

5. Statistische Grundlagen der Simulation

5.1 Zufallszahlenerzeugung

- Der Begriff der Zufallszahl und ihre Bedeutung in der Simulation
- Erzeugung gleichverteilter Zufallszahlen
- Zufallszahlen beliebiger Verteilungen

5.2 Zur Notwendigkeit der Wiederholung von Simulationsexperimenten

5.3 Anlaufphase und Schätzgenauigkeit der Simulationsergebnisse

Literatur:

Page, Diskrete Simulation, Kap. 4.1, 4.4.1-4.4.4

Kosturiak/Gregor, Simulation von Produktionssystemen, Kap. 3

Oakshott, Business Modelling and Simulation, Kap. 3, 7, 10

Statistische Verfahren in der Simulation

- **Stochastische Systemsimulation:**
- **Zufallsexperimente =>
Anwendung statistischer Verfahren**
- (1) Erzeugung von Zufallsprozessen bestimmter statistischer Verteilungen (Input)**
 - » Bestimmung des statistischen Verteilungstyps
- (2) Statistische Auswertung der stochastischen Simulationsergebnisse (Output)**

5.1 Zufallszahlenerzeugung

- **Begriff der Zufallszahl und ihre Bedeutung in der Simulation**
 - Zufallszahlen zur Realisierung von **Zufallsvariablen**
 - Zuweisung eines Wertes aus dem Wertebereich an Variablen
 - Natürliche Zufallszahlen $\leq$ reale physikalische Prozesse
 - Beispiele:
 - » Werfen eines idealen Würfels
 - Augenzahl
 - » Zerfall radioaktiver Substanzen
 - Anzahl der je Zeiteinheit zerfallenen Teilchen
- **Abspeicherung umfangreicher Tabellen natürlicher Zufallszahlen erforderlich**
 - Ineffizienz $\Rightarrow$ Pseudo-Zufallszahlen $\Rightarrow$ Zufallszahlengeneratoren

Zufallszahlenerzeugung

- **Vorteil von Pseudo-Zufallszahlen**
 - Reproduzierbarkeit
- **Ausgangspunkt der Zufallszahlenerzeugung beliebiger statistischer Verteilungstypen**
 - Gleichverteilte Zufallszahlen
 - Gleichverteilung im Intervall (a, b)

$$F(X) = 0 \quad \text{für } X \leq a$$

$$F(X) = \frac{X-a}{b-a} \quad \text{für } a < X < b$$

$$F(X) = 1 \quad \text{für } X \geq b$$

$$E(X) = \frac{a+b}{2} \quad \text{bzw.} \quad \text{Var}(X) = \frac{(b-a)^2}{12}$$

Zufallszahlenerzeugung

• Multiplikative Kongruenzmethode zur Zufallszahlenerzeugung

- Rekursive Rechenvorschrift

$$z_{i+1} = (A * z_i) \bmod P \quad \text{mit } i = 0, 1, \dots$$
- nächste Zufallszahl z_{i+1} aus vorangegangenen z_i durch Multiplikation mit A und durch Modulo - Division durch P
- ganzzahlige z_0 , A aus dem Bereich $[1, P-1]$
- Rechenbeispiel: $A=5$, $P=17$ und $z_0=5$

$$z_{i+1} = (5 * z_i) \bmod 17 \quad \text{mit } i = 0, 1, \dots$$

Multiplikative Kongruenzmethode zur Zufallszahlenerzeugung

Zahlenbeispiel : $A=5$, $P=17$, $Z_0=5$

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	...
A Z _i	25	40	30	65	70	10	50	80	60	45	55	20	15	75	35	5	25	...
Z _{i+1}	8	6	13	14	2	10	16	12	9	11	4	3	15	7	1	5	8	...

- Periodizität der Zahlenfolge im Intervall $[1, (P-1)]$
 » Wiederholung ab $z_i = z_0$
- Geeignete Wahl der Parameter P und A :
 - $P = 2^{31} - 1$ und $A = p^k$
 - mit p = Primzahl als Faktor von $P-1$ und k positiver Exponent
 - bei Rechengenauigkeit von 32 Bit

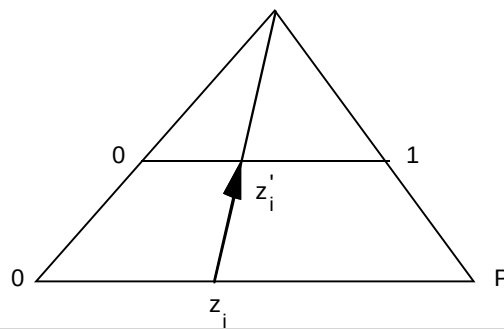
Multiplikative Kongruenzmethode zur Zufallszahlenerzeugung

- Division der z_i durch P

$$z'_i = \frac{z_i}{P} \quad (i=1,2,3,\dots)$$

» Intervall (0,1)

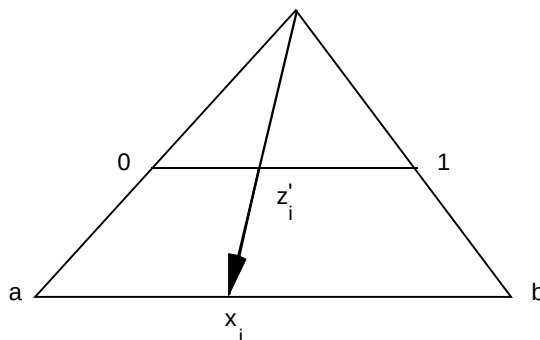
- Transformation zur Erzeugung von (0, 1)- gleichverteilten Zufallszahlen



Erzeugung gleichverteilter Zufallszahlen

- Transformation von im Intervall (a, b) gleichverteilten Zufallszahlen

$$\frac{x_i - a}{b - a} = z'_i \Rightarrow x_i = a + (b - a) z'_i$$



Zufallszahlenerzeugung

- **Weitere Eigenschaften von Generatoren für Pseudo-Zufallszahlen:**

- Stochastische Unabhängigkeit
- Zufälligkeit in der Folge (nicht stellenweise gehäuft)

- **Zufallszahlen beliebiger Verteilungen**

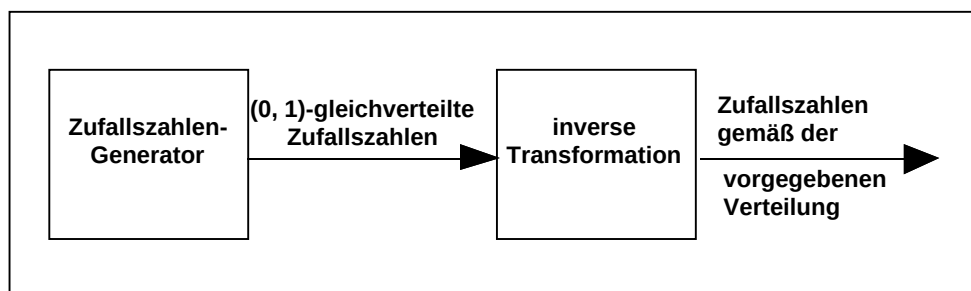
- Transformation von im Intervall (0,1) gleichverteilter Zufallszahlen $z_1, z_2, z_3, \dots$ in Zufallszahlen $x_1, x_2, x_3, \dots$, die nach Verteilungsfunktion $F(X)$ verteilt sind, durch Bildung der **Umkehrfunktion**, sofern existent

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

- Durch Inversion Gewinnung einer Berechnungsfunktion für entsprechend verteilte Zufallszahlen
- **Inverse Transformation**

Zufallszahlen beliebiger Verteilung

- **Erzeugung von Zufallszahlen mit vorgegebener Verteilung**



Zufallszahlen beliebiger Verteilungen

Exponentialverteilte Zufallszahlen

- Typische Verteilung für Zwischenankunftszeiten bzw. Bedienzeiten

- Theoretische Verteilungsfunktion

$$Z = 1 - e^{-\lambda X} = F(X) \quad \Rightarrow \quad X = -\frac{1}{\lambda} \ln(1 - Z)$$

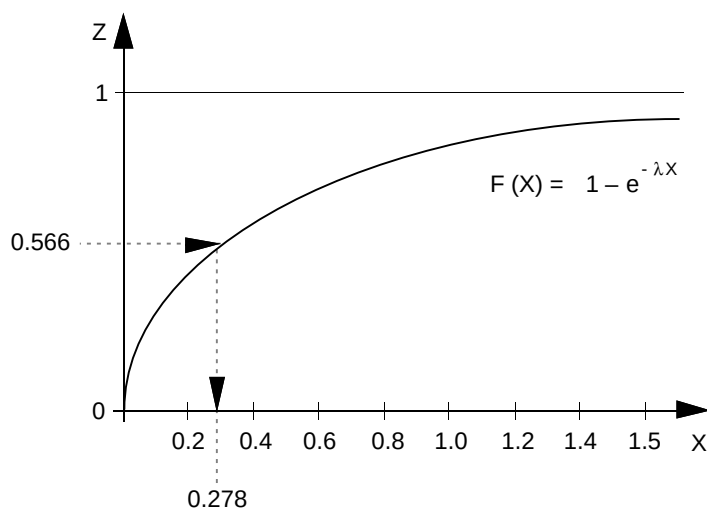
$$x_i = -\frac{1}{\lambda} \ln(1 - z_i)$$

- » mit z_i gleichverteilt in $(0, 1)$
- » Erwartungswert / Standardabweichung = $1 / \lambda$

Zufallszahlen beliebiger Verteilungen

- Exponentialverteilte Zufallszahlen

- Geometrische Veranschaulichung der Transformation gleichverteilter in exponentialverteilte Zufallszahlen



Zufallszahlen beliebiger Verteilungen

Erlang-verteilte Zufallszahlen

- Arithmetisches Mittel aus k unabhängigen, exponentialverteilten Zufallsvariablen mit gleichem Mittelwert
 - » Erlang-Verteilung der Ordnung k mit Parametern k und $m = 1/\lambda$
- Erzeugung durch Generierung k exponential-verteilter Zufallszahlen und Bildung des Mittelwertes

• Konstruktionsformel

$$x_j = - \frac{1}{k \times \lambda} \sum_{i=1}^k \ln(1 - z_i)$$

- » Mittelwert = $1 / \lambda$
- » Varianz = $1 / (k * \lambda^2)$

Zufallszahlen beliebiger Verteilungen

Normalverteilte Zufallszahlen

- Keine explizite Inversion der Verteilungsfunktion der Normalverteilung möglich
 - » Näherungsverfahren
- Zentraler Grenzwertsatz:
 - Summe von n identisch verteilten, unabhängigen Zufallsvariablen mit Mittelwert μ und Varianz σ^2 nähert sich der Normalverteilung mit $(n*\mu, n*\sigma^2)$ asymptotisch.
- Gute Approximation
 - n = 12 Zufallszahlen in (0,1)-gleichverteilt mit $\mu = 1/2$ und $\sigma^2 = 1/12$
 - » NV (6,1)

Zufallszahlen beliebiger Verteilungen

- **Normalverteilte Zufallszahlen**

- N (1,6)

- Standardisierung auf N (0,1)-Verteilung

$$z_j = \sum_{i=1}^{12} z_i - 6$$

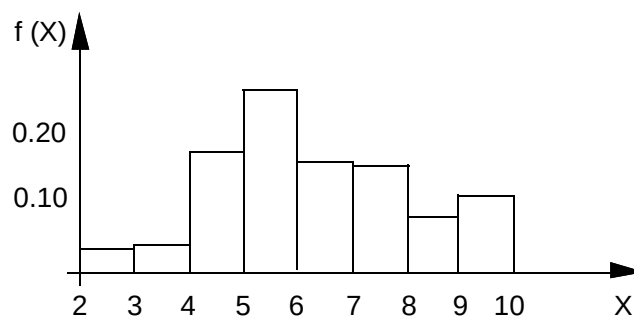
- Transformation auf beliebige N (μ , σ^2)-Verteilung

$$x_j = z_j \times \sigma + \mu$$

Zufallszahlen beliebiger Verteilungen

Zufallszahlen mit empirischer Verteilung

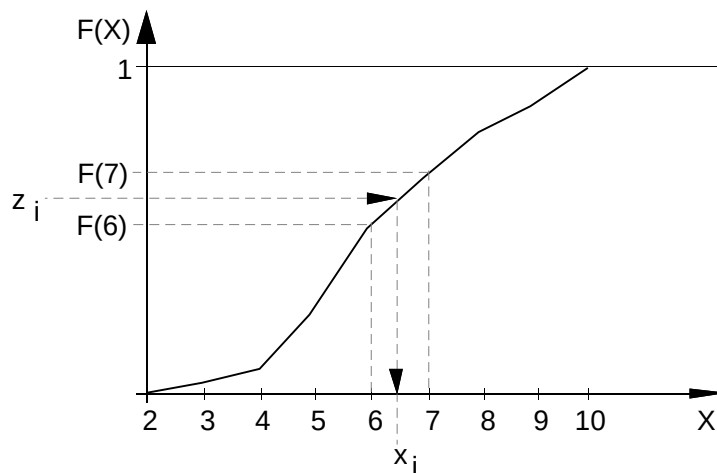
- Empirische Verteilung in Form eines Histogramms
- Annahme: Gleichverteilung in einer Klasse
- Histogramm der empirischen Verteilung
 - » relative Häufigkeit
 - » **Beispiel : Empirische Verteilung von 500 Beobachtungen**



Zufallszahlen beliebiger Verteilungen

Zufallszahlen mit empirischer Verteilung

- Verteilungsfunktion der empirischen Verteilung
 - » Kumulative Häufigkeit



Zufallszahlen beliebiger Verteilungen

• Zufallszahlen mit empirischer Verteilung

- Transformationsverfahren

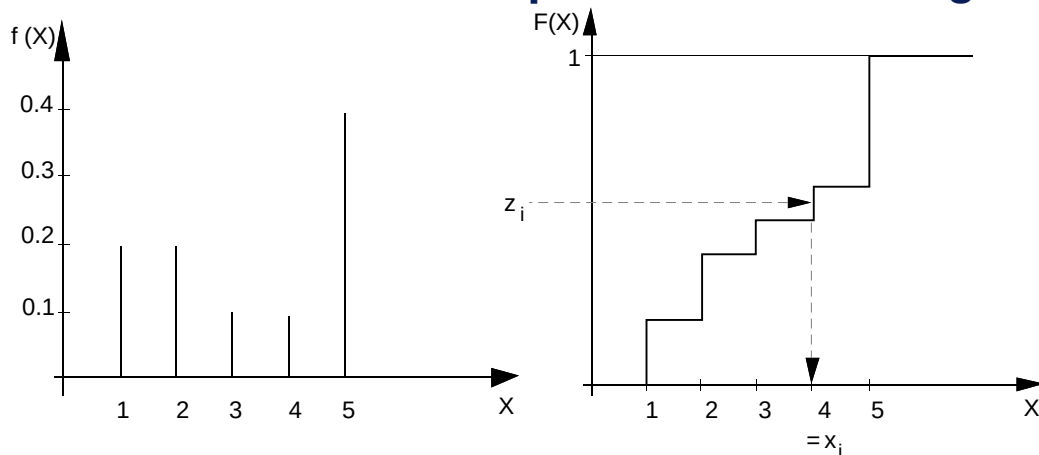
$$x_i = x_j + (z_i - F(x_j)) \frac{(x_{j+1} - x_j)}{(F(x_{j+1}) - F(x_j))}$$

• Zufallszahlen aus diskreten empirischen Verteilungen

- Beispiel: Diskrete Verteilung der Nachfragemenge in einem Lagerhaltungssystem

Zufallszahlen beliebiger Verteilungen

• Zufallszahlen diskreter empirischer Verteilungen



z_i	x_i
$0 < z_i \leq 0.2$	1
$0.2 < z_i \leq 0.4$	2
$0.4 < z_i \leq 0.5$	3
$0.5 < z_i \leq 0.6$	4
$0.6 < z_i \leq 1$	5

5.2 Zur Notwendigkeit der Wiederholung von Simulationsexperimenten

• Simulationsexperiment

- Simulierter Ablauf eines oder mehrerer stochastischer Prozesse über gewisse Zeitspanne

• Untersuchung der stochastischen Eigenschaften eines Simulationsprozesses

- Simulation des Prozessablaufes bis zu einem bestimmten Zeitpunkt und Analyse des sich einstellenden Ergebnisses
 - » Momentaufnahme eines ständig sich entwickelnden Prozesses
- Wiederholung dieser Momentaufnahmen in gewissen zeitlichen Abständen
 - » Bild über zeitlichen Ablauf des Prozesses

Zur Notwendigkeit der Wiederholung von Simulationsexperimenten

- **Beispiel für die Untersuchung der stochastischen Eigenschaften eines Simulationsprozesses**
 - Simulation eines Wartesystems bis t_0
 - » X_{t_0} = Warteschlangenlänge zum Zeitpunkt t_0
 - » Jede Simulation des Prozesses liefert genau einen Wert für X_{t_0}
 - » X_{t_0} zufallsabhängig
 - » Zufallsabhängig bedeutet hier abhängig von Zufallszahlen
 - » Zur Gewinnung genauerer statistischer Schätzwerte für das unbekannte X_{t_0} muss eine Stichprobe mit n Beobachtungen für X_{t_0} durchgeführt werden
- **Wiederholung von Simulationsexperimenten erforderlich !**

Zur Notwendigkeit der Wiederholung von Simulationsexperimenten

- **Nur über zahlreiche Wiederholungen eines Simulationsexperiments Überblick über Variationsbreite der Ergebniswerte X möglich**
 - Konfidenzintervall
 - Über wiederholte stochastische Simulationen vollständige Wahrscheinlichkeitsverteilung der Zufallsvariablen X abschätzbar
 - **Häufig nur Schätzung bestimmter Verteilungsparameter**
 - i.d.R. Mittelwert, Varianz
 - Geringerer Aufwand
- => Unabhängige Wiederholungen der Simulation zur Schätzung notwendig**

5.3 Anlaufphase und Schätzgenauigkeit der Simulationsergebnisse

Anfangszustand und Anlaufphase (Einschwingphase)

- Meistens Untersuchung **stationärer** Prozesse
 - $\neq$ transienter Zustand
- Stationärer stochastischer Prozeß X_t
 - Wkt von X_t unabhängig von t
 - **Beispiel : Warteschlangenlänge X_t**
 - » Wkt ($X_t = a$) = $f(a)$ für die gesamte Simulationszeit (für alle t)
- Konvergenz vieler stochastischer Prozesse gegen stationären Prozeß
- In der Praxis Erreichen des stationären Zustandes nach einer bestimmten Anlaufphase
 - „Einpendeln“ des Systems

Anfangszustand und Anlaufphase

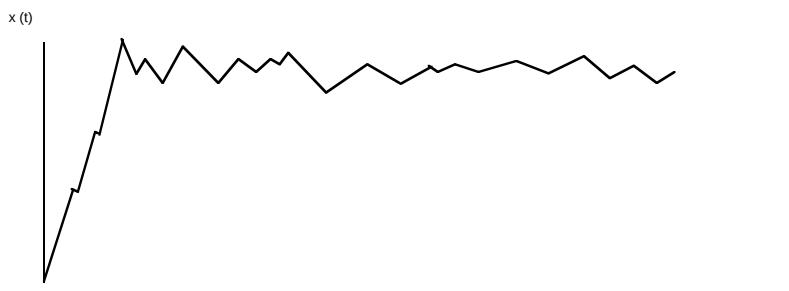
- Keine Berücksichtigung der Beobachtungen der Anlaufphase für die statistische Auswertung der Simulationsergebnisse
 - Beginn des eigentlichen Simulationsexperimentes erst nach Erreichen der stationären Phase
 - » Problem der Bestimmung von Anfangszustand und Anlaufphase
- Zu kurze Anlaufphase
 - » „ Verzerres “ Simulationsergebnis ("Bias")
- Zu lange Anlaufphase
 - » Überhöhter Rechenaufwand bzw. bei vorgegebener Simulationsdauer ungenauere Ergebnisse
- „ Guter “ Anfangszustand = > Kurze Anlaufphase

Anfangszustand und Anlaufphase

- **Keine exakten statistischen Verfahren zur Bestimmung der Anlaufphase verfügbar**
 - » Heuristische Ansätze
- **Heurist. Regeln zur Bestimmung der Anlaufphase**
 - (1) Regel nach Conway :**
 - Ist x_{i-1} Min / Max der Restfolge $x_i, x_{i+1}, \dots$
 - » x_{i-1} zur Anlaufphase
 - $x_i \neq \text{Min / Max von } x_{i+1}, x_{i+2} \dots$
 - » stationäre Phase beginnt
 - (2) Regel nach Tocher :**
 - Zyklische stochastische Prozesse, die ineinandergreifen
 - 3 - 4 Zyklen => Beginn der Stationären Phase

Anfangszustand und Anlaufphase

- **Visuelle Abschätzung der Anlaufphase**
- **Beispiele**
 - Ein nach der Einschwingphase stationärer Prozeß

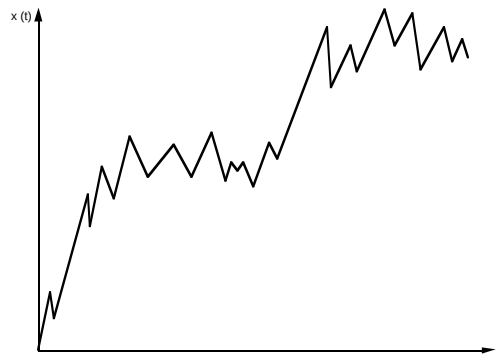


Anfangszustand und Anlaufphase

- Visuelle Abschätzung der Anlaufphase

- Beispiel:

- Zeitweise transiente Phase, z. B. in einer Verkehrs- simulation aufgrund einer "rush-hour"



Schätzgenauigkeit der Simulationsergebnisse - Konfidenzintervalle -

- Aufgabe der Simulation =

- Schätzung des unbekannten Erwartungswertes des Zielkriteriums

- Empirischer Mittelwert („Punktschätzung“) als Ergebnis

- » Keine Aussage über die Schätzgenauigkeit bzw. Schätzzuverlässigkeit

= > **Intervallschätzung erforderlich**

- Stichprobe vom Umfang n = Resultat eines Simulationsexperimentes

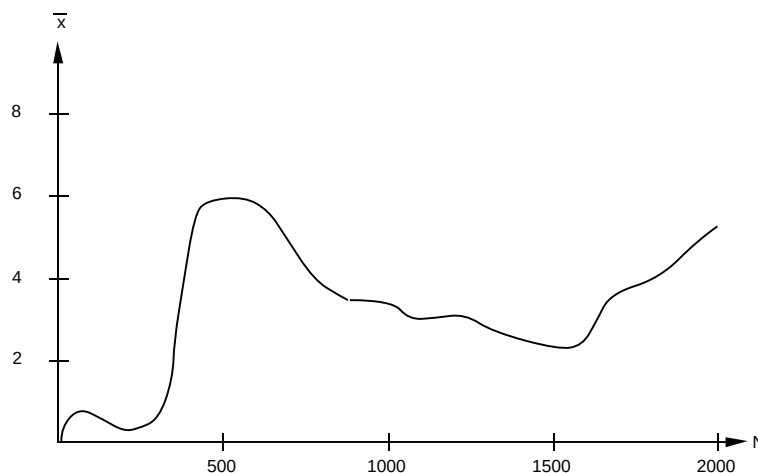
=> Schätzwert für $E(X)$

Schätzgenauigkeit der Simulationsergebnisse - Konfidenzintervalle -

- Schätzgenauigkeit nur mit **Konfidenzintervallen** zu beurteilen
 - Stichprobenmittelwert ohne Konfidenzintervall kaum aussagekräftig
 - » Genauigkeit unklar !
 - **Beispiel für die Bedeutung des Konfidenzintervalls**
 - Bedienungssystem mit einer Bedienungsstelle und exponentialverteilten Zwischenankunfts- bzw. Bedienzeiten
 - (M / M / 1) - Bediensystem mit analytischer Lösung für mittlere Warteschlangenlänge WSL
- $E(WSL) = 4.816$

Beispiel für die Bedeutung des Konfidenzintervalls

- Bedienungssystem mit einer Bedienungsstelle und exponentialverteilten Zwischenankunfts- bzw. Bedienzeiten
- Simulationsergebnisse für die mittlere Warteschlangenlänge in Abhängigkeit der Anzahl der Stichproben N



Berechnung des Konfidenzintervalls

- **Voraussetzung für die Berechnung von Konfidenzintervallen**
 - stationärer Prozeß
- **Unabhängige Stichproben = >**
 - Konventionelle Berechnung der Konfidenzintervalle
- **Autokorrelierte (abhängige) Stichprobenwerte = >**
 - Kompliziertere Auswertung der Simulationsergebnisse
- **Begriff der Autokorrelation :**
 - Abhängigkeit von Stichprobenwerten innerhalb einer Zeitreihe

Berechnung des Konfidenzintervalls

- **Konfidenzintervalle bei unabhängigen Stichproben**

- Schätzwert für Varianz

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- Zufallsgröße "Arithmetisches Mittel" als Schätzer für unbekannten Erwartungswert $E(X)$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i \quad \text{mit Varianz} \quad \text{Var}(\bar{X}) = \frac{\sigma^2}{n},$$

- Dann gilt für das Konfidenzintervall:

$$\left(\bar{X} - z \frac{s}{\sqrt{n}}, \bar{X} + z \frac{s}{\sqrt{n}} \right),$$

mit $z = 1,96$ für Wahrscheinlichkeit = 95% (Konfidenzniveau)

Berechnung des Konfidenzintervalls

Gängige Verfahren zur Gewinnung unabhängiger Stichproben in der Simulation:

(1) Unabhängige Wiederholungsläufe

- Wiederholung der Simulation mit unterschiedlichen Zufallszahlen
- Jeder Lauf eine Beobachtung X_i
- Nachteil: Wiederholtes Durchlaufen der Anlaufphase

(2) Quasi - unabhängige Blöcke in einem langen Simulationslauf

- Zerlegung des Simulationslaufes in Teilintervalle mit eigenen Stichprobenmittelwerten
- Stichprobenmittelwerte neue Beobachtungen, die (möglichst) unabhängig sein sollen
- Nachteil: Bestimmung der adäquaten Blocklänge erforderlich

Länge eines Simulationslaufes

- **Problem: Wahl des Stichprobenumfanges**
= Anzahl der Beobachtungen
- **Über Vorgabe maximale Länge Konfidenzintervall Berechnung des Stichprobenumfanges n aus Konfidenzintervallformel**

- L = halbe Länge des Konfidenzintervalls

$$L = z \frac{s}{\sqrt{n}} f \quad \text{mit } f \geq 1$$

- f = Faktor für autokorrelierte Stichprobenwerte
- $f = 1 \iff$ Unabhängigkeit
- auflösen nach n

$$n = z^2 \frac{s^2}{L^2} f^2$$

Länge eines Simulationslaufes

- **Zuerst Ableitung einer Schätzung s für unbekannte Varianz erforderlich**
 - » Schrittweise Vorgehensweise
- **Schrittweise Berechnung des Stichprobenumfanges**
 - 1) **Simulation $n_1 \leq n$ Stichproben**
 - 2) **Schätzung von Standardabweichung s und ggf. Faktor f sowie Berechnung des Konfidenzintervalls**
 - 3) **Wenn Konfidenzintervall innerhalb der Genauigkeitsgrenzen**
 - » Abbruch

Länge eines Simulationslaufes

Schrittweise Berechnung des Stichprobenumfanges, Forts.

- 4) **Anderenfalls Berechnung von n aus Formel für Stichprobenumfang**

$$n = z^2 \frac{s^2}{L^2} f^2$$

- 5) **Wiederholung von 1) mit $n_2 = n - n_1$**

» Zum Stichprobenumfang n muss noch Anlaufphase hinzugerechnet werden !