Глава 15

Проверка на хипотези

Определение 15.1 *Хипотеза наричаме произволно твърдение за параметър на разпределението на генералната съвкупност.*

Определение 15.2 *Двете допълващи се хипотези в проверката на хипотези се наричат нулева H_0 и алтернативна H_1 или:*
$H_0 : \theta \in \Theta_0$
$H_1 : \theta \in \bar{\Theta}_0$.

Пример 15.1 *θ може да бъде например средната промяна в кръвното налягане на пациенти след приемането на дадено лекарство и тогава хипотезите ще имат вида:*
$H_0 : \theta = 0$
$H_1 : \theta \neq 0$,
или пропорцията на дефектните детайли в дадено производство, където предполагаемата допустима стойност на тази пропорция е θ_0:
$H_0 : \theta \geq \theta_0$
$H_1 : \theta < \theta_0$.

Определение 15.3 *Процедурата (тест, критерий) по проверка на хипотези включва:*
1. Определяне за кои стойности на извадката се взема решение да се приеме нулевата хипотеза.
2. Определяне за кои стойности да се отхвърли нулевата и да се приеме алтернативната хипотеза.

Това означава, че трябва да се определи т.нар. *област на отхвърляне (критична област)*, попадането в която води до отхвърляне на H_0, и съответно *област на приемане*, която я допълва. Обикновено не се разглеждат самите стойности на извадката, а дадена статистика от нея $W(X_1, X_2, \ldots, X_n) = W(\mathbf{X})$ и критичната област се определя за нея. Например $W(\mathbf{X}) = \bar{X}$ и H_0 се отхвърля, ако $\bar{X} \geq 3$.

Определение 15.4 *Тестова статистика на отношение на правдоподобия наричаме:*

$$\lambda(\mathbf{x}) = \frac{sup_{\Theta_0} L(\theta|\mathbf{x})}{sup_{\Theta} L(\theta|\mathbf{x})}.$$

Критерий (тест) на отношение на правдоподобия наричаме всеки тест за проверка на хипотези, който има критична област от вида: $\{\mathbf{x} : \lambda(\mathbf{x}) \leq c, \quad 0 \leq c \leq 1\}$.

Пример 15.2 *Нека* $X_1, X_2, \ldots, X_n \sim N(\mu, 1)$ *и имаме следните хипотези:*
$H_0 : \mu = \mu_0$
$H_1 : \mu \neq \mu_0$.
МПО на μ *е* $\bar{X}$. *Тогава*

$$\lambda(\mathbf{x}) = \frac{(2\pi)^{-\frac{n}{2}} e^{-\sum\limits_{i=1}^{n} \frac{(x_i-\mu_0)^2}{2}}}{(2\pi)^{-\frac{n}{2}} e^{-\sum\limits_{i=1}^{n} \frac{(x_i-\bar{x})^2}{2}}} = \exp\left[-\sum_{i=1}^{n} \frac{(x_i-\mu_0)^2}{2} + \sum_{i=1}^{n} \frac{(x_i-\bar{x})^2}{2} \right]$$

$$= \exp[-\frac{n}{2}(\bar{x}-\mu_0)^2]$$

от $\sum\limits_{i=1}^{n}(x_i-\mu_0)^2 = \sum\limits_{i=1}^{n}(x_i-\bar{x}+\bar{x}-\mu_0)^2 = \sum\limits_{i=1}^{n}(x_i-\bar{x})^2 + \sum\limits_{i=1}^{n}(\bar{x}-\mu_0)^2 + 2\sum\limits_{i=1}^{n}(x_i-\bar{x})(\bar{x}-\mu_0) = \sum\limits_{i=1}^{n}(x_i-\bar{x})^2 + n(\bar{x}-\mu_0)^2 + 0$. *Тогава критерият с отношение на правдоподобия ще има вида:*

$$\left\{ \mathbf{x} : |\bar{x}-\mu_0| \geq \sqrt{-\frac{2\ln c}{n}} \right\},$$

където $0 \leq c \leq 1$ *и следователно* $0 \leq \sqrt{-\frac{2\ln c}{n}} \leq \infty$.
Тези тестове ще отхвърлят $H_0 : \mu = \mu_0$, *ако извадковото средно* $\bar{x}$ *се отличава от* μ_0 *с дадена разлика.*

Пример 15.3 *Нека*

$$X_1, X_2, \ldots, X_n \sim Ex : f(x|\theta) = \begin{cases} e^{-(x-\theta)}, & x \geq \theta \\ 0, & x < \theta \end{cases}$$

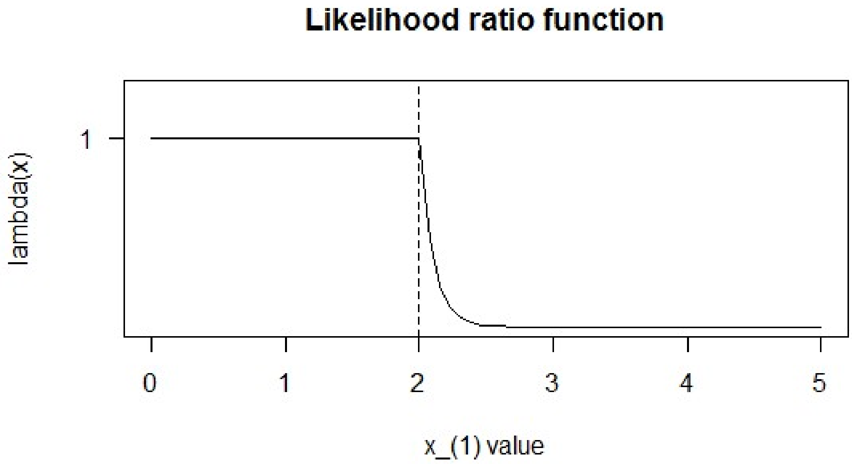
Тогава

$$L(\theta|\mathbf{x}) = \begin{cases} e^{-\sum\limits_{i=1}^{n} x_i + n\theta}, & \theta \leq x_{(1)} \\ 0, & \theta > x_{(1)} \end{cases}$$

където $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ *са наредените статистики, т.е* $x_{(1)} = \min(x_1, x_2, \ldots, x_n)$.

Нека имаме хипотези:
$H_0 : \theta \leq \theta_0$,
$H_1 : \theta > \theta_0$.
Тогава, понеже $L(\theta|\mathbf{x})$ *расте в интервала* $(-\infty, x_{(1)})$, *максимумът ще се достигне в* $x_{(1)}$. *Ако* $x_{(1)} \leq \theta_0$, *то* $\max\limits_{\theta\leq\theta_0} L(\theta|\mathbf{x}) = e^{-\sum\limits_{i=1}^{n} x_i + n x_{(1)}}$, *а ако* $x_{(1)} > \theta_0$, $\max\limits_{\theta\leq\theta_0} L(\theta|\mathbf{x}) = e^{-\sum\limits_{i=1}^{n} x_i + n\theta_0}$. *Окончателно:*

$$\lambda(x) = \begin{cases} 1, & x_{(1)} \leq \theta_0, \\ e^{-n(x_{(1)}-\theta_0)}, & x_{(1)} < \theta_0 \end{cases}$$



Тестът отхвърля H_0 *с критична област* $\{\mathbf{x} : x_{(1)} \geq \theta_0 - \frac{\ln c}{n}\}$.

Когато нулевата хипотеза представлява обединение (сечение) на по-прости хипотези, критичната област може да се намери като сечение (обединение) на техните критични области.

Нека $H_0 : \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma$ и
$H_{0\gamma} : \theta \in \Theta_\gamma$,
$H_{1\gamma} : \theta \in \bar{\Theta}_\gamma$.
Ако съществува тест за всяка Θ_γ с критична област $\{\mathbf{x} : T_\gamma(x) \in R_\gamma\}$, то критичната област на теста за проверка на новата хипотеза е:
$\bigcup_{\gamma \in \Gamma} \{\mathbf{x} : T_\gamma(x) \in R_\gamma\}$.

Пример 15.4 *Нека* $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$ *и*
$H_0 : \mu = \mu_0$,
$H_1 : \mu \neq \mu_0$.
Тогава $H_0 : \{\mu : \mu \leq \mu_0\} \cap \{\mu : \mu \geq \mu_0\}$ *и новата критична област е* $\frac{\bar{X}-\mu_0}{S/\sqrt{n}} \geq t_L \cup \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \leq t_U$ *(виж Пример 15.2), а ако* $t_L = t_U$*, то критичната област има вида* $\frac{|\bar{X}-\mu_0|}{S/\sqrt{n}} \geq t_L$.

Ако имаме $H_0 : \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma$ при горните условия, то новата критичната област е:
$\bigcap_{\gamma \in \Gamma} \{\mathbf{x} : T_\gamma(x) \in R_\gamma\}$.

И така, при дефинирана нулева хипотеза H_0 и алтернатива H_1 имаме следните възможности:

	приемаме H_0	отхвърляме H_0
H_0 е вярна	правилно решение	грешка от I тип
H_1 е вярна	грешка от II тип	правилно решение

Нека сме дефинирали хипотези
$H_0 : \theta \in \Theta_0$,
$H_1 : \theta \in \bar{\Theta}_0$.
и R е критичната област на теста за проверката им. Тогава

$$P_\theta(\mathbf{X} \in R) = \begin{cases} \text{вероятността за грешка от I тип}, & \theta \in \Theta_0 \\ 1 - \text{вероятността за грешка от II тип}, & \theta \in \bar{\Theta}_0 \end{cases}$$

Определение 15.5 *(Функция на) мощност на критерий/тест за проверка на хипотези за параметъра θ с критична област R наричаме функция на θ, дефинирана като $\beta(\theta) = P_\theta(\mathbf{X} \in R)$.*

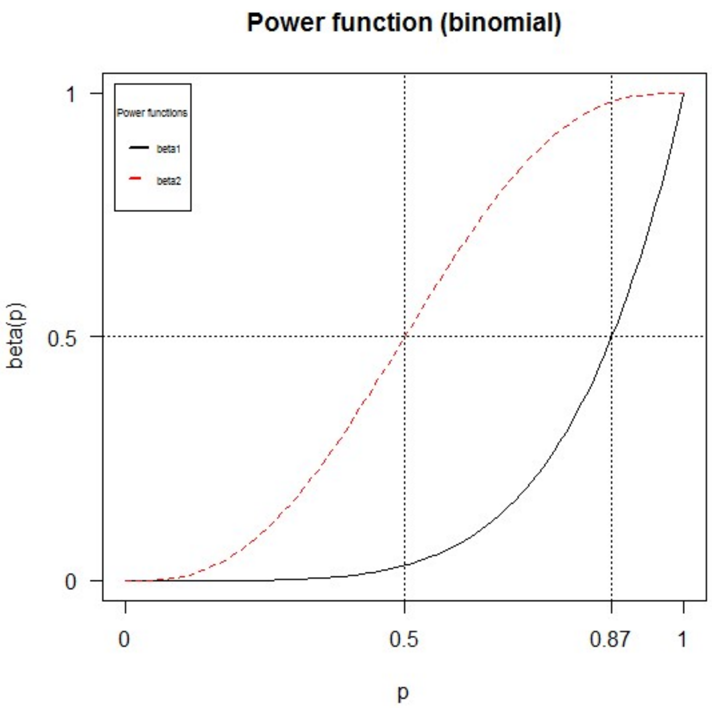
В идеалния случай мощността ще е 0 за $\theta \in \Theta_0$ и 1 за $\theta \in \bar{\Theta}_0$. При определяне на "добрите" тестове ще се стремим към такова поведение на мощността.

Пример 15.5 *Нека $X \sim Bi(n = 5, p)$.*
$H_0 : p \leq \frac{1}{2}$,
$H_1 : p > \frac{1}{2}$.
Ще сравним две възможни критични области:

$$R_1 = \{5\} : \beta_1(p) = P_p(X = 5) = p^5$$

и

$$R_2 = \{3, 4, 5\} : \beta_2(p) = P_p(X \in \{3, 4, 5\}) = \binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p) + \binom{5}{5}p^5$$

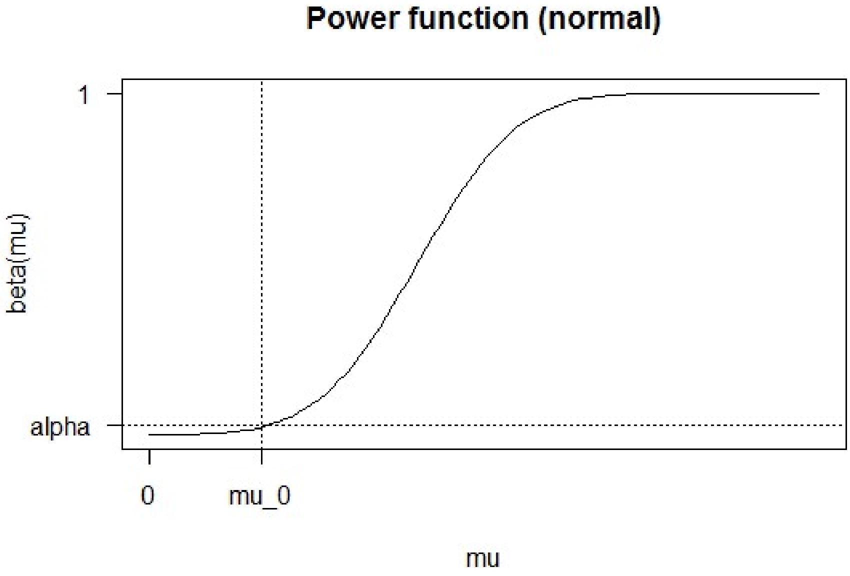


Виждаме, че първата функция е близка до 0 за голяма част от стойностите на параметъра - тя достига $\frac{1}{2}$ едва за $p = 0.87$, докато втората има по-равномерно нарастване. В първия случай почти никога няма да правим грешка от I тип, но за сметка на това доста често ще имаме грешка от II тип. Във втория случай вероятностите за двете грешки за изравнени.

Пример 15.6 *Нека $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$.*
$H_0 : \mu \leq \mu_0$
$H_1 : \mu > \mu_0$.
Ако σ^2 е известна, критичната област има вида $R : \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$. Тогава:

$$\beta(\mu) = P_\mu\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} > c\right) = P_\mu\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > c + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right) = P_\mu\left(z > c + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right),$$

за $\mu \in (-\infty, \infty)$ $\beta(\mu)$ *расте от* 0 *до* 1, $\lim\limits_{\mu \to -\infty} \beta(\mu) = 0$ *и* $\lim\limits_{\mu \to \infty} \beta(\mu) = 1$ *и освен това* $\beta(\mu_0) = \alpha$ *за* $P(Z > c) = \alpha$.



Пример 15.7 (големина на извадката n) *Да предположим, че искаме да ограничим едновременно грешките от I и II тип, например за нормална извадка максималната грешка от I тип да е* 0.1, *а тази от II тип да е* 0.2 *за* $\mu \geq \mu_0 + \sigma$. *Тогава:* $\beta(\mu) = P\left(z > c + \frac{\mu_0 - \mu}{\sigma/\sqrt{n}}\right)$, $\beta(\mu)$ *расте и следователно искаме* $\beta(\mu_0) = 0.1$ *и* $\beta(\mu_0 + \sigma) = 0.8$. *Знаем, че* $P(Z > 1.28) = 0.1$ *и* $P(Z > -0.84) = 0.8$. *Тогава* $\beta(\mu_0) = 0.1$ *за* $c = 1.28$ *и* $\beta(\mu_0 + \sigma) = P(Z > 1.28 - \sqrt{n}) = 0.8$ *за* $1.28 - \sqrt{n} = -0.84$, *откъдето* $n = 5$. *Единственият начин да контролираме едновременно и двата типа грешка, е да имаме свобода да избираме големината на извадката* n.

За фиксирано n не е възможно да изберем произволни стойности на грешките от I и от II тип, но можем да ги контролираме свързано.

Определение 15.6 *За* $0 \leq \alpha \leq 1$ *казваме, че критерий/тест с мощност* $\beta(\theta)$ *има ниво на съгласие* α, *ако* $\sup\limits_{\theta \in \Theta_0} \beta(\theta) \leq \alpha$. *При равенство казваме, че нивото на съгласие е точно* α.

Обикновено α се определя експертно (0.01, 0.05, 0.1 например) и така се контролира грешката от I тип. Тогава експерименталната хипотеза се поставя в H_1, защото малкото α ни предпазва да потвърдим експерименталната хипотеза, когато тя не е вярна.

Ако разгледаме класа на всички тестове с ниво на съгласие α, този от тях с най-малка вероятност за грешка от II тип ще е "най-добър".

Определение 15.7 *Нека* $\mathcal{C}$ *е класът от критерии за проверка на* $H_0 : \theta \in \Theta_0$ *срещу* $H_1 : \theta \in \bar{\Theta}_0$. *Критерий от този клас с мощност* $\beta(\theta)$ *наричаме равномерно най-мощен (РНМ), ако* $\beta(\theta) \geq \beta'(\theta)$, $\forall \theta \in \Theta_0$ *и* $\forall \beta'(\theta)$ *е мощност на критерий от класа* $\mathcal{C}$.

Нека сега $\mathcal{C}$ е класът на всички такива тестове с ниво на съгласие α. РНМ критерий не винаги съществува, но в някои случаи е възможно да се определят условия за това. Конкретен пример за това ни дава следната:

Теорема 15.1 (лема на Нейман – Пиърсън): *Нека имаме $H_0 : \theta = \theta_0$ срещу $H_1 : \theta = \theta_1$ и съответните плътности на разпределението означим с $f(x|\theta_i)$, $i = 1, 2$. Нека имаме тест с критична област R, за която:*
1) $x \in R$ ако $f(x|\theta_1) > kf(x|\theta_0)$ и $x \in \bar{R}$ ако $f(x|\theta_1) < kf(x|\theta_0)$ за някое $k \geq 0$ и
2) $\alpha = P_\theta(X \in R)$
Тогава:
a) (достатъчност) всеки тест, който удовлетворява 1) и 2), е РНМ с ниво на съгласие α;
b) (необходимост) ако съществува тест, удовлетворяващ 1) и 2) с $k > 0$, то всеки РНМ критерий с ниво α е с ниво точно α, и всеки РНМ критерий с ниво α удовлетворява 1), с изключение на множество A, за което $P_{\theta_0}(X \in A) = P_{\theta_1}(X \in A) = 0$.

Доказателство: Първо, всеки тест, удовлетворяващ 2) е с ниво на съгласие α, защото $\sup\limits_{\theta\in\Theta_0} P_\theta(X \in R) = P_{\theta_0}(X \in R) = \alpha$. Ще дефинираме тестова функция (индикатор на критичната област) по следния начин:

$$\phi(x) = \begin{cases} 1, & X \in R \\ 0, & X \in \bar{R} \end{cases}$$

Нека $\phi'(x)$ е тестова функция на друг тест с ниво α. Имаме, че $0 \leq \phi'(x) \leq 1$, следователно, от 1), получаваме че $(\phi(x) - \phi'(x))(f(x|\theta_1) - kf(x|\theta_0)) \geq 0$. Тогава:

$$0 \leq \int [\phi(x) - \phi'(x)][f(x|\theta_1) - kf(x|\theta_0)]dx = \beta(\theta_1) - \beta'(\theta_1) - k\beta(\theta_0) + k\beta'(\theta_0).$$

За a): $\beta(\theta_0) - \beta'(\theta_0) = \alpha - \beta'(\theta_0) \geq 0 \Rightarrow 0 \leq \beta(\theta_1) - \beta'(\theta_1) - k\beta(\theta_0) + k\beta'(\theta_0) \leq \beta(\theta_1) - \beta'(\theta_1) \Rightarrow \beta(\theta_1) \geq \beta'(\theta_1)$.
За b): Нека $\phi'(x)$ е тестова функция за някой РНМ критерий с ниво α. $\phi(x)$ удовлетворява 1) и 2), следователно е РНМ критерий с ниво точно α. Тогава $\beta(\theta_1) = \beta'(\theta_1) \Rightarrow \alpha - \beta'(\theta_0) = \beta(\theta_0) - \beta'(\theta_0) \leq 0$. Но $\beta'(\theta_0) \leq \alpha \Rightarrow$ в горното неравенство имаме равенство, а това е възможно само ако $\phi'(x)$ удовлетворява 1) и 2) с изключение на множество с мярка 0.