Тема 7. РАЗЛИЧЕНИЕ ДВУХ ПРОСТЫХ ГИПОТЕЗ

А) Число наблюдений фиксировано: процедура Неймана-Пирсона

Пусть $x \equiv (x_1, x_2, ..., x_n)$ — результаты *n* наблюдений (не обязательно независимых), являющиеся конкретными значениями многомерной случайной величины $\xi \equiv (\xi_1, \xi_2, ..., \xi_n)$. Имеется два предположения (две гипотезы) относительно закона распределения наблюдений:

$$H_0 : p_0(x) \quad \text{и} \quad H_1 : p_1(x).$$

По наблюдениям x требуется принять одно из двух решений: «верна гипотеза H_0» или «верна гипотеза H_1». Сказанное означает, что нужно построить решающую функцию (процедуру различения) $\delta(x)$, принимающую два значения 0 и 1:

$$\delta(x) = \begin{cases} 1 \text{ (верна гипотеза } H_1\text{), если } x \in \Gamma_1, \\ 0 \text{ (верна гипотеза } H_0\text{), если } x \in \Gamma_0. \end{cases}$$

Область, в которой $\delta(x) = 1$, обозначена Γ_1, а область, где $\delta(x) = 0$, — Γ_0. Таким образом, решающая функция $\delta(x;\Gamma)$ определяется разбиением $\Gamma = (\Gamma_0, \Gamma_1)$ пространства *X* значений x: $\Gamma_0 \cup \Gamma_1 = X$, $\Gamma_0 \cap \Gamma_1 = \varnothing$.

При любой решающей функции $\delta(x)$ возможны ошибки двух типов:

ошибка **первого** рода: «принять H_1 при истинности H_0» и

ошибка **второго** рода: «принять H_0 при истинности H_1»

и соответствующие вероятности

$$\alpha = \mathbf{P}\{\text{принята } H_1 \mid H_0\} = \int_{\Gamma_1} p_0(x)dx$$ —вероятность ошибки **1**-го рода

и

$$\beta = \mathbf{P}\{\text{принята } H_0 \mid H_1\} = \int_{\Gamma_0} p_1(x)dx$$ — вероятность ошибки **2**-го рода.

Характеристикой решающей функции $\delta(x)$ являются эти две вероятности.

Невозможно одновременно уменьшить обе вероятности. Оптимальным (по Нейману-Пирсону) решающим правилом называется такое, для которого при **заданном уровне** α_0 вероятности ошибки **первого** рода,

$$\alpha = \mathbf{P}\{\text{принять } H_1 \mid H_0\} \le \alpha_0,$$

вероятность ошибки **второго** рода **минимальна**:

$$\beta = \mathbf{P}\{\text{принята } H_0 \mid H_1\} \to \mathbf{min}.$$

Оказывается (фундаментальная лемма **Неймана-Пирсона**), оптимальное правило $\delta(x)$ имеет область Γ_1 принятия H_1 следующего вида:

$$\Gamma_1 = \left\{ x: \frac{p_1(x)}{p_0(x)} \geq h \right\}, \qquad \textbf{(7.1)}$$

где h определяется из условия:

$$\alpha(h) = \alpha_0 \qquad \textbf{(7.2)}$$

В задачах на **построение решающей функции** обычно требуется найти:

1) вид решающего правила $\delta(x)$;

2) порог h;

3) вероятность $\beta = \mathbf{P}\{\text{принята } H_0 \mid H_1\}$;

4) если число n наблюдений не задано, то определить количество наблюдений n так, чтобы обеспечить заданный уровень для вероятности ошибки **второго рода**: $\beta(n) = \beta_0$.

Пример 1 (о параметре показательного распределения — задача, актуальная, например, в теории надежности**).**

Пусть $x \equiv (x_1, x_2, ..., x_n)$ — выборка (n независимых одинаково распределённых СВ $(\xi_1, \xi_2, ..., \xi_n)$, распределенных по показательному закону:

$$\xi_i \sim a e^{-a x_i}, \; x_i \geq 0, \; i = \overline{1, n}.$$

Имеется **две гипотезы** относительно параметра a:

$$H_0: a = a_0 = 1,$$
$$H_1: a = a_1 = 1{,}2.$$

Пусть $n = 100$ и задано $\alpha_0 = 0{,}05$. Требуется найти $\delta(x)$, $\Gamma = (\Gamma_0, \Gamma_1)$ и β.

Выписываем два предположения о законе распределения выборки:

$$p_1(x) = a_1^n e^{-a_1 \sum_{i=1}^{n} x_i} \quad \text{и}$$

$$p_0(x) = a_0^n e^{-a_0 \sum_{i=1}^{n} x_i}.$$

1) Определяем решающее правило:

$$\Gamma_1 = \left\{ x: \frac{p_1(x)}{p_0(x)} \geq h \right\} = \left\{ x: \left(\frac{a_1}{a_0}\right)^n e^{-(a_1 - a_0)\sum_{i=1}^{n} x_i} \geq h \right\} =$$

приводим к удобному виду (выделяем решающую статистику):

$$\Gamma_1 = \left\{ x: n \ln\frac{a_1}{a_0} - (a_1 - a_0)\sum_{i=1}^{n} x_i \geq \ln h \equiv h_1 \right\} = \left\{ x: \sum_{i=1}^{n} x_i \leq \frac{h_1 - n\ln\frac{a_1}{a_0}}{a_0 - a_1} \equiv h_2 \right\}.$$

Заметим, что при делении на отрицательное число $(a_0 - a_1)$ знак неравенства изменился. Итак, решающее правило $\delta(x)$ имеет вид:

если $\sum_{i=1}^{n} x_i \le h_2$, то принимается гипотеза H_1;

если $\sum_{i=1}^{n} x_i > h_2$, то принимается гипотеза H_0.

2) Находим порог h из условия $\alpha(h_2) = \alpha_0$:

$$\alpha(h_2) = \mathbf{P}\{\text{принята } H_1 \mid H_0\} = \mathbf{P}\left\{\sum_{i=1}^{n} x_i \le h_2 \mid H_0\right\} \approx$$

учтем, что если верна гипотеза H_0, то $\sum_{i=1}^{n} x_i \sim N\left(n \cdot \frac{1}{a_0}, n \cdot \frac{1}{a_0^2}\right)$.

$$\approx \mathbf{\Phi}\left(\frac{h_2 - n/a_0}{\sqrt{n}/a_0}\right) = \alpha_0 = 0{,}05. \quad \textbf{(7.3)}$$

откуда получаем значение аргумента

$$\frac{h_2 - n/a_0}{\sqrt{n}/a_0} = \mathbf{\Phi}^{-1}(\alpha_0) \approx -1{,}65.$$

$$h_2 \approx \frac{n}{a_0} - 1{,}65\frac{\sqrt{n}}{a_0} = \frac{100}{1} - 1{,}65 \cdot \frac{10}{1} = 83{,}5.$$

Итак, решающая процедура $\delta(x)$ такова:

если $\sum_{i=1}^{n} x_i \le 83{,}5$, то принимается гипотеза H_1; иначе — H_0.

3) Определим вероятность β ошибки второго рода:

$$\beta = \mathbf{P}\{\text{принята } H_0 \mid H_1\} = \mathbf{P}\left\{\sum_{i=1}^{n} x_i > 83{,}5 \mid H_1\right\} \approx$$

$$\approx 1 - \mathbf{\Phi}\left(\frac{83{,}5 - n/a_1}{\sqrt{n}/a_1}\right) = 1 - \mathbf{\Phi}\ \frac{83{,}5 - 100/1{,}2}{10/1{,}2} \approx 0{,}5.$$

Здесь учтено, что если верна гипотеза H_1, то сумма $\sum_{i=1}^{n} x_i$ распределена нормально: $\sum_{i=1}^{n} x_i \sim N\left(\frac{n}{a_1}, \frac{n}{a_1^2}\right)$.

Мы получили $\beta = 0{,}5$. Нас может не устраивать такая вероятность ошибки. Нужно увеличить число наблюдений.

4) Если мы имеем возможность изменять n, то можно уменьшить β, задав эту вероятность на нужном уровне, например, $\beta_0 = 0{,}05$.

Тогда получаем два условия:

$$\alpha = \mathbf{P}\{\text{принята } H_1 \mid H_0\} = \mathbf{\Phi}\left(\frac{h_2 - n/a_0}{\sqrt{n}/a_0}\right) = \alpha_0; \quad \textbf{(7.4)}$$

$$\beta = \mathbf{P}\{\text{принята } H_0 \mid H_1\} = 1 - \mathbf{\Phi}\left(\frac{h_2 - n/a_1}{\sqrt{n}/a_1}\right) = \beta_0. \quad \textbf{(7.5)}$$

Имеем систему из двух уравнений с двумя неизвестными n и h_2. Поскольку $\mathbf{\Phi}^{-1}(\alpha)=\mathbf{\Phi}^{-1}(0{,}05)\approx-1{,}65$ и $\mathbf{\Phi}^{-1}(1-\beta)=\mathbf{\Phi}^{-1}(9{,}95)\approx 1{,}65$, то можно выразить h_2 из (7.4) и (7.5):

$$\begin{cases} h_2=\dfrac{n}{a_0}-1{,}65\dfrac{\sqrt{n}}{a_0}, \\ h_2=\dfrac{n}{a_1}+1{,}65\dfrac{\sqrt{n}}{a_1}. \end{cases}$$

Приравнивая правые части, получаем:

$$n\left(\frac{1}{a_0}-\frac{1}{a_1}\right)=1{,}65\sqrt{n}\left(\frac{1}{a_0}+\frac{1}{a_1}\right).$$

Решая квадратное уравнение, получаем $n\approx 225$, а затем находим

$$h_2=\left(\frac{1}{a_0}\cdot n-1{,}65\sqrt{n}\right)=225-1{,}65\cdot 15\approx 200.$$

Пример 2. Пётр утверждает, что умеет бросать монету так, что вероятность выпадения герба равна $0{,}6$. Это гипотеза

$$H_1:\ P(\Gamma)=p_1=0{,}6.$$

Павел утверждает, что это невозможно, и что вероятность выпадения герба равна $0{,}5$. Это гипотеза

$$H_0:\ P(\Gamma)=p_0=0{,}5.$$

Как установить, кто из них прав?

Решение.

Во-первых, Пётр должен бросить монету некоторое количество $n=\ ?$ раз.

Во-вторых, нужно построить процедуру проверки двух гипотез H_1 и H_0 по n наблюдениям, причём число n выбрать так, чтобы вероятности ошибок 1-го и 2-го рода α и β были малыми и, если **по справедливости**, то одинаковыми, например, $\alpha=\beta=0{,}025$ (по согласию Пётра и Павла).

Будущие результаты наблюдений $\xi\equiv(\xi_1,\xi_2,\ldots,\xi_n)$ — это случайные величины, которые примут некоторые конкретные значения $x\equiv(x_1,x_2,\ \ldots,x_n)$, $x_i\in\{0;1\}$, $i=\overline{1,n}$, с вероятностями

$$\mathbf{P}(x_i)=p^{x_i}(1-p)^{1-x_i}.$$

1) Вид решающего правила.

Запишем два варианта закона распределения наблюдений:

если верна гипотеза H_0, то

$$\mathbf{P}_0(x)=\prod_{i=1}^{n}\mathbf{P}\{x_i\,|H_0\}=\prod_{i=1}^{n}p_0^{x_i}(1-p_0)^{1-x_i}=\left(\frac{p_0}{1-p_0}\right)^{\sum_{i=1}^{n}x_i}(1-p_0)^n;$$

если верна гипотеза H_1, то

$$\mathbf{P}_1(x) = \frac{p_1}{1-p_1}^{\sum\limits_{i=1}^{n} x_i} (1-p_1)^n,$$

где $\sum\limits_{i=1}^{n} x_i$ — число выпадений герба.

Область принятия гипотезы H_1:

$$\Gamma_1 = \quad x:\ \frac{\mathbf{P}_1(x)}{\mathbf{P}_0(x)} \geq h \ ,$$

где **нужно выбрать** h **и** n **так, что** $\alpha\,(n\ h, =\ )\quad ($ **и** $\beta\,(n,h) = 0{,}025$. **Конкретизируем** Γ_1:

$$\Gamma_1 = \left\{x:\ \frac{p_1}{1-p_1}^{\sum\limits_{i=1}^{n} x_i} (1-p_1)^n \Big/ \frac{p_0}{1-p_0}^{\sum\limits_{i=1}^{n} x_i} (1-p_0)^n \geq h\right\} =$$

$$= \left\{x:\ \frac{p_1}{p_0}\cdot\frac{1-p_0}{1-p_1}^{\sum\limits_{i=1}^{n} x_i} \ \frac{1-p_1}{1-p_0}^{n} \geq h\right\} = \left\{x:\ \sum_{i=1}^{n} x_i \geq h_1(n)\right\}$$

где в неизвестное $h_1(n)$ вошли все величины, кроме наблюдений (после логарифмирования и переноса).

Имеем два условия:

$$\begin{cases} \alpha\,(n,h) = \mathbf{P}\{\text{принята } H_1 \,|\, H_0\} = 0{,}025; \\ \beta\,(n,h) = \mathbf{P}\{\text{принята } H_0 \,|\, H_1\} = 0{,}025. \end{cases}$$

$$\alpha = \mathrm{P}\{\text{принять } H_1 \,|\, H_0\} = \mathrm{P}\left\{S_n = \sum_{i=1}^{n} x_i \geq h_1 \,|\, H_0\right\} = 1 - \mathrm{P}\ S_n < h_1 \,|\, H_0 \ =$$

Здесь $S_n = \sum\limits_{i=1}^{n} x_i$ — **решающая статистика** (количество успехов).

Если верна гипотеза H_0, то $S_n = \sum\limits_{i=1}^{n} x_i \sim Bi(n, p_0)$ имеет приближенно нормальное распределение $N\ np_0,\ np_0(1-p_0)$.

$$\approx 1 - \mathbf{\Phi}\left(\frac{h_1 - np_0}{\sqrt{np_0(1-p_0)}}\right) = 0{,}025 \quad \Rightarrow \quad \frac{h_1 - np_0}{\sqrt{np_0(1-p_0)}} = 2.$$

Если же верна гипотеза H_1, то $S_n = \sum\limits_{i=1}^{n} x_i \sim Bi(n, p_1)$ имеет приближенно нормальное распределение $N\ np_1,\ np_1(1-p_1)$.

$$\beta = \mathbf{P}\{\text{принята } H_0 \,|\, H_1\} = \mathbf{P}\ S_n \leq h_1 \,|\, H_1 \ \approx \mathbf{\Phi}\left(\frac{h_1 - np_1}{\sqrt{np_1(1-p_1)}}\right) = 0{,}025 \quad \Rightarrow$$

$$\Rightarrow \quad \frac{h_1 - np_1}{\sqrt{np_1(1-p_1)}} = -2.$$

Итак получили систему из двух уравнений:

$$\begin{cases} h_1 - np_0 = 2\sqrt{np_0(1-p_0)}; \\ h_1 - np_1 = -2\sqrt{np_1(1-p_1)}. \end{cases}$$

Вычтем одно уравнение из другого:

$$n(p_1 - p_0) = 2\sqrt{n}\left(\sqrt{p_0(1-p_0)} + \sqrt{p_1(1-p_1)}\right) \quad \Rightarrow$$

$$n = \left(2\cdot\frac{\sqrt{p_0(1-p_0)} + \sqrt{p_1(1-p_1)}}{p_1 - p_0}\right)^2 = \left(2\cdot\frac{0{,}5+0{,}5}{0{,}1}\right)^2 = 400.$$

Порог: $h_1 = np_0 + 2\sqrt{np_0(1-p_0)} = 200 + 2\cdot\frac{1}{2}\sqrt{400} = 220.$

Итак: $\Gamma_0 = \left\{x\colon \sum_{i=1}^{400} x_i < 220\right\}$. Таким обр азом, решающая процедура имеет вид:

если $S_n = \sum_{i=1}^{400} x_i < 220$, то принимается гипотеза $H_0 = \{\text{Пётр, ты неправ}\}$.

если же $S_n = \sum_{i=1}^{400} x_i \geq 220$, то принимается гипотеза H_1.

Б) Число наблюдений не фиксировано: критерий Вальда (ПКОВ — последовательтый критерий отношения вероятностей)

Во многих практических задачах бывает важно максимально быстро принять решение (задачи военные, испытания в теории надежности, оценка эффективности экономических и других сложных систем). Возникает следующая постановка задачи различения двух гипотез.

Наблюдатель последовательно получает наблюдения $x_1, x_2, ..., x_n, ...$, независимые и одинаково распределенные. Имеется две гипотезы

$$H_0\colon x_i \sim p_0(x_i) \quad \text{и} \quad H_1\colon x_i \sim p_1(x_i).$$

Требуется построить такое решающее правило δ^*, для которого вероятности α и β ошибок находятся на заданном уровне, и при этом среднее число наблюдений минимально. Таким правилом является ПКОВ — последовательтый критерий отношения вероятностей, определяемый двумя порогами A и B, **верхним и нижним**: $0 < B < 1 < A < \infty$ и состоящий в следующем.

На каждом шаге, после получения n-го наблюдения, вычисляется отношение правдоподобия L_n двух гипотез,

$$L_n \equiv \frac{p_1(x_1)}{p_0(x_1)}\cdot\frac{p_1(x_2)}{p_0(x_2)}\cdot\ldots\cdot\frac{p_1(x_n)}{p_0(x_n)} \qquad \textbf{(7.6)}$$

и оно сравнивается с порогами A и B:

1) если $L_n \geq A$, то принимается гипотеза H_1 и наблюдения заканчиваются,
2) если $L_n \leq B$, то принимается гипотеза H_0 и наблюдения заканчиваются,
3) если $B < L_n < A$, то наблюдения продолжаются, делается следующее $(n+1)$-е наблюдение.

Пороги определяются так, чтобы вероятности ошибок $\alpha(A, B)$ и $\beta(A, B)$ были на заданном уровне α и β. Справедлива теорема Вальда и Вольфовица, доказывающая, что такое правило имеет минимальные (среди всех возможных процедур) средние числа наблюдений $\bar{\nu}_0 = M(\nu | H_0)$ и $\bar{\nu}_1 = \mathbf{M}(\nu | H_1)$, где ν — число наблюдений — случайная величина.

Имеются весьма **точные** и **простые** приближенные формулы для порогов:

$$A \approx \frac{1-\beta}{\alpha}, \qquad B \approx \frac{\beta}{(1-\alpha)}, \tag{7.7}$$

а также для средних чисел наблюдений:

$$\mathbf{M}(\nu | H_1) \approx \frac{(1-\beta)\ln A + \beta \ln B}{\mathbf{M}(\zeta | H_1)}, \qquad \mathbf{M}(\nu | H_0) \approx \frac{\alpha \ln A + (1-\alpha)\ln B}{\mathbf{M}(\zeta | H_0)}, \tag{7.8}$$

$\mathbf{M}(\zeta | H_k) = \mathbf{M} \ln \frac{p_1(x_i)}{p_0(x_i)} | H_k$ — информация по Кульбаку, $k = 0, 1$.

Здесь мы использовали обозначение: $\zeta = \ln \frac{p_1(x_i)}{p_0(x_i)}$.

При решении задач обычно требуется найти:

1) вид решающего правила;
2) пороги;
3) средние числа наблюдений (при H_0 и H_1).

Пример 3. Различение гипотез о среднем нормальной совокупности.

На выходе приёмника РЛС в дискретные моменты времени $1, 2, \ldots, n, \ldots$ измеряется напряжение сигнала S с аддитивной нормальной и независимой по времени помехой:

$$\xi_i = S + \varepsilon_i, \ \varepsilon_i \sim N(0, \sigma^2), \ i \in \boldsymbol{N}.$$

Рассмотрим две гипотезы:

H_0: $S = 0$ (нет цели) $\Leftrightarrow$ $\xi_i \sim N(0, \sigma^2)$, $\sigma^2 = 1$, $i \in \boldsymbol{N}$.

H_1: $S = a \neq 0$ (цель есть) $\Leftrightarrow$ $\xi_i \sim N(a, \sigma^2)$, $a = 0{,}1$, $\sigma^2 = 1$, $i \in \boldsymbol{N}$.

Построить последовательную процедуру решения задачи о наличии цели; принять:

$\alpha = \mathrm{P}\{\text{принять } H_1 | H_0\} = 10^{-4}$ (вероятность ложной тревоги);
$\beta = \mathrm{P}\{\text{принять } H_0 | H_1\} = 10^{-2}$ (вероятность пропуска).

1) Решающее правило.

На n-м шаге логарифм отношения правдоподобия:

$\mathbf{M}(\zeta|H_k) = \mathbf{M} \ln\frac{p_1(x_i)}{p_0(x_i)}|H_k$

$$\ln L_n = \sum_{i=1}^{n} \ln\frac{p_1(x_i)}{p_0(x_i)} = \sum_{i=1}^{n} \ln\frac{\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x_i-a)^2}{2\sigma^2}\right\}}{\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{x_i^2}{2\sigma^2}\right\}} =$$

$$= \sum_{i=1}^{n} -\frac{1}{2\sigma^2}[x_i^2 - 2ax_i + a^2 - x_i^2] = \frac{a}{\sigma^2}\left(\sum_{i=1}^{n} x_i - \frac{na}{2}\right).$$

В терминах суммы S_n решающая процедура δ^* на n-м шаге примет вид: здесь переход написать к сумме

если $S_n = \sum\limits_{i=1}^{n} x_i \geq \frac{\sigma^2}{a}\ln A + \frac{na}{2} \equiv a(n)$, то принимается гипотеза H_1;

если $S_n = \sum\limits_{i=1}^{n} x_i \leq \frac{\sigma^2}{a}\ln B + \frac{na}{2} \equiv b(n)$, то принимается гипотеза H_0;

если $b(n) < S_n = \sum\limits_{i=1}^{n} x_i < a(n)$, то следует продолжить наблюдения.

2) Пороги.

$$A \approx A' = \frac{1-\beta}{\alpha} = \frac{1-10^{-2}}{10^{-4}} \approx 10^4, \quad \ln A \approx 4\cdot 2{,}3 = 9{,}2;$$

$$B \approx B' = \frac{\beta}{1-\alpha} \approx 10^{-2}, \quad \ln B \approx -2\cdot 2{,}3 = -4{,}6.$$

3) Среднее число наблюдений.

Введём обозначение: $\zeta_i = \frac{a}{\sigma^2}x_i - \frac{a}{2}$ (это одно слагаемое в формуле для $\ln L_n$).

$$\mathbf{M}(\zeta_i|H_0) = -\frac{a^2}{2\sigma^2}, \quad \mathbf{M}(\zeta_i|H_1) = \frac{a^2}{2\sigma^2};$$

$$\mathbf{M}(\nu|H_1) = \frac{(1-\beta)\ln A + \beta\ln B}{\mathbf{M}(\zeta|H_1)} \approx \frac{\ln A}{a^2/2\sigma^2} = \frac{2\cdot 1\cdot 9{,}2}{(0{,}1)^2} = 1840;$$

$$\mathbf{M}(\nu|H_0) = \frac{\alpha\ln A + (1-\alpha)\ln B}{\mathbf{M}(\zeta|H_0)} \approx \frac{\ln B}{-a^2/2\sigma^2} = \frac{-2\cdot 1\cdot 4{,}6}{-(0{,}1)^2} = 920.$$

Дополнительно построим решающее правило с фиксированным объёмом выборки, имеющее те же α и β. Определим объём выборки и сравним его с числом наблюдений для ПКОВ.

Процедура **Неймана-Пирсона** определяет область принятия гипотезы H_1:

$$\Gamma_1 = x \equiv (x_1, x_2, \ldots, x_n): \frac{p_1(x)}{p_0(x)} \geq h, \quad \alpha(h) = \alpha = 10^{-4};$$

$$\frac{p_1(x)}{p_0(x)} = -\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n}(x_i - a)^2 - \sum_{i=1}^{n} x_i^2\right] = \frac{1}{2\sigma^2}\left[2a\sum_{i=1}^{n} x_i - na^2\right];$$

$$\Gamma_1 = \left\{x:\ \sum_{i=1}^{n} x_i \geq \frac{\sigma^2}{a}\ln h + n\frac{a}{2} \equiv h_1\right\}.$$

Имеем **два условия для вероятностей ошибок**:

а) если **верна гипотеза** H_0, то $S_n = \sum_{i=1}^{n} x_i \sim N(0, n\sigma^2)$, поэтому

$$\alpha = \mathbf{P}\{\text{принята } H_1 \mid H_0\} \approx 1 - \mathbf{\Phi}\frac{h_1 - 0}{\sigma\sqrt{n}} = 10^{-4};$$

б) если **верна гипотеза** H_1, то $S_n = \sum_{i=1}^{n} x_i \sim N(na, n\sigma^2)$, поэтому

$$\beta = \mathbf{P}\{\text{принята } H_0 \mid H_1\} \approx 1 - \mathbf{\Phi}\frac{h_1 - na}{\sigma\sqrt{n}} = 10^{-2}.$$

Из каждого условия определяем порог:

$$h_1 = Q(1 - 10^{-4})\cdot\sigma\sqrt{n} = 3{,}7\sigma\sqrt{n};,$$

$$h_1 = na + Q(10^{-2})\cdot\sigma\sqrt{n} = na - 2{,}35\sigma\sqrt{n}.$$

Приравнивая правые части, получаем:

$$na = \sigma\sqrt{n}(3{,}7 + 2{,}35) \quad\Rightarrow\quad a\sqrt{n} = 6{,}05\sigma \quad\Rightarrow\quad n \approx \frac{\sigma^2}{a^2}6^2 = \frac{1\cdot 36}{10^{-2}} = 3600$$

против 920 и 1840 для ПКОВ.

Пример 4. Различение гипотез о вероятности случайного события**.** Пётр утверждает, что может бросать монету с вероятностью $p_1 = 0{,}6$:

$$H_1:\ P(\Gamma) = p_1 = 0{,}6.$$

Павел утверждает, что это невозможно, и что

$$H_0:\ P(\Gamma) = p_0 = 0{,}5.$$

Построить последовательное решающее правило с $\alpha = \beta = 0{,}025$ и определить среднее число наблюдений.

Решение. Пусть $x \equiv (x_1, x_2, ..., x_n)$ — результаты *n* наблюдений, где $x_i \in \{0; 1\},\ i = \overline{1, n}$.

А) Если **верна гипотеза** H_1, то

$$x_i \sim \mathbf{P}_1(x_i) = p_1^{x_i}(1 - p_1)^{1-x_i} = \left(\frac{p_1}{1-p_1}\right)^{x_i}(1 - p_1),\ i \in \boldsymbol{N}.$$

В) Если **верна гипотеза** H_0, то

$$x_i \sim \mathbf{P}_0(x_i) = p_0^{x_i}(1 - p_0)^{1-x_i} = \left(\frac{p_0}{1-p_0}\right)^{x_i}(1 - p_0),\ i \in \boldsymbol{N}.$$

1) Решающее правило.

Условие **продолжения наблюдений** имеет вид: $\ln B < \ln L_n < \ln A$.

$$\ln L_n = \sum_{i=1}^{n} \ln\frac{\mathbf{P}_1(x)}{\mathbf{P}_0(x)} = \sum_{i=1}^{n}\left[x_i \ln\frac{p_1}{p_0}\frac{(1-p_0)}{(1-p_1)} + \ln\frac{(1-p_1)}{(1-p_0)}\right] = C_1\sum_{i=1}^{n} x_i + nC_2,$$

где $C_1 = \ln\frac{p_1}{p_0}\frac{(1-p_0)}{(1-p_1)} = \ln\frac{0,6}{0,4} = \ln 1,5 \approx 0,4, \quad C_2 \equiv \ln\frac{(1-p_1)}{(1-p_0)} = \ln\frac{0,4}{0,5} \approx -0,22.$

Условие продолжения наблюдений принимает вид:

$$b(n) \equiv \frac{1}{C_1}(\ln B - nC_2) < \sum_{i=1}^{n} x_i < \frac{1}{C_1}(\ln A - nC_2) \equiv a(n).$$

2) Пороги.

$$A \approx A' = \frac{1-\beta}{\alpha} = \frac{1-10^{-2}}{10^{-4}} \approx 40; \qquad B \approx B' = \frac{\beta}{1-\alpha} \approx \frac{1}{40};$$

$$\ln A \approx \ln 4 + \ln 10 = 1,38 + 2,3 = 3,68 \approx 3,7;$$
$$\ln B \approx -3,7.$$

$$a(n) = \ln 40 + n\ln\frac{0,5}{0,4} \cdot \ln^{-1}\frac{0,6}{0,4} \approx (3,68 + 0,22n)\cdot(0,4)^{-1} = 9,2 + 0,55n;$$

$$b(n) = -9,2 + 0,55n.$$

Итак, если для числа $S(n)$ выпадений герба после n бросаний, верны оба неравенства:

$$b(n) = -9,2 + 0,55n < S(n) = \sum_{i=1}^{n} x_i < 9,2 + 0,55n = a(n),$$

то наблюдения продолжаются до момента нарушения хотя бы одного из неравенств (говорят: до момента выхода на **границу**).

Нетрудно определить минимально возможное число наблюдений:

А) если подряд наблюдаются только нули (принятие гипотезы H_0), то

$$S(n) \equiv \sum_{i=1}^{n} x_i = 0, \ \text{и} \ -9,2 + 0,55n = 0 \quad \Leftrightarrow \quad n \approx 17;$$

В) если только единицы (принятие гипотезы H_1), то

$$S(n) \equiv \sum_{i=1}^{n} x_i = n, \ \text{и} \ 9,2 + 0,55n = n \quad \Leftrightarrow \quad n \approx 20.$$

3) Среднее число наблюдений.

$$\mathbf{M}(\nu\,|H_1) \approx \frac{(1-\beta)\ln A + \beta\ln B}{\mathbf{M}(\zeta\,|H_1)} \approx \frac{\ln A}{\mathbf{M}(\zeta\,|H_1)} \approx \frac{3,7}{0,02} = 185,$$

где знаменатель определен следующим образом:

$$\mathbf{M}(\zeta\,|H_1) = \mathbf{M}\left[\ln\frac{\mathbf{P}_1(\xi)}{\mathbf{P}_0(\xi)}\,|H_1\right] = \mathbf{M}\left[C_1\xi + C_2\,|H_1\right] = p_1C_1 + C_2 =$$

$$= 0,6\ln\frac{0,6}{0,4} - \ln\frac{0,5}{0,4} \approx 0,6\cdot 0,4 - 0,22 = 0,24 - 0,22 = 0,02.$$

$$\mathbf{M}(\nu\,|H_0) \approx \frac{\alpha\ln A + (1-\alpha)\ln B}{\mathbf{M}(\zeta\,|H_0)} \approx \frac{\ln B}{\mathbf{M}(\zeta\,|H_0)} \approx \frac{-3,7}{-0,02} = 185,$$

где знаменатель определен следующим образом:

$$\mathbf{M}(\zeta | H_0) = p_0 C_1 + C_2 = 0{,}5 \cdot 0{,}4 - 0{,}22 = -0{,}02.$$

ДОМАШНЕЕ ЗАДАНИЕ

7.1. Случайная величина имеет распределение **Пуассона** с параметром λ. Используя выборку объёма n, определить наилучшую критическую область для проверки гипотезы $H_0: \lambda = \lambda_0$ против альтернативной гипотезы $H_1: \lambda = \lambda_1$. Рассмотреть случаи: а) $\lambda_0 < \lambda_1$; б) $\lambda_0 > \lambda_1$.

Ответ: а) $\sum_{i=1}^{n} x_i > h$; б) $\sum_{i=1}^{n} x_i < h$.

7.2. Случайная величина ξ имеет **нормальное** распределение $N(0, \sigma^2)$. Используя выборку $x \equiv (x_1, x_2, \ldots, x_n)$ этой случайной величины объёма n, определить наилучшую критическую область W_k для проверки гипотезы $H_0: m = m_0$ против альтернативной гипотезы $H_1: m = m_1$.

Ответ. При $m_0 < m_1$ имеем $W_k = \left\{ x: \sum_{i=1}^{n} x_i > h_1(\alpha) \right\}$; при $m_0 > m_1$ имеем $W_k = \left\{ x: \sum_{i=1}^{n} x_i < h_2(\alpha) \right\}$; $h_1(\alpha)$ и $h_2(\alpha)$ — константы, зависящие от уровня значимости α.

7.3. Пусть случайная величина ξ — число «успехов» в независимых испытаниях, а p — вероятность «успеха» в каждом испытании. Определить **наилучшую критическую** область для проверки гипотезы $H_0: p = p_0$ против альтернативной гипотезы $H_1: p = p_1$.

Ответ. При $p_0 > p_1$ имеем $W_k = \left\{ \sum_{i=1}^{n} x_i < b_\alpha \right\}$; при $p_0 < p_1$ имеем $W_k = \left\{ \sum_{i=1}^{n} x_i > c_\alpha \right\}$; α — уровень значимости.

7.4 [2]. Случайная величина ξ имеет **нормальное** распределение $N(m, 1)$. Проверяется гипотеза $H_0: m = 0$ против альтернативной гипотезы $H_1: m = 1$ на уровне значимости $\alpha = 0{,}05$. Сколько наблюдений необходимо провести, чтобы мощность критерия была не меньше 0,90?

Ответ. $n \geq 9$.

Задачи на построение ПКОВ

7.5. Проверка гипотезы о дисперсии нормальных наблюдений.

Построить последовательную процедуру различения следующих гипотез:

$$H_0:\ \xi_i \sim N(0,\sigma_0^2),\quad H_1:\ \xi_i \sim N(0,\sigma_1^2),\ \sigma_0 < \sigma_1,\quad \sigma_0 = 1{,}0,\quad \sigma_0 = 1{,}1.$$

Вероятности ошибок принять $\alpha = \beta = 10^{-2}$.

Ответ. Условие **продолжения наблюдений** имеет вид:

$$b_n = b + c\cdot n < \sum_{i=1}^{n} x_i^2 < a_n = a + c\cdot n,$$

где $a(n) = \ln A - n\ln\frac{\sigma_0}{\sigma_1}\ \frac{2\sigma_1^2\sigma_0^2}{\sigma_1^2-\sigma_0^2}$, $\qquad b(n) = \ln B - n\ln\frac{\sigma_0}{\sigma_1}\ \frac{2\sigma_1^2\sigma_0^2}{\sigma_1^2-\sigma_0^2}$.

Средние числа наблюдений: $\mathbf{M}(\nu\,|H_1) = \mathbf{M}(\nu\,|H_0) \approx 900.$

7.6. Различение гипотез о распределении временного промежутка.

На автостраде измеряются промежутки времени между автомобилями. Имеются две гипотезы

$$H_0:\ \xi_i \sim p_0(x) = \sqrt{\frac{2}{\pi}}e^{-x^2/2},\ x \ge 0;$$

$H_1:\ \xi_i \sim p_1(x) = e^{-x},\ x \ge 0$ (пуассоновский поток).

относительно распределения длины промежутка. Построить последовательное правило принятия решения (проверки гипотез) и определить средние числа наблюдений.

Ответ. Условие **продолжения наблюдений** имеет вид:

$$2(\ln B - nc) < \sum_{i=1}^{n}(x_i-1)^2 < 2(\ln A - nc),\ \text{где } c = \sqrt{\frac{2}{\pi e}}.$$