Stochastik für die Informatik, Vorlesung 18

Inhalt

- Simulation von Zufallsvariablen
- Monte-Carlo-Schätzer

Lernziele

- Simulationsmethoden kennen
- Monte-Carlo-Schätzer berechnen können

Vorkenntnisse Zufallsvariablen, Verteilungsfunktion, Erwartungswert, Integralrechnung

Kapitel 15: Simulation von Zufallsvariablen

Erinnerung: U gleichverteilt auf $[0,1]$ bedeutet, dass U die Verteilungsfunktion

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

bzw. die Dichte $f(t) = 1_{\{t \in [0,1]\}}$ hat.

(Satz 15.1) Sei X auf $[0,1]$ gleichverteilt. Sei F eine invertierbare Verteilungsfunktion. Dann ist

$$Y := F^{-1}(X)$$

eine Zufallsvariable mit Verteilungsfunktion F.

Anwendung dieses Satzes: Simulation von Zufallsvariablen: Sobald man eine auf $[0,1]-$ gleichverteilte Zufallsvariable simulieren kann, kann man im Prinzip jede Zufallsvariable mit bekannter (invertierbarer) Verteilungsfunktion simulieren (also z.B. Exponentialverteilung).

Bernoulli aus gleichverteilt

(Satz 15.1) Sei U eine auf $[0,1]$ gleichverteilte Zufallsvariable. Sei $p \in (0,1)$. Dann ist

$$X(U) := 1_{\{U \leq p\}}$$

eine Bernoulli-verteilte Zufallsvariable mit Parameter p.

- (Beweis)
- (15.1, 15.2)

Pseudozufallszahlen

(Def. 15.1) Ein Pseudozufallszahlengenerator ist ein Algorithmus welcher ausgehend von einem Startwert eine deterministische Folge von Zahlen erzeugt, welche sich "wie eine Realisierung einer Folge von (uniform auf $[0,1]$ verteilten) Zufallsvariablen" verhält.

- Überprüfung des zufälligen Verhaltens mit Hilfe von statistischen Tests, z.B. χ^2-Test auf uniforme Verteilung
- "Echte" Zufallszahlen: Mit Hilfe physikalischer Generatoren, z.B. durch Beobachtung radioaktiver Zerfälle, kosmisches Rauschen...

Pseudozufallszahlen

Beispiel 15.4: Der lineare Kongruenzgenerator

- Wähle Parameter $m \in \mathbb{N}, a \in \mathbb{Z} \setminus \{0\}, c \in \mathbb{Z}$.
- Wähle einen Startwert $x_0 \in \{0, ..., m\}$
- Definiere rekursiv

$$x_n := (ax_{n-1} + c) \bmod m$$

- Definiere für $n \in \mathbb{N}$

$$u_n := \frac{x_n}{m}$$

- Dann ist die Folge $(u_n)_{n \in \mathbb{N}_0}$ eine Folge von Pseudozufallszahlen (uniform auf $[0, 1]$).

Pseudozufallszahlen

Beispiel 15.4: Der lineare Kongruenzgenerator

Eigenschaften:

- Damit können höchstens m verschiedene Zahlen erzeugt werden
- Die erzeugte Folge ist periodisch
- Entsprechend sollte m möglichst groß gewählt werden, auf jeden Fall deutlich größer als die Anzahl zu erzeugender Zufallszahlen
- Wahl von a nach zahlentheoretischen Überlegungen
- Wahl des Startwerts kann große Auswirkung auf die erzeugte Folge haben, deshalb ist die Methode geeignet um Zufallszahlen zu erzeugen.
- Reproduzierbarkeit durch Wahl des Startwerts gegeben

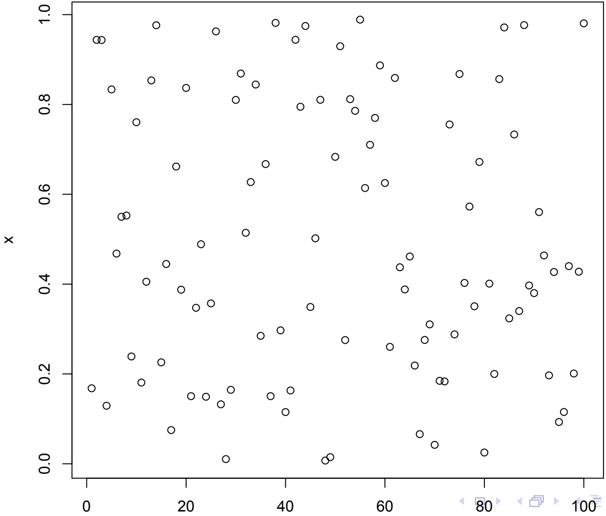
Pseudozufallszahlen

Das Software-Paket R bietet die Auswahl aus verschiedenen Pseudozufallszahlen-Generatoren, der Standard ist der sogenannte Mersenne-Twister (welcher ebenfalls ein rekursives Verfahren benutzt).

- Befehl: `runif`
- `runif(n)`: Vektor von n unabhängigen uniform $[0, 1]$ verteilten Zufallszahlen
- `runif(n, a, b)` : Vektor von n unabhängigen uniform $[a, b]$ verteilten Zufallszahlen
- Mit dem Befehl `set.seed(x)` kann der Startwert x_0 festgelegt werden.
- Wählt man zweimal denselben Startwert, so erhält man zweimal dieselbe Folge
- Simulation von anderen Verteilungen: `rnorm`, `rexp`...

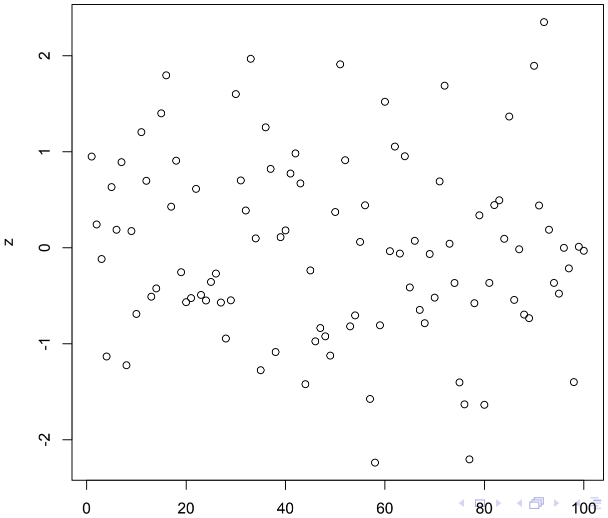
Pseudozufallszahlen

Beispiel: `x<-runif(100)`



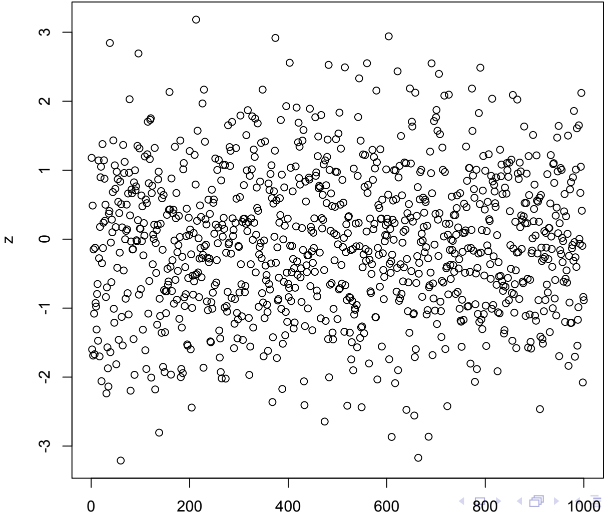
Pseudozufallszahlen

Beispiel: `z<-rnorm(100)`



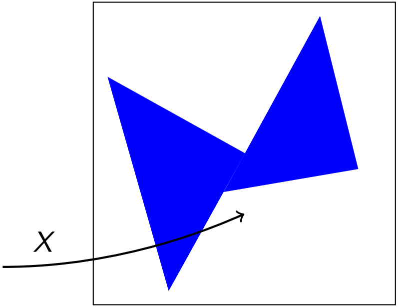
Pseudozufallszahlen

Beispiel: `z<-rnorm(1000)`



Monte-Carlo-Schätzer: Grundidee

Ein elementares Beispiel aus der allerersten Vorlesung



Wie wahrscheinlich ist es, dass eine rein zufällige Auswahl eines Punktes aus dem Quadrat auf die blaue Fläche führt?

"Rein zufällig" war hier umgangssprachlich für gleichverteilt auf dem Quadrat.

Monte-Carlo-Schätzer: Grundidee

Sei Q das Quadrat. $X \sim \mathrm{Unif}(Q)$ bedeutet

$$\mathbb{P}(X \in A) = \frac{\text{Fläche von } A}{\text{Fläche von } Q}$$

Wählen wir eine zufällige Folge $X_1, X_2, X_3, \ldots$ von Punkten aus Q, d.h. $X_i, i \in \mathbb{N}$ sind auf Q gleichverteilt und unabhängig, so können wir nach dem Gesetz der großen Zahlen die Wahrscheinlichkeit für $X \in A$ approximieren durch

$$\mathbb{P}(X \in A) \approx \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \in A\}}$$

Die Auswahl der Punkte kann also als Zufallszahlengenerator verstanden werden.

Monte-Carlo-Schätzer

(Def. 15.2) Sei $U_n, n \in \mathbb{N}$ eine Folge von unabhängigen, auf $[0,1]$ gleichverteilten Zufallsvariablen. Sei $A \subset [0,1]$. Dann ist

$$\hat{\mathbb{P}}(A) := \frac{1}{n} \sum_{i=1}^{n} 1_{\{U_i \in A\}}$$

ein Monte-Carlo-Schätzer für $\mathbb{P}(U_1 \in A)$. Weiter ist für $h : [0,1] \to \mathbb{R}$

$$\hat{\mathbb{E}}[h(U)] := \frac{1}{n} \sum_{i=1}^{n} h(U_i)$$

ein Monte-Carlo-Schätzer für $\mathbb{E}[h(U_1)]$.

- Konkrete Berechnung: Setze für die U_i geeignete Pseudozufallszahlen $u_i, i \in \mathbb{N}$, ein
- Allgemeiner: $h : [a,b] \to \mathbb{R}$, $h : [0,1]^d \to \mathbb{R}$, $[a,b]^d \to \mathbb{R}$
- Satz 15.2: $\hat{P}$ bzw. $\hat{\mathbb{E}}$ sind erwartungstreue und konsistente Schätzer

Monte-Carlo-Approximation von Integralen

Grundidee: U gleichverteilt auf $[0,1]$, hat also Dichte $f_U(x) = 1_{\{x \in [0,1]\}}$. Dann gilt für jede integrierbare Funktion $h : [0,1] \to \mathbb{R}$

$$\int_0^1 h(x)dx = \int_{-\infty}^{\infty} h(x) f_U(x) dx = \mathbb{E}[h(U)].$$

(Satz 15.3) Sei $h : [a,b] \to \mathbb{R}$ eine integrierbare Funktion. Seien $U_i, i \in \mathbb{N}$ unabhängige, auf $[a,b]$ gleichverteilte Zufallsvariablen. Dann ist

$$\hat{I}(h) := \frac{(b-a)}{n} \sum_{i=1}^{n} h(U_i)$$

ein erwartungstreuer Schätzer für das Integral $\int_a^b h(x)dx$. Entsprechend nennt man $\hat{I}(h)$ einen Monte-Carlo Schätzer für das entsprechende Integral. Eine Verallgemeinerung auf mehrdimensionale Integrale ist möglich.

Monte-Carlo-Approximation von Integralen

Grundidee: U gleichverteilt auf $[0,1]$, hat also Dichte $f_U(x) = 1_{\{x\in[0,1]\}}$. Dann gilt für jede integrierbare Funktion $h : [0,1] \to \mathbb{R}$

$$\int_0^1 h(x)dx = \int_{-\infty}^{\infty} h(x) f_U(x)dx = \mathbb{E}[h(U)].$$

(Satz 15.3) Sei $h : [a,b] \to \mathbb{R}$ eine integrierbare Funktion. Seien $U_i, i \in \mathbb{N}$ unabhängige, auf $[a,b]$ gleichverteilte Zufallsvariablen. Dann ist

$$\hat{I}(h) := \frac{(b-a)}{n} \sum_{i=1}^{n} h(U_i)$$

ein erwartungstreuer Schätzer für das Integral $\int_a^b h(x)dx$. Entsprechend nennt man $\hat{I}(h)$ einen Monte-Carlo Schätzer für das entsprechende Integral. Eine Verallgemeinerung auf mehrdimensionale Integrale ist möglich.

Monte-Carlo-Schätzer

Beispiel 15.5

- $h(x) = 3x^2$
- Zur Kontrolle berechnen wir $\mathbb{E}[h(U)]$ analytisch, $U \sim \text{unif}[0,1]$
- Berechnung mit von R generierten Pseudozufallszahlen:
  `x<-runif(100)`
  `mc<-sum(3* x^2)/100` Ergebnisse (3 Wiederholungen):
  0.8689..., 1.1629..., 0.8918...
  Mit 1000 Pseudozufallszahlen: 0.9773

Monte-Carlo-Schätzer

Beispiel 15.5

- $h(x) = 3x^2$
- Zur Kontrolle berechnen wir $\mathbb{E}[h(U)]$ analytisch, $U \sim \text{unif}[0,1]$
- Berechnung mit von R generierten Pseudozufallszahlen:
  `x<-runif(100)`
  `mc<-sum(3* x^2)/100` Ergebnisse (3 Wiederholungen):
  0.8689..., 1.1629..., 0.8918...
  Mit 1000 Pseudozufallszahlen: 0.9773