5. Testování exponenciálního a Poissonova rozložení

5.1. Věta: Věta o provedení testu dobré shody

5.2. Příklad: Byla zjišťována doba životnosti 45 součástek (v hodinách). Ze získaných údajů byl vypočten výběrový průměr m = 99,93 h a výběrový rozptyl $s^2 = 7328{,}9\ h^2$. Máme k dispozici roztříděné údaje:

Doba životnosti	Počet součástek
$\langle 0, 50 \rangle$	15
$\langle 50, 100 \rangle$	14
$\langle 100, 150 \rangle$	6
$\langle 150, 200 \rangle$	5
$\langle 200, 250 \rangle$	2
$\langle 250, 300 \rangle$	1
$\langle 300, 350 \rangle$	1
$\langle 350, 400 \rangle$	1

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že doba životnosti se řídí exponenciálním rozložením.

Řešení: $\hat{\lambda} = \frac{1}{m} = \frac{1}{99{,}93}$, testujeme H_0: $X_1, \dots, X_{45} \sim \mathrm{Ex}\left(\frac{1}{99{,}93}\right)$ proti H_1: non H_0. Počítáme pravděpodobnosti $p_j = \int_{u_j}^{u_{j+1}} \frac{1}{99{,}93} e^{-\frac{x}{99{,}93}} dx$, j = 1, 2, …, 8

j	$\langle u_j, u_{j+1} \rangle$	n_j	p_j	$np_j = 45p_j$
1	$\langle 0, 50 \rangle$	15	0,3937	17,72
2	$\langle 50, 100 \rangle$	14	0,2387	10,74
3	$\langle 100, 150 \rangle$	6	0,1447	6,51
4	$\langle 150, 200 \rangle$	5	0,0878	3,95
5	$\langle 200, 250 \rangle$	2	0,0532	2,39
6	$\langle 250, 300 \rangle$	1	0,0323	1,45
7	$\langle 300, 350 \rangle$	1	0,0196	0,88
8	$\langle 350, 400 \rangle$	1	0,0119	0,53

Vidíme, že pro j = 4, …, 8 nejsou splněny podmínky dobré aproximace. Posledních 5 intervalů tedy sloučíme do jednoho.

Dostaneme novou tabulku

j	$\langle u_j, u_{j+1} \rangle$	n_j	p_j	$np_j = 45p_j$
1	$\langle 0,50 \rangle$	15	0,3937	17,7157
2	$\langle 50,100 \rangle$	14	0,2387	10,7413
3	$\langle 100,150 \rangle$	6	0,1447	6,5127
4	$\langle 150,400 \rangle$	10	0,2046	9,2084

Testová statistika:

$$K = \sum_{j=1}^{r} \frac{(n_j - np_j)^2}{np_j} = \frac{(15-17,7157)^2}{17,7157} + \frac{(14-10,7413)^2}{10,7413} + \frac{(6-6,5127)^2}{6,5127} + \frac{(10-9,2084)^2}{9,2084} =$$

$$= 1,5133$$

Kritický obor: $W = \langle \chi^2_{1-\alpha}(r-p-1), \infty) = \langle \chi^2_{0,95}(4-1-1), \infty) = \langle 5,9915, \infty)$

$K \notin W \Rightarrow$ H_0 nezamítáme na asymptotické hladině významnosti 0,05.

5.3. Poznámka: V MATLABu se test dobré shody pro exponenciální rozložení provádí pomocí funkce tds_exp.m.

5.4. Příklad: Sledujeme rozložení počtu pacientů, kteří během 75 dnů přijdou na zubní pohotovost. Osmihodinovou pracovní dobu rozdělíme do půlhodinových intervalů a v každém intervalu zjistíme počet příchozích pacientů (máme 16 x 75 = 1200 intervalů).

Počet pacientů	0	1	2	3	4	5	6	7	8 a víc
četnost	79	188	282	275	196	114	45	10	11

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že počet příchozích pacientů během půl hodiny se řídí Poissonovým rozložením.

Řešení:

$$\hat{\lambda} = m = \frac{1}{1200}(79 \cdot 0 + 188 \cdot 1 + 282 \cdot 2 + 275 \cdot 3 + 196 \cdot 4 + 114 \cdot 5 + 45 \cdot 6 + 10 \cdot 7 + 11 \cdot 8) = 2,7992$$

testujeme H_0: $X_1, \ldots, X_{1200} \sim Po(2,7992)$ proti H_1: non H_0. Počítáme pravděpodobnosti

$p_j = \frac{2,7992^j}{j!} e^{-2,7992}$, j = 0, 1, …, 7, $p_8 = 1 - \sum_{j=0}^{7} p_j$.

j	0	1	2	3	4	5	6	7	8
n_j	79	188	282	275	196	114	45	10	11
np_j	73,0329	204,4313	286,1186	266,9646	186,8195	104,5878	48,7931	19,5114	9,7406

Podmínky dobré aproximace jsou splněny.

Testová statistika:
$$K = \sum_{j=1}^{r} \frac{(n_j - np_j)^2}{np_j} = \frac{(79-73{,}0329)^2}{73{,}0329} + \frac{(188-204{,}4313)^2}{204{,}4313} + \ldots + \frac{(11-9{,}7406)^2}{9{,}7406} = 8{,}5019$$
Kritický obor: $W = \langle \chi^2_{1-\alpha}(r-p-1), \infty) = \langle \chi^2_{0,95}(9-1-1), \infty) = \langle 14{,}067, \infty)$

$K \notin W \Rightarrow$ H_0 nezamítáme na asymptotické hladině významnosti 0,05.

5.5. Poznámka: V MATLABu se test dobré shody pro Poissonovo rozložení provádí pomocí funkce tds_poiss.m.

5.6. Věta: Darlingův test exponenciálního rozložení

5.7. Příklad: Pro data z příkladu 5.2. proveďte na hladině významnosti 0,05 Darlingův test.
Řešení: n = 45, m = 99,93 h, s^2 = 7328,9 h^2

Testová statistika: $K = \frac{(n-1)S^2}{M^2} = \frac{44 \cdot 7328{,}91}{99{,}93^2} = 32{,}2924$

Kritický obor:

$$W = \langle 0, \chi^2_{\alpha/2}(n-1) \rangle \cup \langle \chi^2_{1-\alpha/2}(n-1), \infty) = \langle 0, \chi^2_{0,025}(44) \rangle \cup \langle \chi^2_{0,975}(44), \infty) =$$
$$= \langle 0{,}27{,}575 \rangle \cup \langle 64{,}202, \infty)$$
Protože se testová statistika nerealizuje v kritickém oboru, hypotézu o exponenciálním rozložení nezamítáme na asymptotické hladině významnosti 0,05.

5.8. Věta: Jednoduchý test Poissonova rozložení

5.9. Příklad: Pro data z příkladu 5.4. proveďte na hladině významnosti 0,05 jednoduchý test Poissonova rozložení.

Řešení: Nejprve musíme vypočítat realizaci výběrového průměru a výběrového rozptylu:
$$\hat{\lambda} = m = \frac{1}{1200}(79 \cdot 0 + 188 \cdot 1 + 282 \cdot 2 + 275 \cdot 3 + 196 \cdot 4 + 114 \cdot 5 + 45 \cdot 6 + 10 \cdot 7 + 11 \cdot 8) = 2{,}7992$$
$$s^2 = \frac{1}{1199}\left[79 \cdot (0-2{,}7992)^2 + 188 \cdot (1-2{,}7992)^2 + \ldots + 1 \cdot (10-2{,}7992)^2\right] = 2{,}6594$$
$$K = \frac{(n-1)S^2}{M} = \frac{1199 \cdot 2{,}6594}{2{,}7882} = 1139{,}1$$
Kritický obor: $W = \langle 0, \chi^2_{\alpha/2}(n-1) \rangle \cup \langle \chi^2_{1-\alpha/2}(n-1), \infty) = \langle 0; 1104{,}93 \rangle \cup \langle 1296{,}86; \infty)$

H_0 nezamítáme na asymptotické hladině významnosti 0,05.

5.10. Poznámka: Darlingův test i jednoduchý test Poissonova rozložení můžeme v MATLABu provést pomocí funkce darling.m.

5.11. Poznámka: Aproximační vzorec pro výpočet kvantilů χ^2- rozložení

5.12. Poznámka: Význam a způsob konstrukce P-P grafu