Глава 1

Дебильник

1.1 Многомерное нормальное распределение

<u>def</u>. Стандартный гауссовский вектор — случайный n-мерный вектор $Z = (Z_1, Z_2, \dots Z_n)$, координаты которого независимы и имеют распределение $\mathcal{N}(0, 1)$.

<u>def</u>. Гауссовский вектор (Нормальный вектор) — вектор, для которого существует матрица $\mathbf{A} \in \mathbb{R}^{n \times m}$, стандартный гауссовский вектор $Z \in \mathbb{R}^m$, и вектор $b \in \mathbb{R}^n$ такие, что $X = \mathbf{A}Z + b$.

<u>def</u>. Распределение нормального вектора $X \in \mathbb{R}^n$ — $\mathcal{N}(\mu, \mathbf{\Sigma})$ или $\mathcal{N}_n(\mu, \mathbf{\Sigma})$, где $\mu = \mathbb{E}X$ и $\mathbf{\Sigma} = \operatorname{cov}(X)$.

<u>def</u>. Распределение хи-квадрат с n степенями свободы — распределение $\chi^2(n)$ величины $\chi^2 = Z_1^2 + Z_2^2 + \ldots + Z_n^2$, где $Z_1, Z_2, \dots Z_n$ — независимы $\mathcal{N}(0, 1)$ величины.

<u>def</u>. Распределение Стьюдента с n степенями свободы — распределение $T(n)$ величины $\frac{\sqrt{n}X}{\sqrt{Y}}$, где $X \sim \mathcal{N}(0, 1)$, $Y \sim \chi^2(n)$ и независимы.

<u>def</u>. Распределение Фишера со степенями свободы n и m — распределение $F(n, m)$ величины $\frac{X/n}{Y/m}$, где $X \sim \chi^2(n)$, $Y \sim \chi^2(m)$ и независимы.

1.2 Условное матожидание

def. Условное матожидание $\mathbb{E}(Y \mid X)$ случайной величины Y при условии случайной величины X — такая измеримая функция g_0 величины X, при которой $\mathbb{E}(Y - g(X))^2$ минимально для всех измеримых функций g.

Условное матожидание — ортогональная проекция Y на линейное пространство всех измеримых функций X. То есть УМО — единственная измеримая функция, которая удовлетворяет условию ортогональности:

$$\forall g \colon \mathbb{E}(Y - \mathbb{E}(Y \mid X))g(X) = 0.$$

1.3 Статистическая модель, выборка

def. Статистическая модель — множество распределений $\mathfrak{P}$, которое, по нашему мнению, адекватно приближает $\mathcal{P}_D$.

def. Данные d — реализация случайного элемента D, имеющего распределение $\mathcal{P}_D$.

Статистические модели делят на:

- параметрические, если $\mathfrak{P} = \{\mathcal{P}_0 \mid \theta \in \Theta \subset \mathbb{R}^k\}$.
  Пример: $\mathfrak{P} = \{\mathcal{N}(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 \geqslant 0\}$.
- непараметрические, если $\mathfrak{P} = \{\mathcal{P}_0 \mid \theta \in \Theta \subset V\}$, где V не обязательно конечномерное.
  Пример: $\mathfrak{P} = \{\mathcal{P}^{\otimes n} \mid \int_{\mathfrak{X}} x\mathcal{P}(dx) = 0\}$
- семипараметрические, если $\mathfrak{P} = \{\mathcal{P}_0 \mid \theta \in \Theta \subset \mathbb{R}^k \times V\}$.
  Пример: линейная регрессия $Y = X\beta + \varepsilon$, $\beta \in \mathbb{R}^k$, $\mathbb{E}\varepsilon = 0$, $\mathbb{D}\varepsilon = \sigma^2$.

Если $D = [X_1, \ldots X_n]$ и X_i независимы и имеют одинаковое распределение $\mathcal{P}_X$, D называется выборкой объема n и обозначается $X_{[n]}$, $\mathcal{P}_X$ — генеральная совокупность. В этом случае модель приобретает вид $\mathfrak{P} = \{\mathcal{P}^{\otimes n} \mid \mathcal{P} \in \mathfrak{P}_X\}$, где $\mathfrak{P}_X$ — модель для $\mathcal{P}_X$.

1.4 Формула Байеса, априорное, апостериорное распределение

- Априорное распределение — наше ощущение относительно значения параметра до проведения эксперимента.
- Апостериорное распределение — ощущение после получения данных эксперимента.

def (Формула Байеса)**.** Здесь p — вероятность, d — данные, θ — параметры.

$$p(\theta \mid d) = \frac{p(d \mid \theta) \cdot p(\theta)}{p(d)}.$$

- $p(\theta \mid d)$ — апостериорное распределение,
- $p(d \mid \theta)$ — правдоподобие,
- $p(\theta)$ — априорное распределение,
- $p(d)$ — вероятность данных.

1.5 Расстояние Кульбака-Лейблера, энтропия

Пусть мы принимаем случайные символы $x_1, \dots x_k$, вероятность появления x_i равна p_i, записываем с помощью битовой строки длины l_i. Тогда средняя длина символа равна

$$l = \sum_{i=1}^{k} p_i \cdot l_i.$$

Чтобы минимизировать l, необходимо подобрать следующие $l_i = -\log_2 p_i$. И тогда средняя длина будет равна $H(x) \coloneqq -\sum_{i=1}^{k} p_i \cdot \log_2 p_i$, эта величина называется двоичной энтропией сообщения. Аналогично можно брать любой другой логарифм, мы будем использовать натуральный.

Для непрерывной величины можно завести дифференциальную энтропию:

$$H(X) = -\int p(x) \log p(x) dx.$$

Пусть случайная величина X имеет функцию вероятности p, но мы кодируем символы, как-будто она имеет функцию вероятности q. Тогда средняя длина сообщения будет равна $-\sum_{i=1}^{k} p_i \cdot \log q_i$, эта величина называется кросс-энтропией $H(p \mid q)$ распределений p и q.

$H(p \mid q)$ всегда будет больше $H(p)$, так как $H(p)$ минимально.

def. Величина потери информации из-за использования q вместо p называется расстоянием Кульбака-Лейблера между p и q:

$$D_{KL}(p, q) = H(p \mid q) - H(p) = -\sum_{i=1}^{k} p_i \cdot \log \frac{q_i}{p_i}.$$

Для непрерывных величин все обобщается следующим образом

$$D_{KL} = -\int p_i \cdot \log \frac{q_i}{p_i}.$$

1.6 Статистика...

1.6.1 Статистика

Параметр или характеристика распределения — функционал от этого распределения.

def. Статистика — функция θ^* от данных d.

Пусть модель $\mathfrak{P}_{[n]} = \{\mathcal{P}^{\otimes n} \mid \mathcal{P} \in \mathfrak{P}\}$, искомая характеристика $\theta \colon \mathfrak{P} \to \mathbb{R}^k$.

1.6.2 Несмещенность

Чему равна оценка как случайная величина в среднем, если она равна характеристике?

def. Оценка Θ^* называется

- несмещенной, если $\forall \mathcal{P} \in \mathfrak{P} \colon \mathbb{E}\theta^*(X_{[n]}) = \theta(\mathcal{P})$, где $X_{[n]} \sim \mathcal{P}^{\otimes n}$,

- асимптотически несмещенной, если $\forall \mathcal{P} \in \mathfrak{P}\colon \mathbb{E}\theta^*(X_{[n]}) \to \theta(\mathcal{P})$.

Смещение — величина $b(\theta^*) = \mathbb{E}(\theta^*(X_{[n]})) - \theta(\mathcal{P})$.

Среднеквадратичная ошибка — величина $\mathrm{MSE}(\theta^*) = \mathbb{E}\left(\theta^*(X_{[n]}) - \theta(\mathcal{P})\right)^2$.

В общем случае
$$\mathrm{MSE}(\theta^*) = \mathbb{D}\theta^*(X_{[n]} + b^2(\theta^*).$$

- Выборочное среднее как оценка матожидания — несмещенная оценка,
- Выборочная дисперсия как оценка дисперсии — асимптотически несмещенная,
- Исправленная выборочная дисперсия как оценка дисперсии — несмещенная оценка.

1.6.3 Состоятельность

def. Оценка θ^* называется

- состоятельной, если $\forall \mathcal{P} \in \mathfrak{P}\colon \theta^*(X_{[n]}) \xrightarrow{\mathbb{P}} \theta(\mathcal{P})$, где $X_{[n]} \sim \mathcal{P}^{\otimes n}$,
- сильно состоятельной, если $\theta^*(X_{[n]}) \xrightarrow{\text{п. н.}} \theta(\mathcal{P})$.

1.6.4 Асимптотическая нормальность

def. Оценка θ^* называется асимптотически нормальной с коэффициентом рассеивания (или просто дисперсией) $\sigma^2\big(\theta(\mathcal{P})\big\rangle 0$, если
$$\sqrt{n}\left(\theta^*(X_{[n]}) - \theta(\mathcal{P})\right) \xrightarrow{d} \eta \sim \mathcal{N}(0, \sigma^2(\theta^*(\mathcal{P}))).$$
В многомерном случае рассматривается ковариационная матрица вместо дисперсии.

- Выборочная дисперсия и второй момент — асимптотически нормальная оценка.
- Из асимптотической нормальности следует состоятельность.

1.6.5 Эффективность

Рассмотрим класс оценок $K = \{\hat{\theta}\}$ параметра θ.

def. Оценка $\theta^* \in K$ называется эффективной в классе K, если *для любой другой оценки* $\hat{\theta} \in K$ *и для любого исследуемого параметра* $\theta \in \Theta$ *выполняется*

$$\mathrm{MSE}_\theta(\theta^*) \leqslant \mathrm{MSE}_\theta(\hat{\theta}).$$

Класс несмещенных оценок

$$K_0 = \{\hat{\theta} \mid \mathbb{E}\hat{\theta} = \theta, \forall\theta \in \Theta\}.$$

def. Эффективная оценка θ^*, если эффективна в классе K_0.

def. Асимптотически эффективной в классе K, если для любой оценки $\hat{\theta} \in K$ и для любого $\theta \in \Theta$ выполняется

$$\overline{\lim_{n\to\infty}} \frac{\mathrm{MSE}(\theta^*)}{\mathrm{MSE}(\hat{\theta}}.$$

1.6.6 Робастность

def. Робастность — свойство оценки быть устойчивой к хвостам распределения.

Пусть F — распределение, $\{G_n\}$ — последовательность распределений, что

$$|F - G_n| \coloneqq \sup_x |F(x) - G_n(x)| \to 0.$$

def. Характеристика θ обладает качественной робастностью, если $\theta(G_n) \to \theta(F)$

Пусть также δ_x — вырожденное распределение в точке x.

def. Загрязненное распределение — смесь $F_{x,\varepsilon} = (1-\varepsilon)F + \varepsilon\delta_x$.

def. Функция влияния характеристики θ — величина

$$IF(x) = \lim_{\varepsilon\to 0+} \frac{\theta(F_{x,\varepsilon}) - \theta(F)}{\varepsilon}.$$

def. Характеристика θ называется B-робастной или инфинитезимально робастной, если $IF(x)$ ограничена.

def. Асимптотическая толерантность характеристики θ —

$$\tau = \inf\left\{\varepsilon \mid \sup_x |\theta(F_{x,\varepsilon} - \theta(F)| = \infty\right\}.$$

1.6.7 Достаточность

def. Статистика $T(x) = \{T_1(x), \dots, T_m(x))\}$ называется достаточной, если для всех

- $\theta \in \Theta$,
- $B \in \mathfrak{P}(\mathbb{R}^n)$ и
- $t = (t_1, \dots, t_m)$

условная вероятность $\mathbb{P}(X_{[n]} \in B \mid T(X_{[n]}) = t)$ не зависит от θ.

То есть информация о θ в выборке полностью содержится в значении $T(x_{[n]})$.

thm (факторизации). *$T(x)$ достаточна, согда существуют функции g и h, что*

$$p(X_{[n]} = x_{[n]} \mid \theta) = g(T(x_{[n]}), \theta) h(x_{[n]}),$$

где p — вероятность или плотность.

1.6.8 Полнота

def. Статистика T называется полной, если для любой *измеримой g верно следствие*

$$\forall \theta \in \Theta\colon \mathbb{E} g(T(X_{[n]})) \equiv 0 \quad \Longrightarrow \quad g(T(X_{[n]})) \overset{п.н.}{=} 0.$$

1.7 Теоремы Колмогорова-Блэкуэлла-Рао и Лемана-Шеффе

thm (Колмогорова-Блэкуэлла-Рао)**.** *Пусть θ^* — оценка параметра θ, T — достаточная статистика. Тогда*

$$\mathrm{MSE}(\theta^*) \geqslant \mathrm{MSE}(\mathbb{E}(\theta^* \mid T)).$$

thm (Лемана-Шеффе)**.** *Пусть θ^* — оценка параметра θ, T — достаточная и полная статистика. Тогда $\mathbb{E}(\theta^* \mid T)$ — единственная эффективная оценка в классе оценок со смещением $b(\theta^*)$.*

1.8 Доверительный интервал

Пусть есть модель $\mathfrak{P}_{[n]} = \{\mathcal{P}^{\otimes n} \mid \mathcal{P} \in \mathfrak{P}\}$ и $\theta \colon \mathfrak{P} \to \mathbb{R}^k$ — искомая характеристика.

def. Доверительный интервал (точный доверительный интервал) с уровнем доверия γ — пара статистик (θ^*_L, θ^*_R), такая что для любого $\mathcal{P} \in \mathfrak{P}$ и $X_{[n]} \sim \mathcal{P}^{\otimes n}$

$$\mathbb{P}\left(\theta^*_L(X_{[n]}) \leqslant \theta(\mathcal{P}) \leqslant \theta^*_R(X_{[n]})\right) = \gamma.$$

Интервал называется

- асимптотическим, если
$$\mathbb{P}\left(\theta^*_L(X_{[n]}) \leqslant \theta(\mathcal{P}) \leqslant \theta^*_R(X_{[n]})\right) \xrightarrow{n\to\infty} \gamma.$$
- центральным, если
$$\mathbb{P}\left(\theta^*_L(X_{[n]}) > \theta(\mathcal{P})\right) = \mathbb{P}\left(\theta^*_R(X_{[n]}) < \theta(\mathcal{P})\right).$$
- левым, если
$$\mathbb{P}\left(\theta^*_L(X_{[n]}) > \theta(\mathcal{P})\right) = 0.$$
- правым, если
$$\mathbb{P}\left(\theta^*_R(X_{[n]}) < \theta(\mathcal{P})\right) = 0.$$

1.9 Бутстреп

1.9.1 Параметрический бутстреп

Если работаем с параметрической моделью, можем заменить $X = X(\theta)$ не на X^*, а на $X(\theta^*)$ и сэмплировать из этого распределения.

1.9.2 Непараметрический бутстреп

Рецепт

1. изготовим N выборок $x^*_{[n],1}, \ldots, x^*_{[n],N}$ из эмпирического распределения (рандом с возвращением)
2. вычисляем $\theta^b_i = \theta^*(x^*_{[n],i}$, получаем бутстреповскую выборку $\theta^b_{[N]}$,
3. по бутстреповской выборке оцениваем, что нужно.

Ограничения

- θ^* — plug-in оценка
- θ^* — достаточно гладкая (обычно дифференцируема)
- у X достаточно много моментов (обычно конечная дисперсия)
- нужно генерировать большие выборки
- на очень больших данных трудозатратен
- на маленьких данных велика неустранимая ошибка

1.10 Гипотеза, альтернатива...

Пусть $\mathfrak{P}$ — модель.

1.10.1 Гипотеза и альтернатива

def. Гипотеза — утверждение вида $H \colon \mathcal{P}_X \in \mathfrak{P}_0 \subset \mathfrak{P}$.

Если $|\mathfrak{P}_0| = 1$, гипотеза называется простой, иначе сложной.

Нулевая гипотеза — гипотеза H_0, которую мы хотим проверить. Проверка гипотезы — процесс принятия решения о том, противоречит ли она наблюдаемой выборке данных.

Альтернатива — гипотеза H_1, которая отражает, какие отклонения от нулевой гипотезы нам интересны.

1.10.2 Критерий

def. Нерандомизированный критерий (критерий) — отображение $\varphi \colon d \to \{\text{принимаем}, \text{отвергаем}\} = \{H_0, H_1\} = \{0, 1\}$.

Часто критерий устроен так: имеется

- статистика критерия T и
- критическое множество C, и

$$\varphi(d) = [T(d) \in C] = [d \in T^{-1}(C)].$$

def. Рандомизированный критерий — отображение $\varphi \colon d \to [0, 1]$. Значение на данных d определяется как реализация случайной величины $D(\varphi(d))$.

Пусть мы согласны отвергать нулевую гипотезу пр условии, что она верна, но хотим делать это не очень часто. Пусть зафиксирован уровень значимости

$$\alpha \coloneqq \mathbb{P}(\varphi(D) = 1 \mid H_0),$$

который обычно является *параметром критерия*, то есть, задавая его, мы определяем *критическое множество* C_α такое, что

$$\mathbb{P}(T(D) \in C_\alpha \mid H_0) = \alpha.$$

Таким образом, для одного критерия определено семейство критических областей $\{C_\alpha \mid \alpha \in [0, 1]\}$, где обычно $C_\alpha \subset C_{\alpha'}$, если $\alpha < \alpha'$.

def. Уровень значимости — параметр критерия, который регулирует, насколько часто мы будем отвергать нулевую гипотезу при условии, что она верна.

1.10.3 p-value

Хотим оценить, насколько гипотеза противоречит наблюдаемым данным.

def. p-value — характеристика противоречия гипотезы наблюдаемым данным:

$$\text{p-value} \coloneqq \arg\min\{\alpha \in [0,1] \mid T(d) \in C_\alpha\}.$$

Другими словами, p-value — минимальное значение уровня значимости для данного значения статистики критерия, при котором H_0 может быть отвергнута.

Чем меньше p-value, тем больше гипотеза противоречит данным.

1.10.4 Ошибки разных родов

def. Ошибка первого рода — событие $\varphi(D) = 1 \mid H_0$. Если уровень значимости совпадает с вероятностью ошибки первого роба, То критерий называется точным.

Уровень значимости — вероятность ошибки первого рода.

def. Ошибка второго рода β — событие $\varphi(D) = 0 \mid H_1$, не отклонили нулевую гипотезу при условии, что была верна альтернатива.

Мощность критерия — вероятность $1 - \beta$ отклонить H_0 при условии, что верна H_1.

Для заданного уровня значимости мы хотим иметь как можно более мощный критерий.

1.10.5 Свойства критериев

def. Несмещенность — мощность всегда не меньше ошибки первого рода, критерий не отдает предпочтение альтернативе. $1 - \beta \geqslant \alpha$ для всех простых гипотез из $\mathfrak{P}_0$ и простых альтернатив из $\mathfrak{P}_1$.

def. Состоятельность — $\beta \xrightarrow{n\to\infty} 0$ для всех простых альтернатив из $\mathfrak{P}_1$.

def. Асимптотичность — $\alpha \xrightarrow{n\to\infty}$ для всех простых гипотез из $\mathfrak{P}_0$.

def. Наиболее мощный критерий для данного уровня значимости α_0 и простой альтернативы — такой критерий φ_1, что для любого критерия φ_2 такого, что $\alpha(\varphi_2) \leqslant \alpha_0$:

$$\beta(\varphi_1) \leqslant \beta(\varphi_2).$$

1.10.6 Размер эффекта

Во многих случаях важна не только информация о p-value, но и величина наблюдаемого эффекта. Размеры эффекта бывают разные, использование того или иного размера эффекта зависит от контекста.

Вместо сравнения p-value с уровнем значимости для принятия статистического решения можно считать размер эффекта, сравнивать с минимальным практически интересным.

1.11 Постановка гипотезы согласия. Критерии Колмогорова и Андерсона-Дарлинга

1.11.1 Постановка гипотезы согласия

def. Гипотеза согласия — гипотеза о соответствии эмпирического распределения теоретическому распределению вероятностей.

Критерии для гипотез согласия бывают

- общие — применимые к любому предполагаемому распределению выборки,
- специальные — применимые к гипотезам, формулирующие согласие с определенным свойством распределений;

- для простых гипотез,
- для сложных гипотез.

1.11.2 Критерий Колмогорова

Сравнивает эмпирическое и истинное распределение. Для простой гипотезы.

Пусть F_0 непрерывна на $\mathbb{R}$. Определим статистику Колмогорова:

$$D_n(x_{[n]}) = \sup_{x\in\mathbb{R}}|F_n^* - F_0(x)|.$$

- Если H_0 верна, то $D_n\left(X_[n]\right) \xrightarrow{\text{п.н.}} 0$;
- Если H_0 неверна, то $D_n\left(X_{[n]}\right) \xrightarrow{\text{п.н.}} \sup_{x\in\mathbb{R}}|F_X(x) - F_0(x)| > 0$.

1.11.3 Критерий Андерсона-Дарлинга

Для простой гипотезы.

Определим статистику критерия Андерсона-Дарлинга:

$$A^2 = n\int_{\mathbb{R}} \frac{\left(F_n^*(x) - F_0(x)\right)^2}{F_0(x)\left(1 - F_0(x)\right)} dF_0(x) =$$
$$= -n - \sum_{i=1}^{n} \frac{2i-1}{n}\left[\ln F_0(X_{(i)} \mid \theta) + \ln\left(1 - F_0(X_{(n+1-i)} \mid \theta)\right]\right.$$

Статистика A^2 при выполнении H_0 и непрерывности F_0 подчиняется табличному распределению. $C_\alpha = (a^2_{1-\alpha}, \infty)$.

1.12 Постановка гипотезы о параметрах, проверка через доверительные интервалы, z-test, t-test, бутстреп из нулевой гипотезы

1.12.1 Постановка гипотезы о параметрах

Пусть θ — параметр ($X \sim F(x,\theta)$) или характеристика ($\theta = \varphi(F_x)$) распределения.

Нулевая гипотеза: $H_0 \colon \theta = \theta_0$. Типичные альтернативы:

- $H_1 \colon \theta = \theta_1 \neq \theta_0$,
- $H_> \colon \theta > \theta_0$,
- $H_< \colon \theta < \theta_0$,
- $H_{\neq} \colon \theta \neq \theta_0$.

Усредненный рецепт:

1. Выбираем оценку θ^* параметра θ, распределение которой приближенно известно при данном θ.
2. В зависимости от альтернативы строим критическое множество:
   - $H_1 \colon \theta = \theta_1 > \theta_0$ или $H_>$, то $C_\alpha = (\theta^*_{1-\alpha}, \infty)$ [1] — правое критическое множество;
   - $H_1 \colon \theta = \theta_1 < \theta_0$ или $H_<$, то $C_\alpha = (-\infty, \theta^*_\alpha)$ — левое критическое множество;
   - $H_{\neq}$, то $C_\alpha = (-\infty, \theta^*_{\frac{\alpha}{2}} \cup (\theta^*_{1-\frac{\alpha}{2}}, \infty)$ — двустороннее критическое множество.
3. Если $\theta^* \in C_\alpha$, то гипотезу можно отклонить, иначе — нельзя.

1.12.2 Проверка через доверительные интервалы

Усредненный рецепт:

1. В зависимости от альтернативы строим доверительный интервал с уровнем доверия $\gamma = 1 - \alpha$:
   - $H_1 \colon \theta = \theta_1 > \theta_0$ или $H_>$, то (θ^*_L, ∞) — правый доверительный интервал;
   - $H_1 \colon \theta = \theta_1 < \theta_0$ или $H_<$, то $(-\infty, \theta^*_R)$ — левый доверительный интервал;
   - $H_{\neq}$, то (θ^*_L, θ^*_R) — центральный доверительный интервал.

[1] Здесь θ^*_x — квантиль уровня x распределения $\theta^* \mid H_0$

1.12.3 z-test

Пусть $X \sim \mathcal{N}(\mu, \sigma^2)$, μ неизвестно, σ^2 известно.

Если H_0 верна, то $Z = \frac{\sqrt{n}(\overline{X}-\mu_0)}{\sigma} \sim \mathcal{N}(0,1)$ и $\overline{X} \sim N(\mu_0, \frac{\sigma^2}{n})$.

В зависимости от альтернативы подбираем критическую область:

	$\theta_1 > \theta_0,\ H_>$	$\theta_1 < \theta_0,\ H_<$	$H_{\neq}$
C_α	$(\mu_0 + z_{1-\alpha}\frac{\sigma}{\sqrt{n}}, \infty)$	$(-\infty, \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}})$	$\mathbb{R} \setminus (\mu_0 \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$
p-value	$1 - \Phi^{-1}(z)$	$\Phi^{-1}(z)$	$2(1 - \Phi^{-1}(\lvert z \rvert))$

Таблица 1.1: Критическая область для альтернативы

1.12.4 t-test

Пусть $X \sim \mathcal{N}(\mu, \sigma^2)$, μ неизвестно, σ^2 неизвестно.

Если H_0 верна, то $T = \frac{\sqrt{n}(\overline{X}-\mu_0)}{s} \sim T(n-1)$.

В зависимости от альтернативы подбираем критическую область:

	$\theta_1 > \theta_0,\ H_>$	$\theta_1 < \theta_0,\ H_<$	$H_{\neq}$
C_α	$(\mu_0 + t_{1-\alpha}\frac{s}{\sqrt{n}}, \infty)$	$(-\infty, \mu_0 + t_\alpha \frac{s}{\sqrt{n}})$	$\mathbb{R} \setminus (\mu_0 \pm t_{1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}})$
p-value	$1 - T^{-1}(t)$	$T^{-1}(t)$	$2(1 - T^{-1}(\lvert t \rvert))$

Таблица 1.2: Критическая область для альтернативы

1.12.5 Бутстреп из нулевой гипотезы

Пусть мы хотим проверить гипотезу $H_0\colon \mathbb{E}X = \theta_0$.

Рецепт:

1. Назначим каждому наблюдению x_i в выборке вероятность p_i.
2. Из пар (x_i, p_i) изготовим дискретное распределение F_p^*.

3. Подберем p_i так, чтобы с одной стороны $\overline{x} = \theta_0$, а с другой p_i максимизировали правдоподобие выборки $\mathcal{L}(p \mid x_{[n]}) = p_1 p_2 \ldots p_n$.

4. Бутстрепим кучу выборок из получившегося F_p^*, считаем по ним выборочное среднее.

5. Построим критическое множество в зависимости от альтернативы и проверим, лежит ли в нем выборочное среднее исходной выборки.