Jaroslav Janko

Theorie ověřování statistických hypotes

Pokroky matematiky, fyziky a astronomie, Vol. 3 (1958), No. 1, 9--21

Persistent URL: http://dml.cz/dmlcz/137389

MATEMATIKA

THEORIE OVĚŘOVÁNÍ STATISTICKÝCH HYPOTES

Prof. J. JANKO, Praha

1. Úvodní pojmy a definice

Theorii ověřování statistických hypotes vybudovali především J. Neyman, E. S. Pearson a A. Wald. Áčkoliv je to dílo posledních třiceti let, patří již ke klasickým partiím matematické statistiky. Poněvadž není dosud v českém jazyku soustavnějšího pojednání o těchto otázkách, má tato práce poskytnout přehled o statistických principech, na nichž je tato theorie založena, s příslušnými odkazy na základní literaturu. Problém ověřování statistické hypotesy (v dalším krátce hypotesy) znamená rozhodnout na základě informace poskytnuté určitým výběrem, zda má být uvažovaná hypotesa zamítnuta nebo přijata. Předběžnou formulaci jeho uvedeme v této formě: Náhodný výběr, který tvoří n prvků $x_1, x_2, \ldots, x_n$, lze představiti v n-rozměrném Euklidovském prostoru bodem o souřadnicích $x_1, x_2, \ldots, x_n$. Tento prostor budeme nazývat *výběrovým prostorem* a bod, představující náhodný výběr, *výběrovým bodem.* Statistickou hypotesu lze vyjádřit zcela obecně takto: Nechť má konečná množina náhodných proměnných $X_1, X_2, \ldots, X_n$ sdruženou distribuční funkci $F(x_1, x_2, \ldots, x_n)$. *Statistická hypotesa je tvrzení*, že tato neznámá distribuční funkce je prvkem určité třídy ω distribučních funkcí. Theorie se omezuje na případ, kdy třída ω je částí k — parametrického svazku distribučních funkcí Ω, což znamená, že je známo o pravdivé, ale neznámé distribuční funkci, že je prvkem k — parametrického svazku funkcí $F(x_1, x_2, \ldots, x_n; \Theta_1, \Theta_2, \ldots, \Theta_k)$ kde $\Theta_1, \Theta_2, \ldots, \Theta_k$ jsou parametry. Specifikací hodnot těchto parametrů je distribuční funkce F úplně určena. Množina k hodnot parametrů může býti představena bodem v k-rozměrném prostoru Euklidovském, který se nazývá *parametrickým prostorem.* Poněvadž vzájemné přiřazení prvků z Ω a bodů parametrického prostoru je jednoznačné, lze identifikovat Ω s parametrickým prostorem. Problém se řeší tím, že se určí ve výběrovém prostoru určitý obor tak, že hypotesa se zamítne, když výběrový bod padne do tohoto oboru nazvaného *kritickým oborem* pro danou hypotesu. Je úkolem určit kritický obor tak, aby byla co nejmenší pravděpodobnost, že výběrový bod padne do kritického oboru, když je ověřovaná hypotesa pravdivá, a co největší v případě, když je nepravdivá. Jedná-li se na př. o hypotesu, že náhodné proměnné $X_1, X_2, \ldots, X_n$ jsou nezávislé a mají všechny totéž normální rozdělení s průměrem nulovým $\mu = 0$ a rozptylem jednotkovým $\sigma^2 = 1$, čili rozdělení $N(0,1)$, má třída ω jediný prvek. V takovém případě říkáme, že uvažovaná hypotesa je jednoduchou; jinak se nazývá složenou. Kdybychom tedy tvrdili jen, že v normálním rozdělení $\mu = 0$ a nespecifikovali σ^2, byla by to hypotesa složená.

Problém ověřování hypotesy pak znamená, že na základě získaných pozorování náhodných proměnných máme rozhodnout, zda se má přijmout či zamítnout hypotesa H, že neznámá distribuční funkce $F(x_1, x_2, \ldots, x_n; \Theta_1, \ldots, \Theta_k)$ patří do třídy $\omega \subset \Omega$. Rčením, že funkce $F(x_1, x_2, \ldots, x_n; \Theta_1, \ldots, \Theta_k)$ je prvkem určité třídy Ω distribučních funkcí, vyjadřujeme své vědomosti o této neznámé distribuční funkci. Na základě těchto vědomostí nejčastěji předpokládáme, že náhodné proměnné $X_1, X_2, \ldots, X_n$ jsou nezávislé a mají totéž rozdělení, jehož typ je znám. Podle toho pak má tato funkce tvar

$$F(x_1, x_2, \ldots, x_n; \Theta_1, \ldots, \Theta_k) = F_1(x_1; \Theta_1, \ldots, \Theta_k)\, F_2(x_2; \Theta_1, \ldots, \Theta_k) \ldots F_n(x_n; \Theta_1, \ldots, \Theta_k)\,,$$

při čemž $F_i(x; \Theta_1, \ldots, \Theta_k) = F_j(x; \Theta_1, \ldots, \Theta_k)$ pro všechna i, j a x.

Pravidlo, které předepíše jednoznačně rozhodnutí o přijetí či zamítnutí ověřované hypotesy pro každý možný bod výběrového prostoru je *testem čili ověřením* statistické hypotesy H. Je to pravidlo induktivního způsobu reagování na popudy vnějšího světa čili krátce pravidlo induktivního závěru. Rozhodneme-li na základě takového pravidla přijmout hypotesu H, neznamená to, že jsme přesvědčeni, že je pravdivá; znamená to jen, že hypotesa H není v rozporu s pozorovanými daty. Zamítneme-li však podle pravidla induktivního závěru hypotesu H v důsledku toho, že pravidlo toto rozhodnutí předpisuje, znamená to, že hypotesa H je velmi pravděpodobně nesprávná, t. j., budeme-li postupovat podle tohoto pravidla, stane se nám velmi zřídka, že zamítneme správnou hypotesu.

Příklad: Vezměme v úvahu jednoduchý příklad, kdy se předpokládá, že víme, že neznámá distribuční funkce je prvkem jednoparametrického svazku $F(x_1, \ldots, x_n; \Theta)$, dále je známo, že náhodné proměnné $X_1, X_2, \ldots, X_n$ jsou nezávislé a mají normální rozdělení s týmž průměrem a jednotkovým rozptylem. Pak lze psát

$$F(x_1, x_2, \ldots, x_n; \Theta) = \frac{1}{(2\pi)^{n/2}} \int_{-\infty}^{x_1} \exp\left\{ - \frac{(u - \Theta)^2}{2} \right\} du \ldots \int_{-\infty}^{x_n} \exp\left\{ - \frac{(u - \Theta)^2}{2} \right\} du\,.$$

Úkolem jest ověřiti hypotesu $H : \Theta = 0$.

V theorii pravděpodobnosti bylo obvyklým postupem, že se zamítla tato hypotesa tehdy a jen tehdy, jestliže $\left| \frac{x_1 + x_2 + \ldots + x_n}{n} \right| \geq h$. Konstanta h se zvolí tak, aby pravděpodobnost $P\left\{ \left| \frac{1}{n}(x_1 + x_2 + \ldots + x_n) \right| > h \right\}$ za předpokladu, že testovaná hypotesa je pravdivá, byla tak malá, abychom byli ochotni testovanou hypotesu zamítat. Považuje-li se za vyhovující na př. pravděpodobnost $P = 0{,}05$, pak je $h = \frac{1{,}96}{\sqrt{n}}$. Kritický obor se snadno znázorní, jedná-li se o výběr jen dvou pozorování x_1, x_2, takže výběrovým prostorem je Euklidovská rovina a kritický obor tvoří všechny body, pro něž je jednak $\frac{1}{2}(x_1 + x_2) > \frac{1{,}96}{\sqrt{2}}$, jednak $\frac{1}{2}(x_1 + x_2) < \frac{-1{,}96}{\sqrt{2}}$.

Zamítne se tedy hypotesa, že aritmetický průměr je roven nule, jestliže průměr dvou pozorování je větší než $\frac{1{,}96}{\sqrt{2}}$ nebo menší než $\frac{-1{,}96}{\sqrt{2}}$. Při tomto postupu se neodůvodňuje, proč se má užít právě tohoto kritického oboru. Dociluje se pouze toho, že pravděpodobnost, že výběrový bod padne do kritického oboru, je 0,05, když ověřovaná hypotesa je pravdivá. Takových oborů je však nekonečně mnoho.

2. Dva typy chyb. Síla kritického oboru

K rozlišení různých kritických oborů dospěli Neyman a Pearson úvahou o možných chybách, kterých se můžeme dopustit, když na základě kritického oboru rozhodujeme. Předpokládejme, že podle našeho pravidla induktivního reagování na popudy vnějšího světa se má dát přednost určité akci a_1 před druhou a_2 vždy, když hypotesa H je na základě pozorování přijata za pravdivou a druhé akci před prvou, když alternativní hypotesa $\overline{H}$ je přijata za pravdivou. Ověřovaná hypotesa H značí na př., že $\Theta = \Theta_0$, což je určitý bod parametrického prostoru a akce, které se má dát přednost, když hypotesa H je pravdivou, je přijetí hypotesy H, kdežto druhá akce je zamítnutí hypotesy H, když pravdivá je hypotesa $\overline{H}$ čili $\Theta \neq \Theta_0$. *Zamítneme-li hypotesu H, když je pravdivá,* dopouštíme se chyby typu I; *přijmeme-li ji, když je nepravdivá,* dopouštíme se chyby typu II. Řešení problému ověřování statistické hypotesy vychází z té prosté myšlenky, která je vyjádřena požadavkem, abychom upravili svůj test tak, aby minimalisoval pravděpodobnost chyb. Praktické vyhovění tomuto požadavku vedlo Neymana a Pearsona k tomu, aby formulovali základní princip své theorie takto: Mezi všemi kritickými obory (t. j. obory zamítnutí hypothesy H) o téže pravděpodobnosti α chyby typu I, čili kritickými obory téhož rozsahu, jest zvoliti ten obor, pro který je pravděpodobnost chyby typu II minimální.

Při konkretním užívání tohoto principu je obtíž v tom, že pravděpodobnost chyby typu II závisí na pravdivém parametrickém bodu Θ. To znamená, že je-li dán určitý kritický obor, pak pravděpodobnost chyby typu II bude funkcí pravdivého parametrického bodu Θ. Tento bod však neznáme, proto chceme užíti toho kritického oboru, pro který je minimální pravděpodobnost chyby typu II vzhledem ke všem možným alternativním hypothesám $\Theta \neq \Theta_0$. Jestliže takový společný nejlepší kritický obor existuje, je problém rozřešen. Takové případy jsou však dosti výjimečné.

Objasněme si tedy postup statistikův při ověřování statistické hypothesy čili při vybírání kritických oborů. Úkol řešíme tak, abychom se co nejlépe vyhnuli možným chybám. Rozhodneme, kterou chybu ze dvou typů chyb je třeba v daném problému považovat za těžší, takže je důležitější se jí vyhnout. Upravíme pak test tak, že testovanou hypotesou bude ta, jejíž zamítnutí, když je pravdivá, je chybou typu I. Prvním požadavkem nyní je, abychom se této chybě typu I co nejčastěji vyhnuli, čili aby test zamítl velmi zřídka testovanou hypotesu, když je pravdivá.

Je-li možno splnit tuto hlavní podmínku několika způsoby, volíme ten, při kterém bude minimální pravděpodobnost, že bude přijata testovaná hypotesa, když je nepravdivá. Začneme obvykle s tím, že stanovíme pevně určité malé číslo α (obyčejně z intervalu $0 < \alpha < 0{,}2$), které nazýváme *hladinou významnosti,* a požadujeme, aby pravděpodobnost, že se dopustíme chyby prvního typu, nepřekročila toto číslo α. Říkáme, že kritický obor, který vyhovuje této podmínce, odpovídá hladině významnosti α.

Když byla zvolena hladina významnosti α, vznikají dva matematické problémy. První spočívá v určení všech možných kritických oborů, které odpovídají této hladině významnosti. Když byl tento problém vyřešen, je třeba z těchto kritických oborů vybrat jeden, který bude nejlépe kontrolovat chyby typu II. Označme tedy w nějaký obor ve výběrovém prostoru a E budiž ten bod výběrového prostoru, který představuje pozorování $x_1, \ldots, x_n$. $P\{w|\Theta'\}$

bude pravděpodobnost, že bod E je v oboru w, počítaná *za předpokladu, že Θ' je pravdivá hodnota* neznámého parametru Θ. Tato pravděpodobnost je vyjádřena integrálem $\int_w \mathrm{d}F(x_1, \ldots, x_n, \Theta')$ přes obor w. Je-li $\Theta = \Theta_0$ ověřovaná hypotesa a zvolíme w za kritický obor pro hypotesu, bude dán rozsah toho kritického oboru výrazem $P\{w|\Theta_0\} = \alpha$. Není-li tato hypotesa pravdivá a jiná hodnota $\Theta = \Theta'$ je pravdivá, potom $P\{w|\Theta'\}$ je pravděpodobnost, že bod E je v kritickém oboru, počítaná za předpokladu, že Θ' je pravdivá hodnota neznámého Θ čili pravděpodobnost, že test založený na kritickém oboru w zamítne testovanou hypothesu Θ_0, když je pravdivá hypothesa Θ', t. j. že neuděláme chybu typu II. Pravděpodobnost $P\{w|\Theta'\}$ je doplňkem na jedničku pravděpodobnosti, že uděláme chybu typu II; nazýváme ji *silou kritického oboru w* vzhledem k alternativní hypothese $\Theta = \Theta'$. Čím je větší $P\{w|\Theta'\}$, tím je kritický obor w lépe vyhovující. Můžeme říci, že velká pravděpodobnost $P\{w|\Theta'\}$ znamená, že kritický obor w má velkou objevnou sílu. V tom případě, kdy existují dva kritické obory w a w', které odpovídají téže hladině významnosti α a při tom $P\{w'|\Theta'\} < P\{w|\Theta'\}$, říkáme, že kritický obor w je silnější než w', je-li pravdivá hypothesa Θ'.

Vzhledem k tomu, že pravděpodobnost $P\{w|\Theta\}$ je funkcí Θ, nazývá se silofunkcí kritického oboru w. Může býti graficky znázorněna křivkou, jejíž pořadnice se rovná rozsahu kritického oboru w, je-li úsečka rovna hodnotě Θ_0; rovná se pak síle kritického oboru w vzhledem k alternativní hypothese $\Theta = \Theta'$, je-li úsečka rovna hodnotě $\Theta' \neq \Theta_0$. Této křivce se pak říká silokřivka oboru w. Pro konkretní případ kritického oboru $|\bar{x}| > \frac{1,96}{\sqrt{n}}$ lze silokřivku snadno počítat.

3. Stejnoměrně silnější kritický obor

Mějme dva kritické obory w a w', které mají požadovaný rozsah α, takže je stejná pravděpodobnost dopustit se chyby typu I, ať se užije w nebo w'. Jestliže však se užije kritického oboru w, bude pravděpodobnost, že uděláme chybu typu II, menší, než užije-li se w'. Probíhá-li silokřivka oboru w nad silokřivkou oboru w' pro každé Θ kromě Θ_0, kde podle předpokladu mají obě společný bod, pak kritický obor w se nazývá *stejnoměrně silnějším* než w'. Test, který by užil kritického oboru w', pak se nazývá *nepřípustným*, poněvadž jeho užití je za všech okolností nepříznivější než užití kritického oboru w. To se obvykle objasňuje úvahou o dvou statisticích S a S', kteří ověřují touž hypotesu a každý z nich užívá k tomu téhož velkého počtu ν náhodných výběrů se stejným rozsahem n pozorování. Předpokládá se, že statistik S užívá k testování kritického oboru w a statistik S' užívá oboru w'. Otázkou je, zda má býti testovaná hypotesa zamítnuta a každý z nich dostane tedy na ni ν odpovědí, z nichž některé budou správné a jiné budou nesprávné. Je nutné rozlišit dva případy.

První případ je ten, kdy testovaná (čili nulová) *hypotesa je pravdivá* a druhý kdy je *nepravdivá*. V prvním případě dostane každý z těch dvou statistiků odpověď buď, že hypotesa má býti přijata — což jsou odpovědi správné — nebo, že hypotesa má být zamítnuta — což jsou odpovědi s chybou typu I. Pravděpodobnost, že uděláme chybu typu I při testování nulové hypotesy pomocí náhodného výběru, se rovná rozsahu α kritického oboru užitého k testování. Je-li počet výběrů velký, pak se očekává, že relativní četnost chyb typu I bude přibližně rovna jejich pravděpodobnosti α, čili rozsahu kritického oboru.

Poněvadž oba obory w a w' mají podle předpokladu stejný rozsah, udělá každý z obou statistiků přibližně týž počet chyb typu I.

V druhém případě, kdy je testovaná *hypotesa nepravdivá*, některé odpovědi ji správně zamítnou, kdežto ostatní ji přijmou, takže udělají chyby typu II. Poněvadž ν je velké, bude relativní četnost správných zamítnutí přibližně rovna síle kritického oboru, což je pravděpodobnost nedopustiti se chyby typu II. Podle předpokladu je síla oboru w větší než síla oboru w' bez ohledu na to, jaká je pravdivá hodnota Θ, jen když $\Theta \neq \Theta_0$. Bude tedy relativní četnost nesprávných odpovědí u statistika S menší než relativní četnost nesprávných odpovědí u S'. V případě, kdy je testovaná hypothesa nepravdivá — ať je pravdivá hodnota Θ jakákoliv — udělá statistik S méně nesprávných rozhodnutí, kdežto je-li testovaná hypothesa pravdivá, udělají oba statistikové přibližně stejný počet nesprávných rozhodnutí. Užití kritického oboru w je tudíž lepším postupem, než užití w'. Tyto úvahy tedy rozhodují o volbě mezi dvěma kritickými obory stejného rozsahu, je-li jeden z nich stejnoměrně silnější než druhý.

4. Stejnoměrně nejsilnější kritický obor

Probíhá-li však silokřivka oboru w nad silokřivkou obou w' pro některé hodnoty Θ a pro jiné zase pod ní, pak není možno zvoliti jeden z těchto dvou oborů, dokud nezavedeme další principy, na nichž bychom založili rozhodnutí. Je-li silokřivka určitého oboru w pro všechny hodnoty Θ nad silokřivkou každého jiného oboru w' stejného rozsahu, pak obor w se nazývá *stejnoměrně nejsilnějším kritickým oborem* a příslušný test k oboru w je stejnoměrně nejsilnějším testem.

Dospíváme tak k prvnímu principu pro výběr určitého testu: *kdykoliv lze najít stejnoměrně nejsilnější test, dá se mu přednost přede všemi jinými testy, které užívají kritických oborů téhož rozsahu.*

Můžeme odvodit nutnou a postačující podmínku pro obor w, aby byl nejlepším kritickým oborem pro testovanou hypotesu H vzhledem k alternativní hypotese $\overline{H}$. Zvolili jsme kritický obor rozsahu α, takže

$$P(w|\Theta) = \int_w \mathrm{d}F(x_1, \ldots, x_n|\Theta) = \alpha \tag{4.1}$$

Chceme pak, aby byla maximem pravděpodobnost $P(w|\Theta')$, čili aby byl maximem $\int_w \mathrm{d}F(x_1, \ldots, x_n|\Theta')$. Uvažujme jednoduchou hypotesu. Tento problém variačního počtu je ekvivalentní se stanovením nepodmíněného maxima integrálu $\int_w (\mathrm{d}F(x_1, \ldots, x_n|\Theta') - k\, \mathrm{d}F(x_1, \ldots, x_n|\Theta)$ nebo, což je totéž, stanovit minimum $\int_w \left(\mathrm{d}F(x_1, \ldots, x_n|\Theta) - \frac{1}{k}\, \mathrm{d}F(x_1, \ldots, x_n|\Theta')\right)$, kde k je konstanta, kterou máme určit ze vztahu (4.1). Podmínka stacionární hodnoty tohoto integrálu je, aby na hranici oboru w bylo

$$\mathrm{d}F(x_1, \ldots, x_n|\Theta) - \frac{1}{k}\, \mathrm{d}F(x_1, \ldots, x_n|\Theta') = 0 \tag{4.2}$$

Je-li toto řešení minimem, pak je uvnitř oboru w

$$\mathrm{d}F(x_1, \ldots, x_n|\Theta) \leqq \frac{1}{k}\, \mathrm{d}F(x_1, \ldots, x_n|\Theta') \tag{4.3}$$

a vně oboru

$$\mathrm{d}F(x_1, \ldots, x_n|\Theta) > \frac{1}{k}\, \mathrm{d}F(x_1, \ldots, x_n|\Theta') \tag{4.4}$$

5. Existence stejnoměrně nejsilnějších testů a vyčerpávajících (suficientních) charakteristik

S hlediska vlastností distribučních funkcí je důležitou otázka, jaké závěry vyplývají o vyčerpávajících (suficientních) charakteristikách z existence stejnoměrně nejsilnějších testů, po případě obráceně. K jejímu objasnění zavedeme podle [9] definice a vlastnosti vyčerpávajících charakteristik.

Definice I. *Určitá funkce T náhodných proměnných $X_1, X_2, \ldots, X_n$ se nazývá charakteristikou*[1]*, má-li tyto vlastnosti:*

a) T je měřitelná funkce;

b) existují takové hodnoty T', že množina bodů $\mathbf{W}(T')$, v níž $T = T'$ je alespoň $(n - 1)$ rozměrná, t. j. počet rozměrů ze o jednu menší než počet rozměrů výběrového prostoru $\mathbf{W}$;

c) T nezávisí na neznámých parametrech, které mohou být obsaženy v zákonu rozdělení pravděpodobnosti.

Definice je formulována tak, aby zahrnovala obvykle užívané charakteristiky. Nebylo by vhodné definovat charakteristiku jen pro spojité funkce, neboť by pak byl vyřazen na př. Studentův poměr $z = \frac{\bar{x}}{s}$ ($\bar{x}$ značí průměr a s^2 rozptyl), který není spojitý a nemá smyslu v žádném bodě přímky $X_1 = X_2 = \ldots = X_n$. Vlastnost a) se považuje za potřebnou z toho důvodu, že by neměly praktickou cenu charakteristiky, pro které není možno počítat zákony rozdělení pravděpodobnoti. Co se týče vlastnosti b), řekneme jen toto: Kdybychom v definici charakteristiky požadovali, aby množina $\mathbf{W}(T')$ měla pro každou hodnotu T' $(n - 1)$ dimensí, vyloučili bychom tím určitý počet obvykle užívaných charakteristik. Na př. množina bodů, v nichž $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = T'$, je prázdná pro každé $T' < 0$, a množina bodů, v nichž $s^2 = 0$, která je představena přímkou $X_1 = X_2 = \ldots = X_n$, má jen jednu dimensi. Také by bylo nevhodné požadovat, aby množina $\mathbf{W}(T)$ měla přesně $(n - 1)$ dimensí, neboť v řadě praktických úloh uvažujeme charakteristiky, kde ta množina je n-rozměrná. To lze ilustrovat případem, kde máme n nezávislých pozorování $x_1, x_2, \ldots, x_n$ jedné náhodné proměnné, která má normální zákon rozdělení se známou směrodatnou odchylkou. K testování hypotés o průměru tohoto zákona rozdělení a k jeho odhadu se někdy užívá charakteristiky T, která se rovná počtu těch x_i, která jsou větší než určitá pevná hodnota x. Volíme ji obyčejně tehdy, když není možno určit skutečné hodnoty x_i, ale lze stanovit počet těch, které překračují hodnotu x. Množina $\mathbf{W}(T)$ ve výběrovém prostoru, v němž je tato charakteristika konstantní, je n-rozměrná, čili má týž počet rozměrů jako výběrový prostor. Objasní se to snadno případem $n = 2$. Výběrový prostor je pak představen rovinou X_1, X_2 a charakteristika T nabývá konstantních hodnot

I/0 pro $X_1 \leqq x$, $X_2 \leqq x$,
II/1 pro $X_1 \leqq x$, $X_2 > x$ nebo $X_1 > x$, $X_2 \leqq x$,
III/2 pro $X_1 > x$, $X_2 > x$.

Tyto čtyři množiny jsou všechny dvojrozměrné. Vlastnost c) rozlišuje mezi neznámým parametrem Θ a jeho určitou hodnotou $\Theta = \Theta_0$, která je specifiko-

[1]) Někdy se užívá také názvu statistika, což není vhodné vzhledem k tomu, že je již vyhrazen pro celou vědu.

vána určitou hypotésou. Tím má býti jen zajištěno, aby bylo možno počítat hodnotu T ve skoro všech bodech výběrového prostoru. Vezměme na př. výraz $z = \frac{\bar{x} - \Theta}{s}$, v němž Θ je neznámý průměr náhodné proměnné X základního souboru, při čemž $\bar{x}$ je průměr a s^2 je rozptyl ve výběru; pak nemůžeme počítat hodnotu z, ačkoliv $\bar{x}$ a s jsou hodnoty známé. Nebude se tedy takový výraz nazývat charakteristikou. Dosadíme-li však do z místo „neznámého parametru" Θ jeho numerickou hodnotu, třeba $\Theta_0 = 1$, jak ji specifikuje určitá hypotesa, pak výraz $z = \frac{\bar{x} - 1}{s}$ se stane funkcí hodnot X, která je určena v každém bodě výběrového prostoru (s výjimkou množiny míry nula, pro níž $x_1 = x_2 = \ldots = x_n$) a bude se nazývat charakteristikou. Dále bude třeba zavésti definice specifické vyčerpávající charakteristiky a děleně vyčerpávající charakteristiky.

Definice II. *Charakteristika T_1 se nazývá specifickou vyčerpávající charakteristikou vzhledem k parametru Θ_1, jestliže pro každou charakteristiku T_2 je podmíněné rozdělení T_2 při daném T_1 nezávislé na Θ_1.*

Definice III. *Charakteristika T_1 se nazývá děleně vyčerpávající charakteristikou parametrů $\Theta_1, \ldots, \Theta_q$, jestliže pro každou charakteristiku T_2 je podmíněné rozdělení T_2 při daném T_1 nezávislé na q parametrech $\Theta_1, \ldots, \Theta_q$, zatím co závisí na zbývajících $l - q$ parametrech $\Theta_{q+1}, \ldots, \Theta_l$.*

Pojem děleně vyčerpávající charakteristiky je užší než pojem specifické vyčerpávající charakteristiky. Je-li totiž T_1 děleně vyčerpávající charakteristika třeba dvou parametrů Θ_1 a Θ_2, pak vyhoví definici specifické vyčerpávající charakteristiky parametru Θ_1 nebo Θ_2 vzaté odděleně. Na druhé straně charakteristika, o níž je známo, že je specificky vyčerpávající vzhledem ku Θ_1, nemusí býti děleně vyčerpávající charakteristikou parametru Θ_1 a Θ_2.
Uvedeme nyní dvě věty, které dokázal Neyman [9] o specificky vyčerpávajících a děleně vyčerpávajících charakteristikách.

Věta 1. *Nutná a postačující podmínka, aby nějaká charakteristika byla specificky vyčerpávající vzhledem k parametru Θ (ve smyslu definice II) je možnost v každém bodě výběrového prostoru, až snad na množinu míry nula, uvést zákon rozdělení pravděpodobnosti proměnných X_i ve tvar součinu*

$$p(x_1, \ldots, x_n|\Theta) = p(T|\Theta)\, \Phi(x_1, \ldots, x_n)|_{T=T(x_1, \ldots, x_n)},$$

kde $p(T|\Theta)$ značí zákon rozdělení pravděpodobnosti T a Φ je funkcí proměnných x_i nezávislou na Θ, nebo ve tvar součinu

$$p(x_1, \ldots, x_n|\Theta) = f(T, \Theta)\, \Phi(x_1, \ldots, x_n),$$

kde funkce f závisí na T a Θ a Φ neobsahuje Θ.

Věta 2. *Nutná a postačující podmínka pro to, aby nějaká charakteristika T byla děleně vyčerpávající charakteristikou parametrů $\Theta_1, \ldots, \Theta_q$ je možnost ve skoro každém bodě prostoru* **W** *uvést zákon rozdělení pravděpodobnosti náhodných proměnných ve tvar součinu*

$$p(x_1, \ldots, x_n|\Theta_1, \ldots, \Theta_q, \ldots, \Theta_l) =$$
$$= p(T|\Theta_1, \ldots, \Theta_q)\, \Phi(x_1, \ldots, x_n; \Theta_{q+1}, \ldots, \Theta_l)|_{T=T(x_1, \ldots, x_n)},$$

kde $p(T|\Theta_1, \ldots; \Theta_q)$ je zákon rozdělení pravděpodobnosti T a funkce Φ nezávisí na $\Theta_1, \ldots, \Theta_q$.

Z theorie vyčerpávajících charakteristik a stejnoměrně nejsilnějších testů [9] pak vyplynuly tyto výsledky:

a) Existuje-li systém stejnoměrně nejsilnějších testů a jsou-li splněny určité podmínky, pak musejí také existovat vyčerpávající charakteristiky buď specificky nebo děleně.

Určitými podmínkami se rozumí především, že systém testů je uspořádaný. Označíme-li $w(\alpha)$ nejlepší kritický obor rozsahu α, který odpovídá některému testu ze systému, pak systém se nazývá uspořádaný, jestliže pro všechna $0 < \alpha_1 < \alpha_2 \leqq \alpha_0$ je nejlepší kritický obor $w(\alpha_1)$ obsažen ve $w(\alpha_2)$. Dále se tím rozumí, že hraniční prostor sdružený s tímto systémem je skoro identický s výběrovým prostorem. Hraniční prostor odpovídající tomuto systému je množina bodů $\mathbf{E}$ majících tu vlastnost, že každý náleží do kladné hranice aspoň jednoho nejlepšího kritického oboru sdruženého s testy obsaženými v systému. Bod E patří do kladné hranice kritického oboru $w(\alpha)$, patří-li do jeho kladné části, kde totiž $p(x_1, x_2, \ldots, x_n | \Theta_1^0, \Theta_2^0, \ldots, \Theta_l^0) > 0$, a jestliže v každém okolí bodu E existuje aspoň jeden bod E', patřící do $w(\alpha)$ a aspoň jeden bod E'', ležící vně $w(\alpha)$, které jsou oba různé od E. Symboly $\Theta_1^0, \Theta_2^0, \ldots, \Theta_l^0$ značí hodnoty parametrů určené hypotesou.

b) Existuje-li systém stejnoměrně nejsilnějších testů, nemusí vůbec existovat vyčerpávající charakteristika.

c) Když existuje vyčerpávající charakteristika, může ale nemusí existovat systém stejnoměrně nejsilnějších testů.

6. Hypotesy složené

Uvažujeme-li hypotesy složené, pak mějme testovanou hypotesu, která závisí na $r + s$ parametrech $\Theta_1, \ldots, \Theta_r, \Theta_{r+1}, \ldots, \Theta_{r+s}$.

Posledních s parametrů je specifikováno a hypotesa určuje distribuční funkci nehledíc k r nespecifikovaným parametrům. Je to složená hypotesa o r stupních volnosti. Na př. hypotesa, že základní soubor je normální se specifikovaným průměrem, při čemž se nic nepředpokládá o jeho rozptylu, je složená s jedním stupněm volnosti. Předpokládá se, že každá přípustná jednoduchá alternativní hypotesa je dána specifikováním ostatních r parametrů a že existuje společný výběrový prostor pro všechny alternativy.

Abychom provedli test složené hypotesy H, musíme především kontrolovat chyby prvního typu stanovením kritického oboru tak, aby

$$\int_w \mathrm{d}F(x_1, \ldots, x_n | \Theta_1, \ldots, \Theta_r, \Theta_{r+1}, \ldots, \Theta_{r+s}) = \alpha\,. \qquad (6.1)$$

Tento případ se ovšem liší od jednoduchého tím, že F se může měnit vzhledem k neznámým parametrům. Abychom tedy byli jisti, že máme chybu pod kontrolou, musíme býti s to najít w tak, aby tento vztah (6.1) platil, ať jsou $\Theta_1, \ldots, \Theta_r$ jakékoliv. Můžeme-li takový obor najít, nazýváme jej oborem podobným výběrovému prostoru a α je jeho rozsah.

Problém testování složených hypotes pak spočívá v tom, že najdeme podobné obory a pak z nich vybereme jeden, který minimalisuje chybu druhého typu pro nějakou jednoduchou přípustnou alternativní hypotesu. Je-li tento obor týž pro všechny přípustné jednoduché hypotesy, je to společný nejsilnější kritický obor.

Pišme tedy zákon rozdělení pravděpodobnosti náhodných veličin $X_1, X_2, \ldots, X_n$ ve tvaru $f(X_1, X_2, \ldots, X_n; \Theta_1, \Theta_2, \ldots, \Theta_r; \Theta_{r+1}, \Theta_{r+2}, \ldots, \Theta_{r+s})$. Složená hypotesa o r stupních volnosti je každá hypotesa $H(\Theta_{r+1} = \Theta_{r+1}^0, \Theta_{r+2} =$

$= \Theta^0_{r+2}, \ldots, \Theta_{r+s} = \Theta^0_{r+s}$), která zahrnuje s parametrů Θ_{r+j}($j = 1, 2, \ldots, s$), ponechává však volnými r parametrů Θ_i ($i = 1, 2, \ldots, r$). Je to tedy hypotesa, která se týká individuelně s parametrů, při čemž nezáleží na tom, jaké jsou ostatní r parametry. Je-li $r = 0$, pak $H(\Theta_1 = \Theta^0_1, \ldots, \Theta_s = \Theta^0_s)$ je s-násobná jednoduchá hypotesa.

Tak jako při jednoduchých hypotesách je třeba k testování složené hypotesy najíti kritické obory ve výběrovém prostoru (společném pro všechny přípustné hypotesy), které přihlížejí ke dvěma zdrojům chyb v usuzování. Tedy kritický obor $w(\Theta^0_{r+1}, \ldots, \Theta^0_{r+s})$ musí splňovat především podmínku

$$\int_{w(\Theta^0_{r+1}, \ldots, \Theta^0_{r+s})} f(x_1, \ldots, x_n; \Theta_1, \ldots, \Theta_r; \Theta^0_{r+1}, \ldots, \Theta^0_{r+s}) \, dx_1, \ldots, dx_n = \alpha$$

a jeho umístění ve výběrovém prostoru by mohlo záviset na parametrech $\Theta^0_{r+1}, \ldots, \Theta^0_{r+s}$ zahrnutých do hypotesy, ale nesmí záviset na t. zv. volných parametrech $(\Theta_1, \ldots, \Theta_r)$.

α je nějaká předem stanovená hladina významnosti $0 \leqq \alpha \leqq 1$ a integrál je n-násobný. Tato podmínka podle předpokladu platí pro všechny hodnoty parametrů $(\Theta_1, \ldots, \Theta_r)$ a $(\Theta^0_{r+1}, \ldots, \Theta^0_{r+s})$. Funkce f je ovšem podstatně kladná a její integrál přes celý výběrový prostor se rovná jedné. Na formě této funkce bude záviseti, zda bude možno najít kritický obor splňující tuto podmínku čili kritický obor podobný výběrovému prostoru, jemuž se také říká *platný test*. Je tudíž možno, aspoň theoreticky, udat nutné a postačující podmínky, které má funkce f splňovat, abychom dostali platný test. Dosud však byly dány jen postačující podmínky Neymanem, které poskytují také cestu, kterou je možno vytvořit nekonečně mnoho podobných kritických oborů, z nichž lze známým Neymanovým-Pearsonovým obratem vybrat ten, který je nejsilnější vzhledem k určité alternativní hypotese $H(\Theta_{r+1} \neq \Theta^0_{r+1}, \ldots, \Theta_{r+s} \neq \Theta^0_{r+s})$ bez zřetele na ostatní parametry.

Neyman uvádí [7] postačující podmínky pro existenci platných kritických oborů pro složenou hypotesu tímto systémem rovnic

$$\frac{\partial^2 \log f}{\partial \Theta_j \partial \Theta_k} = A_{jk} + \sum_{l=1}^{r} B^l_{jk} \frac{\partial \log f}{\partial \Theta_l} \quad (j, k = 1, 2, \ldots, r),$$

kde o A_{jk} a B^l_{jk} se předpokládá, že jsou to funkce všech parametrů, volných i ostatních. Je to systém simultanních lineárních parciálních diferenciálních rovnic, v nichž $\log f$ je závislá proměnná a Θ_i ($i = 1, 2, \ldots, r$) jsou nezávislé proměnné.

Z těchto podmínek pak lze odvoditi [13] postačující podmínky pro získání postupu k sestrojení podobných kritických oborů, jež jsou

$$f \equiv f_1(F_1, \ldots, F_m; \Theta_1, \Theta_2, \ldots, \Theta_r; \Theta_{r+1}, \ldots, \Theta_{r+s}) \, f_2(x_1, \ldots, x_n; \Theta_{r+1}, \ldots, \Theta_{r+s}),$$

kde

$$F_i \equiv F_i(x_1, \ldots, x_n; \Theta_{r+1}, \ldots, \Theta_{r+s}) \quad (i = 1, 2, \ldots, m),$$

dále $m < n$ ale $m \lesseqqgtr r$, při čemž případ $m > r$ je možný jen, je-li $r < n$. Z toho plyne, že m nemusí býti vždy rovno r, aby byl získán základní postup pro sestrojení podobných kritických oborů. $(F_1, F_2, \ldots, F_m)$ pak můžeme nazvat množinou děleně vyčerpávajících charakteristik pro volné parametry $(\Theta_1, \ldots, \Theta_r)$ podmíněnou parametry $(\Theta_{r+1}, \ldots, \Theta_{r+s})$, které jsou známé.

Za hypotesy testované se ovšem v horních dvou identitách Θ_{r+j} nahradí Θ^0_{r+j} $(j = 1, 2, \ldots, s)$. Uvedeme nyní postup pro vytvoření platných testů za horních podmínek. Vezmeme v úvahu m-násobný svazek ploch $S(\mu_1, \ldots, \mu_m)$, vytvořený průsekem ploch

$$F^0_i \equiv F_i(x_1, \ldots, x_n; \Theta^0_{r+1}, \ldots, \Theta^0_{r+s}) = \mu_i \quad (i = 1, 2, \ldots, m) . \tag{6.2}$$

Tento svazek má dvě důležité vlastnosti: a) Umístění ve výběrovém prostoru kteréhokoli prvku svazku může záviset na $(\Theta^0_{r+1}, \ldots, \Theta^0_{r+s})$, ale je nezávislé na volných parametrech $(\Theta_1, \ldots, \Theta_r)$, což je ovšem z rovnice (6.2) zřejmé.

b) Vyjmeme-li z nějaké vrstvy omezené plochami $S(\mu_1, \ldots, \mu_m)$ a $S(\mu_1 + + \mathrm{d}\mu_1, \ldots, \mu_m + \mathrm{d}\mu_m)$ část, jejíž hmota je pro každou množinu hodnot parametrů $Q_1, \ldots, \Theta_r; \Theta^0_{r+1}, \ldots, \Theta^0_{r+s}$ rovna α násobku hmoty celé vrstvy, znamená to podmínku

$$\begin{aligned} &\int_{\Delta} f_2(x_1, \ldots, x_n; \Theta^0_{r+1}, \ldots, \Theta^0_{r+s})\, \mathrm{d}x_1, \ldots, \mathrm{d}x_n = \\ &= \alpha \int_{O} f_2(x_1, \ldots, x_n; \Theta^0_{r+1}, \ldots, \Theta^0_{r+s})\, \mathrm{d}x_1 \ldots \mathrm{d}x_n , \end{aligned} \tag{6.3}$$

kde integrál na pravé straně je přes celou vrstvu, kdežto na levé straně přes příslušnou její část. Poněvadž tato podmínka je zcela nezávislá na volných parametrech $(\Theta_1, \ldots, \Theta_r)$, můžeme vyjmout z té vrstvy část, která splňuje (6.3) a jejíž relativní umístění uvnitř vrstvy je nezávislé na volných parametrech, ale může záviset na parametrech $(\Theta^0_{r+1}, \ldots, \Theta^0_{r+s})$. Podržíme-li nyní parametry $(\Theta^0_{r+1}, \ldots, \Theta^0_{r+s})$ pevné a měníme volné parametry $(\Theta_1, \ldots, \Theta_r)$, bude se stále hmota části rovnat α násobku hmoty celé vrstvy a relativní umístění části uvnitř vrstvy zůstane ovšem podle definice nezměněno. Měníme-li $(\Theta^0_{r+1}, \ldots, \Theta^0_{r+s})$, zůstane úměrný zlomek hmoty stálý při správné volbě w, ale relativní umístění se budou měnit. Tyto dvě skutečnosti se krátce nazývají *princip invariance umístění* a *princip invariance úměrnosti hmoty* určité části každé vrstvy při změně volných parametrů $(\Theta_1, \ldots, \Theta_r)$.

V každé vrstvě lze takové části dostat nekonečně mnoha způsoby. Vezmeme tedy určitou takovou část z každé vrstvy, spojíme pak tyto části ze všech možných vrstev a dostaneme kritický obor, jehož umístění ve výběrovém prostoru je nezávislé na volných parametrech $(\Theta_1, \ldots, \Theta_r)$; může záviset na ostatních parametrech $(\Theta^0_{r+1}, \ldots, \Theta^0_{r+s})$, ale jeho hmota je α pro všechny hodnoty parametrů volných i ostatních. Bude to tedy podobný kritický obor rozsahu α, jichž lze tímto postupem vytvořiti nekonečný počet. Lze ukázat, že jsou-li Neymanovy podmínky splněny, pak nemůže existovat podobný kritický obor, který by nebylo možno vytvořit právě uvedeným postupem. Tato vlastnost se nazývá *principem vyčerpání* vzhledem k platným testům.

Další podrobnosti a důkazy jsou v práci [13] a [14]. Pokud se týče spojitosti mezi dosažitelností platných testů složené hypotesy, v níž jsou některé parametry ponechány volnými a existencí určité množiny děleně vyčerpávajících charakteristik pro volné parametry, tedy děleně vyčerpávající charakteristika vzhledem k určité množině parametrů skutečně zajišťuje za určité dodatečné podmínky existenci stejnoměrně nejsilnějšího testu. Na druhé straně existence množiny děleně vyčerpávajících charakteristik pro některé parametry podmíněné jinými parametry známými zajišťuje dosažitelnost platných testů pro složenou hypotézu, v níž tytéž parametry jsou ponechány volnými a ty jiné vstupují do hypotesy.

7. Další požadavky na testy

Budeme se nyní zase zabývat případem jednoparametrickým. Stejnoměrně nejsilnější kritické obory se vyskytují zřídka. Bylo by tudíž třeba zavésti další požadavek nestrannosti, a to z toho důvodu, že je-li test jednostranný, je pro nějakou hodnotu Θ' v okolí hodnoty Θ_0 síla testu menší než rozsah kritického oboru. Taková situace je však nežádoucí, neboť pravděpodobnost zamítnutí hypotesy $\Theta = \Theta_0$ je větší, když je Θ_0 pravdivá, než když je Θ' pravdivá. Test je nestranný, má-li jeho silofunkce relativní minimum pro hodnotu $\Theta = \Theta_0$, kde Θ_0 je testovaná hypotesa.

Testů založených na nestrannosti existuje obecně nekonečně mnoho, proto se definuje *stejnoměrně nejsilnější nestranný test*, který je aspoň tak silný vzhledem ke všem alternativním hypotesám nebo silnější než každý jiný nestranný test, založený na kritickém oboru stejného rozsahu. Jestliže tedy přijmeme požadavek nestrannosti a existuje-li stejnoměrně nejsilnější nestranný test, je jistě nejvýhodnější. Kritický obor odpovídající takovému testu byl nazván *kritickým oborem typu* A_1.

Tento obor existuje v několika důležitých případech, ale v mnohých dalších neexistuje, proto byl zaveden ještě *obor typu A*. Obor typu A pak je určen tak, že test na něm založený je stejnoměrně nejsilnější nestranný v malém okolí hodnoty Θ_0. Musí pak obor w rozsahu $P\{w|\Theta_0\} = \alpha$ míti silofunkci, která splňuje podmínky

$$\left(\frac{\partial P\{w|\Theta\}}{\partial \Theta}\right)_{\Theta=\Theta_0} = 0 \quad \text{a} \quad \left(\frac{\partial^2 P\{w|\Theta\}}{\partial \Theta^2}\right)_{\Theta=\Theta_0} \geqq \left(\frac{\partial^2 P\{w'|\Theta\}}{\partial \Theta^2}\right)_{\Theta=\Theta_0}$$

pro všechny obory w', které splňují první podmínku a mají týž rozsah jako obor w. Je jasné, že první podmínka vyžaduje, aby kritický obor byl nestranný, a druhá vyžaduje, aby silofunkce oboru typu A měla větší křivost, než silofunkce každého jiného nestranného oboru téhož rozsahu α. Zhruba to znamená, že takový kritický obor je nejsilnějším v okolí Θ_0. Podmínky, za nichž existuje kritický obor typu A jsou splněny ve většině praktických případů. Námitky, které se proti němu vznášejí, spočívají v tom, že máme větší zájem na tom, jak se chová silofunkce v hodnotách Θ, které jsou dále od testované hodnoty Θ_0 než v jejím blízkém okolí.

Tato kritika však postupně ztrácí na významu, když rozsah výběru roste, poněvadž obor typu A se chová pro velká n přibližně jako stejnoměrně nejsilnější nestranný kritický obor. Nesnáze, vznikající z neexistence stejnoměrně nejsilnějších nestranných testů, postupně mizí s rostoucím rozsahem výběrovým, neboť t. zv. *asymptoticky nejsilnější nestranné testy existují prakticky vždycky*. K zavedení těchto testů provedeme úvahy, při nichž se počet pozorování n bude měnit, takže na př. kritický obor v n-rozměrném výběrovém prostoru označíme w_n a jeho silofunkci $P\{w_n|\Theta\}$.

Pro určité pevné Θ a určitý rozsah kritického oboru α označme supremum $P\{w_n|\Theta\}$ vzhledem ke všem oborům w_n rozsahu α symbolem $P_n(\Theta, \alpha)$ čili $P_n(\Theta, \alpha) = \sup\limits_{w_n} P\{w_n|\Theta\}$.

Existuje-li nějaký obor w_n, pro který $P\{w_n|\Theta\} = P_n(\Theta, \alpha)$ pro určité $\Theta \neq \Theta_0$, pak tento obor w_n je nejsilnějším testem rozsahu α vzhledem k alternativě Θ. Funkce $P_n(\Theta, \alpha)$ se pro určité pevné α nazývá *obalovou silofunkcí* rozsahu α pro testování hypotesy $\Theta = \Theta_0$. Je jasné, že pro kterýkoli kritický obor w_n

rozsahu α je splněn vztah $P\{w_n|\Theta\} \leqq P_n(\Theta, \alpha)$ pro všechna Θ. V těch případech, kde existuje stejnoměrně nejsilnější test rozsahu α, splývá zřejmě jeho silofunkce s $P_n(\Theta, \alpha)$. Neexistuje-li stejnoměrně nejsilnější test, pak se stává hořejší vztah aspoň pro jednu hodnotu Θ ostrou nerovností.

Nyní vezměme v úvahu rozsah n výběru jako proměnnou. Budeme nazývat posloupnost kritických oborů $\{w_n\}$, $n = 1, 2, \ldots$ asymptoticky nejsilnějším testem rozsahu α pro testování hypotesy $\Theta = \Theta_0$, jsou-li splněny podmínky

$$P\{w_n|\Theta_0\} = \alpha \quad (n = 1, 2, \ldots)\,,$$
$$\lim_{n \to \infty} [P_n(\Theta, \alpha) - P\{w_n|\Theta\}] = 0$$

stejnoměrně v Θ. Druhá podmínka znamená, že pro libovolné kladné ε existuje n_ε takové, že pro všechna $n > n_\varepsilon$ je absolutní rozdíl mezi skutečnou silofunkcí oboru w_n a obalovou funkcí všude menší než ε. Z toho vyplývá, že je-li posloupnost w_n asymptoticky nejsilnějším testem, pak pro dostatečně velké n je kritický obor w_n prakticky tak dobrý jako stejnoměrně nejsilnější test.

Budeme ještě definovat *nestrannou obalovou silofunkci* [17] $P_n^*(\Theta, \alpha) = \sup_{w_n} P\{w_n|\Theta\}$ pro všechny obory w_n, které jsou nestranné a rozsahu α. Posloupnost $\{w_n\}$ kritických oborů pak se nazývá asymptoticky nejsilnějším nestranným testem rozsahu α pro testování hypotesy $\Theta = \Theta_0$, jestliže

$$P\{w_n|\Theta_0\} = \alpha \quad (n = 1, 2, \ldots)\,,$$
$$\lim_{n \to \infty} [P_n^*(\Theta, \alpha) - P\{w_n|\Theta\}] = 0$$

stejnoměrně v Θ. Lze pak říci opět, že je-li w_n asymptoticky nejsilnější nestranný test, pak je pro dostatečně velká n kritický obor w_n prakticky tak dobrý jako stejnoměrně nejsilnější nestranný kritický obor.

Jsou-li dvě posloupnosti kritických oborů $\{w_n\}$ a $\{w_n'\}$ asymptoticky nejsilnějšími nestrannými testy, pak test založený na w_n pro dostatečně velké n je prakticky ekvivalentní s testem založeným na w_n'. Není-li však výběrový rozsah n tak velký, aby w_n a w_n' byly prakticky ekvivalentní, je třeba rozhodnout, kterého se má užít. K rozhodnutí nás vede tato úvaha: Při dostatečně velkém n jsou silofunkce $P\{w_n|\Theta\}$ a $P\{w_n'|\Theta\}$ v malém okolí obalové silofunkce. Mohou se však chovat různě v tom smyslu, že s rostoucím n se jedna silofunkce, třeba $P\{w_n|\Theta\}$ blíží obalové silofunkci rychleji než druhá. V tom případě se zdá vhodným dáti přednost $\{w_n\}$. Zavádí se tedy *pojem nejostřejšího testu* rozsahu α pro testování hypotesy $\Theta = \Theta_0$, jímž je kritický obor w_n výběrového prostoru n-rozměrného, který má tyto dvě vlastnosti:

$$P\{w_n|\Theta_0\} = \alpha\,,$$
$$\sup_{\Theta} [P_n(\Theta, \alpha) - P\{w_n|\Theta\}] \leqq \sup_{\Theta} [P_n(\Theta, \alpha) - P\{w_n'|\Theta\}]$$

pro každý obor w_n', splňující první vlastnost.

Poněvadž $\{w_n\}$ je asymptoticky nejsilnějším nestranným testem, je-li w_n nejostřejším kritickým oborem pro každý výběrový rozsah n, a lze tedy říci, že přiblížení jeho silofunkce k obalové silofunkci je rychlejší než pro každý jiný asymptoticky nejsilnější nestranný test, jeví se žádoucím užíti nejostřejšího testu.

Dosavadní úvahy vedoucí k řešení problému nejlepšího testu jsou založeny na myšlence, že test t_1 je lepší než t_2, jestliže $P\{x \in w_1|\overline{H}\} \geqq P\{x \in w_2|\overline{H}\}$, kde w_1 je kritický obor pro test t_1 a w_2 je kritický obor pro test t_2, a je při tom splněna rovnice

$$P\{x \in w_1|H\} = P\{x \in w_2|H\} = \alpha .$$

Klademe však na pojem nejlepšího testu ještě další požadavky. Takovými požadavky jsou na př. snadné numerické počítání nebo malá citlivost testu na požadavek normality.

8. Otázka hospodárnosti při volbě testu

Při praktické aplikaci pak je na místě zvážit, jaká je ztráta, jestliže použijeme jiného testu, než je test v horním smyslu nejlepší, nebo jaké výhody získáme, jestliže použijeme právě tohoto jiného testu. Může se totiž stát, že nejsilnější test má přibližně touž silofunkci na téže hladině významnosti, jako daný test, ale při tom nejsilnější test spočívá na menším výběrovém rozsahu. Jinými slovy nejsilnější test založený na n' výběrových hodnotách poskytuje přibližně totéž množství informace, jako daný test založený na n výběrových hodnotách ($n' \leqq n$). Říká se pak, že $n - n'$ výběrových hodnot bylo promarněno nebo ztraceno, když se užije daného testu místo nejsilnějšího testu.

Jednu z možností jak odhadnout ztrátu, která vzniká, když se nepoužije nejlepšího testu, dává t. zv. *koeficient silové vydatnosti* testu K, který je definován rovnicí

$$K = \frac{\bar{n}(\alpha, P_{t_1}\{x \in w_1|\overline{H}\})}{n(\alpha, P_{t_2}\{x \in w_2|\overline{H}\})},$$

kde $\bar{n}$ nemusí být celé číslo a je zvoleno tak, aby se $P_{t_1}\{x \in w_1|\overline{H}\}$ co nejvíce přimykala $P_{t_2}\{x \in w_2|\overline{H}\}$ ve smyslu dále uvedeném pro různé jednoduché hypotesy $\bar{h} \in \overline{H}$. Potom koeficient silové vydatnosti K, který je mezi nulou a jednou, udává množství informace, které jsme získali na jednotku pozorování při testu t_2, jestliže se při testu t_1 koeficient K rovná právě jedné.

Tento postup pro měření silové vydatnosti nějakého testu je odlišný od obvyklé metody měření vydatnosti nějakého odhadu, neboť tuto dostáváme výpočtem poměru rozptylu vydatného odhadu ku rozptylu daného odhadu: tento poměr se vyjadřuje v procentech. Avšak postup, kterým se určuje silová vydatnost nějakého testu, spočívá v tom, že se spojitě mění rozsah výběru příslušného nejsilnějšího testu na téže hladině významnosti, až jsou silofunkce daného testu a nejsilnějšího testu ekvivalentní v tom smyslu, že plocha mezi dvěma silokřivkami, pro kterou silofunkce nejsilnějšího testu převyšuje silofunkci daného testu, se rovná analogické ploše, pro kterou silofunkce nejsilnějšího testu má nižší hodnoty než silofunkce daného testu.

Je tudíž definována silová vydatnost daného testu jako poměr výběrového rozsahu (který nemusí být celým číslem) nejsilnějšího testu s ekvivalentní silofunkcí ku výběrovému rozsahu daného testu, vyjádřený v procentech.

(Dokončení.)