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 $\mathbf{x} = (x_1, \dots, x_N)$ — 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 $N \times N$

$$\operatorname{var}[\mathbf{x}] := \mathbb{E}\left[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^\top\right]$$

Расширяем:

$$\operatorname{var}[\mathbf{x}] = \begin{pmatrix} \mathbb{E}[(x_1-\mu_1)(x_1-\mu_1)] & \cdots & \mathbb{E}[(x_1-\mu_1)(x_N-\mu_N)] \\ \vdots & & \vdots \\ \mathbb{E}[(x_N-\mu_N)(x_1-\mu_1)] & \cdots & \mathbb{E}[(x_N-\mu_N)(x_N-\mu_N)] \end{pmatrix}$$

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. $\operatorname{var}[\mathbf{x}]$ существует и неотрицательно определена,
2. $\operatorname{var}[\mathbf{x}] = \mathbb{E}\left[\mathbf{x}\mathbf{x}^\mathsf{T}\right] - \boldsymbol{\mu}\boldsymbol{\mu}^\mathsf{T}$, и
3. $\operatorname{var}[\mathbf{A}\mathbf{x} + \mathbf{b}] = \mathbf{A}\operatorname{var}[\mathbf{x}]\mathbf{A}^\mathsf{T}$ (для любых постоянных и согласованных $\mathbf{A}, \mathbf{b}$).

Кросс-ковариация между случайными векторами $\mathbf{x}$ и $\mathbf{y}$ определяется как

$$\operatorname{cov}[\mathbf{x}, \mathbf{y}] := \mathbb{E}\left[(\mathbf{x} - \mathbb{E}[\mathbf{x}])(\mathbf{y} - \mathbb{E}[\mathbf{y}])^{\top}\right]$$

Очевидно, $\operatorname{var}[\mathbf{x}] = \operatorname{cov}[\mathbf{x}, \mathbf{x}]$

Факт. (??) Если $\mathbf{z}$ — случайный вектор в $\mathbb{R}^N$, удовлетворяющий $\mathbb{E}[\mathbf{z}\mathbf{z}^{\top}] = \mathbf{I}$ и $\mathbf{A}$ любая постоянная матрица размера $N \times N$, то

$$\mathbb{E}[\mathbf{z}^{\top}\mathbf{A}\mathbf{z}] = \operatorname{trace}\mathbf{A}$$

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 $\mathscr{B}(\mathbb{R}^N)$

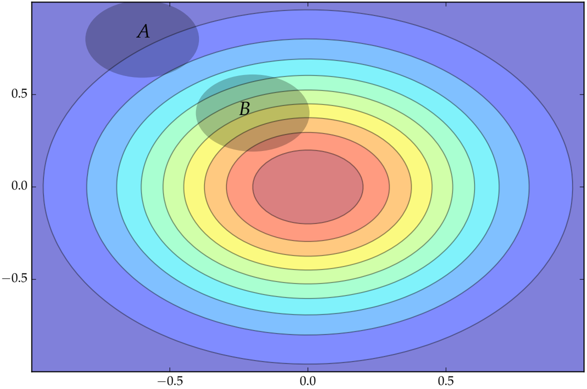

Рис.: Пример распределения и события A и B

Любое распределение P в $\mathbb{R}^N$ характеризуется функцией

$$F(\mathbf{s}) := F(s_1, \ldots, s_N) := P\left(\times_{n=1}^N (-\infty, s_n]\right) \qquad (\mathbf{s} \in \mathbb{R}^N)$$

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

$$F(\mathbf{s}_j) \to 1 \text{ при } \mathbf{s}_j \to \infty$$

$$\text{и} \quad F(s_1, \ldots, s_{nj}, \ldots, s_N) \to 0 \text{ при } s_{nj} \to -\infty$$

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**

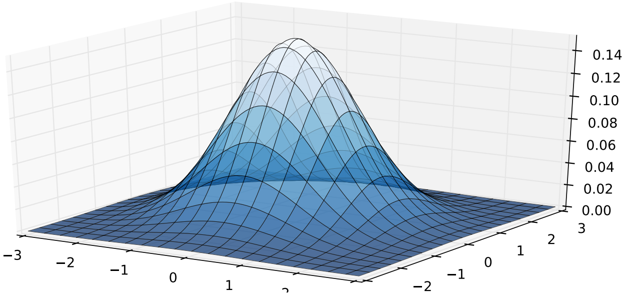

Рис.: Функция плотности двумерного стандартного нормального распределения

Распределение произведения $P_1, \ldots, P_N$ определяется следующим фактом:

Факт. (??) Возьмем распределения $P_1, \ldots, P_N$ в $\mathbb{R}$, существует единственное и определенное распределение $\mathring{P}$ в $\mathbb{R}^N$, такое что

$$\mathring{P}(B_1 \times \cdots \times B_N) = \prod_{n=1}^{N} P_n(B_n) \qquad \text{для всех } B_n \in \mathscr{B}(\mathbb{R}),\ n = 1, \ldots, N$$

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,

$$P_n(B) = P\{\mathbf{s} \in \mathbb{R}^N : \mathbf{s}^\mathsf{T}\mathbf{e}_n \in B\}$$

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 ET)

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:

$$p_1(s_1) = \int_{-\infty}^{\infty} p(s_1, s_2)\, \mathrm{d}s_2$$

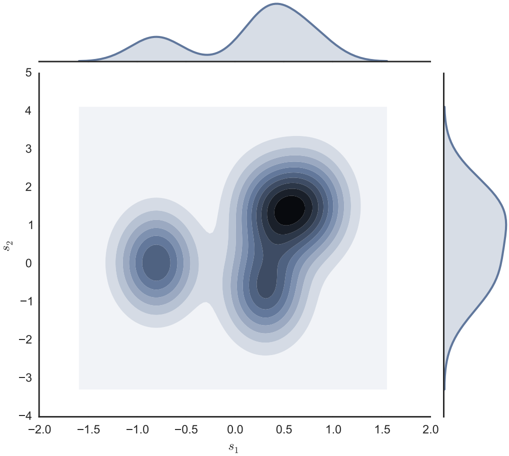

Рис.: Двумерная совместная функция плотности и две ее частных вариации

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

$$\mathbb{E}_P h := \int h(\mathbf{s}) P(d\mathbf{s}) \qquad (1)$$

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 $x_1, \dots, x_N$ **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 $B_1, \dots, B_N$, где каждый B_n — Борелевское подмножество $\mathbb{R}$

Случайные величины $x_1, \dots, x_N$ независимы, когда множества вида $\{x_1 \in B_1\}, \dots, \{x_N \in B_N\}$ 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 $\mathbf{x} = (x_1, \dots, x_N)$ и P_n — его n-ое частное

Так как $\cap_{n=1}^{N} \{x_n \in B_n\} = \{(x_1, \dots, x_N) \in B_1 \times \cdots \times B_N\}$, случайные величины $x_1, \dots, x_N$ 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 $x_1, \ldots, x_N$:

$$F(s_1, \ldots, s_N) = \prod_{n=1}^{N} F_n(s_n)$$

для всех $(s_1, \ldots, s_N) \in \mathbb{R}^N$, где F функция распределения $\mathbf{x}$ и $F_1, \ldots, F_N$ 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 $\boldsymbol{\Sigma}$ 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 $x_1, \dots, x_N$ 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 $\mathbf{x}_1, \ldots, \mathbf{x}_N$ 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 $B_1, \ldots, B_N$, 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

$$C(s_1,\dots,s_N) = \mathbb{P}\{u_1 \le s_1,\dots,u_N \le s_N\} \tag{5}$$

Где $0 \le s_n \le 1$ 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

$$F(s_1, \ldots, s_N) = C(F_1(s_1), \ldots, F_N(s_N))$$

$$(s_n \in \mathbb{R}, n = 1, \ldots, N) \quad (6)$$

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 $F_1, \dots, F_N$, 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

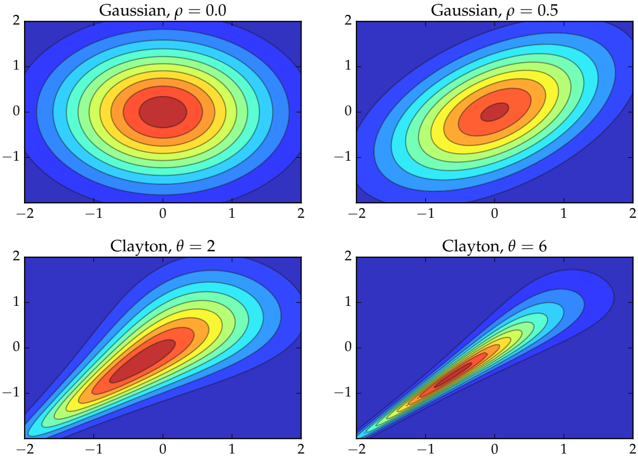

Рис.: Двумерный Гауссовская (вверху) и не-Гауссовская (внизу)

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

$$\mathcal{L}\left(\mathbf{z}^\mathsf{T}\mathbf{A}\mathbf{z}\right) = \chi^2(K) \quad \text{, где} \quad K := \operatorname{trace} \mathbf{A}$$

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

$$\langle x, y \rangle := \mathbb{E}\left[xy\right]$$

Неравенство Коши — Буняковского для случайных величин говорит нам, что $\mathbb{E}\left[xy\right]$ 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

$$L_2 := \{ \text{ все случайные величины } x \text{ в } (\Omega, \mathscr{F}, \mathbb{P}) \text{ с } \mathbb{E}\left[x^2\right] < \infty\}$$

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 ET 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

$$x, y \in Z \text{ и } \alpha, \beta \in \mathbb{R} \implies \mathbb{E}\left[\alpha x + \beta y\right] = \alpha \mathbb{E}\left[x\right] + \beta \mathbb{E}\left[y\right] = 0$$

Как и в $\mathbb{R}^N$, **ортонормированный бизис** 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 $S := \text{span}\{\mathbb{1}, x\}$ 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

$$\langle u_1, u_2 \rangle = \mathbb{E}\,[u_1 u_2] = \mathbb{E}\left[\frac{x-\mu}{\sigma_x}\right] = 0$$

Ясно, что $\|u_1\| = \|u_2\| = 1$, так что эта пара ортонормирована

Также просто показать, что $\text{span}\{u_1, u_2\} = \text{span}\{\mathbb{1}, x\}$, 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 $x \perp y \iff \text{cov}[x, y] = 0$

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 $S := \mathrm{span}\{\mathbb{1}, x\}$

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:

$$\mathrm{span}\{\mathbf{x}\} := \text{ случайные величины вида } \mathbf{x}^\mathsf{T}\mathbf{b} \text{ для некоторых } \mathbf{b} \in \mathbb{R}^K$$

Предположим, что $\mathbb{E}\left[\mathbf{x}^\mathsf{T}\mathbf{x}\right] < \infty$

Факт. (??) Если $\mathbb{E}\left[\mathbf{x}\mathbf{x}^\mathsf{T}\right]$ положительно определена, то проекция $\mathbf{P}y$ любого $y \in L_2$ на $\operatorname{span}\{\mathbf{x}\}$ определяется как

$$\hat{y} = \mathbf{x}^\mathsf{T}\mathbf{b}^* \quad \text{где} \quad \mathbf{b}^* := \mathbb{E}\left[\mathbf{x}\mathbf{x}^\mathsf{T}\right]^{-1}\mathbb{E}\left[\mathbf{x}y\right]$$

Упражнение **??** просит доказать вышеизложенный факт

Положительная определенность $\mathbb{E}\left[\mathbf{x}\mathbf{x}^\mathsf{T}\right]$ обеспечивает обратимость, значит $\mathbf{b}^*$ однозначно определено

По определению ортогональных проекций, $\mathbf{b}^*$ обязательно удовлетворяет

$$\mathbf{b}^* = \operatorname*{argmin}_{\mathbf{a} \in \mathbb{R}^K} \mathbb{E}\left[(y - \mathbf{x}^\mathsf{T}\mathbf{a})^2\right]$$

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 $\mathcal{G} := \{x_1, \dots, x_D\}$ — 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

$$z = g(x_1, \dots, x_D)$$

- 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 $\mathcal{G} = \{x_1, \ldots, x_p\}$, то мы можем взять $y = g(x_1, \ldots, x_p) = \alpha + \sum_{i=1}^{p} 0x_i$

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

$$\mathbb{E}\left[y \mid \mathcal{G}\right] := \operatorname*{argmin}_{z \in L_2(\mathcal{G})} \|y - z\| \qquad (13)$$

$\mathbb{E}\left[y \mid \mathcal{G}\right]$ — 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

$$\mathbb{E}\left[y \mid \mathcal{G}\right] = \mathbf{P}y \quad \text{, когда } \mathbf{P} := \operatorname{proj} L_2(\mathcal{G})$$

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 $\mathbb{E}[y \,|\, x_1, \ldots, x_D]$ или $\mathbb{E}[y \,|\, \mathbf{x}]$

- так же, как $\mathbb{E}[y \,|\, \mathcal{G}]$, 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 $\mathbb{E}\,[y \,|\, x] = x$. 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 $\mathbb{E}[y \mid \mathbf{x}] = f^*(\mathbf{x})$

Частная функция f^*, удовлетворяющая $f^*(\mathbf{x}) = \mathbb{E}[y \mid \mathbf{x}]$ 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

$$\mathbb{E}[y \mid x] = \int t p(t \mid x)\, \mathrm{d}t$$

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: $\mathbb{E}[\alpha x + \beta y \mid \mathcal{G}] = \alpha \mathbb{E}[x \mid \mathcal{G}] + \beta \mathbb{E}[y \mid \mathcal{G}]$
2. Если $\mathcal{G} \subset \mathcal{H}$, то $\mathbb{E}[\mathbb{E}[y \mid \mathcal{H}] \mid \mathcal{G}] \mathbb{E}[y \mid \mathcal{G}]$ и $\mathbb{E}[\mathbb{E}[y \mid \mathcal{G}]] = \mathbb{E}[y]$ (**закон повторных ожиданий**)
3. Если y iне зависима от переменных в $\mathcal{G}$, то $\mathbb{E}[y \mid \mathcal{G}] = \mathbb{E}[y]$.
4. Если y является $\mathcal{G}$-измеримой, то $\mathbb{E}[y \mid \mathcal{G}] = y$
5. Если x является $\mathcal{G}$-измеримой, то $\mathbb{E}[xy \mid \mathcal{G}] = x\mathbb{E}[y \mid \mathcal{G}]$ (**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:

$$f^*(\mathbf{x}) = \underset{g \in G}{\operatorname{argmin}} \, \mathbb{E}[(y - g(\mathbf{x}))^2] \tag{14}$$

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

$$
\begin{aligned}
(y - g(\mathbf{x}))^2 &= (y - f^*(\mathbf{x}) + f^*(\mathbf{x}) - g(\mathbf{x}))^2 \\
&= (y - f^*(\mathbf{x}))^2 + 2(y - f^*(\mathbf{x}))(f^*(\mathbf{x}) - g(\mathbf{x})) \\
&\quad + (f^*(\mathbf{x}) - g(\mathbf{x}))^2
\end{aligned}
$$

Рассмотрим математическое ожидание перемножения разных величин. Из закона повторных ожиданий:

$$
\begin{aligned}
&\mathbb{E}\{(y - f^*(\mathbf{x}))(f^*(\mathbf{x}) - g(\mathbf{x}))\} \qquad (16) \\
&= \mathbb{E}\{\,\mathbb{E}[(y - f^*(\mathbf{x}))(f^*(\mathbf{x}) - g(\mathbf{x})) \mid \mathbf{x}]\,\}
\end{aligned}
$$

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. $\mathrm{cov}[\mathbf{x}, \mathbf{y} \,|\, \mathbf{Z}] := \mathbb{E}\left[\mathbf{x}\mathbf{y}^\intercal \,|\, \mathbf{Z}\right] - \mathbb{E}\left[\mathbf{x} \,|\, \mathbf{Z}\right]\mathbb{E}\left[\mathbf{y} \,|\, \mathbf{Z}\right]^\intercal$
2. $\mathrm{var}[\mathbf{x} \,|\, \mathbf{Z}] := \mathbb{E}\left[\mathbf{x}\mathbf{x}^\intercal \,|\, \mathbf{Z}\right] - \mathbb{E}\left[\mathbf{x} \,|\, \mathbf{Z}\right]\mathbb{E}\left[\mathbf{x} \,|\, \mathbf{Z}\right]^\intercal$

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. $\mathbb{E}\left[\mathbf{Y} \mid \mathbf{Z}\right]^{\mathsf{T}} = \mathbb{E}\left[\mathbf{Y}^{\mathsf{T}} \mid \mathbf{Z}\right]$.
2. $\mathbb{E}\left[\mathbf{A}\mathbf{X} + \mathbf{B}\mathbf{Y} \mid \mathbf{Z}\right] = \mathbf{A}\mathbb{E}\left[\mathbf{X} \mid \mathbf{Z}\right] + \mathbf{B}\mathbb{E}\left[\mathbf{Y} \mid \mathbf{Z}\right]$.
3. $\mathbb{E}\left[\mathbb{E}\left[\mathbf{Y} \mid \mathbf{X}\right]\right] = \mathbb{E}\left[\mathbf{Y}\right]$ и $\mathbb{E}\left[\mathbb{E}\left[\mathbf{Y} \mid \mathbf{X}, \mathbf{Z}\right] \mid \mathbf{X}\right] = \mathbb{E}\left[\mathbf{Y} \mid \mathbf{X}\right]$.
4. Если $\mathbf{X}$ и $\mathbf{Y}$ независимые, то $\mathbb{E}\left[\mathbf{Y} \mid \mathbf{X}\right] = \mathbb{E}\left[\mathbf{Y}\right]$.
5. Если $g(\mathbf{X})$ — матрица, зависимая только от $\mathbf{X}$, то
   5.1 $\mathbb{E}\left[g(\mathbf{X}) \mid \mathbf{X}\right] = g(\mathbf{X})$
   5.2 $\mathbb{E}\left[g(\mathbf{X})\, \mathbf{Y} \mid \mathbf{X}\right] = g(\mathbf{X})\mathbb{E}\left[\mathbf{Y} \mid \mathbf{X}\right]$ и
   $\mathbb{E}\left[\mathbf{Y}\, g(\mathbf{X}) \mid \mathbf{X}\right] = \mathbb{E}\left[\mathbf{Y} \mid \mathbf{X}\right] g(\mathbf{X})$