Байесовские методы. Лекция 1.
Факты из теории вероятностей

Целищев М.А.

МГУ им. М. В. Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра математической статистики

весна 2021

Список литературы

- J. Kruschke. Doing Bayesian Data Analysis, Second Edition. A Tutorial with R, JAGS, and Stan. Academic Press, 2014.
- R. McElreath. Statistical Rethinking. A Bayesian Course with Examples in R and Stan. Chapman and Hall CRC, 2015.
- O. Martin. Bayesian Analysis with Python. Introduction to Statistical Modeling and Probabilistic Programming using PyMC3 and ArviZ, Second Edition. Packt, 2018.
- K. P. Murphy. Machine Learning: A Probabilistic Perspective. The MIT Press, 2012.
- A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, D. Rubin. Bayesian Data Analysis, Third Edition. CRC Press, 2013.
- S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng. Handbook of Markov Chain Monte Carlo. Chapman & Hall/CRC, 2011.

Вероятностное пространство

Тройку $(\Omega, \mathcal{F}, \mathbf{P})$ называют *вероятностным пространством*, где

- Ω — непустое множество произвольной структуры,
- $\mathcal{F}$ — σ-алгебра над Ω, то есть класс подмножеств множества Ω, замкнутый относительно операций разности, пересечения, объединения множеств (в счётном числе), а также содержащий пустое множество $\varnothing$ и само Ω; элементы $\mathcal{F}$ называют *событиями*,
- $\mathbf{P}$ — вероятностная мера на $\mathcal{F}$, то есть такая функция $\mathbf{P}\colon \mathcal{F} \to [0,1]$, что $\mathbf{P}(\varnothing) = 0$, $\mathbf{P}(\Omega) = 1$ и

$$\mathbf{P}\left(\bigsqcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbf{P}\left(A_n\right)$$

  для произвольной последовательности *несовместных* (т.е. попарно непересекающихся) событий $A_n \in \mathcal{F}$.

Условная вероятность, независимость событий

Как определяют условную вероятность $\mathbf{P}(A|B)$? Хотят, чтобы

1. Если $A_1 \cap A_2 = \varnothing$, то $\mathbf{P}(A_1 \sqcup A_2|B) = \mathbf{P}(A_1|B) + \mathbf{P}(A_2|B)$.
2. $\mathbf{P}(B|B) = 1$.
3. Если $A \cap B = \varnothing$, то $\mathbf{P}(A|B) = 0$.
4. Если $A_1, A_2 \subset B$, то $\dfrac{\mathbf{P}(A_1|B)}{\mathbf{P}(A_2|B)} = \dfrac{\mathbf{P}(A_1)}{\mathbf{P}(A_2)}$.

Пусть $\mathbf{P}(B) > 0$. Из приведённых четырёх условий вытекает:

$$\mathbf{P}(A|B) = \mathbf{P}(AB \sqcup AB^c|B) = \mathbf{P}(AB|B) + \mathbf{P}(AB^c|B) = \frac{\mathbf{P}(AB|B)}{\mathbf{P}(B|B)} + 0 = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

События A и B называются *стохастически независимыми*, если $\mathbf{P}(A|B) = \mathbf{P}(A)$, или, $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$.

События $A_1, A_2, \ldots \in \mathcal{F}$ называются *стохастически независимыми в совокупности*, если $\forall n \in \mathbb{N} \quad \forall\, 1 \leqslant i_1 < i_2 < \ldots < i_n$ выполнено

$$\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_n}) = \mathbf{P}(A_{i_1}) \ldots \mathbf{P}(A_{i_n}).$$

Формула умножения вероятностей (цепное правило)

Пусть $A_1, \ldots, A_n \in \mathcal{F}$. Тогда

$$\mathbf{P}(A_1 \ldots A_n) = \mathbf{P}(A_1) \cdot \mathbf{P}(A_2|A_1) \cdot \mathbf{P}(A_3|A_1 A_2) \ldots \mathbf{P}(A_n|A_1 \ldots A_{n-1})$$

Формула полной вероятности

Пусть $A \in \mathcal{F}$ и $B_1, \ldots, B_n$ — *полный набор событий*, т.е. $B_i \in \mathcal{F}$, $B_i \cap B_j = \varnothing$ при $i \neq j$ и $\bigsqcup\limits_{i=1}^{n} B_i = \Omega$. Тогда

$$\mathbf{P}(A) = \sum_{i=1}^{n} \mathbf{P}(B_i) \cdot \mathbf{P}(A|B_i).$$

$\mathbf{P}(A|B) = \dfrac{\mathbf{P}(AB)}{\mathbf{P}(B)}$. Событие A произошло раньше, чем B ???

Формула Байеса

Пусть $A \in \mathcal{F}$ и $B_1, \ldots, B_n$ — полная группа событий. Тогда

$$\mathbf{P}(B_j|A) = \frac{\mathbf{P}(AB_j)}{\mathbf{P}(A)} = \frac{\mathbf{P}(B_j) \cdot \mathbf{P}(A|B_j)}{\sum\limits_{i=1}^{n} \mathbf{P}(B_i) \cdot \mathbf{P}(A|B_i)}, \quad j = 1 \ldots n.$$

Пример

В первой урне 3 белых шара и 4 чёрных, а во второй — 2 белых и 2 чёрных. Наугад выбирают урну и тянут из неё шар.

A=\{вытянули белый шар\}, B_i=\{выбрали i-ую урну\}, $i = 1, 2$.

1) Какова вероятность того, что выбрали первую урну и вытащили белый шар?
$\mathbf{P}(B_1 A) = \mathbf{P}(B_1)\mathbf{P}(A|B_1) = \frac{1}{2} \cdot \frac{3}{7}$

2) Какова вероятность того, что вытащили белый шар?
$\mathbf{P}(A) = \mathbf{P}(B_1)\mathbf{P}(A|B_1) + \mathbf{P}(B_2)\mathbf{P}(A|B_2) = \frac{1}{2} \cdot \frac{3}{7} + \frac{1}{2} \cdot \frac{2}{4}$

3) Вслепую провели эксперимент и оказалось, что шар белый. Какова вероятность, что тянули из первой урны?

$$\mathbf{P}(B_1|A) = \frac{\mathbf{P}(B_1)\mathbf{P}(A|B_1)}{\mathbf{P}(B_1)\mathbf{P}(A|B_1) + \mathbf{P}(B_2)\mathbf{P}(A|B_2)} = \frac{\frac{1}{2} \cdot \frac{3}{7}}{\frac{1}{2} \cdot \frac{3}{7} + \frac{1}{2} \cdot \frac{2}{4}}$$

Парадокс Монти Холла

Игроку предлагается выбрать одну из трёх дверей, за одной из которых автомобиль, а за другими — ничего. После выбора игрока ведущий открывает из двух оставшихся дверей ту, за которой заведомо ничего нет, и спрашивает игрока, не хочет ли тот изменить свой начальный выбор. Что делать игроку?

Случайные величины

Функцию $\xi\colon \Omega \to \mathbb{R}$ называют *случайной величиной*, если

$$\{\,\omega \in \Omega \,:\, \xi(\omega) < a\,\} \in \mathcal{F} \ \text{ для каждого } \ a \in \mathbb{R}.$$

Функцией распределения (ф.р.) случайной величины ξ называют

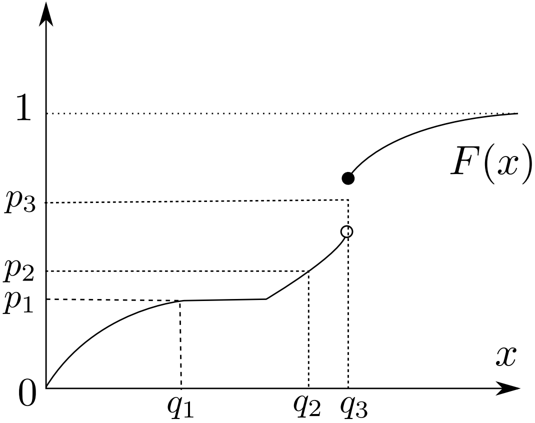
$$F_\xi(x) = \mathbf{P}(\xi < x), \quad x \in \mathbb{R}.$$

Свойства ф.р.:

- $F_\xi(-\infty) = 0$, $F_\xi(+\infty) = 1$,
- монотонность: $F_\xi(x) \leqslant F_\xi(y)$ при $x < y$,
- непрерывность слева: $\lim\limits_{x \uparrow x_0} F_\xi(x) = F_\xi(x_0)$ для всех $x_0 \in \mathbb{R}$.

Квантилью уровня $\alpha \in (0, 1)$ распределения случайной величины ξ называется

$$q_\xi(\alpha) = F_\xi^{-1}(\alpha)$$



С.в. ξ называется *дискретной*, если с вероятностью 1 она принимает значения из не более чем счётного множества $X = \{x_1, x_2, \dots\}$.

Распределение дискретной с.в. задаётся множеством X и $\mathbf{P}(\xi = x_k)$.

Примеры дискретных распределений:

Бернуллиевское $\mathbf{Be}(p)$: $\mathbf{P}(\xi = k) = p^k\,(1-p)^{(1-k)}, \quad k = 0, 1$

Биномиальное $\mathbf{Bi}(n,p)$: $\mathbf{P}(\xi = k) = C_n^k\,p^k(1-p)^{n-k}, \quad k = 0 \dots n$

Геометрическое $\mathbf{Geom}(p)$: $\mathbf{P}(\xi = k) = p\,(1-p)^k, \quad k \in 0, 1, 2, \dots$

Пуассоновское $\mathbf{Pois}(\lambda)$: $\mathbf{P}(\xi = k) = \dfrac{\lambda^k}{k!} e^{-\lambda}, \quad k \in 0, 1, 2, \dots$

https://shiny.ilykei.com/misha/cmc/Binomial/
https://shiny.ilykei.com/misha/cmc/Geometric/
https://shiny.ilykei.com/misha/cmc/Poisson/

Абсолютно непрерывные с.в.

С.в. ξ имеет абсолютно непрерывное распределение, если существует такая функция $p\colon \mathbb{R} \to \mathbb{R}^+$, называемая *плотностью*, что

$$F_\xi(x) = \int_{-\infty}^{x} p(y)\, dy.$$

Примеры абсолютно непрерывных распределений:

Равномерное $\mathbf{U}(a, b)$: $p(x) = \frac{1}{b-a}$, $x \in (a, b)$,

Нормальное $\mathcal{N}(a, \sigma^2)$: $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)$, $x \in \mathbb{R}$,

Показательное $\mathbf{Exp}(\lambda)$: $p(x) = \lambda e^{-\lambda x}$, $x > 0$,

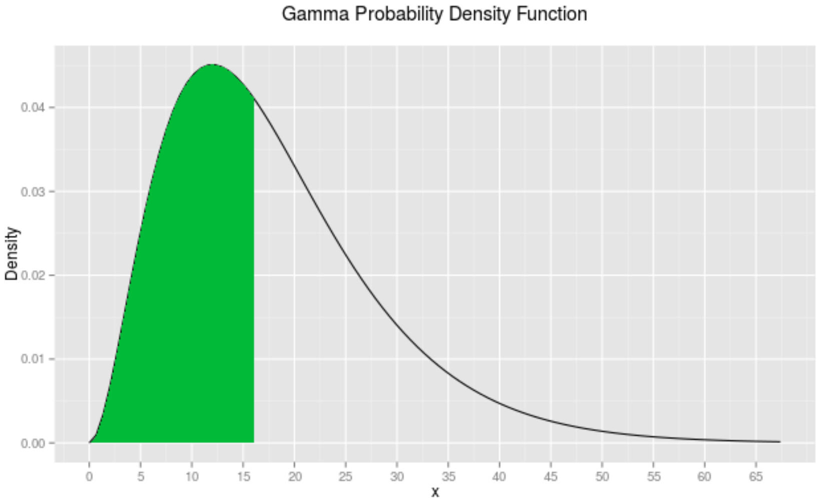
Гамма $\mathbf{\Gamma}(\alpha, \beta)$: $p(x) = \frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)}$, $x > 0$,

Коши $\mathbf{Cauchy}(a, b)$: $p(x) = \frac{b}{\pi(b^2 + (x-a)^2)}$, $x \in \mathbb{R}$.

https://ismay.shinyapps.io/ProbApp/

https://shiny.ilykei.com/misha/cmc/Cauchy/

Квантили в абс. непрерывном случае



Математическое ожидание

Математическим ожиданием дискретной с.в. ξ, принимающей значения $x_1, x_2, \ldots$ называется число

$$\mathbf{E}\xi = \sum_n x_n \mathbf{P}(\xi = x_n),$$

если этот ряд сходится абсолютно (т.е. ряд из модулей сходится).

Математическим ожиданием абс. непрерывной с.в. ξ с плотностью p называется число

$$\mathbf{E}\xi = \int_{-\infty}^{+\infty} x p(x)\, dx.$$

если этот интеграл сходится абсолютно.

Математическое ожидание имеет физический смысл *центра*, вокруг которого сосредоточены возможные значения с.в.

Свойства мат. ожидания

- $\mathbf{E}(a\xi + b\eta) = a\,\mathbf{E}\xi + b\,\mathbf{E}\eta$ для всех $a, b \in \mathbb{R}$
- $\mathbf{E}c = c$ для всех $c \in \mathbb{R}$
- $\xi \geqslant \eta \Rightarrow \mathbf{E}\xi \geqslant \mathbf{E}\eta$
- $|\mathbf{E}\xi| \leqslant \mathbf{E}|\xi|$
- Если существует $\mathbf{E}\xi^m$, то существует $\mathbf{E}\xi^k$ для всех $m > k > 0$

Число $\mathbf{E}\xi^k$ называют k-ым моментом с.в. ξ.

Пусть $g\colon \mathbb{R} \to \mathbb{R}$. Как посчитать $\mathbf{E}g(\xi)$?

в дискр. случае: $\mathbf{E}g(\xi) = \sum\limits_n g(x_n)\mathbf{P}(\xi = x_n)$,

в абс. непр. случае: $\mathbf{E}g(\xi) = \int\limits_{-\infty}^{+\infty} g(x)p(x)\,dx$,

Дисперсия случайной величины

Дисперсией с.в. ξ называется $\mathbf{D}\xi = \mathbf{E}\left(\xi - \mathbf{E}\xi\right)^2$.

Дисперсия характеризует степень разброса случайной величины относительно своего среднего.

Свойства дисперсии:

- $\mathbf{D}\xi \geqslant 0$,
- $\mathbf{D}\xi = 0 \iff \xi$ вырождена, т.е. $\mathbf{P}(\xi = a) = 1$ для некоторого $a \in \mathbb{R}$,
- $\mathbf{D}(b\xi + a) = b^2\mathbf{D}\xi$ для всех $a, b \in \mathbb{R}$.

Стандартным отклонением с.в. ξ называется $\sigma(\xi) = \sqrt{\mathbf{D}\xi}$.

Случайные векторы

Случайным вектором называется вектор из случайных величин. Совместной функцией распределения случайного вектора $\xi = (\xi_1, \ldots, \xi_n)$ называется

$$F_\xi(x_1, \ldots, x_n) = \mathbf{P}(\xi_1 < x_1, \ldots, \xi_n < x_n), \quad x_1, \ldots, x_n \in \mathbb{R}.$$

Случайный вектор $\xi = (\xi_1, \ldots, \xi_n)$ имеет *дискретное* распределение, если все его компоненты — дискретные с.в.
Распределение дискретного случайного вектора задаётся вероятностями

$$\mathbf{P}(\xi = x) \equiv \mathbf{P}(\xi_1 = x_1, \ldots, \xi_n = x_n), \quad x \in \mathbb{R}^n.$$

Случайный вектор ξ имеет *абсолютно непрерывное* распределение, если задана функция $p\colon \mathbb{R}^n \to \mathbb{R}^+$, называемая плотностью, т.ч.

$$F_\xi(x) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_n} p(y_1, \ldots, y_n)\, dy_1 \ldots dy_n, \quad x \in \mathbb{R}^n.$$

Пусть $\xi = (\xi_1, \ldots, \xi_n)$ — случайный вектор и $g\colon \mathbb{R}^n \to \mathbb{R}$. Как посчитать $\mathbf{E}g(\xi)$?

в дискр. случае: $\mathbf{E}g(\xi) = \sum\limits_{x} g(x) \cdot \mathbf{P}(\xi = x)$,

в абс. непр. случае: $\mathbf{E}g(\xi) = \int\limits_{-\infty}^{+\infty} \ldots \int\limits_{-\infty}^{+\infty} g(x)p(x)\, dx$,

Независимость с.в.

Случайные величины $\xi_1, \ldots, \xi_n$ называются *независимыми в совокупности*, если

$$\mathbf{P}(\xi_1 \in B_1, \ldots, \xi_n \in B_n) = \mathbf{P}(\xi_1 \in B_1) \cdot \ldots \cdot \mathbf{P}(\xi_n \in B_n), \quad \forall B_1, \ldots, B_n$$

Это равносильно тому, что

$$F_{\xi_1,\ldots,\xi_n}(x) = F_{\xi_1}(x_1) \ldots F_{\xi_n}(x_n) \text{ для всех } \; x \in \mathbb{R}^n,$$

где F_{ξ_k} — функция распределения случайной величины ξ_k (называемая также *маргинальной* ф.р. компоненты ξ_k случайного вектора ξ). А в абс. непр. случае:

$$p(x_1, \ldots, x_n) = p(x_1) \cdot \ldots \cdot p(x_n) \text{ для всех } \; x_1, \ldots, x_n \in \mathbb{R}.$$

Lemma

Если ξ, η — независимы, то $f(\xi)$ и $g(\eta)$ — тоже независимы.

Lemma

Если ξ, η — независимы, то $\mathbf{E}\xi\eta = \mathbf{E}\xi \cdot \mathbf{E}\eta$.

Ковариация с.в.

$$\mathrm{cov}(\xi, \eta) = \mathbf{E}(\xi - \mathbf{E}\xi)(\eta - \mathbf{E}\eta)$$

Ковариация показывает степень *линейной* зависимости между случайными величинами.

Свойства ковариации:

- $\mathrm{cov}(\xi, \xi) = \mathbf{D}\xi$
- Если ξ и η независимы, то $\mathrm{cov}(\xi, \eta) = 0$ (обратное в общем случае неверно)
- $\mathbf{D}(\xi \pm \eta) = \mathbf{D}\xi + \mathbf{D}\eta \pm 2\,\mathrm{cov}(\xi, \eta)$
- $|\,\mathrm{cov}(\xi, \eta)| \leqslant \sqrt{\mathbf{D}\xi \cdot \mathbf{D}\eta}$

Коэффициентом корреляции между с.в. ξ и η называют

$$\mathrm{corr}(\xi, \eta) = \frac{\mathrm{cov}(\xi, \eta)}{\sqrt{\mathbf{D}\xi}\sqrt{\mathbf{D}\eta}} \in [-1, 1]$$

Если $\mathrm{corr}(\xi, \eta) = \pm 1$, то ξ и η линейно зависимы, то есть
$\xi = \pm b\eta + a$ почти всюду (п.в.) для некоторых $a \in \mathbb{R}, b > 0$.

Условные распределения

Пусть совместное распределение дискр. (ξ, η) задано таблицей,

$\xi \setminus \eta$	y_1	$\dots$	y_j	$\dots$
x_1	p_{11}	$\dots$	p_{1j}	$\dots$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_i	p_{i1}	$\dots$	p_{ij}	$\dots$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$

где $p_{ij} = \mathbf{P}(\xi = x_i, \eta = y_j)$.
Заметим, что
$\mathbf{P}(\xi = x_i) = \sum\limits_j p_{ij}$, $\mathbf{P}(\eta = y_j) = \sum\limits_i p_{ij}$

Условным распределением с.в. ξ при условии $\eta = y_j$ называется

$$\mathbf{P}(\xi = x_i | \eta = y_j) = \frac{\mathbf{P}(\xi = x_i, \eta = y_j)}{\mathbf{P}(\eta = y_j)} = \frac{p_{ij}}{\sum\limits_k p_{kj}}.$$

Условным мат. ожиданием с.в. ξ при условии $\eta = y_j$ называется

$$\mathbf{E}(\xi | \eta = y_j) = \sum_i x_i \cdot \mathbf{P}(\xi = x_i | \eta = y_j).$$

Формула Байеса для с.в.: $\mathbf{P}(\eta = y_j | \xi = x_i) = \dfrac{\mathbf{P}(\eta = y_j) \cdot \mathbf{P}(\xi = x_i | \eta = y_j)}{\sum\limits_m \mathbf{P}(\eta = y_m) \cdot \mathbf{P}(\xi = x_i | \eta = y_m)}.$

Условные распределения (абс. непр. случай)

Пусть совместное распределение абс. непр. (ξ, η) задано плотностью $p(x, y)$. Заметим, что маргинальные плотности считаются как

$$p(x) = \int_{-\infty}^{+\infty} p(x, y)\, dy, \quad p(y) = \int_{-\infty}^{+\infty} p(x, y)\, dx.$$

$$\mathbf{P}(\xi < x | \eta = y) = \frac{\mathbf{P}(\xi < x, \eta = y)}{\mathbf{P}(\eta = y)} = \frac{0}{0} = ???$$  — плохо.

Другой подход (в предположении, что $p(x, y)$ непрерывна):

$$\mathbf{P}(\xi < x | \eta = y) := \lim_{\delta \to +0} \frac{\mathbf{P}\left(\xi < x,\ \eta \in [y, y + \delta)\right)}{\mathbf{P}\left(\eta \in [y, y + \delta)\right)} =$$

$$= \lim_{\delta \to +0} \frac{\int\limits_{-\infty}^{x} du \int\limits_{y}^{y+\delta} p(u, v)\, dv}{\int\limits_{-\infty}^{+\infty} du \int\limits_{y}^{y+\delta} p(u, v)\, dv} = \lim_{\delta \to +0} \frac{\int\limits_{-\infty}^{x} \delta p(u, y)\, du}{\int\limits_{-\infty}^{+\infty} \delta p(u, y)\, du} = \int\limits_{-\infty}^{x} \frac{p(u, y)}{p(y)} du.$$

Условные распределения (абс. непр. случай)

Условная плотность распределения с.в. ξ при условии $\eta = y$:

$$p(x|y) \equiv p(x|\eta = y) = \frac{p(x,y)}{p(y)}$$

При каждом y условная плотность $p(\cdot|y)$ ведёт себя как обычная плотность.

Условным мат. ожиданием с.в. ξ при условии $\eta = y$ называется

$$\mathbf{E}(\xi|\eta = y) = \int\limits_{-\infty}^{+\infty} x p(x|\eta = y)\, dx.$$

Формула Байеса для абс. непр. с.в.:

$$p(y|\xi = x) = \frac{p(y)p(x|y)}{p(x)} = \frac{p(y)p(x|y)}{\int\limits_{-\infty}^{+\infty} p(\hat{y})p(x|\hat{y})\, d\hat{y}}.$$

Далее считаем, что $p(x, y, \ldots)$ — либо совместная плотность с.в. $X, Y, \ldots$, либо $p(x, y, \ldots) = \mathbf{P}(X{=}x, Y{=}y, \ldots)$ в дискретном случае.

Формула умножения вероятностей/плотностей (chain rule):

$$p(x, y, z) = p(x) \cdot p(y, z|x) = p(x) \cdot p(y|x) \cdot p(z|x, y).$$

Формула полной вероятности/плотности:

$$p(x, y) = \int p(x, y, z)\, dz = \int p(x, y|z)\, p(z)\, dz.$$

Формула Байеса:

$$p(y|x) = \frac{p(x, y)}{p(x)} = \frac{p(x, y)}{\int p(x, \hat{y})\, d\hat{y}} = \frac{p(x|y)\, p(y)}{\int p(x|\hat{y})\, p(\hat{y})\, d\hat{y}}.$$

Многомерное нормальное распределение

Случайный вектор $\xi = (\xi_1, \ldots, \xi_d)$ имеет *невырожденное многомерное нормальное распределение* с вектором средних $\mu \in \mathbb{R}^d$ и положительно определённой симметричной ковариационной матрицей $\Sigma \in \mathbb{R}^{d\times d}$, если его плотность представима в виде

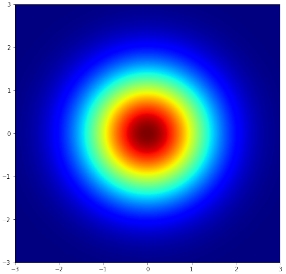
$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)\right), \quad \mathbf{x} \in \mathbb{R}^d.$$

Здесь под векторами понимаются вектор-столбцы, знак T обозначает транспонирование вектора/матрицы, $\det \Sigma$ — определитель матрицы Σ, а Σ^{-1} — матрица, обратная к матрице Σ.

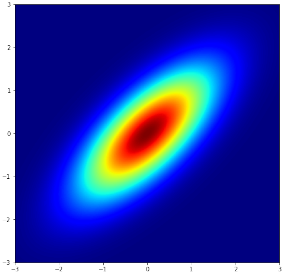
Многомерное нормальное распределение

$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu)\right), \quad \mathbf{x} \in \mathbb{R}^d.$$

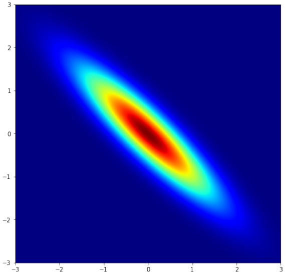
Ниже представлена тепловая карта плотности двумерного нормального распределения с центром $\mu = (0,0)^T$ и разными ковариационными матрицами Σ.



$$\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$



$$\Sigma = \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix}$$



$$\Sigma = \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}$$

Многомерное нормальное распределение

Если d-мерный вектор $\xi \sim \mathcal{N}(\mu, \Sigma)$, то

- каждая его компонента $\xi_j \sim \mathcal{N}(\mu_j, (\Sigma)_{jj}), \;\; j = 1 \dots d$,
- $(\Sigma)_{ij} = \mathrm{cov}(\xi_i, \xi_j), \qquad i, j = 1 \dots n$,
- $(\Sigma)_{ij} = 0 \iff$ компоненты ξ_i и ξ_j независимы.

Если $A \in \mathbb{R}^{m \times d}$, то случайный вектор $\eta = A\xi$ также имеет многомерное нормальное распределение:

$$\eta \sim \mathcal{N}\left(A\mu, A\Sigma A^T\right).$$

Задача. Пусть $\xi = (\xi_1, \xi_2) \sim \mathcal{N}\left(\mu = \begin{pmatrix}\mu_1 \\ \mu_2\end{pmatrix},\ \Sigma = \begin{pmatrix}\sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ \rho\,\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$,

где $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 > 0$, $\rho \in (-1, 1)$.
Найти условное распределение $(\xi_1 | \xi_2)$.

Решение. Совместная плотность $\xi = (\xi_1, \xi_2)$ задаётся формулой

$$p(\mathbf{x}) = \frac{1}{2\pi\sqrt{\det\Sigma}}\ \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)\right), \quad \mathbf{x} \in \mathbb{R}^2,$$

причём $\det\Sigma = \sigma_1^2\sigma_2^2(1-\rho^2)$ и $\Sigma^{-1} = \dfrac{1}{\det\Sigma}\begin{pmatrix}\sigma_2^2 & -\rho\,\sigma_1\sigma_2 \\ -\rho\,\sigma_1\sigma_2 & \sigma_1^2\end{pmatrix}$.

Иными словами, совместная плотность равна $p(x_1, x_2) =$

$$= \frac{1}{2\pi\sqrt{\det\Sigma}}\ \exp\left(-\frac{\sigma_2^2(x_1-\mu_1)^2 - 2\rho\sigma_1\sigma_2(x_1-\mu_1)(x_2-\mu_2) + \sigma_1^2(x_2-\mu_2)^2}{2\det\Sigma}\right) =$$

$$= C\cdot\exp\left(-\ \frac{\left[\sigma_2(x_1-\mu_1) - \rho\sigma_1(x_2-\mu_2)\right]^2 + (1-\rho^2)\sigma_1^2(x_2-\mu_2)^2}{2\det\Sigma}\right)$$

$$p(x_1, x_2) = C \cdot \exp\left(-\ \frac{\big[\sigma_2(x_1-\mu_1) - \rho\sigma_1(x_2-\mu_2)\big]^2 + (1-\rho^2)\sigma_1^2(x_2-\mu_2)^2}{2\det\Sigma}\right)$$

Найдём маргинальную плотность второй компоненты ξ_2:

$$p(x_2) = \int_{-\infty}^{+\infty} p(x_1, x_2)\, dx_1 = \widehat{C} \cdot \exp\left(-\ \frac{(1-\rho^2)\sigma_1^2(x_2-\mu_2)^2}{2\det\Sigma}\right),$$

где $\widehat{C} = C \int\limits_{-\infty}^{+\infty} \exp\left(-\ \frac{\big[\sigma_2(x_1-\mu_1) - \rho\sigma_1(x_2-\mu_2)\big]^2}{2\det\Sigma}\right)\, dx_1$,

причём линейной заменой $z = \sigma_2(x_1-\mu_1) - \rho\sigma_1(x_2-\mu_2)$ показывается, что $\widehat{C}$ действительно константа (не зависит от x_2).

Итак, вспоминая что $\det\Sigma = \sigma_1^2\sigma_2^2(1-\rho^2)$, имеем

$$p(x_2) = \widehat{C} \cdot \exp\left(-\ \frac{(x_2-\mu_2)^2}{2\sigma_2^2}\right), \quad x_2 \in \mathbb{R}.$$

Можно было бы, конечно, посчитать $\widehat{C}$ честно, но зачем? Это ведь нормировочная константа, $\xi_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$, и потому $\widehat{C} = 1/\left(\sqrt{2\pi\sigma_2^2}\right)$.

Вернёмся к нахождению $p(x_1 \mid \xi_2 = x_2)$.

$$p(x_1, x_2) = C \cdot \exp\left(- \frac{\big[\sigma_2(x_1 - \mu_1) - \rho\sigma_1(x_2 - \mu_2)\big]^2 + (1 - \rho^2)\sigma_1^2(x_2 - \mu_2)^2}{2 \det \Sigma} \right),$$

$$p(x_2) = \widehat{C} \cdot \exp\left(- \frac{(x_2 - \mu_2)^2}{2\sigma_2^2} \right),$$

и потому

$$p(x_1 \mid \xi_2 = x_2) = \frac{p(x_1, x_2)}{p(x_2)} = \widetilde{C} \cdot \exp\left(- \frac{\big[\sigma_2(x_1 - \mu_1) - \rho\sigma_1(x_2 - \mu_2)\big]^2}{2 \det \Sigma} \right),$$

где $\widetilde{C} = C/\widehat{C}$, и, опять вспоминая, что $\det \Sigma = \sigma_1^2\sigma_2^2(1 - \rho^2)$, получим

$$p(x_1 \mid \xi_2 = x_2) = \widetilde{C} \cdot \exp\left(- \frac{\Big[x_1 - \mu_1 - \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2)\Big]^2}{2\,\sigma_1^2(1 - \rho^2)} \right),$$

то есть $(\xi_1 \mid \xi_2 = x_2) \sim \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),\ \ \sigma_1^2(1 - \rho^2)\right)$

или, что тоже самое, $(\xi_1 \mid \xi_2) \sim \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(\xi_2 - \mu_2),\ \ \sigma_1^2(1 - \rho^2)\right)$.

Итак, если $\xi = (\xi_1, \xi_2) \sim \mathcal{N}\left(\mu = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix},\ \Sigma = \begin{pmatrix}\sigma_1^2 & \rho\,\sigma_1\sigma_2\\ \rho\,\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$, то

$$(\xi_1 \mid \xi_2 = x_2) \sim \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2),\ \ \sigma_1^2(1-\rho^2)\right).$$

Иными словами, если по какой-то причине стало известно, что $\xi_2 = x_2$, то лучшим прогнозом для ξ_1 будет

$$\mathbf{E}\left(\xi_1 \,|\, \xi_2 = x_2\right) = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2).$$

Выражение $\mathbf{E}\left(\xi_1 \,|\, \xi_2 = x_2\right)$ также называется *регрессией* ξ_1 *по* ξ_2.
Бывает ещё и множественная регрессия:

$$\mathbf{E}\left(\xi_1 \,|\, \xi_2 = x_2, \ldots, \xi_m = x_m\right),$$

но это уже отдельная тема...

Виды сходимости последовательностей с.в.

Говорят, что последовательность с.в. $\xi_1, \xi_2, \ldots$ сходится к с.в. ξ

- **почти всюду (п.в.)**, если $\mathbf{P}(\omega\colon \xi_n(\omega) \to \xi(\omega)) = 1$
- **по вероятности**, если $\forall \varepsilon > 0$ имеет место $\mathbf{P}(|\xi_n - \xi| > \varepsilon) \to 0$ при $n \to \infty$
- **по распределению**, если $F_{\xi_n}(x) \to F_\xi(x)$ при $n \to \infty$ для всех точек x, в которых ф.р. F_ξ непрерывна.
- **слабо**, если $\mathbf{E}f(\xi_n) \to \mathbf{E}f(\xi)$ при $n \to \infty$ для каждой непрерывной ограниченной $f\colon \mathbb{R} \to \mathbb{R}$.

Сходимость п.в. ($\xi_n \xrightarrow{\text{п.в.}} \xi$) влечёт сходимость по вероятности ($\xi_n \xrightarrow{\mathbf{P}} \xi$), а сходимость по вероятности влечёт сходимость по распределению ($\xi_n \xrightarrow{d} \xi$), которая эквивалентна *слабой* сходимости.

Характеристическая функция с.в.

Характеристической функцией (х.ф.) с.в. ξ называют

$$\varphi_\xi(t) = \mathbf{E}e^{it\xi},$$

где i — мнимая единица, а $e^{ix} = \cos x + i \sin x$ для $x \in \mathbb{R}$ (ф-ла Эйлера).

Свойства х.ф.:

- Если ξ и η независимы, то $\varphi_{\xi+\eta}(t) = \varphi_\xi(t)\varphi_\eta(t)$, $t \in \mathbb{R}$.
- Х.ф. однозначно определяет распределение случайной величины, т.е. если $\varphi_\xi(t) = \varphi_\eta(t)$ для всех $t \in \mathbb{R}$, то $\xi \stackrel{d}{=} \eta$ (т.е. $F_\xi(x) = F_\eta(x)$ для всех $x \in \mathbb{R}$)
- $\xi_n \xrightarrow{d} \xi \iff \varphi_{\xi_n}(t) \to \varphi_\xi(t)\ \forall t \in \mathbb{R}$.

Предельные теоремы, ЗБЧ

Обозначение: i.i.d. — независимые одинаково распределённые с.в.

Theorem (Усиленный закон больших чисел для i.i.d.)

Пусть $\xi_1, \xi_2, \ldots$ *— i.i.d. с конечным* $\mathbf{E}\xi_k = a$. *Тогда, если обозначить* $S_n = \xi_1 + \ldots + \xi_n$, *то*

$$\frac{S_n}{n} \xrightarrow{\text{п.в.}} a \quad \text{при} \quad n \to \infty.$$

В частности, ЗБЧ раскрывает физический смысл вероятности события: доля успехов в схеме независимых испытаний сходится п.в. к вероятности успеха в одном испытании.

https://shiny.ilykei.com/misha/cmc/lln/
https://shiny.ilykei.com/misha/cmc/mc/

Предельные теоремы, ЦПТ

Theorem (Центральная предельная теорема в форме Леви)

Пусть $\xi_1, \xi_2, \ldots$ *— i.i.d. с* $\mathbf{E}\xi_k = a$ *и* $0 < \mathbf{D}\xi_k = \sigma^2 < \infty$. *Тогда, если обозначить* $S_n = \xi_1 + \ldots + \xi_n$ *и* $S_n^* = \frac{S_n - \mathbf{E}S_n}{\sqrt{\mathbf{D}S_n}} = \frac{S_n - na}{\sqrt{n}\sigma}$, *то*

$$S_n^* \xrightarrow{d} \mathcal{N}(0,1) \quad \text{при} \quad n \to \infty.$$

Иначе говоря, распределение суммы S_n *большого числа независимых одинаково распределённых слагаемых имеет распределение, близкое к нормальному* $N(\mathbf{E}S_n, \mathbf{D}S_n)$.

https://gallery.shinyapps.io/CLT_mean/

https://xjlc.shinyapps.io/climit/