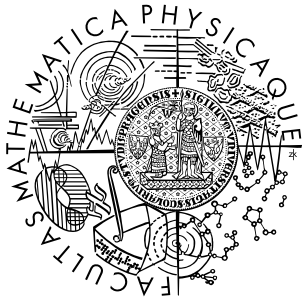
Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

BAKALÁŘSKÁ PRÁCE



Ivo Machek

Kritéria pseudonáhodnosti posloupností

Katedra algebry

Vedoucí bakalářské práce: Mgr. Štěpán Holub, Ph.D.
Studijní program: obecná matematika

2007

Děkuji Mgr. Štěpánu Holubovi, Ph.D za vedení bakalářské práce a jeho cenné rady.

Prohlašuji, že jsem svou bakalářskou práci napsal(a) samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne Jméno Příjmení

Obsah

Název práce: Kritéria pseudonáhodnosti posloupností
Autor: Ivo Machek
Katedra (ústav): Katedra algebry
Vedoucí bakalářské práce: Mgr. Štěpán Holub, Ph.D.
e-mail vedoucího: holub@mff.cuni.cz

Abstrakt: Klíčová slova: pseudonáhodnost, teoretické a empirické testy, spektrální test

Title: Criteria of pseudorandomness
Author: Ivo Machek
Department: Department of algebra
Supervisor: Mgr. Štěpán Holub, Ph.D.
Supervisor's e-mail address: holub@mff.cuni.cz

Abstract: Keywords: pseudorandomness, theoretical and empirical tests, spectral test

Kapitola 1

Pojem náhodnosti

Cílem této kapitoly bude seznámit se s různými pohledy na pojem náhodnost. Nejdříve se dostaneme ke konceptu pseudonáhodnosti, poté uvedeme definici náhodné posloupnosti, jak ji zavedli Kolmogorov, Solomonov a Chaitin.

1.1 Pseudonáhodnost

Náhodně zvolená čísla jsou nedílnou součástí mnoha různých aplikací. Odlišujeme dva přístupy, jak je získávat. Prvním z nich jsou generátory náhodných čísel využívající přírodní zdroj náhodnosti. Jako jednoduchý příklad nám poslouží házení mincí či ruleta. Složitější generátory mohou být založeny například na času mezi emisemi částic při radioaktivním rozpadu. S nástupem výpočetní techniky, ale vzrostl zájem o zdroje náhodnosti, které by byly snadno a rychle přístupné ve velkém množství. To vedlo k rozvoji tzv. pseudonáhodných generátorů.

Uvažujme posloupnost čísel $\{x_n\}$ z množiny $\{0, 1, \ldots, 2047\}$ získanou generátorem tvaru $x_{n+1} = 1365x_n + 1 \bmod 2048$, $x_0 = 1$ (viz. sekce 2.2):

1 1366 911 380 557 498 1883 56 665 462 1895 52 1349 234 1971 1392 ...

Abychom o daném generátoru mohli říct, že je pseudonáhodný, musí splňovat následující podmínky. První podmínkou je „prodloužení náhodnosti“. To znamená, že požadujeme, aby výstupní posloupnost generátoru byla delší než vstupní. Druhou podmínkou je čistě deterministické generování posloupnosti ze vstupní hodnoty, neboli nepovolíme pseudonáhodnému generátoru jiný zdroj náhodnosti než je náhodné zvolení vstupu. Tyto dvě podmínky

předchozí příklad splňuje. Celá posloupnost je vygenerována rekurzivně ze vstupní hodnoty $x_0 = 1$.

Třetí podmínka na pseudonáhodné generátory je nerozlišitelnost jimi vygenerovaných posloupností od posloupností náhodných, neboli požadujeme, aby tyto posloupnosti měly stejné statistické vlastnosti, jaké bychom očekávali u náhodných posloupností. Ověřování této podmínky bude hlavní náplní Kapitoly 2.

V následujícím podáme formální definici pseudonáhodných generátorů a nastíníme otázku jejich existence.

Definice 1. *Řekneme, že funkce $\varepsilon : \mathbb{N} \to \mathbb{N}$ je zanedbatelná, pokud*

$$\forall k > 0 \; \exists n_0 \; \forall n > n_0 : \; \varepsilon(n) < \frac{1}{n^k}$$

Budeme pracovat s náhodnými veličinami s hodnotami v množině $\{0,1\}^n$. Značit je budeme X_n nebo Y_n a nazveme je distribucemi na $\{0,1\}^n$. Volně řečeno tyto náhodné veličiny nám určují pravděpodobnosti jednotlivých prvků množiny $\{0,1\}^n$. Speciální případ uniformní distribuce značíme U_n a definujeme jej vztahem

$$\Pr(U_n = a) := \frac{1}{2^n}$$

pro všechna $a \in \{0,1\}^n$. Dále zavedeme značení:

$$\Pr_{x \in X}(f(x) = a) := \Pr(f(X) = a),$$

kde f je libovolná funkce definovaná na oboru hodnot náhodné veličiny X a a náleží do oboru hodnot funkce f.

Definice 2. *Nechť $p : \mathbb{N} \to \mathbb{N}$ je polynom. Řekneme, že soubory distribucí $\{X_{p(n)}\}_{n \in N}$ a $\{Y_{p(n)}\}_{n \in N}$ jsou výpočetně nerozlišitelné (značíme $\equiv_c$), pokud pro každý pravděpodobnostní polynomiální algoritmus D je funkce*

$$\varepsilon(n) = |\Pr_{x \in X_{p(n)}}(D(x) = 1) - \Pr_{y \in Y_{p(n)}}(D(y) = 1)|$$

zanedbatelná.

Soubory distribucí $\{X_{p(n)}\}_{n \in \mathbb{N}}$ budeme dále značit jen $\{X_n\}$. Soubor uniformních distribucí budeme vždy uvažovat příslušný k souboru distribucí s nímž jej porovnáváme.

Definice 3. *O souboru distribucí* $\{X_n\}$ *řekneme, že je pseudonáhodný, pokud je výpočetně nerozlišitelný od souboru uniformních distribucí, tedy*

$$\{X_n\} \equiv_c \{U_n\}.$$

Definice 4. *Pseudonáhodný generátor je funkce* $g : \{0,1\}^* \to \{0,1\}^*$ *taková, že*

1. *pro* x, $|x| = n$, *je* $|g(x)| = l(n)$, *kde* $l(n) > n$
2. *funkce* g *je polynomiálně vyčíslitelná*
3. *soubor distribucí* $\{g(U_n)\}$ *je pseudonáhodný*

Předpokládejme, že máme pseudonáhodný generátor, který prodlužuje vstupy délky n na výstupy délky $l(n)$. Definujemeli generátor g_1 tak, že na vstup x délky n vydá prvních $(n+1)$ bitů posloupnosti $g(x)$, pak je g_1 opět pseudonáhodný generátor, neboť algoritmus odlišující $\{g_1(U_n)\}$ od $\{U_{n+1}\}$ by odlišoval i $\{g(U_n)\}$ od $\{U_{l(n)}\}$.

Obrácený postup také platí. Pokud máme pseudonáhodný generátor, který vstupy délky n prodlužuje na výstupy délky $n+1$, pak pro každý polynom $l(n)$ existuje generátor, který prodlužuje vstupy délky n na výstupy délky $l(n)$. Konstrukci takového generátoru a důkaz správnosti můžeme najít v knize od O. Goldreicha [2] str. 96-99.

Vezměme si nyní podmínku 3. z Definice 4 a zkusme se na ni podívat zblízka.

$$\{g(U_n)\} \equiv_c \{U_{l(n)}\}$$

znamená, že pokud si vezmeme libovolný $TEST$ (v našem kontextu nejčastěji ověřující nějakou vlastnost typickou pro náhodné posloupnosti), potom

$$|\Pr_{x \in g(U_n)}(TEST(x) = 1) - \Pr_{y \in U_{l(n)}}(TEST(y) = 1)|$$

je funkce zanedbatelná. Neboli pravděpodobnost, že nalezneme danou vlastnost v množině posloupností vygenerovaných generátorem g, je „stejná“ jako pravděpodobnost, že tu vlastnost nalezneme v množině všech posloupností dané délky.

Silnějším pojmem než výpočetní nerozlišitelnost souborů distribucí je jejich statistická blízkost. V tomto případě totiž požadujeme, aby soubory distribucí si byly blízké po prvcích.

Definice 5. *Řekneme, že soubory distribucí* $\{X_n\}$ *a* $\{Y_n\}$ *si jsou statisticky blízké, pokud jejich statistická diference*

$$\Delta(n) = \sum_{\alpha} |\Pr(X_n = \alpha) - \Pr(Y_n = \alpha)|$$

je funkce zanedbatelná v n*.*

Platí, že pokud jsou si soubory distribucí statisticky blízké, potom jsou i výpočetně nerozlišitelné, obrácená implikace ale neplatí (Goldreich [2] str. 89). Tento pojem tedy nemá v oblasti pseudonáhodnosti větší význam, neboť i dvě posloupnosti, které si nejsou statisticky blízké, mohou být výpočetně nerozlišitelné, což je pro nás klíčový pojem.

Definice 6. *Funkci* $f : \{0,1\}^* \to \{0,1\}^*$ *nazveme jednosměrnou, právě když platí:*

1. f *je polynomiálně vyčíslitelná*
2. *exitstuje* $\varepsilon > 0$ *takové, že* $|f(x)| > |x|^{\varepsilon}$
3. *pro každý pravděpodobnostní polynomiální algoritmus* M *platí:*

$$|\Pr_{x \in U_n}(M(f(x)) \in f^{-1}(f(x)))|$$

je funkce zanedbatelná v n*.*

Otázka existence pseudonáhodných generátorů je úzce propojena s otázkou existence jednosměrných funkcí. Dokonce platí následující ekvivalence.

Věta 7. *Existence pseudonáhodných generátorů je ekvivalentní existenci jednosměrných funkcí.*

Pro ilustraci se zacházením s definicí pseudonáhodného generátoru si ukážeme podstatně lehčí implikaci zleva doprava. Dle diskuze za Definicí 4 můžeme bez újmy na obecnosti předpokládat pseudonáhodný generátor, který prodlužuje vstupy délky n na výstupy délky $2n$.

Tvrzení 8. *Nechť* $g : \{0,1\}^n \to \{0,1\}^{2n}$ *je pseudonáhodný generátor, pak* g *je i jednosměrná funkce.*

Důkaz. Pro spor předpokládejme, že existuje pravděpodobnostní polynomiální algoritmus A a $k, n_0 \in \mathbb{N}$ takové, že pro všechna $n > n_0$ platí

$$\begin{aligned} \mathrm{Pr}_{x\in U_n}\Big(A(g(x)) \in g^{-1}(g(x))\Big) &\geq \frac{1}{n^k} \qquad \text{nebo-li} \\ \mathrm{Pr}_{x\in U_n}\Big(g(A(g(x))) = g(x)\Big) &\geq \frac{1}{n^k} \end{aligned}$$

Definujme algoritmus *TEST* následujícím vztahem:

$$TEST(y) = \begin{cases} 1 & g(A(y)) = y \\ 0 & \text{jinak} \end{cases}$$

Potom existuje $l > 0$ takové, že

$$\Big|\mathrm{Pr}_{x\in U_n}\Big(TEST(g(x)) = 1\Big) - \mathrm{Pr}_{y\in U_{2n}}\Big(TEST(y) = 1\Big)\Big| \geq \frac{1}{n^l}$$

neboť první z pravděpodobností je větší než $1/n^k$ z definice *TESTu* a druhá je menší než 2^{-n}, neboť

$$\begin{aligned} &\mathrm{Pr}_{y\in U_{2n}}\Big(TEST(y) = 1\Big) = \mathrm{Pr}_{y\in U_{2n}}\Big(g(A(y)) = y\Big) \leq \\ &\leq \mathrm{Pr}_{y\in U_{2n}}\Big(y \in \mathrm{Rng}(g)\Big) = \frac{|\mathrm{Rng}(g)|}{|U_{2n}|} \leq \frac{2^n}{2^{2n}} \end{aligned}$$

A tedy dostáváme spor s definicí pseudonáhodného generátoru. □

Dále uvedeme jednu z vět, která nám pomůže si utvořit obecnou představu o konstrukci pseudonáhodného generátoru z jednosměrné funkce. Důkaz můžeme najít opět v [2] str. 102-104.

Definice 9. *Nechť $f : \{0,1\}^* \to \{0,1\}^*$ je polynomiálně vyčíslitelná funkce, potom funkci $b : \{0,1\}^* \to \{0,1\}$ nazveme jejím těžkým bitem, právě když pro každý pravděpodobnostní polynomiální algoritmus A je*

$$\mathrm{Pr}_{x\in U_n}\Big(A(f(x)) = b(x)\Big) < \frac{1}{2} + \varepsilon(n),$$

kde ε je funkce zanedbatelná.

Věta 10. *Nechť f je jednosměrná permutace a nechť $b(x)$ je její těžký bit. Pak funkce $x \mapsto f(x) * b(x)$ je pseudonáhodný generátor ($*$ značí spojení dvou posloupností za sebe).*

Stejně jako u pseudonáhodných generátorů nemáme důkaz existence jednosměrných funkcí. Na druhou stranu existence jednosměrných funkcí je snáze uvěřitelná a máme hned několik adeptů, o kterých se předpokládá, že je obtížné je invertovat.

Z nich za předpokladu, že jsou jednosměrné, dokážeme dle předchozí věty zkonstruovat pseudonáhodný generátor, který splňuje naši základní definici. Jako příklad můžeme uvést funkci $\mathrm{RSA}_{N,e} = x^e \bmod N$, o té se předpokládá, že je jednosměrná a za její těžký bit $b(x)$ můžeme vzít první bit rozvoje x.

Velmi důležitou ekvivalentní definici pseudonáhodného generátoru dostaneme díky pojmu nepředpověditelnosti. Důležitost této definice vynikne zvlášť v kryptografii, protože je zásadní otázkou, zda útočník na pseudonáhodný generátor dokáže s nezanedbatelně větší pravděpodobností než $1/2$ předpovídat z části vygenerované posloupnosti bity následující.

Definice 11. *Řekneme, že soubor distribucí $\{X_n\}$ je nepředpověditelný, jestliže*

$$\Pr_{x\in X_n}\Big(A(x) = \mathrm{next}_A(x)\Big) \leq \frac{1}{2} + \varepsilon(n),$$

kde ε je funkce zanedbatelná a kde funkce next_A odpoví $(i+1)$-ní bit x, pokud A přečetl ze vstupu právě $i < |x|$ bitů a odpoví náhodně zvolený bit, pokud A přečetl celý vstup x.

Nyní předvedeme ideu důkazu ekvivalence nepředpověditelnosti a pseudonáhodnosti.

Věta 12. *Nechť $g : \{0,1\}^n \to \{0,1\}^{l(n)}$, kde $l(n) > n$, je polynomiálně vyčíslitelná funkce, potom soubor distribucí $\{g(U_n)\}$ je pseudonáhodný, právě když je nepředpověditelný.*

Důkaz. Pro spor předpokládejme, že $\{g(U_n)\}$ je předpověditelný, pak existuje pravděpodobnostní polynomiální algoritmus A, který s nezanedbatelně větší pravděpodobností než $1/2$ předpovídá $(i+1)$-ní bit $g(x)$ z prvních i bitů $g(x)$. Sestrojíme algoritmus B, který odpoví 1, pokud algoritmus A úspěšně předpoví $(i+1)$-ní bit a jinak 0. Potom máme algoritmus odlišující $\{g(U_n)\}$ od $\{U_{l(n)}\}$ s nezanedbatelnou pravděpodobností, a tedy spor.

Na druhou stranu nechť $\{g(U_n)\}$ je nepředpověditelný. Distribuce $Y^i_{l(n)}$ na $\{0,1\}^{l(n)}$ definujeme tak, že prvních i bitů bude rozděleno stejně jako tomu je u distribuce $g(U_n)$ a posledních $l(n) - i$ bitů bude rozděleno uniformně. Díky nepředpověditelnosti $(i+1)$-ního bitu platí, že $\{Y^i_n\} \equiv_c \{Y^{i+1}_n\}$ $\forall i =$

$0, ..., l(n) - 1$. A tudíž využitím toho, že relace $\equiv_c$ je tranzitivní, dostáváme, že $\{U_{l(n)}\} = \{Y_0\} \equiv_c \{Y_{l(n)-1}\} = \{g(U_n)\}$. □

1.2 Kolmogorovská složitost

Pseudonáhodné posloupnosti nedokážeme od náhodných posloupností výpočetně odlišit, ale z intuitivního pohledu posloupnosti, které jsme získali čistě deterministickou cestou, za náhodné považovat nemůžeme. Proto si pro srovnání s konceptem pseudonáhodnosti uvedeme klasickou definici náhodnosti, kterou zavedli Kolmogorov, Solomonov a Chaitin.

Volně řečeno, binární posloupnost budeme považovat za kolmogorovsky náhodnou, pokud nejkratší program, který ji produkuje, není kratší než sama posloupnost. Na tento program můžeme nahlížet jako na nejkratší vysvětlení či popis jevu, který je popsán danou posloupností. Pro formální přiblížení této definice si představíme pojem kolmogorovské složitosti, která nám právě udává délku nejkratšího programu produkujícího danou posloupnost. Budeme předpokládat znalost základních pojmů z teorie složitosti jako jsou Turingův stroj a rekurzivní funkce.

Nechť X je množina objektů (můžeme ji ztotožnit s $\{0,1\}^*$), kterou lze efektivně očíslovat přirozenými čísly pomocí funkce $n : X \to \mathbb{N}$. Každý prvek množiny X budeme chtít popsat konečnou binární posloupností a protože množinu přirozených čísel můžeme ztotožnit s množinou všech konečných binárních posloupností, tak nás bude zajímat, jestli $n(x)$ je nejúspornější způsob, jak popsat prvky $x \in X$.

Zadefinujme pro $y \in \{0,1\}^*$, $l(y)$ jako počet bitů y. Nechť f je libovolná částečná funkce z $\{0,1\}^*$ do $\mathbb{N}$. Potom pro každý objekt $x \in X$ definujeme jeho složitost vzhledem k popisovací funkci f předpisem:

$$C_f(x) = \min\{l(y) \mid y \in \{0,1\}^* \wedge f(y) = n(x)\}$$

a $C_f(x) = \infty$, pokud žádné takové y neexistuje.

Složitost objektů $x \in X$ můžeme tedy porovnávat vzhledem k dané popisovací funkci. Problém nastane pokud zvolíme odlišnou popisovací funkci, protože jednotlivé složitosti se mohou diametrálně lišit. Abychom mohli říct, že x_1 je složitější než x_2, potřebujeme vědět, že složitost objektu x je vlastností závisející pouze na x a ne na popisovací funkci f. Toho dosáhneme definováním tzv. univerzální popisovací funkce f_0, po které budeme chtít, aby minimalizovala, až na aditivní konstantu, hodnotu $C_f(x)$ pro všechna

$x \in X$, nebo-li

$$C_{f_0}(x) \leq C_f(x) + c_f \qquad \forall x \in X$$

pro všechny f z nějaké podmnožiny částečných funkcí definovaných na množině $\{0,1\}^*$. Poté kolmogorovskou složitost C zadefinujeme vztahem

$$C(x) = C_{f_0}(x).$$

Otázkou ale zůstává, jestli taková univerzální popisovací funkce existuje. Pokud bychom uvažovali třídu všech částečných funkcí $f : \{0,1\}^* \to \mathbb{N}$, tak taková funkce neexistuje (viz [4] str. 97). Pokud se ale omezíme na částečné rekurzivní funkce, potom mezi nimi můžeme univerzální funkci najít.

Tvrzení 13. *V třídě všech částečných rekurzivních funkcí existuje funkce univerzální.*

Důkaz. Nechť f_0 je částečná rekurzivní funkce, která odpovídá univerzálnímu Turingovu stroji U. Vstup stroje U očekáváme ve tvaru: $11\ldots10{*}T{*}p$, kde počet jedniček na začátku vstupu je $l(T)$. Díky nim si stroj U nejdříve rozdělí vstup na posloupnosti T, p a potom simuluje běh stroje T na vstupu p. Určuje-li Turingův stroj T částečnou rekurzivní funkci f_T, potom platí, že

$$C_{f_0}(x) \leq C_{f_T}(x) + c_T,$$

kde c_T můžeme položit rovno $2l(T) + 1$. Jelikož každá částečná rekurzivní funkce odpovídá nějakému Turingovu stroji T, tak jsme s důkazem hotovi. □

Věta 14. *Kolmogorovská složitost není částečně rekursivní funkce.*

Důkaz. Pro spor předpokládejme, že máme algoritmus, který kolmogorovskou složitost počítá. Definujme algoritmus P, který na vstup $n \in \mathbb{N}$ bude pracovat takto:

Algoritmus P

pro $i = 1 \ldots \infty$

přes všechny posloupnosti s délky i

jestliže $C(s) > n$ pak vrať $s \longrightarrow$ konec

Zápis tohoto algoritmu má fixní délku U, kterou můžeme považovat za konstantu. Výstupem algoritmu P je posloupnost s, kde $C(s) > n$. Uvážíme-li ale, že na popsání posloupnosti s jsme potřebovali jen $\log_2(n) + U$ bitů, dostáváme spor, neboť určitě existuje n_0 takové, že pro všechna $n > n_0$

$$C(s) \leq U + \log_2(n) < n < C(s)$$ □

Tato věta nám říká, že stejně jako u pseudonáhodnosti, ani v tomto případě nejsme schopni prakticky ověřit, zda-li je posloupnost náhodná. Pokud ale nalezneme „jednoduché vysvětlení“ můžeme tvrdit, že náhodnou není, což je speciálně případ všech posloupností vygenerovaných pseudonáhodnými generátory.

Kapitola 2

Testy náhodnosti

V předchozí kapitole jsme zmínili pseudonáhodné generátory, jimiž vygenerované posloupnosti nedokáže žádný efektivně pracující algoritmus odlišit od uniformně zvolených posloupností (viz konstrukce pseudonáhodného generátoru z jednosměrné funkce). Takové generátory je vhodné používat pro kryptografické účely nebo v aplikacích, kde není zapotřebí rychlé generování.

V aplikacích, jako jsou různé simulace přírodních jevů či generování náhodných kvantit v numerické analýze, je zapotřebí generovat náhodná čísla rychle, což kryptograficky bezpečné generátory nesplňují. K tomu slouží velká řada různých typů jednoduchých generátorů, u kterých ale nemůžeme říct, že jsou pseudonáhodné ani za předpokladu výpočetní neřešitelnosti nějakého obtížného problému. Proto, abychom je mohli úspěšně využívat, je zapotřebí se ujistit alespoň o jisté míře náhodnosti. Cílem této práce je představení různých přístupů k tomu, jak posuzovat pseudonáhodnost generátorů.

2.1 Konstrukce testu

Jednou z možností, jak testovat pseudonáhodné generátory, by mohlo být využití přímo jejich definice, a pro libovolný pravděpodonostní polynomiální algoritmus zkoumat příslušné pravděpodobnosti. Problémem, na který bychom narazili, je asymtotika, protože námi posuzované generátory mají definiční obor omezen na konečnou množinu (většinou $\mathbb{Z}_m$ nebo $\mathbb{Z}_m^k$). Proto bychom se mohli snažit ověřovat podmínku 3. z Definice 4 pro pevné „malé“ $\varepsilon > 0$:

$$|\mathrm{Pr}_{x \in g(U_n)}(D(x) = 1) - \mathrm{Pr}_{y \in U_{l(n)}}(D(y) = 1)| < \varepsilon.$$

Kdybychom uměli spočítat tyto jednotlivé pravděpodobnosti, můžeme tento postup použít, ale jelikož $\Pr_{x \in g(U_n)}(D(x) = 1)$ je většinou obtížné vyjádřit pouze z parametrů generátoru a hodnotu $D(g(y))$ je časově náročné spočítat pro všechny různé vstupní hodnoty y přímo z posloupností $g(y)$, zaměříme se pouze na testování náhodných vzorků z vygenerovaných posloupností.

Statistická hypotéza je tvrzení o pravděpodobnostním rozdělení jedné či více náhodných veličin. V našem případě hypotéza H tvrdí, že prvky posloupnosti $u_1, \ldots, u_n$ jsou hodnoty nezávislých uniformně rozdělených náhodných veličin. Naším cílem bude sestrojit testy, které by na základě konkrétní posloupnosti získané daným generátorem, buď hypotézu H zamítly nebo nezamítly. Výsledek ale nemůže být definitivní, neboť každá posloupnost dané délky je v uniformním rozdělení stejně pravděpodobná. Výsledek si spíše můžeme vyložit tak, že vzhledem k danému testu je pravděpodobné, že hypotéza H buď platí nebo neplatí.

Každý test je založen na nějaké vlastnosti náhodné posloupnosti, kterou umíme statisticky popsat. To znamená, že na základě této vlastnosti dokážeme sestrojit testovou statistiku, jejíž pravděpodobnostní rozdělení za předpokladu hypotézy H známe, nebo ho dokážeme úspěšně aproximovat.

Jednoduchý příklad nám zprostředkuje následující binární posloupnost $n = 15$ prvků:

$$1\ 1\ 1\ 0\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1$$

Testovou statistiku X zformulujeme jako počet jedniček v posloupnosti u_1, $\ldots, u_n$. Za předpokladu hypotézy H má testová statistika binomické rozdělení s parametry $(n, p) = (15, 1/2)$. Pozorovaná hodnota statistiky je $x_s = 14$.

$$\Pr(X \geq x_s) = \sum_{i=14}^{n} \binom{n}{i} \frac{1}{2^n} = \frac{1}{2048}$$

Horní mez, při které hypotézu H zamítáme se většinou volí mezi 0.05 a 0.01, tedy v tomto konkrétním případě bychom hypotézu H zamítli.

Pro konstrukci statistik, u kterých budeme znát pravděpodobnostní rozdělení za předpokladu hypotézy H, se používají tzv. testy dobré shody. Jedním z nich je χ^2-test, jenž si nyní stručně popíšeme, abychom ho mohli v následujícím příkladu použít. Definici χ^2 rozdělení o m stupních volnosti (χ^2_m) a důkaz následujícího tvrzení lze nalézt v knize od K. Zváry a J. Štěpána [5].

Mějme n nezávislých pozorování, která mohou nabývat m hodnot $A_1, \ldots, A_m$ s pravděpodobnostmi $p_1, \ldots, p_m$. Označme X_i počet pokusů v nichž

nastal výsledek A_i. Potom řekneme, že náhodný vektor $X = (X_1, \ldots, X_m)$ má multinomické rozdělení s parametry $n, p_1, \ldots, p_m$. Platí následující věta:

Věta 15. *Nechť náhodný vektor X má multinomické rozdělení s parametry $n, p_1, \ldots, p_m$. Potom pro velká $n = \sum_{i=1}^{m} X_i$ platí, že*

$$X^2 = \sum_{k=1}^{m} \frac{(X_k - np_k)^2}{np_k} \sim \chi^2_{m-1}.$$

Předchozí věta nám dává návod, jak zkonstruovat testovou statistiku, jejíž rozdělení budeme znát. Hlavní součástí takové konstrukce je znalost pravděpodobností jednotlivých jevů A_i za platnosti hypotézy H.

V následující části si předvedeme jeden konkrétní jednoduchý příklad, který je založen na χ^2-testu. Nebudeme se zabývat rozborem dalších používaných podobných testů, které většinou stojí na složitějších poznatcích z teorie pravděpodobnosti. Rozsáhlý výčet testů, lze nalézt například v knize od D. Knutha [3] str. 61-73.

Příklad (Gap test)
Mějme posloupnost $\{u_n\}$ jejíž prvky náleží do intervalu $[0, 1)$, zvolme parametry $0 \leq \alpha < \beta \leq 1$ a $n \in \mathbb{N}$. V tomto testu nás budou zajímat délky po sobě jdoucích podposloupností jejichž jediný prvek, který náleží do intervalu (α, β), je na začátku podposloupnosti. Například buď $\alpha = 10/100$, $\beta = 32/100$ a mějme následující posloupnost:

$$\left|\frac{11}{100}\right|\frac{18}{100}\frac{77}{100}\frac{60}{100}\left|\frac{31}{100}\frac{58}{100}\frac{57}{100}\right|\frac{20}{100}\frac{51}{100}\frac{98}{100}\frac{37}{100}\frac{80}{100}\frac{91}{100}\frac{38}{100}\left|\frac{17}{100}\frac{40}{100}\frac{91}{100}\right|\ldots$$

Každá z těchto podposloupností odpovídá jednomu náhodnému pokusu a jev A_i nastane, právě když její délka je i. V našem konkrétním příkladě nastane posloupnost jevů $A_1, A_3, A_3, A_7, \ldots$. Nechť X_i je počet podposloupností délky i pro $0 \leq i < t$, kde t je apriori zvolené tak, že pravděpodobnost nalezení podposloupnosti délky t je už malá a nechť X_t je počet výskytů podposloupností délky větší nebo rovny t.

Abychom mohli zkonstruovat statistiku analogicky, jak je tomu ve Větě 15, potřebujeme znát pravděpodobnosti jednotlivých jevů A_i za předpokladu platnosti hypotézy H:

$$p_r = (1-p)^r p, \text{ pro } 0 \leq r < t \text{ a } p_t = (1-p)^t,$$

kde $p = \beta - \alpha$ je pravděpodobnost, že u_i padne do intervalu (α, β). Platí, že $\sum_{r=0}^{t} p_r = 1$ a můžeme porovnávat statistiku X^2 s χ^2-rozdělením o t stupních volnosti.

Můžeme odlišit dva způsoby, jak spočítat testovou statistíku. Nejjednodušší způsob je vygenerování posloupnosti daným generátorem a spočítání testové statistiky přímo z této posloupnosti. Takovému postupu se říká empirický test. Daleko efektivnějším ale u složitějších testů obtížně proveditelným způsobem je vyjádření testové statistiky v závislosti na parametrech daného generátoru. To by nám umožilo jednodušše optimalizovat výběr parametrů generátoru vzhledem k danému testu, aniž bychom museli pro každou různou volbu parametrů generovat nové posloupnosti. Konkrétní ukázku takového postupu představíme na závěr sekce o lineárních kongruenčních generátorech.

V praxi existují velice rozsáhlé sady různých testů snažících se pokrýt nejdůležitější vlastnosti náhodných posloupností. Ale žádné konečné množství testů, úspěšně projitých, nám nemůže zaručit pseudonáhodnost generátoru ve smyslu definice z první kapitoly. Testy, kterým generátor postoupíme, by se v ideálním případě měli volit tak, aby odpovídali použití generátoru, protože potom se snáze vyhneme situaci, kdy by se cílová aplikace mohla sama stát testem, který odhaluje nenáhodnost generátoru.

2.2 Lineární kongruenční generátory a jejich základní vlastnosti

Nejdůležitější částí v konstrukci pseudonáhodného generátoru je jeho matematický rozbor - tzn. teoretické testování. Existují třídy pseudonáhodných generátorů, o nichž máme velice rozsáhlé matematické poznatky a pro které existuje rozsáhlá škála různých měřítek a testů. Jde většinou o generátory lineární, které budou hlavní náplní této části. Na druhé straně pak stojí generátory nelinární, které nemají k dispozici tak rozsáhlý rozbor jejich vlastností. Pro prakrické účely je důležité, že jsou pomalejší než generátory lineární.

Většina používaných generátorů odpovídá následujícímu obecnému rekurzivnímu modelu. Nechť S je konečná množina stavů, μ pravděpodobnostní rozdělení na S, které slouží k vybrání počátečního stavu s_0, $f : S \to S$, U množina všech výstupů a $g : S \to U$. Množina U se většinou volí rovna

intervalu $[0, 1)$ a rozdělení μ jako uniformní rozdělení na množině S. Proto i my zde budeme explicitně předpokládat tyto varianty. Po volbě počátečního stavu s_0 dostaneme následnou posloupnost stavů $\{s_n\}$ rekurentním vztahem $s_i = f(s_{i-1})$. Výstupní posloupnost $\{u_n\}$ získáme vztahem $u_i = g(s_i)$.

Jelikož množina S je konečná, tak pro nějaká $i, r \in \mathbb{N}$ musí platit $s_i = s_{i+r}$. Nejmenší možné takové r nazveme periodou posloupnosti a označíme ρ. Nutně platí, že $\rho \leq |S|$. Dostatečně velká perioda je jednou ze základních podmínek, které zohledňujeme při konstrukci generátoru. Ovšem nemůže to být podmínka postačující.

Nyní definujeme *lineární kongruenční generátory řádu* k, kde $k \in \mathbb{N}$. Nechť $m \in \mathbb{N}$ a $a, a_0, \ldots, a_{k-1} \in \mathbb{Z}_m$, pak funkci f z předchozího obecného modelu definujeme vztahem

$$f : (x_i, \ldots, x_{i+k-1}) \mapsto (x_{i+1}, \ldots, x_{i+k}),$$

kde

$$x_{i+k} = a + a_0 x_i + \ldots + a_{k-1} x_{i+k-1} \bmod m \tag{2.1}$$

V tomto případě se množina stavů S rovná $\{(x_1, \ldots, x_k) \,|\, x_i \in \mathbb{Z}_m\}$ a její mohutnost $|S| = m^k$. Funkci g lze volit například jednoduše $g(x_1, \ldots, x_k) = x_k/m$.

Důležitý speciální případ vztahu 2.1 dostaneme pro $k = 1$. Těmto generátorům se říká *lineární kongruenční generátory* (LCG). Definujeme je vztahem $x_{i+1} = ax_i + c \bmod m$ a po označení $b = a - 1$ indukcí dostaneme

$$x_i = \left(a^i x_0 + \frac{(a^i - 1)c}{b} \right) \bmod m \tag{2.2}$$

Maximální možná perioda těchto generátorů je m a platí následující tvrzení, které nám charakterizuje konkrétní generátory, které maximální periody dosahují. Důkaz můžeme najít v [3] str. 17-19.

Věta 16. *Lineární kongruenční posloupnost definovaná parametry* (x_0, a, c, m) *má periodu délky* m*, právě když platí následující podmínky.*

1. $NSD(c, m) = 1$
2. $a \equiv 1 \bmod p$*, pro každé* p *prvočíslo dělící* m
3. $a \equiv 1 \bmod 4$*, pokud je* m *násobkem* 4

Délka periody je sice velmi důležitá, ale o vlastnostech a struktuře generátoru nám nic neříká. Proto je důležité matematicky rozebrat dané generátory daleko podrobněji. Nyní si předvedeme, jak pro třídu lineárních kongruenčních generátorů (LCG) můžeme matematicky odvodit výsledek jednoho z používaných testů.

Permutační test porovnává četnosti jednotlivých lineárních uspořádání v k-ticích uvažované posloupnosti. Nás bude zajímat jeho nejjednodušší varianta, budeme zkoumat pravděpodobnost, že $x_{i+1} < x_i$. Hodnota, kterou očekáváme od náhodné posloupnosti, je $1/2$. Testovou statistiku odvozenou od této vlastnosti zformulujeme takto:

$$X_P = \sum_{i=1}^{n} [x_{i+1} < x_i],$$

neboli počet případů, kdy $x_{i+1} < x_i$.

Nejdříve si zaveďme dvě pomocné funkce.

Definice 17. *Nechť $x, x_1, \ldots, x_t \in \mathbb{R}$, pak definujme:*

$$\delta(x) = \lfloor x \rfloor + 1 - \lceil x \rceil = [x \text{ je celé číslo}]$$

$$\delta\Big((x_1, \ldots, x_t)\Big) = \delta(x_1) \cdot \ldots \cdot \delta(x_t) \tag{2.3}$$

$$\begin{aligned} ((x)) &= x - \lfloor x \rfloor - \frac{1}{2} + \frac{1}{2}\delta(x) = x - \lceil x \rceil + \frac{1}{2} - \frac{1}{2}\delta(x) \\ &= x - \frac{1}{2}(\lceil x \rceil + \lfloor x \rfloor) \end{aligned}$$

Lemma 18. *Nechť $k, n \in \mathbb{N}$, $x \in \mathbb{R}$, pak platí následující vztahy:*

$$((-x)) = -((x)), \qquad ((x+n)) = ((x))$$

$$((nx)) = ((x)) + ((x + \frac{1}{n})) + \ldots + ((x + \frac{n-1}{n})) \tag{2.4}$$

$$\frac{x}{k} - \left\lfloor \frac{x}{k} \right\rfloor = \frac{x \bmod k}{k} \tag{2.5}$$

Věta 19. *Nechť (x_0, a, c, m) jsou parametry lineární kongruenční posloupnosti s maximální periodou m. Nechť $b = a - 1$ a $d = NSD(m, b)$. Potom*

$$\sum_{i=1}^{m} [x_{i+1} < x_i] = \frac{m}{2} + (c \bmod d) - \frac{d}{2} \tag{2.6}$$

Důkaz. Neboť uvažujeme LCG s maximální periodou, musí splňovat podmínky Věty 16. Platí, že b je dělitelné p pro všechna prvočísla p dělící m, proto $d = NSD(m, b) \neq 1$.

Definujme $f(x) = (ax + c) \bmod m$. Funkce $\lceil (x - f(x))/m \rceil$ je rovna 1, právě když $f(x) < x$ a 0 jinak. Opět díky tomu, že uvažujeme LCG s maximální periodou, můžeme psát:

$$\sum_{i=1}^{m} [x_{i+1} < x_i] = \sum_{0 \leq x < m} \left\lceil \frac{x - f(x)}{m} \right\rceil \tag{2.7}$$

Dosazením z předchozí definice dostáváme:

$$\begin{aligned} \left\lceil \frac{x - f(x)}{m} \right\rceil &= \frac{x - f(x)}{m} - \Big(\Big(\frac{x - f(x)}{m}\Big)\Big) + \frac{1}{2} - \frac{1}{2}\delta\Big(\frac{x - f(x)}{m}\Big) \\ &= \frac{x - f(x)}{m} + \Big(\Big(\frac{bx + c}{m}\Big)\Big) + \frac{1}{2} \end{aligned}$$

neboť $x \neq f(x)$ a tudíž $(x - f(x))/m$ není celé číslo pro všechna x. Dále platí, že

$$\sum_{0 \leq x < m} \frac{x - f(x)}{m} = \sum_{0 \leq x < m} \frac{x}{m} - \sum_{0 \leq x < m} \frac{f(x)}{m} = 0$$

neboť $\{x \,|\, x \in \mathbb{Z}_m\} = \{f(x) \,|\, x \in \mathbb{Z}_m\}$.

Dostáváme:

$$\sum_{0 \leq x < m} \left\lceil \frac{x - f(x)}{m} \right\rceil = \sum_{0 \leq x < m} \Big(\Big(\frac{bx + c}{m}\Big)\Big) + \frac{m}{2}$$

Nechť $b = b_0 d$ a $m = m_0 d$, potom platí, že $(bm_0 = b_0 d m_0 = b_0 m) \bmod m = 0$ a tedy:

$$\Big(\Big(\frac{b(x + jm_0) + c}{m}\Big)\Big) = \Big(\Big(\frac{bx + c}{m} + j\frac{bm_0}{m}\Big)\Big) = \Big(\Big(\frac{bx + c}{m}\Big)\Big)$$

pro všechna $j \in \{0, 1, \ldots, d - 1\}$. A tedy platí:

$$\begin{aligned} \sum_{0 \leq x < m} \Big(\Big(\frac{bx + c}{m}\Big)\Big) &= d \sum_{0 \leq x < m_0} \Big(\Big(\frac{bx + c}{m}\Big)\Big) = d \sum_{0 \leq x < m_0} \Big(\Big(\frac{c}{m} + \frac{b_0 x}{m_0}\Big)\Big) \\ &= d\Big(\Big(m_0 \frac{c}{m}\Big)\Big) = d\Big(\Big(\frac{c}{d}\Big)\Big). \end{aligned}$$

Jelikož $NSD(b_0, m_0) = 1$, tak platí, že funkce $(b_0 x) \bmod m_0$ je permutací množiny $\{0, 1, \ldots, m_0 - 1\}$ a poslední rovnost jsme dostali dle Lemma 18. Nakonec dostáváme:

$$\begin{aligned}\sum_{i=1}^{m}[x_{i+1} < x_i] &= \frac{m}{2} + d\Big(\Big(\frac{c}{d}\Big)\Big) = \frac{m}{2} + d\Big(\frac{c}{d} - \Big\lfloor\frac{c}{d}\Big\rfloor - \frac{1}{2} + \frac{1}{2}\delta\Big(\frac{c}{d}\Big)\Big) = \\ &= \frac{m}{2} + (c \bmod d) - \frac{d}{2}\end{aligned}$$

použitím posledního vzorce z lemmatu a faktu, že $NSD(c, d) = 1$ a tedy $\delta(c/d) = 0$. □

Příklad
Nechť $m = 2048$, uvažujme a a c taková, že daný generátor dosahuje maximální periody. Permutační test budeme provádět na celé periodě a Věta 19 nám dává vzorec pro výpočet testové statistiky. Pro zjednodušení zafixujeme $c = 1$ a budeme se snažit najít a, pro která zamítneme hypotézu H, za které má X_P binomické rozdělení s parametry $(2048, 1/2)$.

$$X_P = \sum_{i=1}^{m}[x_{i+1} < x_i] = \frac{m}{2} + (c \bmod d) - \frac{d}{2} = \frac{m-d}{2} + 1$$

Tedy hodnota testové statistiky závisí na $d = NSD(a-1, 2^{11})$ a čím větší je tento největší společný dělitel, tím nižší hodnotu bude mít testová statistika. Hraničními případy jsou:

$$\begin{aligned}NSD(a-1, 2^{11}) = 64 \ldots X_P = 993 \ldots \Pr(X < X_P) \doteq 0.08 \\ NSD(a-1, 2^{11}) = 128 \ldots X_P = 961 \ldots \Pr(X < X_P) \doteq 0.0025\end{aligned}$$

Hypotézu H bychom zamítli, jestliže $a - 1$ je dělitelné 128.

2.3 Koncept k-distribuovanosti

V následujících sekcích se budeme věnovat jednomu z důležitých teoretických testů - spektrálnímu testu. Základní vlastností náhodných posloupností, na které je tento test založen, je k-distribuovanost. Proto než se pustíme do jeho rozboru, tak si tento pojem přiblížíme.

Uvažujme náhodnou veličinu X s rovnoměrným rozdělením v intevalu $[0, 1)$, potom $\Pr(v \leq X < w) = w - v$ pro pevně zvolená $v, w \in [0, 1]$. Nyní

budeme chtít tuto vlastnost převést na nekonečné posloupnosti. Předpokládejme platnost hypotézy H pro posloupnost $\{u_n\}$ ($u_i \in [0,1)$), pak musí platit, že

$$\lim_{n\to\infty} \frac{\nu(n)}{n} = w - v, \tag{2.8}$$

kde $\nu(n)$ značí počet $i \leq n$ takových, že $v \leq u_i < w$. Řekneme, že posloupnost $\{u_n\}$ je 1-distribuovaná, pokud pro všechna možná v, w platí 2.8.

Otázkou existence předchozí limity, jakožto i limit následujících v této kapitole, se zabývat nemusíme, protože budeme pracovat jen s posloupnostmi periodickými, a tedy existence limity je nutná.

Na příkladu si ukážeme, že rovnoměrné rozdělení prvků v posloupnosti nám ještě nezaručuje její náhodnost. Mějmě dvě 1-distribuované posloupnosti $\{u_n\}$ a $\{v_n\}$. Definujme posloupnost $\{y_n\}$ vztahem $y_{2k} = u_k/2$ a $y_{2k+1} = 1/2 + v_k/2$. Potom posloupnost $\{y_n\}$ je 1-distribuovaná, ale zároveň každý její sudý člen je z intervalu $[0, 1/2)$ a každý její lichý člen z intervalu $[1/2, 1)$, tedy dle každé přirozené definice nemůžeme $\{y_n\}$ považovat za náhodnou. S pojmem 1-distribuovanosti si tedy zdaleka nevystačíme, proto zavedeme obecnější definici.

Definice 20. *Řekneme, že posloupnost $\{u_n\}$ je k-distribuovaná, pokud pro každé $v_0, ..., v_{k-1}$ a $w_0, ..., w_{k-1}$ ($v_i \leq w_i$) platí, že*

$$\lim_{n\to\infty} \frac{|\{i \,|\, i \leq n \wedge (v_{i+j} \leq u_{i+j} < w_{i+j} \ \forall j = 0, ..., k-1)\}|}{n} = \prod_{i=0}^{k-1} (w_i - v_i).$$

Je-li posloupnost k-distribuovaná, pak je l-distribuovaná pro všechna $l \in \{1, \ldots, k\}$, neboť stačí položit $v_{k-1} = \ldots = v_{k-j} = 0$ a $w_{k-1} = \ldots = w_{k-j} = 1$ a dostáváme $(k-j)$-distribuovanost pro všechna $j \in \{1, \ldots, k-1\}$.

Díky volbě $v_0 = v_1 = 0$ a $w_0 = w_1 = 1/2$ vidíme, že posloupnost $\{y_n\}$ z předchozího příkladu není 2-distribuovaná, neboť neexistuje ani jedna dvojice (y_{2k}, y_{2k+1}), která by ležela ve čtverci $[0, 1/2)^2$.

Při zkoumání pseudonáhodných posloupností nás bude zajímat otázka pro jaké nejvyšší k je posloupnost k-distribuovaná. Je zřejmé, že dle každé přirozené definice by měla nekonečná náhodná posloupnost být k-distribuovaná pro všechna k. Takovou posloupnost nazveme ∞-distribuovanou. V [3] str. 159-163 můžeme nalézt diskuzi o definici nekonečné náhodné posloupnosti, která má jako základní stavební prvek právě ∞-distribuovanost.

Pro posloupnosti, které sestávají z prvků konečné množiny M ($|M| = m$), můžeme rozvést stejný koncept:

Definice 21. *Řekneme, že posloupnost $\{x_n\}$ ($x_i \in M$) je k-distribuovaná, pokud pro všechna $z_1, ..., z_k \in M$ platí, že*

$$\lim_{n\to\infty} \frac{|\{i \,|\, i \leq n \wedge (x_{i+j} = z_{i+j} \ \forall j = 0, ..., k-1)\}|}{n} = \frac{1}{m^k}$$

2.4 Odvození spektrálního testu pomocí Fourierových koeficientů

V této kapitole uvažujme $t \in \mathbb{N}$ jako dimenzi testu a $m \in \mathbb{N}$ jako modulus kongruentního generátoru. Funkce $\langle ., . \rangle$ bude značit skalární součin. Definujme posloupnosti $\{x_n\}$ ($x_i \in \mathbb{Z}_m$), $\{u_n \,|\, u_n = x_n/m\}$ a

$$\omega_i = (u_i, \ldots, u_{i+t-1}).$$

Dále pro přehlednost definujme:

$$M_t = \mathbb{Z}_m^t = \{(z_1, \ldots, z_t) \,|\, z_i \in \mathbb{Z}_m\}$$

$$N_t = \mathbb{Z}_m^t/m = \left\{ \left(\frac{z_1}{m}, \ldots, \frac{z_t}{m} \right) \,|\, z_i \in \mathbb{Z}_m \right\}$$

a pro $x \in \mathbb{R}$

$$e(x) = e^{2\pi i x}.$$

Dále budeme využívat vzorec pro výpočet sumy m-tých odmocnin z 1:

$$\frac{1}{m} \sum_{j=0}^{m-1} e\Big(\frac{jq}{m}\Big) = \delta\Big(\frac{q}{m}\Big), \tag{2.9}$$

kde funkci δ jsme definovali v Definici 17. Obecnou analogii platící pro všechna $Q = (q_1, \ldots, q_t) \in M_t$ zformulujeme takto:

$$\begin{aligned} \delta\Big(\frac{Q}{m}\Big) &= \delta\Big(\frac{q_1}{m}\Big) \cdot \ldots \cdot \delta\Big(\frac{q_t}{m}\Big) = \\ &= \frac{1}{m^t} \sum_{j_1}^{m-1} \cdots \sum_{j_t}^{m-1} e\Big(\frac{j_1 q_1 + \ldots + j_t q_t}{m}\Big) \\ &= \frac{1}{m^t} \sum_{K \in M_t} e\Big(\frac{\langle K, Q \rangle}{m}\Big). \end{aligned}$$

Nechť $f_\omega : N_t \to [0,1]$ je definována následujícím vztahem. Ten je ekvivalentní vzorci z Definice 21.

$$f_\omega(z_1,\ldots,z_t) = \lim_{n\to\infty} \frac{|\{i \,|\, i \le n \wedge w_i = (z_1,\ldots,z_t)\}|}{n}$$

Funkce f_ω určuje pravděpodobnostní rozdělení na množině N_t dané posloupností $\{\omega_n\}$. Naším cílem je porovnat toto rozdělení s uniformním rozdělením na množině N_t, neboli s konstantní funkcí f_0 definovanou předpisem $(z_1,\ldots,z_t) \mapsto 1/m^t$. Budeme se toho snažit dosáhnout pomocí tzv. spektrální analýzy.

Jelikož N_t je konečná množina, můžeme pro vyjádření hodnot funkce f_ω využít diskrétní Fourierovu transformaci. Je-li $g : N_t \to \mathbb{R}$ funkce, pak pro všechna $W \in N_t$ platí (odvození lze nalézt v článku od R. R. Coveyoua a R. D. MacPhersona [1] str. 103), že

$$g(W) = \sum_{R\in M_t} \psi(R) e(\langle W, R\rangle),$$

kde $\psi(R)$ jsou jednoznačné a následujícího tvaru:

$$\psi(R) = \frac{1}{m^t} \sum_{Y\in N_t} g(Y) e(\langle -Y, R\rangle).$$

Pro $Q \in M_t$ definujeme

$$\varphi(Q) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} e(\langle Q, \omega_i\rangle) = \tag{2.10}$$

$$= \sum_{R\in N_t} f_\omega(R) e(\langle Q, R\rangle), \tag{2.11}$$

kde $f_\omega(R)$ můžeme nahlédnout jako limitní podíl vektorů R v posloupnosti $\{\omega_n\}$. Díky jednoznačnosti koeficientů funkce φ dostáváme vyjádření f_ω tvaru:

$$f_\omega(W) = \frac{1}{m^t} \sum_{K\in M_t} \varphi(K) e(\langle -K, W\rangle), \tag{2.12}$$

kde záměna množin M_t a N_t v sumách je způsobena tím, že funkci φ uvažujeme na množině M_t. Hodnoty $\varphi(Q)$ nazveme koeficienty funkce f_ω.

Tvrzení 22. *Posloupnost $\{u_n\}$ je t-distribuovaná, právě když pro koeficienty $\varphi(Q)$ funkce f_ω platí, že*

$$\varphi(Q) = \delta\Big(\frac{Q}{m}\Big),$$

neboli

$$\varphi(Q) = 1 \text{ pro } Q = 0 \ \wedge \ \varphi(Q) = 0 \text{ pro } Q \neq 0$$

pro všechna $Q \in M_t$.

Důkaz. Posloupnost $\{u_n\}$ je t-distribuovaná, právě když $f_\omega(W) = 1/m^t$ pro všechna $W \in N_t$.
($\Rightarrow$) Dle vztahu 2.11 platí, že

$$\varphi(Q) = \frac{1}{m^t} \sum_{R \in N_t} e(\langle Q, R\rangle) = \delta\Big(\frac{Q}{m}\Big).$$

($\Leftarrow$) Předpokládáme-li na druhou stranu vztahy pro $\varphi(Q)$ dostaneme z 2.12, že $f_\omega(W) = 1/m^t$ pro všechna $W \in N_t$. □

Předchozí tvrzení nám dává kritérium, jak poznat t-distribuovanou posloupnost. Ovšem uvážíme-li například lineární kongruenční generátory řádu k, tak pro $t > k$ nemohou být posloupnosti jimi vygenerované k-distribuované, neboť perioda takových posloupností může být maximálně m^k, což je ostře menší než $|N_t|$, a tedy existují vektory $W \in N_t$, pro které $f_\omega(W) = 0$.

Proto potřebujeme měřítko, které bychom mohli použít i v těchto případech. Idea a motivace jeho odvození je představena v [1]. Naším cílem bude, aby maximální hodnota $|Q|^{-1}$, pro Q taková, že $\varphi(Q) \neq 0$, byla minimální. Neboli budeme se snažit maximalizovat následující měřítko

$$\nu_t = \min\{|Q| \neq 0 : \varphi(Q) \neq 0\}.$$

Ne vždy je jednoduché spočítat hodnoty $\varphi(Q)$. Ukážeme si na několika příkladech jak se tyto hodnoty dají odvodit pro určité typy lineárních kongruenčních generátorů.

Uvažujme pevné $Q = (q_0, \ldots, q_{t-1})$, nechť $P_t(x) = q_0 + q_1 x + \ldots + q_{t-1}x^{t-1}$ je polynom a ρ označme periodu posloupnosti $\{x_n\}$. Díky periodočnosti $\{x_n\}$ existuje $n_0 \in \mathbb{N}$ takové, že platí

$$\varphi(Q) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} e(\langle Q, \omega_i\rangle) = \frac{1}{\rho} \sum_{i=n_0}^{n_0+\rho-1} e(\langle Q, \omega_i\rangle).$$

A jeli posloupnost $\{x_n\}$ čistě periodická, můžeme psát

$$\varphi(Q) = \frac{1}{\rho}\sum_{i=0}^{\rho-1} e(\langle Q, \omega_i\rangle) = \frac{1}{\rho}\sum_{i=0}^{\rho-1} e\Big(\frac{q_0x_i + q_1x_{i+1} + \ldots + q_{t-1}x_{i+t-1}}{m}\Big).$$

Příklad (LCG)
Uvažujme lineární kongruenční posloupnost s parametry (x_0, a, c, m) a délkou periody ρ. Potom dle vzorce 2.2 a faktu, že pro všechna $l \in \mathbb{Z}$ platí, že $e((l \bmod m)/m) = e(l/m)$, můžeme psát:

$$\begin{aligned}
\varphi(Q) &= \frac{1}{\rho}\sum_{i=0}^{\rho-1} e\bigg(\frac{1}{m}\Big(q_0x_i + q_1(ax_i + c) + \ldots + q_{t-1}\Big(a^{t-1}x_i + c\frac{a^{t-1}-1}{a-1}\Big)\Big)\bigg)\\
&= \frac{1}{\rho}\sum_{i=0}^{\rho-1} e\bigg(\frac{1}{m}\Big(x_iP_t(a) + c\Big(q_1\frac{a-1}{a-1} + q_2\frac{a^2-1}{a-1} + \ldots + q_{t-1}\frac{a^{t-1}-1}{a-1}\Big)\Big)\bigg)\\
&= \frac{1}{\rho}\sum_{i=0}^{\rho-1} e\bigg(\frac{1}{m}\Big(x_iP_t(a) + \frac{c}{a-1}\Big(P_t(a) - P_t(1)\Big)\Big)\bigg)\\
&= e\bigg(\frac{c(P_t(a) - P_t(1))}{m(a-1)}\bigg)\frac{1}{\rho}\sum_{i=0}^{\rho-1} e\bigg(\frac{x_iP_t(a)}{m}\bigg)
\end{aligned}$$

Splňujíli parametry (x_0, a, c, m) lineárního kongruenčního generátoru podmínky Věty 16, posloupnost $\{x_n\}$ má periodu ρ rovnu m a dle vztahu 2.9 platí

$$\frac{1}{m}\sum_{i=0}^{m-1} e\bigg(\frac{x_iP_t(a)}{m}\bigg) = \delta\bigg(\frac{P_t(a)}{m}\bigg)$$

a tedy $\varphi(Q) \neq 0$, právě když je splněna tato kongruence:

$$q_0 + q_1a + \ldots + q_{t-1}a^{t-1} \equiv 0 \bmod m \tag{2.13}$$

Pokud tedy chceme analyzovat lineární kongruenční generátory s maximální periodou pomocí spektrálního testu, bude nás zajímat, zda-li má předcházející kongruence nějaká malá řešení.

Příklad (Fibonacciho posloupnost)
Uvažujme nyní generátor založený na Fibonacciho posloupnosti: $x_{n+2} =$

$x_{n+1} + x_n \bmod m$. Můžeme se opět pokusit spočítat koeficienty $\varphi(Q)$ ze vztahu:

$$\varphi(Q) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n} e(\langle Q, \omega_i \rangle) = \frac{1}{\rho} \sum_{i=n_0}^{n_0+\rho-1} e(\langle Q, \omega_i \rangle).$$

Uvážímeli taková $Q \neq 0$, že polynom $P_t(z) = q_0 + q_1 z + \ldots + q_{t-1} z^{t-1}$ můžeme vyjádřit ve tvaru

$$P_t(z) = p(z)(1 + z - z^2),$$

kde $p(z) = a_0 + a_1 + \ldots + a_{t-3} z^{t-3}$ je celočíselný polynom, potom

$$\begin{aligned}
\langle Q, \omega_i \rangle &= \frac{1}{m}(q_0 x_i + q_1 x_{i+1} + q_2 x_{i+2} + \ldots + q_{t-2} x_{i+t-2} + q_{t-1} x_{i+t-1}) = \\
&= \frac{1}{m}(a_0 x_i + (a_1 + a_0) x_{i+1} + (a_2 + a_1 - a_0) x_{i+2} + \\
&\quad + \ldots + (a_{t-3} - a_{t-2}) x_{i+t-2} + (-a_{t-3}) x_{i+t-1}) = \\
&= \frac{1}{m} \sum_{j=0}^{t-3} a_j (-x_{i+j+2} + x_{i+j+1} + x_{i+j}) \in \mathbb{Z},
\end{aligned}$$

neboť $(-x_{i+j+2} + x_{i+j+1} + x_{i+j}) \equiv 0 \bmod m$. Jelikož $e(z) = 1$ pro $z \in \mathbb{Z}$, tak pro taková Q je $\varphi(Q) = 1$, což nám dává i pro tento typ generátorů částečnou použitelnost výše zmíněného kritéria.

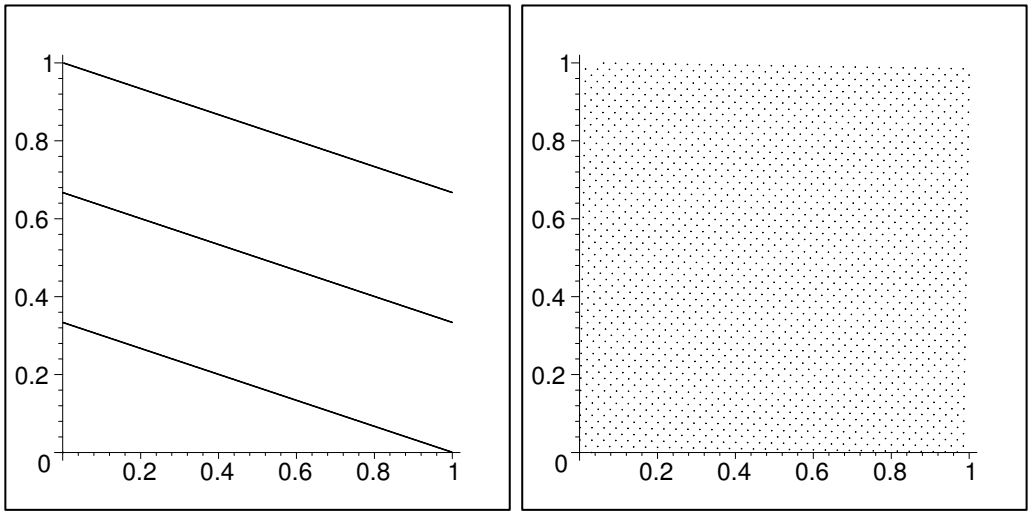
2.5 Spektrální test a mřížová struktura

Nyní si předvedeme odlišný geometrický pohled na spektrální test. Především se opět zaměříme na lineární kongruenční generátory, na které lze tento test aplikovat, protože vektory definované jejich výstupními posloupnostmi tvoří mřížovovou strukturu.

Mějme pseudonáhodný generátor definován množinou stavů S, funkcemi f, g a množinou $U = [0, 1)$. Zvolme pevné $t \in \mathbb{N}$ jako dimenzi testu a posloupnost $\{\omega_n\}$ buď definovaná jako v minulé sekci. Definujme množinu

$$\Psi_t = \{(u_0, \ldots, u_{t-1}) \,|\, s_0 \in S\},$$

a dívejme se na ni jako na podmnožinu jednotkové t-dimenzionální krychle. Ψ_t budeme brát jako multi-množinu, tedy připouštíme násobnost jednotlivých prvků. Volba stavu s_0 náhodně z množiny S odpovídá náhodné volbě vektoru $(u_0, \ldots, u_{t-1})$ z množiny Ψ_t ($|\Psi_t| = |S|$). Předpokládejme navíc, že



Obrázek 2.1: Množiny Ψ_2 LCG generátorů s maximální periodou definovaných parametry (i) a=1365, c=1, m=2048 (ii) a=65, c=1, m=2048

pro každé s_0 je posloupnost $\{u_n\}$ čistě periodická. Nechť $i \in \mathbb{N}$ je libovolné, zvolímeli opět $s_0 \in S$ náhodně, potom na i-tém místě posloupnosti $\{\omega_n\}$ získané z počátečního stavu s_0 můžeme očekávat libovolný prvek z množiny Ψ_t se stejnou pravděpodobností.

Z tohoto důvodu budeme chtít, aby množina Ψ_t byla co nejrovnoměrněji rozdělena v t-dimenzionální jednotkové krychli. Měřítko, které nyní zavedeme, je daleko snáze nahlédnutelné než to vycházející z Fourierových koeficientů.

Pro představu uvažujme nejdřív případ $t = 2$. Mějme množinu $\mathcal{H}$ rovnoběžných přímek, které pokrývají množinu Ψ_2 a označme v jako největší vzdálenost dvou sousedních přímek této množiny. Ta nám ukazuje na největší volný pás nepokrytý množinou Ψ_2. Tedy čím je tato hodnota menší tím lépe. Zadefinujme $1/\nu_2$ jako maximum ze všech hodnot v, které odvodíme ze všech různých množin $\mathcal{H}$ rovnoběžných přímek pokrývajících množinu Ψ_2. Hodnotu $1/\nu_2$ se budeme snažit minimalizovat a hodnotu ν_2 tedy naopak maximalizovat.

Na obrázku 2.1 vidíme porovnání množin Ψ_2 pro dva LCG generátory s maximální periodou. Na prvním obrázku je zakresleno všech 2048 bodů, které leží pouze na třech přímkách. Pokud zvolíme množinu $\mathcal{H}$ rovnu právě těmto třem přímkám, dostaneme hodnotu v, pro kterou se nabývá maxima v definici $1/\nu_2$. Na druhém obrázku je zakreslena množina Ψ_2, která má

dobrou strukturu vzhledem k našemu měřítku.

Obecně $1/\nu_t$ definujeme jako maximální vzdálenost nadrovin, branou přes všechny množiny rovnoběžných $(t-1)$-dimenzionálních nadrovin pokrývajících všechny body množiny Ψ_t.

Nyní si odvodíme fakt, který jsme na začátku uvedli a který můžeme názorně pozorovat na obrázku 2.1(ii). To jest, že množina Ψ_t má pro lineární kongruenční generátory definované vztahem 2.1, mřížovou strukturu. Pro zjenodušení budeme předpokládat, že $a = 0$.

Nechť e_i je i-tý k-dimenzionální jednotkový vektor. Označme $\{x_{n,i}\}$ posloupnost danou vztahem 2.1, pokud počáteční vektor s_0 je roven e_i. Je-li $s_0 = (z_1, \ldots, z_k) \in S$ libovolný, tak $s_0 = z_1e_1 + \ldots + z_ke_k$ a indukcí dostáváme:

$$\begin{aligned} x_{i+k} &= a_0x_i + \ldots + a_{k-1}x_{i+k-1} \bmod m = \\ &= a_0(z_1x_{i,1} + \ldots + z_kx_{i,k}) + \\ &\quad + \ldots + a_{k-1}(z_1x_{i+k-1,1} + \ldots + z_kx_{i+k-1,k}) \bmod m = \\ &= z_1x_{i+k,1} + \ldots + z_kx_{i+k,k} \bmod m \end{aligned}$$

a tedy

$$\{x_n\} = z_1\{x_{n,1}\} + \ldots + z_k\{x_{n,k}\} \bmod m.$$

Zároveň pro každou takovou lineární kombinaci existuje $s_0 = (z_1, \ldots, z_k)$ takový, že posloupnost $\{x_n\}$ je jím definována. Proto platí, že $\Psi_t = L_t \bigcap [0,1)^t$, kde L_t je mříž definovaná následujícími bázovými vektory:

$$\begin{aligned} v_1 &= (1, 0, \ldots, 0, x_{k,1}, \ldots, x_{t-1,1})/m \\ v_2 &= (0, 1, \ldots, 0, x_{k,2}, \ldots, x_{t-1,2})/m \\ &\vdots \\ v_k &= (0, 0, \ldots, 1, x_{k,k}, \ldots, x_{t-1,k})/m \\ v_{k+1} &= (0, 0, \ldots, 0, 1, \ldots, 0) \\ &\vdots \\ v_t &= (0, 0, \ldots, 0, 0, \ldots, 1) \end{aligned}$$

Je-li $t \leq k$, pak L_t je množina všech vektorů, jejichž souřadnice jsou násobky $1/m$ a $|\Psi_t| = m^t$. Zajímavější případ nastává pro $t > k$, pak $|\Psi_t| = m^k$.

Nyní ztotožníme Ψ_t s jejím periodickým rozšířením do prostoru, tedy s množinou L_t. Fakt, že L_t má mřížovou strukturu, nám umožňuje explicitně

popsat všechny množiny $\mathcal{H}$ rovnoběžných nadrovin pokrývajících právě L_t, tedy takových, že je-li P nadrovina patřící do $\mathcal{H}$, pak průnik P a L_t je nenulový a zároveň $\mathcal{H}$ pokrývá celou množinu L_t. Každá množina $\mathcal{G}$ rovnoběžných nadrovin pokrývající L_t obsahuje nějakou $\mathcal{H}$ pokrývající právě L_t, a tedy můžeme bez újmy na obecnosti předpokládat jen tento případ. Dále předpokládejme, že nulový vektor náleží do L_t. Pokud by tomu tak nebylo, můžeme mříž posunout a vhodný vektor, čímž se vlastnosti, které zkoumáme nezmění.

Nechť $u \in \mathbb{R}^t$ je libovolný vektor kolmý na všechny nadroviny množiny $\mathcal{H}$. Nechť $P \in \mathcal{H}$ je jedna konkrétní z nich a nechť $x \in P$, pak $x = y + \lambda u$, kde $\langle y, u\rangle = 0$ a $\|\lambda u\|$ je vzdálenost nadroviny P od počátku. Potom platí, že

$$\langle u, x\rangle = \langle u, y + \lambda u\rangle = \langle u, y\rangle + \langle u, \lambda u\rangle = \lambda\|u\|^2.$$

Platíli naopak, že $\langle u, x\rangle = \lambda\|u\|^2$, tak stejným postupem dostaneme, že $x \in P$ a tedy můžeme zformulovat následující lemma.

Lemma 23. *Nechť $P \in \mathcal{H}$ neprocházející počátkem a nechť λu je vektor kolmý na P, jehož délka se rovná vzdálenosti P od počátku, potom*

$$x \in P \iff \langle u, x\rangle = \lambda\|u\|^2$$

Existují tedy $0 = \lambda_0 < \lambda_1 < \lambda_2 < \lambda_3 \ldots \in \mathbb{R}$ taková, že

$$\mathcal{H} = \{x \in \mathbb{R}^t \mid \langle x, u\rangle = \lambda_i\|u\|^2\}.$$

Nyní využijeme vlastnost mřížové struktury L_t. Nechť $\mathcal{H}$ pokrývá právě L_t, pak všechny sousední nadroviny $P \in \mathcal{H}$ jsou stejně vzdálené a tedy platí, že $\lambda_i = i\lambda_1$. Proto existuje vektor $u_0 = \frac{1}{\lambda_1\|u\|^2}u$ (kde $\lambda_1\|u\|$ je vzdálenost počátku od nejbližší nadroviny $P \in \mathcal{H}$ neprotínající počátek) takový, že

$$\mathcal{H} = \{x \in \mathbb{R}^t \mid \langle x, u_0\rangle \in \mathbb{Z}\}.$$

Definice 24. *Nechť vektory $v_1, \ldots, v_t \in \mathbb{R}^t$ tvoří bázi mříže L_t v t-dimenzionálním prostoru. Potom vektory $v_1^*, \ldots, v_t^*$ definujme vztahy $(v_i, v_j^*) = \delta_{i,j}$ pro všechna $i, j \in \{1, \ldots, t\}$. Mříž tvořenou bází $v_1^*, \ldots, v_t^*$ nazveme duální k mříži L_t a značíme L_t^*.*

Tvrzení 25. *Nechť $\mathcal{H} = \{x \in \mathbb{R}^t \mid \langle x, u\rangle \in \mathbb{Z}\}$, pak*

1. $\mathcal{H} \supset L_t \iff u \in L_t^*$

2. ν_t je rovno délce nejkratšího nenulového vektoru mříže L_t^*.

Důkaz. 1. Jelikož $v_1^*, \ldots, v_t^*$ tvoří bázi prostoru $\mathbb{R}^t$, tak existují $w_i \in \mathbb{R}$ taková, že $u = w_1 v_1^* + \ldots + w_t v_t^*$.

$$\begin{aligned} \langle v_i, u \rangle &= \langle v_i, w_1 v_1^* + \ldots + w_t v_t^* \rangle = \\ &= w_1 \langle v_i, v_1^* \rangle + \ldots + w_t \langle v_i, v_t^* \rangle = w_i \end{aligned}$$

Protože $\mathcal{H} \supset L_t$, tak $v_i \in \mathcal{H}$ a tedy $w_i \in \mathbb{Z}$ pro všechna $i \in \{1, \ldots, t\}$ a tedy $u \in L_t^*$.

Naopak jeli $u \in L_t^*$, pak stejným postupem dostaneme, že $\langle v_i, u \rangle \in \mathbb{Z} \Longrightarrow v_i \in \mathcal{H}$ a tedy $L_t \subset \mathcal{H}$.

2. Mějme $\mathcal{H}$ jako výše, potom vzdálenost mezi sousedními nadrovinami množiny $\mathcal{H}$ je rovna vzdálenosti nadroviny definované vztahem $\langle u, x \rangle = 1$ od počátku, nebo-li

$$\min\{\|x\| \,|\, \langle x, u \rangle = 1\}.$$

Využitím Cauchyho nerovnosti $1 = \langle x, u \rangle \leq \|x\| \, \|u\|$ dostáváme, že minimum se nabývá pro $x = (x_1, \ldots, x_t)$, kde

$$x_i = \frac{u_j}{\|u\|^2}$$

a tedy $\min\{\|x\| \,|\, \langle x, u \rangle = 1\} = 1/\|u\|$. Dostáváme tedy

$$\nu_t = \min\{\|u\| \,|\, \mathcal{H} \supset L_t\} = \min\{\|u\| \,|\, u \in L_t^*\}.$$

□

Jednoduchý odhad na hodnotu ν_t dostáváme z toho, že pro $t > k$ vektor $u = (0, \ldots, 0, a_1, \ldots, a_k, -1,)$ určuje množinu $\mathcal{H}$, která pokrývá L_t, neboť $\langle u, v \rangle \in \mathbb{Z}$ pro všechna $v \in \Psi_t$. A tedy

$$\nu_t \leq 1 + a_1^2 + \ldots + a_k^2.$$

Pokud bychom stejně jako v příkladu z minulé sekce uvažovali LCG generátor s parametry (x_0, a, c, m), který dosahuje maximální periody, dostali bychom stejný výsledek, viz kongruence 2.13. Odvození a algoritmus na výpočet lze nalézt v [3] str. 96-101.

$$\nu_t^2 = \min\{u_1^2 + \ldots + u_t^2 \,|\, (u_1 + a u_2 + \ldots + a^{t-1} u_t \equiv 0 \bmod m) \wedge u_i \in \mathbb{Z}\}.$$

Příklad (konkrétní generátor RANDU)
Jedním z generátorů hojně využívaných v minulosti byl generátor RANDU definovaný vztahem:

$$x_{n+1} = (2^{16} + 3)x_n \bmod 2^{31}.$$

Proto abychom ukázali jeho slabiny, předpokládejme trochu obecnější případ. Mějme lineární kongruenční generátor daný parametry $c = 0$, $m = 2^e$, $a = 2^l + 3$ pro $l \geq e/2$. Pak platí, že

$$\begin{aligned} 9x_n - 6x_{n+1} + x_{n+2} &= 9x_n - 6(2^l + 3)x_n + (2^l + 3)^2 x_n = \\ &= 2^{2l} x_n \equiv 0 \bmod 2^e \end{aligned}$$

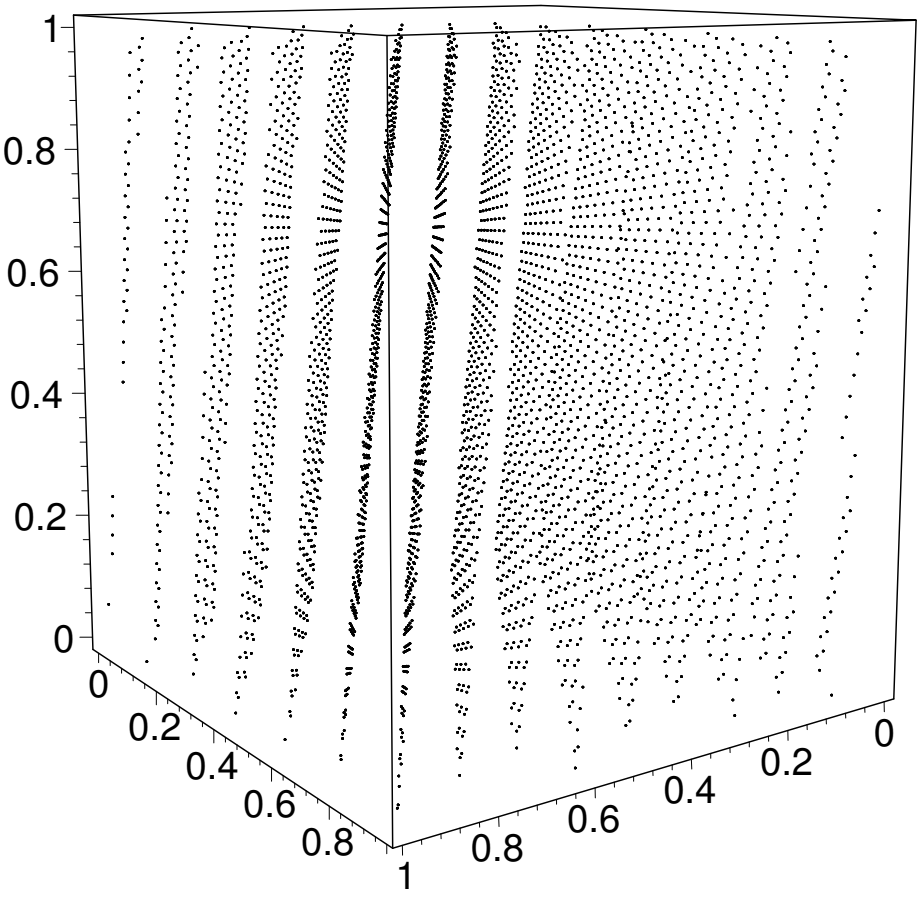
Protože

$$-6 \cdot 2^e < 9x_n - 6x_{n+1} + x_{n+2} < 10 \cdot 2^e,$$

tak body posloupnosti $\{\omega_n\}$ můžeme v jednotkové krychli pokrýt pouhými 15 rovnoběžnými rovinami. Pokud bychom na tento generátor aplikovali spektrální test, tak bychom získali:

$$\nu_3^2 = 9^2 + 6^2 + 1^2 = 118,$$

což je hodnota extrémně nízká v porovnání s lineárními kongruenčními generátory používanými v současnosti (viz. [3] str. 106).



Obrázek 2.2: Body (x_n, x_{n+1}, x_{n+2}) posloupnosti vygenerované generátorem RANDU

Kapitola 3

Závěr

V úvodní kapitole jsme nastínili rozdíl mezi konceptem pseudonáhodnosti a náhodnosti. Uvedli jsme definici pseudonáhodného generátoru, naznačili otázku jeho existence a možnost získat opravdu pseudonáhodné generátory za předpokladu platnosti hypotézy o jednosměrnosti dané funkce.

V hlavní části práce jsme se zabývali tím, jak posuzovat posloupnosti vygenerované pseudonáhodnými generátory, statistickými testy. Testové statistiky můžeme počítat buďto přímo z vygenerované posloupnosti - tento způsob nazýváme empirickým testem - nebo je v některých případech dokážeme matematicky odvodit přímo z parametrů generátoru a takto získat nástroj na posuzování náhodnosti různých generátorů stejné třídy vzhledem k danému testu. Soustředili jsme se hlavně na třídu lineárních kongruenčních generátorů, pro kterou jsme nastínili některé důležité vlastnosti, jako je například charakterizace parametrů, pro které vygenerovaná posloupnost dosahuje maximální periody. Dále jsme se věnovali jednomu z nejdůležitějších teoretických testů pro tuto třídu generátorů - to jest spektrálnímu testu. Představili jsme jeho dvě verze, nejdříve odvození pomocí diskrétní Fourierovy transformace a poté geometrické odvození využívající toho, že vektory lineárních kongruenčních generátorů tvoří mřížovou strukturu.

Literatura

[1] R. R. Coveyou, R. D. MacPherson: *Fourier analysis of random number generators*, Journal of the Association for Computing Machinery, Vol. 14, No. 1, January 1967. pp. 100-11.

[2] O. Goldreich:*Foundations of Cryptography*, Cambridge University Press 2001.

[3] D. E. Knuth: *The art of computer programming, volume 2.*

[4] Ming Li, Paul Vitányi: *An introduction to Kolmogorov complexity and its applications*, 2nd ed.. - New York : Springer, 1997

[5] Karel Zvára, Josef Štěpán: *Pravděpodobnost a matematická statistika*, Matfyzpress, Praha, 2002.