

Příklady na testování hypotéz o exponenciálním a Poissonově rozložení

Příklad 1.: Za 2. světové války byl Londýn bombardován řízenými střelami. Jeho jižní část byla rozdělena na oblasti o ploše $0,25 \text{ km}^2$ a bylo zkoumáno, kolik řízených střel dopadlo na každou z těchto oblastí.

Počet střel	0	1	2	3	4 a víc
Počet oblastí	229	211	93	35	8

Na asymptotické hladině významnosti testujte hypotézu, že počet řízených střel, které dopadly na jednu oblast, se řídí Poissonovým rozložením. Úkol vyřešte

- pomocí testu dobré shody,
- pomocí jednoduchého testu Poissonova rozložení (zde uveďte i odpovídající p-hodnotu).

Výsledky:

Ad a) Test dobré shody

Odhad parametru $\lambda \dots 0,9271$

Teoretické četnosti $np_j \dots 227,9268 \quad 211,3071 \quad 97,9496 \quad 30,2692 \quad 8,5473$

Realizace testové statistiky $\dots K = 1,0301$

Kritický obor $\dots W = \langle 7,8147, \infty \rangle$.

Testová statistika se nerealizuje v kritickém oboru, hypotézu o Poissonově rozložení tedy nezamítáme na asymptotické hladině významnosti 0,05.

Pomocí funkce plot můžeme graficky současně znázornit pozorované relativní četnosti a hodnoty teoretické pravděpodobnostní funkce: `plot(xj,nj./n, 'o', xj,pj, '*')`

Ad b) Jednoduchý test Poissonova rozložení

Realizace testové statistiky $\dots K = 572,6966$

Kritický obor $\dots W = \langle 0,510,4485 \rangle \cup \langle 643,3392; \infty \rangle$

Testová statistika se nerealizuje v kritickém oboru, hypotézu o Poissonově rozložení tedy nezamítáme na asymptotické hladině významnosti 0,05.

Odpovídající p-hodnota je 0,9614.

Upozornění: Jednoduchý test Poissonova (i exponenciálního) rozložení lze provést pomocí funkce `darling.m`, kterou lze stáhnout z Učebních materiálů v ISu.

Popis funkce `darling`:

```
function [zamitnuti,K,p,lam]=darling(X,distrib,alfa)
% TEST K OVERENI EXPONENCIALNIHO A POISSONOVA ROZLOZENI
% function [zamitnuti,K,p,lambda]=DARLING(X,ROZLOZENI,ALFA)
% X muze byt n-vektor pozorovanych velicin, jejichz rozdeleni overujeme;
%   - pro souhrnne zadana data je X tvaru (r x 2), kde prvni sloupec
%     obsahuje jednotlivé varianty a druhy sloupec četnosti;
%   - pro vypoctene statistiky je X=[n,m,s2], kde n=pocet pozorovani,
%     m=vyberovy prumer a s2=vyberovy rozptyl
% ROZLOZENI je 'exp' pro overeni exponencialniho rozlozeni (implicitni)
%   nebo 'poiss' pro overeni Poissonova rozlozeni
% ALFA je hladina vyznamnosti testu (implicitne 0.05)
%
```

% vystup: zamitnuti=1 => ZAMITAME hypotezu o shode rozdeleni
 % zamitnuti=0 => hypotezu o shode rozdeleni NEZAMITAME
 % K = hodnota testoveho kriteria
 % p = p-hodnota testu
 % lambda = odhadnuty parametr rozdeleni

Příklad 2.: Bylo zkoumáno 43 automobilů téže značky a měřena vzdálenost (v tisících km), kterou ujely, než se vyskytla první vážná porucha:

5	48	7	30	15	18	7	1	15	90	25	17	32
3	2	27	19	16	74	9	8	11	12	21	8	9
58	14	24	12	1	5	13	69	23	4	10	3	2
83	6	10	5									

Na asymptotické hladině významnosti testujte hypotézu, že počet km se řídí exponenciálním rozložením s parametrem $\lambda = 0,056$ (tzn., že střední hodnota počtu ujetých kilometrů do první vážné poruchy je 17 857). Úkol vyřešte

- pomocí Darlingova testu exponenciálního rozložení,
- pomocí Kolmogorovova – Smirnovova testu.

Výsledky:

Ad a) $K = 51,8457$, $W = \langle 0;25,9987 \rangle \cup \langle 61,7768; \infty \rangle$, H_0 tedy nezamítáme na asymptotické hladině významnosti 0,05. Odpovídající p-hodnota je 0,2839.

Ad b) K-S test

K-S test má syntaxi `[H,P,KSSTAT,CV] = kstest(X,cdf,alpha,tail)`

Význam výstupních parametrů:

Parametr H ... 0, když nulovou hypotézu nezamítáme na hladině významnosti alpha

H ... 1, když nulovou hypotézu zamítáme na hladině významnosti alpha

Parametr P ... odpovídající p-hodnota

Parametr KSSTAT ... realizace testové statistiky K-S testu (tj. absolutní hodnota maximálního rozdílu mezi distribuční funkcí testovaného spojitého rozložení a výběrovou distribuční funkcí).

Parametr CV ... ta hodnota testové statistiky, pro kterou by se již nulová hypotéza zamítla na hladině významnosti alpha.

Význam vstupních parametrů:

Parametr X ... vektor s realizacemi náhodného výběru.

Parametr cdf ... matice, která má dva sloupce, v prvním jsou hodnoty X, ve druhém hodnoty distribuční funkce testovaného spojitého rozložení v bodech X.

Parametr alpha ... zvolená hladina významnosti testu (implicitně 0,05).

Parametr tail ... specifikuje typ testu. Pokud není udán, je implicitně 0 a provede se oboustranný test. Je-li tail = -1, alternativa tvrdí, že výběrová distribuční funkce je stochasticky menší než testovaná distribuční funkce a je-li tail = 1, alternativa tvrdí, že výběrová distribuční funkce je stochasticky větší než testovaná distribuční funkce.

Upozornění: K-S test předpokládá, že testovaná distribuční funkce je plně specifikována včetně všech případných parametrů. Nejsou-li parametry rozložení předem známy a odhadují se z dat, musí se použít Lilieforsova varianta K-S testu. MATLAB však má implementovanou Lilieforsovu variantu pouze pro normální rozložení.

V našem případě zadáme příkaz:

```
[H,P,KSSTAT,CV] = kstest(x,[x,expcdf(x,1/0.056)])
```

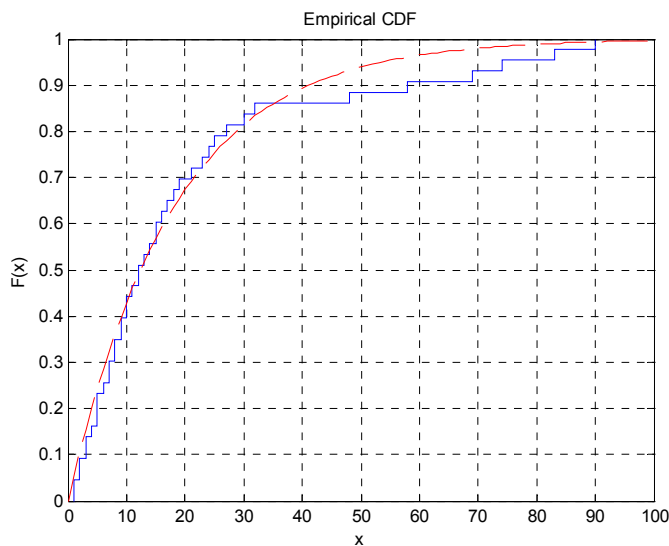
Dostaneme

$H = 0$ (tedy nulovou hypotézu nezamítáme na hladině významnosti 0,05),

$P = 0,9277$ (odpovídající p-hodnota je blízká 1, je podstatně větší než 0,05),
 $KSSTAT = 0,0814$ (naopak testová statistika se realizuje nízkou hodnotou, tedy rozdíl mezi výběrovou distribuční funkcí a distribuční funkcí rozložení $Ex(0,056)$ je malý),
 $CV = 0,2028$ (testová statistika by se musela realizovat hodnotou aspoň 0,2028, abychom na hladině významnosti 0,05 mohli zamítnout hypotézu o rozložení $Ex(0,056)$).

Graficky můžeme posoudit shodu mezi výběrovou distribuční funkcí a distribuční funkcí rozložení $Ex(0,056)$ pomocí příkazů:

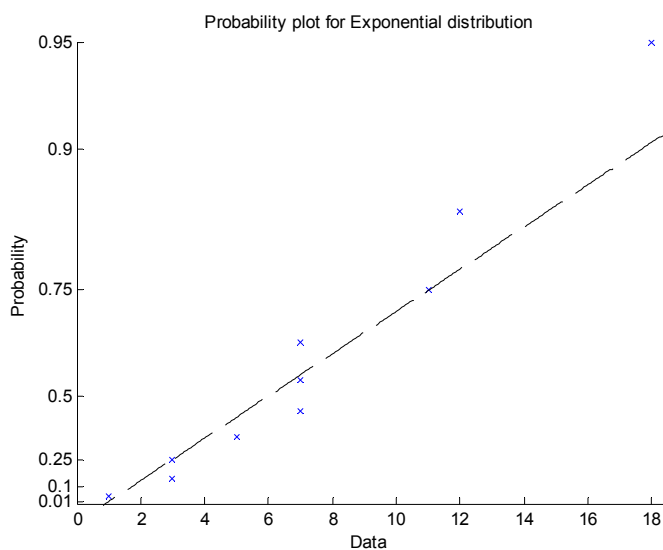
```
xx=[0:1:100]';
cdfplot(x)
hold on
plot(xx,expcdf(xx,1/0.056),'r--')
```



Příklad 3.: Jsou dány realizace náhodného výběru rozsahu 10 ze spojitého rozložení: 12 7 18 5 11 7 1 7 3 3. Shodu mezi výběrovou distribuční funkcí a distribuční funkcí exponenciálního rozložení posuďte pomocí pravděpodobnostně – pravděpodobnostního grafu.

Návod:

Pravděpodobnostně – pravděpodobnostní graf má syntaxi `probplot('distname', X)`, kde `distname` je název testovaného rozložení (v našem případě `exponential`)



Příklad 4.: Byl zjišťován počet poruch určitého zařízení za 100 hodin provozu ve 150 disjunktních 100 h intervalech. Výsledky měření:

Počet poruch za 100 hodin provozu	0	1	2	3	4 a víc
Absolutní četnost	52	48	36	10	4

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že náhodný výběr X_1, \dots, X_{150} pochází z rozložení $Po(1,2)$.

Výsledek: Testová statistika K testu dobré shody nabývá hodnoty 3,053, kritický obor je $W = \langle \chi^2_{1-\alpha}(r-p-1), \infty \rangle = \langle \chi^2_{0,95}(4), \infty \rangle = \langle 9,488; \infty \rangle$, tedy nulovou hypotézu nezamítáme na asymptotické hladině významnosti 0,05.