Занятие 5. Задача различения двух простых гипотез.

Домашнее задание:

отсутствует

Задача 5.1.

Задана выборка $(\xi_1,..., \xi_n)$ из показательного распределения $E(a)$ с неизвестным параметром a. В отношении параметра a выдвигаются две гипотезы: основная гипотеза H_0 утверждает, что $a = a_0 = 1$, альтернативная гипотеза H_1 утверждает, что $a = a_1 = 1.2$.

В условиях задачи необходимо построить критерий Неймана-Пирсона в следующих двух случаях:

1) заданы объем выборки $n = 100$ и вероятность ошибки первого рода $\alpha = 0.05$ (дополнительно вычислить вероятность ошибки второго рода критерия);

2) заданы вероятность ошибки первого рода $\alpha = 0.05$ и вероятность ошибки второго рода $\beta = 0.05$.

Решение:

1) Пусть $K_r(h) = (X_0(h), X_1(h))$ является критерием отношения вероятностей для различения двух простых гипотез H_0 и H_1, тогда согласно определению критерия отношения вероятностей, множество $X_0(h)$ реализаций $(x_1,..., x_n)$ выборки $(\xi_1,..., \xi_n)$ при которых принимается гипотеза H_0 определяется следующим соотношением:

$$X_0(h) = \left\{ (x_1,..., x_n) : \frac{f(x_1,..., x_n \mid a_1)}{f(x_1,..., x_n \mid a_0)} < h \right\},$$

где h – некоторое числовое значение, $f(x_1,..., x_n \mid a_1)$ и $f(x_1,..., x_n \mid a_0)$ – совместные плотности вероятности выборки $(\xi_1,..., \xi_n)$ при значениях параметра a равных a_1 и a_0 соответственно. Преобразуем отношение плотностей вероятности следующим образом:

$$\frac{f(x_1,..., x_n \mid a_1)}{f(x_1,..., x_n \mid a_0)} = \frac{\prod_{i=1}^{n} a_1 e^{-a_1 x_i}}{\prod_{i=1}^{n} a_0 e^{-a_0 x_i}} = \left(\frac{a_1}{a_0}\right)^n e^{-(a_1 - a_0)\sum_{i=1}^{n} x_i},$$

тогда выражение для множества $X_0(h)$ можно представить в следующем виде:

$$X_0(h) = \left\{ (x_1,..., x_n) : \left(\frac{a_1}{a_0}\right)^n e^{-(a_1 - a_0)\sum_{i=1}^{n} x_i} < h \right\} = \left\{ (x_1,..., x_n) : \ln\left(\left(\frac{a_1}{a_0}\right)^n e^{-(a_1 - a_0)\sum_{i=1}^{n} x_i} \right) < \ln h \right\} =$$

$$= \left\{ (x_1,..., x_n) : n \ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)\sum_{i=1}^{n} x_i < \ln h \right\} =$$

$$= \left\{ (x_1,..., x_n) : \sum_{i=1}^{n} x_i > \frac{n \ln\left(\frac{a_1}{a_0}\right) - \ln h}{a_1 - a_0} \right\} = \left\{ (x_1,..., x_n) : \sum_{i=1}^{n} x_i > \tilde{h}(n, h) \right\},$$

где $\tilde{h}(n, h) = \dfrac{n \ln\left(\frac{a_1}{a_0}\right) - \ln h}{a_1 - a_0}$. Соответственно множество $X_1(h)$ реализаций $(x_1,..., x_n)$ выборки $(\xi_1,..., \xi_n)$ при которых принимается гипотеза H_1 определяется соотношением:

$$X_1(h) = \left\{ (x_1,\ldots,x_n) : \frac{f(x_1,\ldots,x_n \mid a_1)}{f(x_1,\ldots,x_n \mid a_0)} \geq h \right\} = \left\{ (x_1,\ldots,x_n) : \sum_{i=1}^{n} x_i \leq \tilde{h}(n,h) \right\}.$$

Вероятность ошибки первого рода $\alpha_r(h)$ критерия отношения вероятностей $K(h)$ является вероятностью принятия гипотезы H_1 при условии, что верна гипотеза H_0, или иначе вероятностью попадания выборки $(\xi_1,\ldots,\xi_n)$ в множество X_1 (принимается гипотеза H_1) при условии, что неизвестный параметр a равен a_0 (верна гипотеза H_0):

$$\alpha_r(h) = P_{a_0}\{(\xi_1,\ldots,\xi_n) \in X_1(h)\} = P_{a_0}\left\{ \frac{f(\xi_1,\ldots,\xi_n \mid a_1)}{f(\xi_1,\ldots,\xi_n \mid a_0)} \geq h \right\} = P_{a_0}\left\{ \sum_{i=1}^{n} \xi_i \leq \tilde{h}(n,h) \right\}.$$

Поскольку вероятность вычисляется при условии, что неизвестный параметр a равен a_0, то каждая случайная величина ξ_i имеет показательное распределение $E(a_0)$ с математическим ожиданием $\frac{1}{a_0}$ и дисперсией $\frac{1}{a_0^2}$. При больших n распределение суммы случайных величин $\sum_{i=1}^{n} \xi_i$ можно считать приближенно нормальным с параметрами $n\frac{1}{a_0}$ и $n\frac{1}{a_0^2}$, тогда нормированная сумма имеет приближенно стандартное нормальное распределение:

$$\frac{\sum_{i=1}^{n} \xi_i - n\frac{1}{a_0}}{\sqrt{n\frac{1}{a_0^2}}} \sim N(0,1).$$

Используя полученное приближение для распределения суммы случайных величин, для вероятности ошибки первого рода можно получить следующее приближенное равенство:

$$\alpha_r(h) = P_{a_0}\left\{ \frac{\sum_{i=1}^{n} \xi_i - n\frac{1}{a_0}}{\sqrt{n\frac{1}{a_0^2}}} \leq \frac{\tilde{h}(n,h) - n\frac{1}{a_0}}{\sqrt{n\frac{1}{a_0^2}}} \right\} \approx \Phi\left(\frac{\tilde{h}(n,h) - n\frac{1}{a_0}}{\sqrt{n\frac{1}{a_0^2}}} \right).$$

Аналогично, вероятность ошибки второго рода $\beta_r(h)$ критерия отношения вероятностей K является вероятностью принятия гипотезы H_0 при условии, что верна гипотеза H_1, или иначе вероятностью попадания выборки $(\xi_1,\ldots,\xi_n)$ в множество X_0 (принимается гипотеза H_0) при условии, что неизвестный параметр a равен a_1 (верна гипотеза H_1):

$$\beta_r(h) = P_{a_1}\{(\xi_1,\ldots,\xi_n) \in X_0(h)\} = P_{a_1}\left\{ \frac{f(\xi_1,\ldots,\xi_n \mid a_1)}{f(\xi_1,\ldots,\xi_n \mid a_0)} < h \right\} =$$

$$= P_{a_1}\left\{ \sum_{i=1}^{n} \xi_i > \tilde{h}(n,h) \right\} = 1 - P_{a_1}\left\{ \sum_{i=1}^{n} \xi_i \leq \tilde{h}(n,h) \right\}.$$

Используя получено ранее приближение для распределения суммы $\sum_{i=1}^{n} \xi_i$, для вероятности ошибки второго рода получим следующее приближенное равенство:

$$\beta_r(h) = 1 - P_{a_1}\left\{ \frac{\sum_{i=1}^{n} \xi_i - n\frac{1}{a_1}}{\sqrt{n\frac{1}{a_1^2}}} \le \frac{\tilde{h}(n,h) - n\frac{1}{a_1}}{\sqrt{n\frac{1}{a_1^2}}} \right\} \approx 1 - \Phi\left(\frac{\tilde{h}(n,h) - n\frac{1}{a_1}}{\sqrt{n\frac{1}{a_1^2}}} \right).$$

Таким образом, получена система из двух приближенных равенств:

$$\begin{cases} \alpha_r(h) \approx \Phi\left(\dfrac{\tilde{h}(n,h) - n\frac{1}{a_0}}{\sqrt{n\frac{1}{a_0^2}}} \right) \\ \beta_r(h) \approx 1 - \Phi\left(\dfrac{\tilde{h}(n,h) - n\frac{1}{a_1}}{\sqrt{n\frac{1}{a_1^2}}} \right) \end{cases} \Leftrightarrow \begin{cases} \Phi^{-1}\left(\alpha_r(h)\right) \approx \dfrac{\tilde{h}(n,h) - n\frac{1}{a_0}}{\sqrt{n\frac{1}{a_0^2}}} \\ \Phi^{-1}\left(1 - \beta_r(h)\right) \approx \dfrac{\tilde{h}(n,h) - n\frac{1}{a_1}}{\sqrt{n\frac{1}{a_1^2}}} \end{cases}.$$

Полученная система является основополагающей для построения критериев отношения вероятностей. Действительно, если, например, заданы объем n и требуемая вероятность ошибки первого рода критерия $\alpha_r(h)$, то из полученной системы можно найти пороговое значение $\tilde{h}(n,h)$ и тем самым построить критерий (дополнительно из системы в этом случае может быть найдена вероятность ошибки второго рода $\beta_r(h)$). Если заданы требуемые значения вероятностей ошибок первого и второго родов $\alpha_r(h)$ и $\beta_r(h)$, то из полученной системы можно определить требуемый объем выборки n и пороговое значение $\tilde{h}(n,h)$.

2) Построим критерий Неймана-Пирсона для заданных значений объема выборки $n = 100$ и вероятности ошибки первого рода $\alpha = 0.05$: согласно определению критерий Неймана-Пирсона является критерием отношения вероятностей $K^* = K_r(h^*)$, для которого вероятность ошибки первого рода $\alpha_r(h^*)$ в точности равна α. Подставляя значение α в систему приближенных равенств для критерия $K_r(h^*)$ получим следующую систему:

$$\begin{cases} \Phi^{-1}(\alpha) \approx \dfrac{\tilde{h}(n,h^*) - n\frac{1}{a_0}}{\sqrt{n\frac{1}{a_0^2}}} \\ \Phi^{-1}\left(1 - \beta_r(h^*)\right) \approx \dfrac{\tilde{h}(n,h^*) - n\frac{1}{a_1}}{\sqrt{n\frac{1}{a_1^2}}} \end{cases} \Leftrightarrow \begin{cases} \tilde{h}(n,h^*) \approx n\dfrac{1}{a_0} + \Phi^{-1}(\alpha)\sqrt{n\dfrac{1}{a_0^2}} \\ \beta_r(h^*) \approx 1 - \Phi\left(\dfrac{\tilde{h}(n,h^*) - n\frac{1}{a_1}}{\sqrt{n\frac{1}{a_1^2}}} \right) \end{cases}$$

Вычислим значения $\tilde{h}(n,h^*)$ и $\beta_r(h^*)$:

$$\tilde{h}(n,h^*) \approx 100\frac{1}{1} + \Phi^{-1}(0.05)\sqrt{100\frac{1}{1^2}} \approx 100 - 1.65 \cdot 10 = 83.5,$$

$$\beta_r(h^*) \approx 1 - \Phi\left(\frac{83.5 - 100\frac{1}{1.2}}{\sqrt{100\frac{1}{1.2^2}}} \right) \approx 1 - \Phi\left(\frac{83.5 - 83.33}{8.33} \right) \approx 1 - \Phi(0.02) \approx 1 - 0.508 = 0.492 .$$

3) Построим критерий Неймана-Пирсона для заданных значений вероятностей ошибки первого рода $\alpha = 0.05$ и второго рода $\beta = 0.05$: согласно определению критерий Неймана-Пирсона является критерием отношения вероятностей $\hat{K} = K_r(\hat{h})$, для которого вероятность ошибки первого рода $\alpha_r(\hat{h})$ в точности равна α . При этом однако требуется, чтобы вероятность ошибки второго рода $\beta_r(\hat{h})$ была равна β . Подставляя значения α и β в систему приближенных равенств, получим систему:

$$\begin{cases} \Phi^{-1}(\alpha) \approx \dfrac{\tilde{h}(n,\hat{h}) - n\dfrac{1}{a_0}}{\sqrt{n\dfrac{1}{a_0^2}}} \\ \Phi^{-1}(1-\beta) \approx \dfrac{\tilde{h}(n,\hat{h}) - n\dfrac{1}{a_1}}{\sqrt{n\dfrac{1}{a_1^2}}} \end{cases} \Leftrightarrow \begin{cases} \tilde{h}(n,\hat{h}) \approx n\dfrac{1}{a_0} + \Phi^{-1}(\alpha)\sqrt{n\dfrac{1}{a_0^2}} \\ \tilde{h}(n,\hat{h}) \approx n\dfrac{1}{a_1} + \Phi^{-1}(1-\beta)\sqrt{n\dfrac{1}{a_1^2}} \end{cases}$$

Вычитая из первого приближенного равенства второе и преобразуя, получим приближенное равенство для объема n :

$$0 \approx n\frac{1}{a_0} + \Phi^{-1}(\alpha)\sqrt{n\frac{1}{a_0^2}} - n\frac{1}{a_1} - \Phi^{-1}(1-\beta)\sqrt{n\frac{1}{a_1^2}},$$

$$0 \approx n\left(\frac{1}{a_0} - \frac{1}{a_1}\right) + \sqrt{n}\left(\frac{\Phi^{-1}(\alpha)}{a_0} - \frac{\Phi^{-1}(1-\beta)}{a_1}\right)$$

$$0 \approx \sqrt{n}\left(\frac{a_1 - a_0}{a_1 a_0}\right) + \left(\frac{a_1\Phi^{-1}(\alpha) - a_0\Phi^{-1}(1-\beta)}{a_1 a_0}\right),$$

$$n \approx \left(\frac{a_1\Phi^{-1}(\alpha) - a_0\Phi^{-1}(1-\beta)}{a_0 - a_1}\right)^2.$$

Таким образом,

$$\begin{cases} n \approx \left(\dfrac{a_1\Phi^{-1}(\alpha) - a_0\Phi^{-1}(1-\beta)}{a_0 - a_1}\right)^2 \\ \tilde{h}(n,\hat{h}) \approx n\dfrac{1}{a_0} + \Phi^{-1}(\alpha)\sqrt{n\dfrac{1}{a_0^2}} \end{cases}.$$

Вычислим значения n и $\tilde{h}(n,\hat{h})$:

$$n \approx \left(\frac{1.2\cdot\Phi^{-1}(0.05) - 1\cdot\Phi^{-1}(1-0.05)}{1-1.2}\right)^2 \approx \left(\frac{1.2\cdot(-1.65) - 1.65}{-0.2}\right)^2 \approx 329.42 < 330$$

$$\tilde{h}(n,\hat{h}) \approx 330\frac{1}{1} + \Phi^{-1}(0.05)\sqrt{330\frac{1}{1^2}} \approx 330 - 1.65\cdot 18.16 \approx 300 .$$

Ответ:

1) Критерий $K^* = (X_0^*, X_1^*)$:

$$X_0^* = \left\{(x_1,\dots,x_n) : \sum_{i=1}^{n} x_i > 83.5\right\},\ X_1^* = \left\{(x_1,\dots,x_n) : \sum_{i=1}^{n} x_i \le 83.5\right\},$$

вероятность ошибки второго рода $\beta(K^*) \approx 0.492$.

2) Критерий $\hat{K} = (\hat{X}_0, \hat{X}_1)$:

$$\hat{X}_0 = \left\{ (x_1,..., x_n) : \sum_{i=1}^{n} x_i > 300 \right\}, \ \hat{X}_1 = \left\{ (x_1,..., x_n) : \sum_{i=1}^{n} x_i \le 300 \right\},$$

где $n = 330$.

Задача 5.2.

В условиях задачи 6.1 и пункта 2 построить последовательный критерий отношения вероятностей с вероятностями ошибок первого и второго родов приближенно равных α и β , и вычислить средние объемы выборки для построенного критерия.

Решение:

1) В качестве требуемого критерия может быть использован последовательный критерий отношения вероятностей $\tilde{K}_s = K_s\left(\frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha}\right)$, условие продолжения для которого имеет следующий вид:

$$\frac{\beta}{1-\alpha} < \frac{f(x_1,..., x_k \mid a_1)}{f(x_1,..., x_k \mid a_0)} < \frac{1-\beta}{\alpha},$$

где $(x_1,..., x_k)$ – реализация выборки $(\xi_1,..., \xi_k)$, которой располагает последовательный критерий на шаге k, $f(x_1,..., x_k \mid a_1)$ и $f(x_1,..., x_k \mid a_0)$ – плотности вероятности выборки $(\xi_1,..., \xi_k)$, при условии, что значение параметра a равно a_1 и a_0 соответственно. Используя преобразование отношения плотностей вероятности из пункта 1) решения задачи 6.1, преобразуем условие продолжения к следующему виду:

$$\frac{\beta}{1-\alpha} < \frac{f(x_1,..., x_k \mid a_1)}{f(x_1,..., x_k \mid a_0)} < \frac{1-\beta}{\alpha},$$

$$\frac{\beta}{1-\alpha} < \left(\frac{a_1}{a_0}\right)^n e^{-(a_1-a_0)\sum_{i=1}^{n} x_i} < \frac{1-\beta}{\alpha},$$

$$\ln\left(\frac{\beta}{1-\alpha}\right) < k\ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)\sum_{i=1}^{k} x_i < \ln\left(\frac{1-\beta}{\alpha}\right),$$

$$\frac{k\ln\left(\frac{a_1}{a_0}\right) - \ln\left(\frac{\beta}{1-\alpha}\right)}{(a_1 - a_0)} > \sum_{i=1}^{k} x_i > \frac{k\ln\left(\frac{a_1}{a_0}\right) - \ln\left(\frac{1-\beta}{\alpha}\right)}{(a_1 - a_0)}.$$

Согласно определению последовательного критерия отношения вероятностей, критерий $\tilde{K}_s$ принимает гипотезу H_0 на шаге k , если:

$$\frac{f(x_1,..., x_k \mid a_1)}{f(x_1,..., x_k \mid a_0)} < \frac{\beta}{1-\alpha},$$

$$\sum_{i=1}^{k} x_i > \frac{k\ln\left(\frac{a_1}{a_0}\right) - \ln\left(\frac{\beta}{1-\alpha}\right)}{(a_1 - a_0)},$$

и принимает гипотезу H_1 на шаге k , если:

$$\frac{1-\beta}{\alpha} < \frac{f(x_1,..., x_k \mid a_1)}{f(x_1,..., x_k \mid a_0)},$$

$$\frac{k\ln\left(\frac{a_1}{a_0}\right) - \ln\left(\frac{1-\beta}{\alpha}\right)}{(a_1 - a_0)} > \sum_{i=1}^{k} x_i .$$

2) Для последовательного критерия отношения вероятностей $\tilde{K}_s$ объем выборки ν, требуемый для принятия решение о том, какую гипотезу принять, является случайной величиной, распределение которой зависит от того какое значение a_0 либо a_1 имеет неизвестный параметр a (какая из гипотез H_0 либо H_1 верна).

Известно, что если значение параметра a совпадает с a_0, то математическое ожидание ν (средний объем выборки, при условии, что верна гипотеза H_0) определяется соотношением:

$$M_{a_0}[\nu] = \frac{\ln\left(\frac{1-\beta}{\alpha}\right)\cdot\alpha + \ln\left(\frac{\beta}{1-\alpha}\right)\cdot(1-\alpha)}{M_{a_0}[\zeta]},$$

где случайная величина ζ имеет вид:

$$\zeta = \ln\left(\frac{f_\xi(\xi \mid a_1)}{f_\xi(\xi \mid a_0)}\right) = \ln\left(\frac{a_1 e^{-a_1\xi}}{a_0 e^{-a_0\xi}}\right) = \ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)\xi$$

и ξ – случайная величина, имеющая показательное распределение $E(a_0)$. Математическое ожидание $M_{a_0}[\zeta]$ имеет вид:

$$M_{a_0}[\zeta] = M_{a_0}\left[\ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)\xi\right] = \ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)M_{a_0}[\xi] = \ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)\frac{1}{a_0},$$

откуда,

$$M_{a_0}[\nu] = \frac{\ln\left(\frac{1-\beta}{\alpha}\right)\cdot\alpha + \ln\left(\frac{\beta}{1-\alpha}\right)\cdot(1-\alpha)}{\ln\left(\frac{a_1}{a_0}\right) - \left(\frac{a_1}{a_0} - 1\right)}.$$

Если же значение параметра a совпадает с a_1, то математическое ожидание ν (средний объем выборки, при условии, что верна гипотеза H_1) определяется соотношением:

$$M_{a_1}[\nu] = \frac{\ln\left(\frac{\beta}{1-\alpha}\right)\cdot\beta + \ln\left(\frac{1-\beta}{\alpha}\right)\cdot(1-\beta)}{M_{a_1}[\zeta]},$$

где случайная величина ζ имеет вид:

$$\zeta = \ln\left(\frac{f_\xi(\xi \mid a_1)}{f_\xi(\xi \mid a_0)}\right) = \ln\left(\frac{a_1 e^{-a_1\xi}}{a_0 e^{-a_0\xi}}\right) = \ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)\xi,$$

и ξ – случайная величина, имеющая показательное распределение $E(a_1)$. Математическое ожидание $M_{a_1}[\zeta]$ имеет вид:

$$M_{a_1}[\zeta] = M_{a_1}\left[\ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)\xi\right] = \ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)M_{a_1}[\xi] = \ln\left(\frac{a_1}{a_0}\right) - (a_1 - a_0)\frac{1}{a_1},$$

откуда,

$$M_{a_1}[\nu] = \frac{\ln\left(\frac{\beta}{1-\alpha}\right)\cdot\beta + \ln\left(\frac{1-\beta}{\alpha}\right)\cdot(1-\beta)}{\ln\left(\frac{a_1}{a_0}\right) - \left(1 - \frac{a_0}{a_1}\right)}.$$

Подставим в полученные для $M_{a_0}[\zeta]$ и $M_{a_1}[\zeta]$ выражения числовые значения $\alpha = 0.05$, $\beta = 0.05$, $a_0 = 1$ и $a_1 = 1$:

$$M_{a_0}[\nu] = \frac{\ln\left(\frac{1-0.05}{0.05}\right)\cdot 0.05 + \ln\left(\frac{0.05}{1-0.05}\right)\cdot(1-0.05)}{\ln\left(\frac{1.2}{1}\right) - \left(\frac{1.2}{1} - 1\right)} \approx 149.9,$$

$$M_{a_1}[\nu] = \frac{\ln\left(\frac{0.05}{1-0.05}\right)\cdot 0.05 + \ln\left(\frac{1-0.05}{0.05}\right)\cdot(1-0.05)}{\ln\left(\frac{1.2}{1}\right) - \left(1 - \frac{1}{1.2}\right)} \approx 169.3.$$

Ответ:

Последовательный критерий отношения вероятностей $K_s\left(\frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha}\right)$, вероятности ошибок которого приближенно равны α и β, на шаге k:

1) принимает гипотезу H_0, если:

$$\sum_{i=1}^{k} x_i > \frac{k\ln\left(\frac{a_1}{a_0}\right) - \ln\left(\frac{\beta}{1-\alpha}\right)}{(a_1 - a_0)}$$

2) продолжается, если:

$$\frac{k\ln\left(\frac{a_1}{a_0}\right) - \ln\left(\frac{\beta}{1-\alpha}\right)}{(a_1 - a_0)} > \sum_{i=1}^{k} x_i > \frac{k\ln\left(\frac{a_1}{a_0}\right) - \ln\left(\frac{1-\beta}{\alpha}\right)}{(a_1 - a_0)}$$

3) принимает гипотезу H_0, если:

$$\frac{k\ln\left(\frac{a_1}{a_0}\right) - \ln\left(\frac{1-\beta}{\alpha}\right)}{(a_1 - a_0)} > \sum_{i=1}^{k} x_i .$$

Для критерия $K_s\left(\frac{\beta}{1-\alpha}, \frac{1-\beta}{\alpha}\right)$ средний объем выборки приближенно равен 150, при условии, что верна гипотеза H_0, и приближенно равен 170, при условии, что верна гипотеза H_1.

Задача 5.3.

В условиях задачи 6.1 задан объем выборки $n = 100$ и требуется построить минимаксный критерий K_m и вычислить вероятности ошибок первого и второго рода построенного критерия.

Решение:

1) Известно, что минимаксный критерий K_m (при некоторых условиях) совпадает с критерием отношения вероятностей $K_r(h_m)$, в котором пороговое значение h_m выбирается из условия равенства функций вероятности ошибки первого и второго рода критериев отношения вероятностей $\alpha_r(h)$ и $\beta_r(h)$:

$$\alpha_r(h_m) = \beta_r(h_m).$$

Ранее в решении задачи 6.1 пункт 1 были получены выражения для вероятностей ошибок критериев отношения вероятностей:

$$\alpha_r(h) \approx \Phi\left(\frac{\tilde{h}(n,h) - n\dfrac{1}{a_0}}{\sqrt{n\dfrac{1}{a_0^2}}}\right),$$

$$\beta_r(h) \approx 1 - \Phi\left(\frac{\tilde{h}(n,h) - n\dfrac{1}{a_1}}{\sqrt{n\dfrac{1}{a_1^2}}}\right).$$

Откуда, приравнивая функции вероятностей ошибок, получим уравнение относительно h_m:

$$\Phi\left(\frac{\tilde{h}(n,h_m) - n\dfrac{1}{a_0}}{\sqrt{n\dfrac{1}{a_0^2}}}\right) \approx 1 - \Phi\left(\frac{\tilde{h}(n,h_m) - n\dfrac{1}{a_1}}{\sqrt{n\dfrac{1}{a_1^2}}}\right).$$

В действительности не обязательно вычислять значение h_m, поскольку для построения критерия $K_r(h_m)$ будет вполне достаточно вычислить значение выражения $\tilde{h}(n,h_m)$.

Используя свойство функции Лапласа ($\Phi(-x) = 1 - \Phi(x)$), преобразуем уравнение к следующему виду:

$$\Phi\left(\frac{\tilde{h}(n,h_m) - n\dfrac{1}{a_0}}{\sqrt{n\dfrac{1}{a_0^2}}}\right) \approx \Phi\left(-\frac{\tilde{h}(n,h_m) - n\dfrac{1}{a_1}}{\sqrt{n\dfrac{1}{a_1^2}}}\right).$$

Поскольку функция Лапласа $\Phi(x)$ является функцией строго возрастающей, то из равенства значений функции следует равенство аргументов:

$$\frac{\tilde{h}(n,h_m) - n\dfrac{1}{a_0}}{\sqrt{n\dfrac{1}{a_0^2}}} \approx -\frac{\tilde{h}(n,h_m) - n\dfrac{1}{a_1}}{\sqrt{n\dfrac{1}{a_1^2}}},$$

$$\frac{a_0\tilde{h}(n,h_m) - n}{\sqrt{n}} \approx -\frac{a_1\tilde{h}(n,h_m) - n}{\sqrt{n}},$$

$$a_0\tilde{h}(n,h_m) - n \approx -a_1\tilde{h}(n,h_m) + n,$$

$$\tilde{h}(n,h_m) \approx \frac{2n}{a_1 + a_0}.$$

Отсюда минимаксный критерий $K_m = K_r(h_m) = (X_0(h_m), X_1(h_m))$ имеет следующий вид:

$$X_0(h_m) = \left\{(x_1,\ldots,x_n) : \sum_{i=1}^{n} x_i > \tilde{h}(n,h_m)\right\} = \left\{(x_1,\ldots,x_n) : \sum_{i=1}^{n} x_i > \frac{2n}{a_1 + a_0}\right\},$$

$$X_1(h_m) = \left\{(x_1,\ldots,x_n) : \sum_{i=1}^{n} x_i \le \tilde{h}(n,h_m)\right\} = \left\{(x_1,\ldots,x_n) : \sum_{i=1}^{n} x_i \le \frac{2n}{a_1 + a_0}\right\}.$$

2) Подставляя величины параметров a_0 и a_1 и значение n, получим приближенное значение для $\tilde{h}(n, h_m)$:

$$\tilde{h}(n, h_m) = \frac{2 \cdot 100}{1.2 + 1} \approx 91 .$$

Для критерия $K_m = (X_0(h_m), X_1(h_m))$ получим следующие множества:

$$X_0(h_m) = \left\{ (x_1,..., x_n) : \sum_{i=1}^{n} x_i > 91 \right\}, \; X_1(h_m) = \left\{ (x_1,..., x_n) : \sum_{i=1}^{n} x_i \le 91 \right\}.$$

3) Используя полученное выражение для $\tilde{h}(n, h_m)$, вычислим величину вероятности ошибки первого рода $\alpha(K_m)$ минимаксного критерия K_m (эта же величина будет и вероятностью ошибки второго рода $\beta(K_m)$, поскольку для минимаксного критерия вероятности ошибок равны):

$$\alpha(K_m) = \alpha_r(h_m) \approx \Phi\left(\frac{\tilde{h}(n,h) - n\frac{1}{a_0}}{\sqrt{n\frac{1}{a_0^2}}} \right) = \Phi\left(\frac{\frac{2n}{a_1 + a_0} - n\frac{1}{a_0}}{\sqrt{n\frac{1}{a_0^2}}} \right),$$

$$\alpha_r(K_m) \approx \Phi\left(\frac{91 - 100}{10} \right) = \Phi\left(-0.9\right) \approx 0.1841 .$$

Ответ:

1) Минимаксный критерий $K_m = (X_0, X_1)$ имеет вид:

$$X_0 = \left\{ (x_1,..., x_n) : \sum_{i=1}^{n} x_i > 91 \right\},$$

$$X_1 = \left\{ (x_1,..., x_n) : \sum_{i=1}^{n} x_i \le 91 \right\}.$$

2) Вероятности ошибок первого рода $\alpha(K_m)$ и второго рода $\beta(K_m)$ минимаксного критерия K_m:

$$\alpha(K_m) = \beta(K_m) \approx 0.1841 .$$