ПРИКЛАДНАЯ СТАТИСТИКА

Теоретический минимум по теории вероятностей*

1. Условная вероятность случайного события

Условной вероятностью события A при условии, что произошло событие B, называется число:

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

Из формулы условной вероятности можно получить формулу для вероятности произведения нескольких событий:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B) \cdot \mathbb{P}(B).$$

Если событий несколько, формулу можно продолжить:

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A \mid B, C) \cdot \mathbb{P}(B \mid C) \cdot \mathbb{P}(C).$$

*Это краткая выжимка с основными определениями из теории вероятностей. Она не претендует на полноту. Частично шпаргалка основана на материале https://github.com/bdemeshev/pr201/tree/master/cheat_sheet

2. Независимость событий

Говорят, что два события попарно **независимы,** если верно следующее:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Говорят, что n случайных событий **независимы в совокупности,** если для любого $1 \leqslant k \leqslant n$ и любого набора различных меж собой индексов $1 \leqslant i_1, \ldots, i_k \leqslant n$ имеет место равенство:

$$\mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdot \ldots \cdot \mathbb{P}(A_{i_k})$$

3. Формула полной вероятности

Пусть событие A происходит вместе с одним из событий $H_1, H_2, \ldots, H_k$. Пусть эти события попарно несовместны (ещё говорят, что они составляют **полную группу.)**

Нам известны вероятности этих событий $\mathbb{P}(H_1), \mathbb{P}(H_2), \ldots, \mathbb{P}(H_k)$, а также условные вероятности события A: $\mathbb{P}(A \mid H_1), \mathbb{P}(A \mid H_2), \ldots, \mathbb{P}(A \mid H_k)$.

Тогда вероятность события A может быть вычислена по формуле:

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(H_i) \cdot \mathbb{P}(A \mid H_i).$$

4. Формула Байеса

Пусть событие A происходит вместе с одним из событий $H_1, H_2, \ldots, H_k$, которые составляют полную группу и попарно несовместны.

Пусть известно, что в результате испытания событие A произошло. Тогда условная вероятность того, что имело место событие H_k, можно пересчитать по формуле:

$$\mathbb{P}(H_k \mid A) = \frac{\mathbb{P}(H_k \cap A)}{P(A)} = \frac{\mathbb{P}(H_k) \cdot \mathbb{P}(A \mid H_k)}{\sum_{i=1}^{\infty} \mathbb{P}(H_i) \cdot \mathbb{P}(A \mid H_i)}$$

5. Функция распределения случайной величины

Функцией распределения случайной величины X называется функция $F_X(x)$, определённая для любого действительного числа $x \in \mathbb{R}$ и выражающая собой вероятность того, что случайная величина X примет значение, лежащее на числовой прямой левее точки x, то есть

$$F_X(x) = \mathbb{P}(X \leqslant x).$$

Любая функция распределения обладает следующими свойствами:

- Принимает значения в диапазоне от 0 до 1, при этом:

$$\lim_{x \to +\infty} F_X(x) = 1 \qquad \lim_{x \to -\infty} F_X(x) = 0$$

- $F_X(x)$ не убывает: $F_X(x_1) \leqslant F_X(x_2) \quad \forall x_1 \leqslant x_2$
- $F_X(x)$ непрерывна справа: $\lim\limits_{x \to x_0^+} F_X(x) = F_X(x_0)$

6. Функция плотности случайной величины и ее свойства

Случайная величина X имеет **абсолютно непрерывное распределение,** если существует неотрицательная функция $f_X(x)$ такая, что

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt.$$

Функция $f_X(x)$ называется **функцией плотности распределения** случайной величины X.

Свойства функции плотности:

- Неотрицательно определена: $f_X(x) \geqslant 0$
- Площадь под плотностью распределения всегда равна единице:

$$\int_{-\infty}^{+\infty} f_X(t)\, dt = 1$$

- С помощью плотности можно найти вероятность того, что случайная величина попадёт в конкретный отрезок:

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_{a}^{b} f_X(x)\, dx = F_X(b) - F_X(a)$$

7. Функция совместного распределения двух случайных величин

Пусть у нас ест две случайные величины X и Y. **Совместной функцией распределения** двумерной случайной величины будет называться функция, определённая $\forall x, y \in \mathbb{R}$ и выражающая собой вероятность одновременного выполнения событий $X \leqslant x$, и $Y \leqslant y$:

$$F(x, y) = P(X \leqslant x, Y \leqslant y)$$

Свойства функции распределения аналогичны одномерному случаю:

- Принимает значения в диапазоне от 0 до 1, при этом:

$$F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0, \qquad F(+\infty, +\infty) = 1$$

- $F_X(x)$ не убывает по каждому из своих аргументов
- $F_X(x)$ непрерывна справа по каждому из своих аргументов
- Если один из аргументов стремится к бесконечности, то получится функция распределения другой составляющей:

$$F(x, +\infty) = \mathbb{P}(X < x, Y < +\infty) = \mathbb{P}(X < x) = F_X(x)$$

Случайные величины X и Y **независимы,** если их функция распределения равна произведению функций распределения составляющих:

$$F(x, y) = F_X(x) \cdot F_Y(y)$$

Для n случайных величин функцию распределения можно задать по аналогии.

8. Совместная плотность распределения двух случайных величин

Случайные величины X, Y имеют **абсолютно непрерывное совместное распределение,** если существует функция $f(x, y) \geqslant 0$ такая, что $\forall x, y$ совместная функция распределения представима в виде:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(t_1, t_2)\, dt_1\, dt_2$$

Если такая функция $f(x, y)$ существует, она называется плотностью совместного распределения случайных величин X и Y.

Свойства совместной функции плотности:

- $f(x, y) \geqslant 0$
- Площадь под совместной плотностью распределения равна единице:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(t_1, t_2)\, dt_1\, dt_2 = 1$$

- Чтобы получить плотность распределения одной из составляющих, можно выинтегрировать из совместной плотности все значения другой:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy$$
$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx$$

Случайные величины X и Y с абсолютно непрерывными распределениями **независимы**, если плотность их совместного распределения существует и равна произведению частных функций плотности:

$$f(x, y) = f_X(x) \cdot f_Y(y)$$

9. Математическое ожидание

Интуитивно: среднее арифметическое значение величины при многократном повторении случайного эксперимента

Математическим ожиданием $\mathbb{E}(X)$ непрерывной случайной величины X называется число:

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t \cdot f_X(t)\, dt,$$

Математическим ожиданием $\mathbb{E}(X)$ дискретной случайной величины X называется число:

$$\mathbb{E}(X) = \sum_k a_k \cdot \mathbb{P}(X = a_k),$$

Свойства:

- $\mathbb{E}(a \cdot X + b \cdot Y + c) = a \cdot \mathbb{E}(X) + b \cdot E(Y) + c$
- Если $X \geqslant Y$ почти наверное, то $\mathbb{E}(X) \geqslant \mathbb{E}(Y)$

- Если X и Y независимы и их математические ожидания существуют, то

$$\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

10. Дисперсия

Интуитивно: мера разброса случайной величины. **Геометрический смысл**: квадрат длины случайной величины.

Дисперсия случайной величины $Var(X)$ — это число

$$Var(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = E(X^2) - E^2(X)$$

Дисперсия — это среднее значение квадрата отклонения случайной величины X от своего среднего.

Свойства:

- $Var(X) \geqslant 0$
- $Var(X) = 0$ равносильно тому, что $\mathbb{P}(X = const) = 1$
- $Var(a \cdot X + b) = a^2 \cdot Var(X)$
- $Var(X + Y) = Var(X) + Var(Y) + 2 \cdot Cov(X, Y)$
- $Var(X + Y) = Var(X) + Var(Y)$, если величины линейно независимы

11. Стандартное отклонение

Стандартным отклонением называют корень из дисперсии:

$$\sigma(X) = \sqrt{Var(X)}$$

Эту величину вводят, так как дисперсия измеряется в квадратах (лет^2, м^2) и т.п.

Свойства:

- $\sigma(X) \geqslant 0$
- $\sigma(X) = 0$ равносильно тому, что $\mathbb{P}(X = const) = 1$
- $\sigma(a \cdot X + b) = |a| \cdot \sigma(X)$

12. Ковариация

Интуитивно: мера линейной связи величин X и Y

Ковариацией между двумя случайными величинами называют величину

$$\mathrm{Cov}(X, Y) = \mathbb{E}\left([X - \mathbb{E}(X)] \cdot [Y - \mathbb{E}(Y)]\right) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot E(Y)$$

Геометрический смысл: скалярное произведение случайных величин

Свойства:

- Если X и Y независимы, то их ковариация равна нулю, но обратное неверно. Нулевая ковариация означает отсутствие линейной взаимосвязи. Взаимосвязь может быть устроена сложнее.
- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
- $\mathrm{Cov}(a \cdot X + b, Y) = a \cdot \mathrm{Cov}(X, Y)$
- $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$

13. Корреляция

Интуитивно: отнормированная мера линейной связи величин X и Y

Коэффициентом корреляции $\mathrm{Corr}(X, Y)$ случайных величин X и Y называется число:

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)} \cdot \sqrt{\mathrm{Var}(Y)}}.$$

Геометрический смысл: косинус угла между случайными величинами

Свойства:

- $-1 \leqslant \mathrm{Corr}(X, Y) \leqslant 1$
- $|\mathrm{Corr}(X, Y)| = 1 \Leftrightarrow \exists a, b \in \mathbb{R} : X = a \cdot Y + b$
- $\mathrm{Corr}(a \cdot X + b, c \cdot Y + d) = sign(ac) \cdot \mathrm{Corr}(X, Y)$
- $\mathrm{Corr}(X, Y) = \mathrm{Corr}(Y, X)$
- $\mathrm{Corr}(X, Y) = 0$ означает отсутствие линейной зависимости между X и Y, но зависимость может быть устроена сложнее.

14. Закон больших чисел в слабой форме

Пусть $X_1, \dots, X_n$ попарно независимые и одинаково распределённые случайные величины с конечным вторым моментом, $\mathbb{E}(X_i^2) < \infty$, тогда имеет место сходимость:

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{p} \frac{\mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n)}{n}$$

Если у случайных величин одинаковое математическое ожидание, тогда:

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{p} \mathbb{E}(X_1)$$

15. Центральная предельная теорема

Пусть $X_1, \dots, X_n$ случайные величины, имеющие одинаковое распределение с конечными математическим ожиданием и дисперсией:

$$X_1, \dots, X_n \sim iid(\mu, \sigma^2).$$

тогда при $n \to \infty$ имеет место сходимость по распределению:

$$\frac{X_1 + \dots + X_n - \mu \cdot n}{\sqrt{n}\sigma} \xrightarrow{d} N(0, 1)$$

16. Сходимость по вероятности

Говорят, что последовательность случайных величин $X_1, X_2, \dots$ сходится к случайной величине X при $n \to \infty$ **по вероятности**, и пишут $X_n \xrightarrow{p} X$, если для любого $\varepsilon > 0$:

$$\mathbb{P}(|X_n - X| \geqslant \varepsilon) \to 0$$

17. Сходимость по распределению

Говорят, что последовательность случайных величин $X_1, X_2, \dots$ сходится к случайной величине X при $n \to \infty$ **по распределению**, и пишут $X_n \xrightarrow{d} X$, если $F_{X_n}(x) \to F_X(x)$ для всех x, в которых $F_X(x)$ непрерывна.

18. Основные распределения

Биномиальное распределение

Биномиальное распределение — дискретное распределение количества успехов среди n испытаний с вероятностью успеха, равной p. Обычно записывают как:

$$X \sim Binom(n, p)$$

Вероятность того, что произойдёт k успехов расчитывается по формуле:

$$\mathbb{P}(X = k) = C_k^n \cdot p^k \cdot (1 - p)^{n-k}, \quad k \in \{0, 1, \ldots, n\}$$

Пример, когда возникает: сколько раз человек попадёт в баскетбольную корзину при n бросках

Свойства:

- $\mathbb{E}(X) = n \cdot p$
- $Var(X) = n \cdot p \cdot (1 - p)$

Распределение Пуассона

Распределение Пуассона — распределение дискретной случайной величины, представляющей собой число событий, произошедших за фиксированное время, при условии, что данные события происходят с некоторой фиксированной средней интенсивностью λ и независимо друг от друга. Хорошо подходит для моделирования счётчиков. Обычно записывают как:

$$X \sim Pois(\lambda)$$

Вероятность того, что произойдёт k событий расчитывается по формуле:

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \{0, 1, \ldots, \}$$

Пример, когда возникает: число лайков под фотографией, любая случайная величина счётчик, которая подчиняется аксиомам простейшего потока событий

Свойства:

- $\mathbb{E}(X) = \lambda$

- $Var(X) = \lambda$

Геометрическое распределение

Распределение Пуассона — распределение дискретной случайной величины, представляющей собой номер первого успеха в серии испытаний Бернулли. Обычно записывают как:

$$X \sim Geom(p)$$

Вероятность того, что номер первого успеха равен k находится как:

$$\mathbb{P}(X = k) = p \cdot (1 - p)^{k-1}$$

Пример, когда возникает: номер попытки, при которой игрок попал в баскетбольную корзину

Свойства:

- $\mathbb{E}(X) = \frac{1}{p}$
- $Var(X) = \frac{1-p}{p^2}$

Равномерное распределение

Равномерное распределение на отрезке $[a; b]$ обладает плотностью распределения:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a; b] \\ 0, & x \notin [a; b] \end{cases}$$

Обычно записывают как:

$$X \sim \mathcal{U}[a; b]$$

Пример, когда возникает: остаток при округлении чисел

Свойства:

- $\mathbb{E}(X) = \frac{a+b}{2}$
- $Var(X) = \frac{(b-a)^2}{12}$

Экспоненциальное распределение

Экспоненциальное распределение обладает плотностью распределения:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & x < 0 \end{cases}$$

Обычно записывают как:

$$X \sim \mathrm{Exp}(\lambda)$$

Пример, когда возникает: время между событиями, имеющими распределение Пуассона (время, пока следующий человек придёт в кассу, время до следующего лайка под фото и тп)

Свойства:

- $\mathbb{E}(X) = \frac{1}{\lambda}$
- $Var(X) = \frac{1}{\lambda^2}$

Нормальное распределение

Говорят, что у случайной величины X **нормальное распределение с параметрами μ и σ^2**, если она обладает плотностью распределения

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Обычно записывают как:

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

Пример, когда возникает: нахождение суммы или среднего большого количества независимых одинаково распределенных величин

Свойства:

- $\mathbb{E}(X) = \mu$
- $Var(X) = \sigma^2$
- Если $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ и $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, тогда

$$a \cdot X + b \cdot Y + c \sim \mathcal{N}(a \cdot \mu_1 + b \cdot \mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2)$$

- Для нормального распределения выполняются правила одной, двух и трёх сигм:

$$\mathbb{P}(\mu - \sigma \leqslant X \leqslant \mu + \sigma) \approx= 0.683$$
$$\mathbb{P}(\mu - 2 \cdot \sigma \leqslant X \leqslant \mu + 2 \cdot \sigma) \approx= 0.954$$
$$\mathbb{P}(\mu - 3 \cdot \sigma \leqslant X \leqslant \mu + 3 \cdot \sigma) \approx= 0.997$$

”Хи-квадрат”распределение

Пусть случайные величины $X_1, \dots, X_k$ независимы и одинаково распределены. Причём нормально с параметрами 0 и 1. Обычно такой факт записывают следующим образом:

$$X_1, \dots, X_k \sim \text{iid } N(0, 1).$$

Буквы iid расшифровываются как identically independently distributed (независимы и одинаково распределены).

Случайная величина $Y = X_1^2 + \dots X_k^2$ имеет **распределение хи-квадрат с k степенями свободы.** Степень свободы это просто название для параметра распределения.

Обычно записывают как:

$$Y \sim \chi_k^2$$

Пример, когда возникает: на практике тесно связано с выборочной дисперсией для нормальных выборок

Свойства:

- $\mathbb{E}(\chi_k^2) = k \cdot \mathbb{E}(X_i^2) = k$
- $Var(\chi_k^2) = k \cdot \mathbb{E}(X_i^4) = 2k$
- Распределение устойчиво к суммированию. То есть, если χ_k^2 и χ_m^2 независимы, тогда $\chi_k^2 + \chi_m^2 = \chi_{k+m}^2$
- $\frac{\chi_k^2}{k} \longrightarrow 1$ по вероятности.

Распределение Стьюдента

Пусть случайные величины

$$X_0, X_1, \dots, X_k \sim \text{iid } N(0, 1),$$

тогда случайная величина

$$Y = \frac{X_0}{\sqrt{\chi_k^2 / k}}$$

имеет **распределение Стьюдента с k степенями свободы.** Обычно записывают как:

$$Y \sim t(k)$$

Пример, когда возникает: на практике тесно связано с отношением выборочного среднего и стандартного отклонения нормальных выборок

Свойства:

- $\mathbb{E}(Y) = 0, k > 1$
- $Var(Y) = \frac{k}{k-2}, k > 2$
- Симметрично относительно нуля
- $t(k) \to N(0, 1)$ по распределению при $k \to \infty$
- При $k = 1$ совпадает с распределением Коши

Распределение Фишера

Говорят, что случайная величина

$$Y = \frac{\chi_k^2 / k}{\chi_m^2 / m}$$

имеет **распределение Фишера с k, m степенями свободы**.

Обычно записывают как:

$$Y \sim F_{k,m}$$

Пример, когда возникает: на практике тесно связано с отношением выборочных дисперсий двух нормальных выборок

Свойства:

- $\mathbb{E}(Y) = \frac{m}{m-2}, m > 2$
- $Var(Y) = \frac{2m^2(m+k-2)}{k(m-2)^2(m-4)}, m > 4$
- Если $Y \sim F(k, m)$, тогда $\frac{1}{Y} \sim F(m, k)$
- При $k \to \infty$ и $m \to \infty$ $F(k, m) \to 1$ по вероятности
- А вот этот факт не раз всплывёт в эконометрике: $t_k^2 = F(1, k)$

Квантильное преобразование

Теорема:

Пусть функция распределения $F_X(x)$ непрерывна. Тогда случайная величина $Y = F(X)$ имеет равномерное распределение на отрезке $[0; 1]$.

Следствие:

Пусть $Y \sim U[0; 1]$, а $F(x)$ произвольная функция распределения. Тогда случайная величина $X = F^{-1}(Y)$ будет иметь функцию распределения $F(x)$.