Zufällige Abbildungen und Faktorisieren via Pollard-Rho

Felix Potthast

02.03.2016

Gliederung

Pseudozufallszahlen

- Arithmetische, deterministische Methoden zur Generierung zufällig wirkender Folgen
- Verwendung z.B. für Monte-Carlo Verfahren oder Kryptographie
- Übliche Form: $x_{i+1} = f(x_i)$ mit Seed x_0, $x_i \in \{0, \cdots, m-1\} = \mathbb{Z}/m\mathbb{Z}$
- Gewünschte Eigenschaften: lange Periode, Gleichverteilung

Periodische Folgen

Definition: Periodische Folge

Sei $(x_i)_{i \geq 0}$ eine Folge. x_i ist periodisch mit der Periodenlänge λ und der Vorperiodenlänge μ (und $\rho = \mu + \lambda$), wenn gilt:
$x_0, x_1, \ldots, x_\mu, \ldots, x_{\mu+\lambda-1}$ sind paarweise verschieden, für $n \geq \mu$ gilt aber: $x_{n+\lambda} = x_n$

Sei $f : \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ beliebig.
Dann ist die Folge $(x_i)_{i \geq 0}$, $x_0 \in \mathbb{Z}/m\mathbb{Z}$, $x_{i+1} = f(x_i)$ periodisch mit $1 \leq \lambda \leq m$ und $0 \leq \mu$ sowie $\mu + \lambda = \rho \leq m$.

Zufällige Abbildungen

Es gibt m^m Abbildungen in obiger Form. Sei f nun eine daraus zufällig gewählte Abbildung. Dann ist die Wahrscheinlichkeit die gewünschte maximale Periodenlänge $\lambda = m$ zu erhalten gegeben durch: $\frac{(m-1)!}{m^m}$
(Erläuterung: $m!$ ist die Anzahl der Permutationen, der Zyklus bleibt aber gleich für m Verschiebungen)
Eine Approximation der Fakultät mit der Stirlingformel ergibt dann:

$$\frac{(m-1)!}{m^m} = \frac{m!}{m^{(m+1)}} \approx \frac{\sqrt{2\pi m}\left(\frac{m}{e}\right)^m}{m^{(m+1)}} = \sqrt{\frac{2\pi}{m}} \cdot \left(\frac{1}{e}\right)^m$$

Zufällige Abbildungen

Die Wahrscheinlichkeit, dass eine zufällig gewählte Abbildung f einen Zyklus der Länge 1 hat, ist gegeben durch:

$$1 - \left(\frac{m-1}{m}\right)^m = 1 - \left(1 - \frac{1}{m}\right)^m > 1 - \frac{1}{e} \approx 0.63$$

Zufällige Abbildungen

Sei f wieder eine zufällig gewählte Abbildung mit nun ebenfalls zufällig gewähltem Seed x_0. Die Wahrscheinlichkeit, eine Folge mit Vorperiodenlänge μ und Periodenlänge λ zu erhalten, ist:

$$P_m(\mu, \lambda) = \frac{1}{m} \prod_{k=1}^{\mu+\lambda} \left(1 - \frac{k}{m}\right)$$

Erläuterung: Die ersten $\mu + \lambda$ Werte müssen paarweise verschieden sein, danach muss ein bestimmter Wert erreicht werden, um λ zu erhalten. Der Durchschnitt für $\mu + \lambda$ ist dann gegeben durch:

$$\sum_{\lambda=1}^{m} \sum_{\mu=0}^{m-\lambda} (\lambda + \mu) P_m(\mu, \lambda) = Q(m) \approx \sqrt{\frac{\pi m}{2}} \approx 1.253 \cdot \sqrt{m}$$

Zufällige Abbildungen

$Q(m)$ ist die Ramanujan Q-Funktion gegeben durch

$$Q(m) = \sum_{k=1}^{m} \frac{m!}{(m-k)!m^k}$$

Für $k = o\left(m^{\frac{2}{3}}\right)$ und $m \to \infty$ gilt:

$$\frac{m!}{(m-k)!m^k} = e^{\frac{-k^2}{2m}} \left(1 + O\left(\frac{k}{m}\right) + O\left(\frac{k^3}{m^2}\right)\right)$$

Für alle k, mit $m \to \infty$ gilt:

$$\frac{m!}{(m-k)!m^k} = e^{\frac{-k^2}{2m}} + O\left(\frac{1}{\sqrt{m}}\right)$$

Beweis:

$$\frac{m!}{(m-k)!m^k} = \frac{\prod_{j=1}^{k-1}(m-j)}{m^k} = \prod_{j=1}^{k-1}\left(1-\frac{j}{m}\right) = e^{\ln\left(\prod_{j=1}^{k-1}\left(1-\frac{j}{m}\right)\right)}$$

$$= e^{\sum_{j=1}^{k-1}\ln\left(1-\frac{j}{m}\right)}$$

Da $k = o(n)$ wenden wir die Taylor-Formel auf den Logarithmus an:
$\ln(1+x) = x + O(x^2)$ mit $x = \frac{-j}{m}$

$$e^{\sum_{j=1}^{k-1}\ln\left(1-\frac{j}{m}\right)} = e^{\sum_{j=1}^{k-1}\left(-\frac{j}{m}+O\left(\frac{j^2}{m^2}\right)\right)} = e^{\left(-\frac{k(k-1)}{2m}+O\left(\frac{k^3}{m^2}\right)\right)}$$

Zufällige Abbildungen

Dann können wir für den Fall $k = o\left(m^{\frac{2}{3}}\right)$ die Taylor-Formel für die Exponentialfunktion anwenden: $e^x = 1 + O(x)$

$$e^{\left(-\frac{k(k-1)}{2m} + O\left(\frac{k^3}{m^2}\right)\right)} = e^{\left(\frac{-k^2}{2m} + \frac{k}{2m} + O\left(\frac{k^3}{m^2}\right)\right)}$$

$$= e^{\left(\frac{-k^2}{2m}\right)} \cdot e^{\left(\frac{k}{2m} + O\left(\frac{k^3}{m^2}\right)\right)} = e^{\frac{-k^2}{2m}} \left(1 + O\left(\frac{k}{m}\right) + O\left(\frac{k^3}{m^2}\right)\right)$$

$\blacksquare$

Zufällige Abbildungen

Dann betrachten wir noch die Approximationsformel für alle k: Sei $k_0 = m^{\frac{3}{5}}$, dann gilt für $k \leq k_0$ die obige Approximationsformel und wir setzen $x = \frac{k}{\sqrt{2m}}$:

$$e^{\frac{-k^2}{2m}}\left(1 + O\left(\frac{k}{m}\right) + O\left(\frac{k^3}{m^2}\right)\right) = e^{\frac{-k^2}{2m}} + e^{-x^2}O\left(\frac{k}{m}\right) + e^{-x^2}O\left(\frac{k^3}{m^2}\right)$$

$$= e^{\frac{-k^2}{2m}} + e^{-x^2}O\left(\frac{x\sqrt{2m}}{m}\right) + e^{-x^2}O\left(\frac{\left(x\sqrt{2m}\right)^3}{m^2}\right)$$

$$= e^{\frac{-k^2}{2m}} + xe^{-x^2}O\left(\frac{1}{\sqrt{m}}\right) + x^3e^{-x^2}O\left(\frac{1}{\sqrt{m}}\right)$$

Es gilt: $xe^{-x^2} = O(1)$ und $x^3e^{-x^2} = O(1)$ und daher: $e^{\frac{-k^2}{2m}} + O\left(\frac{1}{\sqrt{m}}\right)$

Zufällige Abbildungen

Sei nun $k \geq k_0$, dann gilt:

$$\frac{m!}{(m-k_0)!m^{k_0}} \leq \frac{m!}{(m-k)!m^k}$$

und da $k_0 = o\left(m^{\frac{2}{3}}\right)$:

$$\frac{m!}{(m-k_0)!m^{k_0}} = e^{\frac{-k_0^2}{2m}} + O\left(\frac{1}{\sqrt{m}}\right) = e^{-\frac{\sqrt[5]{m}}{2}} + O\left(\frac{1}{\sqrt{m}}\right)$$
$$\Longrightarrow \frac{m!}{(m-k)!m^k} = O\left(\frac{1}{\sqrt{m}}\right)$$

Zufällige Abbildungen

Daraus lässt sich jetzt folgende Asymptotische Formel für $Q(m)$ herleiten:

$$Q(m) = \sum_{k=1}^{m} \frac{m!}{(m-k)!m^k} = \sqrt{\frac{\pi m}{2}} + O(1)$$

Wir benutzen für unterschiedliche Bereiche der Summe unterschiedliche Fehlerabschätzungen: sei $k_0 = o\left(m^{\frac{2}{3}}\right)$

$$Q(m) = \sum_{k=1}^{m} \frac{m!}{(m-k)!m^k} = \sum_{k=1}^{k_0} \frac{m!}{(m-k)!m^k} + \sum_{k=k_0+1}^{m} \frac{m!}{(m-k)!m^k}$$

$$= \sum_{k=1}^{k_0} e^{\frac{-k^2}{2m}} \left(1 + O\left(\frac{k}{m}\right) + O\left(\frac{k^3}{m^2}\right)\right) + \Delta$$

Zufällige Abbildungen

Wir betrachten dabei Δ als vernächlässigbar, da exponentiell klein. Wir nehmen auch an, dass alle weiteren Terme vernachlässigbar sind und betrachten nun: $\sum_{k=1}^{\infty} e^{\frac{-k^2}{2m}} + O(1)$. Nach der Euler-Maclaurin Formel gilt:

$$\sum_{i=m}^{n} f(i) = \int_m^n f(x)\, dx + \frac{f(n) + f(m)}{2}$$
$$+ \sum_{j=1}^{k} \frac{B_{2j}}{(2j)!} \left(f^{(2j-1)}(n) - f^{(2j-1)}(m) \right) + R_{2k}(m, n)$$

Daraus folgt dann:

$$Q(m) = \sum_{k=1}^{\infty} e^{\frac{-k^2}{2m}} + O(1) = \int_0^{\infty} e^{-\frac{k^2}{2m}} dk + O(1)$$

Dann substituieren wir $k = \varphi(x) = \frac{x}{\sqrt{m}}$:

$$\sqrt{m} \int_0^{\infty} e^{-\frac{\left(\frac{k}{\sqrt{m}}\right)^2}{2m}} \left(\frac{1}{\sqrt{m}}\right) dx + O(1) = \sqrt{m} \int_{\varphi(0)}^{\varphi(\infty)} e^{-\frac{x^2}{2}} dx + O(1)$$

wobei

$$\int_0^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{\frac{\pi}{2}}$$

und daher insgesamt:

$$Q(m) = \sqrt{\frac{\pi}{2m}} + O(1)$$

Zufällige Abbildungen

Die Wahrscheinlichkeit, in einem Zyklus der Länge $\lambda = 1$ zu enden, ist dann:

$$\sum_{\mu=0}^{m-1} P_m(\mu, 1) = \frac{1}{m} \sum_{k=1}^{m} \frac{m!}{(m-k)!m^k} \approx \frac{1}{m} \sqrt{\frac{\pi}{2 \cdot m}} \approx \frac{1.25}{m\sqrt{m}}$$

Zufällige Abbildungen: Fazit

- Durchschnittliche Zyklenlänge relativ gering
- Für Zufallszahlengeneratoren sind hohe Zyklenlängen erwünscht
- Zufällige Zufallszahlengeneratoren sind daher nicht Optimal

Zyklensuche nach Floyd

Man vergleiche jeweils x_i mit x_{2i}, es werden also der Reihe nach größere Werte für μ und λ getestet. Wenn $x_k = x_{2k}$ gilt: $\mu < i$ und λ ist ein Vielfaches von i. Dann kann man in einem weiteren Schritt die exakten Werte herausfinden. Der Algorithums braucht mindestens $\mu + 1$ und maximal $\mu + \lambda$ Schritte um einen Zyklus zu finden. Um die exakten Werte von μ und λ zu finden werden nochmals $\mu + \lambda$ Schritte gebraucht. In jedem Schritt wird f einmal ausgewertet und es findet ein Vergleich statt.

Zyklensuche nach Brent

Man vergleiche jeweils x_i mit $x_{\ell(i)-1}$, wobei

Definition:

$$\ell(n) := max\{2^x \mid 2^x \leq n,\ x \in \mathbb{N}_0\} = 2^{\lfloor \log_2(n) \rfloor}$$

Hier testet man also für μ exponentiell ansteigende Werte, für jedes μ welches getestet wird probiert man dann alle λ von 1 bis $\mu + 1$ aus, woraufhin man das μ weiter erhöht. Die Zyklensuche nach Brent braucht nie mehr Schritte als die Zyklensuche nach Floyd und man erhält sofort die exakte Periodenlänge. μ muss auch hier in einem weiteren Schritt erlangt werden. Auch dieser Algorithums braucht mindestens $\mu + 1$ und maximal $\mu + \lambda$ Schritte, um einen Zyklus zu finden. Werden die exakten Werte für μ und λ gebraucht, werden weitere μ Schritte benötigt.

Grundbegriffe

Definition: Teiler

$$a \mid b \iff \exists x \in \mathbb{Z} : a \cdot x = b$$

Definition: Primzahl

$$p \text{ ist Primzahl} \iff p \in \mathbb{P} \iff p > 1 \wedge (a \mid p \implies a = 1 \vee a = p)$$

Definition: Kongruenz

$$a \equiv b \pmod{m} \iff m \mid (a - b)$$

Definition: Modulo-Abbildung

$$\mathrm{mod} : \mathbb{Z} \times \mathbb{N} \to \mathbb{Z}, \quad (a, m) \mapsto a \mod m := a - \left\lfloor \frac{a}{m} \right\rfloor \cdot m$$

Grundbegriffe

Größter gemeinsamer Teiler

$$\mathrm{ggT}(a, b) = \max\{d \mid (d \mid a) \wedge (d \mid b)\}$$

Primfaktorzerlegung

Für jede natürliche Zahl $n \in \mathbb{N}$ gibt es Primzahlen $p_k < p_{k+1}$ und Exponenten α_k, so dass:

$$n = \prod_{k=1}^{M} p_k^{\alpha_k}$$

Faktorisierung nach Pollard-Rho

Sei $f(x)$ ein Polynom mit ganzzahligen Koeffizienten und n die Zahl, die zu faktorisieren ist.
Dann gilt $x_i = x_j \mod n \implies f(x_i) = f(x_j) \mod n$
Nun betrachten wir die Folgen $(x_{a_{i+1}})_{i \geq 0} : x_{i+1} = f(x_i) \mod a$ und $(x_{n_{i+1}})_{i \geq 0} : x_{i+1} = f(x_i) \mod n$ als zwei zufällige periodische Folgen, wobei μ_a, μ_n und λ_a, λ_n die Vorperiodenlängen bzw. Periodenlängen der Folgen sind.
Es gilt $a \mid n$, und daher $\mu_a \leq \mu_n$ und $\lambda_a \leq \lambda_n$. Die Erwartungswerte für $\mu_a + \lambda_a$ sowie $\mu_n + \lambda_n$ haben wir bereits im Abschnitt über zufällige Abbildungen betrachtet.

Faktorisierung nach Pollard-Rho

Um einen Zyklus der Folge $(x_{a_{i+1}})_{i \geq 0}$ zu finden (Floyd oder Brent), obwohl a unbekannt ist, betrachten wir die Folge $(x_{n_{i+1}})_{i \geq 0}$ und prüfen statt der Bedingung $x_i = x_j$ $d := \text{ggT}(x_i - x_j, n) > 1$. d lässt sich dann wie folgt charakterisieren:

$$d = \prod_{\substack{x_i \equiv x_j \pmod p \\ p|n,\ p \in \mathbb{P}}} p$$

Im Mittel findet man den ersten Zyklus für den kleinsten Primteiler p_1 von n, wenn kein anderer Teiler von n an jener Stelle einen Zyklus hat, gilt $d = p_1$. Dann haben wir einen Primteiler gefunden und können den Algorithmus mit $\frac{n}{p_1}$ erneut durchführen.

Faktorisierung nach Pollard-Rho

Ohne x_0 und/oder f abzuändern, kann kein Teiler von d gefunden werden. Insbesondere kann der Algorithmus keinen Teiler von n finden, wenn $d = n$ gilt. Dies definiert man daher als Abbruchbedingung. Üblicherweise setzt man $f(x) := x^2 + c$ mit $c \neq 0, c \neq -2$ Da die Berechnung des ggT jedoch verhältnismäßig viel Zeit (log(n) mal Multiplikation) in Anspruch nimmt, berechnet eine optimierte Variante Q (wobei $log(n) << Q << n^{\frac{1}{4}}$) mal $q = q \cdot (x_i - x_j)$ und dann ggT(q, d) berechnet. Dies nimmt in Kauf, das dazwischen zu findende Teiler multipliziert werden. Um das zu verhindern kann man den alten Wert von x_i abspeichern und die Suche von dort erneut beginnen für den Fall $d > 1$.

Pollard-Rho Faktorisierung

Der Algorithmus scheitert genau dann, wenn $\mu = \mu_p = \mu_n$ und $\lambda = \lambda_p = \lambda_n$. Die Wahrscheinlichkeit dafür ist:

$$\sum_{\lambda=1}^{p} \sum_{\mu=0}^{p-\lambda} P_p(\mu, \lambda) \cdot P_n(\mu, \lambda)$$

$$= \sum_{\lambda=1}^{p} \sum_{\mu=0}^{p-\lambda} \frac{1}{p} \prod_{k=1}^{\mu+\lambda} \left(1 - \frac{k}{p}\right) \cdot \frac{1}{n} \prod_{k=1}^{\mu+\lambda} \left(1 - \frac{k}{n}\right)$$

$$= \frac{1}{n \cdot p} \sum_{\lambda=1}^{p} \sum_{\mu=0}^{p-\lambda} \prod_{k=1}^{\mu+\lambda} \left(1 - \frac{k}{p}\right) \cdot \left(1 - \frac{k}{n}\right)$$

Pollard-Rho Faktorisierung

Der Algorithmus liefert einen Primfaktor p_i genau dann, wenn $\mu_{p_i} + \lambda_{p_i} < \mu_{p_j} + \lambda_{p_j}$ für $i \neq j$. Die Wahrscheinlichkeit dafür ist:

$$\sum_{\substack{\mu_{p_i}+\lambda_{p_i} \leq p_i \\ \mu_{p_i}+\lambda_{p_i} < \mu_{p_j}+\lambda_{p_j} \leq p_j}} \prod_{j=1}^{M} P_{p_j}(\mu_j, \lambda_j)$$

Laufzeitanalyse der Pollard-Rho Faktorisierung

Das Pollard-Rho verfahren ist besonders effizient für Zahlen mit einem kleinen Primfaktor. Die Zeit um einen Faktor p zu finden ist im Mittel $\sqrt{p}$. $\Omega(n)$ ist die Anzahl der (mit Vielfachheit gezählten) Primfaktoren von n. Dann gilt: $\Omega(n) \sim \log(\log(n))$ für $n \to \infty$ (Satz von Hardy-Ramanujan). Es gilt: $p_1 \leq n^{\frac{1}{\Omega(n)}}$. Unter der Annahme, dass die zu faktorisierende Zahl keine Primzahl ist, ist $\Omega(n) \geq 2$. Dann hat man $p_1 \leq \sqrt{n}$ als exakte Schranke.

Referenzen

Robert Sedgewick (1995)
An Introduction to the Analysis of Algorithms
Addison-Wesley

J.M. Pollard (1975)
A Monte Carlo method for factorization
BIT Numerical Mathematics 15(3), 331-334.

R.P. Brent (1980)
An Improved Monte Carlo Factorization Algorithm
BIT Numerical Mathematics 20, 176-184.

Donald E. Knuth (1997)
The Art of Computer Programming, vol. I: Fundamental Algorithms (3rd ed.)
Addison-Wesley

Donald E. Knuth (1997)
The Art of Computer Programming, vol. II: Seminumerical Algorithms (3rd ed.)
Addison-Wesley