

Гипотезы и критерии (продолжение)

В идеале хотелось бы такой критерий, который минимизирует вероятности совершения обеих ошибок. Однако, как правило, такая задача некорректно — при уменьшении вероятности одной ошибки увеличивается вероятность другой. Существует несколько способов постановки задачи выбора критерия на основе минимизации вероятностей ошибок, наиболее распространенный из них выглядит следующим образом. Один тип ошибки объявляется более опасным, и вероятность совершить этот тип ошибки ограничивается сверху некоторым малым числом α , а вероятность совершения ошибки другого типа минимизируется по остаточному принципу. Как правило, более опасной считается ошибка первого рода, соответственно получаем задачу минимизации

$$\begin{cases} P(I_S) \leq \alpha \\ P(II_S) \rightarrow \min_S \end{cases}.$$

Число α называется *уровнем значимости* критерия S . Иначе говоря, величина $\alpha \in (0, 1)$ — уровень значимости критерия S , если $P(I_S) \leq \alpha$. На практике обычно выбирается $\alpha = 0.05$, то есть ошибка первого рода может произойти не более чем в 5% случаев. Вообще говоря, равенство может не достигаться. Величина вероятности ошибки первого рода $P(I_S)$ называется также *реальным уровнем значимости*. Во многих случаях на практике реальный уровень значимости вычисляется с помощью семплирования.

Как правило, основная гипотеза имеет достаточно простой вид, в то время как альтернативная гипотеза устроена существенно сложнее. Например, $H_0: \theta = \theta_0$ vs. $H_1: \theta > \theta_0$ или $H_0: X_i$ имеет нормальное распределение vs. $H_1: \text{распределение } X_i \text{ отличается от нормального}$. При разных распределениях из альтернативной гипотезы поведение различных критериев может различаться. Для сравнения критериев на распределениях, соответствующих альтернативной гипотезе, вводится функция *мощности*, которая определяется как $\beta_S(P) = P(X \in S)$, где $P \in \mathcal{P}_1$. Конечно, можно рассматривать функцию $\beta_S(P)$ и для распределений из $\mathcal{P}_0$, причем в таком случае ее значения будут не превосходить уровня значимости критерия. В параметрическом случае функцию мощности можно рассматривать как функцию от параметра, записывая $\beta_S(\theta)$.

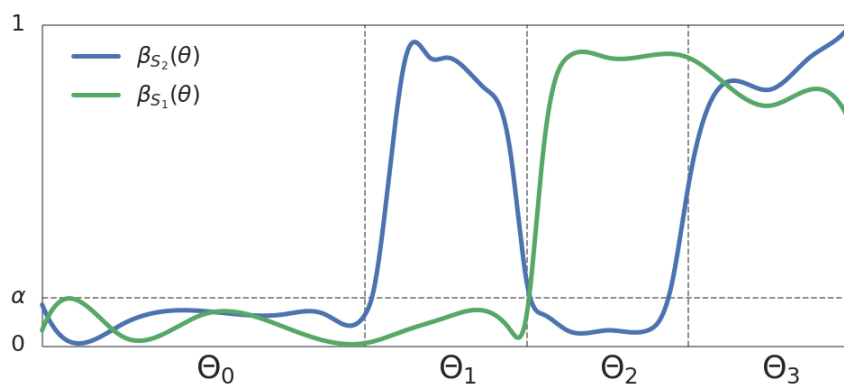


Рис. 1: Пример функций мощности для двух критериев.

На рисунке 1 приведен пример функций мощности для двух критериев S_1 и S_2 для проверки гипотезы $H_0: \theta \in \Theta_0$. На множестве Θ_0 функции мощности обоих критериев не превосходят установленного уровня значимости α , причем критерий S_1 достигает его. Максимальное значение величины $\beta_S(\theta)$ при $\theta \in \Theta_0$ является реальным уровнем значимости. Если в качестве альтернативной гипотезы рассматривается $H_1: \theta \in \Theta_1$, то критерий S_2 можно считать достаточно

мощным для проверки H_0 vs. H_1 , поскольку функция мощности $\beta_{S_2}(\theta)$ принимает достаточно большие значения при $\theta \in \Theta_1$. В то же время вероятность ошибки второго рода тоже достаточно большая, поскольку для ее определения рассматривается наихудший случай при $\theta \in \Theta_1$. Критерий S_1 не стоит использовать для проверки таких гипотез. Если же в качестве альтернативной гипотезы мы рассмотрим $H_1: \theta \in \Theta_2$, то справедлива полностью обратная ситуация. Наконец, в случае $H_1: \theta \in \Theta_3$ оба критерия являются достаточно хорошими, причем вероятность ошибки второго рода не сильно велика.

Пример. Пусть $X \sim \text{Exp}(\theta)$ — выборка из одного наблюдения. Построим критерий уровня значимости α для проверки $H_0: \theta = \theta_0$ vs. $H_1: \theta > \theta_0$. Заметим, что $E_\theta X_1 = 1/\theta$, в следствии чего при увеличении значения θ стоит ожидать меньшие значения величины X . Таким образом, логично определить критерий как $S = \{x \in \mathcal{X} \mid x < c\}$, где критический уровень c подберем из условия на ошибку первого рода. Поскольку основная гипотеза простая, данное условие принимает простой вид $P_{\theta_0}(X < c) \leq \alpha$. Критическая область критерия S изображена на рисунке 2. Пунктирной линией изображена плотность, соответствующая альтернативной гипотезе.

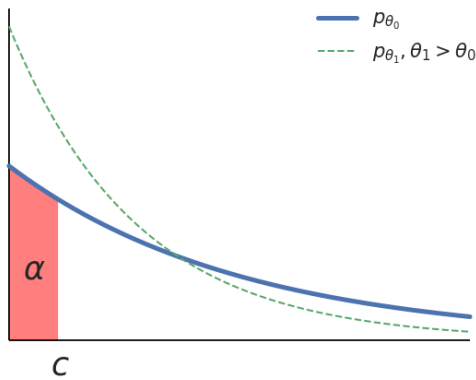


Рис. 2: Плотности распределения

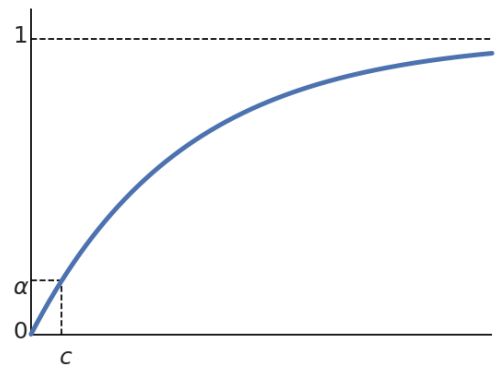


Рис. 3: Мощность критерия

Найдем c в явном виде. Функция распределения имеет вид $P_{\theta_0}(X < c) = 1 - e^{-\theta_0 c}$, откуда мы можем получить $c \leq -\log(1 - \alpha)/\theta_0$. Мощность критерия равна $\beta_S(\theta) = P_\theta(X < c) = 1 - e^{-\theta c}$ и возрастает при увеличении c , соответственно, логично было бы взять $c_0 = -\log(1 - \alpha)/\theta_0$. Отсюда получаем критерий $S = \{x \in \mathcal{X} \mid x < -\log(1 - \alpha)/\theta_0\}$, соответствующая функция мощности равна $\beta_S(\theta) = 1 - e^{\theta \log(1 - \alpha)/\theta_0} = 1 - (1 - \alpha)^{\theta/\theta_0}$, $\theta > \theta_0$, ее график изображен на рисунке 3. Заметим, что $\beta_S(\theta_0) = \alpha$ — уровень значимости, а при увеличении θ к бесконечности функция мощности возрастает к 1. Это означает, что чем сильнее истинное значение θ отличается от предполагаемого основной гипотезой, тем мощнее критерий.

Предположим, что $\theta_0 = 1$, то есть рассматриваются гипотезы $H_0: \theta = 1$ vs. $H_1: \theta > 1$. Возьмем также уровень значимости равным $\alpha = 0.05$. Подставив все величины, получаем значение критического уровня $c_0 = -\log 0.95 \approx 0.051$. В результате

1. Если $X \leq 0.051$, то гипотеза H_0 отклоняется в пользу альтернативы H_1 . Результат является статистически значимым, а факт того, что $X \leq 0.051$ является (статистическим) доказательством против гипотезы H_0 ;
2. Если $X > 0.051$, то гипотеза H_0 не отклоняется, а результат является статистически не значимым. Иначе говоря, с точки зрения статистики результат отсутствует вообще.