JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

www.uni-frankfurt.de

Adaptive Modellierung und Simulation

Kapitel 5: Stochastische Simulation

Rüdiger Brause

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Simulation

- Simulation
  - determin. Simulation
    - kontinuierl.
    - diskret
  - stochast. Simulation
    - Monte-Carlo-Simulation
    - Simulation mit Zufahlszahlen

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Stochastische Simulation

Kont.vs. diskrete Simulation

Erzeugung einer Verteilung

Monte-Carlo-Simulation

Simulationsrahmen

Auswertung

Analytische Simulation

- **Beispiel:** Optimierung einer Lagerhaltung
  - Produktion *R* nutzt ein Lager *L*
  - fester Bearbeitungsaufwand *F* jeder Bestellung *B*
    ⇒ **große** Bestellmengen
  - ungenutztes Material bindet Kapital *H*
    ⇒ **kleine** Bestellmengen

Optimum B* ?

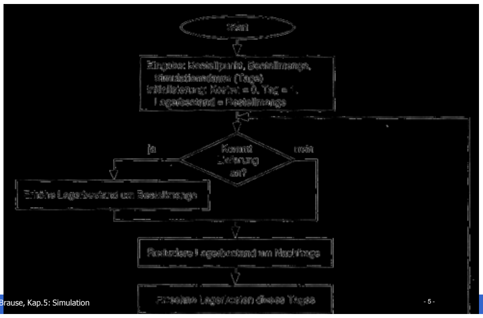
Rechnung

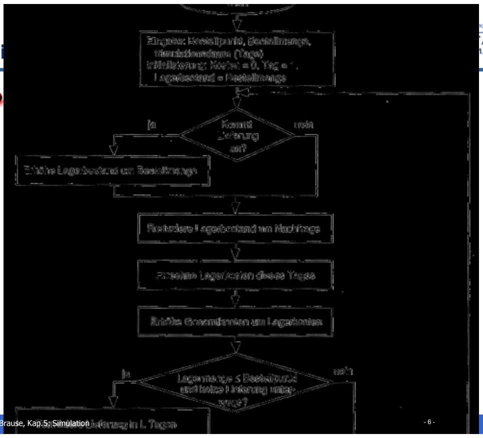
⇒ **$B^* = (2RF/H)^{1/2}$**

Ermittelbar auch durch diskrete Simulation !

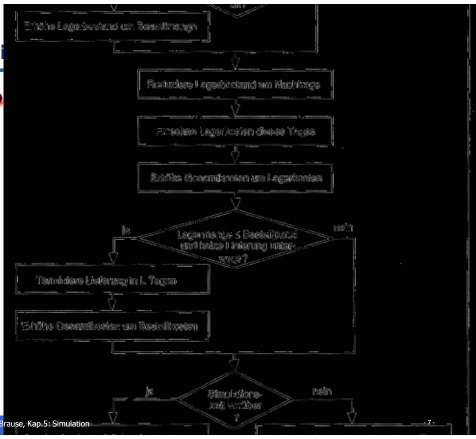
Diskrete Simulation

- **Flussdiagramm Liefersystem**





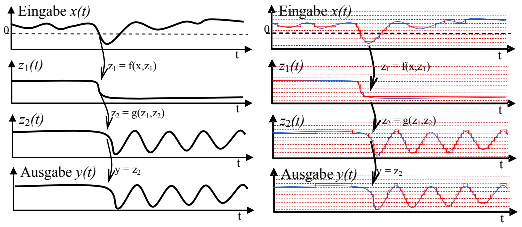
Di



Zeitkontin. vs. zeitdiskrete Simulation

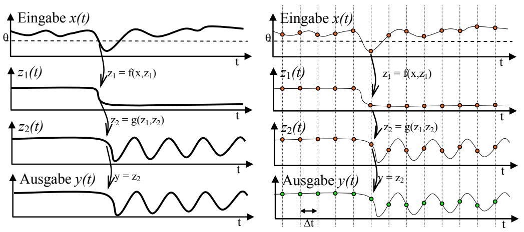
- **Realität: kontinuierlich**
- **Computer-Repräsentation:** Wert-Diskretisierung mit endl. Binärzahl

$$\Re \ni \mathbf{x} \rightarrow \Sigma_i b_i 2^i$$



Zeitkontin. vs. zeitdiskrete Simulation

- **Realität: kontinuierlich**
- **Zeit-Diskretisierung**



Sampling-Probleme

UNIVERSITÄT FRANKFURT AM MAIN

Gleiche Abtastfrequenz

Pro: Zeitsynchrone Abtastwerte

Contra:

- Hohe Samplingrate nötig bei „schnellen“ Änderungen.
- Nyquist-Theorem: höchster Frequenzanteil $f \leq f_s$ Samplingfrequenz
- Hohe Speicher- und CPU-Anforderungen bei hoher Anforderung an alle Sampling-Raten aller Variablen

Individuelle Abtastfrequenz

Pro: Geringerer Speicherbedarf

Geringer Aufwand (Geld, Belastung): Klein. Praxis!

Contra:

- Die Messzeitpunkte sind nicht identisch: Temp. Datenbanken !
- Messwerte neu konstruieren („Resampling“)
- „*Missing values*“ ergänzen (Zufallsverteilung)

Sampling-Probleme

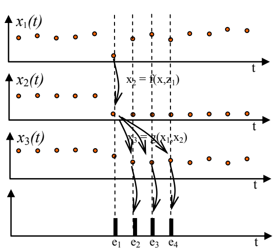
UNIVERSITÄT FRANKFURT AM MAIN

Beispiel: Variable für Sept. Schock-Diagnose

- 140 Var. Existieren – aber nur wenige sind häufig gemessen
- Blutdruck ist wichtig und häufig gemessen
- Thrombozytenzahl nur einmal pro Tag gemessen,
- Sepsis-Dynamik ist innerhalb 6 Std. aktiv, also geringere Diagnosefähigkeit als möglich.

Ereignisorientierte Simulation

UNIVERSITÄT FRANKFURT AM MAIN



Zeitliche Ordnung ⇔ Kausalordnung:
Zeit = Ordnungsindex

Ereignisorientierte Simulation

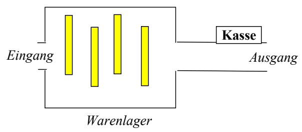
- **Beispiel: Supermarkt**

Eingabe x ∈ {Φ, a,b,c,...}

Einkaufszeit $T_{\alpha L}$

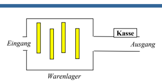
Warteschlange
- Länge z_w
- Zeit T_{aW}

Ausgabe y∈ {Φ, a,b,c,...}



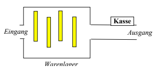
Ereignisorientierte Simulation

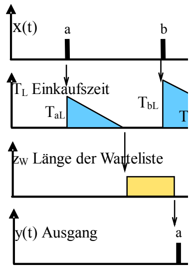
- **Beispiel: Supermarkt**



Ereignisorientierte Simulation

- **Beispiel: Supermarkt**

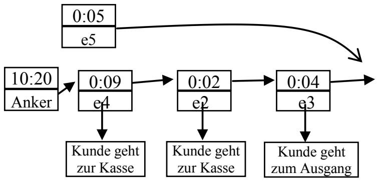




Zentrale Zeitachse

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

- **Ungeordnete Achse, Herunterzählen**

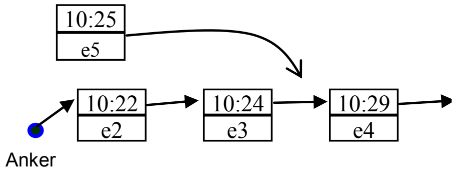


- Schnelles Einhängen
- Langsamer Zyklus

Zentrale Zeitachse

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

- **Geordnete Achse, Vergleich mit aktuell. Zeit**



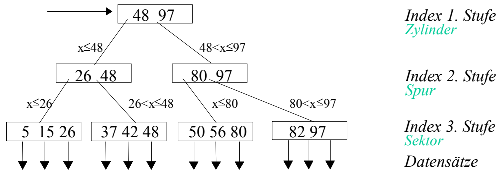
- Langsames Einhängen
- Schneller Zyklus

Indexsequentielle Dateien

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

- **Beispiel Einwohnerdatei,** *Schlüssel = Alter*
  - Auffüllen von Containern der Größe *m* mit Schlüsseln
  - Baumstruktur für indizierten Zugriff:
  - Index *n-1*. Stufe= {max (container $_{\text{Index n-ter Stufe}}$) }
  - Index-sequentielle Speicherung (Zylinder, Spur, Sektor)

Beispiel 12 Record-Schlüssel: 5, 15, 26, 42, 48, 56, 80, 82, 97, 37, 50



Problem: Eingliedern/Ausgliedern von Datensätzen in die Container (Datenblöcke) fester Größe, z.B. „41“

Stochastische Simulation

Kont.vs. diskrete Simulation

Erzeugung einer Verteilung

Monte-Carlo-Simulation

Simulationsrahmen

Auswertung

Stochast. Simulation

- **Eigenschaften von Zufallszahlen**

- **DEF Verteilungsfunktion P(w)**
- **DEF Dichtefunktion p(x)**
- **DEF Erwartungswert μ**
- **DEF Varianz σ^2**

Stochast. Simulation

- **Messung von Zufallszahlen**
  - Implementierung von Histogrammen
  - Fluch der Dimensionen

Stochast. Simulation

UNIVERSITÄT FRANKFURT AM MAIN

Erzeugung von Zufallszahlen

Forderungen

- Unabhängigkeit — *$P(x_i,x_j) = P(x_i)P(x_j)$*
- Konstante Dichte bei U(0,1) — *$p(x) = c$*
- Hohe Besetzungsdichte — *keine Lücken zwischen mögl. Werten*
- Reproduzierbarkeit — *Pseudo-Zufallszahlengenerator*

Linearer kongruenter Generator LCS

UNIVERSITÄT FRANKFURT AM MAIN

Historisch
von Neumann, Metropolis 1940

$X_t = 7\,1\,8\,2 \rightarrow (X_t)^2 = 5\,1\,\boxed{5\,8\,1\,1}\,2\,4$

$X_t = 5\,8\,1\,1 \rightarrow (X_t)^2 = 3\,3\,\boxed{7\,6\,7\,7}\,2\,1$

$X_t = 7\,6\,7\,7 \rightarrow (X_t)^2 = \ldots$

Linearform LCS

$x_{t+1} = (ax_t + b) \textbf{ mod } m$ — *linearer Kongruenter Generator LCG*

$n \textbf{ mod } m := n - \left\lfloor \frac{n}{m} \right\rfloor m$ — *n, m ganze Zahlen > 0*

Linearer kongruenter Generator LCS

UNIVERSITÄT FRANKFURT AM MAIN

Unabhängigkeit
Nicht gegeben bei gleichen Sequenzen

Besetzungsdichte
Problematisch bei schlechten Werten für a, b, m

Forderungen:
- ggT(b,m) = 1
- WENN q Primzahl in m, DANN auch in a-1
- WENN 4 in m, DANN auch in a-1

Implementierung
Modulo-Bildung durch Löschen der k obersten Bits
Erzeugung von x durch Shift und Addition

Verallgemeinerung
$x_{t+1} = g(x_t, x_{t-1}, x_{t-2}, \ldots) \textbf{ mod } m$

Quadratische Generator, multiple-rekursive Generator

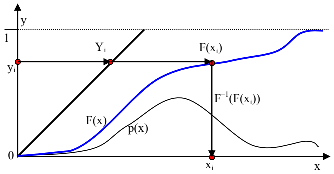
Erzeugung einer Verteilung

- **Gegeben:** Zufallszahl einer Verteilung
- **Gesucht:**
  - Zufallszahl mit neuem Erwartungswert
    $x' = x + \mu$
  - Zufallszahl mit neuer Varianz
    $x' = (x - \mu)\sigma + \mu$
  - Zufallszahl mit neuen Intervallgrenzen x_0', x_1'
    $x' = x_0' + (x - x_0)\,(x_1' - x_0')/\,(x_1 - x_0)$

Erzeugung einer Verteilung

- **Kontinuierliche Verteilung**

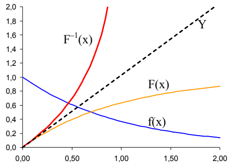
X verteilt nach F(x) mit Dichte p(x)
Y uniform verteilt im Intervall [0,1[



$x_i = F^{-1}(y_i)$

Erzeugung einer Verteilung

- **Beispiel:** Erzeugung einer Exponentialverteilung



$$p(x) = \begin{cases} \lambda e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$$

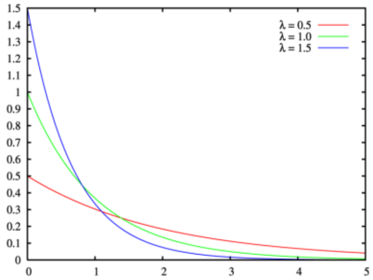
$$F(x) = \int_0^x \lambda e^{-\lambda t}\,dt = 1 - e^{-\lambda x}$$

$$x = F^{-1}(y) = -\frac{\ln(1-y)}{\lambda}$$

Verteilungen

- **Exponential-Verteilung** $f(x) = \begin{cases} \lambda e^{-\lambda x} & , x \geq 0, \\ 0 & , x < 0. \end{cases}$

$E(X) = \frac{1}{\lambda}$ = var(X)



Erzeugung

$T = \frac{-\ln U}{\lambda}$

Erzeugung einer Verteilung

- **Simuliere das Auftreten eines Ereignisses e mit P**
  Wenn x aus U(0,1), so ist P = Prob(x ≤ P)

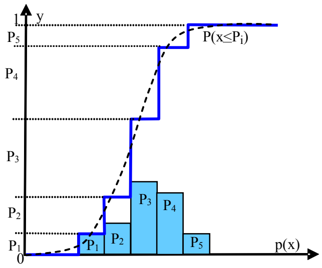
  Also
```
x = UniformRandom(0,1)
IF x ≤ Pi THEN SimulateEvent(i)
```

- **Simuliere das Auftreten von vielen Ereignissen**, dh. eine diskrete Verteilung $P_1, P_2, \ldots, P_n$

Wie ??

Erzeugung einer Verteilung

- **Erzeugung einer diskreten Verteilung**



Erzeugung einer Verteilung

- **Simulation von zufälligen Ereignissen $e_1, e_2, \ldots, e_n$**

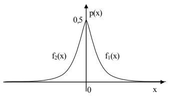
Ein Ereignis mit $P_i = P(x \leq P_i)$ simuliert durch uniforme Zufallszahl aus einem P_i breiten Unter-Intervall von [0,1]

$[0,P_1],]P_1,P_2], \ldots ,]P_{n-1},1]$

```
x = UniformRandom (0,1); e = 0; i = 1;
WHILE i < n DO
  IF x≤A[i]        THEN e = i ; i = n      (*Abbruch*)
                   ELSE i=i+1
ENDWHILE
IF e = 0 THEN e = n;                       (*letztes Intervall*)
SimulateEvent(e);
```


Erzeugung einer Verteilung

- **Zusammensetzen von Verteilungen**



Beispiel: *Laplace-Verteilung*

$$p(x) = \tfrac{1}{2} e^{-|x|} \quad \text{oder} \quad \begin{cases} \frac{1}{2} e^{-x} & \text{für } x \geq 0 \\ \frac{1}{2} e^{+x} & \text{für } x < 0 \end{cases}$$

$$p(x) = \tfrac{1}{2} f_1(x) + \tfrac{1}{2} f_2(x)$$

$$\text{mit} \quad f_1(x) = \begin{cases} e^{-x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases} \qquad f_2(x) = \begin{cases} e^{x} & \text{für } x < 0 \\ 0 & \text{sonst} \end{cases}$$

$$p(x) = \sum_i P_i f_i(x) \quad \text{mit} \quad \sum_i P_i = 1$$

- **Überlagern von Verteilungen**

$$x = X_1 + X_2 + \ldots + X_m$$

X_i aus der selben Verteilung:
m-fache Konvolution eines X_i

Beispiel: *m-Erlang-Verteilung*

Erzeugung einer Verteilung

- **Erzeugen der Normalverteilung**

Problem: $$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

Verteilungsfkt. Φ hat keine analyt. Lösung, nur tabelliert

$\Rightarrow$ **Inverse Funktion ex. nicht**

Lösung: Überlagern von Verteilungen

Nutzung des Zentralen Grenzwertsatzes:

$$\lim_{n\to\infty} x(n) = N(0,1) \quad \text{mit} \quad x(n) = \sum_{i=1}^{n} \frac{(X_i - \mu)}{\bar{\sigma}}$$

mit beliebig verteilten X_i und gemeinsamen $\mu = \langle X_i \rangle$ und $\bar{\sigma}$

X_i uniform aus U(-1/2 ,1/2) $\Rightarrow$ N(0,1) durch n = 12

Erzeugung einer Normalverteilung

- **x_i = #Grauwerte *i* im Bild pdf = Histogramm**


x_i^a


x_i^b


x_i^c


x_i aus PixelMix

Erzeugung einer Normalverteilung


p_i^a [redacted]

0 [redacted] 255 i

\+ 
p_i^b [redacted]

\+ 
p_i^c [redacted]

= 
p_i = $p(x_{mn}=i)$ [redacted] $_{mn}$ + x^c_{mn}

Erzeugung einer Normalverteilung

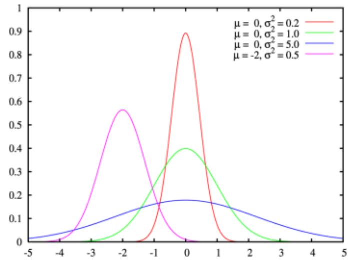
- **Pseudocode zur Erzeugung einer Normalverteilung**

```
PROCEDURE GaussRandom(mu,sigma:REAL):REAL;
      (* generiert einen normalverteilten Zufallswert
        mit Mittelwert μ und Streuung σ*)
VAR n,sum:REAL;
BEGIN
      sum:=0;
      FOR i:=1 TO 12 DO
       sum :=sum + UniformRandom() - 0.5
      END;
RETURN sum*sigma + mu;
```

- **Anwendung:** χ^2-Verteilung, Fisher F(m,n)-Verteilung

Verteilungen

- **Normal-Verteilung** $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$



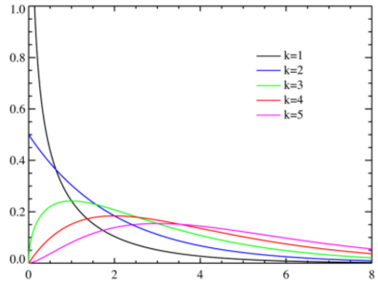
Verteilungen

- **χ^2-Verteilung** $f_k(x) = \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2}$

E(x) = k
Var(x) = 2k

Erzeugung

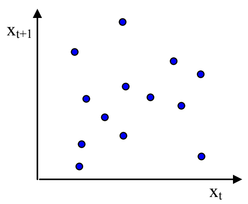
$Y = \sum_{m=1}^{k} X_m^2$



Testen einer Verteilung

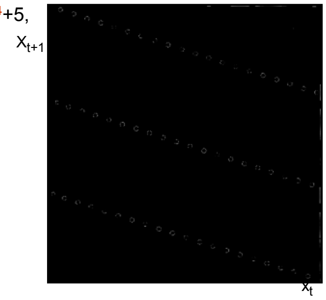
- **Grafische Tests auf Abhängigkeit**

Soll: uniforme Dichte



Testen einer Verteilung

- **LCG**
  $m= 2^8$, $a = 2^4+5$,
  $b = 1$



Testen einer Verteilung

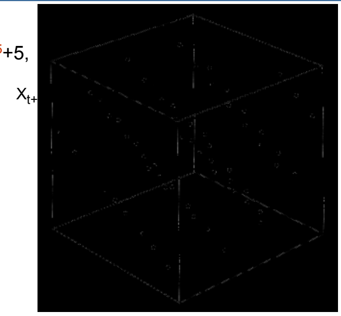
JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

- **LCG**
  $m= 2^8$, $a = 2^5+5$,
  $b = 1$

Testen einer Verteilung

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

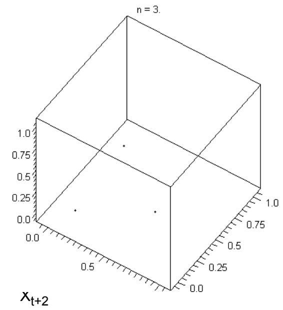
- **LCG**
  $m= 2^8$, $a = 2^5+5$,
  $b = 1$



Testen einer Verteilung

- **LCG**
  m= 2^{31}, a = 2^{16}+3,
  b = 0

FORTRAN
RANDU-Generator
n = 3, 9, 27, 81, 243, 729, 2.187, 6.561, 19.683, 59.049 - Punktdichten



JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Testen einer Verteilung

- **LCG**
  m= 2^{31}, a = 2^{16}+3,
  b = 0

FORTRAN
RANDU-Generator

n = 2000

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Testen einer Verteilung

- **RUN-Tests:** *Erzeugung einer Zeichenkette aus 0,1*

  Beispiel: ansteigend=1 / abfallend=0

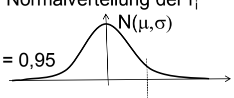
  $5 \rightarrow 7 \rightarrow 8 \rightarrow 3 \rightarrow 2 \rightarrow 6$
  1 1 0 0 1

  $s_i = \begin{cases} 1 & x_t \geq x_{t-1} \\ 0 & x_t < x_{t-1} \end{cases}$

  s = $s_1 s_2 s_3 \ldots s_t$, Unterscheid. von Unterketten r_i mit gleich. Zeichen

  Test: mittl. Kettenlänge $\mu = \langle r_i \rangle = \frac{2N-1}{3}$, Varianz $\sigma^2(r) = \frac{16N-29}{90}$
  Test auf Normalverteilung der r_i

  $P(r \leq r_\alpha) = 0{,}95$



Sind 95% der $r_i < r_\alpha$?

Testen einer Verteilung

- **RUN-Tests:** *Erzeugung einer Zeichenkette aus 0,1*

  Beispiel: größer ½ =1 / kleiner =0

  0,5 → 0,7 → 0,8 → 0,3 → 0,2 → 0,6
  1 1 1 0 0 1

  $s_i = \begin{cases} 1 & x_t \geq 0{,}5 \\ 0 & x_t < 0{,}5 \end{cases}$

  $s = s_1 s_2 s_3 \dots s_t$, Unterscheid. von Unterketten r_i mit gleich. Zeichen

  Test:

  $P(s_i=0) = P(s_i=1) = ½ \Rightarrow P(n \text{ gleiche Zeichen}) = (½)^n$

  Konvergiert P(.) zu $(½)^n$ für alle n?

Testen einer Verteilung

- **Anpassungs-Test**

  Gegeben: „uniformer" Generator mit p(x)

  Test des Histogramms aus *m* Intervallen, *n* Ereignissen

  Soll: P_i = n/m pro Intervall

  Gemessen: Varianz $\chi^2 = \frac{m}{n} \sum_{i=1}^{m} \left(P_i - \frac{n}{m} \right)^2$

  χ^2-Verteilung mit m Freiheitsgraden

  Test: Ist die Varianz zu groß?

  Ist $\chi^2 > \chi^2_{m-1,\alpha}$ mit α Akzeptanzwahrscheinlichkeit ?

  wobei $\chi^2_{m-1,\alpha} \approx (m-1)\,(1 - a + x_\alpha\, a^{1/2})^3$ mit $a = \frac{2}{9(m-1)}$

  und x_α korresp. kritischer Punkt einer $N(0,1)$

Testen einer Verteilung

- ***d*-dimensionaler Anpassungs-Test**

  Gegeben: „uniformer" Generator mit p(x)

  Test des *d*-dim. Histogramms aus *m* Intervallen/dim, *n* Ereignissen

  Soll: $P_{ijk..} = n/m^d$ pro Intervall

  Gemessen: Varianz $\chi^2(d) = \frac{m^d}{n} \sum_{i,j,k,...=1}^{m} \left(P_{ijk...} - \frac{n}{m^d} \right)^2$

  Test: Ist die Varianz zu groß?

  Ist $\chi^2(d) > \chi^2_{m-1,\alpha}$ mit α Akzeptanzwahrscheinlichkeit ?

  Beispiel: $\chi^2_{m-1,\alpha}$ = 4.211,4 für α = 0,9

  RANDU-Generator: $\chi^2(1)$ = 4.202,0 < 4211,4

  $\chi^2(2)$ = 4.202,3 < 4211,4

  $\chi^2(3)$ = 16.252,3 > 4211,4 schlechte Leistung!

Verteilungen

- **Gamma-Verteilung** $f(x|k,\theta) = x^{k-1}\frac{e^{-x/\theta}}{\theta^k\,\Gamma(k)}$ for $x > 0$

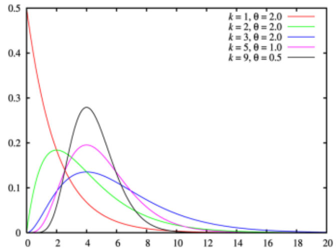
$$\theta = \frac{1}{kN}\sum_{i=1}^{N} x_i$$

Def. Gamma-Funktion

$$\Gamma(z+1) = z!$$
$$\Gamma(z+1) = z\,\Gamma(z)\,.$$

Erzeugung

$$\sum_{k=1}^{n} -\ln U_k \sim \text{Gamma}(n,1)$$

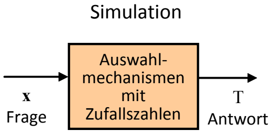


Stochastische Simulation

- Kont.vs. diskrete Simulation
- Erzeugung einer Verteilung
- **Monte-Carlo-Simulation**
- Simulationsrahmen
- Auswertung

Monte-Carlo-Simulation

- **Leistungsschema**

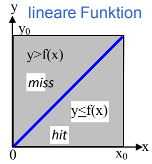


Hit-and-miss Simulation

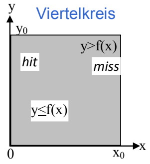
- Simuliere alle Situationen
- Selektiere die „günstigen" mit Hilfe von Auswahlregeln

Beispiel: Integralmessung durch Simulation

lineare Funktion

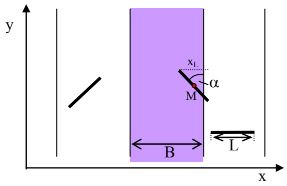


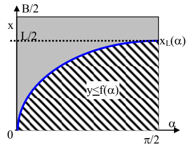
Viertelkreis



Hit-and-miss Simulation: Buffons Problem

Mit welcher Wahrscheinlichkeit *P* kreuzt ein geworfener Stab *L* eine Streifengrenze *B* ?





P = 2L / Bπ

Test auf Genauigkeit Simulation vs. analyt. Ergebnis nötig

Test der Resultate

Gegeben: *N* Zufallsexperimente, Trefferrate T_i

$$\hat{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$$ *Schätzung für Erwartungswert (Trefferrate)*

mit experimentell beobachteter Varianz

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(T_i - \hat{T}\right)^2$$

Frage: Ist $\hat{T}$ normalverteilt um vermutete Trefferrate T_μ?

Test: Liegt $\hat{T}$ im Konfidenzintervall $[\hat{T} - x_\alpha, \hat{T} + x_\alpha]$?

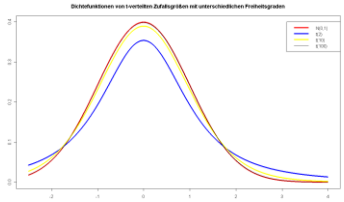
wobei $x_\alpha = N_\alpha \frac{s}{\sqrt{n}}$ *Umrechnung der Grenzen des Konfidenzintervalls* $[-N_\alpha, N_\alpha]$ *von* N(0,1)

JA $\Rightarrow$ $\hat{T}$ ist mit Wahrscheinlichkeit 1-2α um T_μ normalverteilt.
Bei n<40 *t*-Verteilung benutzen!

Verteilungen

- **Studentische t-Verteilung mit *n* Freiheitsgraden**

Stud. t-Verteilung $f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$ $\qquad \Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}\,\mathrm{d}t$



Monte-Carlo-Simulation

- ***Crude* Monte Carlo**
  Übergang auf Gesamtmenge der Zufallszahlen durch Veränderung des Ziels:

$$I = \int f(x)dx = \int \tilde{f}(x)p(x)dx$$

so dass

$$I = \langle \tilde{f}(x) \rangle = \frac{1}{n}\sum_{i=1}^{n} \tilde{f}(x_i) \quad \text{wobei } x_i \in p(x)$$

Stochastische Simulation

Kont.vs. diskrete Simulation

Erzeugung einer Verteilung

Monte-Carlo-Simulation

Simulationsrahmen

Auswertung

Simulationsrahmen

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Problemidentifikation

- **Anwender-Terminologie** ist komplex und undurchsichtig.
- Die **Wahrnehmung** des Problems ist stark von unterschiedlichen Interessen des Auftraggebers bestimmt.
- Es ist schwer, von dem, was die Auftraggeber als Problem ansehen, zu dem **wirklichen Problem** durchzudringen. Wichtige Aspekte anderer Anwender werden vernachlässigt.
- Sind wir zu selbstsicher, so führt ein "**Schnellschuss**" leicht zu einer realitätsfernen Lösung.
- Auch **übermäßige Betroffenheit** bewirkt eine "betriebsblinde" Lösung.

Simulationsrahmen

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Modellspezifikation

- **Anfang**: Glossar mit Grundbegriffen und Beschreibung der Bedeutung und Wechselwirkung
- **Zweiter Schritt**: formale Modelle für die Wechselwirkungen (UML-Diagramme, Modellierung der Eingaben,…)
- Nach der Modellierung folgt die **Evaluation** des Modells.

Spiralenmodell: Radius ~ Detaillierungsgrad

Simulationsrahmen

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Modelldetallierung

- Detaillierung im top-down-Modus
- Nur soweit nötig zur Beantwortung der Fragen
- Nur soweit ein geplantes System schon festgelegt ist

Probleme der Detaillierung

- Zu detailliertes Modell ("*brute force*"): Anzahl der möglichen Simulationsläufe ist aus Rechenzeitgründen begrenzt ➔unsignifikante stoch. Ergebnisse.
- Verhalten des Gesamtsystems wird schwer verständlich und validierbar ➔ Fehler in der Modellierung werden schwer erkannt
- Fertigstellung des Modells wird stark verzögert, so dass das Projekt erfolglos abgebrochen werden muss.

Simulationsrahmen

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Integration realen Systemverhaltens

Parameterschätzungen basieren auf Eingabegrößen
einer diskreten, ereignisorientierten Simulation:

- die Frequenz oder der Zeitabstand auftretender Ereignisse, etwa zwischen Eingabe und Ausgabe
- eine nachgefragte Menge (Ausgabe) pro Zeiteinheit
- die Zeitdauern für die Dienstleistung in Warteschlangen: Bedienung von Menschen, Fertigung von Waren

Warteschlangen im System:
Abweichung vom klassischen Warteschlangenmodell ?

- Gibt es eine zentrale Warteschlange für mehrere Bedienstationen, oder hat jede Station ihre eigene?
- Ist es möglich, sich vorzudrängeln (*cheating*), die Warteschlange plötzlich zu wechseln (*swopping*), das Warten vorzeitig abzubrechen (*reneging*) oder angesichts einer vollen Warteschlange sich gar nicht erst anzustellen (*balking*)?

Simulationsrahmen

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

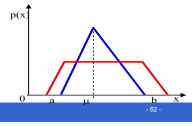
Modellierung der Eingabeverteilungen - Wie?

Verteilung passt am Besten zu den gemessenen Daten
Oder **Annahmen** entsprechen inhaltlich am Besten dem Modell:

- reelle Variable, die als ***Summe*** von unabhängigen Zufallsgrößen entsteht $\Rightarrow$ Normalverteilung.
- beobachtete Variable x als ***Produkt*** unabhängiger Zufallsgrößen y_i, so ist der Logarithmus der beobachteten Variable

$$\ln x = \ln \prod_{i=1}^{m} y_i = \sum_{i=1}^{m} \ln y_i = \sum_{i=1}^{n} X_i \qquad \text{log-}\textit{normalverteilt}$$

- Ex. keine Daten über die fragliche Eingabevariable (System ex. nicht, zu selten, zu wenig Zeit)
Befragung von Experten:
Ermittlung von min, max, Mittelwert ;
Schätzung der Verteilung mit Vertrauensintervall



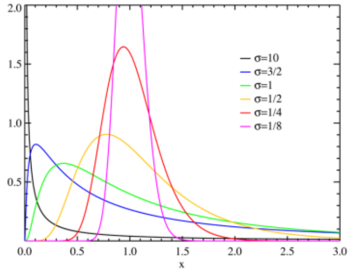
Verteilungen

JOHANN WOLFGANG GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Log-normal-Verteilung $f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x-\mu)^2/2\sigma^2}$

$$\mathrm{E}(X) = e^{\mu+\sigma^2/2}$$

$$\mathrm{var}(X) = (e^{\sigma^2}-1)e^{2\mu+\sigma^2}$$



Simulationsrahmen

- **Modellierung von Ankunftsprozessen**
  Einzelereignisse unabhängig, aber ex. zeitl. Abhängigkeiten

  - **Autoregressive Prozesse** *Frühere Werte AR(N)*
    $x_i = \mu + a_1(x_{i-1}-\mu) + \ldots + a_N(x_{i-N}-\mu) + \varepsilon_i$
    μ = gleit. Erwartungswert, ε normalverteilt **ARMA-Prozess**
    a_i über Parameteranpassung(Box-Jenkins)
    Stationäre Verteilung $F(x_i)$ beobachtet: Simulation von x_i durch
    $x_i = F^{-1}(\Phi(z_i))$, z_i normalverteilt mit ARMA **ARTA-Prozess**

  - **Ankunftsprozesse** *Poisson-Prozess*
    Ereignisse mit Zeiten t_0, t_1, …gegeben.
    DEF Ankunftsprozess = stoch. Prozess, gekennzeichnet durch Anzahl der Ankunftsereignisse t_i, mit $t_0 < t_i \leq t$

Ankunftsprozesse

- **Binominalverteilung**
  p = Prob(Im Zeitabschnitt Δt kommt ein Kunde)
  Verteilung von *k* Kunden in *n* Zeitabschnitten = Binomialverteilung

- **Poissonprozess**
  $\lim_{\substack{\Delta t \to 0 \\ n \to \infty}}$ Binomialverteilung = Poisson-Verteilung

Binomialverteilung

- **Def. Bernoulli-Prozess**

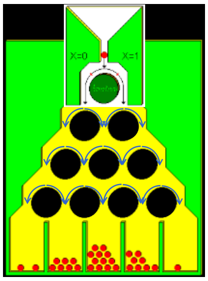
Stochast. Prozess, bei dem

2 Ereignisse bei jedem Versuch möglich sind

z.B. {Erfolg, Misserfolg}, {Zahl,Wappen}, {1,0}, {Wahr,Falsch},...

mit p bzw. q = 1-p

Galton-Brett



Binomialverteilung

Bei n Würfen ergibt dies eine Kette von n Zeichen (Entscheidungen), von denen k gleich sind.

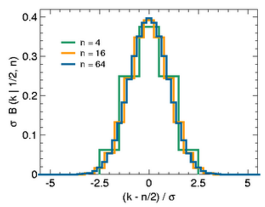
- Die Anzahl von k gleichen Zeichen in unterschiedlicher Reihenfolge ist $\binom{n}{k}$
- Die Wahrscheinlichkeit für k gleiche Zeichen ist $p^k q^{n-k}$

Also :

$B(k,n) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{(n-k)!k!} p^k q^{n-k}$ **Binomialverteilung**

Binomialverteilung

- **Übergang zur Normalverteilung** bei $n \to \infty$



Zentrierte und skalierte Verteilung

Poisson-Verteilung

- **Grenzwert der Binomialverteilung:** $p = \frac{\lambda}{n}$

$$\lim_{n\to\infty} P(X=k) = \lim_{n\to\infty} \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$= \lim_{n\to\infty} \Big(\frac{(n-k)! \cdot}{}$ [redacted] $\frac{}{(n-k)!n^k} \Big) \left(\frac{\lambda^k}{k!}\right) \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}$

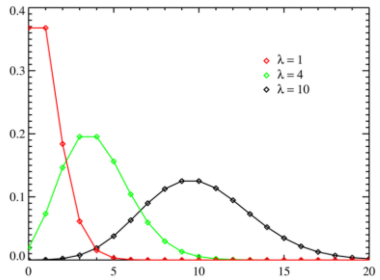
$$= \lim_{n\to\infty} \underbrace{\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)}_{=1} \underbrace{\left(\frac{\lambda^k}{k!}\right)}_{=\frac{\lambda^k}{k!}} \underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{=e^{-\lambda}} \underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{=1}$$

$P_\lambda(n) = \frac{\lambda^n}{n!} e^{-\lambda}$ Poisson-Verteilung

Verteilungen

- **Poisson-Verteilung** $P(N=k)=\frac{e^{-\lambda}\lambda^k}{k!},$

$E(X) = \lambda = var(X)$



Poisson-Prozess

Sei die Anzahl der Kunden, die im Zeitabschnitt [t, t+s] ankommen, mit $k = N(t+s)–N(t)$ bezeichnet.

Dann gelten folgende 3 Bedingungen für einen Poisson-Prozess:

1. Die Kunden kommen nach einander an.
2. Die Anzahl *k* ist unabhängig von der Anzahl N(u), $0 \leq u \leq t$, die vorher angekommen waren.
3. Die Verteilung der Kundenzahl *k* ist unabhängig von *t* (*Stationarität*).

- **Satz**

  $P(k) = \frac{(\lambda s)^k e^{-\lambda s}}{k!}$ mit $\begin{array}{l} k = N(t+s)-N(t) \\ k = 0,1,2,..., \quad t,s \geq 0 \end{array}$

  und $\langle N(s)\rangle = \lambda s.$

Poisson-Prozess

- **Erwartungswert** $\langle N(s)\rangle = \lambda s$, also in s=1 Zeit sind λ Kunden bzw. Zeit pro Kunde $1/\lambda$.

- **Satz**

  Ist $\{N(t), t \geq 0\}$ ein Poisson-Prozess mit der Rate λ, so sind die Zwischenankunftszeiten Δt_i exponentiell verteilt mit dem Erwartungswert $1/\lambda$.

Die Umkehrung des Satzes ist ebenfalls korrekt:
Sind die Δt_i exponentiell verteilt, so liegt ein Poisson-Prozess vor.

- **Problem**: Schätzung der $\lambda(t)$

Poisson-Prozess

- **Schätzen der Parameter**

  Beispiel: *Kundenankunft im Laden*

	Montag	Dienstag	Mittwoch	Donn.	Freitag	
8- 9 Uhr						
10-11 Uhr	9	8	11	10	13	*k(t=10)*
11-12 Uhr						
12-13 Uhr						
13-14 Uhr	5	7	3	5	10	*k(t=13)*
14-15 Uhr						
15-16 Uhr						

Problem: Gruppen von Kunden

Lösung: Verbundener Poisson-Prozess

Stochastische Simulation

Kont.vs. diskrete Simulation

Erzeugung einer Verteilung

Monte-Carlo-Simulation

Simulationsrahmen

Auswertung

Simulationsauswertung

- **Fragestellung der Simulation = Spezifikation**

  Auswertung der Simulation gemäß der Fragestellung, die zu Beginn festgelegt sein muss.

  Aufgabe: Definition der Leistungsmaße

  Beispiel: *Verkehrsströme*

  Auslastung der Verkehrsmittel + Strassen

  Aber: Auslastung = stochastisch.

  Welche Verteilung, Varianz wird toleriert?

Simulationsauswertung

- **Beispiel:** Vergleich von zwei Computersystemen

  System A Leistung = 500 Transakt./sek, Preis= 500€
  1 Warteschlange, 2 Server

  System B Leistung = 1000 Transakt./sek, Preis=1000€
  1 Warteschlange, 1 Server

  Welches soll man kaufen?

  Welches System hat kürzere Wartezeiten?

  Bekannt:
  λ Jobs/Minute, Poisson-Prozess
  Servicezeit T exp. verteilt pro Server mit $\langle T(A) \rangle$ = 1,8 Min.
  $\langle T(B) \rangle$ = 0,9 Min.

Simulationsauswertung

- **Vergleich zweier Systeme** *Beispiel Warteschlangen*

  Ergebnis *Warteschlangentheorie*

  Mittl. Verzögerung System A = 3,7 Min,
  System B = 4,13 Min.

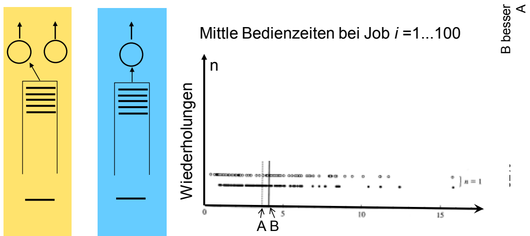
  System A besser, weil zwei Server parallel arbeiten können und so Lastspitzen schneller abbauen.

  Simulationsergebnis ?

Simulationsauswertung

- **Vergleich zweier Systeme** *Simulation der Warteschlangen*

System A *System B* *System A* = O, *System B* = ●



Varianzreduktion durch n-fache Mittelung pro Job *i*

Simulationsauswertung

Problem *Varianz bei 2 Systemen A,B*
beobachtete Varianz zu groß für eindeutige Entscheidung

$$\text{var}(x_A - x_B) = \text{var}(x_A) + \text{var}(x_B) - 2\text{cov}(x_A,x_B)$$

Kleinere Kovarianz $\Rightarrow$ kleinere Varianz.

Wie?
Ergebnisse korrelieren lassen durch gemeinsame Zufallszahlen, z.B. gleicher Startwert des Zufallszahlengenerators bei beiden Systemen A und B

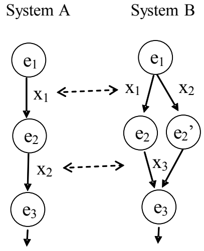
Antithetische Zufallsvariable

Problem: hohe Ergebnis-Varianz bei einem System

Verkleinere beobachtete Varianz durch:
- Bilde Paare von Wiederholungen für dasselbe System
- Wähle negativ korrelierte Zufallszahlen (antithetische Zufallszahlen), sodass Auswirkungen gegenläufig werden

Antithetische Zufallsvariable durch
- beim Wiederholungslauf für jede Zufallsvariable $u \in U(0,1)$ die Variable $1-u$ verwenden

Problem Antithetische Zufallsvariable nur möglich bei gleichen Systemen



Ungleiche Anzahl von Variablen $\Rightarrow$ keine korrespondierende Zufallszahlen

Simulationsauswertung

- **Synchronisierung von Zufallsvariablen**

Synchronisierung zweier Ströme durch

- Zusätzliche Zufallszahlen; Abstimmung der Programmierungen
- Stromzuordnung (*stream dedication*)
  Getrennte Generatoren für verschiedene Zwecke,
  z.B. Ankunftszeiten unabh. von Flussentscheidungen
- Speichern der Zufallszahlen von A und Wiederverwendung in B

Problem: Beeinflussen unterschiedl. Strategien den Zeitablauf, so sind Korrelationen nicht mehr da (Vergrößerung der Varianz!)

Simulationsauswertung

- **Prüfen der Testergebnisse: Konfidenzintervalle**

Vergleich der Ergebnisse y_A von System A mit y_B von System B durch

Definition $z := y_A - y_B$
und Test $\langle z \rangle = 0$?
Ist $\langle z \rangle$ aus Konfidenzintervall $[z_\alpha, z_{1-\alpha}]$?

Vergleich simuliertes System A vs. reales System B:
Unterschiedliche Zahlen n_1, n_2 von Messungen
→approximierte Anzahl von Freiheitsgraden der t-Verteilung
→*Welch-Konfidenzintervall*

Vielen Dank für Ihre Aufmerksamkeit!

Fragen ?