K 18 Simulation
Warum?
(a) In der Statistik

- Wenn wir eine analytische Lösung vermuten:
  Gosset und die Verteilung des Korrelationskoeffizients

- Wenn es nur eine numerische oder eine
  assymptotische analytische Lösung gibt:
  Geburtstags Problem: $P(> 2$ am selben Tag $) =$?

- Wenn wir keine Lösung haben:
  Epidemiologische Probleme

- Bayes und Monte Carlo Markov Chain (MCMC)

$$P(\theta|x) = \frac{P(x|\theta)\pi(\theta)}{\int_\theta P(x|\theta)\pi(\theta)\mathrm{d}\theta}$$

(b) Im allgemeinen
Komplexe Strukturen — Verkehr, Flughäfen, Derivate
Militär — “Wargames”
(und sogar Strategien für Monopoly und Patience)

18.1 Das Geburtstagsproblem

Wenn m Leute zusammenkommen, was ist die Wahrscheinlichkeit, dass alle verschiedene Geburtstage haben?

Nehmen wir an, dass für jede Person, unabhängig von den anderen, alle Tage gleichwahrscheinlich sind:

$$P(\text{Geburtstag ist Tag } i) = \frac{1}{365}$$

Weder säisonale Schwankungen in Geburtsraten noch Schaltjahre werden beachtet.

$$P(\text{keine gemeinsame Geburtstage}) = \prod_{i=0}^{m-1} \left(1 - \frac{i}{365}\right)$$

$$< 0.5 \quad \text{für } m = 23$$

$$< 0.05 \quad \text{für } m = 48$$

Für mehrfache Koinzidenzen (Diaconis und Mosteller):

$$P(\geq k \text{ Geburtstage an einem Tag}) > 0.5$$

k	2	3	5	9	10
m	23	88	313	985	1181

18.2 Zufallszahlen

Zufallszahlen sind für Simulation notwendig.

Was hat man vorher gemacht?
— Tabellen von Zufallszahlen, (z.B. das RAND Buch)
— elektronisches Rauschen (z.B. ERNIE)

Wegen Effizienz und Wiederholbarkeit hat man seit langem versucht, Sequenzen von Zufallszahlen künstlich zu erzeugen.

Eine Sequenz von Pseudozufallszahlen $\{U_i\}$ ist eine deterministische Sequenz von Zahlen aus $[0, 1)$, die die gleichen statistischen Eigenschaften hat, wie eine Sequenz von Zufallszahlen.

John von Neumann

“Any one who considers arithmetical methods of producing Random numbers is, of course, in a state of sin.”

“There is no such thing as a random number — there are only methods to produce random numbers, and an arithmetical procedure is of course not such a method.”

und

“If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.”

Zufallszahlen?

41592653589793238462643383279502884197169399375105820974944592307816406286208998628034825342117067982148086513282306647...

18.3 Eigenschaften von Pseudozufallszahlen

Welche Eigenschaften müssen solche Sequenzen haben?

$$P(U_n = u | U_{n-1}, U_{n-2}, \ldots) = P(U = u)$$

sollte eine Gleichverteilung sein. Wenn wir die Zahlen $\{0, 1 \ldots 9\}$ nehmen, sollten

$\{X_i\}$ gleichverteilt

$\{X_i, X_j\}$ gleichverteilt

$\{X_i, X_j, X_k\}$ gleichverteilt u.s.w.

Corr$(X_i, X_{i+k}) \doteq 0 \quad k = 1, 2, \ldots$

es keine Folgen geben

die Sequenzen dürfen keine Struktur haben,
die die Berechnungen stören könnten.

Für statistische Berechnungen ist es unwichtig, ob die nächste Zahl vorhergesagt werden kann oder nicht. In der Kryptographie ist diese Eigenschaft ausschlaggebend.

18.4 Erzeugung von Pseudozufallszahlen

18.4.1 Lineare kongruente Generatoren

$$x_{j+1} = (ax_j + c) \mod m$$

Alle Größen sind ganzzählig:

a Multiplikator

c Inkrement

m Modulus

x_0 Startwert

z.B.

$$m = 7, c = 2, a = 3, x_0 = 1 \Rightarrow$$

$$x_1 = 5, x_2 = 3, x_3 = 4, x_4 = 0, x_5 = 2, x_6 = 1$$

mit einer Periode von 6 ABER

$$x_0 = 6 \Rightarrow x_1 = 6, \ldots x_n = 6$$

18.4.2 Multiplikative lineare kongruente Generatoren

In der Tat hat man früher oft $c = 0$ gesetzt, um Rechenzeit zu sparen:

$$x_{j+1} = ax_j \quad \text{mod } m$$

z.B. $m = 2^{31} - 1$

Data Desk benutzt $a = 7^5 = 16807$
$(630, 360, 016$ 'soll' besser sein)

Bis c. 1975 hat IBM $m = 2^{31}$ und $a = 65539$ angeboten. Dieses Generator, RANDU, hat große Schwächen gehabt.

Ansi-C ist $2^{31}, 1103515245, 12345, 12345$

Maple hat $10^{12} - 11, 427419669081, 0, 1$

Es müssen die theoretischen Eigenschaften (z.B. Periode) und empirische Resultate untersucht werden.

Und, natürlich, sollten solche Generatoren schnell sein.

c.f. Mersenne Twister von Makoto Matsumoto.
Periode von $2^{19937} - 1$ und
$623-$dimensionale Gleichverteilung.

18.5 Kombinationen von Generatoren

Gegeben zwei Generatoren, die unabhängige Zahlensequenzen $X \sim G[0,1)$ und $Y \sim G[0,1)$ generieren, ist

$$Z = X + Y - \text{Trunc}(X + Y)$$

auch $G[0,1)$ verteilt.

$$f_Z(z) = \int_0^z f_X(x) f_Y(z-x) \mathrm{d}x + \int_z^1 f_X(x) f_Y(1+z-x) \mathrm{d}x$$

$$= z + 1 - z$$

$$= 1$$

Weiterhin gilt

$$\{X_i\} \sim G[0,1) \text{ u.i.v. } \Rightarrow$$

$$Z = \sum X_i - \text{Trunc}(\sum X_i) \sim G[0,1)$$

Die Forschung über Erzeuger von Zufallszahlen ist sehr lebendig, aber es gibt kein optimales Resultat.

18.6 Zufällige Stichproben aus Verteilungen

Zwei Schritte:

(1) u_i wird aus $G[0,1)$ erzeugt.

(2) $x_i = h(u_i)$ für eine passende Transformation h

18.6.1 Diskrete ZV

$$p_i = P(X = x_i) \qquad i = 1, \ldots m$$

$$x_1 < x_2 < \ldots x_m$$

$$\begin{aligned} F(x) &= 0 && X < x_1 \\ &= \sum_{i=1}^{j} p_i && x_j \leq X < x_{j+1} \\ &= 1 && x_m \leq X \end{aligned}$$

Sei die Folge $\{u_i\}$ u.i.v. $\sim G[0,1)$ gegeben, setzen wir

$$X_k = x_j \quad \text{wenn} \quad \sum_{i=1}^{j-1} p_i \leq u_k < \sum_{i=1}^{j} p_i$$

$$X_k = x_1 \quad \text{wenn} \quad 0 \leq u_k < \sum_{i=1}^{j} p_i$$

18.6.2 Stetige ZV

Falls X die Verteilungsfunktion $G(x)$ hat, haben wir schon bewiesen ($\S 8.8.1$), dass für G umkehrbar

$$h = G^{-1}$$

die benötigte Form hat. Dieses Resultat ist die stetige Version des Verfahrens für diskrete ZV.

Wenn $\{u_1, u_2, \ldots, u_n\}$ eine zufällige Stichprobe aus $G[0, 1)$ ist, dann ist $\{x_i = G^{-1}(u_i)\}$ eine zufällige Stichprobe für eine ZV mit Verteilungsfunktion $G(x)$.

Leider klappt diese Methode sehr selten und deshalb werden verteilungsspezifische Methoden vorgeschlagen.

18.6.3 Beispiele mit $G^{-1}(u)$

Exponential $Y \sim E(\lambda)$

$$G(y) = 1 - e^{-\lambda y}$$

$$G^{-1}(u) = -\frac{1}{\lambda} \ln(1-u)$$

Aber, wenn

$$U \sim G[0,1)$$

dann gilt, dass

$$V = 1 - U \sim G[0,1)$$

$\Rightarrow$ wir können genausogut

$$G^{-1}(u) = -\frac{1}{\lambda} \ln u$$

nehmen.

Normal Für $X \sim N(\mu, \sigma^2)$ lässt sich

$$G(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \mathrm{d}x$$

nur numerisch berechnen und umkehren.