Programmiertechnik II

Zufallszahlen

Motivation

- Simulation
  - Frisörbeispiel
- Stichprobenauswahl
  - Telefonumfragen
- Numerische Algorithmen
  - naives Beispiel: Berechnung von Pi
- Automatisiertes Testen
  - Beispiel aus Übungsaufgabe
- "Faire" Entscheidungen
  - Generierung einer Reihenfolge von Seminarvorträgen
- Computerspiele
  - Verteilung von Hindernissen auf dem Spielfeld

Echter Zufall und Pseudozufall

- Zufällige Ereignisse: keine erkennbare Ursache; bei Wiederholung unter "den gleichen Umständen" auch andere Ergebnisse beobachtbar
  - Würfel
- Deterministisches Programm liefert immer gleiches Ergebnis bei gleichem Zustand
  - Zustand aber u.U. nicht nach außen sichtbar; Ergebnis erscheint zufällig -> pseudozufällig

Echter Zufall

- Generierung von echten Zufallszahlen auf Basis zufälliger (physikalischer) Ereignisse
  - Tippett 1927: 40000 Zufallsziffer, gewonnen aus britischer Bevölkerungszählung
  - RAND Corporation 1955: 1000000 Zufallszahlen aus physikalischem Gerät
  - random.org: Fortlaufende Generierung aus atmosphärischem Rauschen
- Zufallszahlen in Betriebssystemen: Verwaltung von Entropiepools
  - Linux: /dev/random, /dev/urandom
  - Windows: CryptoAPI
  - Verwendet zur sicheren Generierung geheimer Schlüssel

Pseudozufallszahlen: Güteforderungen

- Forderungen an Folgen von Zahlen
- statistische Verteilung
  - Gleichverteilung: jede Zahl kommt gleich oft vor
  - andere Verteilungen lassen sich mathematisch ableiten:
    - Normalverteilung/Gaußverteilung
    - Erlangverteilung
    - ...
- Nichtvorhersagbarkeit
  - ideal (bei Gleichverteilung): auf jede Zahl kann prinzipiell jede andere Zahl folgen
- lange Perioden:
  - Nachdem eine Folge von Pseudozufallszahlen beobachtet wurde, dauert es lange, bis die gleiche Folge wieder erscheint

Erste Versuche

- J. von Neumann 1946, Mittquadratmethode
- geg. n-stellige Zahl
- berechne n^2
- nächste Zufallszahl: verwende die mittleren n Stellen
- Algorithmus oft periodisch:
  - unterschiedliche Periodenlänge: 0 -> 0 -> 0...
  - 20-bit-Zahlen: längste Periode 142

Linearer Kongruenzgenerator

- Lehmer 1949
- LCG: Linear Congruential Generator
- wähle 4 feste Zahlen:
  - Modulus m > 0
  - Faktor a (0 <= a < m)
  - Inkrement c (0 <= c < m)
  - Startwert X_0 (0 <= X_0 < m)
- $X_{n+1} = (aX_n + c) \bmod m$
- Periodenlänge hängt von konkreten Werten ab
  - Periode offenbar < m

- Wahl von m: effiziente Modulo-Berechnung?
  - Zweierpotenz, am einfachsten $2^{\text{Wortlänge}}$
- Wahl von a: maximale Periodenlänge
  - Satz von Knuth: Periodenlänge ist m genau dann, wenn
    - c und m teilerfremd sind, und
    - a-1 Vielfaches jedes Primteilers p von m ist, und
    - a-1 ein Vielfaches von 4 ist, falls m ein Vielfaches von 4 ist
- Vorhersagbarkeit der nächsten Zahl:
  - Verwendung von $X_n \bmod m'$ $(m' < m)$
- K&R rand(): $m=2^{32}$, a= 1103515245, $c=X_0=12345$›

Mersenne-Twister

- Makoto Matsumoto (松本 眞) und Takuji Nishimura (西村 拓士) 1997
- übliche Variante: Periode 2^{19937}
- hohe Gleichverteilung
  - bewiesen für bis-zu-623-Tupel
- schnell
- jedes Bit für sich gleichverteilt
- interner Zustand: (N=)624 32-bit Zahlen
  - Startwerte Y_1 bis Y_{624}
- Rechenschritt: $h = Y_{i-N} - Y_{i-N} \% 2^{31} + Y_{i-N+1} \% 2^{31}$
  - $Y_i = Y_{i-227}$ ^ int(h/2) ^ ((h % 2) * 0x99080bdf)

- Berechnung der Zufallszahl: Umordnung der Bits von Y_i
  - $x = Y_i \text{ ^ } int(Y_i/2^{11})$
  - $y = x \text{ ^ } ((x * 2^7) \& 0x9d2c5680)$
  - $z = y \text{ ^}((y * 2^{15}) \& 0xefc60000)$
  - $Z_i = z \text{ ^ } int(z/2^{18})$
- Initialisierung der 624 Zahlen mit echten Zufallszahlen

χ^2-Test

- statistischer Test
  - u.a. auch geeignet, um Güte eines Zufallszahlengenerators zu testen
- testet Verteilungsfunktionen
- einfachster Fall: Gleichverteilung ganzer Zahlen
  - jede Zahl müsste gleich oft vorkommen
  - Durchführung der Messung: wie oft kommt jede Zahl tatsächlich vor
- Zufallsgröße mit m Freiheitsgraden:

$$\chi^2 = \sum_{j=1}^{m} \frac{(n_j - n_{jo})^2}{n_{jo}}$$

- Test: Vergleich mit (1-α)-Quantil der χ^2-Verteilung, $\chi^2(1-\alpha;m-1)$
  - Werte in Tabellen berechnet

Kolmogorov-Smirnov-Test

- Андрей Николаевич Колмогоров, Владимир Иванович Смирнов
- geeignet auch für kleine Stichprobenzahlen
- geeignet auch für kontinuierliche Werte
- einseitiger Test: Vergleich von Ist-Verteilung mit Soll-Verteilung (Hypothese)
  - (kumulative) Verteilungsfunktion F(x): Wie viele Messwerte sind kleiner als x?
- Berechnung der Differenzen zwischen Ist und Soll
- Bildung des Maximums der Beträge
- Vergleich mit Tabellenwert:
  - Irrtumswahrscheinlichkeit 5%: 1.36/sqrt(N) bei N Stichproben

Diehard Tests

- Entwickelt von George Marsaglia
- Erkennung typischer Symptome von Pseudozufallszahlen
- Beispiele:
  - Alle Permutationen von 5 Zahlen sollten gleich oft vorkommen
  - Verteile Punkte zufällig auf der Ebene. Die Abstände sollten normalverteilt sein
  - Generiere Folgen von Gleitkommazahlen; Längen von aufsteigenden und absteigenden Folgen muss gewisser Verteilung unterliegen
  - ...