Testování generátorů náhodných čísel

Při použití generátoru pseudonáhodných čísel v praxi je důležité ověřit, zda získaná posloupnost čísel má vlastnosti náhodného výběru ze zadaného rozložení:

- typ rozložení
- náhodnost
- nezávislost.

Obecné zásady pro testování:

- ověřovaná posloupnost musí být dostatečně rozsáhlá
- závěry se provádějí až po prozkoumání většího počtu posloupností
- testované posloupnosti by měly mít různé výchozí hodnoty.

1. Testy shody rozložení

a) Kolmogorovův – Smirnovův test

Test je založen na porovnání teoretické a empirické distribuční funkce. Velké odchylky mezi těmito dvěma funkcemi budou svědčit o tom, že rozdíl mezi modelovými a vygenerovanými hodnotami není způsoben pouze náhodnými vlivy.
Nechť $X_1, ..., X_n$ je náhodný výběr ze spojitého rozložení s distribuční funkcí $\Phi(x)$.

Výběrovou distribuční funkci označíme $F_n(x)$, tj. pro $\forall x \in R$: $F_n(x) = \frac{1}{n}\mathrm{card}\{i; X_i \le x\}$.

Na hladině významnosti α testujeme nulovou hypotézu
H_0: $\Phi(x) = F_n(x)$ pro $\forall x \in R$
proti alternativě
H_1: $\Phi(x) \neq F_n(x)$ pro aspoň jednu hodnotu x.

Testová statistika má tvar: $D_n = \max\limits_{x \in R} \left|\Phi(x) - F_n(x)\right|$.

Nulovou hypotézu zamítáme na hladině významnosti α, když $D_n > D_{n,\alpha}$, kde $D_{n,\alpha}$ je tabelovaná kritická hodnota.

Pro větší n lze kritickou hodnotu aproximovat výrazem $D_{n,\alpha} \approx \sqrt{\frac{1}{2n}\ln\frac{2}{\alpha}}$.

Upozornění: Při testování generátorů pseudonáhodných čísel přesně známe parametry rozložení, z něhož čísla generujeme, tudíž v K-S testu nemusíme používat modifikované kritické hodnoty.
V MATLABu provádí K-S test funkce kstest.m.

Příklad 1.: Bylo vygenerováno 10 000 pseudonáhodných čísel z Rs(0,1). Vygenerované hodnoty byly roztříděny do 10 ekvidistantních třídicích intervalů. Máme k dispozici jejich meze $\left(u_j, u_{j+1}\right\rangle$ a absolutní četnosti n_j:

$\left(u_j, u_{j+1}\right\rangle$	n_j
$\left(0; 0{,}1\right\rangle$	985
$\left(0{,}1; 0{,}2\right\rangle$	1029
$\left(0{,}2; 0{,}3\right\rangle$	1005
$\left(0{,}3; 0{,}4\right\rangle$	1005
$\left(0{,}4; 0{,}5\right\rangle$	1002
$\left(0{,}5; 0{,}6\right\rangle$	970
$\left(0{,}6; 0{,}7\right\rangle$	1028
$\left(0{,}7; 0{,}8\right\rangle$	1032
$\left(0{,}8; 0{,}9\right\rangle$	960
$\left(0{,}9; 1\right\rangle$	984

Na hladině významnosti 0,05 ověřte K-S testem, že tyto hodnoty skutečně pocházejí z rozložení Rs(0,1).

Řešení: Tabulku absolutních četností doplníme o hodnoty teoretické a výběrové distribuční funkce a vypočteme absolutní hodnoty jejich rozdílů. Největší z těchto absolutních hodnot porovnáme s kritickou hodnotou a pak rozhodneme o nulové hypotéze.

$\langle u_j, u_{j+1}\rangle$	n_j	N_j	$\Phi(x)$	$F_{10000}(x)$	$\lvert\Phi(x) - F_{10000}(x)\rvert$
$\langle 0; 0{,}1\rangle$	985	985	0,1	0,0985	0,0015
$\langle 0{,}1; 0{,}2\rangle$	1029	2014	0,2	0,2014	0,0014
$\langle 0{,}2; 0{,}3\rangle$	1005	3019	0,3	0,3019	0,0019
$\langle 0{,}3; 0{,}4\rangle$	1005	4024	0,4	0,4024	0,0024
$\langle 0{,}4; 0{,}5\rangle$	1002	5026	0,5	0,5026	0,0026
$\langle 0{,}5; 0{,}6\rangle$	970	5996	0,6	0,5996	0,0004
$\langle 0{,}6; 0{,}7\rangle$	1028	7024	0,7	0,7024	0,0024
$\langle 0{,}7; 0{,}8\rangle$	1032	8056	0,8	0,8056	0,0056
$\langle 0{,}8; 0{,}9\rangle$	960	9016	0,9	0,9016	0,0016
$\langle 0{,}9; 1\rangle$	984	10000	1	1	0

Největší rozdíl v absolutní hodnotě je 0,0056. To je realizace testové statistiky, kterou porovnáme s aproximací kritické hodnoty počítané podle vzorce $D_{n,\alpha} \approx \sqrt{\frac{1}{2n}\ln\frac{2}{\alpha}}$:

$$D_{10000;0,05} \approx \sqrt{\frac{1}{2\cdot 10000}\ln\frac{2}{0{,}05}} = 0{,}0136 .$$

Jelikož testová statistika je menší než kritická hodnota, nelze na hladině významnosti 0,05 zamítnout hypotézu, že vygenerovaná data pocházejí z rozložení Rs(0,1).

Kolmogorovův - Smirnovův test pro rozložení Rs(0,1) v MATLABu

Vygenerujeme n = 10000 realizací z Rs(0,1):

```
n=10000;
realizace=unifrnd(0,1,n);
```

Vypočteme hodnoty distribuční funkce rozložení Rs(0,1) v bodech vygenerovaných realizací:

```
Fi=unifcdf(realizace,0,1);
```

Zavoláme funkci kstest:

```
[h,p,ksstat,cv]=kstest(realizace,[realizace,Fi])
```

Vstupní parametry:

`realizace` … sloupcový vektor realizací

`[realizace,Fi]` … matice nx2 obsahující vektor realizací a vektor hodnot distribuční funkce rozložení Rs(0,1)

Výstupní parametry:

`h` … nabývá hodnoty 0, když nulovou hypotézu nezamítáme na hladině významnosti 0,05 a hodnoty 1, když zamítáme na hladině významnosti 0,05

`p` … p-hodnota

`ksstat` … hodnota testové statistiky

`cv` … kritická hodnota

b) Test χ^2 dobré shody

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

Spojitý případ: 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 $p_j = \Phi(u_{j+1}) - \Phi(u_j)$.

Diskrétní případ: 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 $p_j = \Phi\left(x_{[j]}\right) - \lim_{x \to x_{[j]}-} \Phi(x) = P\left(X = x_{[j]}\right)$.

Provedení testu pro diskrétní i spojitý případ

Testová statistika: $K = \sum_{j=1}^{r} \frac{\left(n_j - np_j\right)^2}{np_j}$. 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 testová statistika $K \geq \chi^2_{1-\alpha}(r-1-p)$.

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

Závažnost rozdílu mezi pozorovanými četnostmi a teoretickými četnostmi lze pro každý index j orientačně posoudit porovnáním vypočítaného sčítance $\frac{\left(n_j - np_j\right)^2}{np_j}$

s hodnotou 3,84 (to je kvantil $\chi^2_{0,95}(1)$). Jestliže některý sčítanec převýší tuto hodnotu, lze předpokládat, že odchylka od modelu se nachází právě v této oblasti hodnot.

χ^2 test dobré shody v MATLABu provádí funkce chi2gof.m

Příklad 2.: Na hladině významnosti 0,05 ověřte χ^2 testem dobré shody, zda data z příkladu 1 pocházejí z rozložení Rs(0,1).

Řešení: Tabulku absolutních četností doplníme o pravděpodobnosti p_j, teoretické četnosti np_j a jednotlivé sčítance $\frac{(n_j - np_j)^2}{np_j}$ vstupující do testové statistiky K.

$\langle u_j, u_{j+1} \rangle$	n_j	p_j	np_j	$\frac{(n_j - np_j)^2}{np_j}$
$\langle 0; 0{,}1 \rangle$	985	0,1	1000	0,225
$\langle 0{,}1; 0{,}2 \rangle$	1029	0,1	1000	0,841
$\langle 0{,}2; 0{,}3 \rangle$	1005	0,1	1000	0,025
$\langle 0{,}3; 0{,}4 \rangle$	1005	0,1	1000	0,025
$\langle 0{,}4; 0{,}5 \rangle$	1002	0,1	1000	0,004
$\langle 0{,}5; 0{,}6 \rangle$	970	0,1	1000	0,900
$\langle 0{,}6; 0{,}7 \rangle$	1028	0,1	1000	0,784
$\langle 0{,}7; 0{,}8 \rangle$	1032	0,1	1000	1,024
$\langle 0{,}8; 0{,}9 \rangle$	960	0,1	1000	1,600
$\langle 0{,}9; 1 \rangle$	984	0,1	1000	0,256

Testová statistika:

$$K = \sum_{j=1}^{r} \frac{(n_j - np_j)^2}{np_j} = 0{,}225 + 0{,}841 + \ldots + 0{,}256 = 5{,}684$$

Kritický obor: $W = \langle \chi^2_{0,95}(9), \infty) = \langle 16{,}919; \infty)$

Protože se testová statistika nerealizuje v kritickém oboru, nemůžeme na hladině významnosti 0,05 zamítnout hypotézu, že data pocházejí z rozložení Rs(0,1).

Test dobré shody pro rozložení Rs(0,1) v MATLABu

Funkce chi2gof.m implicitně třídí data do 10 intervalů.

`[h,p] = chi2gof(realizace,'cdf',@unifcdf)`

Vstupní parametry:

`realizace` … sloupcový vektor realizací

`'cdf'` … parametr, který dává funkci na vědomí, že bude použita distribuční funkce nějakého rozložení

`@unifcdf` … označení distribuční funkce rovnoměrného spojitého rozložení

Výstupní parametry:

`h` … nabývá hodnoty 0, když nulovou hypotézu nezamítáme na hladině významnosti 0,05 a hodnoty 1, když zamítáme na hladině významnosti 0,05

`p` … p-hodnota

2. Testy náhodnosti

Nechť $x_1, \ldots, x_n$ je posloupnost navzájem různých čísel generovaných ze spojitého rozložení (jsou-li dvě sousední hodnoty stejné, jednu vyškrtneme). Na hladině významnosti α testujeme nulovou hypotézu H_0: posloupnost je náhodná proti alternativě H_1: posloupnost není náhodná.

a) Test založený na bodech zvratu

Tento test zkoumá, zda kolísání hodnot podle velikosti se v dané posloupnosti mění dostatečně rychle. Není vhodný pro testování existence trendu, protože vychází pouze z lokálních vlastností posloupnosti.

Číslo x_i se nazývá bodem zvratu, když obě sousední čísla jsou současně buď větší než x_i nebo menší než x_i, tj. platí-li buď $x_{i-1} > x_i < x_{i+1}$ nebo $x_{i-1} < x_i > x_{i+1}$.

Konstrukce testové statistiky:

Označme Y celkový počet bodů zvratu v posloupnosti $x_1, \ldots, x_n$. Platí-li H_0, pak statistika Y má asymptoticky normální rozložení se střední hodnotou $E(Y) = \frac{2(n-2)}{3}$ a rozptylem $D(Y) = \frac{16n-29}{90}$,

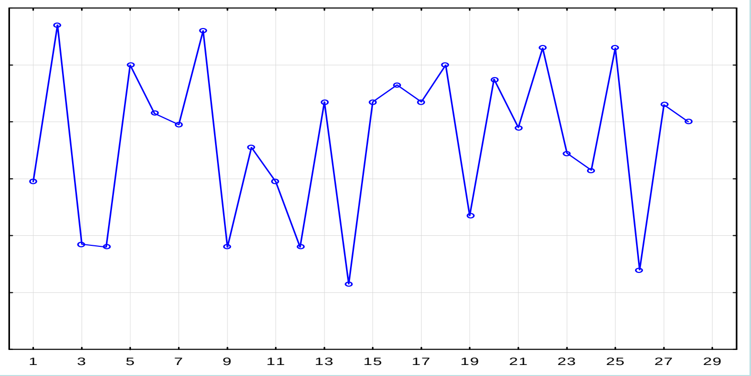
tedy standardizovaná statistika $U = \dfrac{Y - \frac{2(n-2)}{3}}{\sqrt{\frac{16n-29}{90}}} \approx N(0,1)$.

Kritický obor: $W = \left(-\infty, -u_{1-\alpha/2}\right\rangle \cup \left\langle u_{1-\alpha/2}, \infty\right)$

Pokud $U \in W$, nulovou hypotézu zamítáme na asymptotické hladině významnosti α.

Příklad 3: Bylo vygenerováno 28 pseudonáhodných čísel z Rs(0,1):
0,39 0,94 0,17 0,16 0,80 0,63 0,59 0,92 0,16 0,51 0,39 0,16 0,67 0,03 0,67 0,73 0,67 0,80 0,27 0,75 0,58 0,86 0,49 0,43 0,86 0,08 0,66 0,60
Pomocí testu založeného na bodech zvratu ověřte na hladině významnosti 0,05 hypotézu, že tato posloupnost je náhodná.

Řešení:



Celkem zjistíme Y = 21 bodů zvratu.

Testová statistika: $$U = \frac{Y - \frac{2(n-2)}{3}}{\sqrt{\frac{16n-29}{90}}} = \frac{21 - \frac{2(28-2)}{3}}{\sqrt{\frac{16\cdot 28-29}{90}}} = 1{,}699.$$

Kritický obor: $W = \left(-\infty, -u_{0,975}\right\rangle \cup \left\langle u_{0,975}, \infty\right) = \left(-\infty, -1{,}96\right\rangle \cup \left\langle 1{,}96, \infty\right)$.

Protože testová statistika s nerealizuje v kritickém oboru, nelze na hladině významnosti 0,05 zamítnout hypotézu, že daná posloupnost je náhodná.

Výpočet v MATLABu:

Test založený na bodech zvratu je prováděn pomocí funkce body_zvratu.m:

```
function [H,P,U]=body_zvratu(x,alfa)
% funkce testuje nahodnost posloupnosti pomoci bodu zvratu
% synatxe: [H,P,U]=body_zvratu(x,alfa)
% vystupni parametry:
% H ... vysledek testu: 1 ... H0 zamitame, 0 ... H0 ... nezamitame
% P ... vypocitana p-hodnota
% U ... realizace testove statistiky
% vstupni parametry:
% x ... sloupcovy vektor hodnot testovane posloupnosti
% alfa ... hladina vyznamnosti
n=size(x,1);
Y=0;
for i=2:n-1
    if ((x(i)>x(i-1))&(x(i)>x(i+1)))|((x(i)<x(i-1))&(x(i)<x(i+1)))
        Y=Y+1;
    end
end;
U=(Y-(2*n-4)/3)/sqrt((16*n-29)/90);
P=2*min(normcdf(U,0,1),1-normcdf(U,0,1));
if P<=alfa
    H=1;
end
if P>alfa
    H=0;
end
```

Použijeme-li tuto funkci na data z příkladu 3, dostaneme výsledky:

```
[H,P,U]=body_zvratu(x,alfa)

H =
     0
P =
    0.0893
U =
   1.6994
```

Protože p-hodnota je 0,0893, nemůžeme na hladině významnosti 0,05 zamítnout hypotézu o náhodnosti dané posloupnosti.

b) Test znamének diferencí

Tento test zkoumá, zda posloupnost neobsahuje dlouhé řady čísel jdoucích za sebou vzestupně nebo sestupně. Používá se k ověření existence trendu.

Test je založen na počtu kladných 1. diferencí dané posloupnosti, tj. na počtu bodů růstu. Číslo x_i se nazývá bodem růstu, když $x_i < x_{i+1}$.

Konstrukce testové statistiky: Označme Y celkový počet bodů růstu v posloupnosti $x_1, \dots, x_n$.

Platí-li H_0, pak statistika Y má asymptoticky normální rozložení se střední hodnotou

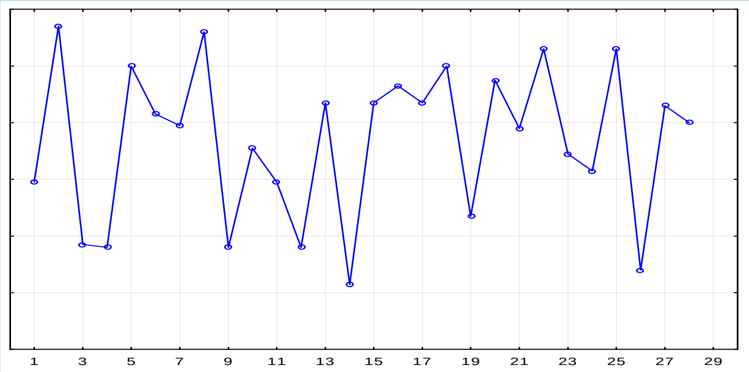
$E(Y) = \frac{n-1}{2}$ a rozptylem $D(Y) = \frac{n+1}{12}$, tedy standardizovaná statistika $U = \dfrac{Y - \frac{n-1}{2}}{\sqrt{\frac{n+1}{12}}} \approx N(0,1)$.

Kritický obor: $W = \left(-\infty, -u_{1-\alpha/2}\right\rangle \cup \left\langle u_{1-\alpha/2}, \infty\right)$

Pokud $U \in W$, nulovou hypotézu zamítáme na asymptotické hladině významnosti α.

Příklad 4.: Na hladině významnosti 0,05 testujte pomocí testu znamének diferencí hypotézu, že data z příkladu 3 jsou náhodná.

Řešení:



Celkem zjistíme Y = 12 bodů růstu.

Testová statistika: $$U = \frac{Y - \frac{n-1}{2}}{\sqrt{\frac{n+1}{12}}} = \frac{12 - \frac{28-1}{2}}{\sqrt{\frac{28+1}{12}}} = -0{,}9649.$$

Kritický obor: $$W = \left(-\infty, -u_{0,975}\right\rangle \cup \left\langle u_{0,975}, \infty\right) = \left(-\infty, -1{,}96\right\rangle \cup \left\langle 1{,}96, \infty\right).$$

Protože testová statistika s nerealizuje v kritickém oboru, nelze na hladině významnosti 0,05 zamítnout hypotézu, že daná posloupnost je náhodná.

Výpočet v MATLABu:

Test znamének diferencí je prováděn pomocí funkce znamenka_diferenci.m:

```
function [H,P,U]=znamenka_diferenci(x,alfa)
% funkce testuje nahodnost posloupnosti pomoci znamenek diferenci
% synatxe: [H,P,U]=znamenka_diferenci(x,alfa)
% vystupni parametry:
% H ... vysledek testu: 1 ... H0 zamitame, 0 ... H0 ... nezamitame
% P ... vypocitana p-hodnota
% U ... realizace testove statistiky
% vstupni parametry:
% x ... sloupcovy vektor hodnot testovane posloupnosti
% alfa ... hladina vyznamnosti
n=size(x,1);
Y=0;
for i=1:n-1
    if x(i)<x(i+1)
        Y=Y+1;
    end
end;
U=(Y-(n-1)/2)/sqrt((n+1)/12);
P=2*min(normcdf(U,0,1),1-normcdf(U,0,1));
if P<=alfa
    H=1;
end
if P>alfa
    H=0;
end
```

Použijeme-li tuto funkci na data z příkladu 3, dostaneme výsledky:

```
[H,P,U]=znamenka_diferenci(x,alfa)
H =
     0
P =
    0.3346
U =
   -0.9649
```

Protože p-hodnota je 0,3346, nemůžeme na hladině významnosti 0,05 zamítnout hypotézu o náhodnosti dané posloupnosti.

c) Test založený na Spearmanově koeficientu

Tento test zkoumá, zda velikost generované hodnoty nezávisí na pořadí, v němž bylo číslo generováno (např. zda na počátku generované posloupnosti nejsou soustředěny nízké hodnoty a na konci vysoké).

Na základě generované posloupnosti $x_1, \ldots, x_n$ utvoříme dvojice $(1, x_1), \ldots, (n, x_n)$. Předpokládáme, že tyto dvojice pocházejí z dvourozměrného rozložení s teoretickým Spearmanovým koeficientem pořadové korelace ρ_S.

Na hladině významnosti α testujeme nulovou hypotézu H_0: $\rho_S = 0$ proti H_1: $\rho_S \neq 0$.

Označme R_i pořadí hodnoty x_i v dané posloupnosti. Vypočteme Spearmanův koeficient pořadové korelace:

$r_S = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{n} (i - R_i)^2$. Nulovou hypotézu zamítáme na hladině významnosti α ve prospěch alternativy, když $|r_S| \geq r_{S,1-\alpha/2}(n)$, kde $r_{S,1-\alpha/2}(n)$ je kritická hodnota, kterou pro $\alpha = 0{,}05$ a $n \leq 30$ najdeme v tabulkách.

Kritické hodnoty pro Spearmanův koeficient pořadové korelace pro n = 5, 6, …, 30, α = 0,05

n	$r_{S;0,975}$	$r_{S;0,95}$
5	0,9	0,8
6	0,8286	0,7714
7	0,745	0,6786
8	0,6905	0,5952
9	0,6833	0,5833
10	0,6364	0,5515
11	0,6091	0,5273
12	0,5804	0,4965
13	0,5549	0,478
14	0,5341	0,4293
15	0,5179	0,4429
16	0,5	0,4265
17	0,4853	0,4118
18	0,4716	0,3994
19	0,4579	0,3895
20	0,4451	0,3789
21	0,4351	0,3688
22	0,4241	0,3597
23	0,415	0,3518
24	0,4061	0,3435
25	0,3977	0,3362
26	0,3894	0,3299
27	0,3822	0,3236
28	0,3749	0,3175
29	0,3685	0,3113
30	0,362	0,3059

Příklad 5: Pro data z příkladu 3 testujte na hladině významnosti 0,05 hypotézu, že velikost generované hodnoty nezávisí na pořadí. Výpočet proveďte pomocí MATLABu.

Řešení: Ve statistickém toolboxu MATLABu je implementována funkce tiedrank(x), která pro daný vektor x poskytne vektor pořadí, přičemž pro skupinky stejných hodnot spočítá průměrné pořadí.
Postup výpočtu:
Do proměnné x vložíme dané hodnoty.
Utvoříme vektor y=[1:28]';
Pomocí funkce tiedrank zjistíme vektor pořadí: R=tiedrank(x);
Pomocí funkce corrcoef spočteme koeficient korelace a odpovídající p-hodnotu: [rs,p]=corrcoef(y,R)
Dostaneme

```
r =                       p =
    1.0000    0.0729         1.0000    0.7124
    0.0729    1.0000         0.7124    1.0000
```

Velikost Spearmanova koeficientu (rs = 0,0729) svědčí o velmi slabé přímé pořadové závislosti, která je na hladině 0,05 nevýznamná (p = 0,7124).

Asymptotické varianty testu

$n > 20$:

lze použít testovou statistiku $T_0 = \frac{r_S\sqrt{n-2}}{\sqrt{1-r_S^{\ 2}}}$, která se v případě platnosti nulové hypotézy asymptoticky

řídí rozložením t(n-2).

Kritický obor: $W = \left(-\infty, -t_{1-\alpha/2}(n-2)\right\rangle \cup \left\langle t_{1-\alpha/2}(n-2), \infty\right)$

Hypotézu, že velikost generovaných hodnot nezávisí na pořadí, zamítáme na asymptotické hladině

významnosti α, když $t_0 \in W$.

$n > 30$:

lze použít testovou statistiku $r_s\sqrt{n-1}$. Platí-li H_0, pak $r_s\sqrt{n-1} \approx N(0, 1)$. Nulovou hypotézu tedy zamítáme

na asymptotické hladině významnosti α ve prospěch alternativy, když

$r_S\sqrt{n-1} \in \left(-\infty, -u_{1-\alpha/2}\right\rangle \cup \left\langle u_{1-\alpha/2}, \infty\right)$.

3. Testy nezávislosti

a) Test založený na koeficientu autokorelace

Tímto testem ověřujeme, zda existuje lineární závislost mezi sousedními nebo i vzdálenějšími členy posloupnosti $x_1, \dots, x_n$.

Pro $k < n$ je výběrový koeficient autokorelace k-tého řádu r_k definován jako výběrový Pearsonův koeficient korelace dvojic $(x_1, x_{k+1}), \dots, (x_{n-k}, x_n)$, o nichž předpokládáme, že pocházejí z dvourozměrného rozložení s koeficientem korelace ρ_k.

Je-li $k = 1$, počítá se koeficient korelace mezi sousedními členy generované posloupnosti. Není vhodné počítat koeficienty autokorelace pro $k > n/4$.

Posloupnost koeficientů autokorelace $r_1, r_2, \dots$ se nazývá korelogram.

Na hladině významnosti α testujeme nulovou hypotézu H_0: $\rho_k = 0$ proti H_1: $\rho_k \neq 0$.

Testová statistika $T_0 = r_k\sqrt{n-k}$ se v případě platnosti nulové hypotézy asymptoticky řídí rozložením N(0,1).

Kritický obor: $W = \left(-\infty, -u_{1-\alpha/2}\right\rangle \cup \left\langle u_{1-\alpha/2}, \infty\right)$

Pokud $T_0 \in W$, nulovou hypotézu zamítáme na asymptotické hladině významnosti α, tedy hodnoty generované posloupnosti po k členech nelze považovat za lineárně nezávislé.

Příklad 6: Pro data z příkladu 3 testujte na hladině významnosti 0,05 hypotézu, že mezi nimi neexistuje autokorelace 1. až 7. řádu.

Řešení:

Tabulka pro výpočet koeficientů autokorelace 1. až 7. řádu

	1 X(i)	2 X(i+1)	3 X(i+2)	4 X(i+3)	5 X(i+4)	6 X(i+5)	7 X(i+6)	8 X(i+7)
1	0,39	0,94	0,17	0,16	0,8	0,63	0,59	0,92
2	0,94	0,17	0,16	0,8	0,63	0,59	0,92	0,16
3	0,17	0,16	0,8	0,63	0,59	0,92	0,16	0,51
4	0,16	0,8	0,63	0,59	0,92	0,16	0,51	0,39
5	0,8	0,63	0,59	0,92	0,16	0,51	0,39	0,16
6	0,63	0,59	0,92	0,16	0,51	0,39	0,16	0,67
7	0,59	0,92	0,16	0,51	0,39	0,16	0,67	0,03
8	0,92	0,16	0,51	0,39	0,16	0,67	0,03	0,67
9	0,16	0,51	0,39	0,16	0,67	0,03	0,67	0,73
0	0,51	0,39	0,16	0,67	0,03	0,67	0,73	0,67
1	0,39	0,16	0,67	0,03	0,67	0,73	0,67	0,8
2	0,16	0,67	0,03	0,67	0,73	0,67	0,8	0,27
3	0,67	0,03	0,67	0,73	0,67	0,8	0,27	0,75
4	0,03	0,67	0,73	0,67	0,8	0,27	0,75	0,58
5	0,67	0,73	0,67	0,8	0,27	0,75	0,58	0,86
6	0,73	0,67	0,8	0,27	0,75	0,58	0,86	0,49
7	0,67	0,8	0,27	0,75	0,58	0,86	0,49	0,43
8	0,8	0,27	0,75	0,58	0,86	0,49	0,43	0,86
9	0,27	0,75	0,58	0,86	0,49	0,43	0,86	0,08
0	0,75	0,58	0,86	0,49	0,43	0,86	0,08	0,66
1	0,58	0,86	0,49	0,43	0,86	0,08	0,66	0,6
2	0,86	0,49	0,43	0,86	0,08	0,66	0,6	
3	0,49	0,43	0,86	0,08	0,66	0,6		
4	0,43	0,86	0,08	0,66	0,6			
5	0,86	0,08	0,66	0,6				
6	0,08	0,66	0,6					
7	0,66	0,6						
8	0,6							

Korelogram a hodnoty testových statistik

i	1 r_i	2 n-i	3 T_0	4 kvantil	5 rozhodnutí
1	-0,3235	27	-1,68096	1,959964	nezamítáme
2	0,0523	26	0,266679	1,959964	nezamítáme
3	0,1708	25	0,854	1,959964	nezamítáme
4	-0,4572	24	-2,23981	1,959964	zamítáme
5	0,3121	23	1,496779	1,959964	nezamítáme
6	-0,2657	22	-1,24624	1,959964	nezamítáme
7	0,0285	21	0,130603	1,959964	nezamítáme

Závěr: V testovaných datech byla na hladině významnosti 0,05 prokázána autokorelace 4. řádu.

b) Cochranův test

Na hladině významnosti 0,05 testujeme nulovou hypotézu
H_0: $\rho_1 = \rho_2 = \dots \rho_k = 0$ (tj. všechny koeficienty autokorelace až do řádu k jsou nulové)
proti alternativě
H_1: $\rho_i \neq 0$ pro aspoň jeden index i.

Testová statistika:

$$Q = n\sum_{i=1}^{k} r_i^2$$

Platí-li nulová hypotéza, Q se asymptoticky řídí rozložením $\chi^2(k)$.
H_0 tedy zamítáme na asymptotické hladině významnosti α, když $Q \geq \chi^2_{1-\alpha}(k)$.
Cochranův test má význam zvláště v případě, kdy jeden z autokorelačních koeficientů je významný (přitom může být významný pouze nahodile) a ostatní významné nejsou.

Příklad 7: Pro data z příkladu 3 proveďte na hladině významnosti Cochranův test.
Řešení: Připomeňme, že koeficienty autokorelace byly

r_i	-0,3235	0,0523	0,1708	-0,4572	0,3121	-0,2657	0,0285

Po dosazení do vzorce pro testovou statistiku dostaneme Q = 14,4034.
Odpovídající kvantil: $\chi^2_{1-\alpha}(k) = \chi^2_{0,95}(7) = 14,067$.
Protože $Q \geq 14,067$, H_0 zamítáme na asymptotické hladině významnosti 0,05.

Výpočet v MATLABu:

Cochranův test je prováděn pomocí funkce Cochran.m:

```
function [Q,chi,h,p]=Cochran(x,rad,alpha)
% funkce provadi Cochranuv test autokorelace
% vstupni parametry: x - vektor realizaci
% rad - nejvyssi rad autokorelace
% alpha - hladina vyznamnosti
% vystupni parametry:
% Q - hodnota testove statistiky
% chi - kriticka hodnota
% h zamitnuti (1)/nezamitnuti (0) nulove hypotezy
% p-hodnota testu
n=size(x,1);
if rad>n/4
    error('Rad autokorelace je prilis velky')
end;
Q=0;
for i=1:rad
    [r,p]=corrcoef(x(1:n-i),x(i+1:n));Q=Q+r(1,2)^2;
end;
Q=n*Q;chi=chi2inv((1-alpha),rad);
if Q<chi
  h=0;
else
    h=1;
end;
p=1-chi2cdf(Q,rad);
```