Testování hypotéz

Formulace úlohy:

V předchozích částech byly demonstrovány postupy vedoucí k nějakému odhadu parametru daného rozdělení. V některých případech je zajímavá „validita“ některého výroku o parametru (či o typu, či o vlastnosti, …) daného (daných) rozdělení pravděpodobnosti. Takové výroky či tvrzení budou nadále nazývány statistickými hypotézami. Budeme zkoumat postupy, ve kterých je „validita“ statistické hypotézy ověřována proti nějakému jinému výroku (statistické alternativě). Test hypotézy H proti její alternativě A se realizuje pomocí některých pozorování a jejich vyhodnocení, v dalším budeme předpokládat, že taková pozorování tvoří náhodný výběr.

Příklad 1.: Běžně je známo, že u daného osiva je podíl neklíčivých semen cca 0,25 (= pravděpodobnost toho, že dané semeno nevyklíčí $p = 0{,}25$), byla vypěstována nová varianta daného osiva, u které je předpoklad lepší klíčivosti. U ní pak testujeme hypotézu $H \equiv p = 0{,}25$ (nové osivo nemá požadovanou vlastnost) proti alternativě $A \equiv p < 0.25$ (nové osivo má požadovanou vlastnost).

Příklad 2.: Před seřízením stroje je zmetkovitost cca 7% (pravděpodobnost zmetku $p = 0{,}07$). Po seřízení testujeme hypotézu $H \equiv p = 0{,}07$ (seřízení nezlepšilo zmetkovitost) proti alternativě $A \equiv p < 0{,}07$ (seřízení má positivní vliv na zmetkovitost).

Příklad 3.: Před nástupem nové skupiny pracovníků byla chybovost produkce $p = 0{,}03$ (= pravděpodobnost vyrobení chybného produktu). Po nástupu nové skupiny pracovníků testujeme hypotézu $H \equiv p = 0{,}03$ (nic se nezměnilo) proti alternativě $A \equiv p \neq 0{,}03$ (došlo ke změně jak positivní, tak i negativní).

Příklad 4.: Máme pozorování náhodné proměnné ξ o níž víme, že její střední hodnota je μ a rozptyl σ^2. Budeme testovat hypotézu o tom, že ξ se řídí normálním rozdělením $H \equiv N(\mu,\sigma^2)$, proti alternativě $A \equiv f_\xi \in G$, kde G je množina všech jedno-modálních hustot na daném nosiči (R_1).

Příklad 5.: Před nástupem nové skupiny pracovníků byla chybovost produkce $p = 0{,}03$ (= pravděpodobnost vyrobení chybného produktu). Po nástupu nové skupiny pracovníků testujeme hypotézu $H \equiv p \le 0{,}03$ (nedošlo k zhoršení) proti alternativě $A \equiv p > 0{,}03$ (došlo k zhoršení).

Hypotézy v příkladech 1, 2, 3, 4 jsou vlastně výrokem o tom, že se jedná o náhodnou proměnnou z jednoho konkrétního rozdělení, takové budou nazývány **jednoduchými.** **Alternativy** v příkladech 1, 2, 3, 4, 5 a **hypotéza** v příkladu 5. jsou výrokem o tom, že se jedná o náhodnou proměnnou, která se řídí některým z více, specifikovaných, rozdělení pravděpodobnosti, takové budou nazývány **složenými.** Někdy se pojem jednoduchá a složená hypotéza používá jen u parametrických testů. Jednoduchá je taková, u které se uvádí některá daná hodnota parametru, složená je taková, u které se uvádí, že hodnot parametru může být „více“ než jedna.

Při testech jednoduchých parametrických hypotéz proti speciálním složeným alternativám se používá dalšího názvosloví:

$H \equiv p = k$ = daný parametr (nebo parametrická funkce) nabývá konkrétní dané hodnoty;

$A \equiv p < k$ levostranná alternativa,

$A \equiv p > k$ pravostranná alternativa,

$A \equiv p \neq k$ oboustranná alternativa.

Poznatky:

1. Hypotézu a alternativu nelze volit libovolně, vyplývají z řešeného problému.
2. Hypotéza a alternativa dohromady nemusí tvořit všechny možnosti.
3. Řešené úlohy jsou dichotomií, „vybíráme" hypotézu nebo alternativu.
4. Nelze obecně vyloučit, že testovaná hypotéza není součástí alternativy. V dalším bude však předpokládáno, že hypotéza není obsažena (a to ani částí) v alternativě.

Mějme náhodný výběr $\{x_1, x_2, ..., x_n\}$ testujeme hypotézu H proti alternativě A, k testu hypotézy použijeme statistiku $T(x_1, x_2, ..., x_n)$, která bude nazývána testovým kritériem. Obor hodnot, kterých může statistika $T(x_1, x_2, ..., x_n)$ nabývat, rozdělíme na dvě disjunktní části W a její doplněk. Pokud hodnota testového kritéria padne do W zamítáme hypotézu H ve prospěch alternativy A. Obor W bude nazýván kritickým oborem. Pokud hodnota T, testového kritéria padne mimo kritický obor W, testovanou hypotézu nezamítáme.

Chyby takového rozhodování

- **Chyba prvního druhu:** Hypotéza H **platí,** ale **je** na základě testového kritéria **zamítnuta** $T(x_1, x_2, ..., x_n) \in W$.
- **Chyba druhého druhu:** Hypotéza H **neplatí,** ale **není** na základě testového kritéria **zamítnuta** $T(x_1, x_2, ..., x_n) \notin W$.

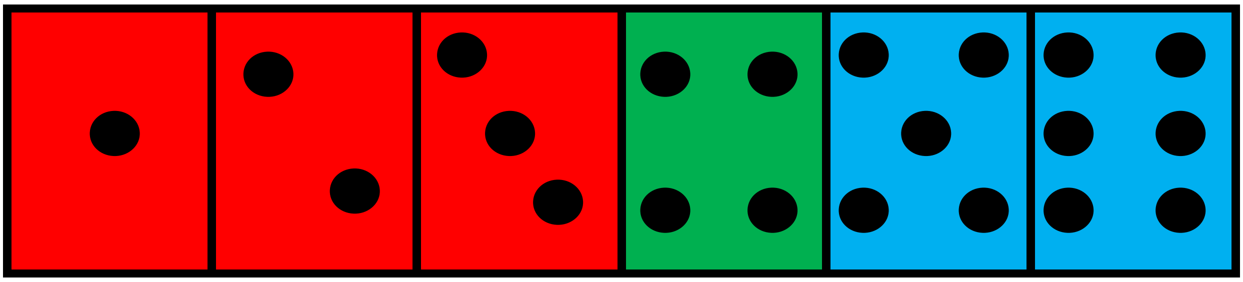
Pravděpodobnosti chyb

Pravděpodobnost chyby prvního druhu: $\alpha = P(T \in W / H)$. α je pravděpodobností chyby prvního druhu = hladinou významnosti testu = velikostí kritického oboru. Běžně bývá dána tato pravděpodobnost a k ní se určuje odpovídající kritický obor. Taková skutečnost bývá značena $\alpha = P(T \in W_\alpha / H)$. Kritický obor je pak řešením této rovnice k danému α vůči W_α. Pro složené hypotézy lze zavést kritický obor následující nerovností $\alpha \geq P_{P \in H}(T \in W_\alpha / H)$. Danému α je pak přiřazen „maximální" kritický obor vyhovující $\alpha \geq P_{P \in H}(T \in W_\alpha / H); \forall P \in H$. Pravděpodobnost chyby prvního druhu je též nazývána ***hladinou významnosti testu***.

p-hodnota testu: je nejmenší hodnotou hladiny významnosti, při konkrétním testu, při které by bylo možno ještě H zamítnout. Detaily viz: http://www.jstor.org/pss/2684655

Pravděpodobnost chyby druhého druhu: $1 - \beta = P(T \notin W_\alpha / A) = 1 - P(T \in W_\alpha / A)$. Pravděpodobnost β je pak nazývána silou testu založeného na kritickém oboru W_α nebo přímo silou kritického oboru W_α. Síla kritického oboru (testu) je pravděpodobností zamítnutí hypotézy za platnosti alternativy.

Demonstrace:



Máme klasickou, korektní, hrací kostku, která má obarvené stěny: čísla 5 a 6 mají modrou **(m)** barvu, číslo 4 zelenou (**z**) a ostatní $\{3,2,1\}$ mají červenou (**c**).

Hrajeme hru, kdy je nám sdělena barva, která padla a na jejím základě máme určit číslo, které padlo.

Varianta a.

Testovanou hypotézou je, že padla 6. její alternativou je pak, že padlo číslo jiné než 6. Kritickým oborem zvolíme $W = \{z, c\}$, oborem přijetí pak je $A_H = \{m\}$.

Pravděpodobnost chyby **1. druhu** = zamítáme H, pokud platí je pak **0**, neboť:
$p(1/z,c) = 1/4;\ p(2/z,c) = 1/4;\ p(3/z,c) = 1/4;\ p(4/z,c) = 1/4$
$p(5/z,c) = 0;\ p(6/z,c) = 0$.

Pravděpodobnost chyby **2. druhu** = přijímáme H pokud neplatí je pak ½, neboť:
$p(1/m) = 0;\ p(2/m) = 0;\ p(3/m) = 0;\ p(4/m) = 0\ \ p(5/m) = 1/2;\ p(6/m) = 1/2$.
Kritická funkce je tedy: $\phi(m) = 0; \phi(z) = 1; \phi(c) = 1;$

Varianta b.

Testovanou hypotézou je, že padla 6. její alternativou je pak, že padlo číslo jiné než 6. Kritickým oborem zvolíme $W = \{m, z, c\}$, oborem přijetí je pak $A_H = \emptyset$.

Pravděpodobnost chyby **1. druhu** = zamítáme H, pokud platí je pak **1,0**, neboť:
$p(1/m,z,c) = 1/6;\ p(2/m,z,c) = 1/6;\ p(3/m,z,c) = 1/6;\ p(4/m,z,c) = 1/6$
$p(5/m,z,c) = 1/6;\ p(6/m,z,c) = 1/6$.

Pravděpodobnost chyby **2. druhu** = přijímáme H pokud neplatí je pak **0**, neboť:v tomto případě nikdy H nepřijmeme.
Kritická funkce je tedy: $\phi(m) = 1; \phi(z) = 1; \phi(c) = 1;$

Varianta c.

Testovanou hypotézou je, že padla 6. její alternativou je pak, že padlo číslo jiné než 6. Kritickým oborem zvolíme $W = \{c\}$, oborem přijetí $A_H = \{m, z\}$.

Pravděpodobnost chyby **1. druhu** = zamítáme H, pokud platí je pak **0**, neboť:
$p(1/c) = 1/3;\ p(2/c) = 1/3;\ p(3/c) = 1/3;\ p(4/c) = 0\ \ p(5/c) = 0;\ p(6/c) = 0$.;

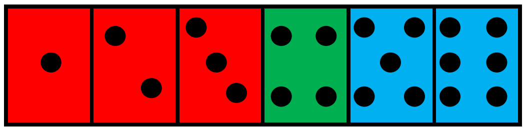
Pravděpodobnost chyby **2. druhu** = přijímáme H pokud neplatí je pak **2/3**, neboť:

$p(1/m,z)=0;\ p(2/m,z)=0;\ p(3/m,z)=0;\ p(4/m,z)=1/3$

$p(5/m,z)=1/3;\ p(6/m,z)=1/3.$

Kritická funkce je tedy:

$\phi(m)=0;\phi(z)=0;\phi(c)=1;$



Varianta … .

Shrnutí:

	Kritický obor - oblast zamítnutí 6	Pravděpodobnost chyby 1. druhu, tj. zamítnutí 6, když padla: α	Pravděpodobnost chyby 2. druhu, tj. přijetí 6, když nepadla: $1\text{-}\beta$	Pravděpodobnost chybného rozhodnutí
1	c, z, m	100,00%	0,00%	16,67%
2	c, z	0,00%	50,00%	41,67%
3	c	0,00%	66,67%	55,56%
4	z	0,00%	80,00%	66,67%
5	*prázdný*	0,00%	100,00%	83,33%
6	c, m	20,00%	100,00%	86,67%
7	z, m	33,33%	100,00%	88,89%
8	m	50,00%	100,00%	91,67%

$P(H\text{ plati})=1/6=16{,}67\%$, $P(H\text{ neplati})=1-1/6=83{,}33\%$

Problém volby vhodného kritického oboru (pokud to lze) je dán především podstatou úkolu, který má statistický postup řešit. Aparát testování hypotéz dává prostředky pro řešení situace, kdy nejsou k dispozici další informace.

Příklad: Na základě diagnostických dat se má rozhodnout o operačním odstranění krčních (?) mandlí či o jejich neodstranění. Pokud budou odstraněny je u pacienta dobrá předpověď ústupu „anginósních" komplikací. Pokud budou odstraněny, vzrůstá riziko různých zhoubných bujení. Tj. hypotéza (rozhodnutí) o odstranění je přijímána pokud rizika případných následků častých „angín" (přetěžování kardiovaskulárního systému) převýší potenciální riziko možných (potenciálně) zhoubných bujení. Přijetí hypotézy $H\equiv$ mandle budou operačně odstraněny je podmíněno hodnotou rozhodovací funkce (statistiky) složené z anamnestických a diagnostických informací. Alternativou A je rozhodnutí o neoperování.

Chyby 1. ho druhu se dopustíme, pokud zamítneme H, když H platí. V tomto případě pokud se rozhodneme **neoperovat** za předpokladu, že se **operovat má**.

Chyby 2. ho druhu se dopustíme, pokud přijmeme H, když H neplatí. V tomto případě pokud se rozhodneme **operovat**, když se **operovat nemá**.

V tomto případě **chyba 2. ho druhu** je **nenapravitelná**, zatím-co **chyba 1. ho druhu** je (alespoň potenciálně později) **napravitelná**.

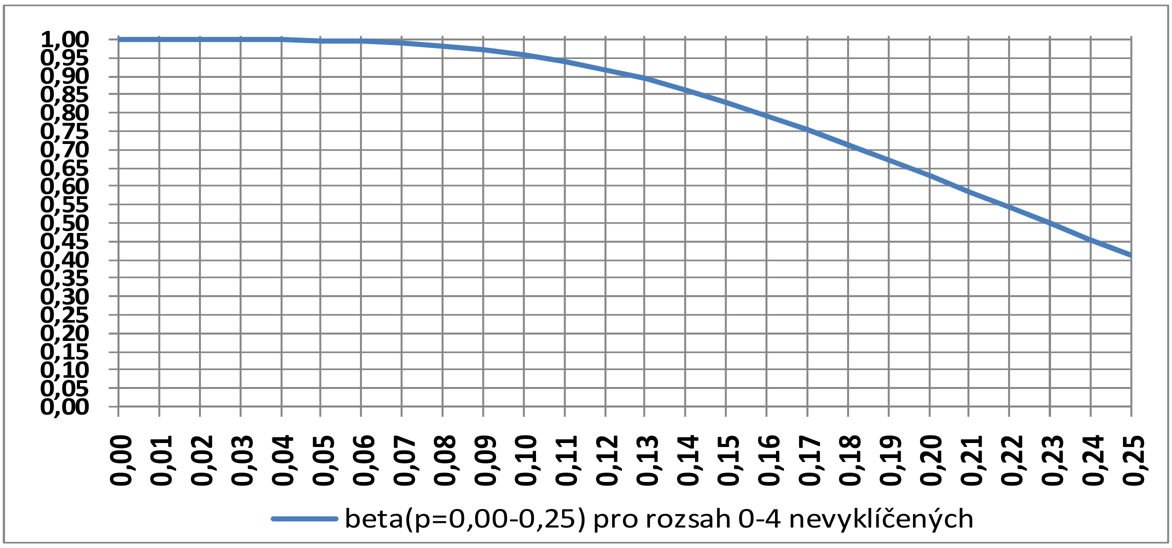
Tj. vlastní podstata problému si vynutí minimalizaci pravděpodobnosti chyby 2. ho druhu při dodržení omezení na pravděpodobnost chyby 1. ho druhu, pokud to jde.

Vlastní chyba rozhodnutí zde nemá významnějšího postavení.

Příklad: Běžně je známo, že u daného osiva je podíl neklíčivých semen cca 0,25 (= pravděpodobnost toho, že dané semeno nevyklíčí $p = 0{,}25$), byla vypěstována nová varianta daného osiva, u které je předpoklad lepší klíčivosti. U ní pak testujeme hypotézu $H \equiv p = 0{,}25$ (nové osivo nemá požadovanou vlastnost) proti alternativě $A \equiv p < 0.25$ (nové osivo má požadovanou vlastnost). Byl učiněn laboratorní experiment s klíčivostí semen. Bylo testováno 20 semen. Počet semen, která nevyklíčila T se pak řídí binomickým rozdělením pravděpodobnosti $P(\text{nevyklici } k) = \binom{20}{k} p^k (1-p)^{20-k}$, kde p je skutečná klíčivost. Intuitivně určíme, že zamítáme hypotézu $H \equiv p = 0{,}25$ v případě kdy nevyklíčí 0 až 4 semena, tedy $W = \{0,1,2,3,4\}$. Potom pravděpodobnost chyby prvního druhu je: $\alpha = \sum_{i=0}^{4} \binom{20}{i} p^i (1-p)^{20-i}; p = 0{,}25; \text{pak } \alpha = 0{,}415 \Rightarrow W = \{0,1,2,3,4\} = W_{0{,}415}$. Pro sílu testu dostáváme:

$$1-\beta = P(T \notin W_\alpha / A) = 1 - P(T \in W_\alpha / A) = 1 - \beta(p) = 1 - \sum_{i=0}^{4} \binom{20}{i} p^i (1-p)^{20-i}; p = \langle 0;0{,}25 \rangle \Rightarrow$$

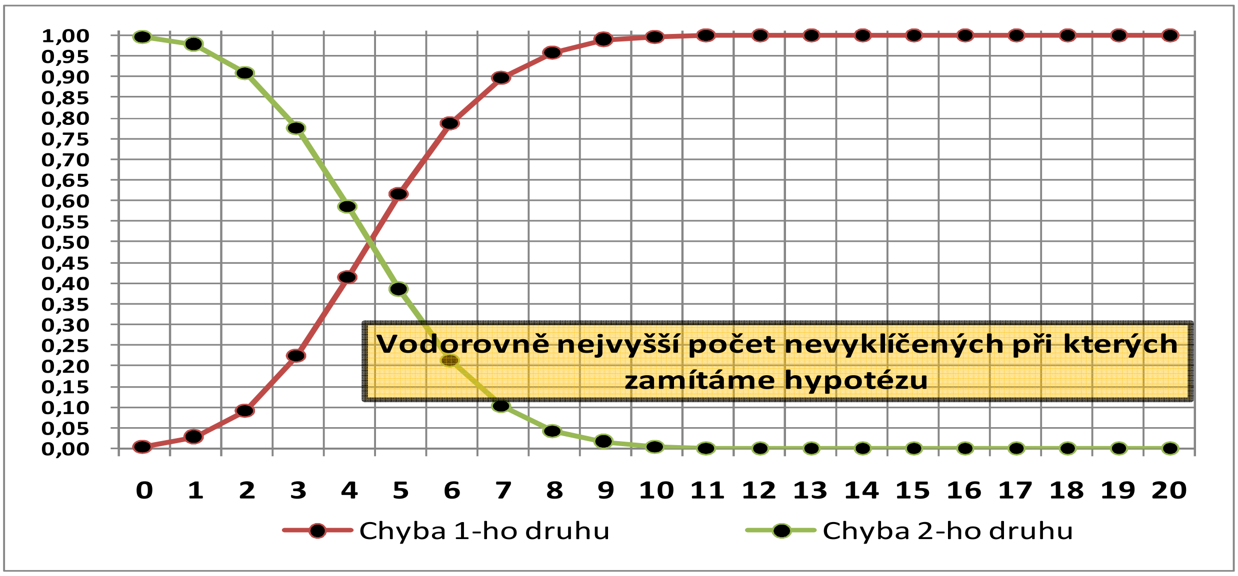
$\beta(p) = \sum_{i=0}^{4} \binom{20}{i} p^i (1-p)^{20-i}; p = \langle 0;0{,}25 \rangle$, její průběh je v následujícím grafu.



Průběh síly testu.

Pro garantovanou chybu druhého druhu pak platí:

$$1 - \beta = \sup_{p \in (0;0{,}25)} P(T \notin W_\alpha / A) = 1 - \inf_{p \in (0;0{,}25)} \sum_{i=0}^{4} \binom{20}{i} p^i (1-p)^{20-i}$$



Závislost chyby prvního a druhého druhu na volbě kritického oboru

Z tohoto příkladu lze tušit, že snižování chyby 1. ho druhu („zmenšování" kritického oboru) vede na zvětšování chyby 2. ho druhu. Toto bude upřesněno v následujícím (pro obecný případ, nikoliv jen tento konkrétní):

Mějme náhodný výběr $\{x_1, x_2, ..., x_n\}$ a statistiku $T(x_1, x_2, ..., x_n)$. Uvažujeme kritický obor W_α, pro který je dle definice $\alpha \geq P(T \in W_\alpha / H)$, dále další kritický obor $W \subset W_\alpha$, takový že: $P(T \in W_\alpha - W / H) > 0; P(T \in W_\alpha - W / A) > 0$.

Potom:

$\alpha_W = P(T \in W / H) = P(T \in W_\alpha / H) - P(T \in W_\alpha - W / H) < P(T \in W_\alpha / H)$ a

$1 - \beta_W = 1 - P(T \in W / A) = 1 - (P(T \in W_\alpha / A) - P(T \in W_\alpha - W / A)) =$

$= 1 - P(T \in W_\alpha / A) + P(T \in W_\alpha - W / A) > 1 - \beta_{W_\alpha}$

Opačně: Mějme kritický obor $W \supset W_\alpha$, takový že: $P(T \in W - W_\alpha / H) > 0; P(T \in W - W_\alpha / A) > 0$.

Potom: $\alpha_W = P(T \in W / H) = P(T \in W - W_\alpha / H) + P(T \in W_\alpha / H) > P(T \in W_\alpha / H)$ a

$1 - \beta_W = 1 - P(T \in W / A) = 1 - (P(T \in W_\alpha / A) + P(T \in W - W_\alpha / A)) =$

$= 1 - P(T \in W_\alpha / A) - P(T \in W - W_\alpha / A) < 1 - \beta_{W_\alpha}$

Odtud plyne jeden z možných přístupů k testování, zvolí se chyba 1-ho druhu $0 < \alpha < 1$ a hledá se takový test (takové testové kritérium, kritický obor W_α), který minimalizuje chybu druhého druhu $1 - \beta_{W_\alpha}$ (= maximalizuje sílu testu β_{W_α}). Toto je zřejmé pokud je alternativa jednoduchá. Pokud je složená, pak hledáme takový test, který maximalizuje sílu testu pro „každý prvek" alternativy (samozřejmě pokud takový test existuje).

Proto: při jednoduché alternativě mluvíme o **nejsilnějším testu** a při složené alternativě mluvíme o **stejnoměrně nejsilnějším testu,** pokud takový existuje.

Neyman – Pearsonovo lemma (viz příloha) říká, že pro nejsilnější test při jednoduché hypotéze H a alternativě A a u náhodného výběru $\{x_1, x_2, ..., x_n\}$ ke každému α existuje k, takové, že nerovnost:

$$\frac{Q(x_1, x_2, ..., x_n; H)}{Q(x_1, x_2, ..., x_n; A)} = \frac{\prod_{i=1}^{n} f(x_i; H)}{\prod_{i=1}^{n} f(x_i; A)} < k$$ definuje kritický obor W_α pro nejsilnější test,

kde $f(x_i; H)$ je hustota (pravděpodobnost v diskrétním případu) při platnosti H a $f(x_i; A)$ je hustota (pravděpodobnost v diskrétním případu) při platnosti A, $P\left(\frac{\prod_{i=1}^{n} f(x_i; H)}{\prod_{i=1}^{n} f(x_i; A)} < k \middle| H \right) = \alpha$. Tedy testovým kritériem je věrohodnostní poměr nebo výraz s ním ekvivalentní (tj. dávající stejnou množinu řešení).

Příklad: Tvar testového kritéria, při parametrickém testu, pro exponenciální typ rozdělení (tj. při testu parametrické hypotézy) $f(x; g) = e^{K(x)Q(g)+R(g)+S(x)} = e^{K(x)Q(g)}e^{R(g)+S(x)}$. Pak $\prod_{i=1}^{n} f(x_i; H) = e^{nR(g_0)} \prod_{i=1}^{n} e^{K(x_i)Q(g_0)+S(x_i)}$ a $\prod_{i=1}^{n} f(x_i; A) = e^{nR(g)} \prod_{i=1}^{n} e^{K(x_i)Q(g)+S(x_i)}; g \in G$ při jednoduché hypotéze a složené alternativě. Testové kritérium bude vypadat:

$$\frac{e^{nR(g_0)} \prod_{i=1}^{n} e^{K(x_i)Q(g_0)+S(x_i)}}{e^{nR(g)} \prod_{i=1}^{n} e^{K(x_i)Q(g)+S(x_i)}} = \frac{e^{nR(g_0)} \prod_{i=1}^{n} e^{K(x_i)Q(g_0)}}{e^{nR(g)} \prod_{i=1}^{n} e^{K(x_i)Q(g)}} =$$

$$= e^{n(R(g_0)-R(g))} \left(\prod_{i=1}^{n} e^{K(x_i)} \right)^{Q(g_0)-Q(g)} = e^{n(R(g_0)-R(g))} e^{(Q(g_0)-Q(g)) \sum_{i=1}^{n} K(x_i)}$$

Kritický obor pak bude vypadat následovně: $\left[e^{n(R(g_0)-R(g))} e^{(Q(g_0)-Q(g)) \sum_{i=1}^{n} K(x_i)} < k \right] \Leftrightarrow$

$$\Leftrightarrow \left[(Q(g_0)-Q(g)) \sum_{i=1}^{n} K(x_i) + n(R(g_0)-R(g)) < lg\, k \right] \Leftrightarrow \left[\sum_{i=1}^{n} K(x_i) < \frac{lg\, k - n(R(g_0)-R(g))}{Q(g_0)-Q(g)} \overset{def}{=} c \right].$$

To vše pokud $Q(g_0) - Q(g) > 0; \forall g \in G$. Pokud $Q(g_0) - Q(g) < 0; \forall g \in G$ je $\sum_{i=1}^{n} K(x_i) > c$

Shrnuto: Tvar kritického oboru bude: $\sum_{i=1}^{n} K(x_i) < c, resp. \left(\sum_{i=1}^{n} K(x_i) > c \right)$, kde číslo c se určí podle volby chyby 1-ho druhu $P\left(\sum_{i=1}^{n} K(x_i) < c \middle| g_0 \right) = \alpha, resp. \left(P\left(\sum_{i=1}^{n} K(x_i) > c \middle| g_0 \right) = \alpha \right)$ v případě $Q(g_0) - Q(g) > 0; \forall g \in G, resp. (Q(g_0) - Q(g) < 0; \forall g \in G)$.

Příklad 1.: Pro normální rozdělení se známým parametrem σ^2 je $K(x)=x; Q(\mu)=\frac{\mu}{\sigma^2}; R(\mu)=-\frac{\mu^2}{2\sigma^2}-0.5ln(2\pi\sigma^2); S(x)=\frac{-x^2}{2\sigma^2}$, (námět: odvoďte a dokažte). Máme náhodný výběr rozsahu n $\{x_1, x_2, ..., x_n\}$.

Budeme testovat jednoduchou hypotézu $\mu=\mu_0$ proti jednoduché alternativě $\mu=\mu_1; \mu_1>\mu_0$, μ_1 je konkrétní dané číslo. Potom: $G=\{\mu_1\}$, $Q(\mu_0)-Q(\mu_1)=\frac{\mu_0-\mu_1}{\sigma^2}<0$,

$\sum_{i=1}^{n} K(x_i)=\sum_{i=1}^{n} x_i = n\bar{x}$, při α a kritickém oboru daném vztahem

$$\alpha = P\left(\sum_{i=1}^{n} K(x_i)>c \middle| \mu_0\right)=P\left(\sum_{i=1}^{n} x_i>c \middle| \mu_0\right)=P\left(n\bar{x}>c \middle| \mu_0\right)=P\left(\bar{x}>\frac{c}{n} \middle| \mu_0\right)=$$

$$=P\left(\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}>\frac{\frac{c}{n}-\mu_0}{\frac{\sigma}{\sqrt{n}}} \middle| \mu_0\right)$$ označíme $\frac{\frac{c}{n}-\mu_0}{\frac{\sigma}{\sqrt{n}}}=c_1$ a dostaneme $\alpha=P\left(\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}>c_1 \middle| \mu_0\right)$.

Náhodná veličina $\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}$ má rozdělení $N(0,1)$, proto

$$\alpha = P\left(\frac{\bar{x}-\mu_0}{\frac{\sigma}{\sqrt{n}}}>c_1 \middle| \mu_0\right)=P\left(\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}>c_1 \middle| \mu_0\right)=1-\Phi(c_1).$$ A odtud $c_1=u_{1-\alpha}$. Pak při $\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}>u_{1-\alpha}$ zamítáme $H\equiv\mu=\mu_0$ ve prospěch $A\equiv\mu=\mu_1$. Nerovnost $\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}>u_{1-\alpha}$ lze postupně upravovat $\left[\frac{\sqrt{n}(\bar{x}-\mu_0)}{\sigma}>u_{1-\alpha}\right]\Leftrightarrow\left[\bar{x}>\mu_0+u_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right]$.

Chyba druhého druhu: $1-\beta=P(T\notin W_\alpha/A)=1-P(T\in W_\alpha/A)=1-P\left(\bar{x}>\mu_0+u_{1-\alpha}\frac{\sigma}{\sqrt{n}} \middle| \mu_1\right)$ a

síla testu je $$\beta=P\left(\bar{x}>\mu_0+u_{1-\alpha}\frac{\sigma}{\sqrt{n}} \middle| \mu_1\right)=P\left(\frac{\sqrt{n}(\bar{x}-\mu_1)}{\sigma}>\frac{\sqrt{n}\left(\mu_0-\mu_1+u_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right)}{\sigma} \middle| \mu_1\right).$$

Za platnosti $A\equiv\mu=\mu_1$ má, opět, náhodná veličina $\frac{\sqrt{n}(\bar{x}-\mu_1)}{\sigma}$ rozdělení $N(0,1)$ a odtud:

$$\beta=1-\Phi\left(\frac{\sqrt{n}\left(\mu_0-\mu_1+u_{1-\alpha}\frac{\sigma}{\sqrt{n}}\right)}{\sigma}\right)=1-\Phi\left(\frac{\sqrt{n}(\mu_0-\mu_1)}{\sigma}+u_{1-\alpha}\right),$$ kde $u_{1-\alpha}$ je $1-\alpha$ kvantil rozdělení $N(0,1)$. Poznamenejme, že: síla kritického oboru (testu) je pravděpodobností

zamítnutí hypotézy za platnosti alternativy. Za uvedených předpokladů se jedná o nejsilnější test.

Námět: Podrobně rozeberte úlohu předcházejícího konkrétního případu, pokud budeme testovat jednoduchou hypotézu $\mu = \mu_0$ proti jednoduché alternativě $\mu = \mu_1; \mu_1 < \mu_0$.

Příklad pokračování předchozího obecného 2.: Pro normální rozdělení se známým parametrem σ^2 budeme testovat jednoduchou hypotézu $\mu = \mu_0$ proti alternativě $\mu = \mu_1; \mu_1 > \mu_0$, kde μ_1 je jakékoliv číslo splňující danou nerovnost. Alternativa je tedy v této modifikaci složená. Odvození jednotlivých vztahů je shodné jako v přecházející konkrétní úloze. Tedy hypotézu zamítáme, pokud $\bar{x} > \mu_0 + u_{1-\alpha} \frac{\sigma}{\sqrt{n}}$ a $\beta(\mu_1) = 1 - \Phi\left(\frac{\sqrt{n}(\mu_0 - \mu_1)}{\sigma} + u_{1-\alpha} \right)$ je v tomto případě silofunkce testu. Protože $\Phi(x)$ je rostoucí funkce na celém R_1, je $\beta(\mu_1)$ klesající funkce na $\mu_1 > \mu_0$ a proto platí $\beta(\mu_1) \le \beta(\mu_0) = 1 - \Phi(u_{1-\alpha}) = \alpha$. Jedná se tedy o test stejnoměrně nejsilnější.

Námět: Odvoďte totéž, pokud budeme testovat jednoduchou hypotézu $\mu = \mu_0$ proti ***složené*** alternativě $\mu = \mu_1; \mu_1 < \mu_0$.

Námět: Rozborem předchozích příkladů dokažte, že pro test jednoduché hypotézy $\mu = \mu_0$ proti složené alternativě $\mu = \mu_1; \mu_1 \neq \mu_0$ neexistuje stejnoměrně nejsilnější test.

Test poměrem věrohodností

V předchozích případech byla studována jednoduchá hypotéza proti jednoduché nebo složené alternativě. Ani v těchto relativně jednoduchých případech není garantována existence stejnoměrně nejsilnějšího testu. Častý je i případ kdy je sama hypotéza složenou. Proto byl navržen test (pro parametrické hypotézy) poměrem věrohodností. Mějme nějaký prostor možných parametrů G, a rozdělení $f(x; g)$. Budeme testovat hypotézu $g \in G_H \subset G$ obecně proti alternativě $g \in G_A \subset G; G_H \cap G_A = \emptyset$. Jako testové kritérium bude užit poměr věrohodností: $L^*(x_1, x_2, ..., x_n) = \dfrac{\sup\limits_{g \in G_H} \prod\limits_{i=1}^{n} f(x_i; g)}{\sup\limits_{g \in G} \prod\limits_{i=1}^{n} f(x_i; g)}$. Pokud existuje maximálně věrohodný odhad $\hat{g}$ parametru g, pak evidentně platí: $L^*(x_1, x_2, ..., x_n) = \dfrac{\sup\limits_{g \in G_H} \prod\limits_{i=1}^{n} f(x_i; g)}{\prod\limits_{i=1}^{n} f(x_i; \hat{g})}$. Kritický obor je pak zaveden standardním způsobem $P\left(L^*(x_1, x_2, ..., x_n) \le L^*_\alpha\right) = \alpha$ a obdobně silofunkce testu $\beta(g) = P\left(L^*(x_1, x_2, ..., x_n) \le L^*_\alpha\right); g \in G_A$. Opět je připomínáno, že síla testu (silofunkce) je pravděpodobností zamítnutí hypotézy při příslušnosti testovaného parametru „alternativní" množině. Pro $L^*(x_1, x_2, ..., x_n)$ platí $0 < L^*(x_1, x_2, ..., x_n) \le 1$. Takový test sám o sobě je svou konstrukcí intuitivní (pokud se poměr blíží jedné, je předpoklad, že alternativní množina je blízká maximálně věrohodnému odhadu {pokud existuje}; Neyman, J., Pearson, E.S.: On the use and interpretation of certain test kriteria for purposes of statistical inference. Biometrika 20A, str.175-240, 263-294, rok 1928).

Pro exponenciální skupinu rozdělení, při existenci maximálně věrohodného odhadu je: $f(x,g)=e^{K(x)Q(g)+R(g)+S(x)}$ a

$$L^*(x_1,x_2,\dots,x_n)=\frac{\sup\limits_{g\in G_H}\prod\limits_{i=1}^{n}e^{K(x_i)Q(g)+R(g)+S(x_i)}}{\prod\limits_{i=1}^{n}e^{K(x_i)Q(\hat g)+R(\hat g)+S(x_i)}}=\sup_{g\in G_H}\frac{\prod\limits_{i=1}^{n}e^{K(x_i)Q(g)+R(g)+S(x_i)}}{\prod\limits_{i=1}^{n}e^{K(x_i)Q(\hat g)+R(\hat g)+S(x_i)}}=\sup_{g\in G_H}\prod_{i=1}^{n}e^{K(x_i)(Q(g)-Q(\hat g))+(R(g)-R(\hat g))};$$

vzhledem k tomu, že logaritmus je rostoucí funkce lze pracovat s logaritmem testového kritéria:

$$\lg L^*(x_1,x_2,\dots,x_n)=\sup_{g\in G_H}\left(\lg\prod_{i=1}^{n}e^{K(x_i)(Q(g)-Q(\hat g))+(R(g)-R(\hat g))}\right)=\sup_{g\in G_H}\left\{(Q(g)-Q(\hat g))\left[\sum_{i=1}^{n}K(x_i)\right]+n(R(g)-R(\hat g))\right\}$$

Příklad: Test o parametru alternativního rozdělení p. $A(p)\cong f(x)=p^x(1-p)^{1-x}=e^{x\lg(\frac{p}{1-p})+\lg(1-p)}, x=0,1;$ $K(x)=x; Q(p)=\lg\left(\frac{p}{1-p}\right); R(p)=\lg(1-p); S(x)=0$. Logaritmus testového kritéria bude:

$$\lg L^*(x_1,x_2,\dots,x_n)=\sup_{p\in G_H}\left\{(Q(p)-Q(\hat p))\left[\sum_{i=1}^{n}K(x_i)\right]+n(R(p)-R(\hat p))\right\}=$$

$$=\sup_{p\in G_H}\left\{\left(\lg\left(\frac{p}{1-p}\right)-\lg\left(\frac{\hat p}{1-\hat p}\right)\right)\sum_{i=1}^{n}x_i+n(\lg(1-p)-\lg(1-\hat p))\right\}=\sup_{p\in G_H}\left\{\lg\left(\frac{p(1-\hat p)}{\hat p(1-p)}\right)\sum_{i=1}^{n}x_i+n\lg\frac{(1-p)}{(1-\hat p)}\right\}$$

kde $\hat p=\frac{1}{n}\sum_{i=1}^{n}x_i$.

Námět: sestrojte test parametrické hypotézy $p<0.5$ proti alternativě $p\geq 0.5$ pro alternativní rozdělení, včetně kritického oboru a průběhu silofunkce pro různé pozorované hodnoty $k=\sum_{i=1}^{n}x_i;\ \hat p=\frac{k}{n}$ a pro různá α (přípustné hodnoty chyby prvního druhu).

Poznámka: Silofunkce může být v některých případech prostředek ke stanovení potřebného rozsahu náhodného výběru použitého v testu.

Doporučená a zdrojová literatura:

Jiří Reif	Metody matematické statistiky, ZČU v Plzni 2004
Jaroslav Hátle, Jiří Likeš	Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974.
C. Radhakrishna Rao	Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978
Alfréd Rényi	Teorie pravděpodobnosti, ACADEMIA, Praha 1972
Jana Jurečková	Testy parametrických hypotéz, skripta MFF-UK, Praha 1982.
Mark J. Schervish	P Values: What They Are and What They Are Not. The American Statistician, Vol. 50, No. 3 (Aug., 1996), pp. 203-206. http://www.jstor.org/stable/2684655

Příloha: Neyman-Pearsonovo lemma

Zdroj-kopie z: Jana Jurečková: Testy parametrických hypotéz, skripta MFF-UK, Praha 1982, str. 30-33.

2.2. Neyman-Pearsonovo fundamentální lemma

Systém rozdělení pravděpodobnosti nazveme jednoduchý, jestliže obsahuje právě jeden prvek. Systém, který není jednoduchý, nazveme složený.

Jestliže je hypotéza i alternativa jednoduchá, odpadá problém stejnoměrnosti v úloze (2.7)-(2.8). Úloha pak má řešení za velmi obecných podmínek a řešení má jednoduchý explicitní tvar.

Následující věta je základem teorie testování statistických hypotéz; má četné aplikace nejen při hledání optimálního testu, ale i při jiných optimalizačních úlohách.

VĚTA 2.1. (Neyman-Pearsonovo lemma). Nechť P_0 a P_1 jsou dvě rozdělení pravděpodobnosti na $(\mathfrak{X}, \mathfrak{A})$, která mají hustoty p_0 a p_1 vzhledem k míře μ (může být i $\mu = P_0 + P_1$).

(1) Pak pro lib. $\alpha \in (0,1)$ existuje test Φ hypotézy $H : \{P_0\}$ proti alternativě $K : \{P_1\}$, který splňuje

$$(2.9) \qquad \Phi(x) = \begin{cases} 1 & \text{když} \quad p_1(x) > k\, p_0(x) \\ 0 & \text{když} \quad p_1(x) < k\, p_0(x) \end{cases}$$

kde konstanta k a hodnoty $\Phi(x)$ na $\{x : p_1(x) = k \cdot p_0(x)\}$ jsou určeny tak, že platí

$$(2.10) \qquad E_0 \Phi(X) = \alpha .$$

Test Φ je nejsilnějším testem H proti K velikosti $\leq \alpha$.

(2) Je-li Φ nejsilnější test H proti K velikosti $\leq \alpha$, pak Φ vyhovuje vztahu (2.9) pro nějakou konstantu k s.v. $[\mu]$. Pokud neexistuje test velikosti $< \alpha$ a síly 1, pak Φ vyhovuje také (2.10).

Důkaz. (1) Uvažujme funkci $\alpha(c) = P_0\left\{p_1(X) > c\, p_0(X)\right\}$; tato funkce je nerostoucí, zprava spojitá a taková, že $\lim_{c \to 0+} \alpha(c) = 1$ a $\lim_{c\to\infty} \alpha(c) = 0$. Pak k libovolnému $\alpha \in (0,1)$ existuje $k \geq 0$ tak, že

$$\alpha(k-0) \geq \alpha \geq \alpha(k) \tag{2.11}$$

kde jsme označili $\alpha(k-0) = \lim_{c\to k-} \alpha(c)$.

Uvažujme test

$$\Phi(x) = \begin{cases} 1 & \text{když } p_1(x) > k\, p_0(x) \\ \dfrac{\alpha - \alpha(k)}{\alpha(k-0) - \alpha(k)} & \text{když } p_1(x) = k\, p_0(x) \\ 0 & \text{když } p_1(x) < k\, p_0(x) \end{cases} \tag{2.12}$$

Střední výraz na (2.12) není definován, pokud je k bodem spojitosti $\alpha(c)$; pak je však $P_0\left\{p_1(X) = k\, p_0(x)\right\} = 0$. Velikost testu (2.12) je

$$E_0 \Phi(X) = P_0\left\{\frac{p_1(X)}{p_0(X)} > k\right\} + \frac{\alpha - \alpha(k)}{\alpha(k-0) - \alpha(k)} P_0\left\{\frac{p_1(X)}{p_0(X)} = k\right\} = \alpha \tag{2.13}$$

a tedy test Φ splňuje (2.9) a (2.10).

Zbývá dokázat, že test Φ je nejsilnější. Nechť Φ^* je libovolný jiný test takový, že $E_0 \Phi^*(X) \leq \alpha$. Nechť $S^+ = \left\{x \in \mathcal{X}: \Phi(x) - \Phi^*(x) > 0\right\}$ a $S^- = \left\{x \in \mathcal{X}: \Phi(x) - \Phi^*(x) < 0\right\}$.

Jestliže $x \in S^+$, je $\Phi(x) > 0$, a tedy $p_1(x) \geq k\, p_0(x)$; podobně je $p_1(x) \leq k\, p_0(x)$ pro $x \in S^-$; odtud plyne

$$\int (\Phi - \Phi^*)(p_1 - k\, p_0)\,d\mu = \int_{S^+ \cup S^-} (\Phi - \Phi^*)(p_1 - kp_0)\,d\mu \geq 0 \tag{2.14}$$

a tedy

$$\int (\Phi - \Phi^*)\,p_1\,d\mu \geq k \int (\Phi - \Phi^*)\,p_0\,d\mu \geq 0$$

a test Φ^* je nejvýše stejně silný jako Φ.

(2) Nechť Φ^* je nejsilnější test H proti K velikosti $\leq \alpha$ a nechť Φ vyhovuje (2.9) a (2.10). Nechť S^+ a S^- jsou definovány stejně jako v důkazu části (1) a nechť $S = (S^+ \cup S^-) \cap \{x : p_1(x) \neq k\, p_0(x)\}$.
Předpokládejme $\mu(S) > 0$. Protože je $(\Phi(x) - \Phi^*(x)) \cdot (p_1(x) - k\, p_0(x)) > 0$ pro $x \in S$, platí

$$\int_{S^+ \cup S^-} (\Phi - \Phi^*)(p_1 - k\, p_0)\,d\mu = \int_S (\Phi - \Phi^*)(p_1 - k\, p_0)\,d\mu > 0,$$

a tedy

$$E_1\left[\Phi(x) - \Phi^*(x)\right] > k\, E_0\left[\Phi(x) - \Phi^*(x)\right] \geq 0$$

a tedy Φ je silnější než Φ^* proti alternativě K. To je však spor, a tedy $\mu(S) = 0$, což bylo třeba dokázat.

Kdyby velikost Φ^* byla $< \alpha$ a síla < 1, přidáním dalších bodů (nebo částí bodů) do kritického oboru bychom zvyšovali sílu Φ^* tak dlouho, dokud by buď velikost nedosáhla α nebo síla nedosáhla 1. Odtud plyne, že platí buď $E_0 \Phi^*(x) = \alpha$ nebo $E_1 \Phi^*(x) = 1$. Q.E.D.

<u>Důsledek.</u> <u>Nechť</u> Φ <u>je nejsilnější test</u> $H : \{P_0\}$ <u>proti</u> $K : \{P_1\}$ <u>velikosti</u> $\leq \alpha$, $0 < \alpha < 1$. <u>Pak platí buď</u> $\beta = E_1 \Phi(X) > \alpha$ <u>nebo</u> $P_0 = P_1$.

<u>Důkaz</u>: Test $\Phi_1(x) \equiv \alpha$ má velikost α i sílu α; test Φ je alespoň stejně silný jako Φ_1, a tedy $\beta = E_1 \Phi(X) \geq \alpha$. Jestliže je $\alpha = \beta < 1$, je test $\Phi_1(x) \equiv \alpha$ nejsilnější a podle věty 2.1, část (2), musí vyhovovat (2.9). Odtud plyne, že $p_0(x) = p_1(x)$ s.v. $[\mu]$, a tedy $P_0 = P_1$.