МАТЕМАТИЧЕСКАЯ СТАТИСТИКА
Лекция 10

Повтор

§ 10. Различение двух простых гипотез
10.1. Фиксированный объем наблюдений

Пусть имеется совокупность наблюдений

$x = (x_1, x_2 ... x_n)$ - реализации случайных величин

$\xi \equiv (\xi_1, \xi_2 \ldots \xi_n)$,

Имеется два предположения (гипотезы) относительно распределения:

1. H_0: ξ распределена по закону $p_0(x)$;
2. H_1: ξ - по закону $p_1(x)$

($p_0(x)$, $p_1(x)$ — плотности, если ξ — непрерывна, если дискретна — вероятности).

По x требуется принять одно из двух решений:

«*верна H_0*» (это решение обозначим «0»),

«*верна H_1*» (решение «1»).

Ясно, что задача сводится к определению решающей функции $\delta(x)$, имеющей два значения 0 и 1, т.е. к определению разбиения $\Gamma = (\Gamma_0, \Gamma_1)$ пространства X всех возможных значений x:

$$\delta(x) = \begin{cases} 0, & \text{если } x \in \Gamma_0, \\ 1, & \text{если } x \in \Gamma_1. \end{cases} \qquad \Gamma_0 \cup \Gamma_1 = X, \quad \Gamma_0 \cap \Gamma_1 = \varnothing$$

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

1) ошибка первого рода — принятие H_1 при истинности H_0;
2) ошибка второго рода — принятие H_0 при истинности H_1.

Любая решающая функция характеризуется двумя условными вероятностями:

$$\alpha = P(\text{принять } H_1 \mid H_0) = \int_{\Gamma_1} p_0(x)dx, \tag{1}$$

$$\beta = P(\text{принять } H_0 \mid H_1) = \int_{\Gamma_0} p_1(x)dx,$$

Байесовский подход

Будем считать, что ***многократно сталкиваемся*** с необходимостью ***выбора между H_0 и H_1.*** В этом случае можно говорить о частоте, с которой истинна H_0 (или H_1), т.е. о том, что истинность H_0 (или H_1) — событие случайное, причем вероятность события, когда верна H_0, а когда верна H_1, известны:

$$P(H_0) = q_0, \quad P(H_1) = q_1, \qquad q_0 + q_1 = 1.$$

Кроме того, будем считать, что при каждой ошибке несем потери.

При ошибке первого рода:

$$W_0 \qquad \text{с вер.} \qquad P(H_0)\, P(\text{принять } H_1 \mid H_0)),$$

а при ошибке второго рода потери

W_1 с вер. $P(H_1)\,P$(принять $H_0 \mid H_1$)).

Если пользуемся правилом $\delta(\Gamma)$ (с разбиением Γ), то средний штраф от однократного использования:

$$R(\Gamma) = q_0 \cdot \alpha(\Gamma) \cdot W_0 + q_1 \cdot \beta(\Gamma) \cdot W_1 . \qquad (1а)$$

Назовем правило δ (соответственно, разбиение $\Gamma \equiv (\Gamma_0, \Gamma_1)$) **в байесовском смысле оптимальным,** если

$$R(\Gamma) = \min_{\Gamma'} R(\Gamma').$$

Оказывается справедливой следующая теорема.

Теорема. *Оптимальным является правило, для которого область* Γ_1 *принятия гипотезы* H_1 *определяется соотношением*:

$$\Gamma_1 = \left\{ x : \frac{p_1(x)}{p_0(x)} \geq h \equiv \frac{q_0 W_0}{q_1 W_1} \right\}. \qquad (2)$$

Действительно, пусть Т = (T_0, T_1) — произвольное разбиение; для него средние потери

Конец повтора

Подход Неймана-Пирсона

Оптимальным (в смысле Неймана-Пирсона) назовем такое правило, которое имеет заданную вероятность α ошибки первого рода, а вероятность β ошибки второго рода при этом минимальна, т.е. правило δ (соответственно разбиение Г) оптимально, если

$$\beta(\Gamma) = \min_{\Gamma'} \beta(\Gamma'), \qquad (3а)$$

при условии $\alpha(\Gamma') \leq \alpha_0$.

(3б)

Оказывается справедливой следующая теорема.

Теорема. *Оптимальным является правило, для которого область* Γ_1 *определяется соотношением*:

$$\Gamma_1 = \left\{ x : \frac{p_1(x)}{p_0(x)} \geq h \right\}, \qquad (3в)$$

где h определяется из условия

$$\alpha(h) = \alpha_0. \qquad (4а)$$

Заметим, что статистика осталась той же самой

Действительно, во-первых, будем предполагать, что можно подобрать такое h, при котором **в (4а) имеем равенство.** В пользу этого говорит то, что $\alpha(h)$ является невозрастающей по h функцией:

если $h_2 > h_1$, то $\Gamma_1(h_2) \subseteq \Gamma_1(h_1)$, и потому $\alpha(h_2) \leq \alpha(h_1)$,

причем $\alpha(0) = 1$ и $\alpha(\infty) = 0$ (поскольку $\Gamma_1(0) = X$, $\Gamma_1(\infty) = \varnothing$).

Далее, пусть Т = (T_0, T_1) — произвольное разбиение, причем

$$\alpha(T) = P(\text{принять } H_1|H_0, T) = \int\limits_{T_1} p_0(x)dx \le \alpha_0. \quad (4б)$$

Покажем, что $\beta(T) > \beta(\Gamma)$. Оценим разность $\beta(T) - \beta(\Gamma)$. Обозначим

$U = T_0 \cap \Gamma_0$ — общую часть T_0 и Γ_0,

которую можно убрать из T_0 и Γ_0 при оценке разности:

$$\beta(T) - \beta(\Gamma) = P(\text{пр.} H_0 \,|\, H_1, T) - P(\text{пр. } H_0 |H_1, \Gamma) = \int\limits_{T_0} p_1(x)dx - \int\limits_{\Gamma_0} p_1(x)dx$$

$$= \left(\int\limits_{T_0} p_1(x)dx - \int\limits_{U} p_1(x)dx\right) - \left(\int\limits_{\Gamma_0} p_1(x)dx - \int\limits_{U} p_1(x)dx\right) = \int\limits_{(T_0 \setminus U) \subseteq \Gamma_1} p_1(x)dx - \int\limits_{\Gamma_0 \setminus U \subseteq \Gamma_0} p_1(x)dx .$$

Учтем, что из T_0 убрали точки, входящие в Γ_0, так что $(T_0 \setminus U) \subseteq \Gamma_1$,
из Γ_0 убрали некоторые точки, но $(\Gamma_0 \setminus U) \subseteq \Gamma_0$,
и потому в силу (3в) для точек из Γ_1 справедливо

$$p_1(x) \ge h \cdot p_0(x),$$

а для точек из Γ_0:

$$p_1(x) < h \cdot p_0(x).$$

Учтем неравенства, добавим $\left(\int\limits_U p_0 dx - \int\limits_U p_0 dx\right)$ и учитывая (4), получаем

$$\beta(T) - \beta(\Gamma) > h\left[\int\limits_{T_0 \setminus U} p_0(x)dx - \int\limits_{\Gamma_0 \setminus U} p_0(x)dx\right] = h\left[\int\limits_{T_0} p_0(x)dx - \int\limits_{\Gamma_0} p_0(x)dx\right] =$$

$$= h\{P(\text{пр. } H_0|H_0,T) - P(\text{пр. } H_0|H_0,\Gamma)\} =$$

$$= h\{[1-\alpha(T)] - [1-\alpha(\Gamma)]\} = h\{\alpha(\Gamma) - \alpha(T)\} = h\{\alpha_0 - \alpha(T) \ge 0.$$

Замечание. Приведенный результат есть частный случай фундаментальной леммы Неймана-Пирсона, справедливый при условии, что существует корень h уравнения (4а). Можно, однако, привести примеры, когда $\alpha(h)$ изменяется скачком, и тогда (3в) требует некоторого простого уточнения.

10.2. Пример. Различение гипотез о среднем нормальной совокупности (*аналогичные примеры нужно уметь решать)*

На вход канала связи подается сигнал S, который может принимать два значения:

$S = 0$ (сигнала нет: нет цели), $S = a \neq 0$ (сигнал есть: есть цель).

В канале действует аддитивная случайная ошибка ε, нормально распределенная со средним $M\varepsilon = 0$ и дисперсией $D\varepsilon = \sigma^2$; результатом является

$$x' = S + \varepsilon.$$

Измерения повторяются n раз, так что на выходе имеются наблюдения

$$(x_1, x_2 \ldots x_n) \equiv x,$$

по которым нужно решить,

есть ли сигнал (H_1: $S = a$) или нет (H_0: $S = 0$).

Требуется построить решающее правило δ, имеющее заданную вероятность α_0 ошибки первого рода (вероятность ложной тревоги)

$$\alpha \equiv P(\text{пр. } H_1\text{: есть цель} \mid H_0\text{: нет цели}) = \alpha_0$$

при минимальном значении вероятности β ошибки второго рода (вероятности пропуска):

$$\beta \equiv P(\text{пр. } H_0\text{: нет цели} \mid H_1\text{: есть цель}) \to \min$$

Считая ошибки независимыми, имеем две плотности для наблюдений:

$$p_1(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-a)^2}{2\sigma^2}}, \qquad p_0(x) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_i^2}{2\sigma^2}}.$$

В соответствии с теоремой, решение о наличии сигнала нужно принять, т.е. принять H_1, если x попадает в Γ_1, где (логарифм не изменяет область)

$$\Gamma_1 = \left\{ x : \ln\frac{p_1(x)}{p_0(x)} \geq \ln h \equiv h_1 \right\} = \left\{ x : \frac{1}{2\sigma^2}\left(2a\sum_{i=1}^{n} x_i - na^2\right) \geq h_1 \right\} = \left\{ x : \sum_{i-1}^{n} x_i \geq h_2 \right\},$$

где обозначено $h_2 \equiv \left(h_1 2\sigma^2 + na^2\right)/2a$;.

Итак, если

$$\sum_{i=1}^{n} x_i \geq h_2, \tag{5}$$

то принимается H_1, в противном случае принимается H_0. Порог h_2 определяется из заданного α_0:

$$\alpha(h_2) = P\{\text{пр. } H_1 / H_0\} = P\left(\sum_{i=1}^{n} x_i \geq h_2 / H_0\right) = \alpha_0.$$

Если верна H_0, то $\sum_{i=1}^{n} x_i \sim N(0, n\sigma^2)$, и потому:

$$\alpha(h_2) = 1 - \Phi(h_2/\sigma\sqrt{n}) = \alpha_0, \quad \Rightarrow \quad (h_2/\sigma\sqrt{n}) = \Phi^{-1}(1-\alpha_0) = Q(1-\alpha_0),$$

откуда

$$h_2 = \sigma\sqrt{n}\, Q(1 - \alpha_0), \tag{6}$$

где $\Phi(x)$ — функция нормального $N(0, 1)$ распределения; $Q(1 - \alpha_0)$ — квантиль порядка $(1 - \alpha_0)$ этого распределения.

Определим вероятность β ошибки второго рода для нашей процедуры. Если верна H_1, то $\sum_{i=1}^{n} x_i \sim N(na, n\sigma^2)$, и потому

$$\beta = P(\text{пр. } H_0 / H_1) = P\{\sum_{i=1}^{n} x_i < h_2 / H_1\} = \Phi\left(\frac{h_2 - na}{\sigma\sqrt{n}}\right) = \Phi(Q - \frac{a}{\sigma}\sqrt{n}).$$

Посмотрим результаты в числах

Положим, a = 0,2, σ = 1,0 (т.е. ошибка σ в 5 раз больше сигнала a), n = 500, $\alpha_0 = 10^{-2}$; при этом

$h_2 = 1\ \sqrt{500}\cdot 2{,}33 = 52$, $\beta = \Phi(2{,}33 - 0{,}2 \cdot 22{,}4) = \Phi(-2{,}14) = 1{,}6\cdot 10^{-2}$.

Как видим, вероятности ошибок порядка 10^{-2}.

10.3. Последовательное различение двух простых гипотез (последовательный критерий отношения вероятностей Вальда)

Многие практические задачи для достаточно надежного принятия решения требуют уменьшения времени набора данных, например: испытания надежности, военные задачи, оценка эффективности экономических и других сложных систем. Возникает задача: разработать *ускоренные статистические процедуры*, имеющие заданные характеристики качества. Общая идея построения таких процедур состоит в том, чтобы

не фиксировать объем наблюдений, а накапливать наблюдения до тех пор, пока не будет достаточно информации для получения выводов требуемой надежности.

Соответствующие процедуры разрабатываются в разделе математической статистики, который называется *последовательным анализом*. Рассмотрим задачу, с которой началось это направление в мат. стат-ке:

последовательное различение двух простых гипотез.

Мы хотели бы пользоваться таким правилом различения, которое имело бы заданные уровни вероятностей ошибок α и β, и при этом требовало бы минимальное в среднем число наблюдений.

Пусть x_1, $x_2 \dots x_n$ ...— последовательность (не фиксированной длины) независимых, одинаково распределенных случайных величин. Относительно распределения имеется два предположения (две гипотезы):

H_0: наблюдения распределены с плотностью $p_0(x)$;

H_1: наблюдения распределены с плотностью $p_1(x)$;

(если наблюдения дискретны, то $p_0(x)$, $p_1(x)$ — вероятности, а аргумент дискретный).

После каждого наблюдения статистику предоставляется выбор из трех возможных решений:

— принять H_0 и закончить наблюдения;

— принять H_1 и закончить наблюдения;

— не принимать ни одну из гипотез и продолжить наблюдения (т.е. сделать еще одно наблюдение).

А. Формулировка решающего правила и его оптимальность. Рассмотрим следующую процедуру δ^*, называемую *последо-*

вательный критерий отношения вероятностей (ПКОВ). Она определяется двумя порогами:

верхним А и нижним В, при этом $0 < B < 1 < A$.

Пусть уже получено n наблюдений (n = 1, 2...); запишем отношение правдоподобия:

$$L_n(x_1, \ldots, x_n) = \frac{p(x_1, x_2, \ldots x_n | H_1)}{p(x_1, x_2, \ldots x_n | H_0)} = \prod_{i=1}^{n} \frac{p_1(x_i)}{p_0(x_i)} \equiv \frac{p_{1n}(x^n)}{p_{0n}(x^n)}, \qquad (7)$$

где обозначено $p_{1n}(x^n) = \prod_{i=1}^{n} p_1(x_i)$, $p_{0n}(x^n) = \prod_{i=1}^{n} p_0(x_i)$.

Процедура δ^* на очередном шаге n такова:

1) если $L_n(x_1, x_2 \ldots x_n) \le B$, то принимается H_0 и наблюдения заканчиваюся;

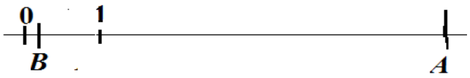
2) если $L_n(x_1, x_2 \ldots x_n) \ge A$, то принимается H_1 и наблюдения заканчиваются;

Примечание [MA1]: Номер формулы на месте? Здесь и должен быть?

3) если

$$B < L_n(x_1, x_2 \ldots x_n) < A, \qquad (8)$$

то наблюдения продолжаются (т.е. делается еще одно наблюдение).



Очевидно, эта процедура характеризуется некоторыми вероятностями ошибок и средними числами наблюдений:

$$\alpha = \alpha(A, B) = P\{\text{пр. } H_1 / H_0\}, \quad \beta = \beta(A, B) = P\{\text{пр. } H_0 / H_1\},$$

$$n_0 = n_0(A, B) = M(\nu / H_0), \qquad n_1 = n_1(A, B) = M(\nu / H_1),$$

где ν — число наблюдений (случайная величина) до принятия окончательного решения. Если α_0 и β_0 заданы, то можно найти пороги A и B, т.е. правило δ^*. Оказывается, такое правило обладает свойством оптимальности.

Теорема (Вальд и Вольфовиц, 1948 г.). Среди всех решающих правил δ', обладающих свойством

$$\alpha(\delta') \le \alpha_0, \quad \beta(\delta') \le \beta_0, \qquad (9)$$

последовательный критерий отношения вероятностей δ^* ***имеет минимальные средние числа наблюдений:***

$$n_0(\delta^*) \le n_0(\delta'), \quad n_1(\delta^*) \le n_1(\delta'). \qquad (10)$$

Заметим, что минимальность достигается сразу по двум характеристикам (см. [8, Леман. *Проверка статистических гипотез*]).

Б. Связь порогов с вероятностями ошибок.

Итак, $\alpha = \alpha(A, B)$, $\beta = \beta(A, B)$

Легко показать справедливость неравенств, связывающих пороги с вероятностями ошибок:

$$A \le \frac{1-\beta}{\alpha}, \qquad B \ge \frac{\beta}{1-\alpha}. \tag{11}$$

Действительно, обозначим через R_n множество тех последовательностей $x_1, x_2...x_n$ длины n, для которых проверочная процедура заканчивается на n-м шаге принятием H_1: (т.е., на всех предыдущих шагах L_n между порогами, а на n-м $\ge A$)

$$R_n = \left\{(x_1,...,x_n) \equiv x^n : B < \frac{p_{1k}(x^k)}{p_{0k}(x^k)} < A,\ k = 1,...,n-1,\ \frac{p_{1n}(x^n)}{p_{0n}(x^n)} \ge A\right\}.$$

В R_n справедливо: $\frac{p_{1n}(x^n)}{p_{0n}(x^n)} \ge A$

Оценим α:

$$\alpha = \alpha(A, B) = P\{\text{пр. } H_1 / H_0\} = \sum_{n=1}^{\infty} P\{\text{пр. } H_1 \text{ на } n\text{-м шаге} \mid H_0\} =$$

$$= \sum_{n=1}^{\infty}\int_{R_n} p_{0n}(x^n)dx^n \le \frac{1}{A}\sum_{n=1}^{\infty}\int_{R_n} p_{1n}(x^n)dx^n = \frac{1}{A}\sum_{n=1}^{\infty} P\{\text{пр. } H_1 \text{ на } n\text{-м шаге} \mid H_1\} =$$

$$= \frac{1}{A} P\{\text{пр. } H_1 \mid H_1\} = \frac{1-\beta}{A},$$

что дает первое из неравенств в (11).

Аналогично показывается справедливость второго неравенства введением S_n : множества тех последовательностей $x_1, x_2... x_n$ длины n, для которых проверочная процедура заканчивается на n-м шаге принятием H_0:

$$S_n = \left\{(x_1, x_2...x_n) \equiv x^n : B < \frac{p_{1k}(x^k)}{p_{0k}(x^k)} < A,\ k = 1,2...n-1,\ \frac{p_{1n}(x^n)}{p_{0n}(x^n)} \le B\right\}.$$

В. Выбор порогов. Вместо неизвестных значений A и B возьмем их приближенные значения A' и B':

$$A \approx A' = \frac{1-\beta}{\alpha}, \qquad B \approx B' = \frac{\beta}{1-\alpha}. \tag{12}$$

При таком выборе порогов вероятности ошибок α' и β' не равны α и β. Но для них справедливы неравенства (11):

$$A' \le \frac{1-\beta'}{\alpha'}, \qquad B' \ge \frac{\beta'}{1-\alpha'}$$

$$\frac{1-\beta}{\alpha} = A' \le \frac{1-\beta'}{\alpha'}, \qquad \frac{\beta}{1-\alpha} = B' \ge \frac{\beta'}{1-\alpha'} \tag{13}$$

Из них получаем связь между α', β' и α, β:

$$\alpha' \le \frac{1-\beta'}{1-\beta}\alpha \le \frac{\alpha}{1-\beta}, \qquad \beta' \le \frac{1-\alpha'}{1-\alpha}\beta \le \frac{\beta}{1-\alpha}.$$

Поскольку знаменатели справа незначительно меньше 1, значения α' и β', если и превышают заданные значения α и β, то весьма незначительно. Более того, если неравенства (13), записанные в виде

$$\alpha'(1-\beta) \le \alpha(1-\beta'),$$

$$\beta'(1-\alpha) \le \beta(1-\alpha')$$

сложить, то получим, что сумма новых не дольше суммы старых:

$$\alpha'+\beta' \le \alpha+\beta.$$

На самом деле, как показывает более детальный анализ [8], не только суммы, но и сами слагаемые α' и β' меньше заданных вероятностей α и β.

Г. Тождество Вальда. Пусть $\zeta_1, \zeta_2 \ldots \zeta_{k\text{-}1}$ — последовательность независимых, одинаково распределенных случайных величин, и пусть ν — момент остановки наблюдений, причем случайное событие $\{\nu \ge k\}$ (быть или не быть следующему k-му наблюдению) определяется значениями $\zeta_1, \zeta_2 \ldots \zeta_{k\text{-}1}$, т.е.

$$\{\nu \ge k\} = f(\zeta_1, \zeta_2 \ldots \zeta_{k\text{-}1}).$$

Справедливо следующее утверждение: среднее значение суммы в момент остановки равно

$$\mathrm{M}\sum_{k=1}^{\nu}\zeta_k = \mathrm{M}\nu\mathrm{M}\zeta. \qquad (14)$$

Доказательство.

Введем случайные величины

$$\eta_k = \begin{cases} 1, \text{ если } \nu \ge k, \\ 0, \text{ если } \nu < k. \end{cases} = g(\zeta_1, \zeta_2, \ldots, \zeta_{k\text{-}1}).$$

Примечание [MA2]: Верно ли записано?

Тогда

$$\sum_{k=1}^{\nu}\zeta_k = \sum_{k=1}^{\infty}\zeta_k\eta_k \quad \text{и}$$

$$\mathrm{M}\sum_{k=1}^{\nu}\zeta_k = \sum_{k=1}^{\infty}\mathrm{M}(\zeta_k\eta_k) = \sum_{k=1}^{\infty}\mathrm{M}\zeta_k\mathrm{M}\eta_k = \mathrm{M}\zeta\cdot\mathrm{M}\sum_{k=1}^{\infty}\eta_k = \mathrm{M}\nu\cdot\mathrm{M}\zeta.$$