Algorithms of Information Security: PRNG cvičení

Olha Jurečková, Martin Jureček
{jurecolh,jurecmar}@fit.cvut.cz

Faculty of Information Technology
Czech Technical University in Prague

September 23, 2020

Golombovi postuláty náhodnosti

Pro posouzení náhodnosti požadovaných pseudonáhodných sekvencí existují Golombovi pseudohonáhodné postuláty.

Definition

Nechť $s = s_0, s_1, s_2, \ldots$ je nekonečná posloupnost. Podposloupnost, která obsahuje prvních n členů posloupnosti s značíme $s^n = s_0, s_1, s_2, \ldots, s_{n-1}$.

Příklad. Nechť $s = 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$ je nekonečná posloupnost, potom $s^1 = 0$, $s^2 = 0, 1$ atd.

Definition

Řekneme, že posloupnost $s = s_0, s_1, s_2, \ldots$ je *N-periodická,* pokud $s_i = s_{i+N}$ pro všechna $i \geq 0$. Posloupnost s se nazývá *periodická,* pokud je N-periodická pro nějaké kladné celé N. *Periodou* posloupnosti s, je nejmenší kladné celé N, pro které je posloupnost s N-periodická. Pokud s je periodická posloupnost s periodou N, potom *cyklem* posloupnosti s je podposlounost s^N.

Příklad. Nechť $s = 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$ Potom $s_i = s_{i+3} = 0, 1, 1$. Tudíž, s je periodická posloupnost s periodou 3 a cyklus posloupnosti s je podposloupnost $s^3 = 0, 1, 1$.

Definition

Nechť s je posloupnost. *Run* posloupnosti s je podposloupnost s obsahující po sobě jdoucí (v řádě za sebou jdoucí) 0 nebo po sobě jdoucí 1, kterým ani nepředchází stejný symbol ani nenasleduje stejný symbol. Run 0 se nazývá *mezera* a run 1 se nazývá *blok*.

Příklad. Nechť $s = 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$ je posloupnost. Potom 0 je run délky 1 a nazývá se mezera, a 1,1 je run délky 2 a nazývá se blok.

Definition

Nechť $s = s_0, s_1, s_2, \ldots$ je periodická posloupnost s periodou N. Potom *autokorelační funkce* posloupnosti s je funkce $C(t)$ definovana následujícím způsobem:

$$C(t) = \frac{1}{N} \sum_{i=0}^{N-1} (2s_i - 1)(2s_{i+t} - 1), \text{for } 0 \leq t \leq N - 1.$$

Autokorelační funkce $C(t)$ měří množství (počet) podobnosti mezi posloupnosti s a posunem posloupnosti s o t pozic.

*Příklad.*Nechť $s = 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$ je posloupnost. Najděte:

- periodu a cyklus dané posloupnosti
- mezeru a blok
- $C(0), C(1), C(2)$ a $C(3)$.

Řešení. 1. $s = 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$ je daná posloupnost, potom $s_i = s_{i+3} = 0, 1, 1$. Tudíž, s je periodická posloupnost s periodou 3 a cyklus posloupnosti s je podposloupnost $s^3 = 0, 1, 1$.
2.Nechť $s = 0, 1, 1, 0, 1, 1, 0, 1, 1, \ldots$ je daná posloupnost, potom 0 je run délky 1 a nazývá se mezera, a 1,1 je run délky 2 a nazývá se blok.

3.Potom perioda N je 3, tj. $N = 3$.

$$C(t) = \frac{1}{N}\sum_{i=0}^{N-1}(2s_i - 1)(2s_{i+t} - 1), \text{for } 0 \leq t \leq N-1.$$

Potom

$$C(0) = \frac{1}{3}\sum_{i=0}^{2}(2s_i - 1)(2s_{i+0} - 1) =$$

$$= \frac{1}{3}((2s_0 - 1)(2s_0 - 1) + (2s_1 - 1)(2s_1 - 1) + (2s_2 - 1)(2s_2 - 1)).$$

$$C(0) = \frac{1}{3}((2\cdot 0-1)(2\cdot 0-1)+(2\cdot 1-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 1-1)) =$$

$$= \frac{1}{3}(1+1+1) = 1.$$

$$C(1) = \frac{1}{3}((2\cdot 0-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 0-1)) =$$

$$= \frac{1}{3}(-1+1-1) = \frac{1}{3}.$$

$$C(2) = \frac{1}{3}((2\cdot 0-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 0-1)+(2\cdot 1-1)(2\cdot 1-1)) =$$

$$= \frac{1}{3}(-1-1+1) = -\frac{1}{3}.$$

$$C(3) = \frac{1}{3}((2\cdot 0-1)(2\cdot 0-1)+(2\cdot 1-1)(2\cdot 1-1)+(2\cdot 1-1)(2\cdot 1-1)) =$$

$$= \frac{1}{3}(1+1+1) = 1.$$

Definition

Nechť s je periodická posloupnost s periodou N. *Golombovi postuláty náhodnosti* jsou následující:

- R1: V cyklu s^N posloupnosti s, počet 1 liší se od počtu 0 maximálně o 1.
- R2: V cyklu s^N, aspoň $\frac{1}{2}$ runů má délku 1, aspoň $\frac{1}{4}$ má délku 2, aspoň $\frac{1}{8}$ má délku 3, etc., tak dlouho, pokud počet runů dané délky je alespoň 1. Navíc, pro každou z těchto délek existuje (teměř) stejně mnoho mezer a bloků.
- R3: Autokorelační fukce $C(t)$ má 2 hodnoty. Pro nějaké celé K,

$$N \cdot C(t) = \sum_{i=0}^{N-1} (2s_i - 1)(2s_{i+t} - 1) = \begin{cases} N, & \text{if } t = 0, \\ K, & \text{if } 1 \leq t \leq N-1 \end{cases}$$

Příklad. Nechť s je periodická posloupnost s periodou $N = 15$ a cyklem

$$s^{15} = 0, 1, 1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1.$$

Ověřte Golombovi postuláty náhodnosti.

Řešení.

- R1: Počet 0 je 7 a počet 1 je 8. Rozdil mezi počtem 0 a 1 je 8-7=1. Tudiž R1 je splněna.
- R2: s^{15} má 8 runů. Máme 4 runy délky 1 (2 mezery a 2 bloky): s_0, s_5, s_{13}, s_{14}. Počet mezer je 2 (s_0, s_{14}) a počet bloků je taky 2 (s_5, s_{13}).
  Počet runů délky 2 je 2 $(s_1 s_2, s_3 s_4)$. Počet mezer je 1 $(s_3 s_4)$ a počet bloků je taky 1 $(s_1 s_2)$.
  Počet runů délky 3 je 1 $(s_6 s_7 s_8)$. Počet mezer je 1 $(s_6 s_7 s_8)$ a počet bloků je 0.
  Počet runů délky 4 je 1 $(s_9 s_{10} s_{11} s_{12})$. Počet mezer je 0 a počet bloků je $(s_9 s_{10} s_{11} s_{12})$.
  Celkový počet runů je 4+2+1+1=8. $\frac{1}{2}$ runů je délky 1, $\frac{1}{4}$ runů je délky 2 a $\frac{1}{8}$ runů je délky 3. Potom R2 je splněno.

- R3: Spočteme $C(0)$.

$$C(0) = \frac{1}{15}\sum_{i=0}^{14}(2s_i - 1)(2s_{i+0} - 1) =$$

$$= \frac{1}{15}((2s_0 - 1)(2s_0 - 1) + (2s_1 - 1)(2s_1 - 1) + (2s_2 - 1)(2s_2 - 1) + \ldots +$$
$$+(2s_{13} - 1)(2s_{13} - 1) + (2s_{14} - 1)(2s_{14} - 1)) =$$
$$= \frac{1}{15}((2 \cdot 0 - 1)(2 \cdot 0 - 1) + (2 \cdot 1 - 1)(2 \cdot 1 - 1) +$$
$$+(2 \cdot 1 - 1)(2 \cdot 1 - 1) + (2 \cdot 0 - 1)(2 \cdot 0 - 1) +$$
$$+(2 \cdot 0 - 1)(2 \cdot 0 - 1) + (2 \cdot 1 - 1)(2 \cdot 1 - 1) +$$
$$+(2 \cdot 0 - 1)(2 \cdot 0 - 1) + (2 \cdot 0 - 1)(2 \cdot 0 - 1) +$$
$$+(2 \cdot 0 - 1)(2 \cdot 0 - 1) + (2 \cdot 1 - 1)(2 \cdot 1 - 1) +$$
$$+(2 \cdot 1 - 1)(2 \cdot 1 - 1) + (2 \cdot 1 - 1)(2 \cdot 1 - 1) +$$
$$+(2 \cdot 1 - 1)(2 \cdot 1 - 1) + (2 \cdot 0 - 1)(2 \cdot 0 - 1) +$$
$$+(2 \cdot 1 - 1)(2 \cdot 1 - 1)) =$$

$$= \frac{1}{15}(1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1) = 1.$$

$$C(0) = \frac{1}{15}(1+1+1+1+1+1+1+1+1+1+1+1+1+1+1) = 1.$$

$$C(1) = \frac{1}{15}(-1+1-1+1-1-1+1+1-1+1+1+1-1-1-1) = -\frac{1}{15}.$$

$$C(1) = C(2) = \ldots = C(14) = -\frac{1}{15}$$

$$15.C(t) = \begin{cases} 15, & t = 0 \text{ nebo } t = 15 \\ -1, & \text{pro } 1 \leq t \leq 14 \end{cases}$$

Potom R3 je splněno.

χ^2 test dobré shody

Test dobré shody testuje shodu ***empirických četností*** (skutečné četnosti) $X_1, \ldots, X_n$ jevů $A_1, \ldots, A_n$ se středními hodnotami těchto četností (tzv. ***očekávané četnosti***) $mp_1, \ldots, mp_n$, kde pravděpodobnosti $p_1, \ldots, p_n$ jsou určeny z platnosti nějakého pravděpodobnostního modelu.
Nulová hypotéza říká, že pravděpodobnosti jevů $A_1, \ldots, A_n$ jsou po řadě rovny $p_1, \ldots, p_n$ a testová statistika má tvar:

$$X^2 = \sum_{i=1}^{n} \frac{(X_i - mp_i)^2}{mp_i}.$$

Náhodná veličina X^2 má přibližně χ^2-rozdělení o $n-1$ stupních volnosti. Nulovou hypotézu zamítáme na hladině významnosti α, jestliže $X^2 > \chi^2_{1-\alpha;\, n-1}$, kde hodnota $\chi^2_{1-\alpha;\, n-1}$ je kvantil χ^2-rozdělení o $n-1$ stupních volnosti. Z toho potom můžeme odvodit, že naše statistika nemá χ^2-rozdělení a pravděpodobnosti jevů jsou různé od pravděpodobností $p_1, \ldots, p_n$.
Poznamenejme, že χ^2 test dobré shody je asymptotický, a proto ho možno doporučit jen při dostatečně velkém rozsahu výběru m. V literatuře se obvykle uvádí, že musí platit $mp_i \geq 5$ pro každé $i = 1, \ldots, n$.

Chi-Square Goodness-of-Fit Test

Příklad. Máme hrací kostku a chceme ověřit, jestli je kostka pravidelná. Hodíme kostkou 48krát a získáme následující četnosti hodů:

Hodnota	1	2	3	4	5	6
Četnost	10	6	14	2	4	12

Zjistěte, zda daná kostka je homogenní?

Řešení. Nejprve si definujme nulovou a alternativní hypotézu:

H_0: kostka je pravidelná

H_A: kostka není pravidelná

Jestli je kostka pravidelná, potom pravděpodobnost každé hodnoty je $\frac{1}{6}$. Tedy očekávané četnosti jednotlivých hodnot jsou stejné a rovné 8. Naměřené četnosti jsou $(X_1, X_2, X_3, X_4, X_5, X_6) = (10, 6, 14, 2, 4, 12)$. Dále aplikujeme test dobré shody a spočteme hodnotu:

$$X^2 = \sum_{i=1}^{n} \frac{(X_i - mp_i)^2}{mp_i} = \sum_{i=1}^{6} \frac{(X_i - 8)^2}{8} =$$

$$= \frac{(10-8)^2}{8} + \frac{(6-8)^2}{8} + \frac{(14-8)^2}{8} + \frac{(2-8)^2}{8} + \frac{(4-8)^2}{8} + \frac{(12-8)^2}{8} =$$

$$= 14.$$

Pracujeme na hladině významnosti $\alpha = 0,05$ a uvažujeme $n - 1 = 5$ stupňů volnosti. Kvantil $\chi^2_{1-\alpha;\, n-1}$ najdeme v příslušné statistické tabulce a je rovný $\chi^2_{0,95;\, 5} = 11,071$. Máme $X^2 > 11,071$, proto nulovou hypotézu H_0 můžeme zamítnout na dané hladině významnosti.

Příklad. Pomocí χ^2 testu dobré shody odovďte statistiku pro Serial test.

Řešení. Předpokládejme, že máme posloupnost bitů délky n.
Nejprve si označme:

- n_0 počet 0 v dané posloupnosti
- n_1 počet 1 v dané posloupnosti
- n_{00} počet podposloupností 00 v dané posloupnosti
- n_{01} počet podposloupností 01 v dané posloupnosti
- n_{10} počet podposloupností 10 v dané posloupnosti
- n_{11} počet podposloupností 11 v dané posloupnosti

Chceme odovodit následující statistiku

$$X^2 = \frac{4}{(n-1)}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_{00}^2 + n_{11}^2) + 1$$

Potřebujeme zjistit jaké jsou očekávané počty podposloupnosti v posloupnosti délky n.

- $\frac{n}{2}$ je očekávaný počet 0 v dané posloupnosti
- $\frac{n}{2}$ je očekávaný počet 1 v dané posloupnosti
- $\frac{n-1}{4}$ je očekávaný počet podposloupností 00 v dané posloupnosti
- $\frac{n-1}{4}$ je očekávaný počet podposloupností 01 v dané posloupnosti
- $\frac{n-1}{4}$ je očekávaný počet podposloupností 10 v dané posloupnosti
- $\frac{n-1}{4}$ je očekávaný počet podposloupností 11 v dané posloupnosti

Potom

$$X^2 = \frac{(n_0 - \frac{n}{2})^2}{\frac{n}{2}} + \frac{(n_1 - \frac{n}{2})^2}{\frac{n}{2}} +$$
$$+ \frac{(n_{00} - \frac{n-1}{4})^2}{\frac{n-1}{4}} + \frac{(n_{01} - \frac{n-1}{4})^2}{\frac{n-1}{4}} +$$
$$+ \frac{(n_{10} - \frac{n-1}{4})^2}{\frac{n-1}{4}} + \frac{(n_{11} - \frac{n-1}{4})^2}{\frac{n-1}{4}}$$

Poznamenejme, že platí:

$$n_{00} + n_{01} + n_{10} + n_{11} = (n - 1).$$

Potom

$$X^2 = \frac{4}{n-1}\left(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2 - \right.$$

$$\left. -2n_{00}\frac{n-1}{4} - \ldots - 2n_{11}\frac{n-1}{4} + 4(\frac{n-1}{4})^2\right) +$$

$$+\frac{2}{n}(n_0^2 - n_0 n + \frac{n^2}{4} + n_1^2 - n_1 n + \frac{n^2}{4})$$

Poznamenejme, že platí:

$$n_1 = n - n_0.$$

Nakonec získáváme

$$X = \frac{4}{(n-1)}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_{00}^2 + n_{11}^2) + 1.$$

Příklad. Nechť je dana posloupnost s délky $n = 160$, která obsahuje 4krát opakující se následující podposloupnost:

$$1110001100010001010011101111001001001001.$$

Aplikujte frekvenční(monobit) a serial test.

Řešení.

Frekvenční test.

Definujme si nulovou hypotézu H_0 a alternativní hypotézu H_A:

- H_0 : počet jedniček se rovná $\frac{n}{2}$
- H_A : počet jedniček se nerovná $\frac{n}{2}$.

Použijeme následující statistiku:

$$X^2 = \frac{(n_0 - \frac{n}{2})^2}{\frac{n}{2}} + \frac{(n_1 - \frac{n}{2})^2}{\frac{n}{2}} = \frac{(n_0 - n_1)^2}{n}.$$

Potřebujeme spočitat počet nul a jedniček v dané posloupnosti. $n_0 = 84$ a $n_1 = 76$. Potom

$$X^2 = \frac{(84 - 80)^2}{80} + \frac{(76 - 80)^2}{80} = 0,4.$$

Pracujeme na hladině významnosti $\alpha = 0,05$ a uvažujeme 1 stupeň volnosti. Kvantil $\chi^2_{0,95;\ 1}$ najdeme v příslušné statistické tabulce a je rovný $\chi^2_{0,95;\ 1} = 3,8415$. Máme $X^2 < 3,8415$, proto testována posloupnost prošla Frekvenčním monobit testem.

Serial test.

Nápověda. Máme $n_0 = 84$ a $n_1 = 76$. Dále $n_{00} = 44, n_{01} = 40, n_{10} = 40$ a $n_{11} = 35$. Spočteme X^2 a dostaneme $X^2 = 0,6252$.

Pracujeme na hladině významnosti $\alpha = 0,05$ a uvažujeme 2 stupně volnosti. Kvantil $\chi^2_{0,95;\ 2} = 5,9915$.

Daná posloupnost prošla i Serial testem.

Příklad.

Nechť $s = 1, 0, 1, 1, 0, 1, 1, 0, 1, \ldots$ je posloupnost. Najděte:

- periodu a cyklus dané posloupnosti
- $C(0), C(1), C(2)$ a $C(3)$.

Příklad.

Máme hrací kostku a chceme ověřit, jestli je kostka pravidelná. Hodíme kostkou 300krát a získáme následující četnosti hodù:

Hodnota	1	2	3	4	5	6
Četnost	40	55	51	49	46	59

Zjistěte, zda daná kostka je homogenní?

Příklad.

Nechť máme nějaký generátor pseudonáhodných bitů, který vyprodukoval následující posloupnost $x = (1110100111101100)$. Otestujte tento generátor pomocí Frekvenčního monobit testu.

Linear congruential(LCG) PRBG

Algorithm 1 Algorithm LCG PRBG

Input: $N-$length generated sequence

Output: $x_1, x_2, \ldots, x_N \in \mathbb{Z}_2$ a pseudorandom bit sequence

1: select a random integer x_0 (the seed) and select parameters a, c, m, where a, c are in the interval $[1, m]$
2: **for** $i = 1$ **to** l **do**
3: $\quad x_i = ax_{i-1} + c \bmod m$
4: **end for**
5: **return** the output sequence $x_1, x_2, \ldots, x_l$

LCG. Naprogramujte LCG s následujícími parametry:

- $x_0 = 20170705$ je semínko angl.seed
- $a = 742938285$ činitel
- $e = 31$
- $m = 2^e - 1$ modulus

Pomocí daného LCG vygenerujte posloupnost délky $N = 100000$. Dále naprogramujte Frekvenční (monobit) test a ověřte pomocí něj vygenerovanou posloupnost.