Ověřování exponenciálního a Poissonova rozložení

Test dobré shody
Testujeme hypotézu, která tvrdí, že náhodný výběr $X_1, ..., X_n$ pochází z rozložení s distribuční funkcí $\Phi(x)$.

a) Je-li distribuční funkce spojitá, pak data rozdělíme do r třídicích intervalů $\left(u_j, u_{j+1}\right\rangle$, j = 1, ..., r. Zjistíme absolutní četnost n_j j-tého třídicího intervalu a vypočteme pravděpodobnost p_j, že náhodná veličina X s distribuční funkcí $\Phi(x)$ se bude realizovat v j-tém třídicím intervalu. Platí-li nulová hypotéza, pak $p_j = \Phi(u_{j+1}) - \Phi(u_j)$.
b) Má-li distribuční funkce nejvýše spočetně mnoho bodů nespojitosti, pak místo třídicích intervalů použijeme varianty $x_{[j]}$, j = 1, …, r. Pro variantu $x_{[j]}$ zjistíme absolutní četnost n_j a vypočteme pravděpodobnost p_j, že náhodná veličina X s distribuční funkcí $\Phi(x)$ se bude realizovat variantou $x_{[j]}$. Platí-li nulová hypotéza, pak
$p_j = \Phi(x_{[j]}) - \lim\limits_{x \to x_{[j]}-} \Phi(x) = P(X = x_{[j]})$.

Testová statistika: $K = \sum_{j=1}^{r} \frac{(n_j - np_j)^2}{np_j}$. Platí-li nulová hypotéza, pak $K \approx \chi^2(r-1-p)$, kde p je počet odhadovaných parametrů daného rozložení. (Např. pro normální rozložení p = 2, protože z dat odhadujeme střední hodnotu a rozptyl.) Nulovou hypotézu zamítáme na asymptotické hladině významnosti α, když $K \geq \chi^2_{1-\alpha}(r-1-p)$. Aproximace se považuje za vyhovující, když $np_j \geq 5$, j = 1, ..., r.

Upozornění: Hodnota testové statistiky K je silně závislá na volbě třídicích intervalů. Navíc při nesplnění podmínky $np_j \geq 5$, j = 1, ..., r je třeba některé intervaly resp. varianty slučovat, což vede ke ztrátě informace.

Jednoduchý test exponenciálního rozložení
Testujeme hypotézu, která tvrdí, že náhodný výběr $X_1, ..., X_n$ pochází z exponenciálního rozložení. Označme M výběrový průměr a S^2 výběrový rozptyl tohoto náhodného výběru. Víme, že střední hodnota náhodné veličiny $X \sim Ex(\lambda)$ je $E(X) = 1/\lambda$ a rozptyl je $D(X) = 1/\lambda^2$.
Test založíme na statistice $K = \frac{(n-1)S^2}{M^2}$, která se v případě platnosti H_0 asymptoticky řídí rozložením $\chi^2(n-1)$. Kritický obor: $W = \left\langle 0, \chi^2_{\alpha/2}(n-1)\right\rangle \cup \left\langle \chi^2_{1-\alpha/2}(n-1), \infty\right)$. Jestliže $K \in W$, H_0 zamítáme na asymptotické hladině významnosti α.

Jednoduchý test Poissonova rozložení
Testujeme hypotézu, která tvrdí, že náhodný výběr $X_1, ..., X_n$ pochází z Poissonova rozložení. Označme M výběrový průměr a S^2 výběrový rozptyl tohoto náhodného výběru. Víme, že střední hodnota náhodné veličiny $X \sim Po(\lambda)$ je $E(X) = \lambda$ a rozptyl je $D(X) = \lambda$. Test založíme na statistice $K = \frac{(n-1)S^2}{M}$, která se v případě platnosti H_0 asymptoticky řídí rozložením $\chi^2(n-1)$. Kritický obor: $W = \left\langle 0, \chi^2_{\alpha/2}(n-1)\right\rangle \cup \left\langle \chi^2_{1-\alpha/2}(n-1), \infty\right)$.

Příklad 1.: Byla zkoumána doba životnosti 45 součástek (v hodinách). Výsledky jsou uvedeny v tabulce rozložení četností:

Doba životnosti	Počet součástek
(0, 50]	15
(50, 100]	14
(100, 150]	6
(150, 200]	5
(200, 250]	2
(250, 300]	1
(300, 350]	1
(350, 400]	1

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že daný náhodný výběr pochází z exponenciálního rozložení. Použijte a) test dobré shody, b) jednoduchý test exponenciálního rozložení (využijte toho, že z původních dat byl vypočten průměr m = 99,93 a rozptyl s^2 = 7328,91).

Výsledky:

ad a) Protože nejsou splněny podmínky dobré aproximace pro j = 4, 5, 6, 7, 8, je třeba sloučit třídicí intervaly 4 až 8 do jednoho. K = 1,5153, $W = \langle 5{,}9915, \infty)$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

ad b) K = 32,2924, $W = \langle 0; 27{,}575 \rangle \cup \langle 64{,}202; \infty)$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Příklad 2.: Studujeme rozložení počtu pacientů, kteří během 75 dnů přijdou na pohotovost. Osmihodinovou pracovní dobu rozdělíme do půlhodinových intervalů a v každém intervalu zjistíme počet příchozích pacientů:

Počet pacientů	Pozorovaná četnost
0	79
1	188
2	282
3	275
4	196
5	114
6	45
7	10
8	7
9	3
10	1

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že daný náhodný výběr pochází z Poissonova rozložení. Použijte a) test dobré shody, b) jednoduchý test Poissonova rozložení.

Výsledky:

ad a) Protože nejsou splněny podmínky dobré aproximace, je třeba sloučit poslední tři varianty do jedné. K = 8,5217, $W = \langle 14{,}067, \infty)$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

ad b) K = 1158,47, $W = \langle 0; 1104{,}93 \rangle \cup \langle 1296{,}86; \infty)$, H_0 nezamítáme na asymptotické hladině významnosti 0,05.