Generovanie náhodných čísel

Náhodné čísla sú dôležitá súčasť výpočtov v:

- modelovaní a simuláciách
- numerickej analýze
- rozhodovaní
- počítačovej grafike
- kryptografii
- dátovej komunikácií
- ...

Základné spôsoby získania postupnosti náhodných čísel:

1) Hardwarové generátory (snímanie hodnôt nejakého fyzikálneho deja)
2) Tabuľky náhodných čísel (uložené napr. na CDROM)
3) Generovanie náhodných čísel určitým algoritmom z počiatočných hodnôt.

Súčasnosť ➔ používa sa najmä spôsob 3 - generovanie náhodných čísel algoritmom

Výhody

- rýchla SW implementácia s malými nárokmi na pamäť
- Oproti spôsobu 1 má výhodu opakovateľnosti
- oproti spôsobu 2 výhodu kompaktnosti

Vlastnosti

- získaná postupnosť čísel je v podstate deterministická ➔ , hovoríme o generovaní, resp. generátoroch ***pseudonáhodných*** čísel (GPČ)
- vlastnosti odpovedajúceho generátora je treba poznať, resp. starostlivo testovať, aby sme mohli posúdiť jeho vhodnosť resp. nevhodnosť pre danú oblasť použitia.
- Získané postupnosti náhodných čísel majú zvyčajne ***rovnomerné resp. kváziromerné (=diskrétne rovnomerné) rozdelenie***
- Je možné dosiahnuť iba ***konečnú periódu*** generovaných dát

Generátory pseudonáhodných čísel (GPČ)

Najpoužívanejšie riešenie - metódy využívajúce ***rekurenciu***, t.j. vzťah, kde výstupné hodnoty sú generované na základe vzťahu

$$x_n = f(x_{n-1}, \ldots, x_{n-k}) \qquad 0 \leq k \leq n$$

História:
Prvá metóda tohto typu - ***metóda stredných rádov*** (autor J. von Neuman 1946).
Nasledovné číslo bolo tvorené strednými číslicami druhej mocniny čísla predchádzajúceho.

***Príklad**:*

Generujteposupnosť pseudonáhodných čísiel pomocou metódy stredných rádov. Začnite z čísla 6100.

***Riešenie**:*

6100^2=37210000; 2100^2=4410000; 4100^2=16810000; 8100^2=65610000

T.j. výsledkom jepostupnosť (6100, 2100, 4100, 8100, 6100, ...).
Vidíme, že dĺžka cyklu je 5.

Pravidlá pri generovaní pseudonáhodných čísel:

- pracujeme vo všeobecnosti v ***matematike modulo m***, t.j. na množine čísel

$$Z_m = \{0,1,2,\ldots,m-1\}$$

- pojem rovnosť nahrádza pojem ***kongruencia***
  Ak m je prirodzené číslo, potom hovoríme:
  a a b sú ***kongruentné modulo m*** ak majú po delení číslom m ten istý zvyšok.
  Píšeme:

$$a \equiv b \pmod{m}$$

Postupnosť pseudonáhodných čísel u_n (pseudo)náhodnej premennej U s rovnomerným rozdelením zvyčajne generujeme v ***dvoch fázach***:

1) rekurentný vzťah: $x_n = f(x_{n-1}, \ldots, x_{n-k})$

2) výstupná hodnota: $u_n = x_n / m, \quad u_n \in <0,1)$

Pre strednú hodnotu a rozptyl náhodnej premennej U s ***kvázirovnomerným*** rozdelením platí:

$$E(U) = \sum_{j=0}^{m-1} u_j p_j = \sum_{j=0}^{m-1} \frac{j}{m} \frac{1}{m} = \frac{1}{2}\left(1 - \frac{1}{m}\right) \qquad D(U) = E(U^2) - (E(U))^2 = \frac{1}{12}\left(1 - \frac{1}{m^2}\right)$$

Základné vlastnosti, ktoré by mal dobrý generátor spĺňať:

1) ***dlhá perióda*** – v ideálnom prípade nekonečne dlhá, v súčasnosti napr. generátor Mersenne Twister dosahuje periódu $2^{19937}-1$

2) ***rovnomerné a s rastúcou dĺžkou sekvencie čoraz lepšie zaplnenie intervalu*** (pokiaľ možno malo by platiť aj pre ľubovoľné subsekvencie)

3)***opakovatelnosť*** schopnosť opakovane generovať tú istú sekvenciu pseudonáhodných čísel pomocou jednoducho špecifikovaných počiatočných podmienok

4) ***čas generovania*** – zanedbateľný oproti času operácií ktoré vytvorenú sekvenciu používajú

5) ***požiadavky na pamäť a portabilita*** – implementácia nenáročná na pamäť a vo vyššom programovacom jazyku

6) ***dobré štatistické vlastnosti*** – generátor musí úspešne zdolať súbor štatistických testov, teoretických (ak sú k dispozícii) aj empirických, ktoré je výhodné cielene voliť vzhľadom na oblasť použitia generátora.

Základné typy GPČ.

Lineárne kongruenčné generátory (LCG)

Generátory tohoto typu zaviedol Lehmer v r.1949. Hodnoty sú generované pomocou rekurentného vzťahu:

$$x_n = (ax_{n-1} + c) \bmod m$$

Označujeme ho ako: ***LCG(m,a,c,x_0).***

Platí:

- Pri $c = 0$, hovoríme o ***multiplikatívnom*** LCG
- pri $c \neq 0$ o ***zmiešanom*** LCG.

Maximálna perióda LCG je *m* a je dosiahnutá, keď:

a) *c* a *m* sú nesúdeliteľné

b) *a-1* je násobkom každého prvočiniteľa *m*

c) *a-1* je násobkom *4*, ak *m* je násobkom *4*

Potencia

Jednoduchým kritériom na posudzovanie náhodnosti LCG - koncept tzv. ***potencie***. Potencia je najmenšie číslo *s*, pre ktoré platí

$$(a-1)^s \equiv 0 \mod m$$

za dostatočne dobrú je považovaná hodnota $s \geq 5$.

Všeobecne známe LCG a ich koeficienty

Názov, použitie LCG	m	a	c	X_0
RANDU –počítače IBM(1960)	2^{31}	65539	0	X_0
ANSI C – funkcia rand()	2^{31}	1103515245	12345	12345
Program DERIVE	2^{32}	3141592653	1	0
Jazyk SIMULA	2^{35}	5^{15}	0	1
ANSI C – funkcia drand48()	2^{48}	25214903917	11	0
Program MAPLE	10^{12}-11	427419669081	0	1
Program MATHEMATICA	a^{48}- a^{8}+1	$a = 2^{31}$	0	1

Špeciálne prípady LCG a ich vlastnosti

Prípad	max. perióda	Dosiahnutie max. periódy	Poznámka
$M=2^M,\ c > 0$	2^M	a mod $4 = 1$ c je nepárne	najnižšie bity nemajú náhodný charakter
$M=2^M,\ c = 0$	2^{M-2}	a mod $8 = 3$ alebo 5 x_0 je nepárne	najnižšie bity nemajú náhodný charakter
M – prvočíslo	$m-1$	$x_0 \neq 0$ alebo $c \neq 0$ a je primitívny element*) Z_m	najnižšie bity môžu mať náhodný charakter

primitívny element - prvok, ktorý je schopný svojimi mocninami generovať Z_m, t.j. $Z_m = \{a^0, a^1, a^2, ..., a^{m-1}\}$.

Ako zistiť či číslo p je primitívny element?

> Ak m je prvočíslo a c_i prvočinitele $m-1$, potom p je primitívny element Z_m, ak pre žiadny c_i neplatí $p^{(m-1)/c_i} \equiv 1$

Spektrálne charakteristiky LCG

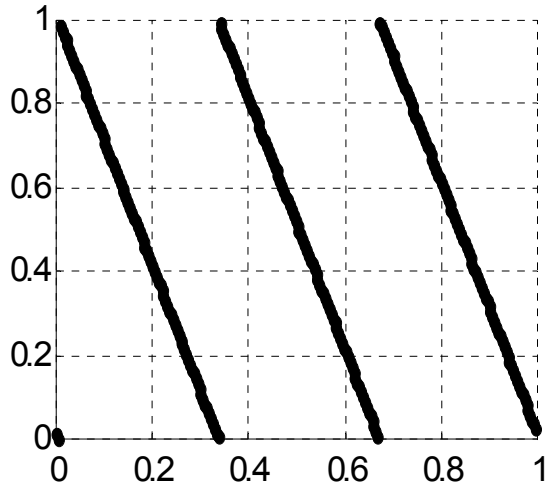
Označme u_n výstupy z LCG a vytvorme *N*-tice:

$$S_n^N = \left(u_n, u_{n+1}, \ldots, u_{n+N-1}\right)$$

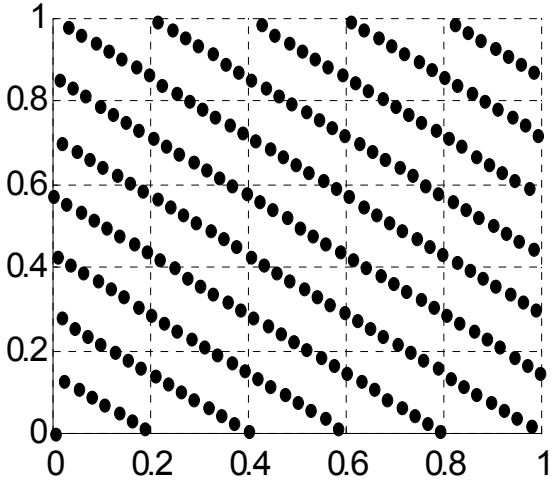
Použijeme ich ako súradnice bodov v *N*-rozmernom priestore.

Výsledok:

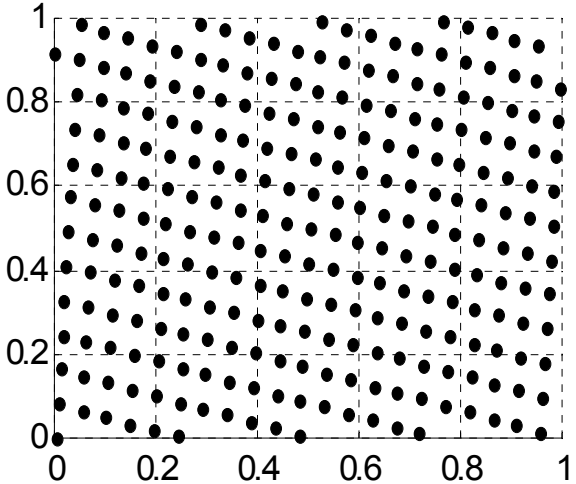
- *mriežková štruktúra* pri vypĺňaní *N*-rozmerného intervalu
- *t.j.* body S_n^N ležia na množine *ekvidištančných paralelných hyperrovín*.
- aj napriek tomu sú LCG generátory vo všeobecnosti najviac používané.



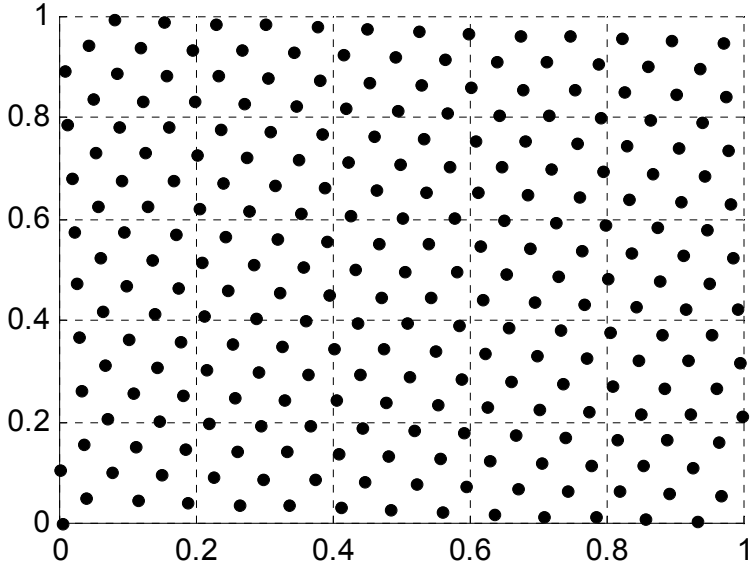
a)LCG(256,85,1,0)



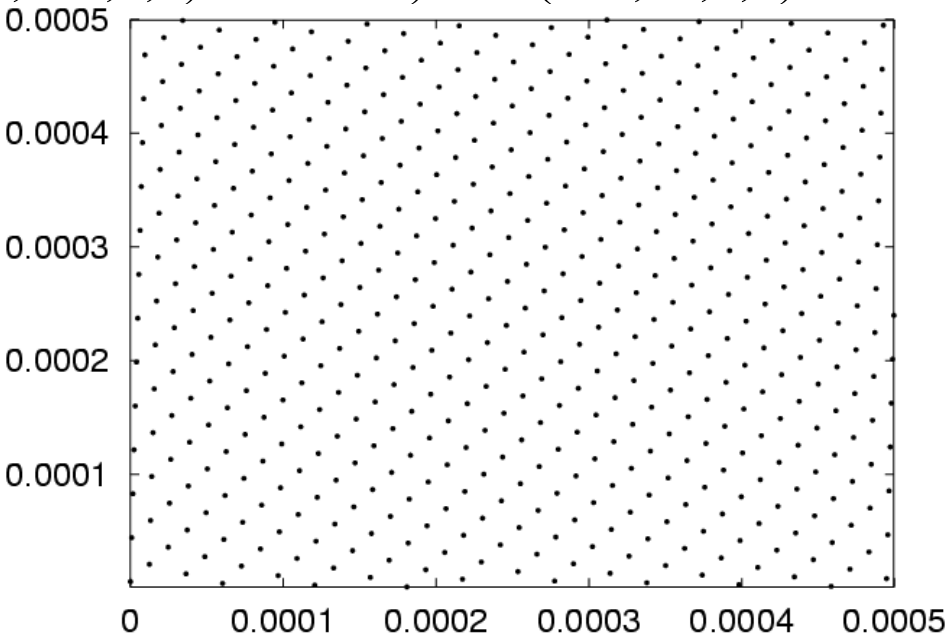
b)LCG(256,101,1,0)



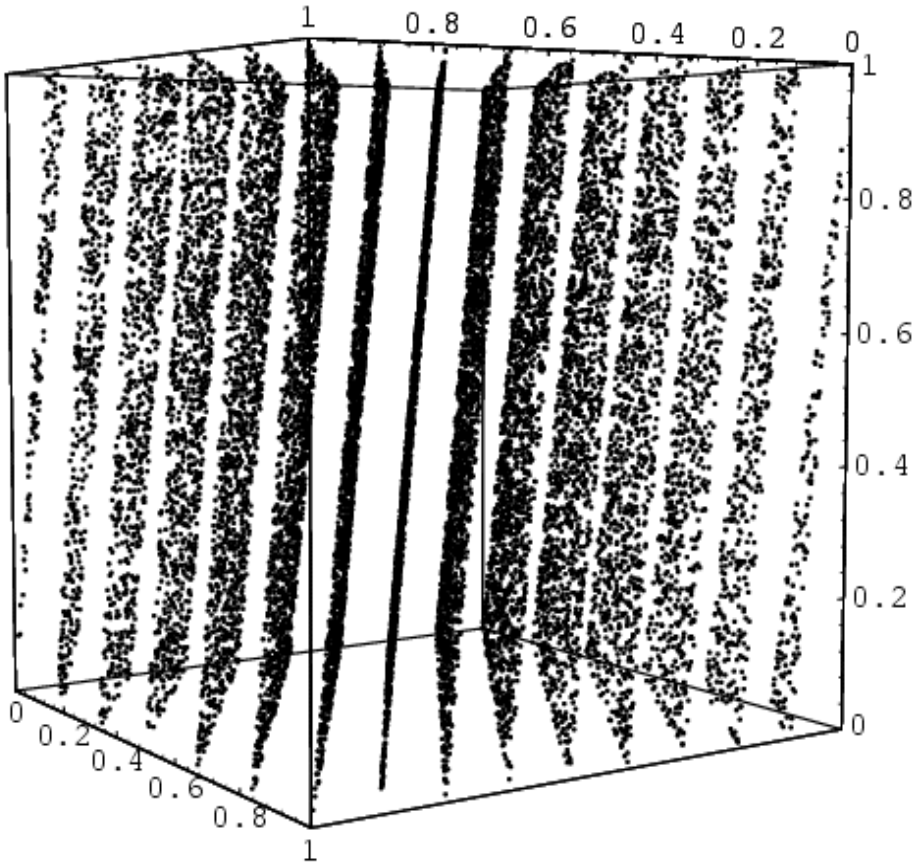
c)LCG(256,61,1,0)



d) LCG(256,237,1,0)



e) LCG: ANSI C - funkcia `rand()`



Zlé vyplnenie priestoru pomocou 15 rovín pri LCG generátore RANDU

Fibbonaciho generátory (LFG)

Rekurentný vzťah má tvar:

$$x_n = (x_{n-l} + x_{n-k}) \bmod m \; ; \quad l > k > 0$$

- Väčšinou sa volí $m=2^M$ ➔ max. perióda môže dosiahnuť $(2^l - 1) \times 2^{M-1}$
- Vhodná voľba je napr. l=55, k=24, M=31
- Namiesto operácie + sa používajú aj operácie
  - "×" lepšie vlastnosti avšak štvrtinová max. perióda
  - XOR horšie vlastnosti, jednoduchý výpočet
- Na výpočet potrebujeme posledných l hodnôt x_n.
- Z l počiatočných hodnôt $x_0, \ldots, x_{l-1}$ musí byť aspoň jedna nepárna.

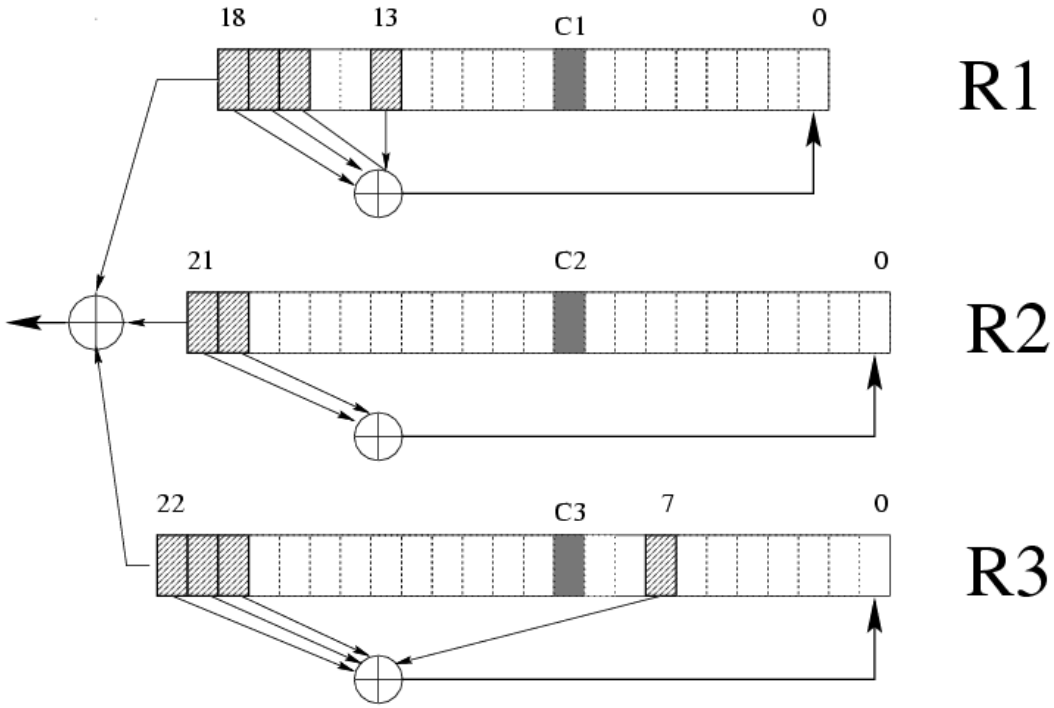
Zložené rekurzívne generátory (MRG)

Rekurentný vzťah:

$$x_n = (a_1 x_{n-1} + \cdots + a_k x_{n-k}) \bmod m$$

Max. perióda je $m^k - 1$.

Špeciálny prípad ➔ LFSR (Linear Feedback Shift Register) generátory, kde $a_i = \{0,1\}$ a $m = 2$.
Sú použité napr. v GSM algoritme A5/1.



Kombinácia troch LFSR generátorov R1, R2, R3 o veľkosti 18, 21 a 22 bitov použitá na generovanie pseudonáhodných bitov v kódovacom algoritme A5/1 pre GSM. V A5/1 je naviac v aktuálnom kroku taktovaný iba generátor, ktorého bit C súhlasí s majoritnou hodnotou bitov C.

Tausworthove generátory

Tausworthove generátory konštruujú výstupnú hodnotu z L hodnôt x_n

$$u_n = \sum_{j=1}^{L} x_{ns+j-1} 2^{-j} \quad , \quad L \leq s$$

Ak sú hodnoty x_n generované pomocou MRG s periódou ρ a $nsd(s, \rho) = 1$, potom aj postupnosť u_n má periódu ρ.

Nelineárne generátory

Výhody

- lepšie autokorelačné vlastnosti ako lineárne generátory
- lepšie spektrálne vlastnosti (výstupné hodnoty nemajú mriežkovú štruktúru)
- lepšie kryptografické vlastnosti

Nevýhody

- pomalšie

Inverzné kongruenčné generátory (ICG, EICG)

Nech m je prvočíslo a pre $x \in Z_m$ nech

$\bar{x} = 0$ ak $x = 0$ a

$\bar{x} = x^{-1} = x^{m-2} \bmod m$ (analogické k Fermatovej vete: $x^m \equiv x \bmod m$), ak $m \neq 0$. Potom platí

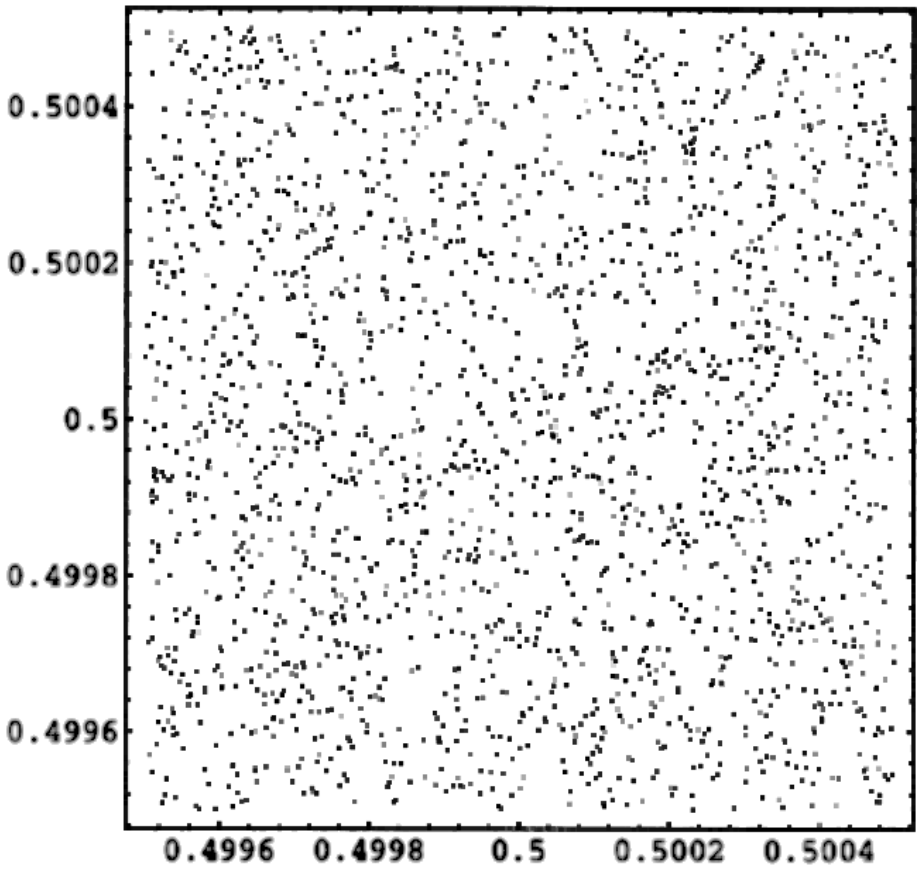
ICG: $x_n = (a\bar{x}_{n-1} + c) \bmod m$

EICG: $x_n = \overline{a(n + n_0) + c} \bmod m$

, kde ICG označuje inverzné a EICG explicitne inverzné kongruenčné generátory.

Vlastnosti

- Maximálne dosiahnuteľná perióda je m
- Inverzný prvok sa počíta inverzným Euklidovým algoritmom na nájdenie celočíselných riešení rovnice $\bar{x} \cdot x + k \cdot m = 1$
- ICG, EICG majú podstatne lepšie autokorelačné vlastnosti ako LCG
- Sú vhodné na použitie pre paralelné algoritmy.



Spektrálne charakteristiky ICG(2^{31}-1 , 1288490188 , 1 , 0)

Mocninové generátory bitov

Generátory tohoto typu sú vhodné predovšetkým na kryptografické účely.

RSA generátor

Nech $m = p \cdot q$ je súčinom dvoch veľkých prvočísiel.
Náhodne zvoľme b také, že $nsd(\phi(m), b) = 1$, kde $\phi(m) = (p-1)(q-1)$.

Potom:

$$x_n = x_{n-1}^b \bmod m$$

pričom $x_o \in \langle 1, m-1 \rangle$. Výstupom je najmenej významný bit x_n.

BBS (Blum Blum Shub) generátor

Postupnosť výstupných bitov b_n generovaná pomocou

$$x_n = x_{n-1}^2 \bmod m \qquad\qquad b_n = x_n \cdot z \bmod 2$$

kde

- x_0 je nesúdeliteľné s m
- m je tvorené súčinom dvoch veľkých prvočísiel, ktoré sa dajú vyjadriť v tvare $4k+3$
- $x_n \cdot z$ predstavuje skalárny súčin po bitoch s náhodnou bitovou maskou z

Vlastnosti:

- pre náhodne zvolené z, m, x_0 uspeje generátor vo všetkých štatistických testoch, ktoré sú menej náročnejšie ako faktorizovanie čísla m, t.j. generované bity sú do tejto miery neodlíšitelné od postupnosti skutočne náhodných bitov.

Metódy zlepšenia vlastností GPČ

- Vlastnosti už existujúcich generátorov, môžeme zlepšiť a nevhodným algoritmom resp. voľbou konštnánt aj zhoršiť

a) ***Aritmetickým skladaním*** výstupov z viacerých generátorov.

$$z_n = (x_n + y_n) \bmod m$$

kde x_n resp. y_n sú výstupy z pôvodných generátorov. Vhodné je voliť ***nesúdeliteľné*** veľkosti periód jednotlivých generátorov.

b) ***Premiešavaním (shuffling).***
Touto metódou môžeme dodatočne zlepšiť vlastnosti existujúceho generátora, ktorý generuje vstupné hodnoty pre premiešavací algoritmus. Treba si však uvedomiť, že ***premiešavaním môžeme zmeniť iba poradie vstupných hodnôt, nie hodnoty samotné***.

Existujú viaceré verzie metódy premiešavania, napríklad:

I) Prvých k vstupných hodnôt x_k uložíme do poľa. Potom z každej ďalšej vstupnej dvojice náhodných čísel pomocou prvého čísla náhodne vyberieme hodnotu z poľa, ktorá bude našim výstupom a druhé vložíme na uvoľnené miesto.

II) Prvých k hodnôt x_k uložíme do poľa a $(k+1)$-tu hodnotu do pomocnej premennej idx. Výstupné hodnoty potom generujeme v dvoch krokoch nasledovne:

1) pomocou premennej idx vyberieme hodnotu z poľa ktorá bude našim výstupom a na uvoľnené miesto vložíme novú vstupnú hodnotu
2) výstupnú hodnotu okopírujeme do premennej idx .

Výhoda***: Oproti metóde I) sme schopní generovať z jednej vstupnej hodnoty jednu hodnotu výstupnú.***

Kvázináhodné postupnosti čísel

➔ také postupnosti, ktoré rovnomerne vypĺňajú daný interval, pričom medzi po sebe nasledujúcimi hodnotami môže existovať evidentná závislosť.

Príklad:

$$x_n = (x_{n-1} + 1) \bmod 10 \qquad (\#.4.1)$$

- Prakticky sa používajú iba také algoritmy, ktoré sa snažia o rovnomerné zapĺňanie intervalu už od začiatku postupnosti, t.j. mohli by sme hocikedy prestať a interval by bol danými hodnotami vyplnený rovnomerne.
- platí, že ľubovoľná sub-sekvencia zaplní interval približne rovnako rovnomerne ako iná rovnako dlhá sub-sekvencia

Príklad:

➔ Haltonove postupnosti

Haltonove postupnosti

V jednom rozmere na intervale <0,1) sa *j*-ty člen Haltonovej postupnosti H_j generuje nasledovne:

1) zvoľme si prvočíslo *b*, ktoré bude predstavovať základ číselnej sústavy (napr. 2)
2) vyjadrime hodnotu *j* v číselnej sústave zo základom b. (napr. pre j=14, b=2 je výsledok 1100 pri základe 2)
3) hodnotu H_j dostaneme tak, že získané číslice napíšeme za desatinnú čiarku a v opačnom poradí (t.j. z príkladu dostaneme 0.0011 pri základe 2)

➔ Keď chceme generovať N-tice v N-rozmernom priestore, treba použiť v jednotlivých rozmeroch ako základy *rôzne prvočísla*. Obyčajne sa zvykne používať prvých n prvočísiel.

Príklad: Vygenerujte Haltonovu postupnosť na intervale $\langle 0,1)$ pre b=2 s periódou 8

Riešenie: Podľa krokov 2,3 postupne vypĺňame tabuľku:

j	0	1	2	3	4	5	6	7
$(j)_2$	000	001	010	011	100	101	110	111
$(H_j)_2$	0,000	0,100	0,010	0,110	0,001	0,101	0,011	0,111
H_j	0	0,5	0,25	0,75	0,125	0,625	0,375	0,875

Takže výsledná postupnosť je {0; 0,5; 0,25; 0,75; 0,125; 0,625; 0,375; 0,875 }.

Generovanie rovnomerného rozdelenia na ľubovoľnom intervale $< a,b)$:

Používame vzťah:

$$u_{ab_n} = a + (b-a).u_n, \qquad u_n \in < 0,1) \qquad \rightarrow \qquad u_{ab_n} \in < a,b)$$

Tvorba náhodných premenných s nerovnomerným rozdelením

Vstup: jedna alebo viac postupností s rovnomerným rozdelením
Výstup: náhodná premenná so želaným rozdelením

Niektoré užitočné vlastnosti, ktoré môžeme využiť:

A) Ak X_1, X_2 sú nezávislé náhodné premenné s distribučnými funkciami $F_1(x)$, $F_2(x)$:

$\max(X_1, X_2)$ má rozdelenie $F_1(x)F_2(x)$
$\min(X_1, X_2)$ má rozdelenie $F_1(x) + F_2(x) - F_1(x).F_2(x)$

B) Nech x je náhodná premenná s funkciou hustoty pravdepodobnosti $f(x)$ Pomocou monotónnej funkcie r vytvorme náhodnú premennú $Y = r(X)$ a označme $g(x)$ jej funkciou hustoty pravdepodobnosti. Potom:

$$g(x) = f\left[r^{-1}(x)\right] \cdot \left|\left(r^{-1}\right)'(x)\right|$$

Generovanie náhodných premenných so spojitým rozdelením

Najznámejšie všeobecné metódy:

- ***metóda inverznej funkcie***
- ***vylučovacia metóda***

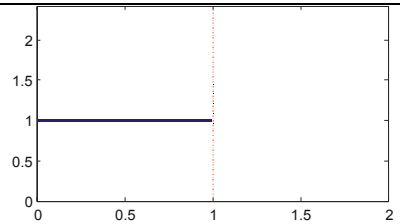
Okrem nich existujú ***špeciálne metódy*** ➔ navrhnuté na jedno cieľové rozdelenie náhodnej premennej.

Príklad: ***Polárna metóda*** na generovanie náhodných premenných s normálovým rozdelením

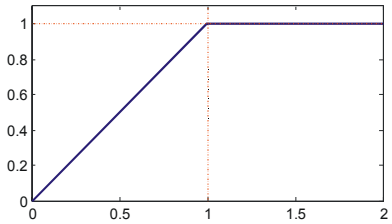
Príklady spojitých a diskrétnych rozdelení

Rovnomerné rozdelenie

f(x)



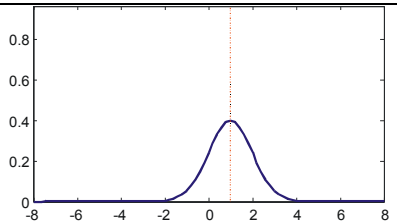
F(x)



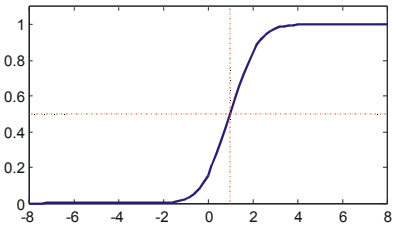
$x \in (0,1>$

Normálne rozdelenie

f(x)



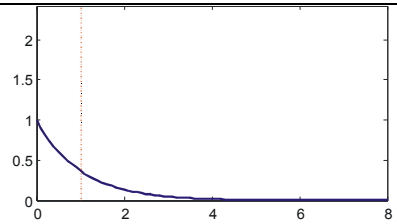
F(x)



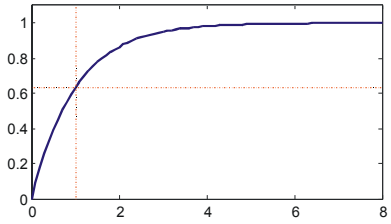
$\mu = 1,\ \sigma = 1$

Exponenciálne rozdelenie

f(x)



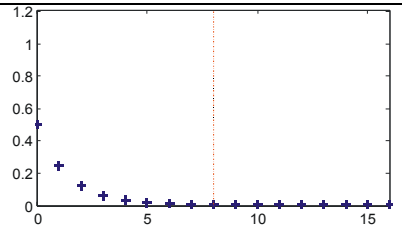
F(x)



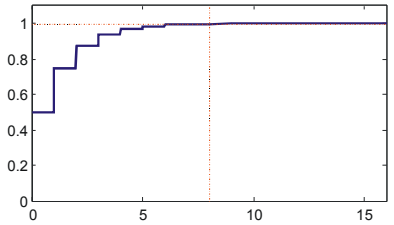
$\mu = 1$

Geometrické rozdelenie

f(x)



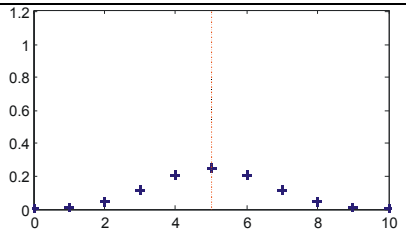
F(x)



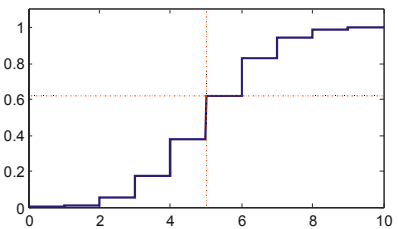
prevdepodobnosť 0.5

Binomické rozdelenie

f(x)



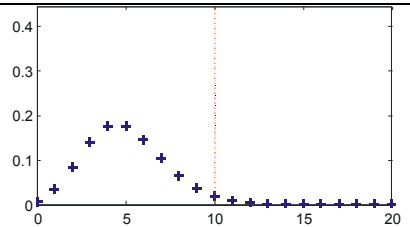
F(x)



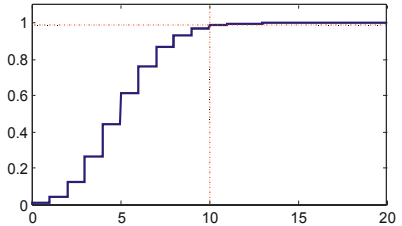
10 pokusov, prevdepodobnosť 0.5

Poissonove rozdelenie

f(x)



F(x)



$\lambda = 5$

Metóda inverznej funkcie

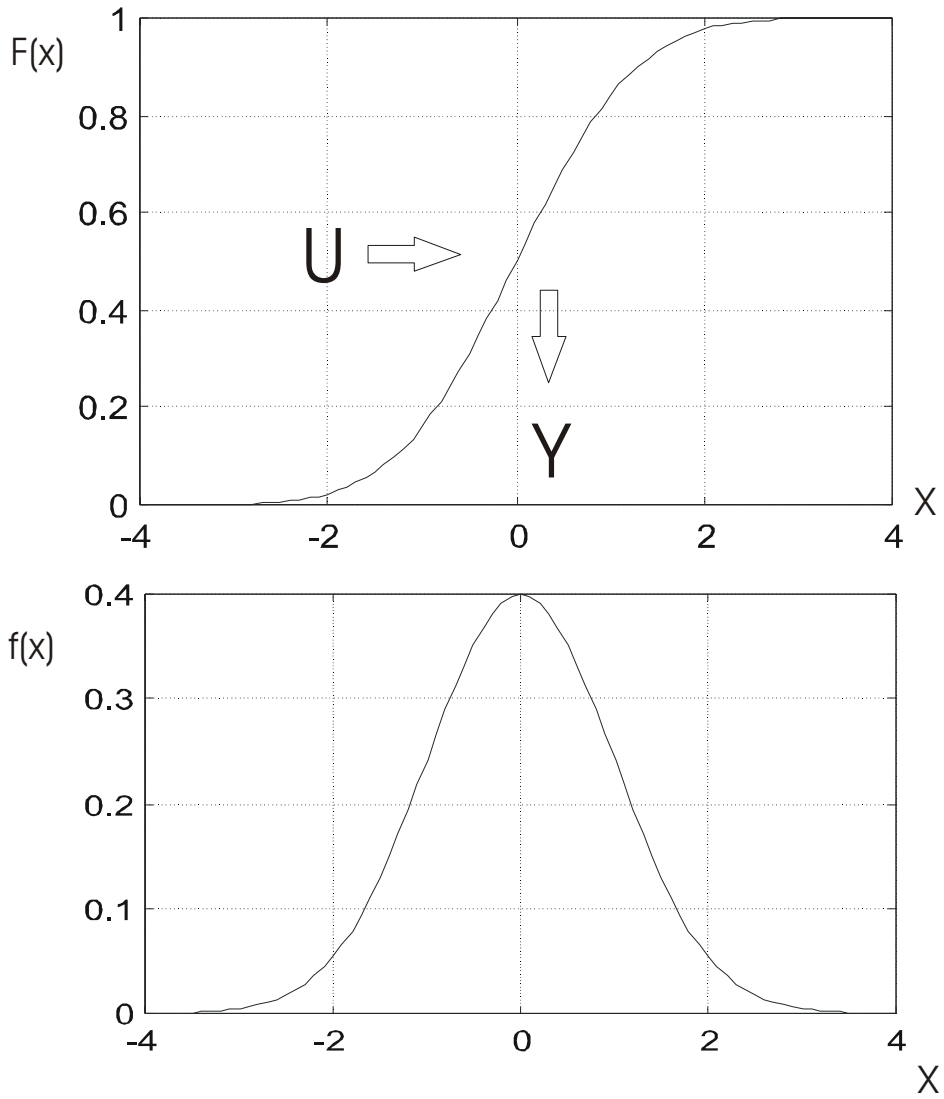
Ak spojitá náhodná premenná Y má distribučnú funkciu F, potom hodnoty y náhodnej premennej Y môžeme vytvoriť pomocou náhodnej premennej U s rovnomerným rozdelením na intervale <0,1) a hodnotami u použitím vzťahu:

$$Y = F^{-1}(U)$$

<u>Dôkaz:</u> *Distribučná funkcia náhodnej premennej Y je daná pravdepodobnosťou* $P(Y < y)$ *:*

$$P(Y < y) = P(F^{-1}(U) < y) = P(U < F(y)) = \int_0^{F(y)} dx = F(y)$$

teda náhodná premenná Y má distribučnú funkciu F.



Metóda inverznej funkcie: mapovanie náhodnej premennej U s rovnomerným rozdelením na náhodnú premennú Y s normálovým rozdelením $N(0,1)$.

Príklad: Vytvorte spojitú náhodnú premennú X s exponenciálnym rozdelením pomocou metódy inverznej funkcie.

Riešenie: Funkcia hustoty pravdepodobnosti náhodnej premennej s exponenciálnym rozdelením je daná ako $f(x) = \lambda\, e^{-\lambda x}$ ($x > 0$). Tomu odpovedá distribučná funkcia $F(x) = 1 - e^{-\lambda x}$. Inverziou dostávame $F^{-1}(x) = -\ln(1-x)/\lambda$. Teda ak máme náhodnú premennú U s rovnomerným rozdelením, potom náhodná premenná $X = -\ln(1-U)/\lambda$ má exponenciálne rozdelenie s parametrom λ. Keď uvážime, že 1-U má taktiež rovnomené rozdelenie, môžeme výsledok zjednodušiť na $X = -\ln(U)/\lambda$.

Vylučovacia metóda I.

Nech spojitá náhodná premenná X *s hodnotami* $x \in (a,b)$ má funkciu hustoty pravdepodobnosti $f(x) \leq d$ a distribučnú funkciu $F(x)$. Nech Y a Z sú náhodné premenné nasledovne vytvorené z dvoch nezávislých náhodných premenných s rovnomerným rozdelením U_y a U_z:

$$Y = a + (b-a)U_y$$

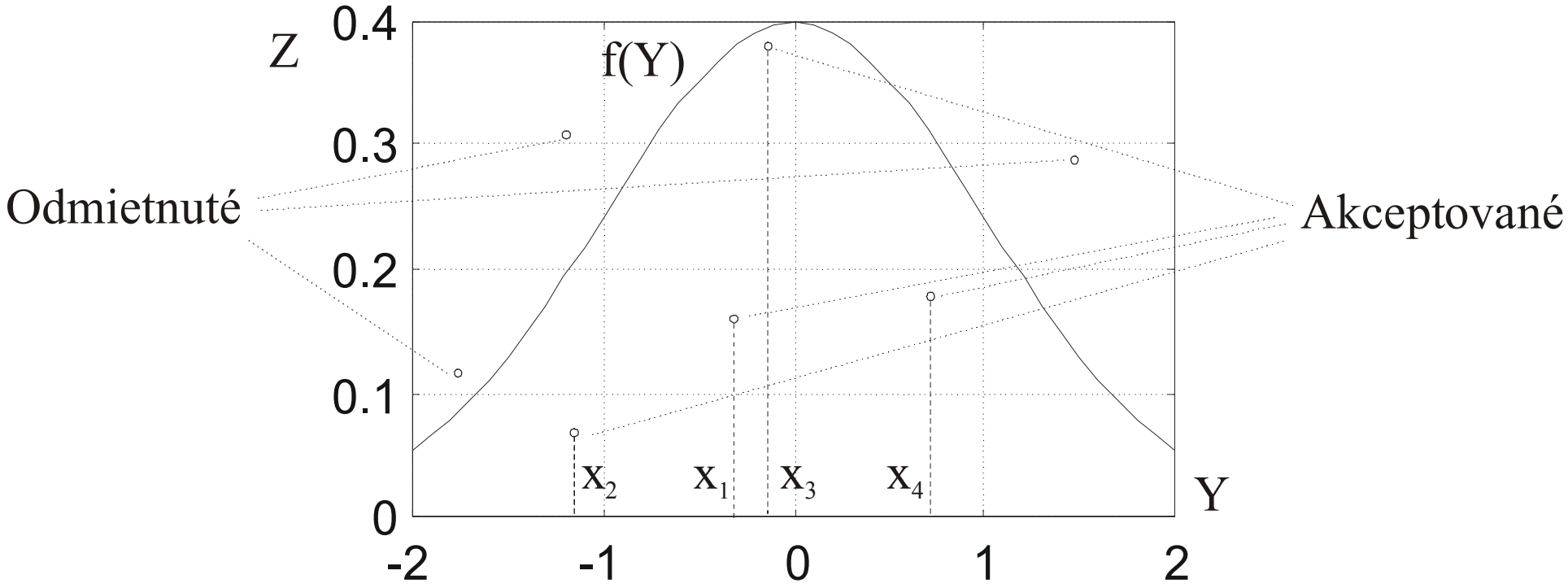
$$Z = dU_z$$

Potom hodnoty náhodnej premennej X sú výberom takých hodnôt Y, pre ktoré platí:

$$Z \leq f(Y)$$

Priemerný počet opakovaní na výber jednej hodnoty je:

$$d(b-a)/\big(F(b)-F(a)\big)$$



Príklad: Vylučovacia metóda I pre generovanie náhodnej premennej s normálovým rozdelením $N(0,1)$ na intervale $x \in (-2,2)$. Vygenerované sú prvé 4 hodnoty X.

Vylučovacia metóda II

- Je zovšeobecnením metódy I
- Môžeme ju použiť, keď chceme transformovať náhodnú premennú Y so známym rozdelením na náhodnú premennú X so želaným rozdelením.

Nech spojité náhodné premenné X a Y majú funkcie hustoty pravdepodobnosti $f(x)$ a $g(y)$ a nech ***c*** je konštanta pre ktorú platí:

$$f(y) \leq c \cdot g(y) \qquad \text{pre všetky} \quad y$$

Potom hodnoty náhodnej premennej X môžeme vypočítať pomocou hodnôt Y a náhodnej premennej U s rovnomerným rozdelením v nasledovných krokoch:

1) Generuj po jednej hodnote z Y s daným $g(y)$ a jednej hodnote z U
2) Ak $f(Y) \geq Ucg(Y)$, vyber aktuálnu hodnotu Y, t.j. $X = Y$
3) Opakuj od kroku 1

Pri výpočte treba nájsť čo najmenšie c.

Príklad: Vytvorte náhodnú premennú s normálovým X rozdelením N(0,1) pomocou exponenciálneho rozdelenia použitím vylučovacej metódy II.

Riešenie: $N(0,1)$ má absolutnu hodnotu funkciu hustoty prevdepodobnosti:

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2}, \qquad 0 < x < \infty$$

Zvoľme:

$$g(x) = e^{-x}, \qquad 0 < x < \infty$$

Hľadaním maxima podielu týchto dvoch funkcií dostaneme

$$c = \sqrt{2e/\pi} \approx 1.32$$

Hodnoty náhodnej premennej X s $N(0,1)$ vytvoríme v nasledovných krokoch:

1) generujme hodnotu z Y s exponenciálnym rozdelením s $\lambda = 1$ a hodnotu U s rovnomerným rozdelením na (0,1).
2) Ak $-\log U \geq (Y-1)^2/2$, vyber $|X|$, t.j. $|X| = Y$, ináč opakuj od kroku 1)
3) S rovnakou pravdepodobnosťou vyber $-X$ alebo X. Opakuj od kroku 1)

Poznamka: ak chceme generovať náhodnú premennú Z s rozdelením $N(\mu, \sigma^2)$, potom je treba výsledok transformovať pomocou $Z = \mu + \sigma X$

Polárna metóda

Je príkladom špeciálnej metódy, pomocou ktorej generujeme dve nezávislé náhodné premenné X_1, X_2 s normálovým rozdelením.

Postup:

1) Vygeneruj po jednej hodnote z dvoch nezávislých náhodných premenných U_1, U_2 s rovnomerným rozdelením
2) Vytvor premenné $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$, ktoré majú rovnomerne rozdelenie na intervale < -1, 1)
3) Ak pre $S = V_1^{\,2} + V_2^{\,2}$ platí $S \geq 1$, opakuj postup znovu od bodu (1)
4) Vygeneruj výstupné hodnoty náhodných premenných s normálovým rozdelením

$$X_1 = V_1 \sqrt{(-2 \ln S)/S} \qquad X_2 = V_2 \sqrt{(-2 \ln S)/S}$$

a opakuj postup od bodu (1)

Generovanie vybraných diskrétnych rozdelení:

➔ generovanie diskrétnych náhodných premenných použítím spojitých náhodných premenných U s rovnomerným rozdelením:

Geometrické rozdelenie - má náhodná premenná G, ktorá predstavuje počet nezávislých pokusov medzi výskytmi nejakej udalosti, ktorá sa vyskytuje z pravdepodobnosťou p.

Hodnoty G sú rovné n s pravdepodobnosťou

$$(1-p)^{n-1}\,p\,.$$

Postup:

Hodnoty náhodnej premennej G môžeme generovať pomocou:

$$G = \left\lceil \ln U / \ln(1-p) \right\rceil$$

, kde operátor $\lceil\ \rceil$ znamená najvyššie bližšie celé číslo.

Binomické rozdelenie (t,p) – má N, počet výskytov udalosti, ktorá môže nastať s pravdepodobnosťou p v každom z t nezávislých pokusov, ktoré sú k dispozícii.

Platí

$$P(N=n)=\binom{t}{n}p^n(1-p)^{t-n}.$$

Postup:

Generujeme po jednej hodnote z t generátorov $U_1,U_2,\ldots,U_t$ a počítame koľko z nich má hodnotu menšiu ako p.

Vlastnosti:

Algoritmus je efektívny pre malé t.

Poissonove rozdelenie so strednou hodnotou μ .

Medzi *exponenciálnym* a *Poissonovým* rozdelením je podobný vzťah ako medzi *geometrickým* a *binomickým* rozdelením:

- Reprezentuje počet výskytov udalosti za jednotkový čas.
- Udalosť sa môže vyskytnúť hocikedy.

Postup:

Sčítavame po jednej hodnote náhodných premenných $X_1, X_2, \cdots$ s exponenciálnym rozdelením s so strednou hodnotou $1/\mu$ a zisťujeme koľko ich treba na splnenie podmienky $X_1 + X_2 + \ldots X_m \geq 1$. Počet N=m-1má Poissonove rozdelenie.

Poznámka: Ekvivalentná podmienka je $U_1 \cdot U_2 \cdot \ldots \cdot U_m \leq e^{-\mu}$

Vlastnosti:

Algoritmus je efektívny pre malé μ .

Testy GPČ

Cieľom testov GPČ ➔ zistiť, či sa generované postupnosti správajú dostatočne náhodne.

- Existuje nekonečne veľa vlastností, ktoré môžeme testovať
- v praxi sa používajú tie, ktoré sa ukázali byť najužitočnejšie.
- aj keď generátor uspeje v *N* testoch nemôžeme si byť istý, že v *N+1* teste úplne nezlyhá. (t.j. pre každý generátor existuje test, pri ktorom úplne zlyhá)
- S rastúcim počtom úspešne zdolaných testov však môžeme byť čoraz spokojnejší s náhodnosťou generovanej sekvencie a predpokladať že náhodná aj je.

Základné druhy testov

- ***Teoretické***
- ***Empirické***

Teoretické aj empirické testy na testovanie náhodnosti používajú testovacie procedúry na testovanie hypotéz a to najmä

- χ^2 test
- KS (Kolmogorov - Smirnov) test

Ekvivalentný test pre nerovnomerné rozdelenia

Ak chceme zistiť či náhodná veličina Y má predpokladané rozdelenie so známou distribučnou funkciou $F(y) = P(Y \leq y)$, môžeme uskutočniť ekvivalentný test, kde testujeme či náhodná veličina

$$Z = F(Y)$$

má rovnomerné rozdelenie.

χ^2 (Chi-kvadrát) test.

Pri tomto teste testujeme hypotézu H_0:

Postupnosť pseudonáhodných čísel má dané rozdelenie.

Postup:

1. Z testovanej postupnosti použijeme ľubovoľných n za sebou nasledujúcich hodnôt $x_1, x_2, ..., x_n$, ktoré rozdelíme do k disjunktných tried.
2. Počty v jednotlivých triedach označme z_j, $j=1,2,...k$. V praxi volíme n také aby v každej triede bolo $z_j > 5$.
3. Ak označíme p_j = *P(hodnota je z j-tej triedy)*, očakávame, že do tejto triedy padne $n\,p_j$ hodnôt.
4. Testovacie kritérium je založené na ***rozdiele očakávaných a skutočných počtov v triedach:***

$$V = \sum_{j=1}^{k} \frac{(z_j - np_j)^2}{np_j}$$

- Pre veľké n má V ***vždy*** (t.j. nezávisle od očakávaného rozdelenia vstupných dát) približne rozdelenie ***chi-kvadrát s k-1 stupňami voľnosti*** χ^2_{k-1}.

- Hodnotu V testujeme na zvolenej ***hladine testu*** α =*P(správnu* H_0 *zamietneme, t.j. nastane chyba I. druhu)* pomocou kritických hodnôt $\chi^2_{k-1}(\alpha)$. To sú hodnoty, ktoré náhodná veličina Y s rozdelením χ^2_{k-1} prekročí s pravdepodobnosťou α, t.j. $P(Y \geq \chi^2_{k-1}(\alpha)) = \alpha$.
- V praxi volíme α =*0.01-0.05*. Hodnoty $\chi^2_{k-1}(\alpha)$ sú pre dané *k-1* a α a uvedené v tabuľkách
- Proti H_0 svedčia
  - nielen príliš veľké hodnoty V (veľký rozdiel oproti očakávanému rozdeleniu)
  - ale aj príliš malé hodnoty (nedostatočnej náhodnosť testovanej postupnosti)

  Hypotézu H_0 potom ***zamietneme na hladine*** α ak platí

$$V \leq \chi^2_{k-1}(\alpha/2) \qquad V \geq \chi^2_{k-1}(1-\alpha/2)$$

Pri overovaní rovnomerného rozdelenia postupnosti pseudonáhodných čísel sa používa *k=10-100* postupne pre rôzne úseky generovanej postupnosti.

Kritické hodnoty náhodnej premennej V s rozdelením chi-kvadrát s k-1 stupňami voľnosti.

	$p = 1\%$	$p = 5\%$	$p = 25\%$	$p = 50\%$	$p = 75\%$	$p = 95\%$	$p = 99\%$
$\nu = 1$	0.00016	0.00393	0.1015	0.4549	1.323	3.841	6.635
$\nu = 2$	0.02010	0.1026	0.5754	1.386	2.773	5.991	9.210
$\nu = 3$	0.1148	0.3518	1.213	2.366	4.108	7.815	11.34
$\nu = 4$	0.2971	0.7107	1.923	3.357	5.385	9.488	13.28
$\nu = 5$	0.5543	1.1455	2.675	4.351	6.626	11.07	15.09
$\nu = 6$	0.8721	1.635	3.455	5.348	7.841	12.59	16.81
$\nu = 7$	1.239	2.167	4.255	6.346	9.037	14.07	18.48
$\nu = 8$	1.646	2.733	5.071	7.344	10.22	15.51	20.09
$\nu = 9$	2.088	3.325	5.899	8.343	11.39	16.92	21.67
$\nu = 10$	2.558	3.940	6.737	9.342	12.55	18.31	23.21
$\nu = 11$	3.053	4.575	7.584	10.34	13.70	19.68	24.72
$\nu = 12$	3.571	5.226	8.438	11.34	14.85	21.03	26.22
$\nu = 15$	5.229	7.261	11.04	14.34	18.25	25.00	30.58
$\nu = 20$	8.260	10.85	15.45	19.34	23.83	31.41	37.57
$\nu = 30$	14.95	18.49	24.48	29.34	34.80	43.77	50.89
$\nu = 50$	29.71	34.76	42.94	49.33	56.33	67.50	76.15
$\nu > 30$	$\nu + \sqrt{2\nu}x_p + \frac{2}{3}x_p^2 - \frac{2}{3} + O(1/\sqrt{\nu})$						
$x_p =$	−2.33	−1.64	−.674	0.00	0.674	1.64	2.33

Príklad: GPČ sme použili na simuláciu hádzania dvomi hracími kockami. Po 144 hodoch sme zistili nasledovné padnuté počty bodov:

Získaný počet bodov		2	3	4	5	6	7	8	9	10	11	12	Suma
Početnosť hodov (z_j)	GPČ A	4	10	10	13	20	18	18	11	13	14	13	144
	GPČ B	3	7	11	15	19	24	21	17	13	9	5	144

Podľa daných výsledkov otestujte oba generátory na hladine α =0,05.

Riešenie: V našom prípade je počet tried *k=11*. Pomocou počtu pravdepodobnosti si najprv si zistíme očakávané (resp. ideálne) hodnoty početností (pri počte hodov 144):

Trieda (*j*)	1	2	3	4	5	6	7	8	9	10	11
Získaný počet bodov	2	3	4	5	6	7	8	9	10	11	12
Pravdepodobnosť výskytu (p_j)	1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36
Odpovedajúca početnosť (np_j)	4	8	12	16	19	24	21	17	13	9	5

Z tabuliek dostaneme $\chi^2_{10}(0{,}025) = 3{,}25$ a $\chi^2_{10}(0{,}975) = 20{,}5$. Pre generátor A dostaneme $V_A = 29.242$, pre generátor B $V_B = 1{,}142$. Vidíme, že generátor A v teste neuspel, lebo $V_A \geq \chi^2_{10}(0{,}975)$ a takisto musíme pre generátor B hypotézu, že generovaná postupnosť má očakávané rozdelenie, zamietnuť na hladine α =0,05 lebo $V_B \leq \chi^2_{10}(0{,}025)$.

Kolmogorov-Smirnov (KS) test.

- Pri tomto type testu nedelíme hodnoty testovanej postupnosti do tried
- test je vhodný na testovanie spojitých rozdelení
- Testujeme zhodu empirickej distribučnej funkcie $F_n(x)$ s očakávanou distribučnou funkciou $F(x)$.

Postup:

- Z n hodnôt $x_1, x_2, ..., x_n$ testovanej postupnosti vytvoríme distribučnú funkciu:

  $F_n(x) =$ (počet hodnôt $x_1, x_2, ..., x_n$, pre ktoré $< x$) / n
- Na vykonanie testu potrebujeme 2 hodnoty K_n^+ a K_n^-:

  $K_n^+ = \sqrt{n}\max\left(F_n(x) - F(x)\right), \qquad x \in (-\infty, \infty)$

  $K_n^- = \sqrt{n}\max\left(F(x) - F_n(x)\right), \qquad x \in (-\infty, \infty)$
- Na danej hladine α testu potom tieto hodnoty porovnávame s ***kritickými hodnotami*** z tabuliek, podobne ako tomu bolo pri χ^2 teste.
- Avšak kritické hodnoty sú vypočítané pre konkrétne n a α a nie sú použité ako aproximácie (na rozdiel od $\chi^2_{k-1}(\alpha)$ v χ^2 teste).

Kritické hodnoty pri KS teste

SELECTED PERCENTAGE POINTS OF THE DISTRIBUTIONS K_n^+ AND K_n^-

	$p = 1\%$	$p = 5\%$	$p = 25\%$	$p = 50\%$	$p = 75\%$	$p = 95\%$	$p = 99\%$
$n = 1$	0.01000	0.05000	0.2500	0.5000	0.7500	0.9500	0.9900
$n = 2$	0.01400	0.06749	0.2929	0.5176	0.7071	1.0980	1.2728
$n = 3$	0.01699	0.07919	0.3112	0.5147	0.7539	1.1017	1.3589
$n = 4$	0.01943	0.08789	0.3202	0.5110	0.7642	1.1304	1.3777
$n = 5$	0.02152	0.09471	0.3249	0.5245	0.7674	1.1392	1.4024
$n = 6$	0.02336	0.1002	0.3272	0.5319	0.7703	1.1463	1.4144
$n = 7$	0.02501	0.1048	0.3280	0.5364	0.7755	1.1537	1.4246
$n = 8$	0.02650	0.1086	0.3280	0.5392	0.7797	1.1586	1.4327
$n = 9$	0.02786	0.1119	0.3274	0.5411	0.7825	1.1624	1.4388
$n = 10$	0.02912	0.1147	0.3297	0.5426	0.7845	1.1658	1.4440
$n = 11$	0.03028	0.1172	0.3330	0.5439	0.7863	1.1688	1.4484
$n = 12$	0.03137	0.1193	0.3357	0.5453	0.7880	1.1714	1.4521
$n = 15$	0.03424	0.1244	0.3412	0.5500	0.7926	1.1773	1.4606
$n = 20$	0.03807	0.1298	0.3461	0.5547	0.7975	1.1839	1.4698
$n = 30$	0.04354	0.1351	0.3509	0.5605	0.8036	1.1916	1.4801
$n > 30$	$y_p - \frac{1}{6}n^{-1/2} + O(1/n)$, where $y_p^2 = \frac{1}{2}\ln(1/(1-p))$						
$y_p =$	0.07089	0.1601	0.3793	0.5887	0.8326	1.2239	1.5174

Opakovaný Kolmogorov-Smirnov (KS) test.

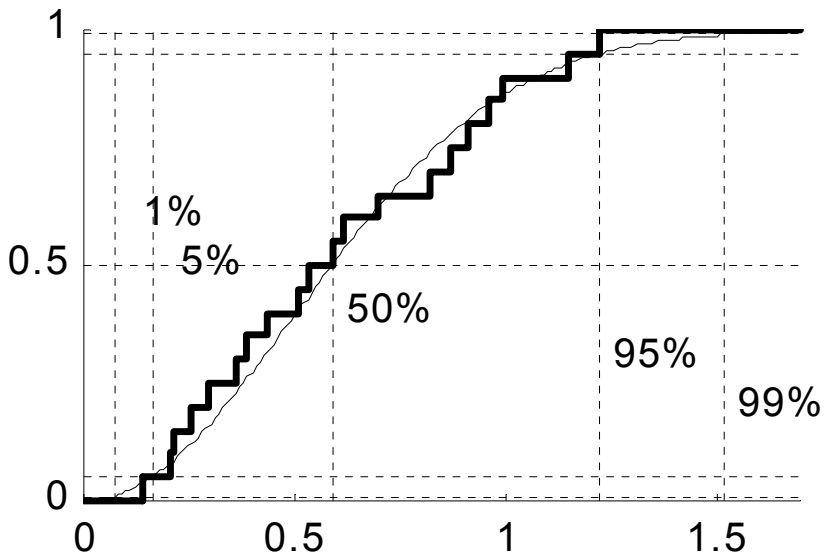
KS test môžeme ľahko znovu aplikovať na výsledky KS testu, keďže distribučná funkcia *F(x)* rozdelenia hodnôt K_n^+ alebo K_n^- sa dá pre veľké *n* aproximovať pomocou:

$$F_\infty(x) = 1 - e^{-2x^2}, \qquad x \geq 0$$

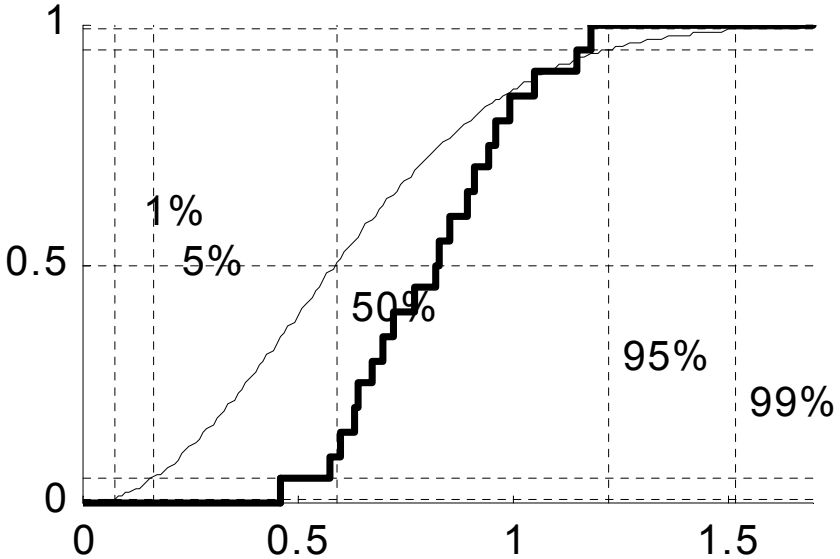
Takýmto spôsobom môžeme

detekovať naraz lokálne aj globálne nenáhodné správanie sa.

Napr. rozdelíme testované hodnoty na 20 skupín po 1000, zistíme hodnoty K^{+}_{1000} a K^{-}_{1000} pre každú skupinu a z nich následne K^{+}_{20} a K^{-}_{20} (globálne správanie). T.j. napr. testujeme distribučnú funkciu $F_n(x)$ získanú z 20 hodnôt K^{+}_{1000} oproti očakávanej $F^{+}_{\infty}(x)$



a)



b)

a) $F_n(x)$ je v norme

b) sú evidentné nedostatky (nevyhovuje K^{-}_{20}), hoci každý dielčí výsledok K^{+}_{1000} je v intervale 5-95%, t.j. vyhovel na hladine $\alpha = 0.1$.

Empirické testy

- testy na overovanie náhodnosti ***vygenerovanej*** postupnosti pseudonáhodných čísel

- testy aplikujeme na postupnosť $u_n = u_0, u_1, u_2, \ldots$ reálnych čísel, o ktorých predpokladnáme, že sú rovnomerne rozdelené na intervale <0,1)

- V prípade, že testy boli navrhnuté ako celočíselné, použijeme pomocnú testovaciu postupnosť :

$$y_n = \lfloor d u_n \rfloor$$

,v ktorej sú hodnoty rovnomerne rozdelené na $Z_d = \{0,1,\ldots, d-1\}$

- Známe batérie testov obsahujúce rozne druhy epmirických testov sú napr. DIEHARD a NIST.

Test rovnomernosti rozdelenia (test frekvencie)

Základný predpoklad pri všetkých GPČ je, že výstupné hodnoty majú ***rovnomerné rozdelenie***. Pri testovaní tohoto predpokladu sú v zásade dve možnosti:

a) použijeme KS test, kde $F_n(x) = x$ pre $0 \le x \le 1$, prípadne opakovaný KS test
b) použijeme χ^2 test , pričom hodnoty transformujeme pomocou $y_n = \lfloor du_n \rfloor$. Počet tried sa potom rovná d.

Sériový test

V sériovom teste overujeme ***správanie sa N-tíc***. Z n hodnôt testovanej postupnosti vyberáme postupne neprekrývajúce sa N-tice, ktoré považujeme za súradnice bodov v oblasti v N-rozmernom priestore. Transformujeme ich použitím $y_n = \lfloor du_n \rfloor$ a aplikujeme χ^2 test. Počet tried je potom d^N, t.j. treba voliť minimálne $n > 5d^N$. V praxi sa najčastejšie takto testujú dvojice resp. trojice čísel.

Test maxima z t hodnôt

Označme $v_j = \max(u_{tj}, u_{tj+1}, \ldots, u_{tj+(t-1)})$. Potom môžeme postupovať nasledovne:

a) postupnosť v_j má mať rozdelenie s distribučnou funkciou $F(x) = x^t$, čo overíme KS testom

b) použijeme test rovnomernosti rozdelenia na postupnosť hodnôt v_j^t.

Test minimálnej vzdialenosti

Na testovanú postupnosť sa pozeráme ako na súradnice bodov v rovine, pričom testujeme minimálnu vzdialenosť medi nimi. *N* krát opakuj: a) zvoľ *n* náhodných bodov na jednotkovom štvorci b)Nájdi minimálnu vzdialenosť *d* medzi 2 bodmi zo všetkých $(n^2 - n)/2$ párov. Ak body boli náhodne rozdelené, potom druhé mocniny získaných *N* hodnôt *d* majú exponenciálne rozdelenie so strednou hodnotou μ, t.j. $1 - e^{-d^2/\mu}$ má mať rovnomerné rozdelenie na <0,1), čo sa overí KS testom.

Test bitového prúdu

Na testovanú postupnosť sa pozeráme ako na prúd bitov *b1, b2, ... bN*, kde *N*=počet hodnôt postupnosti krát bitová náročnosť jednej hodnoty. Z tohoto prúdu bitov vyberáme prekrývajúce sa slová o veľkosti 20 bitov (t.j. *1.slovo=b1,b2,...,b20, 2.slovo=b2,b3,...,b21, ...*) a počítame počet chýbajúcich slov. Z celkového počtu 2^{20} možných slov pri výbere 2^{21} slov zo vstupného prúdu bitov očakávame, že počty *J* chýbajúcich slov bude mať približne normálové rozdelenie s $\mu = 141909$ a $\sigma = 428$. Overiť to KS testom môžeme a) hneď b) *J* normujeme transformáciou (*J*-141909)/428) a následne prevedieme na rovnomerné rozdelenie na <0,1) a testujeme až potom.

Teoretické testy

- poskytujú možnosť odhaliť nedostatky generátorov už pred ich empirickým testovaním.
- umožňujú predpovedať a porozumieť ich správaniu sa za špecifických okolností.
  - Pracuje sa s celou periódou generátora ➔ niektoré testy (napr. test na rovnomernosť rozdelenia) nemajú veľký zmysel

Najčastejšie používajú

- spektrálny test - použiteľný iba na generátory s mriežkovou štruktúrou výstupných hodnôt.
- test diskrepancie - všeobecnejší avšak vo viacerých rozmeroch príliš náročný na výpočet

Spektrálny test

- veľmi dôležitý test na kontrolu LCG
- Zatiaľ všetky LCG generátory o ktorých sa zistilo, že sú nevyhovujúce, neprešli ani týmto testom.
- Spektrálny test je úzko spätý so *sériovým testom* avšak výsledky sú spoľahlivejšie, lebo sú rotačne invariantné vzhľadom na orientáciu mriežkovej štruktúry vstupných údajov.

Označme u_n výstupy z LCG a vytvorme *N*-tice:

$$S_n^N = (u_n, u_{n+1}, \ldots, u_{n+N-1})$$

Použijeme ich ako súradnice bodov v *N*-rozmernom priestore.

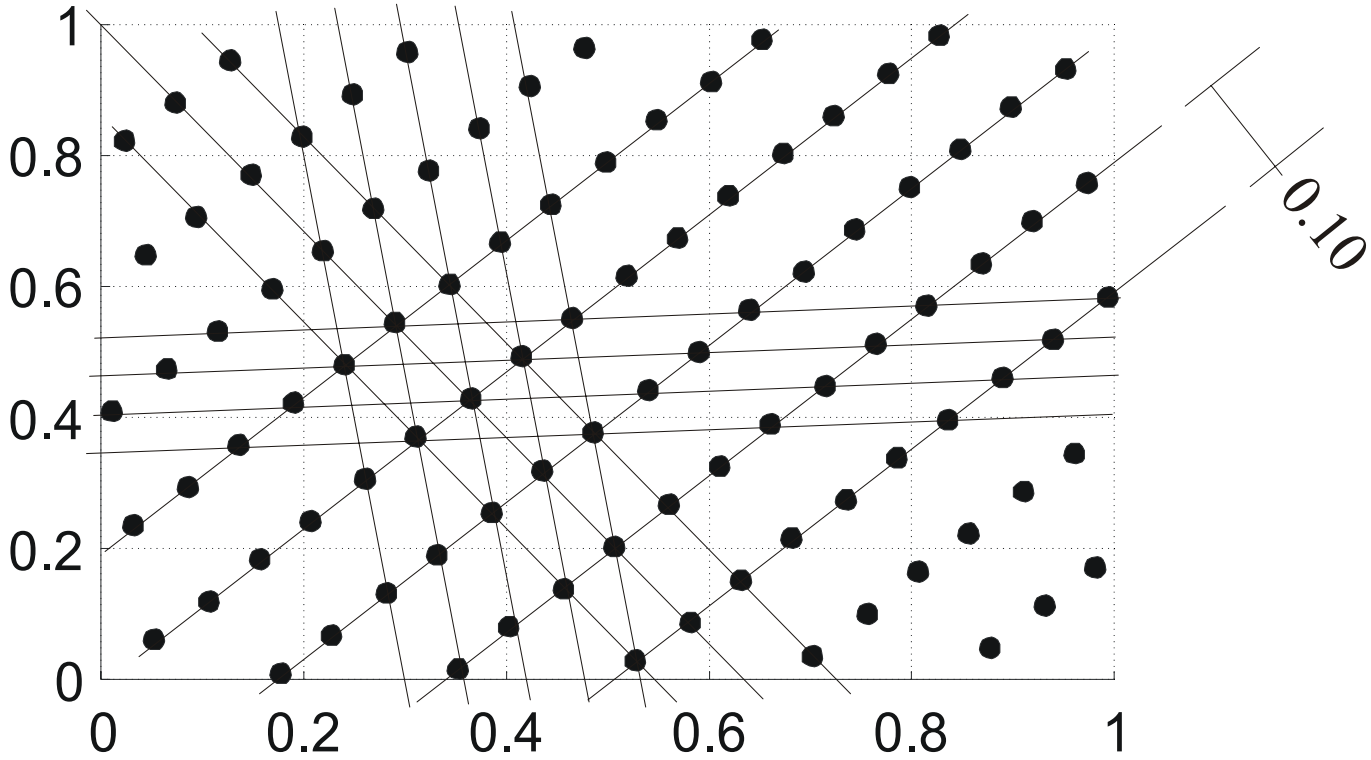
Je evidentná *mriežková štruktúra* pri vypĺňaní *N*-rozmerného intervalu
➔ Body S_n^N ležia na množine *ekvidištančných paralelných hyperrovín*.

Spektrálny test nám poskytuje numerickú hodnotu ν_N vyplnenia N-rozmerného intervalu definovaním $1/\nu_N$ ako *maximálnej vzdialenosti medzi hyperrovinami* zo všetkých množín paralelných (N-1)-rozmerných hyperrovín prekrývajúcich všetky body $\{S_n^N\}$.

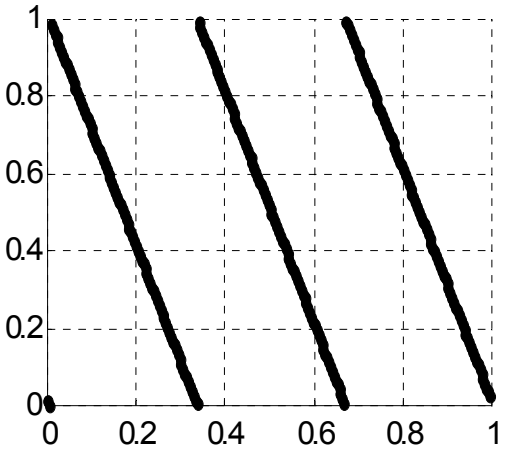
Hodnoty ν_N môžeme normovať na $\nu_N^* \in <0,1>$ pomocou:

$$\nu_N^* = \nu_N / (\gamma_N^{1/2} m^{1/N})$$

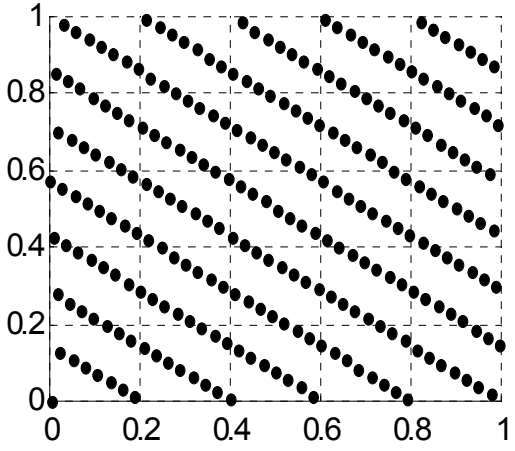
γ_N sú Hermitove konštanty ($\gamma_2 = (4/3)^{1/2}, \gamma_3 = 2^{1/2}, \ldots$).



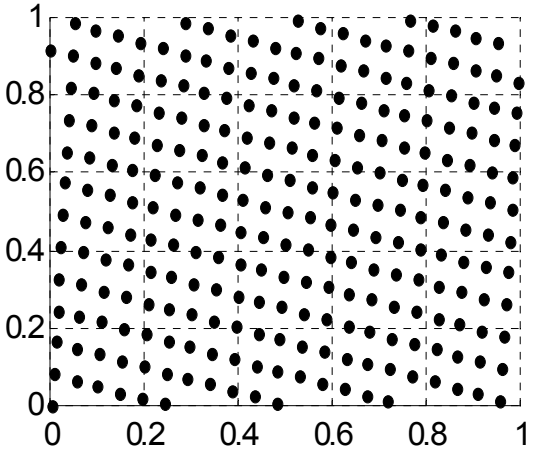
Príklad spektrálnych charakteristík GPČ (generátor LCG(97,17,0,1)), vynesené u_n voči u_{n-1} . Naznačené sú viaceré smery paralelných priamok a vzdialenosť medzi priamkami pre množinu s maximálnou vzdialenosťou.



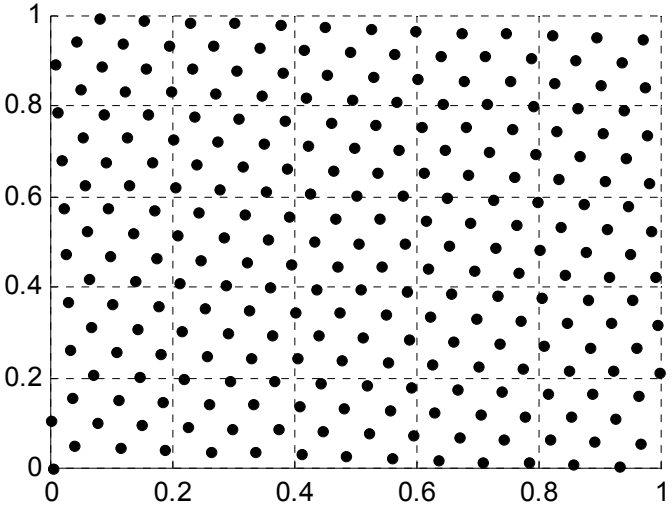
a)LCG(256,85,1,0)



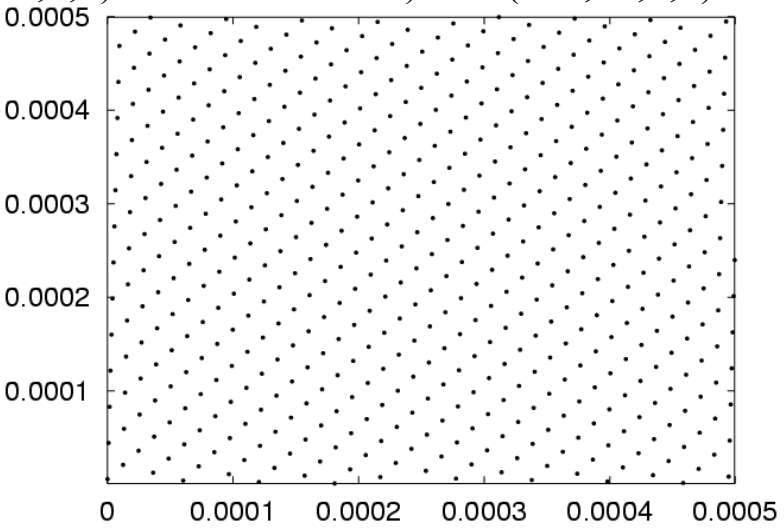
b)LCG(256,101,1,0)



c)LCG(256,61,1,0)



d) LCG(256,237,1,0)



e) LCG: ANSI C - funkcia `rand()`

Normovaná hodnota spektrálneho testu pre generátory a)–d) sa zlepšuje: a) 0,1839 b) 0,5003 c) 0,7357 d) 0,9196