольной с.в. x

Грубо говоря, $\mathbb{E}[x] :=$ "сумма" всех возможных значений x, взвешенных по их вероятностям.

"Сумма" в кавычках, потому что может быть бесконечное число слагаемых

Мы используем современный, формальный и строгий подход к определению ожиданий

Для конечных случайных величин дано вероятностное пространство $(\Omega, \mathscr{F}, \mathbb{P})$ и случайная величина x, принимающая только конечное количество различных значений $s_1, \ldots, s_J$, **ожидание** x определяется как

$$\mathbb{E}\, x = \sum_{j=1}^{J} s_j \, \mathbb{P}\{x = s_j\} \tag{9}$$

Пример. Давайте применим это определение к простейшему возможному случаю, в котором случайная величина x удовлетворяет $x(\omega) = \alpha$ для всех $\omega \in \Omega$, где α — некоторое постоянное скалярное значение. В этом случае сумма в (9) имеет только одно слагаемое и

$$\mathbb{E}x = \alpha\mathbb{P}\{x = \alpha\} = \alpha\mathbb{P}\{\omega \in \Omega : x(\omega) = \alpha\} = \alpha\mathbb{P}(\Omega) = \alpha$$

Пример. Чтобы оценить ожидание двоичной случайной величины x, мы применим (9), получаем

$$\mathbb{E}\,x = 1 \times \mathbb{P}\{x = 1\} + 0 \times \mathbb{P}\{x = 0\} = \mathbb{P}\{x = 1\}$$

Пример. Рассмотрим N подбрасываний честной монеты

Пространство элементарных событий — это $\Omega := \{0,1\}^N$, события такие, что $\mathscr{F} :=$ все подмножества Ω, и $\mathbb{P}(A) := 2^{-N}|A|$ для всех $A \in \mathscr{F}$

Пусть $x(\omega) = x(b_1, \ldots, b_N) = \sum_{n=1}^{N} b_n$

Прежде всего заметьте, что $0 \leq x \leq N$

По определению $\mathbb{P}$, для любого k мы имеем
$\mathbb{P}\{x = k\} = 2^{-N}|A_k|$, где

$$A_k := \{x = k\} = \left\{(b_1, \dots, b_N) \in \Omega \; : \; \sum_{n=1}^{N} b_n = k\right\}$$

Из комбинаторики, $|A_k| = \binom{N}{k}$, где правая сторона называется **биномиальным коэффициентом** для N, k, что удовлетворяет $\sum_{k=0}^{N} k\binom{N}{k} = N2^{N-1}$ для всех N

Ожидания x:

$$\mathbb{E}\, x = \sum_{k=0}^{N} k\, 2^{-N} |A_k| = 2^{-N} \sum_{k=0}^{N} k \binom{N}{k} = \frac{N}{2}$$

Для обычного x, приблизительная произвольная случайная величина с конечными случайными величинами:

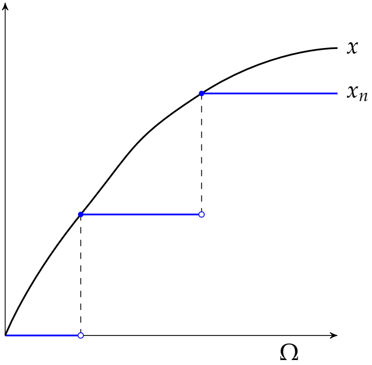


Рис.: Конечное приближение к общей случайной величине

Мы можем улучшать приближение без ограничений – Пусть x_n принимает большее и большее количество различных значений

Процесс дает последовательность *конечных* случайных величин x_n, сходящихся к x

Определим ожидания x как

$$\mathbb{E}\, x := \lim_{n\to\infty} \mathbb{E}\, x_n$$

$\mathbb{E}x$ также упоминается как **Интеграл Лебега** с.в. x относительно $\mathbb{P}$, с альтернативными обозначениями $\mathbb{E}x = \int x(\omega)\mathbb{P}(\mathrm{d}\omega)$

Существует ли последовательность аппроксимации случайной величины? Да, смотрите страницу 94 в ET и **dudley2002real**, утверждение 4.1.5

Если x принимает отрицательные значения, то пишут
$x = x^{+} - x^{-}$

где $x^{+} := \max\{x, 0\}$ и $x^{-} = \min\{x, 0\}$

Определим ожидания как

$$\mathbb{E}x := \mathbb{E}x^{+} - \mathbb{E}x^{-}$$

Сфокусируемся на **интегрируемых** случайных величинах: всех случайных величинах x, таких что $\mathbb{E}|x| < \infty$

- мы имеем $x^{+} \leq |x|$ и $x^{-} \leq |x|$
- таким образом, $\mathbb{E}x := \mathbb{E}x^{+} - \mathbb{E}x^{-}$ хорошо определены (почему?)

Свойства ожидания

Факт. (??) Возьмем любое вероятностное пространство $(\Omega, \mathscr{F}, \mathbb{P})$, в нем существует однозначно определенная функция $\mathbb{E}$, что отображает каждую интегрируемую случайную величину x на $(\Omega, \mathscr{F}, \mathbb{P})$ в качестве значения

$$\mathbb{E}x = \int x(\omega)\mathbb{P}(d\omega) \tag{10}$$

в $\mathbb{R}$, называемой **математическим ожиданием** x. Функция имеет следующие свойства:

1. $\mathbb{E}\alpha = \alpha$ для всех $\alpha \in \mathbb{R}$
2. $\mathbb{E}\mathbb{1}_A = \mathbb{P}(A)$ для всех $A \in \mathscr{F}$
3. $x \leq y \implies \mathbb{E}x \leq \mathbb{E}y$
4. $\mathbb{E}[\alpha x + \beta y] = \alpha\mathbb{E}x + \beta\mathbb{E}y$ для всех интегрируемых x, y и констант α, β

Чтобы напомнить себе об основной вероятностной мере $\mathbb{P}$, мы можем написать $\mathbb{E}_{\mathbb{P}} x$ вместо $\mathbb{E}\, x$

Обратите внимание на выражение $\mathbb{E}\, \alpha$ понимаемое как ожидание постоянной случайной величины, равной α

- следует из 4. при $x = \mathbb{1}_{\Omega}$ и $\beta = 0$

Упражнение: проверьте 3. для $x(\omega) := \mathbb{1}\{\omega \in A\}$ и $y(\omega) := \mathbb{1}\{\omega \in B\}$

Подсказка: что означает $x \leq y$ относительно A и B?

Для дальнейших подробности и ссылок на доказательства вышеуказанного факта, смотрите страницу 96 в ET

Мы сейчас докажем, что если x — конечная случайная величина с диапазоном $\{s_j\}_{j=1}^{J}$ и h — любая $\mathscr{B}$-измеримая функция, то

$$\mathbb{E}h(x) = \sum_{j=1}^{J} h(s_j)\mathbb{P}\{x = s_j\} \tag{11}$$

Сначала заметьте, что $\sum_{j=1}^{J} \mathbb{1}\{x = s_j\} = 1$, и значит мы можем записать $h(x)$ как

$$h(x) = h(x)\sum_{j=1}^{J} \mathbb{1}\{x = s_j\} = \sum_{j=1}^{J} h(s_j)\mathbb{1}\{x = s_j\}$$

Используем линейность ожиданий:

$$\mathbb{E}h(x) = \sum_{j=1}^{J} h(s_j)\mathbb{E}\,\mathbb{1}\{x = s_j\}$$

Применение части 2. факта **??** приводит к (11)

Неравенство Чебышёва: **Факт.** (??) Для любой неотрицательной случайной величины x и любой $\delta > 0$, мы имеем

$$\mathbb{P}\{x \geq \delta\} \leq \frac{\mathbb{E}\,x}{\delta} \tag{12}$$

Распространенная разновидность неравенства Чебышева имеет вид

$$\mathbb{P}\{|x| \geq \delta\} \leq \frac{\mathbb{E}\,x^2}{\delta^2} \tag{13}$$

Смотрите упражнение **??** для доказательства

Моменты

Пусть x — случайная величина и $k \in \mathbb{N}$. Если x^k интегрируема, то

- $\mathbb{E}\left[x^k\right]$ называется **k-ым моментом** x
- $\mathbb{E}\left[(x - \mathbb{E}\,x)^k\right]$ называется **k-ым центральным моментом** x

Если $\mathbb{E}\left[|x|^k\right] = \infty$, то говорят, что k-ый момент не существует. Для некоторых случайных величин даже первый момент не существует

В ином случае, каждый момент существует

Факт. (??) Если k-ый момент x существует, то также существует и j-ый для всех $j \leq k$

Доказательство: Упражнение **??**

Неравенство Коши — Буняковского для случайных величин:

Факт. (??) Если x и y — случайные величины с конечным вторым моментом, то

$$|\,\mathbb{E}\,[xy]\,| \leq \sqrt{\mathbb{E}\,[x^2]\mathbb{E}\,[y^2]} \tag{14}$$

Вторым центральным моментом x называется **дисперсия** x:

$$\operatorname{var} x := \mathbb{E}\left[(x - \mathbb{E}\, x)^2\right]$$

Стандартное отклонение x:

$$\sigma_x := \sqrt{\operatorname{var} x}$$

Ковариация случайной величины x и y:

$$\operatorname{cov}[x, y] := \mathbb{E}\left[(x - \mathbb{E}\, x)(y - \mathbb{E}\, y)\right]$$

Факт. (??) Если x и y имеют конечные вторые моменты, то

1. $\operatorname{var} x$ и $\operatorname{cov}[x,y]$ конечны
2. $\operatorname{var} x = \mathbb{E}[x^2] - [\mathbb{E}\, x]^2$, и
3. $\operatorname{cov}[x,y] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$

Часть 1. следует из 2.–3., неравенства Коши — Буняковского и факта ??

Части 2.–3. следуют из линейности $\mathbb{E}$ и нескольких простых манипуляций

Факт. (??) Если $x_1, \dots, x_N$ — случайные величины и $\alpha_0, \alpha_1, \dots, \alpha_N$ — постоянные скаляры, то

$$\operatorname{var}\left[\alpha_0 + \sum_{n=1}^{N} \alpha_n x_n\right] = \sum_{n=1}^{N} \alpha_n^2 \operatorname{var}[x_n] + 2 \sum_{n<m} \alpha_n \alpha_m \operatorname{cov}[x_n, x_m]$$

Некоторые простые выводы:

1. $\operatorname{var}[\alpha + \beta x] = \beta^2 \operatorname{var}[x]$ и
2. $\operatorname{var}[\alpha x + \beta y] = \alpha^2 \operatorname{var}[x] + \beta^2 \operatorname{var}[y] + 2\alpha\beta \operatorname{cov}[x, y]$.

Корреляция x и y:

$$\operatorname{corr}[x,y] := \frac{\operatorname{cov}[x,y]}{\sigma_x \sigma_y}$$

Если $\operatorname{corr}[x,y] = 0$, то x и y **некоррелированы**

Положительная корреляция означает, что $\operatorname{corr}[x,y]$ положительна, и отрицательная корреляция значит, что $\operatorname{corr}[x,y]$ отрицательна

Факт. (??) Возьмем любые две случайные величины x, y и положительные константы α, β, мы имеем

$$-1 \leq \operatorname{corr}[x,y] \leq 1 \quad \text{и} \quad \operatorname{corr}[\alpha x, \beta y] = \operatorname{corr}[x,y]$$

первая часть следует из факта ??; вторая — из алгебры

Лучшие линейные предсказатели

Рассмотрим задачу предсказания значения случайной величины y, учитывая знание значения второй случайной величины x

Мы ищем функцию f, такую что $f(x)$ в среднем близка к y

Для измерения последнего мы будем использовать **среднеквадратичную ошибку**, которая в данном случае составляет

$$\mathbb{E}\left[(y - f(x))^2\right]$$

В §**??**, чтобы получить минимизатор среднеквадратичного отклонения по всем функциям x, мы выбираем

$$f(x) = \mathbb{E}[y \mid x]$$

Здесь мы рассмотрим поиск хорошего предсказателя y среди класса "линейных" функций

$$\mathcal{H}_\ell := \{ \text{ все функции вида } \ell(x) = \alpha + \beta x\}$$

Рассмотрим:

$$\min_{\ell \in \mathcal{H}_\ell} \mathbb{E}\left[(y - \ell(x))^2\right] = \min_{\alpha, \beta \in \mathbb{R}} \mathbb{E}\left[(y - \alpha - \beta x)^2\right] \tag{15}$$

Если α и β решают (15), то функция

$$\ell^*(x) := \alpha^* + \beta^* x \tag{16}$$

называется **лучшим линейным предсказателем** y при данном x

Пример. (??) Взаимосвязь между доходностью данного актива R_a и рыночной доходностью R_m называется коэффициентом **бета**

Измеряет подверженность системному риску в отличие от нерыночного риска.

Бета R_a часто определяется как коэффициент β^* в лучшем линейном предсказании (16), когда x — это рыночная доходность и $y = R_a$

Чтобы решить (15), раскроем квадрат с правой стороны и используем линейность $\mathbb{E}$, чтобы написать целевую функцию как

$$\psi(\alpha,\beta) := \mathbb{E}\left[y^2\right] - 2\alpha\mathbb{E}\left[y\right] - 2\beta\mathbb{E}\left[xy\right] + 2\alpha\beta\mathbb{E}\left[x\right] + \alpha^2 + \beta^2\mathbb{E}\left[x^2\right]$$

Вычисление производных и решение условий первого порядка:

$$\beta^* := \frac{\operatorname{cov}[x,y]}{\operatorname{var}[x]} \quad \text{and} \quad \alpha^* := \mathbb{E}\left[y\right] - \beta^*\mathbb{E}\left[x\right] \tag{17}$$

Смотрите упр. **??**

Распределения

Возьмем случайную величину x на вероятностное пространство $(\Omega, \mathscr{F}, \mathbb{P})$

Вероятность, x примет значение из Борелевского множества B

$$\mathbb{P}\{x \in B\}$$

На практике, удобнее представлять вероятность как *распределение* в $\mathbb{R}$

Специфицируем Ω в $\mathbb{R}$ и возьмем множество событий в $\mathbb{R}$ как $\mathscr{B}(\mathbb{R})$

- Вероятностная мера, определенная в $\mathscr{B}(\mathbb{R})$, называется **законом** или **распределением**

Формально, распределение P — отображение из $\mathscr{B}(\mathbb{R})$ в $[0,1]$, такое что

1. $P(\mathbb{R}) = 1$ и
2. $P(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n)$ для любой несовместной последовательности $\{B_n\}$

Если существует Борелевское множество S с $P(S) = 1$, то мы говорим, что S является **носителем распределения** P

Распределения характеризуются **функцией распределения**, или функция распределения, которой может быть любая функция $F\colon \mathbb{R} \to [0, 1]$, удовлетворяющая следующим условиям:

1. монотонность: $s \leq s'$ подразумевает, что $F(s) \leq F(s')$,
2. непрерывность справа: $F(s_n) \downarrow F(s)$ всюду, где $s_n \downarrow s$, и
3. $\lim_{s \to -\infty} F(s) = 0$ и $\lim_{s \to \infty} F(s) = 1$.

Функция распределения и распределение в $\mathbb{R}$ можно поставить во взаимно однозначное соответствие

Распределение P полностью характеризуется значениями функции

$$F(s) := P((-\infty, s]) \qquad (s \in \mathbb{R}) \tag{18}$$

Факт. (??) Следующие утверждения верны:

1. Если P — любое распределение в $\mathbb{R}$, то функция F в (18) — функция распределения.
2. Возьмем любую функция распределения F в $\mathbb{R}$, существует ровно одно распределение P, удовлетворяющее (18)

Для полного доказательства, смотрите **williams1991probability**, лемма 1.6, или **dudley2002real**, теорема 9.1.1.

Здесь ограничимся демонтрацией, что функция F в (18) удовлетворяет части 1. определения функция распределения

- заметим, что $s \leq s'$ подразумевает, что $(-\infty, s] \subset (-\infty, s']$
- вспомним, что $P(A) \leq P(B)$, если $A \subset B$
- тогда мы имеем $P((-\infty, s]) \leq P((-\infty, s'])$ и $F(s) \leq F(s')$, как и было заявлено

Пример. **Одномерные нормальные распределения** или **распределения Гаусса** относятся к распределениям классов, обозначенным как функция распределения вида

$$F(s) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{s} \exp\left\{-\frac{(t-\mu)^2}{2\sigma^2}\right\} \mathrm{d}t \qquad (s \in \mathbb{R})$$

где $\mu \in \mathbb{R}$ и $\sigma > 0$

Мы представляем распределение, связанное с (μ, σ) с помощью $\mathrm{N}(\mu, \sigma^2)$

Распределение $\mathrm{N}(0, 1)$ называется **стандартным нормальным распределением**

Мы используем символ Φ для его функции распределения

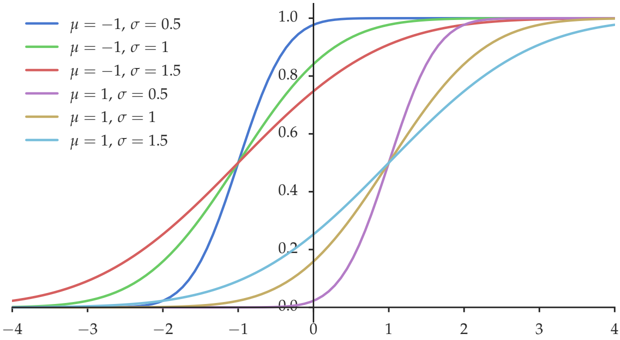


Рис.: Функция распределения для нормальных распределений

Пример. **Распределение Парето** — одномерное распределение с функция распределения вида

$$F(s) = \begin{cases} 0 & \text{, если } s < s_0 \\ 1 - \left(\frac{s_0}{s}\right)^{\alpha} & \text{, если } s_0 \le s \end{cases} \qquad (s \in \mathbb{R},\ s_0,\ \alpha > 0)$$

Распределения Парето часто используются для моделирования явлений с тяжелым правым хвостом, таких как распределение богатства или дохода.

Пример. Класс **функции распределения бета** дан с помощью

$$F(s) = \begin{cases} 0 & \text{, если } s \leq 0 \\ \frac{1}{B(\alpha,\beta)} \int_0^s u^{\alpha-1}(1-u)^{\beta-1}\,\mathrm{d}u & \text{, если } 0 < s < 1 \\ 1 & \text{, если } 1 \leq s \end{cases}$$

где $\alpha, \beta > 0$.

В этом примере $B(\alpha,\beta)$ — **функция бета**

$$B(\alpha,\beta) := \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad \text{where} \qquad \Gamma(a) := \int_0^\infty u^{a-1}e^{-u}\,\mathrm{d}u$$

Функция Γ называется **функция гамма**.

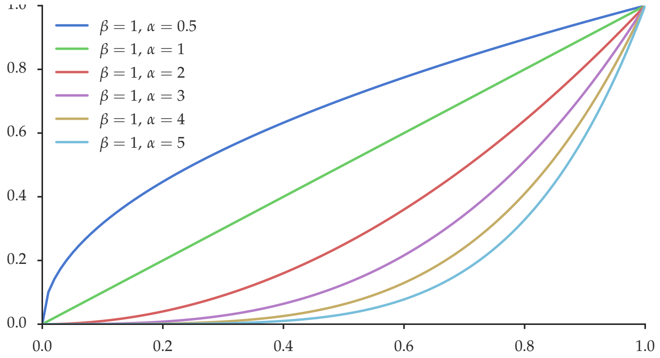


Рис.: Функции распределения бета

Пример. класс **распределений Коши** определяется как

$$F(s) = \frac{1}{\pi} \arctan\left(\frac{s-\tau}{\gamma}\right) + \frac{1}{2} \qquad (s \in \mathbb{R})$$

параметры $\tau \in \mathbb{R}$ и $\gamma > 0$ — параметры местоположения и масштаба соответственно

Если $\tau = 0$ и $\gamma = 1$, то F называется **стандартным распределением Коши**

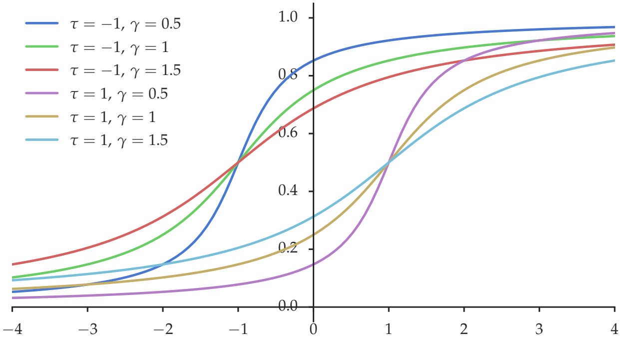


Рис.: Функции распределения Коши

Пример. Возьмем $a < b$, **равномерная функция распределения** на промежутке $[a, b]$ — это

$$F(s) = \begin{cases} 0 & \text{, если } s \leq a \\ \frac{s-a}{b-a} & \text{, если } a < s < b \\ 1 & \text{, если } b \leq s \end{cases}$$

мы обозначаем это распределение как $U[a, b]$

Плотности и функции вероятности

Два удобных частных случая

- дискретный: функция распределения просто скачет (ступенчатая функция)
- абсолютно непрерывный случай: функция непрерывна, без скачков

Дискретный случай

Распределение P называется **дискретным**, если оно имеет носитель распределения в счетном множестве; that is, если существует счетное множество $\{s_j\}_{j\geq 1}$ с $P(\{s_j\}_{j\geq 1}) = 1$

для такого P пусть

$$p_j := P\{s_j\} := P(\{s_j\}) = \text{вероятность в одной точке } s_j$$

Функция вероятности — любая неотрицательная последовательность (конечная или бесконечная), сумма которой равна единице.

Упражнение: покажите, что $\{p_j\}_{j\geq 1}$ является **функцией вероятности**

Мы можем показать связь функции распределения с P как:

$$F(s) = \sum_{j \geq 1} \mathbb{1}\{s_j \leq s\} p_j \tag{19}$$

так как

$$F_x(s) := \mathbb{P}\{x \leq s\} = \mathbb{P} \bigcup_{j \text{ s.t. } s_j \leq s} \{x = s_j\}$$

$$= \sum_{j \text{ s.t. } s_j \leq s} \mathbb{P}\{x = s_j\} = \sum_{j=1}^{J} \mathbb{1}\{s_j \leq s\} p_j$$

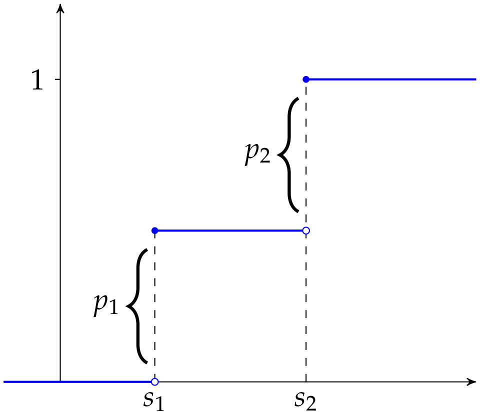


Рис.: Дискретная функция распределения

Пример. Возьмем $N \in \mathbb{N}$ и $\pi \in (0,1)$, последовательность $\{p_0, \dots, p_N\}$, определяемая как

$$p_j = \binom{N}{j} \pi^j (1-\pi)^{N-j}$$

называется **биномиальной функцией вероятности**

Значение p_j — вероятность, j успехов в N независимых испытаниях, вероятность успеха каждого случая равна π

Абсолютно непрерывный случай

Функция плотности — неотрицательная функция p в $\mathbb{R}$, которая интегрируется в 1

Распределение P **определяется функцией плотности** p, если p — функция плотности и

$$P(B) = \int_B p(s)\,\mathrm{d}s \qquad \text{для всех } B \in \mathscr{B}(\mathbb{R})$$

Заметим, что

$$\int_B p(s)\,\mathrm{d}s := \int_{-\infty}^{\infty} \mathbb{1}_B(s) p(s)\,\mathrm{d}s$$

Точное необходимое и достаточное условие существования функции плотности — абсолютная непрерывность

Распределение P в Борелевских подмножествах $\mathbb{R}$ называется **абсолютно непрерывным**, если $P(B) = 0$ всюду, где мера Лебега B равна нулю (смотрите §**??**)

- Любое счетное подмножество $\mathbb{R}$ имеет меру Лебега равную нулю

Факт. (**??**) Если P абсолютно непрерывная, то $P(C) = 0$ всюду, где C счетно

Если распределение абсолютно непрерывно:

- вероятность в каждой точке равна нулю
- соответствующая функция распределения не содержит скачков
- теорема Ньютона — Лейбница говорит, что $F(s)$ дифференцируема во всех точках непрерывности p, и:

$$F'(s) = p(s) \qquad \text{для всех } s \in \mathbb{R}, \text{ таких что } p \text{ непрерывна в } s$$

Пример. Нормальные функции распределения дифференцируемы для всех μ, σ, с функцией плотности

$$p(s) = F'(s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(s-\mu)^2}{2\sigma^2}\right\}$$

Мы используем символ ϕ для стандартного нормального распределения

Рис.: Функции плотности нормального распределения

Пример. Функция распределения Коши имеет функцию плотности

$$p(s) = \frac{1}{\pi\gamma}\left[1 + \left(\frac{s-\tau}{\gamma}\right)^2\right]^{-1} \qquad (s \in \mathbb{R},\ \gamma > 0,\ \tau \in \mathbb{R})$$

Функции плотности Коши более остроконечны около своих мод и имеют большую массу в хвосте, чем нормальная функция плотности

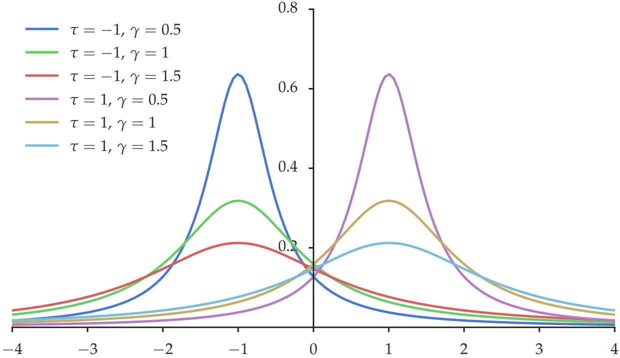


Рис.: Функции плотности Коши

Пример. Бета имеет функцию плотности, определяемую как

$$p(s) = \frac{s^{\alpha-1}(1-s)^{\beta-1}}{B(\alpha,\beta)} \qquad (\alpha,\beta > 0)$$

Когда $0 < s < 1$ и 0 в ином случае

Пример. $U[a, b]$ распределение представлено функцией плотности

$$p(s) = \frac{1}{b-a}\mathbb{1}\{a \leq s \leq b\} \qquad (s \in \mathbb{R},\ a, b \in \mathbb{R},\ a < b)$$

Пример. **Гамма-распределение** с параметром формы α и параметром масштаба β — распределение с функцей плотности

$$p(s) = \frac{s^{\alpha-1}e^{-s/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \qquad (\alpha, \beta > 0)$$

Когда $0 < s < 1$ и 0 в ином случае

Пример. **Хи-квадрат распределение с k степенями свободы** — распределение с функцией плотности

$$p(s) := \frac{1}{2^{k/2}\Gamma(k/2)} s^{k/2-1} e^{-s/2} \qquad (s > 0,\ k \in \mathbb{N})$$

Это распределение представлено символом $\chi^2(k)$

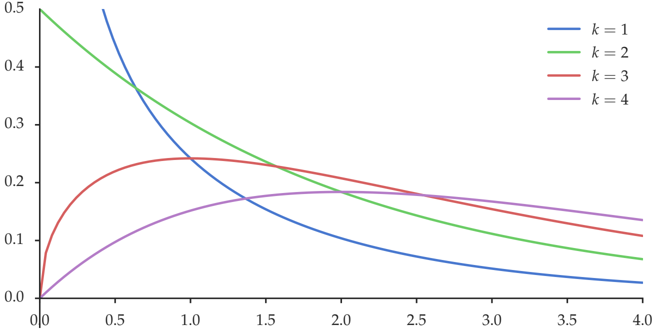


Рис.: Функции плотности хи-квадрата

Пример. **Распределение Стьюдента t с k степенями свободы**, или, проще, t-распределение с k степенями свободы, — распределение в $\mathbb{R}$ с функцией плотности

$$p(s) := \frac{\Gamma(\frac{k+1}{2})}{(k\pi)^{1/2}\Gamma(\frac{k}{2})}\left(1 + \frac{s^2}{k}\right)^{-(k+1)/2} \qquad (s \in \mathbb{R},\ k > 0)$$

Пример. F-**распределение** с параметрами k_1, k_2 — распределение с функцией плотности

$$p(s) := \frac{\sqrt{(k_1 s)^{k_1} k_2^{k_2} / [k_1 s + k_2]^{k_1 + k_2}}}{s B(k_1/2, k_2/2)} \qquad (s \geq 0,\ k_1,\ k_2 > 0)$$

F-распределение возникает при проверке ряда гипотез, как обсуждается ниже.

Интегрирование

Рассмотрим обычный интергал $\int_a^b h(s)\,\mathrm{d}s$ функции h на некотором интервале $[a,b]$

Предположим, мы хотим сделать этот интергал взвешенным, придав больше массы различным частям $[a,b]$:

$$\int_a^b h(s)p(s)\,\mathrm{d}s$$

Например:

- h — функция благосостояния и p плотность агентов
- p — плотность, указывающая на вероятности исходов, h функция прибыли

Предположим, что P не имеет функции плотности, но мы все еще хотим взвесить интеграл с помощью P

- мы хотим определить $\int h(s)P(\mathrm{d}s)$

Возьмем распределение P в $\mathbb{R}$ и рассмотрим $(\mathbb{R}, \mathscr{B}(\mathbb{R}), P)$ как вероятностное пространство

$(\mathbb{R}, \mathscr{B}(\mathbb{R}), P)$ имеет собственный оператор ожидания $\mathbb{E}_P$

Предположим, что h — случайная величина в $(\mathbb{R}, \mathscr{B}(\mathbb{R}), P)$, тогда

$$\mathbb{E}_P h =: \int h(s)P(\mathrm{d}s) := \text{ожидания } h \text{ при } P$$

Факт. Пусть $h\colon \mathbb{R} \to \mathbb{R}$ — $\mathscr{B}$-измерима и P — распределение в $\mathbb{R}$

Если P дискретна, с функцией вероятности $\{p_j\}_{j\geq 1}$ и носителем распределения $\{s_j\}_{j\geq 1}$, то

$$\int h(s)P(\mathrm{d}s) = \sum_{j\geq 1} h(s_j)p_j$$

Если P абсолютно непрерывно с функцией плотности p, то

$$\int h(s)P(\mathrm{d}s) = \int_{-\infty}^{\infty} h(s)p(s)\,\mathrm{d}s$$

Распределения и случайные переменные

Каждая случайная переменная определяет распределение on $\mathbb{R}$

Пусть x — случайная величина на некотором вероятностном пространстве $(\Omega, \mathscr{F}, \mathbb{P})$:

- вероятность $\mathbb{P}\{x \in B\}$ определена для каждого $B \in \mathscr{B}(\mathbb{R})$ (смотрите (6))
- множество функций P определено как

$$P(B) = \mathbb{P}\{x \in B\} \qquad (B \in \mathscr{B}(\mathbb{R})) \tag{20}$$

является **распределением** x

Функция распределения соответствующая распределению P случайной переменной x удовлетворяет

$$F(s) = \mathbb{P}\{x \leq s\} \qquad (s \in \mathbb{R}) \tag{21}$$

мы пишем $\mathcal{L}(x) = F$, чтобы подчеркнуть, что F означает распределение x

Факт. (??) Если $\mathcal{L}(x) = F$, то $\mathbb{P}\{a < x \leq b\} = F(b) - F(a)$ для любых $a \leq b$

Доказательство выполните в качестве упражнения (или смотрите страницу 111 в ET)

Для каждой функции распределения F, существует вероятностное пространство $(\Omega, \mathscr{F}, \mathbb{P})$ и случайная величина $x\colon \Omega \to \mathbb{R}$, такая что $\mathcal{L}(x) = F$; §**??** показывает построение

Если $\mathcal{L}(x) = P$ и P имеет функцию плотности p, мы говорим, что x **имеет функцию плотности** p

Если распределение x дискретно, мы будем называть x **дискретной случайной величиной**

Факт. Если x имеет функцию плотности, то $\mathbb{P}\{x = s\} = 0$ для всех $s \in \mathbb{R}$, и для любых $a < b$,

$$\mathbb{P}\{a < x < b\} = \mathbb{P}\{a < x \leq b\}$$
$$= \mathbb{P}\{a \leq x < b\} = \mathbb{P}\{a \leq x \leq b\}$$

Распределения преобразований

Факт. (??) Если $\mathcal{L}(x) = F$ и $y := \psi(x)$, где $\psi\colon \mathbb{R} \to \mathbb{R}$ строго возрастает, то $\mathcal{L}(y) = G$, где $G(s) := F(\psi^{-1}(s))$.

Доказательство. При таких гипотезах ψ^{-1} существует и (медленно) возрастает. Следовательно

$$\mathbb{P}\{y \leq s\} = \mathbb{P}\{\psi(x) \leq s\} = \mathbb{P}\{x \leq \psi^{-1}(s)\} = F(\psi^{-1}(s))$$

Заметьте, как монотонность используется во втором равенстве

Пример. Если $\mathcal{L}(x) = F$ и $y := \exp(x)$, то функция распределения y — это $G(s) := F(\ln(s))$

Факт. (??) Если x имеет плотность p в $\mathbb{R}$ и $y := \psi(x)$, где ψ — диффеоморфизм в $\mathbb{R}$, то распределение y абсолютно непрерывное с функцией плотности

$$q(s) = p(\psi^{-1}(s)) \left| \frac{\mathrm{d}\psi^{-1}(s)}{\mathrm{d}s} \right| \qquad (s \in \mathbb{R})$$

термин **диффеоморфизм** значит, что ψ — биекция в $\mathbb{R}$ и оба ψ и его обратное дифференцируемы

Пример. Если x имеет функцию плотности p в $\mathbb{R}$, и μ и σ — константы с $\sigma > 0$, то функция плотности $y := \mu + \sigma x$ — это

$$q(s) = p\left(\frac{s-\mu}{\sigma}\right)\frac{1}{\sigma} \qquad (s \in \mathbb{R})$$

Когда x стандартное нормальное: $y = \mu + \sigma x$ is $\mathrm{N}(\mu, \sigma^2)$

Почему?

- Возьмем p функцию плотности стандартного нормального распределения ϕ
- Вспомним

$$p(s) = F'(s) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(s-\mu)^2}{2\sigma^2}\right\}$$

Пусть x — случайная величина в вероятностном пространстве $(\Omega, \mathscr{F}, \mathbb{P})$

Распределение x кодирует всю информацию для расчета ожидания x или любого $\mathscr{B}$-измеримого преобразования $h(x)$

Во-первых, пусть x конечно. Предположим, что

- $\mathcal{L}(x) = P$
- функция $h\colon \mathbb{R} \to \mathbb{R}$ — любая $\mathscr{B}$-измеримая функция
- P помещает все вероятности в конечное множество $\{s_j\}_{j=1}^J$

Исспользуем $\mathbb{P}\{x = s_j\} = P\{s_j\}$ и определение ожиданий:

$$\mathbb{E}h(x) = \sum_{j=1}^J h(s_j)\mathbb{P}\{x = s_j\} = \sum_{j=1}^J h(s_j)P\{s_j\} = \sum_{j=1}^J h(s_j)p_j$$

Ожидания $h(x)$ в $(\Omega, \mathscr{F}, \mathbb{P})$ равны ожиданиям h в $(\mathbb{R}, \mathscr{B}(\mathbb{R}), P)$

Верно также и для бесконечного случая:

Факт. Пусть x — случайная величина в некотором вероятностном пространстве $(\Omega, \mathscr{F}, \mathbb{P})$, пусть $\mathcal{L}(x) = P$ и h — $\mathscr{B}$-измеримая функция, такая что $h(x)$ интегрируема. Ожидания $\mathbb{E}h(x)$ полностью определены h и P. В частности,

$$\mathbb{E}h(x) = \int h(s)P(\mathrm{d}s)$$

где $\int h(s)P(\mathrm{d}s)$ — ожидания h в $(\mathbb{R}, \mathscr{B}(\mathbb{R}), P)$

Пример. Пусть x — случайная величина, чье распределение P является равномерным распределением в $[a, b]$

Применить определение функции плотности равномерного распределения

$$\mathbb{E}\,x = \int sP(\mathrm{d}s) = \int sp(s)\,\mathrm{d}s = \int_{-\infty}^{\infty} \frac{s}{b-a}\mathbb{1}\{a \le s \le b\}\,\mathrm{d}s$$

Решение интеграла дает $\mathbb{E}\,x = \mu := (a+b)/2$. Дисперсия равна

$$\mathrm{var}[x] = \int (s-\mu)^2 P(\mathrm{d}s)$$

$$= \int_a^b \left(s - \frac{a+b}{2}\right)^2 \frac{1}{b-a}\,\mathrm{d}s = \frac{1}{12}(b-a)^2$$

Пример. Предположим, что $\mathcal{L}(x) = \mathrm{N}(\mu, \sigma)$

Если $\sigma > 0$, среднее значение может быть вычислено с помощью

$$\mathbb{E}\, x = \int_{-\infty}^{\infty} s \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(s-\mu)^2}{2\sigma^2}\right\} \mathrm{d}s = \mu$$

Дисперсия определяется как:

$$\mathrm{var}[x] = \int_{-\infty}^{\infty} (s-\mu)^2 \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(s-\mu)^2}{2\sigma^2}\right\} \mathrm{d}s = \sigma^2$$

Моменты распределений

Любые две случайные величины с одинаковыми распределениями имеют одинаковые моменты

Следовательно моменты лучше всего рассматривать как свойство распределения, не как случайную величину

Таким образом, мы определяем

- **среднее** P как $\mu = \int sP(\mathrm{d}s)$,
- k**-ый момент** P как $\int s^k P(\mathrm{d}s)$,
- **дисперсию** P как $\int (s-\mu)^2 P(\mathrm{d}s)$,

и так далее

Функция квантилей

Пусть F — строго возрастающая функция распределения в $\mathbb{R}$

Возьмем $\tau \in (0, 1)$, **τ-ая квантиль** F — это $\xi \in \mathbb{R}$, который является решением $F(\xi) = \tau$

Согласно нашим предположениям о F, такое ξ существует и однозначно определено

0.5-ая квантиль называется **медианой** F

Функция квантилей:

$$F^{-1}(\tau) := \text{ единственный } \xi, \text{ такой что } F(\xi) = \tau \qquad (0 < \tau < 1)$$

Пример. Функция квантилей связанная со стандартным распределением Коши — это $F^{-1}(\tau) = \tan[\pi(\tau - 1/2)]$

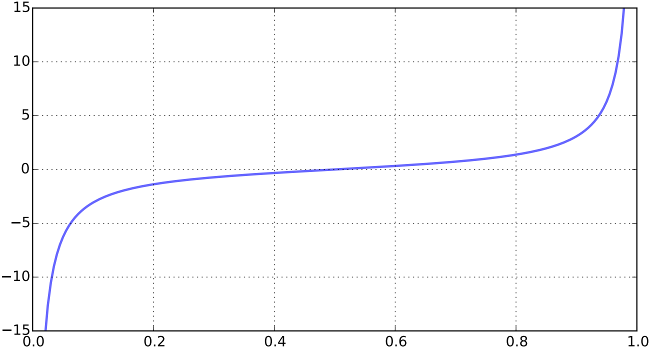


Рис.: Функция квантилей распределения Коши (горизонтальная ось — это $\tau \in (0,1)$)

Когда F не строго возрастающая, F^{-1} не определено

Мы можем задать:

$$F^{-1}(\tau) := \inf\{s \in \mathbb{R} : F(s) \geq \tau\} \qquad (0 < \tau < 1) \tag{22}$$

Функция плотности p симметрична, если $p(s) = p(-s)$ для всех $s \in \mathbb{R}$

Распространенный сценарий проверки гипотез

Факт. (**??**) Пусть x — случайная величина с функцией плотности p. Если p симметрична, то функция распределения G $y := |x|$ определяется как

$$G(s) := \mathbb{P}\{y \leq s\} = \begin{cases} 2F(s) - 1 & \text{, если } s \geq 0 \\ 0 & \text{в ином случае} \end{cases}$$

Докажите в качестве упражнения **??**

Факт эквивалентен $F(s) = 1 - F(-s)$

Возьмем случайную величину x с $\mathcal{L}(x) = F$ и заданной константой $\alpha \in (0, 1)$

Рассмотрим c, являющийся решением $\mathbb{P}\{-c \leq x \leq c\} = 1 - \alpha$

Факт. Если $\mathcal{L}(x) = F$, x имеет симметричную функцию плотности и F — строго возрастающая, то

$$c = F^{-1}(1 - \alpha/2) \quad \Longrightarrow \quad \mathbb{P}\{-c \leq x \leq c\} = 1 - \alpha \tag{23}$$

Когда F стандартная нормальная функция распределения Φ, c обычно обозначается как $z_{\alpha/2}$:

$$z_{\alpha/2} := \Phi^{-1}(1 - \alpha/2) \tag{24}$$

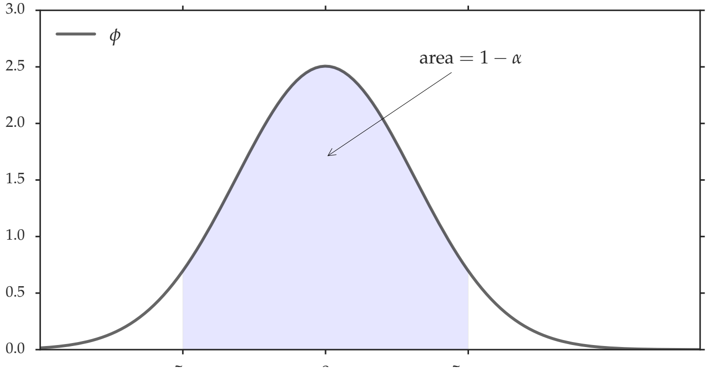


Рис.: Критические значения для стандартной нормальной плотности

Учебник по Эконометрике

Лекция 4: Моделирование зависимости

Джон Стачурски

Лекции: Акшай Шенкер
Перевел: Алексей Кедо

1 октября 2020 г.

Случайный вектор

Случайный вектор $\mathbf{x}$ в $\mathbb{R}^N$ — это функция из Ω в $\mathbb{R}^N$ со следующим свойством

$$\{\omega \in \Omega : \mathbf{x}(\omega) \in B\} \in \mathscr{F} \quad \text{для всех } B \in \mathscr{B}(\mathbb{R}^N)$$

Мы можем также определить **случайный вектор** $\mathbf{x}$ в $\mathbb{R}^N$ как список из N случайных переменных $(x_1, \ldots, x_N)$

Пишем случайные векторы в строки или столбцы по удобству

- во время умножения матриц случайные векторы по умолчанию будут векторами-столбцами

Пример. Вспомните эксперимент с обезьяной с завязанными глазами

Пространство элементарных событий — это единичный диск $\Omega := \{(h, v) \in \mathbb{R}^2 : \|(h, v)\| \leq 1\}$ и пространство событий — это Борелевские множества в Ω

Если $\mathbf{x}$ тождественен Ω, то он просто сообщает результат (h, v) — случайный вектор

Пример. Рассмотрим случайную выборку с перечислением доходов y_n индивидов $n = 1, \ldots, N$

Вектор $(y_1, \ldots, y_N)$ который сообщает результат этой выборки, можно рассматривать как случайный вектор в $\mathbb{R}^N$

Измеримость

Определение случайного вектора гарантирует, что $\{\mathbf{x} \in B\}$ — определенное событие для каждого $B \in \mathscr{B}(\mathbb{R}^N)$

Чтобы убедиться, что $\mathbf{y} = f(\mathbf{x})$ — случайный вектор:

- функция $f\colon \mathbb{R}^N \to \mathbb{R}^M$ должна удовлетворять $f^{-1}(B) \in \mathscr{B}(\mathbb{R}^N)$ для всех $B \in \mathscr{B}(\mathbb{R}^M)$

Ожидания

Ожидания определяются поэлементно

Если $\mathbf{x} = (x_1, \ldots, x_N)$ — случайный вектор в $\mathbb{R}^N$, то

$$\mathbb{E}\,\mathbf{x} = \mathbb{E}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} := \begin{pmatrix} \mathbb{E}\,x_1 \\ \mathbb{E}\,x_2 \\ \vdots \\ \mathbb{E}\,x_N \end{pmatrix}$$

Random Matrix

Случайная матрица $\mathbf{X}$ размера $M \times N$ — массив случайных величин размера $M \times N$

Его ожидание определяется как

$$\mathbb{E}\,\mathbf{X} := \begin{pmatrix} \mathbb{E}\,x_{11} & \cdots & \mathbb{E}\,x_{1N} \\ \vdots & & \vdots \\ \mathbb{E}\,x_{M1} & \cdots & \mathbb{E}\,x_{MN} \end{pmatrix}$$

Из линейности ожиданий (факт ??):

Факт. (??) Если $\mathbf{X}$ и $\mathbf{Y}$ — случайные матрицы или векторы, и $\mathbf{A}$ и $\mathbf{B}$ постоянны и согласованны, то

$$\mathbb{E}\left[\mathbf{AX}+\mathbf{BY}\right]=\mathbf{A}\mathbb{E}\left[\mathbf{X}\right]+\mathbf{B}\mathbb{E}\left[\mathbf{Y}\right]$$

Ковариационная матрица

Ковариационная матрица случайного вектора $\mathbf{x}$ в $\mathbb{R}^N$ с $\boldsymbol{\mu} := \mathbb{E}\mathbf{x}$ является матрицей размера $N \times N$

$$\mathrm{var}[\mathbf{x}] := \mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^{\top}\right]$$

Расширяем:

$$\mathrm{var}[\mathbf{x}] = \begin{pmatrix} \mathbb{E}\left[(x_1-\mu_1)(x_1-\mu_1)\right] & \cdots & \mathbb{E}\left[(x_1-\mu_1)(x_N-\mu_N)\right] \\ \vdots & & \vdots \\ \mathbb{E}\left[(x_N-\mu_N)(x_1-\mu_1)\right] & \cdots & \mathbb{E}\left[(x_N-\mu_N)(x_N-\mu_N)\right] \end{pmatrix}$$

j,k-ый является скалярной ковариацией между x_j и x_k, и главная диагональ содержит дисперсию каждого x_n

Факт. Для любого случайного вектор $\mathbf{x}$ с $\mathbb{E}\left[\mathbf{x}^{\mathsf{T}}\mathbf{x}\right] < \infty$,

1. $\mathrm{var}[\mathbf{x}]$ существует и неотрицательно определена,
2. $\mathrm{var}[\mathbf{x}] = \mathbb{E}\left[\mathbf{x}\mathbf{x}^{\mathsf{T}}\right] - \boldsymbol{\mu}\boldsymbol{\mu}^{\mathsf{T}}$, и
3. $\mathrm{var}[\mathbf{A}\mathbf{x} + \mathbf{b}] = \mathbf{A}\,\mathrm{var}[\mathbf{x}]\mathbf{A}^{\mathsf{T}}$ (для любых постоянных и согласованных $\mathbf{A}, \mathbf{b}$).

Кросс-ковариация между случайными векторами $\mathbf{x}$ и $\mathbf{y}$ определяется как

$$\operatorname{cov}[\mathbf{x}, \mathbf{y}] := \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^\intercal\right]$$

Очевидно, $\operatorname{var}[\mathbf{x}] = \operatorname{cov}[\mathbf{x}, \mathbf{x}]$

Факт. (??) Если $\mathbf{z}$ — случайный вектор в $\mathbb{R}^N$, удовлетворяющий $\mathbb{E}[\mathbf{z}\mathbf{z}^\intercal] = \mathbf{I}$ и $\mathbf{A}$ любая постоянная матрица размера $N \times N$, то

$$\mathbb{E}\left[\mathbf{z}^\intercal \mathbf{A} \mathbf{z}\right] = \operatorname{trace} \mathbf{A}$$

Доказательство — решенное упражнение (смотрите упр. **??**)

Совместные распределения

Распределение или **закон** P в $\mathbb{R}^N$ — вероятностная мера Борелевских множеств $\mathscr{B}(\mathbb{R}^N)$

По определению, оно удовлетворяет $P(\mathbb{R}^N) = 1$ и $P(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} P(B_n)$ для любых непересекающихся последовательностей $\{B_n\}$ в $\mathscr{B}(\mathbb{R}^N)$

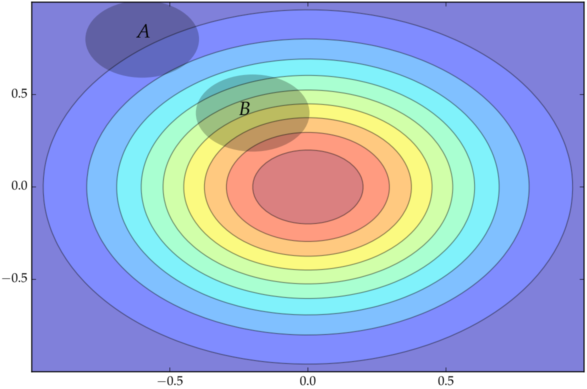


Рис.: Пример распределения и события A и B

Любое распределение P в $\mathbb{R}^N$ характеризуется функцией

$$F(\mathbf{s}) := F(s_1, \dots, s_N) := P\left(\times_{n=1}^N (-\infty, s_n]\right) \qquad (\mathbf{s} \in \mathbb{R}^N)$$

Функция F — **функция совместного распределения**, которая является функцией $F\colon \mathbb{R}^N \to [0,1]$ со следующими свойствами

1. непрерывна справа по каждому из своих аргументов,
2. возрастает по каждому из своих аргументов, и
3. удовлетворяет

$$F(\mathbf{s}_j) \to 1 \text{ при } \mathbf{s}_j \to \infty$$

$$\text{и} \quad F(s_1, \dots, s_{nj}, \dots, s_N) \to 0 \text{ при } s_{nj} \to -\infty$$

Распределение P в $\mathbb{R}^N$:

- **дискретно**, если P имеет носитель распределения в счетном подпространстве $\mathbb{R}^N$
- **абсолютно непрерывно**, если $P(B) = 0$ всюду, где B имеет меру Лебега равную нулю

Опять же, абсолютная непрерывность необходима и достаточна для существования функции плотности:

$$P(B) = \int_B p(\mathbf{s})\ \mathrm{d}\mathbf{s} \qquad \text{для всех } B \in \mathscr{B}(\mathbb{R}^N)$$

справа — многомерный интеграл, который мы можем записать как

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbb{1}_B(s_1, \ldots, s_N) p(s_1, \ldots, s_N)\, \mathrm{d}s_1 \cdots \mathrm{d}s_N$$

Если p — любая функция плотности в $\mathbb{R}^N$, то вышенаписанное определяет распределение

Пример. **Многомерное нормальное распределение** или **многомерное распределение Гаусса** в $\mathbb{R}^N$ — функция p вида

$$p(\mathbf{s}) = (2\pi)^{-N/2}\det(\boldsymbol{\Sigma})^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{s}-\boldsymbol{\mu})^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{s}-\boldsymbol{\mu})\right\}$$

где $\boldsymbol{\mu}$ — любой вектор размера $N \times 1$ и $\boldsymbol{\Sigma}$ — положительно определенная матрица размера $N \times N$

Представим это распределение как $\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Случай $\mathrm{N}(\mathbf{0}, \mathbf{I})$ называется **многомерным стандартным нормальным распределением**

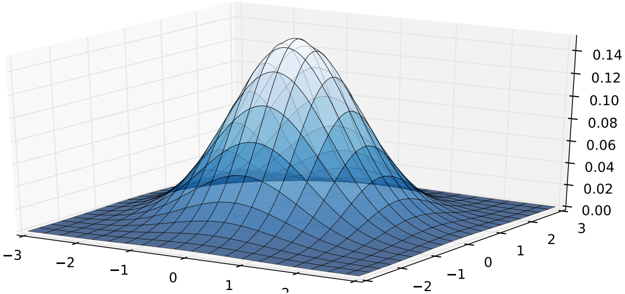


Рис.: Функция плотности двумерного стандартного нормального распределения

Распределение произведения $P_1, \dots, P_N$ определяется следующим фактом:

Факт. (??) Возьмем распределения $P_1, \dots, P_N$ в $\mathbb{R}$, существует единственное и определенное распределение $\mathring{P}$ в $\mathbb{R}^N$, такое что

$$\mathring{P}(B_1 \times \cdots \times B_N)$$

$$= \prod_{n=1}^{N} P_n(B_n) \qquad \text{для всех } B_n \in \mathscr{B}(\mathbb{R}),\ n = 1, \dots, N$$

Единственное, потому что распределения однозначно закреплены цилиндрическими множествами $\mathbb{R}^N$ (смотрите страницу 128 в ET)

Возьмем любое распределение P в $\mathbb{R}^N$, n-ое **частное распределение** P — это распределение в $\mathbb{R}$ определенное как

$$P_n(B) = P(\mathbb{R} \times \cdots \times \mathbb{R} \times B \times \mathbb{R} \times \cdots \times \mathbb{R})$$

Здесь B — n-ый элемент Декартого произведения

Эквивалентно,

$$P_n(B) = P\{\mathbf{s} \in \mathbb{R}^N : \mathbf{s}^\intercal \mathbf{e}_n \in B\}$$

Из P_n мы можем также получить **частную функцию распределения** F_n с помощью

$$F_n(s) := P_n((-\infty, s]) \qquad (s \in \mathbb{R})$$

(смотрите страницу **??** в ЕТ)

Если P_n абсолютно непрерывная, она имеет функцию плотности p_n

Если совместное распределение P имеет функцию плотности p, частное распределение P_n имеет функцию плотности p_n – "интегрировать по другим переменным"

Например, двумерный случай:

$$p_1(s_1) = \int_{-\infty}^{\infty} p(s_1, s_2)\, \mathrm{d}s_2$$

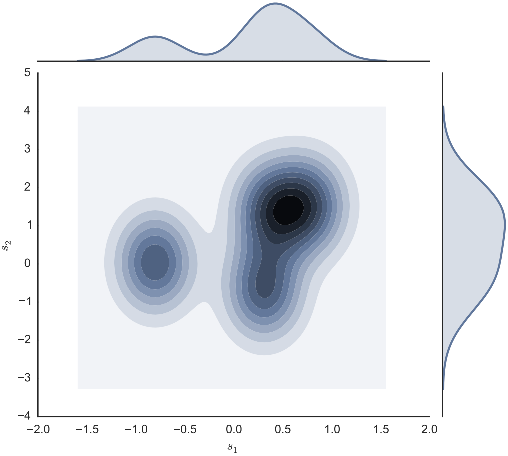


Рис.: Двумерная совместная функция плотности и две ее частных вариации

Совместное распределение не может быть получино только из частных

- частные не говорят нам о своем взаимодействии

Исключение составляют случаи отсутствия взаимодействия - случай произведения функций распределения

Распределения случайных векторов

Пусть $\mathbf{x}$ — случайный вектор в $\mathbb{R}^N$

Распределение $\mathbf{x}$ является вероятностной мерой P на $\mathscr{B}(\mathbb{R}^N)$ определяемая как

$$P(B) = \mathbb{P}\{\mathbf{x} \in B\} \qquad (B \in \mathscr{B}(\mathbb{R}^N))$$

P здесь также называется **совместным распределением** $x_1, \ldots, x_N$, и мы пишем $\mathcal{L}(\mathbf{x}) = P$

Совместное распределение представлено многомерной функцией распределения $F\colon \mathbb{R}^N \to [0,1]$:

$$F(s_1, \ldots, s_N) = \mathbb{P}\{x_1 \leq s_1, \ldots, x_N \leq s_N\}$$

или, в векторной форме

$$F(\mathbf{s}) = \mathbb{P}\{\mathbf{x} \leq \mathbf{s}\} \qquad (\mathbf{s} \in \mathbb{R}^N)$$

Когда распределение P абсолютно непрерывно, существует неотрицательная функция p в $\mathbb{R}^N$, такая что

$$\int_B p(\mathbf{s})\, \mathrm{d}\mathbf{s} = \mathbb{P}\{\mathbf{x} \in B\} \qquad (B \in \mathscr{B}(\mathbb{R}^N))$$

функция p — **совместная функция плотности** $\mathbf{x}$

Для выполнения вышеизложенного достаточно, чтобы

$$\int_{-\infty}^{s_N} \cdots \int_{-\infty}^{s_1} p(t_1, \ldots, t_N)\, \mathrm{d}t_1 \cdots \mathrm{d}t_N = F(s_1, \ldots, s_N)$$

для всех $s_n \in \mathbb{R}$, $n = 1, \ldots, N$

Если $\mathbf{x} = (x_1, \ldots, x_N)$ — случайный вектор в $\mathbb{R}^N$, то каждый x_n — случайная величина в $\mathbb{R}$

Пусть $P_n = \mathcal{L}(x_n)$, тогда:

$$P_n(B) = \mathbb{P}\{x_n \in B\} \qquad (B \in \mathscr{B}(\mathbb{R}),\ n = 1, \ldots, N)$$

P_n называется **частным распределением x_n**

Если $P_1 = P_2 = \cdots = P_N$, то $x_1, \ldots, x_N$ **одинаково распределены**

Нормальные случайные векторы

Случайная переменная x **нормально распределена**, если $x = \mu + \sigma z$ для некоторых $\sigma \geq 0$

Мы пишем $\mathcal{L}(x) = \mathrm{N}(\mu, \sigma)$

Случайный вектор $\mathbf{x}$ в $\mathbb{R}^N$ **многомерный нормальный**, если

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{C}\mathbf{z}$$

где $\mathbf{z}$ — стандартный нормальный случайный вектор размера $K \times 1$, матрица $\mathbf{C}$ имеет размер $N \times K$ и вектор $\boldsymbol{\mu}$ имеет размер $N \times 1$

Если $\mathbf{x}$ многомерный нормальный, то мы пишем $\mathcal{L}(\mathbf{x}) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, где

$$\boldsymbol{\mu} := \mathbb{E}\,\mathbf{x} \quad \text{и} \quad \boldsymbol{\Sigma} := \operatorname{var}\mathbf{x}$$

Имеется $\boldsymbol{\Sigma} = \mathbf{C}\mathbf{C}^{\mathsf{T}}$ (вспомним факт 5.1.2 в ET)

$\mathcal{L}(\mathbf{x}) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ не подразумевает, что $\mathbf{x}$ имеет многомерную нормальную функцию плотности

- распределение $\mathbf{x}$ может и не быть абсолютно непрерывным, например, если $\mathbf{C} = \mathbf{0}$

Абсолютная непрерывность распределения $\mathbf{x}$ совпадает с условиями, где $\boldsymbol{\Sigma} := \operatorname{var} \mathbf{x}$ несингулярна – несингулярность $\boldsymbol{\Sigma}$ будет верна тогда и только тогда, когда $\mathbf{C}^\mathsf{T}$ имеет полный ранг столбцов

Факт. (??) Пусть $\mathbf{x}$ — случайный вектор в $\mathbb{R}^N$. Следующие утверждения верны:

1. вектор $\mathbf{x}$ многомерный нормальный тогда и только тогда, когда $\mathbf{a}^\intercal \mathbf{x}$ нормально распределено в $\mathbb{R}$ для каждого постоянного вектора $\mathbf{a}$ размера $N \times 1$
2. Если $\mathcal{L}(\mathbf{x}) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, то
$$\mathcal{L}(\mathbf{A}\mathbf{x} + \mathbf{b}) = \mathrm{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\intercal)$$
для всех постоянных согласованных $\mathbf{A}, \mathbf{b}$

Следствие: если $\mathbf{x} = (x_1, \ldots, x_N)$ многомерный нормальный, то частное распределение x_n одномерное нормальное

Всегда ли совместное распределение N одномерных нормальных случайных величин является многомерным нормальным?

- Ответ: нет

Ожидания из распределений

Пусть $h\colon \mathbb{R}^N \to \mathbb{R}$ — любая $\mathscr{B}$-измеримая функция и P — распределение в $\mathbb{R}^N$

Функция h теперь рассматривается как а случайная переменная в $(\mathbb{R}^N, \mathscr{B}(\mathbb{R}^N), P)$

Математическое ожидание h может быть записано как

$$\mathbb{E}_P h :=: \int h(\mathbf{s}) P(d\mathbf{s}) \tag{1}$$

Факт. (??) Пусть $h\colon \mathbb{R}^N \to \mathbb{R}$ $\mathscr{B}$-измерима и P — распределение в $\mathbb{R}^N$. Если P дискретное, с вероятностной функцией $\{p_j\}_{j\geq 1}$ и носителем распределения $\{\mathbf{s}_j\}_{j\geq 1}$, то

$$\int h(\mathbf{s})P(\mathrm{d}\mathbf{s}) = \sum_{j\geq 1} h(\mathbf{s}_j)p_j \tag{2}$$

Если P абсолютно непрерывное с функцией плотности p, то

$$\int h(\mathbf{s})P(\mathrm{d}\mathbf{s}) = \int h(\mathbf{s})p(\mathbf{s})\,\mathrm{d}\mathbf{s} \tag{3}$$

правую сторону (3) следует понимать как

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(s_1,\ldots,s_N)\, p(s_1,\ldots,s_N)\, \mathrm{d}s_1 \cdots \mathrm{d}s_N$$

Как и в одномерном случае, такие объекты, как моменты, являются свойствами распределения

Например, пусть $\mathbf{x}$ — случайный вектор в $\mathbb{R}^K$ с $\mathcal{L}(\mathbf{x}) = P$

Ковариационная матрица $\mathrm{var}[\mathbf{x}]$ имеет i, j-ый элемент равный $\mathbb{E}\left[x_i x_j\right] - \mathbb{E}\left[x_i\right]\mathbb{E}\left[x_j\right]$

Мы можем записать $\mathrm{var}[\mathbf{x}]$ относительно P. Если

$$\boldsymbol{\Sigma}_P = (\sigma_{ij}) \quad \text{, где} \quad \sigma_{ij} := \int (s_i s_j) P(\mathrm{d}\mathbf{s}) - \int s_i P(\mathrm{d}\mathbf{s}) \cdot \int s_j P(\mathrm{d}\mathbf{s})$$

тогда $\boldsymbol{\Sigma}_P = \mathrm{var}[\mathbf{x}]$

Независимость случайных величин

Множество N случайных величин $x_1, \ldots, x_N$ **независимо**, если

$$\mathbb{P} \bigcap_{n=1}^{N} \{x_n \in B_n\} = \prod_{n=1}^{N} \mathbb{P}\{x_n \in B_n\} \tag{4}$$

для любых $B_1, \ldots, B_N$, где каждый B_n — Борелевское подмножество $\mathbb{R}$

Случайные величины $x_1, \ldots, x_N$ независимы, когда множества вида $\{x_1 \in B_1\}, \ldots, \{x_N \in B_N\}$ являются независимыми событиями

Бесконечное множество случайных величин $\{x_n\}_{n=1}^{\infty}$ независимо, если любое конечное подмножество $\{x_n\}_{n=1}^{\infty}$ независимо

Эквивалентное определение независимости с использованием распределений

Пусть P — совместное распределение $\mathbf{x} = (x_1, \ldots, x_N)$ и P_n — его n-ое частное

Так как $\cap_{n=1}^{N}\{x_n \in B_n\} = \{(x_1, \ldots, x_N) \in B_1 \times \cdots \times B_N\}$, случайные величины $x_1, \ldots, x_N$ независимы, если

$$P(B_1 \times \cdots \times B_N) = \prod_{n=1}^{N} P_n(B_n)$$

Элементы случайного вектора независимы тогда и только тогда, когда их совместное распределение равняется произведению их частных распределений

Необходимое и достаточное условие независимости $x_1, \dots, x_N$:

$$F(s_1, \dots, s_N) = \prod_{n=1}^{N} F_n(s_n)$$

для всех $(s_1, \dots, s_N) \in \mathbb{R}^N$, где F функция распределения $\mathbf{x}$ и $F_1, \dots, F_N$ частные функциии распределения (почему?)

Если распределение $\mathbf{x}$ абсолютно непрерывное, мы можем также проверить независимость с помощью его функции плотности:

Факт. (??) Если $\mathbf{x} = (x_1, \ldots, x_N)$ имеет совместную функцию плотности p и частные $p_1, \ldots, p_N$, то $x_1, \ldots, x_N$ независимы тогда и только тогда, когда

$$p(s_1, \ldots, s_N) = \prod_{n=1}^{N} p_n(s_n) \qquad \text{для всех} \quad (s_1, \ldots, s_N) \in \mathbb{R}^N$$

Пример.

Пусть $\mathcal{L}(\mathbf{x}) = \mathcal{L}(x_1, \ldots, x_N) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Предположим также, что $\boldsymbol{\Sigma}$ диагональна, с n-ым диагональным элементом $\sigma_n > 0$, тогда $x_1, \ldots, x_N$ независимые

Чтобы убедиться в этом, проверим для любых $\mathbf{s} = (s_1, \ldots, s_N) \in \mathbb{R}^N$, имеется

$$p(\mathbf{s}) = (2\pi)^{-N/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{s} - \boldsymbol{\mu})^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{s} - \boldsymbol{\mu})\right\}$$

$$= \frac{1}{(2\pi)^{N/2} \prod_{n=1}^{N} \sigma_n} \exp\left\{-\frac{1}{2} \sum_{n=1}^{N} (s_n - \mu_n)^2 \sigma_n^{-2}\right\}$$

Пример. (прод.) Вычисление определителя и обратной матрицы $\mathbf{\Sigma}$ с помощью фактов ?? и ??

Последнее выражение можно разложить дальше

$$p(\mathbf{s}) = \prod_{n=1}^{N} \frac{1}{(2\pi)^{1/2}\sigma_n} \exp\left\{ \frac{-(s_n - \mu_n)^2}{2\sigma_n^2} \right\} = \prod_{n=1}^{N} p_n(s_n)$$

где p_n — функция плотности $\mathrm{N}(\mu_n, \sigma_n^2)$

Факт. (??) Если $x_1, \ldots, x_N$ независимые и каждый x_n интегрируемый, то

$$\mathbb{E}\left[\prod_{n=1}^{N} x_n\right] = \prod_{n=1}^{N} \mathbb{E}\left[x_n\right]$$

Независимость случайных векторов

Случайные векторы $\mathbf{x}_1, \dots, \mathbf{x}_N$ в $\mathbb{R}^K$ называются **независимыми**, если

$$\mathbb{P} \bigcap_{n=1}^{N} \{\mathbf{x}_n \in B_n\} = \prod_{n=1}^{N} \mathbb{P}\{\mathbf{x}_n \in B_n\}$$

для любых $B_1, \dots, B_N$, где каждый B_n — Борелевское подмножество $\mathbb{R}^K$

Факт. (??) Если $\mathbf{x}_1, \ldots, \mathbf{x}_N$ — независимые случайные векторы в $\mathbb{R}^K$ и $f_1, \ldots, f_N$ — любые $\mathscr{B}$-измеримые функции, то $f_1(\mathbf{x}_1), \ldots, f_N(\mathbf{x}_N)$ также независимы.

Доказательство. Заметим, что $f_n(\mathbf{x}_n) \in B_n$ тогда и только тогда, когда $\mathbf{x}_n \in f^{-1}(B_n)$. Это ведет к

$$\bigcap_{n=1}^{N} \{f_n(\mathbf{x}_n) \in B_n\} = \bigcap_{n=1}^{N} \{\mathbf{x}_n \in f^{-1}(B_n)\}$$

Применяем независимость $\mathbf{x}_1, \ldots, \mathbf{x}_N$

$$\mathbb{P} \bigcap_{n=1}^{N} \{f_n(\mathbf{x}_n) \in B_n\}$$

$$= \prod_{n=1}^{N} \mathbb{P}\{\mathbf{x}_n \in f^{-1}(B_n)\} = \prod_{n=1}^{N} \mathbb{P}\{f_n(\mathbf{x}_n) \in B_n\}$$

Факт. (??) Если $\mathbf{x}$ и $\mathbf{y}$ независимые, то $\operatorname{cov}(\mathbf{x}, \mathbf{y}) = 0$.

Обратное не верно: можно найти примеры зависимых векторов с нулевой ковариацией. Однако,

Факт. (??) Если $\mathbf{x}$ многомерно нормально распределен и $\mathbf{A}$ и $\mathbf{B}$ согласованные постоянные матрицы, то $\mathbf{Ax}$ и $\mathbf{Bx}$ независимые тогда и только тогда, когда $\operatorname{cov}(\mathbf{Ax}, \mathbf{Bx}) = \mathbf{0}$

Факт. (??) Пусть S — любое линейное подпространство $\mathbb{R}^N$, $\mathbf{P} := \operatorname{proj} S$ и $\mathbf{M}$ — остаточная проекция. Если $\mathcal{L}(\mathbf{z}) = \mathrm{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ в $\mathbb{R}^N$ для некоторых $\sigma^2 > 0$, то $\mathbf{Pz}$ и $\mathbf{Mz}$ независимые

Факт. (??) Если $w_1, \ldots, w_N$ независимые с $\mathcal{L}(w_n) = \mathrm{N}(\mu_n, \sigma_n^2)$ для всех n, то

$$\mathcal{L}\left[\alpha_0 + \sum_{n=1}^{N} \alpha_n w_n\right] = \mathrm{N}\left(\alpha_0 + \sum_{n=1}^{N} \alpha_n \mu_n, \ \sum_{n=1}^{N} \alpha_n^2 \sigma_n^2\right)$$

Суммы произвольных нормальных не всегда нормальны — нам требуется многомерное нормальное распределение

В факте (??) выше:

$$\mathcal{L}(w_1, \ldots, w_N) = \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

где $\mathbf{e}_n^\top \boldsymbol{\mu} = \mu_n$, и

$$\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_N^2)$$

Копула C в $\mathbb{R}^N$ — многомерная функция распределения определённая на единичном гиперкубе $[0,1]^N$, такая что каждое ее частное распределение равномерно на $[0,1]$

C — функция вида

$$C(s_1, \dots, s_N) = \mathbb{P}\{u_1 \leq s_1, \dots, u_N \leq s_N\} \tag{5}$$

Где $0 \leq s_n \leq 1$ и $\mathcal{L}(u_n) = U[0,1]$ для всех n

Пока каждый u_n имеет фиксированное частное распределение, существует бесконечно много способов составить совместное распределение

Пример. Функция $C(s_1, s_2) = s_1 s_2$ on $[0, 1]^2$ называется **независимая копула**

Частные распределения $C(s_1, 1) = s_1$ и $C(1, s_2) = s_2$ как и требуется

(Это функции распределения для $U[0, 1]$ распределения)

Пример. **Копула Гумбеля** — класс функций в $[0,1]^2$, определяемый как

$$C(s_1, s_2) = \exp\left\{-\left[(-\ln s_1)^\theta + (-\ln s_2)^\theta\right]^{1/\theta}\right\}, \quad (\theta \geq 1)$$

Копула Клейтона определяется как

$$C(s_1, s_2) = \left\{\max\left[s_1^{-\theta} + s_2^{-\theta} - 1,\, 0\right]\right\}^{-1/\theta}, \quad (\theta \geq -1,\, \theta \neq 0)$$

Обе они принадлежат к общему классу, называемому **Архимедовы копулы**

Мы можем взять равномерные функции распределения $F_1, \ldots, F_N$ и копулу C, чтобы создать многомерную функцию распределения в $\mathbb{R}^N$ с помощью

$$F(s_1, \ldots, s_N) = C(F_1(s_1), \ldots, F_N(s_N)) \qquad (s_n \in \mathbb{R},\, n = 1, \ldots, N) \quad (6)$$

Польза: разделяем определение частных и определение совместного распределения

Пример. **bonhomme2009assessing** использует копулы для моделирования одного компонента динамики заработка в исследовании, основанном на трехлетних панельных данных (French Labor Force Survey)

Разделы относительно большие (около 30 000), что позволяет гибко моделировать частные распределения с помощью смеси нормальных

Однако, размер временного ряда короткий, поэтому используется семейство копул с одним параметром для привязки частных во времени не трудозатратным способом

Теорема. (**??**) Если F — некоторая функция распределения в $\mathbb{R}^N$ с частными $F_1, \ldots, F_N$, то существует копула C, такая что (6) выполняется. Если каждый F_n непрерывен, то это представление является единственным.

Если $F_1, \dots, F_N$ равномерные нормальные, то $C(F_1(s_1), \dots, F_N(s_N))$ будут равняться многомерной нормальной функции распределения для одного варианта копулы, называемой Гауссовой копулой

Другие варианты приводят к другим распределениям

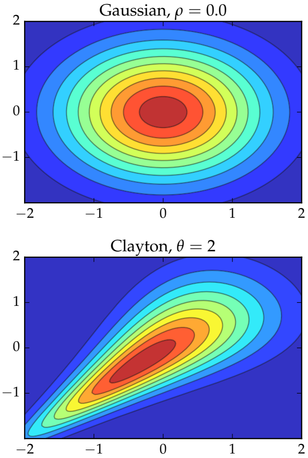


Рис.: Двумерный Гауссовская (вверху) и не-Гауссовская (внизу)

Свойства именных распределений

Факт. (??) Если $x_1, \ldots, x_N$ независимые и $\mathcal{L}(x_n) = \chi^2(k_n)$, то $\mathcal{L}(\sum_n x_n) = \chi^2(\sum_n k_n)$

Факт. (??) Если z и x независимые с $\mathcal{L}(z) = \mathrm{N}(0,1)$ и $\mathcal{L}(x) = \chi^2(k)$, то

$$z\sqrt{\frac{k}{x}} \quad \text{распределено как } t \text{ с } k \text{ степенями свободы}$$

Факт. (??) Если $\mathcal{L}(z_1, \ldots, z_N) = \mathrm{N}(\mathbf{0}, \mathbf{I})$, то $\mathcal{L}(\sum_{n=1}^{N} z_n^2) = \chi^2(N)$.

Факт. (??) Если $\mathcal{L}(\mathbf{z}) = \mathrm{N}(\mathbf{0}, \mathbf{I})$ и $\mathbf{A}$ симметрична и идемпотентна, то

$$\mathcal{L}\left(\mathbf{z}^\top \mathbf{A} \mathbf{z}\right) = \chi^2(K) \quad \text{, где} \quad K := \operatorname{trace} \mathbf{A}$$

Упражнение: получите факт (??) из факта (??). (Смотрите страницу ?? в ET)

Условия и ожидание

Условное ожидание — одно из важнейших понятий как в экономической теории, так и в эконометрике

В этом разделе дается построение математического ожидания, основанное на проекции:

- условное математическое ожидание как оптимальное предсказание с учетом ограниченной информации

Условные функции плотности

Сначала обсуждение условных функций плотности

Пусть x_1 и x_2 — случайные величины. **Условиная функция плотности** x_2 при заданном $x_1 = s_1$ определяется как

$$p(s_2 \mid s_1) := \frac{p(s_1, s_2)}{p(s_2)}$$

Здесь p может обозначать совместную, частную или условную функцию плотности, определяемую аргументом

Закон полной вероятности расширяется до случая с функциями плотности следующим образом: Если (x_1, x_2) — случайный вектор в $\mathbb{R}^2$, то

$$p(s_2) = \int_{-\infty}^{\infty} p(s_2 \mid s_1) p(s_1) \, \mathrm{d}s_1 \qquad (s_2 \in \mathbb{R})$$

Доказательство. Зафиксируем $s_2 \in \mathbb{R}$ и проинтегрируем совместную функцию плотности, чтобы получить частную, получается

$$p(s_2) = \int_{-\infty}^{\infty} p(s_1, s_2) \, \mathrm{d}s_1$$

Сочетаем с $p(s_2 \mid s_1) = p(s_1, s_2) / p(s_1)$, чтобы получить результат

Закон Байеса также расширяется до случая с функциями плотности:

$$p(s_2 \mid s_1) = \frac{p(s_1 \mid s_2)p(s_2)}{p(s_1)}$$

Условная функция плотности $x_{k+1}, \ldots, x_N$ при $x_1 = s_1, \ldots, x_k = s_k$ определяется как

$$p(s_{k+1}, \ldots, s_N \mid s_1, \ldots, s_k) = \frac{p(s_1, \ldots, s_N)}{p(s_1, \ldots, s_k)}$$

Перегруппируйте, чтобы получить полезное разложение совместной функции плотности:

$$p(s_1, \ldots, s_N) = p(s_{k+1}, \ldots, s_N \mid s_1, \ldots, s_k) p(s_1, \ldots, s_k)$$

Предположим, мы хотим предсказать случайную переменную y с помощью другой переменной x

Возьмем x такой, что x и y, как ожидается, будут близки при большинстве реализаций неопределенности

Но что значит "ожидаются близкими"?

Среднеквадратическая ошибка (MSE)

$$\mathbb{E}\left[(x-y)^2\right]$$

Среднеквадратическое отклонение:

$$\|x-y\| := \sqrt{\mathbb{E}\left[(x-y)^2\right]} \tag{7}$$

Есть много параллелей между обычным векторным пространством с эвклидовой нормой и множеством случайных величин в сочетании с "нормой", определенной в (7) — мы формализуем эти идеи далее

Первым геометрическим понятием, которое мы определили для векторов, было скалярное произведение

Аналогично, определим **скалярное произведение между двумя случайными величинами** x и y

$$\langle x, y \rangle := \mathbb{E}[xy]$$

Неравенство Коши — Буняковского для случайных величин говорит нам, что $\mathbb{E}[xy]$ должен быть конечным и определенным всюду, где x и y оба имеют конечные вторые моменты

Множество случайных величин с конечными вторыми моментами обычно обозначается как L_2

$$L_2 := \{ \text{ все случайные величины } x \text{ в } (\Omega, \mathscr{F}, \mathbb{P}) \text{ с } \mathbb{E}[x^2] < \infty \}$$

Факт. (??) Для любых $\alpha, \beta \in \mathbb{R}$ и любых $x, y, z \in L_2$ следующие утверждения верны:

1. $\langle x, y \rangle = \langle y, x \rangle$.
2. $\langle \alpha x, \beta y \rangle = \alpha\beta \langle x, y \rangle$.
3. $\langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle$.

Свойства следуют из определения скалярного произведения и линейности $\mathbb{E}$

Сравните приведенное выше с фактом ?? в ЕТ для векторов в эвклидовом пространстве

Определим L_2 **норму** как

$$\|x\| := \sqrt{\langle x, x \rangle} := \sqrt{\mathbb{E}\left[x^2\right]} \qquad (x \in L_2)$$

Норма дает понятие расстояния $\|x - y\|$ между случайными величинами что согласуется с понятием среднеквадратического отклонения

Факт. (??) Для любых $\alpha \in \mathbb{R}$ и любых $x, y \in L_2$, следующие утверждения верные:

1. $\|x\| \geq 0$ и $\|x\| = 0$ тогда и только тогда, когда $x = 0$
2. $\|\alpha x\| = |\alpha| \|x\|$
3. $\|x + y\| \leq \|x\| + \|y\|$
4. $|\langle x, y \rangle| \leq \|x\| \|y\|$

Свойство 2. приведенного выше факта является непосредственным из определения нормы и линейности $\mathbb{E}$

Свойство 3. называется **неравенством треугольника**, как и в векторном случае

Свойство 4. — это просто **неравенство Коши — Буняковского** для случайных величин со страницы ??

Как и в векторном случае, неравенство треугольника доказывается неравенством Коши - Буняковского (смотрите упражнение ??)

Относительно 1., $\|x\| = 0$ не подразумевает, что $x(\omega) = 0$ для всех $\omega \in \Omega$

Мы можем сказать, что если $\|x\| = 0$, то $\mathbb{P}\{x = 0\} = 1$

При работе с L_2, принято соглашение не различать случайные величины, различающиеся с нулевой вероятностью.

Линейные подпространства в L_2

Любая **линейная комбинация** случайных величин с конечной дисперсией

$$\alpha_1 x_1 + \cdots \alpha_K x_K, \qquad \alpha_k \in \mathbb{R},\ x_k \in L_2 \tag{8}$$

снова в L_2

Когда X — подпространство L_2, множество конечных линейных комбинаций, которое может быть сформировано из элементов X, называется **линейной оболочкой** X, и оюозначается как $\operatorname{span} X$

Пример. Если $x \in L_2$ и $\mathbb{1} := \mathbb{1}_\Omega$ постоянная случайная переменная, всегда равная 1, то $\operatorname{span}\{\mathbb{1}, x\}$ — множество случайных величин

$$\alpha + \beta x := \alpha \mathbb{1} + \beta x \quad \text{для скаляров} \quad \alpha, \beta \tag{9}$$

Это множество $\mathcal{L}$ введенное, когда мы обсуждали лучшие линейные предикторы

Подмножество S множества L_2 называется **линейным подпространством** L_2, если оно замкнуто относительно сложения и умножения на скаляр

- для каждого $x, y \in S$ и $\alpha, \beta \in \mathbb{R}$, мы имеем $\alpha x + \beta y \in S$

Пример. Линейная оболочка любого множества элементов L_2 — линейное подпространство в L_2

Пример. Множество $Z := \{x \in L_2 : \mathbb{E}\, x = 0\}$ является линейным подпространством L_2, так как

$$x, y \in Z \text{ и } \alpha, \beta \in \mathbb{R} \implies \mathbb{E}\,[\alpha x + \beta y] = \alpha \mathbb{E}\,[x] + \beta \mathbb{E}\,[y] = 0$$

Как и в $\mathbb{R}^N$, **ортонормированный базис** линейного подпространства S пространства L_2 — множество $\{u_1, \ldots, u_K\} \subset S$ со свойством

$$\langle u_j, u_k \rangle = \mathbb{1}\{j = k\}$$

$$\text{и} \quad \operatorname{span}\{u_1, \ldots, u_K\} = S$$

Пример. Пусть $x \in L_2$ такой, что $S := \operatorname{span}\{\mathbb{1}, x\}$ множество случайных величин

$$\alpha + \beta x := \alpha \mathbb{1} + \beta x \quad \text{for scalars} \quad \alpha, \beta \tag{10}$$

Если мы определим

$$u_1 := \mathbb{1} \quad \text{и} \quad u_2 := \frac{x - \mu}{\sigma_x}$$

Тогда

$$\langle u_1, u_2 \rangle = \mathbb{E}\left[u_1 u_2\right] = \mathbb{E}\left[\frac{x-\mu}{\sigma_x}\right] = 0$$

Ясно, что $\|u_1\| = \|u_2\| = 1$, так что эта пара ортонормирована

Также просто показать, что $\operatorname{span}\{u_1, u_2\} = \operatorname{span}\{\mathbb{1}, x\}$, значит $\{u_1, u_2\}$ ортонормированный базис для S

Проекции в L_2

Как и в евклидовом случае, если $\langle x, y \rangle = 0$, мы говорим, что x и y **ортогональны**, и пишем $x \perp y$

Факт. Если $x, y \in L_2$ и $\mathbb{E}\, x = 0$ или $\mathbb{E}\, y = 0$, то $x \perp y \iff \operatorname{cov}[x, y] = 0$

Возьмем $y \in L_2$ и линейное подпространство $S \subset L_2$, мы ищем ближайший элемент $\hat{y}$ множества S к y

Близость по норме L_2, так что $\hat{y}$ — решение минимизации $\|y - z\|$ для всех $z \in S$

Мы ищем

$$\hat{y} = \operatorname*{argmin}_{z \in S} \|y - z\| = \operatorname*{argmin}_{z \in S} \sqrt{\mathbb{E}\left[(y - z)^2\right]} \tag{11}$$

Следующая теорема имитирует теорему об ортогональной проекции, которую мы уже видели:

Теорема. (??) Пусть $y \in L_2$ и S — любое непустое замкнутое линейное подпространство L_2

Следующие утверждения верны:

1. задача оптимизации (11) имеет ровно одно решение
2. $\hat{y} \in L_2$ является единственным решением

Утверждение, что S замкнуто значит, что $\{x_n\} \subset S$ и $x \in L_2$ с $\|x_n - x\| \to 0$ подразумевает $x \in S$ — условие истинное для всех линейных подпространств, с которыми мы хотим работать

Аналогично в случае $\mathbb{R}^N$, случайная переменная $\hat{y}$ выше называется **ортогональной проекцией y на S**

Возьмем фиксированное S, операция

$$y \mapsto \text{ортогональная проекция } y \text{ на } S$$

— функция из L_2 в L_2:

- функция называется **ортогональной проекцией на S**
- функция обозначается как $\mathbf{P}$
- мы пишем $\mathbf{P} = \operatorname{proj} S$

Для каждого $y \in L_2$, $\mathbf{P}y$ отображение y с помощью $\mathbf{P}$, которое является ортогональной проекцией $\hat{y}$

- интерпретируем $\mathbf{P}y$ как *лучший преликтор* y *из множества случайных величин, содержащегося в* S

Факт. (??)

Если S — любое линейное подпространство L_2, и $\mathbf{P} = \operatorname{proj} S$, то

1. $\mathbf{P}$ — линейная функция.

Более того, для любых $y \in L_2$, получается

2. $\mathbf{P}y \in S$,
3. $y - \mathbf{P}y \perp S$,
4. $\|y\|^2 = \|\mathbf{P}y\|^2 + \|y - \mathbf{P}y\|^2$,
5. $\|\mathbf{P}y\| \leq \|y\|$, и
6. $\mathbf{P}y = y$ тогда и только тогда, когда $y \in S$.

В 1, $\mathbf{P}$ линейна значит, что $\mathbf{P}(\alpha x + \beta y) = \alpha \mathbf{P}x + \beta \mathbf{P}y$ для всех $x, y \in L_2$ и $\alpha, \beta \in \mathbb{R}$

Факт. (??) Пусть S_i — линейное подпространство L_2 для $i = 1, 2$ и $\mathbf{P}_i = \operatorname{proj} S_i$. Если $S_1 \subset S_2$, то $\mathbf{P}_1 \mathbf{P}_2 y = \mathbf{P}_1 y$ для всех $y \in L_2$

Факт. (??) Если $\{u_1, \ldots, u_K\}$ — ортонормированный базис S, то для всех $y \in L_2$,

$$\mathbf{P} y = \sum_{k=1}^{K} \langle y, u_k \rangle \, u_k \tag{12}$$

Пример. (**??**) Среднее случайной переменной x можно рассматривать как "лучший предиктор x среди множества констант"

Пусть $S := \operatorname{span}\{\mathbb{1}\}$, где $\mathbb{1} := \mathbb{1}_{\Omega}$, и $\mathbf{P} := \operatorname{proj} S$

Объект $\mathbf{P}x$ как раз лучший предиктор x в классе постоянных случайных величин

Не удивительно, что $\mathbf{P}x = \mu\mathbb{1}$, где $\mu := \mathbb{E}\, x$

Самый простой способ проверить это — заметить, что $\{\mathbb{1}\}$ является ортонормированным множеством, охватывающим S, и следовательно, по (12),

$$\mathbf{P}x = \langle x, \mathbb{1}\rangle\ \mathbb{1} = \mathbb{E}\,[x\mathbb{1}]\mathbb{1} = \mathbb{E}\,[x]\mathbb{1} = \mu\,\mathbb{1}$$

Вы можете также проверить утверждение, что $\mu\mathbb{1}$ — проекция x на S, проверив условия в (ii) теоремы **??**

Пример.

Зафиксируем $x, y \in L_2$ и рассмотрим проецирование y на $S := \operatorname{span}\{\mathbb{1}, x\}$

Множество S является множеством случайных величин

$$\alpha + \beta x := \alpha \mathbb{1} + \beta x \quad \text{для скаляров} \quad \alpha, \beta$$

Задача проецирования y на S is эквивалентна задаче поиска лучшего линейного предиктора из §??

Для реализации отзыва проекции

$$u_1 := \mathbb{1} \quad \text{и} \quad u_2 := \frac{x - \mu}{\sigma_x}$$

сформируем ортонормированный базис для S

Пусть $\mathbf{P} = \operatorname{proj} S$, применим факт (**??**), получаем

$$\mathbf{P}y = \langle y, u_1 \rangle u_1 + \langle y, u_2 \rangle u_2 = \mathbb{E}[y] + \frac{\operatorname{cov}[x, y]}{\operatorname{var}[x]}(x - \mathbb{E}[x])$$

Альтернативно

$$\mathbf{P}y = \alpha^* + \beta^* x$$

$$\text{где} \quad \beta^* := \frac{\operatorname{cov}[x, y]}{\operatorname{var}[x]} \quad \text{и} \quad \alpha^* := \mathbb{E}[y] - \beta^* \mathbb{E}[x]$$

Регрессия населения

Рассмотрим расширение задачи поиска лучшего линейного предиктора, описанной выше, до задачи, в которой информация для прогнозирования y — случайный вектор $\mathbf{x}$ в $\mathbb{R}^K$

Мы ищем L_2 ортогональную проекцию y на линейное подпространство:

$$\text{span}\{\mathbf{x}\} := \text{ случайные величины вида } \mathbf{x}^\mathsf{T}\mathbf{b} \text{ для некоторых } \mathbf{b} \in \mathbb{R}^K$$

Предположим, что $\mathbb{E}\left[\mathbf{x}^\mathsf{T}\mathbf{x}\right] < \infty$

Факт. (??) Если $\mathbb{E}[\mathbf{x}\mathbf{x}^\mathsf{T}]$ положительно определена, то проекция $\mathbf{P}y$ любого $y \in L_2$ на $\operatorname{span}\{\mathbf{x}\}$ определяется как

$$\hat{y} = \mathbf{x}^\mathsf{T}\mathbf{b}^* \quad \text{где} \quad \mathbf{b}^* := \mathbb{E}[\mathbf{x}\mathbf{x}^\mathsf{T}]^{-1}\mathbb{E}[\mathbf{x}y]$$

Упражнение **??** просит доказать вышеизложенный факт

Положительная определенность $\mathbb{E}[\mathbf{x}\mathbf{x}^\mathsf{T}]$ обеспечивает обратимость, значит $\mathbf{b}^*$ однозначно определено

По определению ортогональных проекций, $\mathbf{b}^*$ обязательно удовлетворяет

$$\mathbf{b}^* = \operatorname*{argmin}_{\mathbf{a}\in\mathbb{R}^K} \mathbb{E}[(y - \mathbf{x}^\mathsf{T}\mathbf{a})^2]$$

Задача линейного прогнозирования рассматривается также под названием **линейная регрессия населения**

- "население", потому что мы используем истинное совместное распределение $(\mathbf{x}, y)$, когда считаем ожидания

У регрессии популяции есть аналог, называемый многомерной линейной регрессией, основанный на наблюдениях $(\mathbf{x}, y)$ – мы обсудим это в главе **??**

Измеримость

Мы не всегда хотим ограничиваться линейными прогнозами

Чтобы отказаться от требования линейности, изменим линейное подпространство, используемое для проецирования, из множества линейных функций $\mathbf{x}$ на множество произвольных функций $\mathbf{x}$

В результате, лучший предиктор — это условное математическое ожидание относительно $\mathbf{x}$

Подпространством произвольных действительных функций от $\mathbf{x}$ называются $\mathbf{x}$-измеримые функции

Пусть $\mathcal{G} := \{x_1, \ldots, x_D\}$ — любое множество случайных величин и z — любая другая случайная величина

Переменная z **$\mathcal{G}$-измерима**, если существует $\mathscr{B}$-измеримая функция $g\colon \mathbb{R}^D \to \mathbb{R}$, такая что

$$z = g(x_1, \ldots, x_D)$$

- равенство между случайными величинами следует интерпретировать поточечно

$\mathcal{G}$ иногда упоминается как **информационное множество**

Мы также будем писать $\mathbf{x} = (x_1, \ldots, x_D)$ и говорить, что z является $\mathbf{x}$-измеримым

Аналогичная терминология будет использоваться для скаляров и матриц

- например, если $\mathbf{X}$ — случайная матрица, то $\mathbf{X}$-измеримость означает $\mathcal{G}$-измеримость, когда $\mathcal{G}$ содержит все элементы $\mathbf{X}$

Интуиция: $\mathcal{G}$-измеримость z значит, что z полностью определяется элементами в $\mathcal{G}$

Пример. Пусть x, y и z — случайные величины и пусть α и β — скаляры

Если $z = \alpha x + \beta y$, то z $\{x, y\}$-измеримо (возьмем $g(s,t) := \alpha s + \beta t$)

Пример. Если $x_1, \ldots, x_N$ — случайные величины и $\mathcal{G} := \{x_1, \ldots, x_N\}$, то выборочное среднее $\bar{x}_N := \frac{1}{N} \sum_{n=1}^{N} x_n$ является $\mathcal{G}$-измеримым.

Пример. Пусть $\mathbf{x}$ и y независимые и невырожденные

Тогда y не является $\mathbf{x}$-измеримым. Если бы он таким являлся, мы бы имели $y = g(\mathbf{x})$ для некоторой функции g, противоречащее независимости $\mathbf{x}$ и y

Пример. Пусть $y = \alpha$, где α — константа

Эта вырожденная случайная величина является $\mathcal{G}$-измеримой для любых информационных множеств $\mathcal{G}$, потому что y уже детерминированный

Например, если $\mathcal{G} = \{x_1, \dots, x_p\}$, то мы можем взять $y = g(x_1, \dots, x_p) = \alpha + \sum_{i=1}^{p} 0x_i$

Факт. (??) Пусть α, β — любые скаляры, и пусть x и y — случайные величины. Если x и y являются $\mathcal{G}$-измеримыми, то $u := xy$ и $v := \alpha x + \beta y$ также являются $\mathcal{G}$-измеримыми

Предположим, что $\mathcal{G} \subset L_2$ и рассмотрим множество

$$L_2(\mathcal{G}) := \{\text{все } \mathcal{G}\text{-измеримые случайные величины в } L_2\}$$

С учетом факта **??**:

Факт. Для любых $\mathcal{G} \subset L_2$, множество $L_2(\mathcal{G})$ — линейное подпространство L_2

Это дает нам подпространство для проецирования, что позволяет нам определять условные математические ожидания

Факт. (??) Если $\mathcal{G} \subset \mathcal{H}$ и z является $\mathcal{G}$-измеримой, то z является $\mathcal{H}$-измеримой.

Если z известен, когда переменные в $\mathcal{G}$ известны, то он точно известен, когда дополнительная информация, предоставленная $\mathcal{H}$, доступна

Пример. Пусть x_1, x_2 и y — случайные величины и пусть

$$\mathcal{G} := \{x_1\} \subset \{x_1, x_2\} =: \mathcal{H}$$

Если y является $\mathcal{G}$-измеримой, то $y = g(x_1)$ для некоторых $\mathscr{B}$-измеримых g. Но тогда y будет также являться $\mathcal{H}$-измеримой. Например, мы можем написать $y = h(x_1, x_2)$, где $h(x_1, x_2) = g(x_1) + 0x_2$.

Факт. (5.2.12) Если $\mathcal{G} \subset \mathcal{H}$, то $L_2(\mathcal{G}) \subset L_2(\mathcal{H})$

Условное математическое ожидание

Пусть $\mathcal{G} \subset L_2$ и y — некоторая случайная величина L_2

Условное математическое ожидание y при данном $\mathcal{G}$ записывается как $\mathbb{E}\left[y \mid \mathcal{G}\right]$ или $\mathbb{E}^{\mathcal{G}}[y]$ и определяется как

$$\mathbb{E}\left[y \mid \mathcal{G}\right] := \underset{z \in L_2(\mathcal{G})}{\operatorname{argmin}} \|y - z\| \tag{13}$$

$\mathbb{E}\left[y \mid \mathcal{G}\right]$ — лучший предиктор y при данной информации, содержащейся в $\mathcal{G}$

Решение такой задачи минимизации вообще существует? И является ли оно единственным?

- да и да

Имеется

$$\mathbb{E}\,[y \mid \mathcal{G}] = \mathbf{P}y \quad \text{, когда } \mathbf{P} := \operatorname{proj} L_2(\mathcal{G})$$

По теореме об ортогональной проекции, проекция существует и является единственной

Альтернативное (и эквивалентное) определение условного математического ожидания

Функция $\hat{y}$, где $\hat{y} \in L_2$, — **условное математическое ожидание** y при данном $\mathcal{G}$, если

1. $\hat{y}$ является $\mathcal{G}$-измеримой и
2. $\mathbb{E}[\hat{y}\, z] = \mathbb{E}[yz]$ для всех $\mathcal{G}$-измеримых $z \in L_2$.

Для удобства мы также будем использовать такие символы, как $\mathbb{E}[y \mid x_1, \dots, x_D]$ или $\mathbb{E}[y \mid \mathbf{x}]$

- так же, как $\mathbb{E}[y \mid \mathcal{G}]$, когда $\mathcal{G}$ определяется как информационное множество, содержащее переменные, на которые мы ставим условие

Пример. Если x и u независимые, $\mathbb{E}\, u = 0$ и $y = x + u$, то $\mathbb{E}\,[y \mid x] = x$. Чтобы доказать это, нам нужно показать, что x удовлетворяет условиям 1–2 выше

Ясно, что x является x-измеримой

Для 2. мы должны показать, что $\mathbb{E}\,[x\, z] = \mathbb{E}\,[y\, z]$ для всех x-измеримых z. Это означает утверждение

$$\mathbb{E}\,[xg(x)] = \mathbb{E}\,[(x+u)g(x)]$$

для любых $\mathscr{B}$-измеримых g, которое является верным из-за независимости и $\mathbb{E}\, u = 0$

Факт. (??) Возьмем $\mathbf{x} \in \mathbb{R}^D$ и y в L_2, существует $\mathscr{B}$-измеримая функция $f^* \colon \mathbb{R}^D \to \mathbb{R}$, такая что $\mathbb{E}\left[y \mid \mathbf{x}\right] = f^*(\mathbf{x})$

Частная функция f^*, удовлетворяющая $f^*(\mathbf{x}) = \mathbb{E}\left[y \mid \mathbf{x}\right]$ называется **функцией регрессии** y при данном $\mathbf{x}$

Пример. Если x и y — случайные величины и $p(y \mid x)$ —условная функция плотности y при данном x, то

$$\mathbb{E}\left[y \mid x\right] = \int t p(t \mid x)\, \mathrm{d}t$$

Докажите в качестве упражнения **??** в ЕТ

Факт. (??) Пусть x и y — случайные величины в L_2, пусть α и β — скаляры, и пусть $\mathcal{G}$ и $\mathcal{H}$ — подмножества L_2. Следующие свойства выполняются:

1. Линейность: $\mathbb{E}\left[\alpha x + \beta y \mid \mathcal{G}\right] = \alpha\mathbb{E}\left[x \mid \mathcal{G}\right] + \beta\mathbb{E}\left[y \mid \mathcal{G}\right]$
2. Если $\mathcal{G} \subset \mathcal{H}$, то $\mathbb{E}\left[\mathbb{E}\left[y \mid \mathcal{H}\right] \mid \mathcal{G}\right]\mathbb{E}\left[y \mid \mathcal{G}\right]$ и $\mathbb{E}\left[\mathbb{E}\left[y \mid \mathcal{G}\right]\right] = \mathbb{E}\left[y\right]$ (**закон повторных ожиданий**)
3. Если y iне зависима от переменных в $\mathcal{G}$, то $\mathbb{E}\left[y \mid \mathcal{G}\right] = \mathbb{E}\left[y\right]$.
4. Если y является $\mathcal{G}$-измеримой, то $\mathbb{E}\left[y \mid \mathcal{G}\right] = y$
5. Если x является $\mathcal{G}$-измеримой, то $\mathbb{E}\left[xy \mid \mathcal{G}\right] = x\mathbb{E}\left[y \mid \mathcal{G}\right]$ (**условный детерминизм**)

Резюмирем: при данном $y \in L_2$ и случайном векторе $\mathbf{x}$ в $\mathbb{R}^D$, условное математическое ожидание $\mathbb{E}[y \mid \mathbf{x}]$ — функция f^* переменной $\mathbf{x}$, называемая функцией регрессии y при данном $\mathbf{x}$, такая что:

$$f^*(\mathbf{x}) = \operatorname*{argmin}_{g \in G} \mathbb{E}[(y - g(\mathbf{x}))^2] \tag{14}$$

где G — множество функций из $\mathbb{R}^D$ в $\mathbb{R}$ с $g(\mathbf{x}) \in L_2$

Для любых $g \in G$, мы также имеем

$$\mathbb{E}[(y - g(\mathbf{x}))^2] = \mathbb{E}[(y - f^*(\mathbf{x}))^2] + \mathbb{E}[(f^*(\mathbf{x}) - g(\mathbf{x}))^2] \tag{15}$$

Это подразумевает, что (14), потому что $(f^*(\mathbf{x}) - g(\mathbf{x}))^2 \geq 0$

Чтобы доказать (15), пусть f^* — функция регрессии, возьмем любой $g \in G$ и заметим, что

$$(y - g(\mathbf{x}))^2 = (y - f^*(\mathbf{x}) + f^*(\mathbf{x}) - g(\mathbf{x}))^2$$

$$= (y - f^*(\mathbf{x}))^2 + 2(y - f^*(\mathbf{x}))(f^*(\mathbf{x}) - g(\mathbf{x}))$$

$$+ (f^*(\mathbf{x}) - g(\mathbf{x}))^2$$

Рассмотрим математическое ожидание перемножения разных величин. Из закона повторных ожиданий:

$$\mathbb{E}\{(y - f^*(\mathbf{x}))(f^*(\mathbf{x}) - g(\mathbf{x}))\} \tag{16}$$

$$= \mathbb{E}\{\mathbb{E}[(y - f^*(\mathbf{x}))(f^*(\mathbf{x}) - g(\mathbf{x})) \mid \mathbf{x}]\}$$

Используем условный детерминизм, перепишем часть в фигурных скобках справа как

$$(f^*(\mathbf{x}) - g(\mathbf{x}))\mathbb{E}\left[(y - f^*(\mathbf{x})) \mid \mathbf{x}\right]$$

Для второй части данного умножения

$$\mathbb{E}\left[y - f^*(\mathbf{x}) \mid \mathbf{x}\right] = \mathbb{E}\left[y \mid \mathbf{x}\right] - \mathbb{E}\left[f^*(\mathbf{x}) \mid \mathbf{x}\right] = \mathbb{E}\left[y \mid \mathbf{x}\right] - f^*(\mathbf{x}) = 0$$

Значит математическое ожидание в (16) равно нулю — Уравнение (15) следует

Векторный случай

Возьмем случайные матрицы $\mathbf{X}$ и $\mathbf{Y}$

$$\mathbb{E}\left[\mathbf{Y} \,|\, \mathbf{X}\right] := \begin{pmatrix} \mathbb{E}\left[y_{11} \,|\, \mathbf{X}\right] & \cdots & \mathbb{E}\left[y_{1K} \,|\, \mathbf{X}\right] \\ \vdots & & \vdots \\ \mathbb{E}\left[y_{N1} \,|\, \mathbf{X}\right] & \cdots & \mathbb{E}\left[y_{NK} \,|\, \mathbf{X}\right] \end{pmatrix}$$

Мы также определим

1. $\mathrm{cov}[\mathbf{x}, \mathbf{y} \,|\, \mathbf{Z}] := \mathbb{E}\left[\mathbf{x}\mathbf{y}^{\mathsf{T}} \,|\, \mathbf{Z}\right] - \mathbb{E}\left[\mathbf{x} \,|\, \mathbf{Z}\right]\mathbb{E}\left[\mathbf{y} \,|\, \mathbf{Z}\right]^{\mathsf{T}}$
2. $\mathrm{var}[\mathbf{x} \,|\, \mathbf{Z}] := \mathbb{E}\left[\mathbf{x}\mathbf{x}^{\mathsf{T}} \,|\, \mathbf{Z}\right] - \mathbb{E}\left[\mathbf{x} \,|\, \mathbf{Z}\right]\mathbb{E}\left[\mathbf{x} \,|\, \mathbf{Z}\right]^{\mathsf{T}}$

Свойства скалярных условных математических ожиданий в факте ?? переходят к случаю с матрицами

Неполный список:

Факт. (??) Если $\mathbf{X}$, $\mathbf{Y}$ и $\mathbf{Z}$ — случайные матрицы и $\mathbf{A}$ и $\mathbf{B}$ постоянные и согласованные, то

1. $\mathbb{E}[\mathbf{Y} \mid \mathbf{Z}]^{\mathsf{T}} = \mathbb{E}[\mathbf{Y}^{\mathsf{T}} \mid \mathbf{Z}]$.
2. $\mathbb{E}[\mathbf{AX} + \mathbf{BY} \mid \mathbf{Z}] = \mathbf{A}\mathbb{E}[\mathbf{X} \mid \mathbf{Z}] + \mathbf{B}\mathbb{E}[\mathbf{Y} \mid \mathbf{Z}]$.
3. $\mathbb{E}[\mathbb{E}[\mathbf{Y} \mid \mathbf{X}]] = \mathbb{E}[\mathbf{Y}]$ и $\mathbb{E}[\mathbb{E}[\mathbf{Y} \mid \mathbf{X}, \mathbf{Z}] \mid \mathbf{X}] = \mathbb{E}[\mathbf{Y} \mid \mathbf{X}]$.
4. Если $\mathbf{X}$ и $\mathbf{Y}$ независимые, то $\mathbb{E}[\mathbf{Y} \mid \mathbf{X}] = \mathbb{E}[\mathbf{Y}]$.
5. Если $g(\mathbf{X})$ — матрица, зависимая только от $\mathbf{X}$, то
   - 5.1 $\mathbb{E}[g(\mathbf{X}) \mid \mathbf{X}] = g(\mathbf{X})$
   - 5.2 $\mathbb{E}[g(\mathbf{X})\, \mathbf{Y} \mid \mathbf{X}] = g(\mathbf{X})\mathbb{E}[\mathbf{Y} \mid \mathbf{X}]$ и $\mathbb{E}[\mathbf{Y}\, g(\mathbf{X}) \mid \mathbf{X}] = \mathbb{E}[\mathbf{Y} \mid \mathbf{X}]\, g(\mathbf{X})$