Глава 9

Проверка гипотез. Лемма Неймана-Пирсона

9.1 Базовая часть

9.1.1 Постановка задачи

Определение 1. *Гипотезой* мы будем называть утверждение H_0 вида $\theta \in \Theta_0$, где $\Theta_0 \subset \Theta$, альтернативой гипотезе H_0 — утверждение H_1 вида $\theta \in \Theta_1$, $\Theta_1 \subset \Theta$, $\Theta_1 \cap \Theta_0 = \emptyset$.

В случае, когда Θ_0 (Θ_1) содержат только одну точку θ, гипотезу (альтернативу) называют *простой*, в противном случае — *сложной*.

На основе данных мы будем соглашаться (принимать) или не соглашаться (отвергать) с основной гипотезой H_0 в пользу альтернативы.

При этом у нас возможны два вида ошибок. Мы можем отвергнуть верную гипотезу H_0 (такая ошибка называется ошибкой I рода) или не отвергнуть неверную гипотезу H_0 (ошибка II рода).

Пример 1. Представим себе, что у нас есть гипотеза о том, что параметр θ для $X_i \sim \mathcal{N}(\theta, 1)$ равен 0, и альтернатива — $\theta = 1$. Рассмотрим статистику $\overline{X}$. Как мы знаем, она распределена $\mathcal{N}(\theta, 1/n)$. Тогда если $\theta = 0$, то

$$\mathbf{P}_0(\sqrt{n}(\overline{X} - \theta) < z_{1-\alpha}) = 1 - \alpha$$

для любого $\alpha \in [0, 1]$, где z_p — p-квантиль распределения $\mathcal{N}(0, 1)$. Пусть статистика $\overline{X}$ не попала в луч $(-\infty, z_{1-\alpha} n^{-1/2})$. Значит, наше предположение $\theta = 0$ выглядит достаточно неправдоподобным при малых α, так как для выборки из $\mathcal{N}(0, 1)$ вероятность непопадания $\overline{X}$ в наш луч равна α. Будем отвергать наше предположение, если $\overline{X}$ не попала в $(-\infty, z_{1-\alpha} n^{-1/2})$ и принимать (или, скорее, не отвергать) в противном случае. При этом вероятность того, что при настоящем значении параметра 0 мы отвергли H_0 (т.е. вероятность ошибки первого рода) будет α, а вероятность того, что параметр был 1, а мы приняли H_0 (ошибки второго рода) — $\Phi(z_{1-\alpha} - n^{1/2})$.

В общем случае выбор критерия равносилен заданию в выборочном пространстве подмножества D, такого, что при попадании выборки $X_1, \ldots, X_n$ в D мы отвергаем гипотезу H_0 в пользу альтернативы H_1, а в противном случае — принимаем.

Определение 2. Такое множество D называют *критическим. Критической функцией* $\psi(x_1, \ldots, x_n)$ называют в этом случае индикаторную функцию I_D.

Критическая функция определяет, какую гипотезу мы принимаем: если она равна 0, то H_0, если 1, то H_1.

При этом $\mathbf{P}_\theta(D)$ — вероятность отвержения гипотезы. Если $\theta \in \Theta_0$, то гипотеза верна, и, значит, эта вероятность будет вероятностью ошибки первого рода, соответственно, вероятность $\mathbf{P}_\theta(\overline{D})$ при $\theta \in \Theta_1$ — вероятностью ошибки второго рода.

Определение 3. Функцию $\beta(\theta) = \mathbf{P}_\theta(D)$ при $\theta \in \Theta_1$ называют *мощностью критерия.*

Определение 4. Критерий (а вернее последовательность критериев) называют *состоятельным*, если $\beta(\theta) \to 1$ при $n \to \infty$ для всех $\theta \in \Theta_1$.

Состоятельность означает, что с увеличением размера выборки мы будем все реже допускать ошибки второго рода при данном уровне значимости.

Как мы видим на примере 1, уменьшая вероятность ошибки первого рода, мы увеличиваем вероятность ошибки второго рода, поэтому выбрать критерий, минимизирующий обе ошибки, не представляется возможным. В случае "сравнимых" цен ошибок, скажем, в ситуации, когда ошибка первого рода обходится нам в некоторую сумму p_1, а ошибка второго рода — в сумму p_2, мы можем минимизировать $p_1\alpha_1 + p_2\alpha_2$, где α_1, α_2 — ошибки 1, 2 рода.

Зачастую наши ошибки принципиально неравнозначны:

Пример 2. Предположим, что численные показатели в анализах пациентов имеют некоторое распределение $F_0(x)$ для больного и $F_1(x)$ для здорового человека. Семейство $F_0(x), F_1(x)$ представим в виде $F_\theta(x), \theta = 0, 1$. Если мы рассматриваем гипотезу H_0 о том, что $\theta = 0$, т.е. пациент болен, с альтернативой $H_1 : \theta = 1$, то у нас существует 4 возможных варианта. Пациент может быть больным и мы признаем его больным, пациент может быть здоров и мы признаем его здоровым, пациент может быть больным, но мы признаем его здоровым и, наконец, пациент может быть здоровым, а мы признаем его больным. Первые два случая нас целиком и полностью устраивают, в третьем мы случае мы отвергаем гипотезу H_0, хотя она верна, т.е. допускаем ошибку I рода, в четвертом — не отвергаем H_0, хотя она ошибочна, т.е. допускаем ошибку II рода. При этом мы бы хотели, чтобы больные пациенты, которых мы признали здоровыми, были достаточно редки, т.е. вероятность ошибки первого рода не превышала некоторый заданный уровень α (его называют уровнем значимости). Здоровые люди, признанные больными, это тоже неприятно, но значительно менее существенно.

Наша задача — построить критерий (т.е. критическое множество) с уровнем значимости не более α, т.е. такой что $\mathbf{P}_\theta(D) \leq \alpha$ при всех $\theta \in \Theta_0$. Здесь $\mathbf{P}_\theta(D)$ — вероятность отвержения гипотезы, если $\theta \in \Theta_0$, то гипотеза верна и эта вероятность будет вероятностью ошибки первого рода, а соответственно, вероятность $\mathbf{P}_\theta(\overline{D})$ при $\theta \in \Theta_1$ — вероятностью ошибки второго рода. Мы могли бы признавать всех больными и ошибок I рода не было бы вообще, однако тогда вероятность ошибки второго рода будет равна 1.

Итак, мы будем предполагать, что нам задан критерий с заданным уровнем значимости α, то есть такой, что $\mathbf{P}_\theta(D) \leq \alpha$ при всех $\theta \in \Theta_0$.

Определение 5. Критерий называют *несмещенным*, если $\beta(\theta) \geq \alpha$ для всех $\theta \in \Theta_1$.

Несмещенность означает, что вероятность отвержения гипотезы H_0, когда она верна, не больше, чем когда она неверна:

$$\mathbf{P}_{H_0}(\vec{X} \in D) = \alpha \leq \mathbf{P}_{H_1}(\vec{X} \in D).$$

9.1.2 Наиболее мощный критерий

Вполне естественно среди критериев с уровнем значимости α выбирать те, у которых минимальна ошибка II рода.

Определение 6. Если существует критерий с критическим множеством D, такой что его уровень значимости не превосходит α и для любого другого критерия D^* с уровнем значимости не более α при всех $\theta \in \Theta_1$ справедливо соотношение

$$\mathbf{P}_\theta(D) > \mathbf{P}_\theta(D^*),\ \theta \in \Theta_1$$

то критерий D называют равномерно наиболее мощным (РНМ) критерием для проверки H_0 против H_1.

Если H_1 простая (т.е. одноточечная), то критерий называют наиболее мощным (НМ).

Оказывается, что в случае простой гипотезы и простой альтернативы существует НМ критерий, называемый критерием Неймана-Пирсона

Теорема 1 (Лемма Неймана-Пирсона)**.** *Пусть $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$, распределения F_{θ_0}, F_{θ_1} имеют положительные плотности. Тогда если среди критериев с критическим множеством*

$$D = \{x_1, \ldots, x_n : L(\vec{x}; \theta_1) > CL(\vec{x}; \theta_0)\}$$

найдется такой критерий, что $\mathbf{P}_{\theta_0}(D) = \alpha$, то он будет НМ на уровне значимости α.

Здесь $L(\vec{x}; \theta)$ — функция правдоподобия.

Пример 3. Для примера 1 критическое множество критерия Неймана-Пирсона имеет вид

$$D = \left\{ x_1, \ldots, x_n : \frac{(2\pi)^{-n/2} \exp\left(-\sum\limits_{i=1}^{n}(x_i - 1)^2/2\right)}{(2\pi)^{-n/2} \exp\left(-\sum\limits_{i=1}^{n} x_i^2/2\right)} > C \right\},$$

т.е. $D_C = \{\exp(\sum_{i=1}^n x_i - n/2) > C\}$. Таким образом, критические множества в нашем критерии имеют вид $D_{\tilde{C}} = \{\sum_{i=1}^n x_i > \tilde{C}\}$. Какой-либо привязки к исходному параметру C у нас нет, поэтому нам удобнее работать с параметром $\tilde{C}$. Найдем $\tilde{C}_\alpha$ из условия $\mathbf{P}_{\theta_0}(D_{\tilde{C}_\alpha}) = \alpha$. Величина $X_1 + \cdots + X_n$ при $\theta = 0$ имеет распределение $\mathcal{N}(0, n)$, т.е.

$$\mathbf{P}_{\theta_0}\left(\sum_{i=1}^{n} X_i > \tilde{C}\right) = \mathbf{P}\left(Z > \frac{\tilde{C}}{\sqrt{n}}\right) = 1 - \Phi\left(\frac{\tilde{C}}{\sqrt{n}}\right) = \alpha,$$

где $Z \sim \mathcal{N}(0, 1)$, а значит наш критерий имеет критическое множество

$$\{x_1, \ldots, x_n : x_1 + \cdots + x_n > \sqrt{n} z_{1-\alpha}\},$$

что совпадает с примером 1.

В случае, когда условия теоремы не выполнены, зачастую бывает полезным следующий подход.

Определение 7. Назовем критерий рандомизированным, если его критическая функция может принимать значения между 0 до 1, которые можно интерпретировать как вероятность отвержения гипотезы. Если, скажем, на нашей выборке критическая функция приняла значение $1/3$, то мы должны подбросить монету с вероятностью успеха $1/3$ и, в случае удачи, выбрать H_1, иначе — H_0.

В этом случае вероятность ошибки первого рода есть $\mathbf{E}_\theta \psi(X_1, \ldots, X_n)$, $\theta \in \Theta_0$, второго рода — $1 - \mathbf{E}_\theta \psi(X_1, \ldots, X_n)$, $\theta \in \Theta_1$.
Рандомизация позволяет нам добиться нужной ошибки первого рода для случаев, когда вероятность попадания в критическое множество не изменяется непрерывно, например, для дискретных выборок. Оказывается, что для рандомизированных семейств справедлив аналог леммы Неймана-Пирсона для сложной альтернативы:

Теорема 2. *Пусть $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1 > \theta_0$, распределения F_{θ_0}, F_{θ_1} имеют плотности или дискретны. Тогда среди критериев с критической функцией*

$$\psi(x_1, \ldots, x_n) = \begin{cases} 1, & L(\vec{x}; \theta_1) > CL(\vec{x}; \theta_0), \\ \gamma \leq 1, & L(\vec{x}; \theta_1) = CL(\vec{x}; \theta_0), \\ 0, & L(\vec{x}; \theta_1) < CL(\vec{x}; \theta_0), \end{cases}$$

найдется критерий, такой что $\mathbf{E}_{\theta_0} \psi(X_1, \ldots, X_n) = \alpha$, и он будет НМ на уровне значимости α.

Доказательство. Доказательство того, что такой критерий найдется достаточно просто и по факту приведено в примере ниже. Покажем, что он НМ. Заметим, что для любого ψ^* верно соотношение

$$\int_{\mathbb{R}^n} (\psi(\vec{x}) - \psi^*(\vec{x}))(L(\vec{x};\theta_1) - CL(\vec{x};\theta_0))d\vec{x} \geq 0,$$

поскольку там, где $L(\vec{x};\theta_1) = CL(\vec{x};\theta_0)$ подынтегральная функция 0, там где $L(\vec{x};\theta_1) > CL(\vec{x};\theta_0)$ второй сомножитель положителен, а первый неотрицателен, поскольку $\psi(\vec{x}) = 1$, а $\psi^*(\vec{x}) \leq 1$. Наконец, там где $L(\vec{x};\theta_1) < CL(\vec{x};\theta_0)$ второй множитель отрицателен, а первый множитель неположителен, поскольку $\psi(\vec{x}) = 0$.

Следовательно,

$$\mathbf{E}_{\theta_1}\psi(\vec{X}) - \mathbf{E}_{\theta_1}\psi^*(\vec{X}) - C\mathbf{E}_{\theta_0}\psi(\vec{X}) + C\mathbf{E}_{\theta_0}\psi^*(\vec{X}).$$

При этом в правой части стоят ошибки I рода, первая из которых равна α, а вторая его не больше. Значит,

$$\mathbf{E}_{\theta_1}\psi(\vec{X}) \geq \mathbf{E}_{\theta_1}\psi^*(\vec{X}),$$

причем равенство возможно только в двух случаях:

1. если $\mathbf{E}_{\theta_0}\psi^*(\vec{X}) = \alpha$ и $\psi(\vec{x}) = \psi^*(\vec{x})$ при п.в. $\vec{x}$, при которых $L(\vec{x};\theta_1) \neq CL(\vec{x};\theta_0)$;
2. если $C = 0$, $\psi(\vec{x}) = \psi^*(\vec{x})$ при $\vec{x} : L(\vec{x};\theta_1) > 0$

Таким образом, НМ критерий единственен с точностью до указанных выше изменений. □

Пример 4. Пусть $X_i \sim Bernoulli(\theta)$. Тогда построим критерий для проверки гипотезы $H_0 : \theta = \theta_0$, $H_1 : \theta = \theta_1$, $\theta_0 < \theta_1$.

$$\left\{x : \frac{\theta_1^{x_1+\dots+x_n}(1-\theta_1)^{n-x_1+\dots+x_n}}{\theta_0^{x_1+\dots+x_n}(1-\theta_0)^{n-x_1+\dots+x_n}} > C\right\} = \left\{x : \left(\frac{(1-\theta_1)^{-1}-1}{(1-\theta_0)^{-1}-1}\right)^{x_1+\dots+x_n} > \tilde{C}\right\}.$$

Т.к. величина под знаком степени больше одного, то это множество представимо в виде $\{x_1 + \dots + x_n > \widehat{C}\}$, откуда критическая функция имеет вид

$$\begin{cases} 1, & x_1 + \dots + x_n > \widehat{C}, \\ \gamma, & x_1 + \dots + x_n = \widehat{C}, \\ 0, & x_1 + \dots + x_n < \widehat{C}, \end{cases}$$

Остается найти $\widehat{C}$ (очевидно, можно считать его целым). Величина $X_1 + \dots + X_n$ при $\theta = \theta_0$ имеет биномиальное распределение (n, θ), откуда

$$\mathbf{E}_{\theta_0}\psi(X_1, \dots, X_n) = \sum_{i=\tilde{C}+1}^{n} C_n^i \theta_0^i (1-\theta_0)^{n-i} + \gamma C_n^{\tilde{C}} \theta_0^{\tilde{C}} (1-\theta_0)^{n-\tilde{C}} = \alpha.$$

Поскольку $\gamma < 1$, то $\tilde{C}$ находится из соотношения

$$\sum_{i=\tilde{C}+1}^{n} C_n^i \theta_0^i (1-\theta_0)^{n-i} \leq \alpha < \sum_{i=\tilde{C}}^{n} C_n^i \theta_0^i (1-\theta_0)^{n-i},$$

которое можно представить в виде

$$\sum_{i=0}^{\tilde{C}} C_n^i \theta_0^i (1-\theta_0)^{n-i} \geq 1 - \alpha > \sum_{i=0}^{\tilde{C}-1} C_n^i \theta_0^i (1-\theta_0)^{n-i},$$

а γ, соответственно, равно

$$1-\left(\frac{1-\alpha-\sum\limits_{i=0}^{\tilde{C}-1} C_n^i\theta_0^i(1-\theta_0)^{n-i}}{C_n^{\tilde{C}}\theta_0^{\tilde{C}}(1-\theta_0)^{n-\tilde{C}}}\right).$$

При этом вероятность ошибки второго рода есть

$$1-\mathbf{E}_{\theta_1}\psi(X_1,\ldots,X_n)=\sum_{i=0}^{\tilde{C}-1} C_n^i\theta_1^i(1-\theta_1)^{n-i}+(1-\gamma)C_n^{\tilde{C}}\theta_1^{\tilde{C}}(1-\theta_1)^{n-\tilde{C}}.$$

9.2 Факультатив

9.2.1 Простая гипотеза и доверительное оценивание

Будем рассматривать простую гипотезу $H_0:\theta=\theta_0$ с различными альтернативами. Одним из вариантов построения критерия в этом случае является использование доверительного интервала. Действительно, если $\widehat{\theta}_1(X_1,\ldots,X_n)$, $\widehat{\theta}_2(X_1,\ldots,X_n)$ — доверительный интервал уровня $1-\alpha$, то

$$\mathbf{P}_{\theta_0}(\theta_0\notin(\widehat{\theta}_1(X_1,\ldots,X_n),\widehat{\theta}_2(X_1,\ldots,X_n)))=\alpha,$$

то есть если в качестве критического множества D выбрать множество

$$D=\{x_1,\ldots,x_n:\widehat{\theta}_1(X_1,\ldots,X_n)\geq\theta_0\}\cup\{x_1,\ldots,x_n:\widehat{\theta}_2(X_1,\ldots,X_n)\leq\theta_0\},$$

то мы получим критерий с вероятностью ошибки первого рода α.

Пример 5. Для $X_i\sim\mathcal{N}(\theta,1)$ мы можем построить для θ доверительные интервалы вида

$$\left(\overline{X}-z_{1-\alpha+\beta}\frac{1}{\sqrt{n}},\overline{X}-z_{\beta}\frac{1}{\sqrt{n}}\right).$$

Соответственно для проверки гипотезы $H_0:\theta=0$ мы можем использовать критическое множество

$$D=\{\vec{x}:0\in(\overline{x}-n^{-1/2}z_{1-\alpha+\beta},\overline{x}-n^{-1/2}z_{\beta})\}.$$

Какой именно интервал лучше выбрать — зависит от альтернативы. По существу мы делим возможную вероятность ошибки α на две части β и $\alpha-\beta$ и распределяем ее между двумя возможностями $\overline{x}$ мало и $\overline{x}$ велико. При $H_1:\theta\neq\theta_0$ обе эти возможности для нас существенны и мы берем $\beta=\alpha/2$, критическое множество при этом примет вид

$$\{\overline{x}<z_{\alpha/2}\sqrt{n}\}\cup\{\overline{x}>z_{1-\alpha/2}\sqrt{n}\}.$$

При $H_1:\theta>\theta_0$ нас беспокоят большие значения $\overline{x}$, то есть $\beta=0$ и критическое множество будет иметь вид $D=\overline{x}>n^{-1/2}z_{1-\alpha}$. Аналогично при $H_1:\theta<\theta_0$ нужно действовать с точностью до наоборот.

9.2.2 Равномерно наиболее мощные критерии

Заметим, что в критериях примеров 1 и 4 критические функции никак не зависели от θ_1, поэтому данные критерии являлись НМ для всех $H_1:\theta_1>\theta_0$, а значит РНМ для $H_1':\ \theta>\theta_0$.
И наоборот, РНМ критерий для, скажем, $H_0:\theta=\theta_0$ против $H_1:\theta>\theta_0$ найдется только тогда, когда для каждой $H_1':\ \theta=\theta_1>\theta_0$ критерий будет одним и тем же.
В ряде моделей удается воплотить это в приведенную ниже теорему.

Определение 8. Говорят, что для дискретных или абсолютно-непрерывных ф.р. $F_\theta(x)$ выполнено свойство "монотонности отношения правдоподобий", если все плотности (распределения) имеют одинаковый носитель и существует статистика $T(X_1, \dots, X_n)$, такая, что при любых $\theta_0 < \theta_1$

$$\frac{f_{\theta_1}(x_1) \cdot \dots \cdot f_{\theta_1}(x_n)}{f_{\theta_0}(x_1) \cdot \dots \cdot f_{\theta_0}(x_n)}$$

есть монотонная функция от $T(x_1, \dots, x_n)$.

Теорема 3. *Пусть $F_\theta(x)$ — семейство распределений с монотонным отношением правдоподобий, причем упомянутая в определении функция монотонно возрастает по T. Тогда РНМ критерий уровня α для проверки гипотезы $H_0 : \theta \leq \theta_0$ с альтернативой $H_1 : \theta > \theta_0$ задается критической функцией вида*

$$\psi(x_1, \dots, x_n) = \begin{cases} 1, & T(x_1, \dots, x_n) > C, \\ \gamma, & T(x_1, \dots, x_n) = C, \\ 0, & T(x_1, \dots, x_n) < C, \end{cases}$$

где C, γ определяются соотношением $\mathbf{E}_{\theta_0}\psi(X_1, \dots, X_n) = \alpha$.

Аналогичные результаты для убывающих по T отношений правдоподобия или областей вида $\theta > \theta_0$ получаются заменой знака в неравенствах в определении $\psi(x)$.

Отсюда, скажем, для нормальных величин $\mathcal{N}(0, \theta)$ тот же критерий примера 1 является РНМ для проверки гипотезы $H_0 : \theta = \theta_0$ с альтернативой $\theta > \theta_0$.

В некоторых задачах ниже требуется построить РНМ критерий, но такого критерия не существует. В таком случае это и является ответом на поставленный вопрос.