Algorithmen und Datenstrukturen

Master

Probabilistische Algorithmen

Inhalt

- Die Idee
- Zufallszahlen
- MC Methoden

Die Idee

- Ein Algorithmus verwendet den Zufall als Teil des Lösungswegs.
- Input sind gleichverteilte Zufallszahlen
  - Annahme: im Durchschnittsfall wird so die Lösung schneller gefunden als bei vollständiger Lösung
  - Sichere Lösung bei variabler Zeit
  - Unsichere Lösung nach fixer Zeit

Die Idee

- Nutze Zufallszahlen, um Speicherplatz oder Laufzeit zu reduzieren
- Nutze Zufallszahlen, um das Ergebnis bis auf einen abschätzbaren Fehler anzunähern
  - Monte Carlo: Fehler/Unschärfen sind erlaubt
  - Las Vegas: sicheres Ergebnis oder keines

Beispiel: Random Search

- Gegeben ist ein Array von n Zahlen mit Hälfte ‘a‘ und Hälfte ‘b‘
- Problem: finde ein ‘a‘ im Array

- Las Vegas Ansatz:

```
char c;
do {
  c = Array[rand()];
} while (c != ‘a‘);
```

Laufzeit unbestimmt, liefert immer ein Ergebnis

Beispiel Random Search

- Monte Carlo Ansatz: gib Maximalanzahl der Versuche mit an:

```
char c;
int i, max;
for (i=0; i<max; ++i) {
  c = Array[rand()];
  if (c == ‘a‘) break;
}
```

Fixe Laufzeit (über max), kann aber falsches Ergebnis liefern

Geschichtliches

- Michael O. Rabin entwickelt 1976 einen Algorithmus zur Bestimmung nächster Nachbarn (computational geometry)
- 1997 Miller-Rabin Test zur Findung von grossen Primzahlen
  - Ursprünglich John Selfridge1974
  - Publiziert von Miller und Rabin 1976

Der Zufall

- Von Zufall spriht man dann, wenn für ein einzelnes Ereignis oder das Zusammentreffen mehrerer Ereignisse keine kausale Erklärung gibt.
- Eine Zahlenfolge ist „zufällig“, wenn aus der Folge der bereits erschienenen Zahlen nicht auf die nächste geschlossen werden kann. Sie ist nicht vorhersagbar.

Zufall in der Natur

- In der Phsyik spielt der Zufall auch eine Rolle:
- Makroskopisch gibt es Naturgesetze, die ein gutes Modell der Natur darstellen.
- Je kleiner man die Betrachtung macht, desto unschärfer wird das genau bestimmte Verhalten.
- Das exakte Messergebnis wird durch eine Wahrscheinlichtkeit abgelöst, die das Eintreffen eines bestimmen Ergebnisses beschreibt (Quantentheorie, Thermodynamik)

Zufälligkeit

- Zufälligkeit ist subjektiv
  - Eine Bitfolge mag für einen Beobachter zufällig erscheinen, für einen anderen (der die Kryptographie beherrscht) durchaus Sinn ergeben.
  - Zufällig ≠ Nicht Vorhersagbar
- In der Mathematik spricht man von Pseudozufallszahlen, wenn ihre Folge zwar zufällig erscheint, es aber dennoch eine Beschreibung ihrer Entstehung gibt.

Erzeugen des Zufalls

- Drei Mechanismen sind verantwortlich für (scheinbar) zufälliges Verhalten eines Systems:
  - Zufall durch die Umgebung
    - Brown'sche Molekularbewegung
    - Hardware Zufallszahlen Generator
  - Zufall aufgrund von Anfangsbedingungen
    - Würfel, Roulette
  - Zufall durch das System erzeugt
    - Pseudozufallszahlen

Zufallszahlengenerator

- Random Number Generator (RNG)
- Wichtig für
  - Kryptographie
  - Simulation
  - Spiele
- Für wirkliche Zufallszahlen werden Hardware-Generatoren herangezogen
- Pseudozufallszahlen erscheinen zufällig

Hardware RNG

- Früher Würfel
- Heute beruhen diese Generatoren auf atomaren Effekten
  - Radioaktiver Zerfall
  - Thermisches Rauschen
  - Zener Dioden (Durchbruch Rauschen)
  - Atmosphärisches Rauschen

Berechnungsmethoden

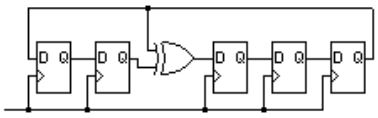
- Pseudo-RNG (PRNG)
- Berechnung erzeigt eine Folge von gut gemischten Zahlen, die sich aber nach gewisser Zeit wiederholt
- Linear kongruenter Generator

$$X_{n+1} = (aX_n + b)\ mod\ m$$

Linear Feedback Shift Register

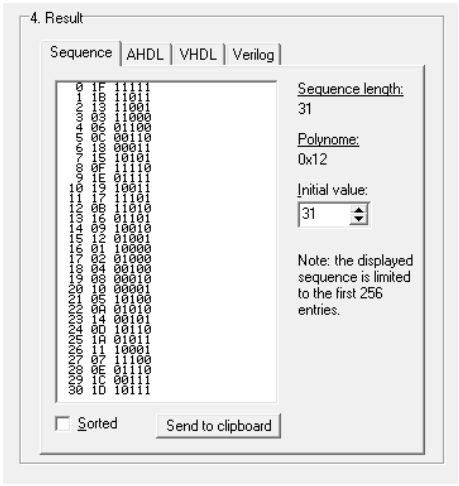
- Rückgekoppeltes Schieberegister zur Erzeugung einer pseudozufälligen Bitfolge.
- Initialisierung mit Seed
- Zahlenfolge ist durch Rückkopplung vorbestimmt.
- Rückkopplung bestimmt Länge der Zahlenfolge.
- Alles 0 ist ungültiger Zustand

Linear Feedback Shift Register



Beispiel: 5 Bit
Sequenzlänge 31

Weitere Details



LFSR Anwendungen

- Pseudo-Zufallssequenzen
- Zähler

- Kryptographische Scambler
  - Klartext xor Bitfolge
  - Empfänger muss Bitfolge kennen!

- DVB-T, CDMA (Handy),
- SATA, SDI
- IEEE 802.11a (WLAN), IEEE 802.15 (Bluetooth)

PRNG

- Von Neumann 1946
- Middle-square method

- Wähle Zahl, quadriere sie und wähle die mittleren Ziffern als neue Zahl

- 1111 -> 01234321
- 2343 -> 05489649
- 4896 -> ...

Mersenne Twister

- 1997 M. Matsumoto, T. Nishimura
- Heute Standard in vielen Simulationsumgebungen (z.B. Matlab)
- Liefert 32 bit Integer Folgen
- Twisted generalized feedback shift register
- Algorithmus Details

Cryptographic secure PRNG

- Problem bei PRNG ist die Möglichkeit, durch lange Beobachtung der Folge die nächste Zahl bestimmen zu können
- Für kryptographische Sicherheit muss dies unterbunden werden (oder so aufwändig sein, dass es praktisch unmöglich ist)
- Windows: CryptGenRandom

Nicht-gleichverteilte Zufallszahlen

- Erhält man mittels Zufallszahlengenerator und einer Verteilungsfunktion
- Zufallszahl wird ermittelt
- Über die Berechnung der inversen Zielverteilung erhält man eine Abbildung der Zufallszahl auf die Zielverteilung

Las Vegas Algorithmen

- Algorithmuen, der Zufallszahlen benutzt und immer korrekte Ergebnisse liefern oder einen Fehler.
- 1979 Laszlo Babai

- Eingesetzt wenn
  - Die Anzahl der möglichen Lösungen eher klein ist
  - Die Richtigkeit der Lösung leicht überprüft werden kann
  - Die Berechnung der Lösung komplex ist

Monte Carlo Algorithmen

- Randomisierte Algorithmen, die mit einer nach oben hin beschränkten Wahrscheinlichkeit ein falsches Ergebnis liefern dürfen.
- Meist effizienter als deterministische Algorithmen.
- Durch vielfaches Wiederholen mit anderen Zufallszahlen kann die Fehlerwahrscheinlichkeit weiter beschränkt werden.

Beispiele

- Miller Rabin Test
- Testet ob eine ungerade Ausgangszahl eine Primzahl ist
  - Sicher zusammengesetzt
  - Wahrscheinlich prim
- Ist auch heute noch gebräuchlich für die Bestimmung von Primzahlen in der Kryptographie

Beispiele

- Berechnung von PI

- Bestimme beliebigen Punkt in X,Y im Intervall [0,1]
- Berechne, ob der Punkt innerhalb des Einheitskreises liegt
- Flächenverhältnis Kreis zu Quadrat:

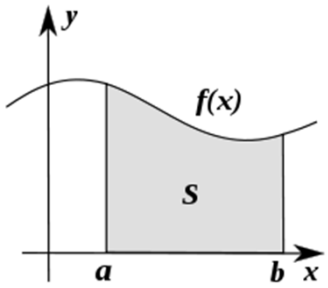
$$\frac{A_{Kreis}}{A_{Fläche}} = \frac{r^2\pi}{4r^2} = \frac{\pi}{4} = \frac{Treffer\ im\ Kreis}{Punkte\ im\ Quadrat}$$

- Demo

Beispiele

- Numerische Integration

- Eine Funktion wird durch zufällig bestimmte Sützstellen und Flächenberechnung angenähert



Monte Carlo Methoden

- 1930 Enrico Fermi (erste Ideen)
- 1946 Stanislaw Ulam (Los Alamos)
- Von Neumann gab dem Verfahren den Codenamen „Monte Carlo“

- Wichtig für Simulation und Rekonstruktion
  - Wetter
  - Klima
  - Produktfertigung

  - Kernphysik (CERN Teilchenbeschleuniger)
  - Medizin (Röntgen und Nuklearmedizin)