

Лекция 8

Основные критерии проверки статистических гипотез

Артемов А. В.
мФТиАД ФКН ВШЭ

5 апреля 2018

1 Ключевые статистики и тесты в теории принятия решений. Дискретное время. Критерий Неймана-Пирсона

Наша ближайшая цель состоит в том, чтобы естественным образом подойти к описанию тех «достаточных» статистик от наблюдаемых данных, на основании которых принимаются «оптимальные» решения.

Начнём с задачи различения двух гипотез. Пусть $(\xi_1, \xi_2, \dots, \xi_n, \dots)$ — независимые и одинаково распределённые случайные величины. Над ними было проведено наблюдение и в результате получилась реализация $(x_1, x_2, \dots, x_n, \dots)$. Но нужно уточнить, что понимать в данном случае под «независимостью и одинаковой распределённостью».

Пусть (Ω, \mathcal{F}) — измеримое пространство (пространство элементарных исходов и сигма-алгебра событий). На нём вводится две вероятностные меры P_0 и P_∞ . Предположение независимости случайных величин $(\xi_1, \xi_2, \dots, \xi_n, \dots)$ означает независимость относительно обеих мер, то есть для любого $n \in \mathbb{N}$ и любых $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ (будем считать, что мы смотрим на одномерные случайные величины)

$$P_\theta(\xi_1 \in A_1, \dots, \xi_n \in A_n) = P_\theta(\xi_1 \in A_1) \cdot \dots \cdot P_\theta(\xi_n \in A_n), \text{ где } \theta = 0 \text{ или } \infty.$$

Далее мы будем считать, что случайные величины ξ_k имеют функции распределения $F = F_\theta(x) (= P_\theta(\xi_k \leq x))$, у которых есть плотность $f = f_\theta(x)$:

$$dF_\theta(x) = f_\theta(x)\mu(dx),$$

где μ — некоторая (σ -конечная) мера. В качестве такой меры всегда можно взять $\mu(dx) = (P_0(dx) + P_\infty(dx))/2$. Впрочем, часто будем полагать, что $\mu(dx) = dx$, то есть μ — это мера Лебега (в этом случае говорят, что функция распределения *абсолютно непрерывна*).

Из независимости и одинаковой распределённости следует, что плотность $p_\theta(x_1, \dots, x_n)$ совместного распределения $F_\theta(x_1, \dots, x_n) = P_\theta(\xi_1 \leq x_1, \dots, \xi_n \leq x_n)$ равна¹

$$p_\theta(x_1, \dots, x_n) = f_\theta(x_1) \dots f_\theta(x_n) \quad (1.1)$$

Одну из главных ролей будет играть *отношение правдоподобия*, которое вводилось в курсе математической статистики:

$$L_n = \frac{f_0(x_1)f_0(x_2)\dots f_0(x_n)}{f_\infty(x_1)f_\infty(x_2)\dots f_\infty(x_n)}. \quad (1.2)$$

¹В дискретном случае полагайте, что $f_\theta(x) = P_\theta(\xi_1 = x)$.

Как вам известно, оно проявляется в задаче различения двух простых гипотез H_0 и H_∞ о том, какую плотность, f_0 или же f_∞ , имеют наблюдаемые случайные величины ξ_1, \dots, ξ_N . Решение этой задачи даёт *лемма Неймана-Пирсона*, один из вариантов которой состоит в следующем.

Пусть есть N наблюдений x_1, \dots, x_N над случайными величинами ξ_1, \dots, ξ_N . По этим наблюдениям нужно сделать вывод, какая из гипотез — H_0 ($\theta = 0$) или же H_∞ ($\theta = \infty$) — имеет место. Будем предполагать, что соответствующими плотностями функций распределения $F_\theta(x_1, \dots, x_N)$ являются $p_\theta(x_1, \dots, x_N)$:

$$F_\theta(x_1, \dots, x_N) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_N} p_\theta(y_1, \dots, y_N) \mu(dy_1, \dots, dy_N),$$

где μ — некоторая σ -конечная мера на $\mathcal{B}(\mathbb{R})$. В случае, когда ξ_k независимы и одинаково распределены плотности p_θ (по мере $\mu(dy_1, \dots, dy_N) = \mu(dy_1) \dots \mu(dy_N)$) задаются формулой (1.1).

С задачей различения двух гипотез связано одно определение:

Определение 1. Всякую (измеримую) функцию $d(x_1, \dots, x_N)$, принимающую два значения: \mathbb{H}_0 (верна гипотеза H_0) и \mathbb{H}_∞ (верна гипотеза H_∞), будем называть *решающей функцией*.

Наряду с решающей функцией в статистике вводятся понятия ошибок первого и второго рода.

Определение 2. Пусть $d(x_1, \dots, x_N)$ — решающая функция. Вероятностью *ошибки первого рода* $\alpha(d)$ будем называть $\alpha(d) = \mathbb{P}(\text{приняли } H_0 \mid \text{верна } H_\infty)$. Аналогично, вероятность *ошибки второго рода* $\beta(d)$ равна $\beta(d) = \mathbb{P}(\text{приняли } H_\infty \mid \text{верна } H_0)$.

Как же выбрать «оптимальное» решающее правило d ? В данном случае не понятно, что вкладывается в слово «оптимальное» — больно много трактовок. Будем считать, что в нашем случае «оптимальность» означает следующее: решающее правило d^* считается оптимальным, если сумма вероятностей ошибок первого и второго рода минимальна:

$$\alpha(d^*) + \beta(d^*) = \inf_d [\alpha(d) + \beta(d)] \triangleq \text{Er}(N; H_0, H_\infty). \quad (1.3)$$

Стоит заметить, что $\text{Er}(N; H_0, H_\infty)$ можно посчитать по следующей формуле:

$$\text{Er}(N; H_0, H_\infty) = 1 - \frac{1}{2} \| \mathbb{P}_0^{(N)} - \mathbb{P}_\infty^{(N)} \|,$$

где $\mathbb{P}_\theta^{(N)}(dx_1, \dots, dx_N) = f_\theta(x_1) \dots f_\theta(x_N) dx_1 \dots dx_N$, а $\| \cdot \|$ есть вариация меры (со знаком):

$$\|Q\| \triangleq 2 \sup_A |Q(A)|.$$

Из этого можно сделать следующий неформальный вывод: если меры $\mathbb{P}_0^{(N)}$ и $\mathbb{P}_\infty^{(N)}$ «сидят» на разных множествах, то $\| \mathbb{P}_0^{(N)} - \mathbb{P}_\infty^{(N)} \| = 2$ и $\text{Er}(N; H_0, H_\infty) = 0$. Это означает, что возможно безошибочное разделение гипотез. Если же меры близки, то $\| \mathbb{P}_0^{(N)} - \mathbb{P}_\infty^{(N)} \| \sim 0$ и $\text{Er}(N; H_0, H_\infty) \sim 1$.

Есть ещё одна формулировка, которую обычно называют *условной*. Пусть $D_a = \{d : \alpha(d) \leq a\}$ — множество решающих правил с вероятностью ошибки первого рода не больше

a . Требуется найти d_a^* из D_a (если такое существует), что оно минимизирует вероятность ошибки второго рода:

$$\beta(d_a^*) = \inf_{d \in D_a} \beta(d). \quad (1.4)$$

Теперь уместно рассказать о *рандомизированных* решающих правилах. Пусть $\varphi = \varphi(x_1, \dots, x_N)$ принимает значения в $[0, 1]$. Будем интерпретировать $\varphi(x_1, \dots, x_N)$, как вероятность принять гипотезу H_0 , если были получены наблюдения x_1, \dots, x_N над случайными величинами ξ_1, \dots, ξ_N . Теперь введём два обозначения, считая, что E_θ есть матожидание, взятое по вероятностной мере P_θ ,

$$\alpha(\varphi) = E_\infty[\varphi(\xi_1, \dots, \xi_N)], \quad \beta(\varphi) = E_0[1 - \varphi(\xi_1, \dots, \xi_N)].$$

Несложно понять, что это есть ни что иное, как вероятности ошибок первого и второго рода соответственно.

Оптимальность в данном случае вводится почти так же, как в условной формулировке. Пусть $\Phi_a = \{\varphi : \alpha(\varphi) \leq a\}$ — множество рандомизированных критериев с вероятностью ошибки первого рода не больше a . Тогда решающая функция φ_a^* будет называться оптимальным (рандомизированным) тестом, если

$$\beta(\varphi_a^*) = \inf_{\varphi \in \Phi_a} \beta(\varphi). \quad (1.5)$$

А теперь можно дать лемму Неймана-Пирсона:

Лемма (Нейман, Пирсон). Для любого $a \in [0, 1]$ найдутся такие константы λ_a^* и h_a^* , что рандомизированный критерий

$$\varphi^*(x_1, \dots, x_N) = \begin{cases} 1, & p_0(x_1, \dots, x_N) > h_a^* p_\infty(x_1, \dots, x_N), \\ \lambda_a^*, & p_0(x_1, \dots, x_N) = h_a^* p_\infty(x_1, \dots, x_N), \\ 0, & p_0(x_1, \dots, x_N) < h_a^* p_\infty(x_1, \dots, x_N) \end{cases} \quad (1.6)$$

является оптимальным в классе Φ_a .

Из (1.6) понятно, что ключевая статистика в лемме Неймана-Пирсона — это отношение правдоподобия (если знаменатель не обращается в ноль, конечно):

$$L_N = \frac{p_0(x_1, \dots, x_N)}{p_\infty(x_1, \dots, x_N)}. \quad (1.7)$$

Казалось бы, зачем вводятся рандомизированные тесты? Оказывается, что инфимум в (1.5) достигается на тесте, для которого вероятность ошибки первого рода *в точности* равна a . Этого, в общем случае, нельзя достичь без рандомизированных критериев, что будет явно использоваться при доказательстве леммы.

Доказательство. Для начала предположим, что мы нашли λ_a^* и h_a^* для критерия (1.6) такие, что $E_\infty[\varphi^*(\xi_1, \dots, \xi_N)] = a$. Докажем, что для любого другого критерия φ из класса Φ_a вероятность ошибки второго рода не меньше:

$$\beta(\varphi^*) = E_0[1 - \varphi^*(\xi_1, \dots, \xi_N)] \leq E_0[1 - \varphi(\xi_1, \dots, \xi_N)] = \beta(\varphi). \quad (1.8)$$

Это, в свою очередь, равносильно тому, что

$$E_0[\varphi^*(\xi_1, \dots, \xi_N)] \geq E_0[\varphi(\xi_1, \dots, \xi_N)].$$

Распишем разность матожиданий:

$$\mathbb{E}_0[\varphi^*(\xi_1, \dots, \xi_N) - \varphi(\xi_1, \dots, \xi_N)] = \int_{\mathbb{R}^N} (\varphi^*(\mathbf{x}) - \varphi(\mathbf{x})) p_0(\mathbf{x}) \mu(d\mathbf{x}),$$

где $\mu(d\mathbf{x}) = \mu(dx_1, \dots, dx_N)$.

Теперь сделаем детур и докажем, что

$$\int_{\mathbb{R}^N} (\varphi^*(\mathbf{x}) - \varphi(\mathbf{x})) (p_0(\mathbf{x}) - h_a^* p_\infty(\mathbf{x})) \mu(d\mathbf{x}) \geq 0. \quad (1.9)$$

Действительно, разобьём его на два (для компактности опустим аргументы функций):

$$\begin{aligned} \int_{\mathbb{R}^N} (\varphi^* - \varphi)(p_0 - h_a^* p_\infty) \mu(d\mathbf{x}) &= \int_{\{\mathbf{x} \in \mathbb{R}^N : \varphi^* > \varphi\}} (\varphi^* - \varphi)(p_0 - h_a^* p_\infty) \mu(d\mathbf{x}) + \\ &+ \int_{\{\mathbf{x} \in \mathbb{R}^N : \varphi^* < \varphi\}} (\varphi^* - \varphi)(p_0 - h_a^* p_\infty) \mu(d\mathbf{x}). \end{aligned}$$

Мы можем сказать, что на множестве $\{\mathbf{x} \in \mathbb{R}^N : \varphi^* > \varphi\}$ критерий не обращается в ноль: $\varphi^* > 0$ (так как φ принимает значения в $[0, 1]$). Если $\varphi^* > 0$, то из определения критерия получаем, что $p_0 \geq h_a^* p_\infty$. Отсюда получаем, что первый интеграл неотрицателен.

Аналогично, мы можем сказать, что на множестве $\{\mathbf{x} \in \mathbb{R}^N : \varphi^* < \varphi\}$ критерий не обращается в единицу: $\varphi^* < 1$. Следовательно, $p_0 \leq h_a^* p_\infty$ и второй интеграл тоже неотрицателен. Комбинируя эти результаты, получаем утверждение (1.9). Из него сразу же получается (1.8):

$$\int_{\mathbb{R}^N} (\varphi^* - \varphi) p_0 \mu(d\mathbf{x}) \geq h_a^* \int_{\mathbb{R}^N} (\varphi^* - \varphi) p_\infty \mu(d\mathbf{x}) = h_a^* (\mathbb{E}_\infty[\varphi^*] - \mathbb{E}_\infty[\varphi]) \geq 0. \quad (1.10)$$

Рассуждение выше опиралось на то, что существуют такие λ_a^* и h_a^* , что вероятность ошибки первого рода равна a : $\mathbb{E}_\infty[\varphi^*(\xi_1, \dots, \xi_N)] = a$. Теперь докажем, что они действительно существуют.

Введём функцию $g(h) = \mathbb{P}_\infty(p_0(\boldsymbol{\xi}) > h p_\infty(\boldsymbol{\xi}))$, где $\boldsymbol{\xi} = (\xi_1, \dots, \xi_N)$. Что мы можем сказать про эту функцию? На самом деле много: она не возрастает, она непрерывна справа, $g(h) = 1$ при $h < 0$ и $g(h) \rightarrow 0$ при $h \rightarrow \infty$. Далее, заметим, что

$$g(h) = \int_{\{\mathbf{x} \in \mathbb{R}^N : \frac{p_0(\mathbf{x})}{p_\infty(\mathbf{x})} > h\}} p_\infty(\mathbf{x}) \mu(d\mathbf{x}). \quad (1.11)$$

Для $a \in (0, 1)$ положим h_a^* , равным минимальному h , для которого выполнено $g(h) \leq a \leq g(h - 0)$. Далее, положим

$$\lambda_a^* = \frac{a - g(h_a^*)}{g(h_a^* - 0) - g(h_a^*)}. \quad (1.12)$$

Теперь покажем, что с такими значениями $E_\infty[\varphi^*] = a$. Действительно,

$$\begin{aligned} E_\infty[\varphi^*] &= \int_{\mathbb{R}^N} \varphi^*(\mathbf{x}) p_\infty(\mathbf{x}) \mu(d\mathbf{x}) = \int_{\{\mathbf{x} \in \mathbb{R}^N : \frac{p_0(\mathbf{x})}{p_\infty(\mathbf{x})} \geq h_a^*\}} \varphi^*(\mathbf{x}) p_\infty(\mathbf{x}) \mu(d\mathbf{x}) = \\ &= \int_{\{\mathbf{x} \in \mathbb{R}^N : \frac{p_0(\mathbf{x})}{p_\infty(\mathbf{x})} > h_a^*\}} p_\infty(\mathbf{x}) \mu(d\mathbf{x}) + \lambda_a^* \int_{\{\mathbf{x} \in \mathbb{R}^N : \frac{p_0(\mathbf{x})}{p_\infty(\mathbf{x})} = h_a^*\}} p_\infty(\mathbf{x}) \mu(d\mathbf{x}) = \\ &= g(h_a^*) + \frac{a - g(h_a^*)}{g(h_a^* - 0) - g(h_a^*)} [g(h_a^* - 0) - g(h_a^*)] = a. \end{aligned}$$

Теперь нужно описать граничные случаи. Если $a = 0$, то ошибок первого рода быть не должно. Как этого достичь? Всегда принимать H_∞ . Другими словами, полагаем, что $\varphi^*(\mathbf{x}) = 0$ и $h_a^* = \infty$. Аналогично, для $a = 1$ мы должны всегда принимать H_0 . Следовательно, нужно положить $h_a^* = 0$ и $\lambda_a^* = 1$.

Тем самым был получен тест φ^* , для которого $E_\infty[\varphi^*] = a$ и он является оптимальным в классе Φ_a . \square

Примечание. В случае, когда случайные величины ξ_1, \dots, ξ_N независимы и одинаково распределены, для совместной плотности верна формула (1.1). В этом случае удобно ввести следующие обозначения:

$$\zeta_k = \log \frac{f_0(x_k)}{f_\infty(x_k)} \text{ и } Z_k = \log L_n = \sum_{k=1}^n \zeta_k.$$

Теперь рассмотрим пару примеров применения леммы Неймана-Пирсона.

Пример 1 (Бернуллиевские случайные величины). Пусть гипотезы H_0 и H_∞ утверждают, что ξ_1, \dots, ξ_N — выборка из распределения Бернулли $\text{Bern}(p_0)$ или $\text{Bern}(p_\infty)$ соответственно, то есть для любого $1 \leq k \leq N$

$$P_\theta(\xi_k = 1) = p_\theta, \quad P_\theta(\xi_k = 0) = 1 - p_\theta \equiv q_\theta, \quad \theta = 0 \text{ или } \infty.$$

В таком случае отношение правдоподобия, согласно формуле (1.7), равно

$$L_n = \prod_{k=1}^N \left(\frac{p_0}{p_\infty} \right)^{x_k} \left(\frac{q_0}{q_\infty} \right)^{1-x_k}.$$

Возьмём от этого логарифм:

$$\begin{aligned} Z_n &= (x_1 + \dots + x_N) \log \frac{p_0}{p_\infty} + (N - x_1 + \dots + x_N) \log \frac{q_0}{q_\infty} = \\ &= (x_1 + \dots + x_N) \log \frac{p_0 q_\infty}{p_\infty q_0} + N \log \frac{q_0}{q_\infty}. \end{aligned}$$

Если ввести обозначение $X_N = x_1 + \dots + x_N$, то оптимальный критерий (1.6) будет выглядеть так:

$$\varphi^*(x_1, \dots, x_N) = \begin{cases} 1, & X_N > h_a^*, \\ \lambda_a^*, & X_N = h_a^*, \\ 0, & X_N < h_a^*, \end{cases} \quad (1.13)$$

где константы λ_a^* и h_a^* находятся из предположения о том, что вероятность ошибки первого рода для этого критерия равна a : $E_\infty[\varphi^*] = a$.

Вообще говоря, необходимость обращаться к рандомизированным тестам обычно связана с дискретностью распределения. В случае, когда распределения имеют плотности $f_\theta(x)$, можно строить и детерминированный тест (так как вероятность равенства нулевой). Рассмотрим это на следующем примере.

Пример 2 (Нормальные случайные величины). Пусть гипотезы H_0 и H_∞ утверждают, что ξ_1, \dots, ξ_n — выборка из нормального распределения с параметрами (μ_0, σ^2) и (μ_∞, σ^2) соответственно. Как известно, плотность нормального распределения $\mathcal{N}(\mu_\theta, \sigma^2)$ равна

$$f_\theta(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu_\theta)^2}{2\sigma^2} \right\}.$$

Посчитаем логарифм отношения правдоподобия:

$$\begin{aligned} Z_n &= \sum_{k=1}^n \left[\frac{(x_k - \mu_\infty)^2}{2\sigma^2} - \frac{(x_k - \mu_0)^2}{2\sigma^2} \right] = \sum_{k=1}^n \frac{(\mu_0 - \mu_\infty)(2x_k - \mu_0 - \mu_\infty)}{2\sigma^2} = \\ &= \frac{\mu_0 - \mu_\infty}{\sigma^2} \left(X_N - \frac{\mu_0 - \mu_\infty}{2} n \right). \end{aligned}$$

Для простоты описания скажем, что $\mu_0 = \mu$, $\mu_\infty = 0$. Тогда

$$Z_n = \frac{\mu}{\sigma^2} \left(X_n - \frac{\mu}{2} n \right).$$

Следовательно, оптимальный критерий (1.6) выглядит так:

$$\varphi^*(x_1, \dots, x_n) = \begin{cases} 1, & Z_n \geq h, \\ 0, & Z_n < h. \end{cases}$$

Немного преобразуем условие, введя обозначение $H = \sigma^2 h / \mu$:

$$\varphi^*(x_1, \dots, x_n) = \begin{cases} 1, & X_n - \mu n / 2 \geq H, \\ 0, & X_n - \mu n / 2 < H. \end{cases} \quad (1.14)$$

Чему равны ошибки первого и второго рода для этого критерия? Для их подсчёта вспомним, что ξ_1, \dots, ξ_n — независимые и одинаково распределённые случайные величины. Если верна гипотеза H_∞ , то $E_\infty[\xi_1 + \dots + \xi_n] = 0$ и $D_\infty[\xi_1 + \dots + \xi_n] = \sigma^2 n$. Следовательно,

$$\begin{aligned} \alpha &= E_\infty[\varphi^*(\xi)] = P_\infty \left(\sum_{k=1}^n \xi_k - \frac{\mu n}{2} \geq H \right) = P_\infty \left(\frac{\xi_1 + \dots + \xi_n}{\sigma \sqrt{n}} \geq \frac{H + \mu n / 2}{\sigma \sqrt{n}} \right) = \\ &= 1 - \Phi \left(\frac{H + \mu n / 2}{\sigma \sqrt{n}} \right), \end{aligned}$$

где Φ — функция распределения стандартной нормальной случайной величины. Аналогично считается и β , только в данном случае $E_0[\xi_1 + \dots + \xi_n] = \mu n$ и $D_0[\xi_1 + \dots + \xi_n] = \sigma^2 n$:

$$\begin{aligned} \beta &= E_0[1 - \varphi^*(\xi)] = P_0 \left(\sum_{k=1}^n \xi_k - \frac{\mu n}{2} < H \right) = \\ &= P_0 \left(\frac{\xi_1 + \dots + \xi_n - \mu n}{\sigma \sqrt{n}} < \frac{H - \mu n / 2}{\sigma \sqrt{n}} \right) = \Phi \left(\frac{H - \mu n / 2}{\sigma \sqrt{n}} \right). \end{aligned}$$

В итоге мы получаем, что

$$\alpha = 1 - \Phi\left(\frac{H + \mu n/2}{\sigma\sqrt{n}}\right), \quad \beta = \Phi\left(\frac{H - \mu n/2}{\sigma\sqrt{n}}\right). \quad (1.15)$$

Далее, пусть C_γ — это квантиль порядка γ для стандартного нормального распределения, то есть корень уравнения $\Phi(C_\gamma) = \gamma$. Тогда, комбинируя это с (1.15), получаем, что

$$\frac{H + \mu n/2}{\sigma\sqrt{n}} = C_{1-\alpha}, \quad \frac{H - \mu n/2}{\sigma\sqrt{n}} = C_\beta. \quad (1.16)$$

Теперь несложно получить связь между (α, β) и (n, h) :

$$(C_{1-\alpha} - C_\beta)^2 = \left(\frac{\mu}{\sigma}\right)^2 n \implies n = \frac{(C_{1-\alpha} - C_\beta)^2}{(\mu/\sigma)^2}. \quad (1.17)$$

$$\frac{2H}{\sigma} = \sqrt{n}(C_{1-\alpha} + C_\beta) \implies H = \frac{C_{1-\alpha}^2 - C_\beta^2}{2\mu/\sigma^2} \implies h = \frac{C_{1-\alpha}^2 - C_\beta^2}{2}. \quad (1.18)$$

Следовательно, если мы хотим, чтобы у теста φ^* были вероятности ошибок первого и второго рода, равные α и β соответственно, то число наблюдений n и порог h будут задаваться формулами (1.17) и (1.18) соответственно.

2 Последовательные тесты

Постановка рассмотренной выше задачи различения двух статистических гипотез (H_0 и H_∞) предполагала, что решение принимается по заданному числу наблюдений N . При этом в классе $\Phi_a = \{\varphi: E_\infty[\varphi] \leq a\}$ оптимальный тест φ^* определялся формулой (1.6).

Давайте немного изменим постановку задачи. Пусть

$$\Phi_{\alpha,\beta} = \{\varphi: E_\infty[\varphi] \leq \alpha, E_0[1 - \varphi] \leq \beta\}. \quad (2.1)$$

То есть $\Phi_{\alpha,\beta}$ — это класс тех тестов φ , для которых вероятности ошибок первого и второго рода не превосходят α и β соответственно. Ранее мы показали, что для гипотез относительно среднего значения в наблюдениях, подчиняющихся гауссовскому распределению, число N необходимых наблюдений и соответствующий порог h задавались формулами (1.17) и (1.18) соответственно. Стоит заметить, что в данном случае число наблюдений является *неслучайной* величиной.

Рассмотрим теперь еще одну, новую, постановку задачи, принадлежащую А. Вальду, а именно задачу *последовательного* различения гипотез. В сущности, именно эта задача дала импульс развитию теории последовательного анализа и теории оптимальных правил остановки.

Но начнём с определений.

Определение 3. Пусть $x \in \mathbb{R}^\infty$, то есть $x = (x_1, \dots, x_n, \dots)$ — числовая последовательность с $x_i \in \mathbb{R}$. Борелевской σ -алгеброй в \mathbb{R}^∞ будем называть минимальную σ -алгебру $\mathcal{B}(\mathbb{R}^\infty)$, порождённую множествами вида

$$\{x: x_1 \in I_1, \dots, x_n \in I_n\}, \quad n \geq 1, \quad (2.2)$$

где I_1, \dots, I_n — это (борелевские) множества из $\mathcal{B}(\mathbb{R})$.

Будем считать, что на $(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$ заданы две вероятностные меры P_0 и P_∞ . Пусть координатно заданные случайные величины $\xi_k = \xi_k(x)$, где $\xi_k(x) = x_k$, являются независимыми и одинаково распределенными по каждой из мер P_0 и P_∞ . Дополнительно скажем, что у них есть плотность $f_\theta(x)$.

Предположим, что шаг за шагом мы получаем данные $x_1, x_2, \dots, x_n, \dots$, являющиеся наблюдениями над случайными величинами $\xi_1, \xi_2, \dots, \xi_n, \dots$. Мы хотим различить две гипотезы H_0 и H_∞ о том, какое «действует» распределение, P_0 или P_∞ , используя *последовательные тесты*, определение которого мы сейчас дадим. Но для него нужно знать, что такое марковский момент.

Определение 4. Пусть (Ω, \mathcal{F}, P) — вероятностное пространство с заданной на ней фильтрацией $\mathbb{F} = (\mathcal{F}_t)_{t \in T}$, где $T \subseteq [0, +\infty)$. Случайную величину τ , принимающую значения в $T \cup \{+\infty\}$, будем называть *марковским моментом* относительно фильтрации \mathbb{F} , если для любого $t \in T$ событие $\{\tau \leq t\}$ содержится в \mathcal{F}_t .

Определение 5. *Последовательным тестом* δ будем называть пару (τ, φ) , где

- $\tau = \tau(x)$ — марковский момент (или момент остановки) относительно потока $\{\mathcal{F}_n, n \geq 1\}$, где $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$.
- $\varphi = \varphi(x)$ — \mathcal{F}_τ -измеримая функция со значениями в $[0, 1]$, где $\mathcal{F}_\tau = \sigma(\xi_1, \dots, \xi_\tau)$.

Момент τ интерпретируется как момент прекращения наблюдений с последующим принятием решения $\varphi = \varphi(x)$, интерпретируемого как вероятность принятия гипотезы H_0 , когда наблюдениями являются x_1, \dots, x_τ .

Наиболее важными характеристиками тестов $\delta = (\tau, \varphi)$ являются средние длительности наблюдений $E_0[\tau]$ и $E_\infty[\tau]$ и вероятности ошибок первого и второго рода $\alpha(\varphi) = E_\infty[\varphi]$ и $\beta(\varphi) = E_0[1 - \varphi]$.

Пусть α и β — это какие-то два числа из $[0, 1]$. Пусть

$$\Delta(\alpha, \beta) = \{\delta = (\tau, \varphi) \mid E_\infty[\tau] < \infty, E_0[\tau] < \infty, \alpha(\varphi) \leq \alpha, \beta(\varphi) \leq \beta\}.$$

Другими словами, $\Delta(\alpha, \beta)$ — это класс последовательных тестов, для которых математическое ожидание момента остановки конечно, а вероятности ошибок первого и второго рода ограничены сверху α и β соответственно.

Определение 6. Будем говорить, что тест $\delta^* = (\tau^*, \varphi^*)$ *оптимален* в классе $\Delta(\alpha, \beta)$, если для любого другого теста $\delta \in \Delta(\alpha, \beta)$

$$E_\infty[\tau^*] \leq E_\infty[\tau] \text{ и } E_0[\tau^*] \leq E_0[\tau].$$

А. Вальд установил, что при определённых условиях такой оптимальный тест $\delta^* = (\tau^*, \varphi^*)$ действительно существует. Этот результат совершенно неочевиден, так как тест δ^* минимизирует два математических ожидания *одновременно*.

Доказательство этого факта весьма трудоёмко, так что ограничимся доказательством того, что существует *почти оптимальный* тест (смысл этого выражения будет объяснён позднее). Более того, мы упростим задачу, ограничившись только детерминированными решающими функциями $\varphi(x) = d(x)$, которые принимают только значения 0 и 1. Дальнейший анализ покажет, что такое ограничение легально.

Следующая лемма важна для доказательства почти оптимальности вальдовского теста — она дает оценку снизу для $E_\infty[\tau]$ и $E_0[\tau]$.

Лемма. Пусть $\delta = (\tau, \varphi)$ — последовательный тест с вероятностями ошибок первого и второго рода, равными α и β соответственно, причём $0 < \alpha + \beta < 1$. Тогда

$$\mathbb{E}_\infty[\tau] \geq \frac{\omega(\alpha, \beta)}{\rho_\infty} \text{ и } \mathbb{E}_0[\tau] \geq \frac{\omega(\beta, \alpha)}{\rho_0},$$

где

$$\omega(x, y) = x \log \frac{x}{1-y} + (1-x) \log \frac{1-x}{y}$$

$$\rho_\infty = \mathbb{E}_\infty \left[\log \frac{f_\infty(\xi_1)}{f_0(\xi_1)} \right], \quad \rho_0 = \mathbb{E}_0 \left[\log \frac{f_0(\xi_1)}{f_\infty(\xi_1)} \right].$$

Предполагается, что ρ_0 и ρ_∞ конечны.

Но для начала докажем одно простое утверждение.

Теорема 1 (Тождество Вальда). Пусть $\xi_1, \dots, \xi_n, \dots$ — последовательность независимых и одинаково распределённых величин, а τ — это случайная величина, не зависящая от ξ_k и принимающая натуральные значения, причём у всех случайных величин конечное матожидание. Тогда

$$\mathbb{E} \left[\sum_{k=1}^{\tau} \xi_k \right] = \mathbb{E}[\tau] \mathbb{E}[\xi_1].$$

Доказательство. Воспользуемся тем, что τ принимает значения в \mathbb{N} и тем, что она не зависит от ξ_k :

$$\begin{aligned} \mathbb{E} \left[\sum_{k=1}^{\tau} \xi_k \right] &= \sum_{n=1}^{\infty} \mathbb{E} \left[I_{\tau=n} \sum_{k=1}^{\tau} \xi_k \right] = \sum_{n=1}^{\infty} \mathbb{E} \left[I_{\tau=n} \sum_{k=1}^n \xi_k \right] = \\ &= \sum_{n=1}^{\infty} \mathbb{P}(\tau = n) \mathbb{E} \left[\sum_{k=1}^n \xi_k \right] = \mathbb{E}[\xi_1] \sum_{n=1}^{\infty} n \mathbb{P}(\tau = n) = \mathbb{E}[\xi_1] \mathbb{E}[\tau]. \end{aligned}$$

Тем самым мы получили желаемое. \square

Теперь приступим к доказательству самой леммы.²

Доказательство. Пусть L_n — это отношение правдоподобия для выборки размера n : $L_0 = 1$ и

$$L_n = \frac{f_0(x_1) f_0(x_2) \dots f_0(x_n)}{f_\infty(x_1) f_\infty(x_2) \dots f_\infty(x_n)}, \quad n \geq 1.$$

Далее, введём следующие обозначения:

$$\zeta_k = \log \frac{f_0(x_k)}{f_\infty(x_k)}, \quad Z_n = \log L_n = \sum_{k=1}^n \zeta_k.$$

Доказывать будем только неравенство для $\mathbb{E}_0[\tau]$ — неравенство для $\mathbb{E}_\infty[\tau]$ доказывается аналогично.

Рассмотрим $\mathbb{E}_0[Z_\tau]$. По тождеству Вальда

$$\mathbb{E}_0[Z_\tau] = \mathbb{E}_0 \left[\sum_{k=1}^{\tau} \zeta_k \right] = \mathbb{E}_0[\zeta_1] \mathbb{E}_0[\tau] = \rho_0 \mathbb{E}_0[\tau] \implies \mathbb{E}_0[\tau] = \frac{\mathbb{E}_0[Z_\tau]}{\rho_0}. \quad (2.3)$$

²Доказательство во всю использует интегралы по бесконечномерным пространствам. В формальности лезть не будем — это достаточно жёсткая математика — и скажем, что это рассуждение корректно и такие интегралы можно рассматривать.

Мы хотим доказать, что

$$\mathbb{E}_0[Z_\tau] \geq \omega(\beta, \alpha), \text{ где } \omega(x, y) = x \log \frac{x}{1-y} + (1-x) \log \frac{1-x}{y}. \quad (2.4)$$

Представим $\mathbb{E}_0[Z_\tau]$ в следующем виде:

$$\mathbb{E}_0[Z_\tau] = \int_{\{x: d(x)=0\}} Z_\tau d\mathbb{P}_0 + \int_{\{x: d(x)=1\}} Z_\tau d\mathbb{P}_0.$$

Далее, вспомним неравенство Йенсена: если $\varphi(\cdot)$ — выпуклая (вниз) борелевская функция, то

$$\varphi(\mathbb{E}[\xi]) \leq \mathbb{E}[\varphi(\xi)]. \quad (2.5)$$

Пользуясь им и тем, что $\beta = \mathbb{P}_0(d(x) = 0)$, получаем, что

$$\begin{aligned} \int_{\{x: d(x)=0\}} Z_\tau d\mathbb{P}_0 &= \mathbb{P}_0(d(x) = 0) \int_{\mathbb{R}^\infty} Z_\tau \mathbb{P}_0(dx | d(x) = 0) = \\ &= \beta \int_{\mathbb{R}^\infty} \log L_\tau \mathbb{P}_0(dx | d(x) = 0) = \\ &= -\beta \int_{\mathbb{R}^\infty} \log \frac{1}{L_\tau} \mathbb{P}_0(dx | d(x) = 0) = \\ &\geq -\beta \log \int_{\mathbb{R}^\infty} \frac{1}{L_\tau} \mathbb{P}_0(dx | d(x) = 0) \\ &= -\beta \log \left[\frac{1}{\mathbb{P}_0(d(x) = 0)} \int_{\{x: d(x)=0\}} \frac{1}{L_\tau} \mathbb{P}_0(dx) \right]. \end{aligned} \quad (2.6)$$

Теперь докажем, что

$$\int_{\{x: d(x)=0\}} \frac{1}{L_\tau} \mathbb{P}_0(dx) = \mathbb{P}_\infty(d(x) = 0) = 1 - \alpha.$$

Имеем

$$\int_{\{x: d(x)=0\}} \frac{1}{L_\tau} \mathbb{P}_0(dx) = \sum_{n=1}^{\infty} \int_{A_n} \frac{1}{L_\tau} \mathbb{P}_0(dx),$$

где $A_n = \{x: d(x) = 0, \tau(x) = n\}$. Теперь рассмотрим член суммы:

$$\begin{aligned} \int_{A_n} \frac{1}{L_\tau} \mathbb{P}_0(dx) &= \int_{A_n} \frac{1}{L_n} \mathbb{P}_0(dx) = \int_{A_n} \frac{f_\infty(x_1) f_\infty(x_2) \dots f_\infty(x_n)}{f_0(x_1) f_0(x_2) \dots f_0(x_n)} \mathbb{P}_0(dx) = \\ &= \int_{A_n} \mathbb{P}_\infty(dx) = \mathbb{P}_\infty(A_n) = \mathbb{P}_\infty(\{d = 0\} \cap \{\tau = n\}). \end{aligned}$$

Следовательно,

$$\int_{\{x: d(x)=0\}} \frac{1}{L_\tau} \mathbb{P}_0(dx) = \mathbb{P}_\infty(d = 0) = \mathbb{E}_\infty(1 - \alpha) = 1 - \alpha.$$

Отсюда следует, что

$$\int_{\{x: d(x)=0\}} Z_\tau dP_0 \geq -\beta \log \frac{1-\alpha}{\beta} = \beta \log \frac{\beta}{1-\alpha}.$$

Аналогично показывается, что

$$\int_{\{x: d(x)=1\}} Z_\tau dP_0 \geq (1-\beta) \log \frac{1-\beta}{\alpha}.$$

Комбинируя всё вышесказанное, получаем желаемое. \square

Эти неравенства полезны тем, что если мы сможем (для заданных α и β) построить тест $\delta^* = (\tau^*, d^*)$, для которого $E_\infty[\tau^*]$ и $E_0[\tau^*]$ равны $\omega(\alpha, \beta)/\rho_\infty$ и $\omega(\beta, \alpha)/\rho_0$, то этот тест будет оптимальным. Мы увидим далее, что в случае непрерывного времени (в задаче различения гипотез относительно среднего значения броуновского движения) такой тест действительно можно построить. Но ясно также, что если суметь построить тест $\tilde{\delta} = (\tilde{\tau}, \tilde{d})$, у которого значения $E_\infty[\tilde{\tau}]$ и $E_0[\tilde{\tau}]$ близки к предельным значениям, то это будет говорить о том, что такой тест “почти оптимален”. Займемся конструкцией таких тестов.

Пусть $A < 0$ и $B > 0$ — это какие-то константы. Положим

$$\tau_{AB} = \inf\{n \in \mathbb{N}: Z_n \geq B \text{ или } Z_n \leq A\}.$$

Далее, введём следующее решающее правило:

$$d_{AB} = \begin{cases} 1, & \text{если } Z_{\tau_{AB}} \geq B \\ 0, & \text{если } Z_{\tau_{AB}} \leq A \end{cases}$$

Тест $\delta_{AB} = (\tau_{AB}, d_{AB})$ был предложен А. Вальдом и называется *последовательным критерием отношений вероятностей*.

Теперь нужно понять, чему равны основные характеристики такого последовательного теста: вероятности ошибок первого и второго рода и матожидания момента остановки. Начнём с матожидания. Опять же, воспользуемся тождеством Вальда и пренебрежём перескоком за границы, то есть будем считать, что если $d_{AB} = 1(0)$, то $Z_{\tau_{AB}} = B(A)$:

$$\begin{aligned} E_0[\tau_{AB}] &= \frac{E_0[Z_{\tau_{AB}}]}{E_0[\zeta_1]} \approx \frac{B P_0(d_{AB} = 1) + A P_0(d_{AB} = 0)}{\rho_0} = \\ &= \frac{B E_0[d_{AB}] + A E_0[1 - d_{AB}]}{\rho_0} = \frac{B(1 - \beta) + A\beta}{\rho_0}. \end{aligned}$$

Аналогично считается $E_\infty[\tau_{AB}]$:

$$\begin{aligned} E_\infty[\tau_{AB}] &= \frac{E_\infty[Z_{\tau_{AB}}]}{E_\infty[\zeta_1]} \approx -\frac{B P_\infty(d_{AB} = 1) + A P_\infty(d_{AB} = 0)}{\rho_\infty} = \\ &= -\frac{B E_\infty[d_{AB}] + A E_\infty[1 - d_{AB}]}{\rho_\infty} = -\frac{B\alpha + A(1 - \alpha)}{\rho_\infty}. \end{aligned}$$

Займемся отысканием формул связи ошибок (α, β) с порогами (A, B) . С этой целью обратимся к так называемому *фундаментальному тождеству Вальда*. Пусть

$$g_\theta(\lambda) = E_\theta[e^{\lambda \zeta_1}] = E_\theta \left[\left(\frac{f_0(\xi_1)}{f_\infty(\xi_1)} \right)^\lambda \right], \quad \lambda \in \mathbb{R}.$$

Понятно, что для независимых и одинаково распределенных величин $\zeta_1, \zeta_2, \dots, \zeta_n$ выполнено равенство

$$\mathbb{E}_\theta[e^{\lambda(\zeta_1 + \dots + \zeta_n)}] = [g_\theta(\lambda)]^n, \text{ или же } \mathbb{E}_\theta\left[\frac{e^{\lambda Z_n}}{(g_\theta(\lambda))^n}\right] = \mathbb{E}_\theta[\exp\{\lambda Z_n - n \log g_\theta(\lambda)\}] = 1.$$

Будем предполагать, что все выражения, приведенные здесь и ниже, определены и конечны по крайней мере для “нужных” нам в дальнейшем значений λ .

Теорема 2 (Фундаментальное тождество Вальда). Для марковского момента τ и $\lambda \in \mathbb{R}$

$$\mathbb{E}_\theta\left[\frac{e^{\lambda Z_\tau}}{(g_\theta(\lambda))^\tau}\right] = \mathbb{E}_\theta[\exp\{\lambda Z_\tau - \tau \log g_\theta(\lambda)\}] = 1.$$

Доказательство этой теоремы требует знания теории мартингалов, поэтому оставим её без доказательства.

Пока что предположим, что мы нашли $\lambda = \lambda_0$ такое, что $g_\theta(\lambda_\theta) = 1$. Очевидным образом это выполняется при $\lambda_\theta = 0$, но есть и нетривиальные λ_θ . Действительно, распишем функцию по определению

$$g_\theta(\lambda) = \mathbb{E}_\theta\left[\left(\frac{f_0(\xi_1)}{f_\infty(\xi_1)}\right)^\lambda\right] = \int_{-\infty}^{+\infty} \left(\frac{f_0(x)}{f_\infty(x)}\right)^\lambda f_\theta(x) dx.$$

Отсюда понятно, что $g_0(-1) = \mathbb{E}_\infty[1] = 1$. Аналогично, $g_\infty(1) = \mathbb{E}_0[1] = 1$. Из этого следует, что

$$\mathbb{E}_0[\exp\{-Z_{\tau_{AB}}\}] = \mathbb{E}_\infty[\exp\{Z_{\tau_{AB}}\}] = 1.$$

Распишем матожидания, снова пренебрегая перескоком:

$$\begin{aligned} e^{-B} \mathbb{P}_0(d_{AB} = 1) + e^{-A} \mathbb{P}_0(d_{AB} = 0) &\approx 1, \\ e^B \mathbb{P}_\infty(d_{AB} = 1) + e^A \mathbb{P}_\infty(d_{AB} = 0) &\approx 1. \end{aligned}$$

Заменим вероятности на вероятности ошибок первого и второго рода:

$$\begin{aligned} e^{-B}(1 - \beta) + e^{-A}\beta &\approx 1, \\ e^B\alpha + e^A(1 - \alpha) &\approx 1. \end{aligned}$$

Отсюда получаем, что

$$\alpha \approx \frac{1 - e^A}{e^B - e^A} = \frac{e^{-A} - 1}{e^{B-A} - 1}, \quad \beta \approx \frac{1 - e^{-B}}{e^{-A} - e^{-B}} = \frac{e^B - 1}{e^{B-A} - 1}.$$

Теперь выразим A и B через α и β :

$$\begin{aligned} \frac{\beta}{1 - \alpha} &\approx \frac{e^B - 1}{e^{-A} - e^{B-A}} = e^A \implies A \approx \ln \frac{\beta}{1 - \alpha}, \\ \frac{1 - \beta}{\alpha} &\approx \frac{e^{B-A} - e^B}{e^{-A} - 1} = e^B \implies B \approx \ln \frac{1 - \beta}{\alpha}. \end{aligned}$$

Теперь подставим полученные значения для A и B в формулы для $\mathbb{E}_\theta[\tau_{AB}]$:

$$\begin{aligned} \mathbb{E}_0[\tau_{AB}] &\approx \frac{B(1 - \beta) + A\beta}{\rho_0} \approx \frac{(1 - \beta) \log \frac{1 - \beta}{\alpha} + \beta \log \frac{\beta}{1 - \alpha}}{\rho_0} = \frac{\omega(\beta, \alpha)}{\rho_0}. \\ \mathbb{E}_\infty[\tau_{AB}] &\approx -\frac{B\alpha + A(1 - \alpha)}{\rho_\infty} \approx -\frac{\alpha \log \frac{1 - \beta}{\alpha} + (1 - \alpha) \log \frac{\beta}{1 - \alpha}}{\rho_\infty} = \\ &= \frac{\alpha \log \frac{\alpha}{1 - \beta} + (1 - \alpha) \log \frac{1 - \alpha}{\beta}}{\rho_\infty} = \frac{\omega(\alpha, \beta)}{\rho_\infty}. \end{aligned}$$

Итак, в предположении пренебрежения эффектом перескока через границу

$$E_0[\tau_{AB}] \approx \frac{\omega(\beta, \alpha)}{\rho_0}, \quad E_\infty[\tau_{AB}] \approx \frac{\omega(\alpha, \beta)}{\rho_\infty}.$$

Теперь вспомним об **ограничении снизу**. Ясно, что в классе тестов (τ, d) таких, что $E_0[\tau] < \infty$, $E_\infty[\tau] < \infty$ и $P_0(d = 0) \leq \alpha$, $P_\infty(d = 1) \leq \beta$, тест (τ_{AB}, d_{AB}) является “почти оптимальным” (с точностью до пренебрежения эффектом перескока процессом $(Z_n)_{n \geq 1}$ порогов A и B).