Глава 2. ТЕОРИЯ СТАТИСТИЧЕСКОЙ ПРОВЕРКИ ГИПОТЕЗ

2. 1. Основные понятия теории статистической проверки гипотез

Рассмотрим следующую статистическую задачу. Пусть имеется случайная выборка

$$X = (x_1^{\mathbf{T}}, \ldots, x_n^{\mathbf{T}})^{\mathbf{T}} \in \mathcal{X} \subseteq \mathbb{R}^{nN}$$

из некоторого N-мерного распределения вероятностей $\mathbf{P}_\theta(\cdot)$, заданного на измеримом пространстве $(\Omega, \mathcal{F})$, где $\theta \in \Theta \subseteq \mathbb{R}^m$ — неизвестное истинное значение векторного параметра, Θ — параметрическое пространство, $\mathcal{X} \subseteq \mathbb{R}^{nN}$ — выборочное пространство. Пусть задано некоторое натуральное число $K \geq 2$ и определено разбиение параметрического пространства Θ на K областей:

$$\Theta = \bigcup_{k=0}^{K-1} \Theta_k,\ \Theta_k \cap \Theta_l = \emptyset,\ k \neq l.$$

Определение. Утверждение о неизвестном значении параметра θ, состоящее в том, что $\theta \in \Theta_k$, называется k-й гипотезой:

$$H_k:\ \theta \in \Theta_k, \quad k = 0, \ldots, K-1.$$

Гипотеза H_k называется простой, если мощность $|\Theta_k| = 1$, в противном случае — сложной (составной).

Если в действительности значение параметра $\theta \in \Theta_\nu$, $\nu \in \{0, \ldots, K-1\}$, то говорят, что гипотеза H_ν истинна.

Задача статистической проверки гипотез

$$H_0, \ldots, H_{K-1}$$

состоит в том, чтобы по наблюдаемой выборке X оптимальным образом оценить номер ν истинной гипотезы: $d = d(X) = k$ — выносим решение в пользу гипотезы H_k ($d = d(X)$ — статистическая оценка для ν). Возможно K решений ($k \in \{0, 1, \ldots, K-1\}$). Множество возможных решений обозначим

$$D = \{0, 1, \ldots, K-1\}, \ |D| = K,$$

и назовем *пространством решений*.

Определение. Решающим правилом (решающей функцией, критерием, тестом) в выше сформулированной задаче статистической проверки гипотез называется функциональное отображение выборочного пространства $\mathcal{X}$ в пространство решений D:

$$\mathcal{X} \overset{d(\cdot)}{\to} D. \tag{1}$$

Существуют два основных типа решающих правил (РП) (1).

Определение. Нерандомизированным РП называется отображение (1) следующего вида:

$$d = d(X) = \begin{cases} 0, & X \in \mathcal{X}_0; \\ \vdots & \\ K-1, & X \in \mathcal{X}_{K-1}, \end{cases}$$

где $\{\mathcal{X}_0, \ldots, \mathcal{X}_{K-1}\}$ — некоторое детерминированное борелевское разбиение выборочного пространства:

$$\mathcal{X} = \bigcup_{k=0}^{K-1} \mathcal{X}_k,\ \mathcal{X}_k \cap \mathcal{X}_l = \emptyset,\ k \neq l.$$

При этом, если выборка X фиксирована, то решение $d = d(X)$ не случайно.

Определение. Рандомизированным РП называется случайное отображение (1) следующего вида:

$$d = d(X, \omega),\ \omega \in \Omega,\ X \in \mathcal{X},\ d \in D,$$

причем если выборка X фиксирована, то решение $d = d(X, \omega)$ является дискретной случайной величиной с множеством значений D и некоторым дискретным распределением вероятностей:

$$\varphi_i = \varphi_i(X) = \mathbf{P}\{d = i|X\},\ i \in D.$$

При этом борелевские функции $\varphi_i = \varphi_i(X)$, $i \in D$, удовлетворяют следующим ограничениям:

$$0 \leq \varphi_i(X) \leq 1,\ i \in D; \quad \sum_{i \in D} \varphi_i(X) = 1, \quad X \in \mathcal{X},$$

и называются критическими функциями.

Укажем алгоритм принятия решения с помощью рандомизированного решающего правила.

1. По выборке X вычисляем значения критических функций: $\varphi_i = \varphi_i(X)$, $i \in D$, и определяем дискретное распределение вероятностей $\{\varphi_0(X), \varphi_1(X), \ldots, \varphi_{K-1}(X)\}$.

2. Проводим случайный эксперимент (жребий) со множеством исходов D и дискретным распределением вероятностей, найденным на шаге 1.

3. Регистрируем исход k этого жребия и принимаем решение $d = k$.

   Нерандомизированное решающее правило есть частный случай рандомизированного решающего правила, если критические функции принимают одно из двух возможных значений:

   $\varphi_i(X) \in \{0, 1\},\ X \in \mathcal{X};\ \mathcal{X}_i = \{X : \varphi_i(X) = 1\},$

2. 2. Решающее правило Неймана – Пирсона

Рассмотрим задачу проверки двух гипотез. Пусть имеется случайная выборка

$$X = (x_1^{\mathbf{T}}, \ldots, x_n^{\mathbf{T}})^{\mathbf{T}} \in \mathbb{R}^{nN},$$

образованная наблюдениями с плотностью распределения вероятностей

$$p(x;\theta),\ x \in \mathbb{R}^N,\ \theta \in \Theta,$$

где θ – неизвестное истинное значение параметра. Совместная плотность распределения всей выборки X определяется равенством:

$$p(X;\theta) = \prod_{l=1}^{n} p(x_l;\theta);$$

вероятностная мера, определенная при значении параметра θ:

$$\mathbf{P}_\theta(A) = \int_A p(X;\theta)dX,\ \ A \in \mathcal{B}^{nN};$$

а пространство параметров Θ разбито на две непересекающиеся области:

$$\Theta = \Theta_0 \cup \Theta_1,\ \ \Theta_0 \cap \Theta_1 = \emptyset.$$

Определены две гипотезы:

$$\begin{array}{lcl} H_0 & : & \theta \in \Theta_0, \\ H_1 & : & \theta \in \Theta_1. \end{array}$$

Так как $H_1 = \overline{H_0}$, то H_1 называется *альтернативой*, а H_0 — *основной* или *нулевой* гипотезой.

Задача заключается в построении решающего правила (теста) для проверки гипотез H_0, H_1 по выборке X. Построим рандомизированное решающее правило (РП общего вида):

$$d = d(X, \omega),\ X \in \mathbb{R}^{nN},\ \omega \in \Omega, \tag{2}$$

где $d \in D = \{0, 1\}$,

$$\begin{aligned} \mathbf{P}\{d(X,\omega) = 1|X\} &= \varphi(X), \\ \mathbf{P}\{d(X,\omega) = 0|X\} &= 1 - \varphi(X), \end{aligned} \tag{3}$$

а $0 \le \varphi(X) \le 1$ — некоторая произвольная критическая функция.

Существует бесконечно много РП $d(\cdot)$, различающихся критическими функциями $\varphi(\cdot)$. Как найти оптимальную критическую функцию $\varphi_*(\cdot)$? Сформулируем критерий оптимальности.

Определение. Принято говорить, что при принятии решений имеет место ошибка I рода, если на самом деле верна гипотеза H_0

($\nu = 0$), а принято решение $d = 1$ в пользу H_1. При этом вероятностью ошибки I рода называется число

$$\alpha = \alpha(\theta) ::= \mathbf{P}_\theta\{d(X,\omega) = 1\} = \mathbf{E}_\theta\{\varphi(X)\} =$$

$$= \int_{\mathbb{R}^{nN}} \varphi(X) p(X;\theta) dX, \ \theta \in \Theta_0.$$

Определение. Принято говорить, что имеет место ошибка II рода, если на самом деле верна гипотеза H_1 ($\nu = 1$), а принято решение $d = 0$ в пользу H_0. При этом вероятностью ошибки II рода называется число

$$\beta = \beta(\theta) ::= \mathbf{P}_\theta\{d(X,\omega) = 0\} = \mathbf{E}_\theta\{1 - \varphi(X)\} =$$

$$= 1 - \int_{\mathbb{R}^{nN}} \varphi(X) p(X;\theta) dX, \ \theta \in \Theta_1.$$

Определение. Мощностью решающего правила $d = d(X,\omega)$ называется вероятность правильного принятия альтеративы H_1 ($\nu = d = 1$):

$$w = w(\theta) = \int_{\mathbb{R}^{nN}} \varphi(X) p(X;\theta) dX, \ \theta \in \Theta_1.$$

Отметим, что $w(\theta) = 1 - \beta(\theta)$.

Было бы целесообразно найти функцию $\varphi_*(\cdot)$ так, что $\alpha = \beta = 0$. Очевидно, что в условиях регулярности это невозможно (также как невозможно было построить статистическую оценку с нулевой вариацией $V = 0$). Как поступать в этой ситуации?

Определение. Критическуюфункцию $\varphi(X)$ в рандомизированном решающем правиле (2), (3) надлежит выбирать таким образом, чтобы вероятность ошибки I рода не превосходила некоторого наперед заданного числа $\varepsilon \in (0, 1)$, а вероятность ошибки II рода была минимальной, то есть как решение следующей экстремальной задачи:

$$\sup_{\theta\in\Theta_0} \alpha(\theta) \le \varepsilon, \qquad \sup_{\theta\in\Theta_1} \beta(\theta) \to \min_{0\le\varphi(\cdot)\le 1},$$

что эквивалентно:

$$\sup_{\theta\in\Theta_0} \alpha(\theta) \le \varepsilon, \qquad \inf_{\theta\in\Theta_1} w(\theta) \to \max_{0\le\varphi(\cdot)\le 1}. \tag{4}$$

При этом оптимальная критическая функция $\varphi_*(X)$ и соответствующее ей решающее правило $d_*(X, \omega)$ называются соответственно критической функцией Неймана — Пирсона и решающим правилом Неймана — Пирсона, величина $\sup_{\theta\in\Theta_0} \alpha(\theta)$ — размером теста, а ε — уровнем значимости теста.

Отметим, что экстремальная задача (4) — задача вариационного исчисления. Нейман и Пирсон решили эту задачу лишь для случая, когда H_0, H_1 — простые гипотезы: $\Theta = \{\theta_0, \theta_1\}$, а $\Theta_0 = \{\theta_0\}$ и $\Theta_1 = \{\theta_1\}$ — одноточечные множества.

Упростим обозначения:

$$H_i: \ \theta = \theta_i, \ p_i(X) = p(X; \theta_i), \ X \in \mathbb{R}^{nN}; \quad i = 0, 1;$$

$$\theta_0, \theta_1 \in \Theta, \ \theta_0 \neq \theta_1;$$

$$\mathbf{P}_i(B) = \mathbf{P}_{\theta_i}(B), \ Q(B) = \frac{1}{2}(\mathbf{P}_0(B) + \mathbf{P}_1(B)), \ B \in \mathcal{B}^{n}$$

$$\mathbf{E}\{\cdot\} \equiv \mathbf{E}_{\theta_i}\{\cdot\}.$$

Определение. Статистикой отношения правдоподобия для проверки простых гипотез H_0, H_1 называется статистика

$$L = L(X) = \frac{p_1(X)}{p_0(X)} \geq 0.$$

Теорема. Для выше сформулированной задачи проверки простых гипотез H_0, H_1 справедливы следующие три утверждения:

1. Для любого наперед заданного ε, $0 < \varepsilon < 1$, найдутся такие постоянные $c_* > 0$, $\varkappa_* \in [0, 1]$, что решающее правило (тест) Неймана — Пирсона $d = d_*(X, \omega)$ с критической функцией

$$\varphi_*(X) = \begin{cases} 0, & L(X) < c_*; \\ \varkappa_*, & L(X) = c_*; \\ 1, & L(X) > c_*, \end{cases} \tag{5}$$

имеет размер, в точности равный ε;

2. Тест Неймана — Пирсона $d = d_*(X, \omega)$, определяемый (5), имеет наибольшую мощность w_* среди всех тестов, размер которых не превосходит ε;

3. Тест $d = d_*(X, \omega)$ — единственный с точностью до множества B нулевой меры: $Q(B) = 0$.

Доказательство. Запишем экстремальную задачу (4) с учетом определений , в явном виде:

$$\alpha = \int_{\mathbb{R}^{nN}} \varphi(X) p_0(X) dX \le \varepsilon,$$

$$w = \int_{\mathbb{R}^{nN}} \varphi(X) p_1(X) dX \to \max_{\varphi(\cdot)}. \qquad (6)$$

Доказательство разобьем на три части.

1. Введем в рассмотрение функцию распределения статистики $L = L(X)$ при верной гипотезе H_0:

$$F_0(y) = \mathbf{P}_0\{L(X) < y\}, \ y \geq 0.$$

Считая произвольными c_* и $\varkappa_*$, вычислим для теста (5) вероятность ошибки I рода:

$$\alpha = \mathbf{E}_0\{\varphi(X)\} = \varkappa_* \mathbf{P}_0\{L = c_*\} + \mathbf{P}_0\{L > c_*\} =$$

$$= \varkappa_*(F_0(c_*+0) - F_0(c_*)) + (1 - F_0(c_*+0)). \ (7)$$

В этом соотношении подберем произвольные константы c_*, $\varkappa_*$ таким образом, чтобы

$$\alpha = \alpha_* = \varepsilon.$$

Для этого выберем c_* так, чтобы

$$F_0(c_*) \leq 1 - \varepsilon \leq F_0(c_* + 0).$$

Возможны два случая:

а) c_* – точка непрерывности функции $F_0(\cdot)$; тогда

$$c_* = F_0^{-1}(1-\varepsilon),$$

и из (7) имеем

$$\alpha_* = \varepsilon \quad \forall \varkappa_*;$$

б) c_* – точка разрыва; в этом случае выберем

$$\varkappa_* = \frac{F_0(c_*+0) - (1-\varepsilon)}{F_0(c_*+0) - F_0(c_*)}.$$

Подставляя это значение в (7), получаем

$$\alpha_* = \varepsilon.$$

2. Для критической функции $\varphi_*(\cdot)$ имеем

$$\alpha_* = \varepsilon, \quad w_* = \int_{\mathbb{R}^{nN}} \varphi_*(X) p_1(X) dX. \tag{8}$$

Выберем произвольную критическую функцию $\tilde{\varphi}(X)$, для которой обозначим

$$\tilde{\alpha} = \int_{\mathbb{R}^{nN}} \tilde{\varphi}(X) p_0(X) dX \le \varepsilon,$$

$$\tilde{w} = \int_{\mathbb{R}^{nN}} \tilde{\varphi}(X) p_1(X) dX.$$

Покажем: $\tilde{w} \le w_*$. Легко проверить поточечно, используя (5), что

$$\Delta = \Delta(X) ::= (p_1(X) - c_* p_0(X)) \times \\ \times (\varphi_*(X) - \tilde{\varphi}(X)) \ge 0, \ \forall X \in \mathbb{R}^{nN}.$$

Отсюда с учетом (8) получаем

$$\int_{\mathbb{R}^{nN}} \Delta(X) dX = w_* - c_* \alpha_* - \tilde{w} + c_* \tilde{\alpha} = \\ = w_* - \tilde{w} - c_*(\varepsilon - \tilde{\alpha}) \ge 0.$$

Следовательно,

$$w_* \ge \tilde{w} + c_*(\varepsilon - \tilde{\alpha}) \ge \tilde{w}.$$

3. В экстремальной задаче (6) значения интегралов не изменятся, если значение $\varphi_*(X)$ изменить на множестве B: $Q(B) = 0$.

■

Следствие. Если функция распределения $F_0(\cdot)$ статистики отношения правдоподобия

$$L = L(X)$$

при гипотезе H_0 – непрерывная функция, то РП Неймана – Пирсона является нерандомизированным и имеет следующий вид:

$$d = d_*(X) = \begin{cases} 0, & L(X) < c_*; \\ 1, & L(X) \geq c_*, \end{cases}$$

где

$$c_* = F_0^{-1}(1 - \varepsilon).$$

Доказательство. Обратимся к первому пункту доказанной выше теоремы. Если $F_0(\cdot)$ непрерывна, то $\varkappa_*$ можно выбирать произвольно на отрезке $[0, 1]$, поэтому выберем

$$\varkappa_* = 1,$$

тогда, согласно (5),

$$\varphi_*(X) \in \{0; 1\},$$

то есть жребий исчезает и имеет место детерминированное (нерандомизированное) решающее правило.

■

Пример. Есть два нормальных распределения с одинаковыми дисперсиями

$$\sigma_1 = \sigma_2 = \sigma$$

и разными математическими ожиданиями

$$\theta_1 < \theta_0.$$

Для случайной выборки

$X = (x_1, \ldots, x_n)^{\mathbf{T}} \in \mathcal{X} \subseteq \mathbb{R}^n,\ x_t \in \mathbb{R},\ n = 1, \ldots, N,$

нужно построить решающее правило Неймана-Пирсона с уровнем значимости ε для решения задачи:

$$\begin{array}{lcl} H_0 & : & \theta = \theta_0, \\ H_1 & : & \theta = \theta_1. \end{array} \tag{9}$$

Р е ш е н и е. Отношение правдоподобия в данном случае имеет вид:

$$L(X) = \frac{p_1(X)}{p_0(X)} = \frac{\prod_{t=1}^{n} n_1(x_t \,|\, \theta_1, \sigma^2)}{\prod_{t=1}^{n} n_1(x_t \,|\, \theta_0, \sigma^2)}$$

где

$$n_1(z \,|\, \theta_i, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z-\theta_i)^2}{2\sigma^2}\right),$$

а $z \in \mathbb{R},\ i \in \{0,\, 1\}$.

Таким образом, отношение правдоподобия для этого примера после ряда очевидных элементарных преобразований примет вид:

$$L(X) = \frac{\prod_{t=1}^{n} \exp\left(-\frac{(x_t-\theta_1)^2}{2\sigma^2}\right)}{\prod_{t=1}^{n} \exp\left(-\frac{(x_t-\theta_0)^2}{2\sigma^2}\right)} =$$

$$= \prod_{t=1}^{n} \exp\left(\frac{(x_t-\theta_0)^2}{2\sigma^2} - \frac{(x_t-\theta_1)^2}{2\sigma^2}\right) =$$

$$= \exp\left(n\frac{\theta_1-\theta_0}{\sigma^2}\overline{x}_t - n\frac{\theta_1^2-\theta_0^2}{2\sigma^2}\right)$$

Очевидно, что $L(X)$ является возрастающей функцией от $\overline{x}$. Поэтому для каждого $c_* > 0$ существует единственное число $\varkappa_* \in [0,1]$, при котором условие $L(X) > c_*$ равносильно $\overline{x} > \varkappa_*$. Заметим, что при истинном математическом ожидании θ_0, то есть в случае, когда истинной является гипотеза H_0, выборочное среднее

$$\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$$

имеет нормальное распределение

$$\mathcal{L}\{\overline{x}\} = \mathcal{N}\left(\theta_0, \frac{\sigma^2}{n}\right).$$

Действительно, так как

$$\mathcal{L}\{x_j\} = \mathcal{N}\left(\theta_0,\, \sigma^2\right),\; j = 1, 2, \ldots, n,$$

то характеристическая функция случайной величины x_j вычисляется по формуле

$$f_{x_j}(t) = \exp\{-i\theta_0 t\}\cdot\exp\left\{-\frac{1}{2}\sigma^2 t^2\right\},\; j = 1, 2, \ldots, n, \tag{10}$$

характеристическая функция суммы

$$S = \sum_{j=1}^{n} x_j \; -$$

по формуле

$$f_S(t) = \prod_{j=1}^{n} f_{x_j}(t) = \exp\{-in\theta_0 t\}\cdot\exp\left\{-\frac{1}{2}n\,(\sigma t)^2\right\},$$

а характеристическая функция выборочного среднего $\overline{x}$:

$$\overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \; -$$

по формуле

$$f_{\overline{x}}(t) = \exp\{-i\theta_0 t\}\cdot\exp\left\{-\frac{1}{2}\frac{\sigma^2}{n}t^2\right\}, \tag{11}$$

Сравнивая равенства (11) и (10) видим, что функция $f_{\overline{x}}(t)$ из (11) – характеристическая функция нормального распределения $\mathcal{N}\left(\theta_0, \frac{\sigma^2}{n}\right)$

Аналогично рассуждая, можно показать, что случайная величина

$$y = \sqrt{n}\frac{\overline{x} - \theta_0}{\sigma}$$

имеет стандартное нормальное распределение:

$$\mathcal{L}\{y\} = \mathcal{N}\left(0, 1\right).$$

А неравенство

$$\overline{x} > \varkappa_*$$

равносильно неравенству

$$y > \zeta,$$

где

$$\zeta = \sqrt{n}\frac{\varkappa_* - \theta_0}{\sigma}.$$

В решающем правиле Неймана-Пирсона вероятность ошибки первого рода должна равняться ε то есть:

$$\alpha = \mathbf{P}_0\left\{L(X) > c_*\right\} = \mathbf{P}_0\left\{\overline{x} > \varkappa_*\right\} =$$

$$= \mathbf{P}_0\left\{y > \zeta\right\} = \varepsilon.$$

С другой стороны,

$$\mathbf{P}_0\left\{y > \zeta\right\} = 1 - \Phi(\zeta).$$

Следовательно,

$$1 - \Phi(\zeta) = \varepsilon.$$

Отсюда мы находим квантиль

$$\zeta = \Phi^{-1}(1 - \varepsilon),$$

критическое значение

$$\varkappa_* = \theta_0 + \zeta \frac{\sigma}{\sqrt{n}},$$

и в итоге получаем решающее правило

$$\begin{cases} \theta = \theta_0, & \text{если } \overline{x} \le \varkappa_* = \theta_0 + \zeta \frac{\sigma}{\sqrt{n}}, \\ \theta = \theta_1, & \text{если } \overline{x} > \varkappa_* = \theta_0 + \zeta \frac{\sigma}{\sqrt{n}}, \end{cases}$$

где

$$\zeta = \Phi^{-1}(1 - \varepsilon).$$

2. 3. Проверка простой гипотезы против сложной альтернативы

Пусть на $\mathbb{R}^N$ задано семейство вероятностных мер $\{\mathbf{P}_\theta \,|\, \theta \in \Theta\}$, и в множестве параметров Θ фиксировано одно значение θ_0. Требуется проверить простую гипотезу

$$H_0: \;\theta = \theta_0$$

против альтернативы общего вида

$$H_1: \;\theta \neq \theta_0.$$

Обычно это делают так. Выбирают какую-нибудь скалярную статистику $T(X)$, для которой при $\theta = \theta_0$ известна функция распределения

$$F(z) = \mathbf{P}_{\theta_0}\{T(X) \le z\}.$$

В области изменения $T(X)$ находят такой отрезок $[\Delta_1, \Delta_2]$, что

$$\mathbf{P}_{\theta_0}\{T(X) \in [\Delta_1, \Delta_2]\} \ge 1 - \varepsilon.$$

По нему определяют решающее правило

$$\begin{cases} \theta = \theta_0, & \text{если } T(X) \in [\Delta_1, \Delta_2], \\ \theta \neq \theta_0, & \text{если } T(X) \notin [\Delta_1, \Delta_2]. \end{cases} \tag{12}$$

По построению оно имеет уровень значимости ε. Затем для каждого $\theta \neq \theta_0$ проверяют, будет ли вероятность

$$\mathbf{P}_\theta \{T(X) \in [\Delta_1, \Delta_2]\}$$

стремиться к нулю при возрастании объема выборки X. Если это так, то решающее правило (12) состоятельно.

Предположим, что функция распределения

$$F(z) = \mathbf{P}_{\theta_0} \{T(X) \leq z\}.$$

непрерывна. Тогда решающее правило (12) можно записать в равносильной форме

$$\begin{cases} \theta = \theta_0, & \text{если } F(T(X)) \in [\delta_1, \delta_2], \\ \theta \neq \theta_0, & \text{если } F(T(X)) \notin [\delta_1, \delta_2], \end{cases} \tag{13}$$

где $\delta_i = F(\Delta_i)$, $i \in \{0, 1\}$.

Замечание. По определению случайная величина x является функцией на каком-то вероятностном пространстве

$$x : \Omega \to \mathbb{R}.$$

Рассмотрим новую случайную величину $F(x)$, являющуюся результатом подстановки случайной величины x в качестве аргумента ее функции распределения $F(\cdot)$, то есть композицию двух функций

$$x : \Omega \to \mathbb{R}.$$

и

$$F : \mathbb{R} \to [0, 1].$$

Полученную таким образом случайную величину $F(x)$ называют p-уровнем исходной случайной величины x. Известно, что если функция распределения случайной величины x непрерывна на вещественной оси, то ее p-уровень $F(x)$ равномерно распределен на отрезке [0, 1].

Следовательно, случайная величина $F\left(T(X)\right)$ является p-уровнем статистики $T(X)$. И она равномерно распределена на отрезке [0, 1] (при $\theta = \theta_0$). Поэтому

$$\delta_2 - \delta_1 = \mathbf{P}_{\theta_0}\left\{F\left(T(X)\right) \in [\delta_1, \delta_2]\right\} \geq 1 - \varepsilon.$$

Если функция распределения $F(z)$ разрывна, то вместо решающего правила (13) можно использовать следующее:

$$\begin{cases} \theta = \theta_0, & \text{если } F\left(T(X)\right) > \delta_1 \\ & \text{и } F\left(T(X) - 0\right) < \delta_2, \\ \theta \neq \theta_0, & \text{если } F\left(T(X)\right) \leq \delta_1 \\ & \text{или } F\left(T(X) - 0\right) \geq \delta_2, \end{cases} \qquad (14)$$

где

$$\delta_2 - \delta_1 = 1 - \varepsilon.$$

Замечание. Известно,что неравенство

$$F\left(T(X)\right) > \delta_1$$

выполняется с вероятностью не меньше $1 - \delta_1$, а неравенство

$$F\left(T(X) - 0\right) < \delta_2$$

выполняется с вероятностью не меньше δ_2. Вероятность того, что хотя бы одно из них нарушается, не превосходит

$$\delta_1 + (1 - \delta_2) = \varepsilon.$$

Это доказывает, что последний тест имеет уровень значимости ε.

Пример 1. Пусть

$X = (x_1, \ldots, x_n)^{\mathbf{T}} \in \mathcal{X} \subseteq \mathbb{R}^n$, $x_t \in \mathbb{R}$, $t = 1, \ldots, N$, –

случайная выборка из нормального распределения $\mathcal{N}(a, \sigma^2)$ с известной дисперсией σ^2. Построить критерий для проверки гипотезы $a = a_0$.

Р е ш е н и е. Известно, что при $a = a_0$ статистика

$$T(X) = \frac{1}{\sigma\sqrt{n}} \sum_{t=1}^{n} (x_t - a_0) = \frac{\sqrt{n}}{\sigma} (\overline{x} - a_0)$$

имеет стандартное нормальное распределение $\mathcal{N}(0,1)$

$$\mathcal{L}\{T(X)\} = \mathcal{N}(0,1).$$

Подберем такое число Δ, при котором

$$\mathbf{P}\{|T(X)| \le \Delta\} = 1 - \varepsilon.$$

Очевидно, это

$$\Delta = \Phi^{-1}\left(1 - \frac{\varepsilon}{2}\right).$$

Соответствующий критерий с уровнем значимости ε имеет вид

$$\begin{cases} a = a_0, & \text{если } |T(X)| \le \Delta, \\ a \ne a_0, & \text{если } |T(X)| > \Delta, \end{cases} \qquad (15)$$

где $\Delta = \Phi^{-1}\left(1 - \frac{\varepsilon}{2}\right)$.

2. 4. Краткое описание нескольких часто используемых вероятностных распределений

χ^2-распределение с n степенями свободы. Пусть $\xi_1, \xi_2, \ldots, \xi_n$ – независимые случайные величины, распределенные по стандартному нормальному закону:

$$\mathcal{L}\{\xi_i\} = \mathcal{N}(0,\, 1),\, i = \overline{1,\, n}.$$

Тогда случайная величина

$$\chi_n^2 = \xi_1^2 + \xi_1^2 + \ldots + \xi_n^2$$

имеет распределение «хи-квадрат» с n степенями свободы

$$\mathcal{L}\{\chi_n^2\} = \chi_n^2.$$

Его математическое ожидание равно n, а дисперсия $2n$.

t^2-распределение Стьюдента с n степенями свободы. Пусть $\xi_0, \xi_1, \ldots, \xi_n$ – независимые случайные величины со стандартным нормальным распределением

$$\mathcal{L}\{\xi_i\} = \mathcal{N}(0,\, 1),\, i = \overline{0,\, n}.$$

Тогда распределение случайной величины

$$t_n = \sqrt{n}\frac{\xi_0}{\sqrt{\xi_1^2 + \xi_1^2 + \ldots + \xi_n^2}}$$

называется распределением Стьюдента с n степенями свободы. При возрастании n оно сходится к стандартному нормальному распределению.

Распределение Фишера с n и m степенями свободы. Пусть χ_n^2 и χ_m^2 – две независимые случайные величины, имеющие распределения «хи-квадрат» с n и m степенями свободы соответственно. Тогда распределение случайной величины

$$F_{n,m} = \frac{m\chi_n^2}{n\chi_m^2}$$

называется распределением Фишера с n и m степенями свободы.

2. 5. Критерии согласия

Предположим, что мы наблюдаем выборку $X = \{x_1, x_2, \dots, x_n\}$ из некоторого распределения $\mathcal{L}\{x\}$ с функцией распределения $F(z)$. Требуется определить, совпадает ли $F(z)$ с другой (заранее заданной) функцией распределения $F_0(z)$. Иначе говоря, нам нужно проверить простую гипотезу

$$H_0 : F(z) \equiv F_0(z)$$

против альтернативы общего вида

$$H_1 : F(z) \not\equiv F_0(z).$$

При такой постановке задачи H_0 называется гипотезой согласия (потому что она утверждает, что выборка X «согласуется» с распределением $F_0(z)$), а соответствующее ей решающее правило называется критерием согласия. Наиболее популярны два критерия согласия: χ^2-критерий Пирсона и критерий Колмогорова.

χ^2-критерий Пирсона. Произвольным образом разобьем вещественную прямую на K частей точками

$$-\infty = b_0 < b_1 < \dots < b_{K-1} < b_K = \infty.$$

Положим

$$\Gamma_k = (b_{k-1},\, b_k].$$

Вероятность того, что случайная величина x с функцией распределения $F_0(z)$ попадет в Γ_k, равна

$$\mathbf{P}\{x \in \Gamma_k\} = p_k = F_0(b_k) - F_0(b_{k-1}),\; k = \overline{1,\, K}.$$

Будем предполагать, что все $p_k > 0$, а для каждой выборки $X = \{x_1, x_2, \ldots, x_n\}$ обозначим через ν_k число выборочных значений x_i, попавших в Γ_k, а через $\widehat{p}_k$ – эмпирические вероятности

$$\widehat{p}_k = \frac{\nu_k}{n}.$$

Определение. χ^2-статистика Пирсона – это

$$\chi^2(X) = \sum_{k=1}^{K} \frac{(\nu_k - np_k)^2}{np_k} = n \sum_{k=1}^{K} \frac{(\widehat{p}_k - np_k)^2}{p_k}. \tag{16}$$

Лемма. Если верна гипотеза H_0, то при $n \to \infty$ имеет место сходимость к χ^2-распределению с $K-1$ степенями свободы:

$$\mathcal{L}\{\chi^2\} \to \chi^2_{K-1},$$

или

$$\mathbf{P}_{H_0}\{\chi^2 < y\} \to F_{\chi^2_{m-1}}(y), \ n \to +\infty, \quad y \geq 0, \tag{17}$$

где $F_{\chi^2_{K-1}}(\cdot)$ – функция χ^2-распределения с $K-1$ степенями свободы.

Д о к а з а т е л ь с т в о. Обозначим нормированные частоты, входящие в (16):

$$\nu_k^*(n) = \frac{\nu_k - np_k}{\sqrt{n}}, \ k = 1, \ldots, K; \quad \sum_{k=1}^{K} \nu_k^*(n) \equiv 0, \tag{18}$$

и определим $(K-1)$-вектор нормированных частот

$$\nu^*(n) = (\nu_1^*(n), \ldots, \nu_{K-1}^*(n))^{\mathbf{T}} \in \mathbb{R}^{K-1}.$$

Заметим, что в силу условия нормировки в (18) $\nu_K^*(n)$ линейно зависит от $\nu^*(n)$:

$$\nu_K^*(n) = -\sum_{l=1}^{K-1} \nu_l^*(n). \tag{19}$$

В асимптотике $n \to +\infty$ воспользуемся центральной предельной теоремой для $\nu^*(n)$ (многомерный аналог теоремы Муавра – Лапласа):

$$\mathcal{L}\{\nu^*(n)\} \to \mathcal{N}_{K-1}(\mathbf{0}_{K-1}, \Sigma), \tag{20}$$

где

$$\Sigma = (\sigma_{ij})_{i,j=1}^{K}, \text{ –}$$

асимптотическая ковариационная матрица, для элементов которой справедлива формула:

$$\sigma_{ij} = p_i(\delta_{ij} - p_j),\ i, j = 1, \ldots, K-1.$$

Непосредственной проверкой можно убедиться, что в этом случае для обратной матрицы справедливо выражение

$$\Sigma^{-1} = (\overline{\sigma}_{ij})_{i,j=1}^{K},$$

$$\overline{\sigma}_{ij} = \frac{1}{p_m} + \frac{\delta_{ij}}{p_i},\ i, j = 1, \ldots, m-1. \qquad (21)$$

Из (16), (18)–(21) имеем представление для χ^2-статистики в виде квадратичной формы:

$$\chi^2 = \sum_{k=1}^{K-1} \frac{(\nu_k^*(n))^2}{p_k} + \frac{(\sum_{l=1}^{K-1} \nu_l^*(n))^2}{p_K} =$$

$$= \sum_{k=1}^{K-1} \sum_{l=1}^{K-1} \nu_k^*(n) \nu_l^*(n) \overline{\sigma}_{kl} =$$

$$= (\nu^*(n))^{\mathbf{T}} \Sigma^{-1} \nu^*(n).$$

Введем в рассмотрение случайный вектор $\xi(n) = (\xi_1(n), \xi_2(n), \ldots, \xi_{K-1}(n))^{\mathbf{T}}$ такой, что

$$\xi(n) = (\Sigma^{1/2})^{-1}\nu^*(n) \in \mathbb{R}^{K-1},$$

где матрица $\Sigma^{1/2}$ – решение по Y матричного уравнения

$$Y^{\mathbf{T}}\Sigma^{-1}Y = \mathbf{I}_{K-1}$$

или

$$\Sigma = \Sigma^{1/2}(\Sigma^{1/2})^{\mathbf{T}}.$$

Тогда χ^2-статистика представима в виде суммы квадратов:

$$\chi^2 = \xi^{\mathbf{T}}(n)\xi(n) = \sum_{k=1}^{K-1} \xi_k^2(n),$$

причем в силу (20)

$$\mathcal{L}\{\xi(n)\} \to \mathcal{N}_{K-1}(\mathbf{0}_{K-1}, \mathbf{I}_{K-1}).$$

Отсюда по определению χ^2-распределения получаем (17).

$\blacksquare$

Определение. χ^2-критерий согласия Пирсона имеет вид

$$d = d(X) = \begin{cases} 0, & \chi^2 < \Delta, \\ 1, & \chi^2 \geq \Delta, \end{cases} \quad \Delta = F^{-1}_{\chi^2_{K-1}}(1-\varepsilon), \tag{22}$$

где $F^{-1}_{\chi^2_{K-1}}(1-\varepsilon)$ – квантиль χ^2-распределения с $K-1$ степенЯМИ свободы уровня $1-\varepsilon$, $0 < \varepsilon < 1$.

Из теоремы вытекает, что вероятность ошибки первого рода

$$\mathbf{P}\left\{\chi^2(X) \geq \Delta \,|\, H_0\right\} \to \varepsilon$$

при $n \to \infty$. Мы всегда можем снизить эту вероятность до приемлемого уровня за счет выбора достаточно малого ε.

С другой стороны, предположим, что выборка X подчиняется распределению $F(z)$, отличному от $F_0(z)$. Тогда числа

$$q_k = F(b_k) - F(b_{k-1})$$

скорее всего будут отличаться от

$$p_k = F_0(b_k) - F_0(b_{k-1}).$$

По усиленному закону больших чисел эмпирические вероятности $\widehat{p}_k$ почти наверное будут сходиться к q_k при $n \to \infty$. Поэтому отношение

$$\frac{\chi^2(X)}{n}$$

будет сходиться к числу

$$c = \sum_{k=1}^{K} \frac{(q_k - p_k)^2}{p_k} > 0,$$

сама случайная величина $\chi^2(X)$ будет неограниченно возрастать, и выбор будет делаться в пользу гипотезы

$$H_1 : F(z) \not\equiv F_0(z).$$

Другими словами, рассматриваемый критерий состоятелен.

К сожалению, вероятность ошибки второго рода стремится к нулю не равномерно по отношению к распределению $F(z) \not\equiv F_0(z)$. Чем меньше отличается $F(z)$ от $F_0(z)$, тем больше нужно брать размер выборки, чтобы гарантировать малость этой вероятности. А в случае, когда распределение $F(z)$ может быть сколь угодно близко к $F_0(z)$, мы ни при каком n не можем утверждать, что вероятность ошибки второго рода мала.

Замечание. χ^2-критерий согласия Пирсона справедлив и для N-мерных ($N > 1$) функций распределения $F(x)$, $x \in \mathbb{R}^N$.

Обобщение χ^2-критерия согласия Пирсона для сложных гипотез согласия. χ^2-критери Пирсона допускает обобщение для случая сложных гипотез согласия:

$$H_0 : F(\cdot) \in F_0,$$

где

$$F_0 = \{F_0(z;\theta), z \in \mathbb{R}^N : \theta \in \Theta \subseteq \mathbb{R}^m\} -$$

некоторое параметрическое семейство функций распределения с m-мерным параметром θ. Появляется дополнительное ограничение: $K > m + 1$, и критерий согласия имеет вид

$$d = d^0(X) = \begin{cases} 0, & \inf\limits_{\theta} \chi^2(\theta) < \Delta, \\ 1, & \inf\limits_{\theta} \chi^2(\theta) \geq \Delta, \end{cases}$$

где

$$\Delta = F^{-1}_{\chi^2_{K-m-1}}(1 - \varepsilon) -$$

квантиль χ^2-распределения с $K - m - 1$ степенями свободы уровня $1 - \varepsilon$, $0 < \varepsilon < 1$,

$$\chi^2(\theta) = \sum_{k=1}^{K} \frac{(\nu_k - np_k(\theta))^2}{np_k(\theta)} \geq 0,$$

$$p_k(\theta) = \mathbf{P}_\theta\{x_i \in \Gamma_k\},\ k = 1, \ldots, K.$$

Статистика

$$\tilde{\theta} = \arg\min_\theta \chi^2(\theta)$$

называется χ^2-оценкой параметра θ.

Критерий Колмогорова. Будем предполагать, что гипотеза согласия $H_0 : F(\cdot) \equiv F_0(\cdot)$ – простая, причем функция $F_0(\cdot)$ непрерывна. Построим по наблюдаемой выборке X выборочную функцию распределения, изученную ранее:

$$\widehat{F}(z) = F_n(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(z - x_i) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(z - x_{(i)}),$$

где $z \in \mathbb{R}^1, x_{(1)} < x_{(2)} < \ldots < x_{(n)}$ – вариационный ряд выборки X.

Выборочная функция распределения – сильно состоятельная оценка истинной функции распределения:

$$F_n(z) \xrightarrow[n \to +\infty]{\mathbf{P}=1} F(z),\ z \in \mathbb{R}^1.$$

Определение. Расстоянием Колмогорова между выборочной функцией распределения $F_n(x)$

и гипотетической функцией распределения $F_0(x)$ называется величина

$$D_n = \sup_{-\infty<z<+\infty} |F_n(z) - F_0(z)|, \qquad (23)$$

$0 \le D_n \le 1$.

Лемма. Если верна гипотеза H_0, то случайные величины $u_1 = F_0(x_1)$, $u_2 = F_0(x_2)$, ..., $u_n = F_0(x_n)$ независимы и одинаково распределены со стандартным равномерным распределением: $\mathcal{L}_{H_0}\{u_i\} = \mathcal{R}[0,1]$, $i = 1, \ldots, n$.

Доказательство. Независимость $\{u_i\}$ вытекает из независимости $\{x_i\}$ и соответствующей теоремы о борелевских функциональных преобразованиях. Вычислим функцию распределения для случайной величины u_i при условии, что верна гипотеза H_0:

$$F_{u_i}(y) ::= \mathbf{P}_{H_0}\{u_i < y\} = \mathbf{P}_{H_0}\{x_i < F_0^{-1}(y)\} =$$

$$= F_0(F_0^{-1}(y)) = y,\ 0 \le y \le 1.$$

Это и есть функция распределения стандартного равномерного закона $\mathcal{R}[0,1]$.

$\blacksquare$

Теорема. Если верна нулевая гипотеза H_0, то распределение вероятностей статистики Колмогорова D_n, определяемой формулой (23), не зависит от гипотетической функции распределения $F_0(\cdot)$:

$$\mathcal{L}_{H_0}\{D_n\} = \mathcal{L}\{D_n^*\},$$

$$D_n^* = \sup_{0 \le u \le 1} |\Psi_n(u) - u|;$$

$$\Psi_n(u) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(u - u_i),$$

где $\{u_i\}$ – независимые в совокупности случайные величины, имеющие стандартное равномерное распределение: $\mathcal{L}\{u_i\} = \mathcal{R}[0,1]$, $i = 1, \ldots, n$.

Доказательство вытекает из леммы и замены переменных $x = F_0^{-1}(u)$, $u = F_0(x)$ в экстремальной задаче (23).

$\blacksquare$

Теорема. Если верна гипотеза H_0, то при бесконечном увеличении объема выборки n

($n \to +\infty$) функция распределения нормированной статистики $\sqrt{n}D_n$ сходится к распределению Колмогорова:

$$\mathbf{P}_{H_0}\{\sqrt{n}D_n < z\} \to K(z) = \sum_{j=-\infty}^{+\infty} (-1)^j e^{-2j^2z^2},$$

$z \geq 0$.

Определение. Пусть ε – некоторое число: $0 < \varepsilon < 1$, а $K^{-1}(1-\varepsilon)$ – квантиль распределения Колмогорова уровня $1-\varepsilon$. Тогда тест

$$d = d(X) = \begin{cases} 0, & \sqrt{n}D_n \;<\; K^{-1}(1-\varepsilon); \\ 1, & \sqrt{n}D_n \;\geq\; K^{-1}(1-\varepsilon), \end{cases} \qquad (24)$$

называется критерием согласия Колмогорова.

Замечание. Распределение статистики D_n не зависит от гипотетической функции распределения $F_0(\cdot)$, в результате чего имеет место универсальность теста Колмогорова.

Следствие. При $n \to \infty$ асимптотический размер теста Колмогорова (24) совпадает с наперед заданным уровнем значимости $\varepsilon \in (0, 1)$.

Доказательство основывается на соотношениях (24), теореме и заключается в проверке следующего факта: $\alpha \to \varepsilon$, где α – вероятность ошибки I рода.

■

2. 6. Критерий отношения правдоподобия для проверки сложных гипотез

Рассмотрим теперь универсальный критерий проверки сложных гипотез – критерий отношения правдоподобия, суть которого состоит в следующем.

Пусть наблюдается случайная выборка

$$X \in \mathbb{R}^{nN}$$

объема n из некоторого N-мерного распределения вероятностей с плотностью

$$p(x;\theta),\ x \in \mathbb{R}^N,\ \theta \in \Theta \subseteq \mathbb{R}^m.$$

Определены две, в общем случае сложные, гипотезы:

$$\begin{array}{lcl} H_0 & : & \theta \in \Theta_0 \subset \Theta; \\ H_1 & : & \theta \in \Theta_1 = \Theta \backslash \Theta_0. \end{array}$$

Аналогично критерию Неймана – Пирсона введем в рассмотрение статистику отношения правдоподобия:

$$\tilde{\Lambda}_n = \tilde{\Lambda}_n(X;\Theta_0) = \frac{\sup\limits_{\theta \in \Theta_1} p(X;\theta)}{\sup\limits_{\theta \in \Theta_0} p(X;\theta)}, \qquad (25)$$

где

$$p(X;\theta) = \prod_{k=1}^{n} p(x_k;\theta) = L(\theta) -$$

функция правдоподобия. Часто рассматривают эквивалентные статистики:

$$\Lambda_n = \Lambda_n(X;\Theta_0) = \frac{\sup\limits_{\theta\in\Theta} p(X;\theta)}{\sup\limits_{\theta\in\Theta_0} p(X;\theta)} \geq 1; \qquad (26)$$

$$\lambda_n = \lambda_n(X;\Theta_0) = \frac{1}{\Lambda_n(X;\Theta_0)} =$$

$$= \frac{\sup\limits_{\theta\in\Theta_0} p(X;\theta)}{\sup\limits_{\theta\in\Theta} p(X;\theta)} \in [0,1].$$

Статистики (26) взаимно-однозначно функционально связаны между собой и статистикой отношения правдоподобия (25):

$$\Lambda_n = \max\{1, \tilde{\Lambda}_n\}, \quad \lambda_n = \min\left\{1, \frac{1}{\tilde{\Lambda}_n}\right\} \in [0,1].$$

Определение. Критерием отношения правдоподобия (КОП) в задаче проверки сложных гипотез H_0, H_1 по выборке X объема n называется следующий статистический критерий

$(C \in [0,1])$:

$$d = d(X) = \begin{cases} 0, & \lambda_n(X;\Theta_0) > C; \\ 1, & \lambda_n(X;\Theta_0) \le C, \end{cases}$$

или в эквивалентном виде

$$d = d(X) = \begin{cases} 0, & -2\ln\lambda_n(X;\Theta_0) < \delta; \\ 1, & -2\ln\lambda_n(X;\Theta_0) \ge \delta, \end{cases} \qquad (27)$$

где критическое (пороговое) значение $\delta \ge 0$ выбирается так, чтобы критерий имел наперед заданный уровень значимости ε:

$$\mathbf{P}_\theta\{d(X)=1\} = \int\limits_{\lambda_n(X;\Theta_0)\le C} p(X;\theta)dX =$$

$$= \mathbf{P}_\theta\{-2\ln\lambda_n \ge \delta\} \le \varepsilon, \ \forall\theta \in \Theta_0. \qquad (28)$$

Как было показано ранее теорема, известная как фундаментальная лемма Неймана – Пирсона), для простых гипотез КОП оптимален при любом n. Для сложных гипотез это, вообще говоря, не так. Тем не менее КОП широко применяется на практике. Оказывается, что он асимптотически оптимален при $n \to \infty$.

Рассмотрим вначале случай простой гипотезы H_0 и сложной альтернативы H_1.

Теорема. Пусть определены простая гипотеза

$$H_0 : \theta = \theta_0,$$

где $\theta_0 = (\theta_{0i}) \in \Theta$ – фиксированная точка m-мерной области Θ, и сложная альтернатива

$$H_1 : \theta \neq \theta_0,$$

тогда при выполнении условий регулярности, принятых в теории оценок максимального правдоподобия, асимптотический ($n \to \infty$) размер КОП (27), в котором $\delta = F^{-1}_{\chi^2_m}(1 - \varepsilon)$ – квантиль уровня $1 - \varepsilon$ от χ^2-распределения с m степенями свободы, совпадает с ε:

$$\mathbf{P}_{\theta_0}\{d(X) = 1\} = \mathbf{P}_{\theta_0}\{-2 \ln \lambda_n(X; \theta_0) \geq \delta\} \to \varepsilon. \tag{29}$$

Доказательство. В силу (28) для доказательства (29) достаточно показать сходимость к χ^2-распределению:

$$\mathcal{L}_{H_0}\{-2 \ln \lambda_n(X; \theta_0)\} \to \chi^2_m, \; n \to \infty,$$

где из (26) следует

$$-2 \ln \lambda_n(X; \theta_0) = 2 \left(l(\widehat{\theta}) - l(\theta_0) \right), \tag{30}$$

$l(\theta) = \ln p(X; \theta)$ – логарифмическая функция правдоподобия, а $\widehat{\theta} = \arg \max\limits_{\theta \in \Theta} p(X; \Theta)$ – ОМП, которая в условиях регулярности обладает свойством сильной состоятельности:

$$\widehat{\theta} \xrightarrow{\text{п.н.}} \theta^0, \ n \to \infty.$$

Следовательно, $\forall \varepsilon > 0 \ \exists \ \overline{n} = \overline{n}(\varepsilon)$, что $\forall n \geq \overline{n}$ с вероятностью единица: $|\widehat{\theta} - \theta_0| \leq \varepsilon$. В силу условий регулярности применим к $l(\theta_0)$ в (30) квадратичную формулу Тейлора с остаточным членом в форме Лагранжа (учитывая, что $\nabla_\theta l(\theta)|_{\theta=\widehat{\theta}} = \mathbf{0}_m$):

$$-2\ln\lambda_n(X;\theta_0) = n(\widehat{\theta}-\theta_0)^{\mathbf{T}}\left(-\frac{1}{n}\nabla^2_{\theta^*} l(\theta^*)\right)(\widehat{\theta}-\theta_0).$$

Здесь θ^* – «промежуточная точка»:

$$|\theta^* - \theta_0| < |\widehat{\theta} - \theta_0|,$$

следовательно,

$$\theta^* \xrightarrow{\text{п.н.}} \theta_0.$$

В силу условий регулярности выполняется усиленный закон больших чисел:

$$-\frac{1}{n}\nabla^2_\theta l(\theta^*) = \frac{1}{n}\sum_{k=1}^{n}\left(-\nabla^2_{\theta^*}\ln p(x_k;\theta^*)\right) \xrightarrow{\text{п.н.}}$$

$$\xrightarrow{\text{п.н.}} \mathbf{E}_{\theta_0}\left\{-\nabla^2_{\theta_0} = \ln p(x_k;\theta_0)\right\} =$$

$$= \mathcal{I}(\theta_0),$$

где $\mathcal{I} = \mathcal{I}(\theta_0)$ – информационная матрица Фишера для выборочного значения. Из свойств ОМП

$$\mathcal{L}_{H_0}\{\sqrt{n}(\widehat{\theta} - \theta_0)\} \to \mathcal{N}_m(\mathbf{0}_m, \mathbb{I}^{-1}(\theta_0)),\ n \to \infty,$$

и по свойствам линейного преобразования многомерного нормального распределения:

$$\mathcal{L}_{H_0}\{\sqrt{n}(\mathcal{I}^{\frac{1}{2}}(\theta_0))^{\mathbf{T}}(\widehat{\theta} - \theta_0)\} =$$

$$= \mathcal{L}\{\xi_n\} \to \mathcal{N}_m(\mathbf{0}_m, \mathcal{I}_m) =$$

$$= \mathcal{L}\{\xi\}.$$

В результате имеем

$$\mathcal{L}_{H_0}\{-2\ln\lambda_n(X;\theta_0)\} \to \mathcal{L}\{\xi^{\mathbf{T}}\xi\} = \chi^2_m,\ n \to \infty.$$

■

Замечание. Если H_0 является сложной гипотезой вида

$$H_0 : \theta \in \Theta_0,\ \Theta_0 = \{\theta : h_j(\theta) = 0,\ j = \overline{1, m-s}\},$$

где $1 \le s \le m-1$, $\{h_j(\cdot)\}_{j=0}^{m-s}$ – фиксированные непрерывные функции, то результат теоремы остается в силе, только

$$\delta = F^{-1}_{\chi^2_{m-s}}(1 - \varepsilon).$$

2. 7. Байесовское решающее правило

Предположим, что параметр θ – случайная величина, принимающая одно из двух возможных значений:

$$\theta \in \Theta = \{\theta_0, \theta_1\};$$

$$\mathbf{P}\{\theta = \theta_i\} = \Pi_i,\ 0 < \Pi_i < 1,\ i = 0, 1;$$

$$\Pi_0 + \Pi_1 = 1.$$

Наблюдается случайная выборка

$$X = (x_1^{\mathbf{T}}, \ldots, x_n^{\mathbf{T}})^{\mathbf{T}} \in \mathbb{R}^{nN}$$

объема n из некоторого распределения вероятностей с условной плотностью $p(x|\theta)$, $x \in \mathbb{R}^N$, $\theta \in \Theta$.

Обозначим:

$$p_i(X) = \prod_{j=1}^{n} p(x_j|\theta_i),\ i = 0, 1,\ –$$

условная плотность распределения выборки X при условии, что $\theta = \theta_i$.

Истинное значение θ неизвестно и определены две простые гипотезы:

$$H_i: \ \theta = \theta_i, \ i = 0, 1.$$

Задача заключается в построении теста для проверки H_0, H_1 по выборке X.

Эту задачу можно решить с помощью теста Неймана – Пирсона, но в таком случае игнорируется информация о $\{\Pi_i\}$. Поэтому рассмотрим другой подход.

Построим рандомизированное решающее правило:

$$d = d(X, \omega) \in D = \{0, 1\}, \ X \in \mathbb{R}^{nN}, \ \omega \in \Omega; \tag{31}$$

$$\mathbf{P}\{d(X, \omega) = 1 | X\} = \varphi(X),$$

$$\mathbf{P}\{d(X, \omega) = 0 | X\} = 1 - \varphi(X),$$

где $\varphi(X)$ – произвольная критическая функция, $0 \leq \varphi(X) \leq 1$.

Обозначим: $\nu = \nu(\omega) \in \{0, 1\}$ номер истинной гипотезы H_ν. $\nu = \nu(\omega) \in \{0, 1\}$ – случайная величина Бернулли. В силу случайности θ

$$\mathbf{P}(H_i) = \mathbf{P}\{\theta = \theta_i\} = \Pi_i, \ i = 0, 1,$$

поэтому Π_i принято называть *априорной вероятностью* i-й гипотезы.

Определение. Функцией потерь в рассматриваемой задаче проверки двух гипотез H_0, H_1 называется функция двух переменных:

$$w = w(i,j) \geq 0, \;\; i,j \in D = \{0,1\},$$

где $w(i,j)$ – величина потерь, которые несет статистик в ситуации, когда на самом деле $\nu = i$ (верна H_i), а принято решение $d = j$ в пользу гипотезы H_j.

Определение. Принято говорить, что имеет место $(0-1)$-функция потерь, если

$$w(i,j) = 1 - \delta_{ij} = \begin{cases} 0, & i = j; \\ 1, & i \neq j. \end{cases}$$

Функцию потерь удобно задавать в виде матрицы потерь: $W = (w_{ij})$, $w_{ij} = w(i,j)$. В случае $(0-1)$-матрицы потерь имеем

$$W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Определение. Функционалом риска называется математическое ожидание случайных потерь (средние потери)

$$r = r(\varphi(\cdot)) = \mathbf{E}\{w(\nu, d(X, \omega))\} \geq 0. \qquad (32)$$

Определение.(Байесовский принцип оптимальности.) Критическую функцию $\varphi(\cdot)$ в рандомизированном решающем правиле (31) надлежит выбирать таким образом, чтобы функционал риска (32) достигал минимального значения:

$$r(\varphi^*(\cdot)) = \inf_{\varphi(\cdot)} r(\varphi(\cdot)). \qquad (33)$$

При этом критическая функция $\varphi^*(\cdot)$, определяемая (33), называется байесовской критической функцией, а соответствующее решающее правило $d^*(X, \omega)$, определяемое (31), – байесовским решающим правилом (БРП).

Теорема. Пусть в сформулированной выше задаче проверки простых гипотез H_0, H_1 функция потерь имеет следующий вид:

$$w(i, j) = \begin{cases} 0, & i = j; \\ w_0, & i = 0, \ j = 1; \\ w_1, & i = 1, \ j = 0, \end{cases} \qquad (34)$$

где $w_0 > 0$, $w_1 > 0$ – некоторые заданные величины.

Тогда байесовская критическая функция задается соотношением ($X \in \mathbb{R}^{nN}$):

$$\varphi^*(X) = \begin{cases} 0, & L(X) < C^*; \\ \varkappa^*, & L(X) = C^*; \\ 1, & L(X) > C^*, \end{cases} \qquad (35)$$

где

$$L(X) = \frac{p_1(X)}{p_0(X)} \geq 0, \quad C^* = \frac{\Pi_0 w_0}{\Pi_1 w_1} \geq 0, \quad \varkappa^* \in [0, 1$$

Д о к а з а т е л ь с т в о. Прежде всего заметим, что в силу произвола в выборе $\varkappa^*$ байесовская критическая функция неоднозначна. Сформулируем экстремальную задачу (33) в явном виде и решим ее. Из (32) и (34) имеем

$$r = r(\varphi(\cdot)) =$$

$$= 0 \cdot \mathbf{P}\{d(X, \omega) = \nu\} + w_0 \cdot \mathbf{P}\{\nu = 0, d = 1\} +$$

$$+ w_1 \cdot \mathbf{P}\{\nu = 1, d = 0\} =$$

$$= w_0 \cdot \mathbf{P}\{\nu = 0\} \cdot \mathbf{P}\{d = 1 | H_0\} +$$

$$+w_1 \cdot \mathbf{P}\{H_1\} \cdot \mathbf{P}\{d = 0 | H_1\} =$$

$$= w_0 \cdot \Pi_0 \cdot \alpha + w_1 \cdot \Pi_1 \cdot \beta.$$

Используя интегральные представления вероятностей ошибок α, β, приходим к задаче:

$$r = r(\varphi(\cdot)) =$$

$$= \Pi_0 w_0 \int_{\mathbb{R}^{nN}} \varphi(X) p_0(X) dX +$$

$$+\Pi_1 w_1 \int_{\mathbb{R}^{nN}} (1 - \varphi(X)) p_1(X) dX =$$

$$= \Pi_1 w_1 - \int_{\mathbb{R}^{nN}} \varphi(X) G(X) dX \to \min_{\varphi(\cdot)}, \qquad (37)$$

где

$$G(X) = \Pi_1 w_1 p_1(X) - \Pi_0 w_0 p_0(X).$$

Из (37) получаем эквивалентную экстремальную задачу:

$$\int_{\mathbb{R}^{nN}} \varphi(X) G(X) dX \to \max_{\varphi(\cdot)},$$

являющуюся линейной задачей вариационного исчисления, причем ограничения носят точечный, локальный характер: $0 \le \varphi(X) \le 1$,

$X \in \mathbb{R}^{nN}$, поэтому решение очевидно:

$$\varphi^*(X) = \begin{cases} 0, & G(X) < 0; \\ \varkappa^*, & G(X) = 0; \\ 1, & G(X) > 0. \end{cases}$$

С учетом обозначений (36) последнее соотношение эквивалентно (35). Найдем минимум риска, который при этом достигается:

$$r^* = r(\varphi^*) = \Pi_1 w_1 - \int_{G(X)>0} G(X)dX.$$

■

Следствие. Среди байесовских решающих правил (35) существует нерандомизированное решающее правило:

$$d = d^*(X) = \begin{cases} 0, & L(X) < C^*; \\ 1, & L(X) \geq C^*. \end{cases}$$

Д о к а з а т е л ь с т в о. Для доказательства достаточно выбрать $\varkappa^* = 1$.

■

Следствие. Если имеет место (0–1)-функция потерь, то есть $w_0 = w_1 = 1$, и гипотезы H_0, H_1 равновероятны ($\Pi_0 = \Pi_1 = \frac{1}{2}$), то БРП имеет вид

$$d = d^*(X) = \begin{cases} 0, & p_0(X) > p_1(X); \\ 1, & p_1(X) \geq p_0(X). \end{cases} \qquad (38)$$

Д о к а з а т е л ь с т в оИз формулы (36) имеем: $C^* = 1, \varkappa^* = 1$.

■

Заметим в заключение, что решающее правило (38) часто называют *тестом максимального правдоподобия*.