Testování hypotéz, sekvenční testy, více výběrové testy, testy založené na bayesovských postupech.

Waldovy testy – opakování.

Tradiční Waldův text má následující strukturu (pro jednoduchou hypotézu a pro jednoduchou alternativu):

1. Přijmeme hypotézu ***H***, pokud: $\prod_{i=1}^{n} \frac{f_A(x_i)}{f_H(x_i)} \le b$,
2. přijmeme alternativu ***A*** (přijmeme hypotézu ***H***), pokud: $\prod_{i=1}^{n} \frac{f_A(x_i)}{f_H(x_i)} \ge a$,
3. pokračujeme v pozorování (doplníme náhodný výběr), pokud: $b < \prod_{i=1}^{n} \frac{f_A(x_i)}{f_H(x_i)} < a$,

kde, čísla a, b jsou volena tak, aby byly dodrženy požadované hodnoty chyb pravděpodobností prvého a druhého druhu.

Pravděpodobnost chyby prvního druhu: $\alpha = P(zamitame\ H / H \text{ plati})$. α je pravděpodobností chyby prvního druhu = hladinou významnosti testu = velikostí kritického oboru. Pravděpodobnost chyby prvního druhu je též nazývána ***hladinou významnosti testu***.

Pravděpodobnost chyby druhého druhu: $1-\beta = P(zamitame\ \ A / A \text{ plati})$. Pravděpodobnost β je pak nazývána silou testu Síla kritického oboru (testu) je pravděpodobností zamítnutí hypotézy za platnosti alternativy.

Za daného typu úlohy je důležité číslo $N = \min\left\{n; \prod_{i=1}^{n} \frac{f_A(x_i)}{f_H(x_i)} \notin (b,a)\right\}$, tj. nejmenší počet kroků za který testování skončí.

Problémem je určit daná čísla a, b. Lze však pro ně dostat jejich aproximace $\hat{b} = \frac{1-\beta}{1-\alpha}; \hat{a} = \frac{\beta}{\alpha}$. Jejich odvození a diskusi lze nalézt v: Marie Hušková: Sekvenční analýza, skripta MFF-UK, Praha 1982, str. 10-11, 33-34 (*poznámka, tam je pravděpodobnost chyby druhého druhu značena* β) nebo C. Radhakrishna Rao: Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978, str. 518-520. Výhodou takové aproximace je, že nezávisí na typu rozdělení použitého pro formulaci hypotézy a alternativy. Naopak nevýhodou může být, že oproti teoretickým mezím může docházet k zvýšení potřebného počtu kroků k ukončení testu. Použití takové aproximace také dává omezení na možnou volbu hodnot chyb prvního a druhého druhu. Protože musí platit: $\hat{b} < \hat{a}$ je potřebné, aby bylo splněno $\frac{1-\beta}{1-\alpha} < \frac{\beta}{\alpha} \Leftrightarrow 0 < \alpha < \beta \le 1$. Tj. pravděpodobnost chyby prvního druhu je menší než síla testu. Taková podmínka nebude v praktických úlohách činit většinou problémy. Také použití aproximací $\hat{b} = \frac{1-\beta}{1-\alpha}; \hat{a} = \frac{\beta}{\alpha}$ v praktických úlohách nepovede k velkému zvýšení počtu kroků testu. Proto budeme nadále ztotožňovat $\hat{b} \triangleq b; \hat{a} \triangleq a$. V této přednášce nebudou řešeny problémy skončení testu a problémy existence „lepšího testu“. Zde je na místě odkázat např. na přednášky z předmětu SA1, nebo lépe: Marie Hušková: Sekvenční analýza, skripta MFF-UK, Praha 1982.

Waldův test hypotéz (y) o parametru alternativního rozdělení

Množina hodnot náhodné proměnné je, v tomto konkrétním případě, $\{0,1\}$. Pravděpodobnost takového rozdělení je pak: $P(x) = p^x (1-p)^{1-x}$. Poměr $\prod_{i=1}^{n} \frac{f_A(x_i)}{f_H(x_i)}$ bude $\prod_{i=1}^{n} \frac{p_A^{x_i}(1-p_A)^{1-x_i}}{p_H^{x_i}(1-p_H)^{1-x_i}}$

Rozhodovací pravidla v základním tvaru jsou:

1. Přijmeme hypotézu ***H***, pokud: $\prod_{i=1}^{n} \frac{p_A^{x_i}(1-p_A)^{1-x_i}}{p_H^{x_i}(1-p_H)^{1-x_i}} \le b$,
2. přijmeme alternativu ***A*** (zamítneme hypotézu ***H***), pokud: $\prod_{i=1}^{n} \frac{p_A^{x_i}(1-p_A)^{1-x_i}}{p_H^{x_i}(1-p_H)^{1-x_i}} \ge a$,
3. pokračujeme v pozorování (doplníme náhodný výběr), pokud: $b < \prod_{i=1}^{n} \frac{p_A^{x_i}(1-p_A)^{1-x_i}}{p_H^{x_i}(1-p_H)^{1-x_i}} < a$.

Uvedené nerovnosti lze zjednodušit na základě následujících konvencí a úvah:

$n_0 = |\{x_i : x_i = 0\}|; n_1 = |\{x_i : x_i = 1\}|; n_0 + n_1 = n$ a

$$\prod_{i=1}^{n} \frac{p_A^{x_i}(1-p_A)^{1-x_i}}{p_H^{x_i}(1-p_H)^{1-x_i}} \prec c \Leftrightarrow \lg \prod_{i=1}^{n} \frac{p_A^{x_i}(1-p_A)^{1-x_i}}{p_H^{x_i}(1-p_H)^{1-x_i}} \prec \lg c \Leftrightarrow \sum_{i=1}^{n} \left(x_i \lg\left(\frac{p_A}{p_H}\right) + (1-x_i)\lg\left(\frac{1-p_A}{1-p_H}\right)\right) \prec C,$$

kde $\prec \in \{<,>,\le,\ge\}$ a $C = \lg c$, ale

$$\sum_{i=1}^{n} \left(x_i \lg\left(\frac{p_A}{p_H}\right) + (1-x_i)\lg\left(\frac{1-p_A}{1-p_H}\right)\right) = n_1 \lg\left(\frac{p_A}{p_H}\right) + n_0 \lg\left(\frac{1-p_A}{1-p_H}\right).$$

Proto:

1. Přijmeme hypotézu ***H***, pokud: $n_1 \lg\left(\frac{p_A}{p_H}\right) + n_0 \lg\left(\frac{1-p_A}{1-p_H}\right) \le B$,
2. přijmeme alternativu ***A*** (zamítneme hypotézu ***H***), pokud: $n_1 \lg\left(\frac{p_A}{p_H}\right) + n_0 \lg\left(\frac{1-p_A}{1-p_H}\right) \ge A$,
3. pokračujeme v pozorování, pokud: $B < n_1 \lg\left(\frac{p_A}{p_H}\right) + n_0 \lg\left(\frac{1-p_A}{1-p_H}\right) < A$.

Odtud: za předpokladu $\frac{p_A}{p_H} \ne 1$, bude platit, že jedno z čísel $\lg\left(\frac{p_A}{p_H}\right) = L_1$, $\lg\left(\frac{1-p_A}{1-p_H}\right) = L_0$ bude kladné a druhé záporné.

Námět: Dokažte. Návod $\frac{p_A}{p_H} > 1 \Rightarrow \frac{1-p_A}{1-p_H} < 1$ a naopak.

Budeme řešit úlohu: **Pokud test dosud (=za *n* pozorování) nerozhodl, „kolik dalších pozorování máme doplnit“[1], aby test s dost velkou pravděpodobností (předem danou) rozhodl.**

Rozhodovací pravidla lze přepsat do tvaru:

1. Přijmeme hypotézu ***H***, pokud: $n_1L_1 + n_0L_0 \leq B$,
2. přijmeme alternativu ***A*** (zamítneme hypotézu ***H***), pokud: $n_1L_1 + n_0L_0 \geq A$,
3. pokračujeme v pozorování, pokud: $B < n_1L_1 + n_0L_0 < A$.

Dále budeme předpokládat, že $0 < p_H < p_A < 1$. Odtud plyne: $L_1 > 0$ a $L_0 < 0$.
Pro hypotézu $H \equiv p = p_H$ a alternativu $A \equiv p = p_A$ je situace zřejmá. Pro hypotézu $H \equiv p \leq p_H$ a alternativu $A \equiv p \geq p_A$ je situace analogická (Námět: zdůvodněte).
Po $n_0 + n_1 = n$ pozorováních (experimentech) tedy platí $B < n_1L_1 + n_0L_0 < A$.

Příklad (motivační, pro názornost):
$\alpha = 0{,}05;\ \ 1-\beta = 0.05;\ \ \ p_H = 0{,}025;\ \ p_A = 0{,}030$
Interpretace $p_H = 0{,}025$ je akceptovatelná pravděpodobnost produkce zmetku, $p_A = 0{,}030$ je již nepřijatelná pravděpodobnost. Meze rozhodnutí $a = \dfrac{\beta}{\alpha} = \dfrac{0{,}95}{0{,}05} = 19; b = \dfrac{1-\beta}{1-\alpha} = \dfrac{0{,}05}{0{,}95} = 0{,}05263$,
jejich logaritmy $A = \lg\dfrac{\beta}{\alpha} = \lg\dfrac{0{,}95}{0{,}05} = 2{,}944; B = \lg\dfrac{1-\beta}{1-\alpha} = \lg\dfrac{0{,}05}{0{,}95} = -2.944$
$L_1 = \lg\left(\dfrac{p_A}{p_H}\right) = \lg\left(\dfrac{0{,}030}{0{,}025}\right) = 0.18232$, $L_0 = \lg\left(\dfrac{1-p_A}{1-p_H}\right) = \lg\left(\dfrac{1-0{,}030}{1-0{,}025}\right) = -0.069$
Rozhodovací pravidla pro tento příklad:

1. Přijmeme hypotézu ***H***, pokud: $0{,}18232n_1 - 0.069n_0 \leq -2.944$,
2. přijmeme alternativu ***A*** (zamítneme hypotézu ***H***), pokud: $0{,}18232n_1 - 0.069n_0 \geq +2{,}944$,
3. pokračujeme v pozorování, pokud: $-2.944 < 0{,}18232n_1 - 0.069n_0 < +2{,}944$.

Interpretace testu (možná varianta realizace):

Potvrzovací, dvoustupňový, test:
Vybere se náhodný výběr o rozsahu $n = \left\lceil \dfrac{-2.944}{-0{,}069} \right\rceil = 43$, pokud mezi nimi není žádný zmetek (tedy $n = n_0$), test končí konstatováním, že „produkce zmetků“ je na akceptovatelné úrovni. Pokud platí $0{,}18232n_1 - 0.069n_0 \geq +2{,}944$, test končí konstatováním, že „produkce zmetků“ je na nepřijatelné úrovni. Pokud platí $-2.944 < 0{,}18232n_1 - 0.069n_0 < +2{,}944$, výběr se doplní nejmenším počtem produktů, tak aby byla potenciální možnost potvrzení akceptovatelné úrovně. Tj. hledá se nejmenší číslo n_{dop} pro které by platilo $0{,}18232n_1 - 0.069(n_0 + n_{dop}) \leq -2.944$,
tj. $n_{dop} = \left\lceil \dfrac{2{,}944}{0{,}069} + \dfrac{0{,}18232}{0{,}069}n_1 - n_0 \right\rceil = \left\lceil 42{,}\bar{6} + 2{,}64232n_1 - n_0 \right\rceil$. Výpočet n_{dop} je dán algoritmem v nějaké „mírně“ programovatelné kalkulačce nebo tabulkou. …. .

[1] Formulace je velice nepřesná ale názorná, počet nutných pozorování je náhodná proměnná.

Test končí konstatováním o ***akceptovatelné úrovni zmetků*** nebo konstatováním o ***nepřijatelné úrovni*** nebo „slabším výrokem“, že ***„akceptovatelnou úroveň zmetků“*** [2] ***nelze vyvrátit***.
Silný výrok o neakceptovatelnosti, znamená nepřijetí dodávky. Slabší výrok může být následován bezplatným doplněním dodávky o určitý počet kusů (**Námět:** pokuste se stanovit jak velký by měl být objem takového doplnění).
Exaktnější pojetí takového dvoustupňového testu je v tom, že druhý stupeň nebude Waldovým testem ale klasickým testem bez „intervalu neurčitosti“.

Následující tabulka demonstruje typy rozhodnutí, pro dané konkrétní zadání, podle počtu pozorovaných zmetků (vad) v prvním stupni testu:

Počet zmetků	Počet ne-zmetků	Hodnota kritéria	Výrok
0	43	-2,9670	**Akcept**
1	42	-2,7157	**Doplňkový test**
...	...	...	**Doplňkový test**
23	20	2,8134	**Doplňkový test**
24	19	3,0647	**Nepřij**
...	...	...	**Nepřij**
43	0	7,8398	**Nepřij**

Následuje tabulka rozsahu doplňkového výběru:

Počet zmetků v prvním stupni	Počet ne-zmetků v prvním stupni	Rozsah doplňkového výběru
1	42	4
2	41	7
3	40	11
4	39	15
5	38	18
6	37	22
7	36	26
8	35	29
9	34	33
10	33	37
11	32	40
12	31	44
13	30	48
14	29	51
15	28	55
16	27	58
17	26	62
18	25	66
19	24	69
20	23	73
21	22	77
22	21	80
23	20	84

[2] Akceptovatelná úroveň zmetků ≡ AQL≡Acceptable Quality Level.

Diskuze: V oblasti takových provozních testů existují dvě základní úlohy

1. Návrh takového testu

2. Analýza hotového testu. Zde např. součástí odběratelské smlouvy budou odpovídající tabulky pro přejímací test. Úlohou je zjistit (popř. překontrolovat) jakou (v testu implicitně vyjádřenou „kvalitu“) předpokládá a zdali jsme jako dodavatelé jeho požadavky splnit.

Bayesovsky inspirované modifikace

Klasické formulace takových Waldových testů mají jeden problém, jsou použitelné jen pro nějaké složené hypotézy a alternativy.
Složená parametrická hypotéza (např. $H \equiv p \le p_H$) může být „matematicky“ formulovaná jako jednoduchá hypotéza o „rozdělení pravděpodobnosti parametru“.
Tedy: rozdělení pozorovaných hodnot je reprezentováno hustotou (pravděpodobností, v diskrétním případě) závislou na parametru $f(x;p)$. Takovou hustotu můžeme chápat jako podmíněnou hustotu při daném (již náhodném) parametru $f(x;p) \overset{def}{=} f(x/p)$. Hypotézu (či alternativu) lze pak formulovat hustotou výskytu daného parametru $w_H(p)$. Např. pro $H \equiv p \le p_H$ bude jednou z přirozených formulací $w_H(p) = \frac{1}{p_H} \Leftrightarrow 0 < p < p_H, \, jinak \; w_H(p) = 0$, tj. rovnoměrné rozdělení hodnot parametru pro interval $0 < p < p_H$. Potom rozdělení za platnosti takové hypotézy. Bude rozdělení dat $f_H(x) = \int_{def\ p} f(x/p) w_H(p) dp$. Obdobně, pokud je alternativa formulována hustotou $w_A(p)$, dostáváme pro rozdělení dat, při platnosti alternativy $f_A(x) = \int_{def\ p} f(x/p) w_A(p) dp$. Aby byla taková úloha formulována korektně, je nutné, aby existovala množina $P^* \subset def\ p$, taková že, $p \in P^* \Rightarrow w_H(p) \ne w_A(p)$ a navíc platí $\int_{P^*} w_H(p) dp > 0$ nebo $\int_{P^*} w_A(p) dp > 0$. Tj. obě „parametrové hustoty“ se musí lišit na množině nenulové pravděpodobnosti.

Testové kritérium pak bude vypadat $\prod_{i=1}^{n} \frac{f_A(x_i)}{f_H(x_i)} = \prod_{i=1}^{n} \frac{\int_{def\ p} f(x_i/p) w_A(p) dp}{\int_{def\ p} f(x_i/p) w_H(p) dp}$. Rozhodovací pravidla se pak budou lišit podle toho, zda se jedná o Waldův, či jiný, sekvenční test nebo o test klasický.

Jinou, exaktněji formulovanou a analyzovanou modifikaci, lze nalézt v Hušková M.: Sekvenční analýza. Praha 1982, skripta MFFUK, str. 59-62.

Příklad: Test (y) o parametru alternativního rozdělení. Množina hodnot náhodné proměnné je, v tomto konkrétním případě, $\{0,1\}$. Pravděpodobnost takového rozdělení je pak: $P(x) = p^x (1-p)^{1-x}$. Klasicky formulovaná bodová hypotéza a alternativa nechť jsou $0 < p_H < p_A < 1$. Takovým formulacím mohou odpovídat „váhové parametrové hustoty“

$w_H(p) = \dfrac{1}{p_H} \Leftrightarrow 0 < p < p_H; w_H(p) = 0 \text{ jinak}$ a $w_A(p) = \dfrac{1}{1-p_A} \Leftrightarrow p_A < p < 1; w_A(p) = 0 \text{ jinak}$.

Potom: $f_H(x) = \left(\dfrac{p_H}{2}\right)^x \left(\dfrac{2-p_H}{2}\right)^{1-x}$ a $f_A(x) = \left(\dfrac{1+p_A}{2}\right)^x \left(\dfrac{1-p_A}{2}\right)^{1-x}$.

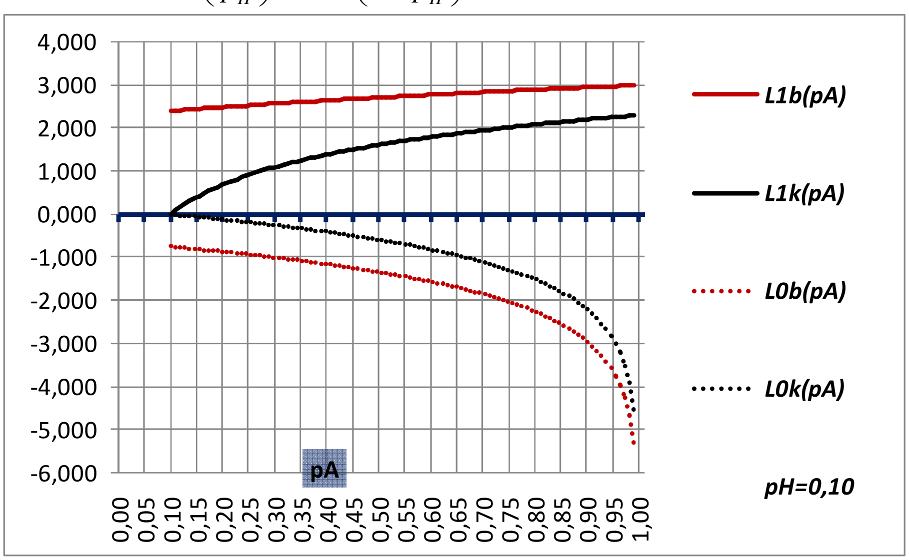
Testové kritérium pak bude:

$$V(x_1,\ldots,x_n) = \prod_{i=1}^{n} \frac{f_A(x_i)}{f_H(x_i)} = \prod_{i=1}^{n} \frac{\left(\dfrac{1+p_A}{2}\right)^{x_i} \left(\dfrac{1-p_A}{2}\right)^{1-x_i}}{\left(\dfrac{p_H}{2}\right)^{x_i} \left(\dfrac{2-p_H}{2}\right)^{1-x_i}} = \left(\frac{1+p_A}{p_H}\right)^{n_1} \left(\frac{1-p_A}{2-p_H}\right)^{n_0}$$ a

v logaritmickém vyjádření: $L(x_1,\ldots,x_n) = n_1 \lg\left(\dfrac{1+p_A}{p_H}\right) + n_0 \lg\left(\dfrac{1-p_A}{2-p_H}\right)$, tedy

$L(x_1,\ldots,x_n) = n_1 L_1 + n_0 L_0$, kde $L_1 = \lg\left(\dfrac{1+p_A}{p_H}\right)$ a $L_0 = \lg\left(\dfrac{1-p_A}{2-p_H}\right)$.

Námět: srovnejte $L_b(x_1,\ldots,x_n) = n_1 \lg\left(\dfrac{1+p_A}{p_H}\right) + n_0 \lg\left(\dfrac{1-p_A}{2-p_H}\right)$ v „bayesovském pojetí“ s klasickou formou $L_k(x_1,\ldots,x_n) = n_1 \lg\left(\dfrac{p_A}{p_H}\right) + n_0 \lg\left(\dfrac{1-p_A}{1-p_H}\right)$. Viz následující obrázek.



Testy hypotéz o parametru střední hodnoty, normálně rozděleného náhodného výběru

Klasický test: Testové kritérium bude:

$$V(x_1,\dots,x_n)=\prod_{i=1}^{n}\frac{f_A(x_i)}{f_H(x_i)}=\frac{\dfrac{1}{\left(\sqrt{2\pi}\sigma\right)^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_A)^2}}{\dfrac{1}{\left(\sqrt{2\pi}\sigma\right)^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu_H)^2}}=e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((x_i-\mu_A)^2-(x_i-\mu_H)^2\right)}$$ a jeho logaritmus:

$$L(x_1,\dots,x_n)=\lg V(x_1,\dots,x_n)=-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left((x_i-\mu_A)^2-(x_i-\mu_H)^2\right)=\frac{(\mu_A-\mu_H)}{2\sigma^2}\sum_{i=1}^{n}(2x_i-\mu_H-\mu_A)$$

$$\Rightarrow L(x_1,\dots,x_n)=\frac{n(\mu_A-\mu_H)}{2\sigma^2}\left(2\bar{x}-(\mu_A+\mu_H)\right)$$

Bayesovsky motivovaný test:

$$f_A(x)=\int_{def\ \mu}f(x;\mu)w_A(\mu)d\mu=\int_{-\infty}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}w_A(\mu)d\mu$$

Budeme déle pracovat s klasicky formulovanou hypotézou $H\equiv\mu>\mu_H$ a alternativou $A\equiv\mu<\mu_A;\mu_A<\mu_H$.

Jako „váhových hustot užijeme“ nevlastních hustot $w_A(\mu)=1\Leftrightarrow\mu<\mu_A;w_A(\mu)=0$ jinak a $w_H(\mu)=1\Leftrightarrow\mu>\mu_H;w_H(\mu)=0$ jinak . Proto:

$$f_A^*(x)=\int_{def\ \mu}f(x;\mu)w_A(\mu)d\mu=\int_{-\infty}^{\mu_A}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}d\mu=1-\Phi\left(\frac{x-\mu_A}{\sigma}\right)$$ a

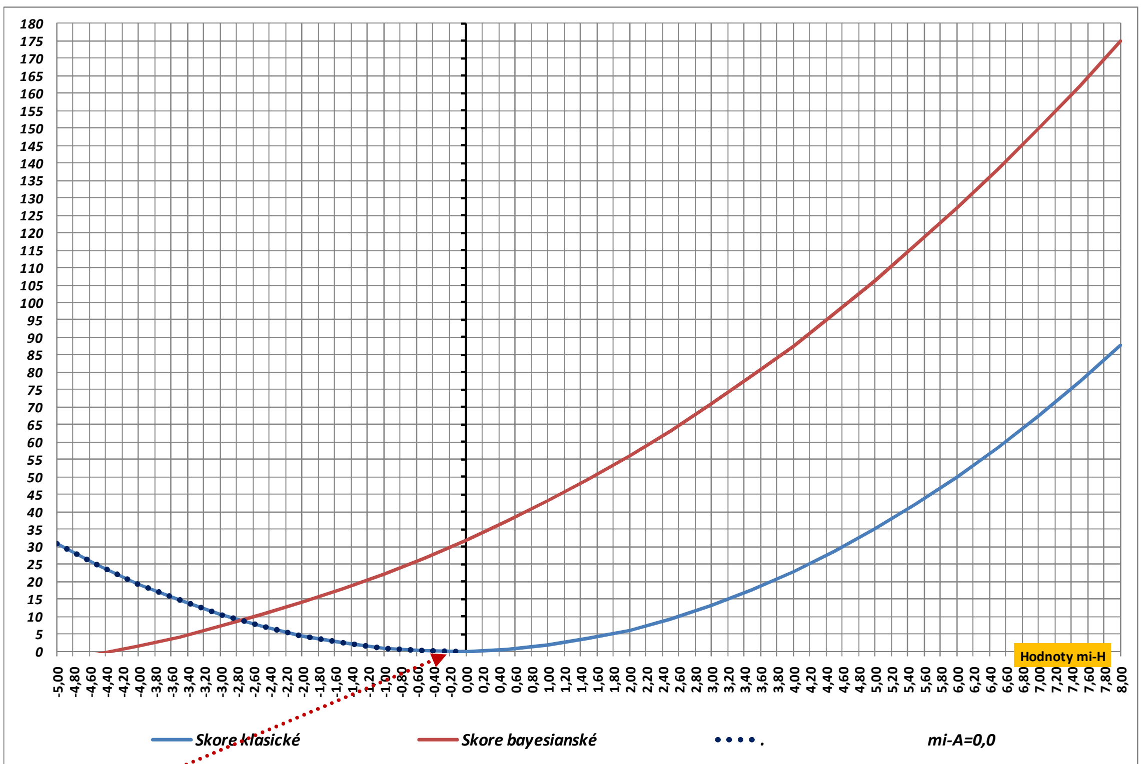
$$f_H^*(x)=\int_{def\ \mu}f(x;\mu)w_A(\mu)d\mu=\int_{\mu_H}^{+\infty}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}d\mu=\Phi\left(\frac{x-\mu_H}{\sigma}\right).$$ Za uvedených předpokladů

bude testové kritérium: $V(x_1,\dots,x_n)=\prod_{i=1}^{n}\frac{f_A^*(x_i)}{f_H^*(x_i)}=\prod_{i=1}^{n}\frac{1-\Phi\left(\frac{x_i-\mu_A}{\sigma}\right)}{\Phi\left(\frac{x_i-\mu_H}{\sigma}\right)}$ a jeho logaritmovaná

verze $L(x_1,\dots,x_n)=\sum_{i=1}^{n}\lg\frac{\Phi\left(\frac{x_i-\mu_A}{\sigma}\right)}{1-\Phi\left(\frac{x_i-\mu_H}{\sigma}\right)}=\sum_{i=1}^{n}\left[\lg\left(1-\Phi\left(\frac{x_i-\mu_A}{\sigma}\right)\right)-\lg\Phi\left(\frac{x_i-\mu_H}{\sigma}\right)\right]$

Názor na „klasifikační“ schopnost obou verzí logaritmovaných testových kritérií si lze udělat z následujícího obrázku, ve kterém „pozorovaná“ (generovaná) data mají následující parametry:

	Pozorování
Minimum	**-8,411**
Průměr	**-0,162**
Medián	**-0,605**
Maximum	**10,653**
StD	**4,269**
Počet	**48**



Průběh obou skóre v závislosti na formulaci hypotézy, při pevné alternativě. Hodnoty klasické skórovací funkce pod hodnotou pevné alternativy jsou nekorektní, mají zde demonstrovat možnost bayesovského přístupu testovat i „nedisjunktní“ hypotézu a alternativu.

Poznámka: Bayesovsky modifikovaný přístup k testování „umožňuje nahradit“ „porovnávání“ složené hypotézy a alternativy pomocí jednoduchých, formulovaných však ne v oblastech předpokládaného výskytu parametru ale v hustotách výskytu parametru. Ve ***většině*** reálných úloh se taková pojetí liší jen v matematickém přístupu při formulaci úlohy.

Doporučená a zdrojová literatura:

Jaroslav Hátle, Jiří Likeš	Základy počtu pravděpodobnosti a matematické statistiky. SNTL Praha 1974.
Alfréd Rényi	Teorie pravděpodobnosti, ACADEMIA, Praha 1972
C. Radhakrishna Rao	Lineární metody statistické indukce a jejich aplikace, ACADEMIA, Praha 1978
Machek J.	Teorie odhadu, SPN Praha 1974, skripta MFFUK
Hušková M.	Sekvenční analýza. Praha 1982, skripta MFFUK