Cvičení 5 s návodem

Příklady na testování exponenciálního a Poissonova rozložení

Teoretická část

I. Test dobré shody

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

Testová statistika $K = \sum_{j=1}^{r} \frac{\left(n_j - np_j\right)^2}{np_j}$ se za platnosti nulové hypotézy asymptoticky řídí rozložením $\chi^2(r\text{-}p\text{-}1)$, kde p je počet odhadovaných parametrů daného rozložení.

Přitom

n_j je absolutní četnost j-tého třídicího intervalu pro veličinu X resp. j-té varianty veličiny X,

np_j je teoretická četnost j-tého třídicího intervalu pro veličinu X resp. j-té varianty veličiny X.

Platí-li nulová hypotéza, pak $p_j = \Phi(u_{j+1}) - \Phi(u_j)$ resp. $p_j = \Phi\left(x_{[j]}\right) - \lim\limits_{x \to x_{[j]}-} \Phi(x) = P\left(X = x_{[j]}\right)$.

Kritický obor: $W = \left\langle \chi^2_{1-\alpha}(r - p - 1), \infty \right)$.

Jestliže $K \in W$, H_0 zamítáme na asymptotické hladině významnosti α.

Aproximace se považuje za vyhovující, když $np_j \geq 5$, j = 1, ..., r.

Při nesplnění podmínky $np_j \geq 5$, j = 1, ..., r je třeba některé intervaly resp. varianty slučovat.

II. Jednoduchý test exponenciálního rozložení (Darlingův test)

Testujeme hypotézu, která tvrdí, že náhodný výběr $X_1, ..., X_n$ pochází z exponenciálního rozložení.

Testová statistika $K = \frac{(n-1)S^2}{M^2}$, která se v případě platnosti H_0 asymptoticky řídí rozložením $\chi^2(n\text{-}1)$.

Přitom M je výběrový průměr a S^2 je výběrový rozptyl daného náhodného výběru.

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.

III. Jednoduchý test Poissonova rozložení

Testujeme hypotézu, která tvrdí, že náhodný výběr $X_1, ..., X_n$ pochází z Poissonova rozložení.

Testová statistika $K = \frac{(n-1)S^2}{M}$, která se v případě platnosti H_0 asymptoticky řídí rozložením $\chi^2(n\text{-}1)$.

Přitom M je výběrový průměr a S^2 je výběrový rozptyl daného náhodného výběru.

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.

Příklad 1.: V systému hromadné obsluhy byla sledována doba obsluhy 70 zákazníků (v min). Výsledky jsou uvedeny v tabulce rozložení četností:

Doba obsluhy	Počet zákazníků
(0, 3]	14
(3,6]	16
(6,9]	10
(9,12]	9
(12,15]	8
(15,18]	5
(18,21]	3
(21,24]	5

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

Výsledek:
Ad a) Podmínky dobré aproximace nejsou splněny, sloučíme intervaly (15,18] a (18,21] .
Testová statistika K = 4,6175, p-hodnota = 0,4643, lambda = 0,1103
Protože p-hodnota je větší než hladina významnosti 0,05, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Ad b) Testová statistika K = 35,7265, p-hodnota = 0,0006, lambda = 0,1122
Protože p-hodnota je menší než hladina významnosti 0,05, H_0 zamítáme na asymptotické hladině významnosti 0,05.

Příklad 2.: Na jistém nádraží byl sledován počet přijíždějících vlaků za 1 h. Pozorování bylo prováděno celkem 15 dnů (tj. 360 h) a výsledky jsou uvedeny v tabulce:

Počet vlaků za 1 hodinu	0	1	2	3	4	5	6	7 a víc
četnost	27	93	103	58	50	21	6	2

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že počet přijíždějících vlaků za 1 h se řídí Poissonovým rozložením, a to a) testem dobré shody, b) jednoduchým testem Poissonova rozložení.

Výsledek:
Ad a) Podmínky dobré aproximace nejsou splněny, sloučíme varianty 6 a 7 a víc.
Testová statistika K = 9,6033, p-hodnota = 0,0873, lambda = 2,2944
Protože p-hodnota je větší než hladina významnosti 0,05, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Ad b)
Testová statistika K = 331,1304, p-hodnota = 0,2968, lambda = 2,3
Protože p-hodnota je větší než hladina významnosti 0,05, H_0 nezamítáme na asymptotické hladině významnosti 0,05.

Další možnosti ověřování exponenciálního rozložení:
využití funkce probplot (pravděpodobnostně – pravděpodobnostní graf),
Kolmogorovův – Smirnovův test (funkce kstest, musíme znát parametr lambda).

Použití K-S testu a P-P plotu:
Vygenerujeme 100 hodnot z exponenciálního rozložení se střední hodnotou 2:
x=exprnd(2,100,1);
Provedeme porovnání výběrové distribuční funkce s distribuční funkcí exponenciálního rozložení se střední hodnotou 2:
[h,p,ksstat]=kstest(x,[x,expcdf(x,2)])
Význam výstupních parametrů:
h = 0, když nezamítáme hypotézu o exponenciálním rozložení Ex(2) na hladině významnosti 0,05, h = 1, když tuto hypotézu zamítáme.
p je odpovídající p-hodnota
ksstat je hodnota testové statistiky.

probplot(‘Exponential‘, x)

Příklady k samostatnému řešení:
1. Máme k dispozici 10 údajů o době mezi poruchami určitého zařízení (v hodinách):
14 25 196 205 64 237 162 84 121 38
Na hladině významnosti 0,05 rozhodněte pomocí Darlingova testu, zda lze rozložení doby do poruchy považovat za exponenciální. [Nulovou hypotézu nezamítáme na hladině významnosti 0,05, p-hodnota = 0,2546]

2. Česká obchodní inspekce provedla šetření ve 22 sběrnách druhotných surovin. Zjišťovala počet závad, které se v jednotlivých sběrnách vyskytly. Výsledky jsou uvedeny v tabulce:

Počet závad	0	1	2	3
Počet sběren	7	5	4	6

Na hladině významnosti 0,05 rozhodněte pomocí a) testu dobré shody (ověřte splnění podmínek dobré aproximace), b) jednoduchého testu, zda lze rozložení počtu závad považovat za Poissonovo. [Nulovou hypotézu nezamítáme na hladině významnosti 0,05, a) p-hodnota = 0,211, b) p-hodnota = 0,7732]