Poznámky k testování hypotéz

Chceme otestovat nějakou hypotézu $\mathbf{H}_0$ o rozdělení náhodné veličiny X (tzv. nulovou hypotézu). Je dobré uvědomit si následující věci:

- Obecně máme vždy k dispozici jen konečně mnoho dat. Pokusů (nebo výsledků) máme ale teoreticky neomezeně, takže principiálně nikdy nemůžeme obsáhnout pomocí konečně mnoha dat všechny možnosti (i při házení mincí se dvěma výsledky, rub nebo líc, vzniká nekonečná posloupnost naměřených dat a všech možných posloupností je navíc nespočetně). Hypotézu tak *nemůžeme potvrdit*, ale nanejvýše v určitém smyslu "vyvrátit" (viz dále).

  Určitou analogií by mohla být situace, že máme zjistit, zda nějaká vlastnost $V(n)$ platí pro všechna přirozená čísla $n \in \mathbb{N}$, a my nemůžeme dělat nic jiného než je postupně všechna procházet. Po konečně mnoha pokusech pak buď narazíme na protipříklad n_0 (a pak určitě víme, že vlastnost V neplatí) anebo ne. Ve druhém případě to může být buď proto, že jsme na protipříklad ještě nenarazili anebo proto, že V zkrátka platí. To vede k tomu, že o závěru takovéhoto pokusu má smysl mluvit jen jako o *zamítnutí* nebo *nezamítnutí* (tj. nemůžeme mluvit o potvrzení). Nezamítnout pak znamená, že nemáme (případné) podklady pro zamítnutí.

  Ovšem to byl příklad, kdy jsme u pokusů (pro konkrétní $n \in \mathbb{N}$) měli vždy odpověď typu ano/ne. U náhodných veličin je to zkomplikováno tím, že i netypická data (která se zdají být v rozporu s $\mathbf{H}_0$) mohou (díky náhodě) pocházet z platnosti námi uvažované hypotézy $\mathbf{H}_0$ (byť s malou pravděpodobností).

- Z výše uvedeného vidíme, že to, jestli $\mathbf{H}_0$ (ve statistice) opravdu platí nebo ne, se tudíž *nikdy nedozvíme* (alespoň ne v rámci testování). A na druhou stranu, pokud bychom to odněkud přece jen věděli, bylo by pak nějaké testování v tomto ohledu už zbytečné.

  Protože ale nějaký závěr z testování potřebujeme udělat, je pak naším *ROZHODNUTÍM* buď zamítnutí nebo nezamítnutí $\mathbf{H}_0$ a to na základě nějakého zvoleného kritéria a pochopitelně s určitou možnou chybou (přesněji, se dvěma chybami), které s sebou toto rozhodnutí přináší:

($\mathbf{H}_0$)	nezamítáme	zamítáme
platí :	správné rozhodnutí	chyba 1. druhu
neplatí :	chyba 2. druhu	správné rozhodnutí

  Máme tak

  $$\text{“nastává chyba 1. druhu”} \quad \overset{\text{(pokud } \mathbf{H}_0 \text{ platí)}}{\Longleftrightarrow} \quad \text{zamítáme } \mathbf{H}_0$$

  $$\text{“nastává chyba 2. druhu”} \quad \overset{\text{(pokud } \mathbf{H}_0 \text{ NEplatí)}}{\Longleftrightarrow} \quad \text{NEzamítáme } \mathbf{H}_0$$

  a zdůrazněme, že ty předpoklady napsané nad ekvivalencemi jsou podstatné. Pokud totiž při testování dojdeme k závěru, že máme hypotézu $\mathbf{H}_0$ zamítnout (na základě nějakého kritéria), NEZNAMENÁ to, že jsem udělali chybu 1. druhu! A to proto, že my zkrátka nevíme, jestli $\mathbf{H}_0$ platí nebo ne.

  A navíc, z toho, že chyba 1. druhu opravdu nastává, okamžitě plyne, že $\mathbf{H}_0$ pak musí platit. A to, jak už bylo řečeno výše, prostě nevíme a právě proto také děláme to testování.

- Pravděpodobnosti, že nastanou uvedené chyby se pochopitelně snažíme minimalizovat (volbou kritéria pro zamítání), ale protože není možné obě dvě současně snižovat, musíme si vybrat tu chybu, kterou pokládáme za důležitější. Jelikož celé testování probíhá za předpokladu platnosti $\mathbf{H}_0$ (tj. výpočty, posuzování jejich výsledků atd.), tak za větší prohřešek pokládáme to, že bychom se nakonec spletli a zamítli $\mathbf{H}_0$, která by platila (tj. udělali bychom chybu 1. druhu). Kritérium pro zamítnutí proto volíme tak, aby pravděpodobnost chyby 1. druhu byla omezena předem danou hodnotou α (tzv. hladinou významnosti). Hodnota α by tudíž měla být malé číslo (např. 5% nebo 1% apod.)

- Na tomto místě je ještě potřeba zdůraznit a ujasnit si následující věc: Pravděpodobnosti jednotlivých chyb i správných rozhodnutí jsou VŽDY VZTAŽENY vzhledem k platnosti nebo neplatnosti hypotézy $\mathbf{H}_0$. Důležité ovšem je, že nejde o podmíněnou pravděpodobnost, protože NEMÁME nijak URČENU pravděpodobnost, jestli $\mathbf{H}_0$ platí nebo ne (a takové rozdělení by ani obecně nemuselo dávat smysl). Abychom tuto odlišnost (od podmíněné pravděpodobnosti) trochu více zdůraznili, označíme to způsobem uvedeným níže. Přitom α bude pravděpodobnost chyby 1. druhu a β bude pravděpodobnost chyby 2. druhu:

$$
\begin{aligned}
P_{(H_0 \text{ platí})}\Big(\text{nastává chyba 1. druhu}\Big) &= P_{(H_0 \text{ platí})}\Big(\text{zamítáme } \mathbf{H}_0\Big) &&= \alpha \\
P_{(H_0 \text{ platí})}\Big(\text{správné rozhodnutí}\Big) &= P_{(H_0 \text{ platí})}\Big(\text{nezamítáme } \mathbf{H}_0\Big) &&= 1-\alpha \\
P_{(H_0 \text{ NEplatí})}\Big(\text{nastává chyba 2. druhu}\Big) &= P_{(H_0 \text{ NEplatí})}\Big(\text{nezamítáme } \mathbf{H}_0\Big) &&= \beta \\
P_{(H_0 \text{ NEplatí})}\Big(\text{správné rozhodnutí}\Big) &= P_{(H_0 \text{ NEplatí})}\Big(\text{zamítáme } \mathbf{H}_0\Big) &&= 1-\beta
\end{aligned}
$$

S podmínkou platnosti/neplatnosti hypotézy jsme tady vlastně ve stejné situaci jako v **Příkladu 2.4**, kde byla na výběr dvě pořadí protihráčů v tenise (tj. dva různé modely nebo také dva různé "světy"). Stejný přístup jsme také používali u metody maximální věrohodnosti, kdy značení $P_\lambda(X_1 = x_1, \ldots, X_n = x_n)$ znamenalo, že hodnota pravděpodobnosti závisí na volbě parametru λ (jenž také určoval, který model zrovna používáme). A ani v případě této metody jsme pochopitelně vůbec neuvažovali o tom, že by zmíněný parametr λ měl nějaké rozdělení pravděpodobnosti.

- Poznamenejme ještě že, pokud jsme $\mathbf{H}_0$ zamítli (při testu s hladinou α), pak jsme se buďto strefili (protože opravdu neplatila) a to pak bylo s pravděpodobností $1-\beta$ (viz výše) anebo jsme udělali chybu 1. druhu (ale jen s malou pravděpodobností α). Tyto dvě pravděpodobnosti ovšem obecně doplňkové nejsou.

- Kritérium zamítnutí pro danou hypotézu $\mathbf{H}_0$, daný počet n naměřených hodnot a danou hladinu α je určeno pomocí tzv. *kritického oboru* $W \subseteq \mathbb{R}^n$ (který tudíž závisí na $\mathbf{H}_0$, n a α) a to prostě tak, že

$$\mathbf{H}_0 \text{ zamítáme (na hladině } \alpha) \quad \Leftrightarrow \quad (x_1, \ldots, x_n) \in W$$

kde $(x_1, \ldots, x_n)$ je náš soubor naměřených hodnot veličiny X (při nezávislých pokusech), který pro test používáme.

Volbu kritického oboru W pak v daném případě obvykle děláme pomocí určité náhodné veličiny T (tzv. *testovací statistiky*), která má rozdělení, které známe v jistém smyslu "úplně" (např. $N(0,1)$; t-rozdělení, χ^2-rozdělení atd.) na rozdíl od rozdělení původní veličiny X, u kterého známe zpravidla jen typ rozdělení (např. normální), ale konkrétní parametry už nevíme. (Právě tyto parametry jsou obvykle předmětem našich hypotéz.)

Výše popsaný kritický obor W (který spadá do $\mathbb{R}^n$) je ale spíše teoretická záležitost, protože v praxi si vystačíme s tím, že zamítací kritérium určíme pomocí rozmezí hodnot statistiky T, tedy pomocí nějaké podmnožiny v $\mathbb{R}$ (např. tak, že $T > c$, kde c je nějaká mezní hodnota pro zamítnutí hypotézy $\mathbf{H}_0$).

Další možností (jak vyjádřit zamítací kritérium) je použití *intervalu spolehlivosti*, který je ovšem nakonec jen jinak přepsaná podmínka pro zvolenou testovací statistiku T.

Jak je vidět, všechny uvedené možnosti (tj. kritický obor, podmínka pro T, podmínka pro interval spolehlivosti) jsou jen ekvivalentními verzemi stejné zamítací podmínky. Neboli (pro danou hladinu α) máme

zamítáme $\mathbf{H}_0 \quad \Leftrightarrow \quad (x_1, \ldots, x_n) \in W \quad \Leftrightarrow$ platí podmínka pro $T \quad \Leftrightarrow$ platí podmínka pro int. spolehl.

- Na závěr si (pro lepší představu) ukážeme volbu ekvivalentních zamítacích podmínek na konkrétním příkladu: pro veličinu $X \sim N(\mu, \sigma^2)$; nulovou hypotézu $\mathbf{H}_0 : \mu \leq \mu_0$ a její alternativu $\mathbf{H}_A : \mu > \mu_0$. Máme soubor naměřených hodnot $(x_1, \ldots, x_n)$. Zde volíme testovací statistiku $T = \frac{\overline{\boldsymbol{X}} - \mu_0}{S_{\boldsymbol{X}}} \sqrt{n}$.

  Veličina T je tady také zapsatelná jako $T = h(X_1, \ldots, X_n)$ pro nějakou reálnou funkci $h : \mathbb{R}^n \to \mathbb{R}$ a naše nezávislé veličiny $X_1, \ldots, X_n$ představující jednotlivé pokusy. Konkrétní tvar funkce h teď nebude vypisovat. Pro hladinu α si dále ještě označme $c := t_{1-\alpha;\ n-1}$ jistou hodnotu (tzv. kvantil). Pak tedy máme:

  - podmínka pro T:
  $$t := h(x_1, \ldots, x_n) = \frac{\overline{\boldsymbol{x}} - \mu_0}{s_{\boldsymbol{x}}} \sqrt{n} \ > \ c$$
  - kritický obor:
  $$W := \{(y_1, \ldots, y_n) \in \mathbb{R}^n \mid h(y_1, \ldots, y_n) \ > \ c\}$$
  - podmínka pro kritický obor:
  $$(x_1, \ldots, x_n) \in W$$
  - interval spolehlivosti:
  $$\langle \mu_L,\ +\infty) := \left\langle \ \overline{\boldsymbol{x}} - \frac{s_{\boldsymbol{x}}}{\sqrt{n}} \cdot c \quad , \quad +\infty \ \right)$$
  - podmínka pro interval spolehlivosti:
  $$\mu_0 \notin \langle \mu_L,\ +\infty)$$