Network Simulation

Zusammenfassung

PER NATZSCHKA

Inhaltsverzeichnis

Abkürzungen

DES Discrete-Event Simulation
PMF Probability Mass Function
CDF Cumulative Distribution Function
PDF Probability Density Function

Author's Contact Information: Per Natzschka, per.natzschka1@mailbox.tu-dresden.de.

1 Network Simulation

1.1 Systemuntersuchung

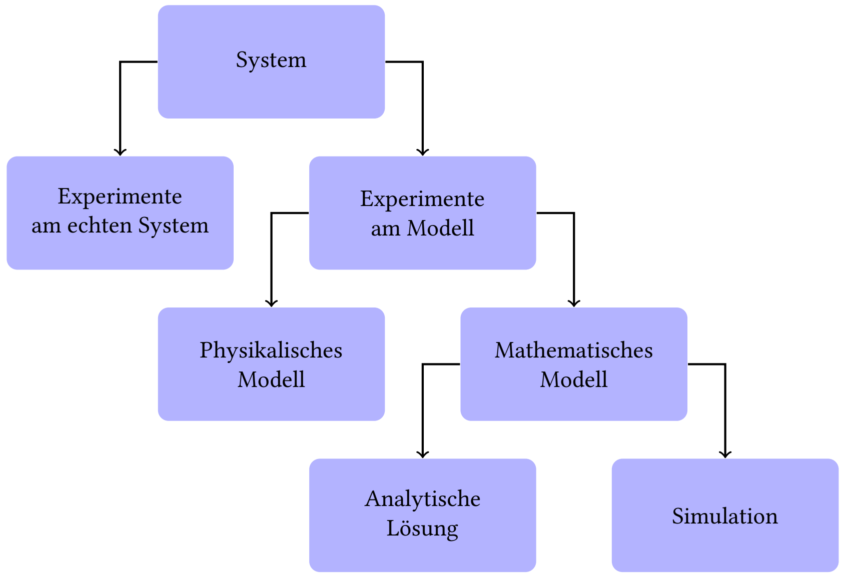

Fig. 1. Möglichkeiten der Systemuntersuchung

- Experimente am echten System
  - teuer
  - System existiert möglicherweise nicht
- Experimente am physikalischen Modell
  - untypisch bei Netzwerksimulationen
  - begrenzte Einsicht durch Feldtests
- Analytische Lösung eines Mathematischen Modells
  - zu präferieren
  - Modelle schnell zu komplex
  - hohe Abstraktion nötig
- Simulation
  - letzter Ausweg
  - Mittelweg zwischen Analytischer Lösung und Physikalischem Modell

1.2 Modelle

- Modell: Repräsentation eines Systems, um es zu untersuchen
- statisch/dynamisch
- deterministisch/stochastisch
- diskret/kontinuierlich

1.3 Simulationen

1.3.1 *Klassifikation von Simulationen.*

- Klassifikation abhängig von Modell
  - statisch/dynamisch
  - deterministisch/stochastisch
  - diskret/kontinuierlich
- Discrete-Event Simulation (DES)
  - dynamisch, stochastisch, diskret
  - trace-driven
  - Objekt-/Prozessorientiert
  - parallel/verteilt
- Terminierende Simulationen
  - spezifische Start- und Endbedingungen
  - Messungen hängen von Start- und Endbedingungen ab
  - ähnlich zu transienter Analyse (Impulsantwort)
  - z. B. Simulation der Flugbahn eines fallenden Balles, bis er den Boden berührt
- Steady-State Simulationen
  - kein natürliches Event legt Simulationslänge fest
  - überprüfen von Langzeitverhalten im eingeschwungenen Zustand
  - korrespondiert zur Steady-State Analyse
  - z. B. Simulation eines Gewichtes an einer Feder

1.3.2 *Schritte einer Simulationsuntersuchung.*

(1) Problemformulierung und Planung der Untersuchung
(2) Daten sammeln und Modelldefinition
(3) Validierung des konzeptuellen Modells
(4) Programmerstellung und Verifikation
(5) Testdurchläufe
(6) Validierung des programmierten Modells
(7) Experimente designen
(8) Simulation durchführen
(9) Output analysieren
(10) Dokumentieren, präsentieren, Ergebniss nutzen

1.3.3 *Vor- und Nachteile.*

- Vorteile
  - meist einzige Möglichkeit
  - erlaubt Annäherung an Systemverhalten unter geplanten Bedingungen
  - Vergleich unterschiedlicher Designs
  - Kontrolle über Bedingungen
  - erlaubt Zeitlupe/-raffung

- Nachteile
  - stochastische Modelle geben nur Schätzungen der wahren Charakteristika
  - teuer und zeitaufwendig zu entwickeln
  - lange Laufzeiten
  - massenhafte Outputdaten und Animationen lassen Ergebnisse glaubwürdiger erscheinen als sie sind

1.3.4 Pitfalls.

- ungenau definierte Objekte/unnötige Details
- Fehlkommunikation mit Management
- Fokus auf Programmierung ("nur eine Programmierübung")
- Zufälligkeit nicht einberechnet
- keine/falsche Daten gesammelt
- unpassende Simulationssoftware (undokumentierte Features?)
- Zweckentfremdung von Animation
- Outpudaten als die einzig wahre Antwort bewerten

1.3.5 Aufbau von DES.

Objekt	Typ	Beschreibung
Systemzustand	Variablenmenge	beschreibt System zu bestimmten Zeitpunkt
Simulationsuhr	Variable	gibt aktuelle Simulationszeit t an
Eventliste	Liste	enthält nächste Auftrittszeit je Eventtyp
Statistische Zähler	Variablenmenge	enthält statistische Informationen
Initialisierungsroutine	Subprogramm	initialisiert Simulationsmodell bei $t = 0$
Zeitablaufsroutine	Subprogramm	bestimmt nächstes Event und setzt t auf dessen Eintrittspunkt
Eventroutine	Subprogramm	updated Systemzustand, wenn bestimmtes Event auftritt
Bibliotheksroutine	Subprogramm	generiert zufällige Beobachtungen aus Wahrscheinlichkeitsverteilungen
Berichtsgenerator	Subprogramm	berechnet Schätzungen der gewünschten Messungen und generiert daraus Bericht am Simulationsende
Hauptprogramm	Subprogramm	startet Zeitablaufsroutine, um nächstes Event zu bestimmen und gibt Kontrolle an entsprechende Eventroutine

Tabelle 1. Elemente von DES

1.3.6 Statistische Aspekte.

- beobachtete Daten als Input
  - trace-driven
    - direkt und nahe an echtem System
    - kann nur historische Inputs reproduzieren → unflexibel
  - empirische Verteilung
    - Datenwerte als Verteilung interpolieren
    - recht valide, einfach, recht direkt
    - kann generierte Varianten begrenzen, schwer zu ändern
  - theoretische Verteilung
    - an theoretische Verteilung anpassen
    - kompakte Repräsentation mit wenigen Parametern, Daten werden "geglättet"
    - eventuell schwer, passende Verteilung zu finden
- Verteilungsfindung
  (1) für Familie entscheiden (exponentiell, gamma, Weibull, ...)
  (2) Parameter schätzen (z. B. Maximum Likelihood Estimation)
  (3) Repräsentativität bewerten (Diagramme, Test, ...)
- Statistische Analyse
  - Terminierende Simulationen
    - n unabhängige Wiederholungen
    - selbe Startbedingungen
    - selbe Endbedingungen
    - unterschiedliche Zufallszahlen
    - Verteilung der Mittelwerte bilden (Normalverteilungsannahme)
  - Steady-State Simulationen
    - anfängliche Warm-Up-Phase (keine Messungen)
    - Konvergenz gegen Steady State
      - weiterhin schwankende, korrelierende Beobachtungen
      - Simulation lang genug für festgelegte Präzision
- Probleme
  - Länge der Warm-Up-Phase
    - Daumenregeln
    - graphische Verfahren
    - statistische Test
  - Analyse der korrelierenden Daten
    - Endkriterien, um Simulation bei gewünschter Präzision zu beenden
    - Umgang mit korrelierenden Daten
    - konzeptuell: unabhängige Wiederholungen, Abschnittsmittel
    - zudem: andere, komplizierte statistische Methoden

2 M/M/1 Queues

2.1 Aufbau

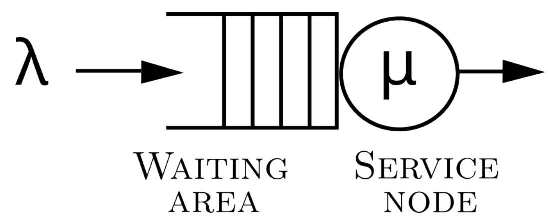

Fig. 2. Aufbau der M/M/1 Queue

- Kendalls Notation
- **M**
  - **M**emoryless arrival
  - exponentielle Verteilung der Zeit zwischen Ankünften → Poisson-Prozess
- **M**
  - **M**emoryless service time
  - exponentielle Verteilung der Bearbeitungszeiten
- **1**
  - **1** Server

2.2 Messungen

Messung	Beschreibung
Mittlere Verzögerung D	Durchschnitt der Verzögerungen $D_i = W_i + S_i$ W_i …Wartezeit, S_i …Bearbeitungszeit
Mittlere Queuelänge N	Durchschnitt von $N(t)$ für $t \to \infty$ $N(t)$ …Anzahl Kunden im System
Auslastung U	Durchschnittlicher Anteil der Zeit mit $N(t) > 0$
Durchsatz X	Durchschnittliche Anzahl abgeschlossener Bearbeitungen pro Zeitschritt

Tabelle 2. Typische Messungen bei M/M/1 Queues

2.3 Analyse

- Darstellung als Markov-Kette mit kontinuierlicher Zeit
  - Zustand i: Anzahl Kunden im System
  - $i \rightarrow i+1$: Ankunftsrate λ
  - $i \rightarrow i-1$: Bearbeitungsrate μ
- Zustandswahrscheinlichkeiten: $\pi_n = \lim_{t \to \infty} P(N(t) = n)$
- Balancegleichungen
  - $\lambda \pi_0 = \mu \pi_1$
  - $(\lambda + \mu)\pi_i = \lambda \pi_{i-1} + \mu \pi_{i+1}$
- Lösung im eingeschwungenen Zustand
  - $\rho = \frac{\lambda}{\mu}$
  - $\pi_i = \rho^i \pi_0$
  - $\displaystyle\sum_{i=0}^{\infty} \pi_i = \sum_{i=0}^{\infty} \rho^i \pi_0 = \frac{1}{1-\rho}\pi_0 = 1 \Rightarrow \pi_0 = 1 - \rho \Rightarrow \pi_i = \rho^i(1-\rho)$
  - modifizierte geometrische Verteilung
- Messungen
  - Durchschnittliche Anzahl Kunden im System: $E[N] = \frac{\rho}{1-\rho}$
  - Durchschnittliche Anzahl Kunden in der Queue: $E[N_q] = \frac{\rho^2}{1-\rho}$
  - Little's Law: $E[N] = \lambda E[T]$
    * Im eingeschwungenen Zustand gilt für Durchsatz: $X = \lambda$
    * Durchschnittliche Verzögerung: $E[T] = \frac{E[N]}{\lambda} = \frac{\frac{1}{\mu}}{1-\rho}$
    * Durchschnittliche Wartezeit: $E[W] = \frac{E[N_q]}{\lambda} = \frac{\frac{\rho}{\mu}}{1-\rho}$
- linear steigende Verzögerung bei niedrigem U
- unbegrenzte Verzögerung für $\rho = U \rightarrow 1$

3 Wahrscheinlichkeitstheorie

- Simulationen modellieren stochastische Prozesse
- statistische Mehoden nötig, um
  - Wahrscheinlichkeitsverteilungen und deren Parameter zu finden (Input-Modellierung)
  - Simulationsergebnisse zu analysieren

3.1 Grundlegendes

Begriff	Beschreibung
Zufallsexperiment	ein Prozess dessen Ergebnis nicht mit sicherheit feststeht
Ergebnisraum S	die Menge aller möglichen Ergebnisse eines Zufallsexperiments
Ergebnis	ein Element des Ergebnisraums S
Ereignis A	eine Menge von Ergebnissen, $A \subseteq S$

Tabelle 3. Grundbegriffe der Wahrscheinlichkeitstheorie

- $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$
- $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- $P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A) \cdot P(B) \Leftrightarrow A$ und B unabhängig

3.2 Zufallsgrößen und deren Verteilung

- Zufallsgröße: $X : S \to \mathbb{R}_0^+$
- diskrete Zufallsgröße: $|X(S)| \leq |\mathbb{N}|$
  - Wahrscheinlichkeitsfunktion/Probability Mass Function (PMF)
    - $p_i = P(X = x_i)$
  - Verteilungsfunktion/Cumulative Distribution Function (CDF)
    - $F(x) = P(X \leq x) = \sum\limits_{x_i \leq x} p_i$
    - $0 \leq F(x) \leq 1$
    - $F(x)$ monoton steigend ($x_1 \leq x_2 \Rightarrow F(x_1) \leq F(x_2)$)
    - $\lim\limits_{x \to -\infty} F(x) = 0, \lim\limits_{x \to +\infty} F(x) = 1$
    - $F(x)$ ist rechtskontinuierlich ($\lim\limits_{x \to x_0+} f(x) = f(x_0)$)
- kontinuierliche Zufallsgröße: $|X(S)| > |\mathbb{N}|$
  - CDF gleich definiert
  - Dichtefunktion/Probability Density Function (PDF)
    - $f(x) = \frac{d}{dx} F(x)$
    - Verteilung der Wahrscheinlichkeiten über Werte der Zufallsgröße
    - $\int_a^b f(x)dx = P(a \leq X \leq b)$

3.3 Momente und Quantile

- CDF $F(x)$ und PDF definieren Zufallsgröße vollständig
- Funktionen aber oft zu komplex
- wenige Zahlen besser zur Beschreibung

3.3.1 *Erwartungswert.*

- Erwartungswert $m = E[X]$
- Berechnung
  - X diskret: $E[X] = \sum_{i=1}^{\infty} x_i p_i$
  - X kontinuierlich: $E[X] = \int_{-\infty}^{\infty} x f(x) dx$
- Linearität: $E[aX + bY] = aE[X] + bE[Y]$
- Funktion einer Zufallsvariable, $Y = g(X)$
  - X diskret: $E[Y] = E[g(X)] = \sum_{i=1}^{\infty} g(x_i) p_i$
  - X kontinuierlich: $E[Y] = E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$

3.3.2 *Varianz.*

- Varianz $\sigma^2 = Var[X]$
- $Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$
- Standardabweichung σ
- Eigenschaften
  - $Var[aX] = a^2 Var[X]$
  - $Var[X + Y] = Var[X] + Var[Y]$ (wenn X und Y unabhängig sind)

3.3.3 *Momente.*

- Moment n-ter Ordnung: $E[X^n], n \geq 1$
- Zentrales Moment n-ter Ordnung: $E[(X - E[X])^n], n \geq 1$
- Beispiele
  - Moment 1. Ordnung: Erwartungswert
  - Zentrales Moment 2. Ordnung: Varianz
  - Moment 3. Ordnung: Schiefe (Maß für Asymmetrie)
- Verteilung kann auch durch Reihe von Momenten definiert werden (wenn diese existiert)

3.3.4 *Median.*

- kleinster Wert $x_{0.5}$, sodass $F(x_{0.5}) \geq 0.5$
- alternative Möglichkeit, Mittelwert anzugeben
- kann sinnvoll sein, wenn Verteilung extreme Werte annehmen kann

3.3.5 *Quantile.*

- für $0 < q < 1$ ist das q-Quantil der kleinste Wert x_q, sodass $F(x) \geq q$
- wenn X kontinuierlich und $F(x)$ streng monoton steigend: $F(x_q) = q, x_q = F^{-1}(q)$
- Median ist 0.5-Quantil

3.4 Verteilungen

3.4.1 *Geometrische Verteilung.*

- Experiment: Wiederhole Bernoulli-Versuche, bis zum ersten Erfolg
- Zufallsvariable: Anzahl an Versuchen
- PMF: $p_i = p(1-p)^{i-1}$
- CDF: $F(i) = \sum_{j=1}^{i} p_j = 1 - (1-p)^i$
- Erwartungswert: $E[i] = \sum_{j=1}^{\infty} j p_j = \frac{1}{p}$

3.4.2 *Exponentialverteilung.*

- PDF: $f(x) = \lambda e^{-\lambda x}, x \geq 0$
- CDF: $F(x) = 1 - e^{-\lambda x}$
- einziger Parameter: Rate λ
- Erwartungswert: $E[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$
- Varianz: $Var[X] = \int_0^{\infty} (x - \frac{1}{\lambda})^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2}$

3.4.3 *Normalverteilung.*

- PDF: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$
- CDF hat keine geschlossene Form
- Notation: $X \sim N(\mu, \sigma^2)$
- Standardnormalverteilung: $Z \sim N(0, 1)$
- Projektion: $F_X(x) = F_Z(\frac{x-\mu}{\sigma})$

3.5 Hypothesentests

3.5.1 *Statistische Hypothese.*

- Behauptung, die eine oder mehrere Populationen betrifft
- Verifikation nur durch Betrachtung der gesamten Population (bei Simulationen unmöglich)
- Falsifizierung
  - Beweis durch Gegenbeispiel
  - Folgt aus hoher Wahrscheinlichkeit → Hypothesentest

3.5.2 *Vorgehen.*

- Vorgehen
  (1) Beginn bei originaler Hypothese (Alternativhypothese H_1)
  (2) Logisches Komplement (Nullhypothese H_0) formulieren
  (3) H_0 widerlegen
  (4) Aus Falsifizierung von H_0 folgt Verifikation von H_1
- Folgerungen
  (1) H_0 widerlegt
    - ausreichend Hinweise in Daten
    - Wahrscheinlichkeit für Wahrheit von H_0 unter gewählter Grenze α (z. B. $\leq 5\%$)
    - H_0 ablehnen ist nach Konstruktion äquivalent zum annehmen von H_1
  (2) H_0 nicht widerlegt
    - durch nicht ausreichende Hinweise in Daten
    - sicherer bei H_0 zu bleiben ("sicherer Standardfall")

3.5.3 *Varianten.*

- One-/Two-Sample
  - Test, ob Mittelwert einer Population von einem gegebenen Wert abweicht (one-sample)
  - Test, ob sich zwei Populationen unterscheiden (two-sample)
- Unpaired/paired two-sample
  - paired Tests sind statistisch aussagekräftiger
  - anwendbar, wenn Proben aus A und B abhängig sind
  - z. B. bei Vorher-Nachher Vergleichen desselben Gebiets
- Normal-/Unnormal verteilte Daten
  - bei Normalverteiltungen (oder Zentraler Grenzwertsatz): t-Test
  - sonst: Wilcox-Test (weniger aussagekräftig)

4 Validierung

- Konzeptuelle Validierung: Überprüfung des konzeptuellen Modells
  - Sind die Abstraktionen, Vereinfachungen, Annahmen des Modells korrekt?
  - Bauen wir das richtige Modell?
  - Problem: Absolute Validation zu teuer
- Verifikation: Überprüfung der Implementation
  - Ist die Implementation korrekt (sprich: bugfrei)?
  - Bauen wir das Modell richtig?
  - Problem: Algorithmisch nicht lösbar (Halteproblem)
- keine Garantien für Validität möglich
- Tests durchführen, bis man sicher genug ist

4.1 Durchführende

- Modellentwickler
- Modellnutzer (geleitet durch Entwickler)
- Dritte (während oder nach Entwicklung)
- Bewertungsmodell
  - Subjektive Punkte für verschiedene Validierungsaspekte
  - Kombination von Einzel-, Kategorie- und Gesamtbewertungen
  - Schwächen
    * Bestandspunktzahl ist subjektiv
    * kein guter Indikator für Korrektheit
    * kann zu hohes Vertrauen in Modell verursachen

4.2 Schritte

- Datenvalidität
  - ausreichend, genau
  - Umformungen korrekt (z. B. dB ↔ lineare Skala)
  - Außenseiter finden und auf Korrektheit überprüfen
- Konzeptuelle Modellvalidierung (z. B. Linearität, Unabhängigkeit von Prozessen)
- Computergestützte Modellverifikation
  - Korrektheitsbeweise
  - Struktur überprüfen
- Funktionelle Validität
  - Simulationsdaten mit echtem System vergleichen (Vergleich der Validationsmöglichkeiten der Funktionalität)

	beobachtbares System	nicht-beobachtbares System
subjektiv	• graphische Anzeigen • Modellverhalten erforschen	• Vergleich zu anderen Modellen • Modellverhalten erforschen
objektiv	• statistische Tests	• Vergleich zu anderen Modellen durch statistische Tests

Tabelle 4. Vergleich der Validationsmöglichkeiten der Funktionalität

4.3 Techniken

4.3.1 *Sehr typisch.*

- Animation
- Vergleich zu anderen Modellen
  - einfache Fälle: analytische Modelle
  - sonst: Ergebnisse anderer (validierter) Modelle (z. B. aus anderen Simulationsframeworks)
- Degenerate Tests
  - Input- und interne Parameter auf degenerierende Fälle setzen
  - z. B. $\lambda > \mu \Rightarrow$ monoton steigende Verzögerung
- Ereignisvalidität
  - Auftritt von Simulationsevents mit echten Events vergleichen

4.3.2 *Typisch.*

- Test bei Extrembedingungen
  - Ausgabe sollte für jede Kombination extremer/unwahrscheinlicher Faktoren plausibel sein
  - z. B. es kommt lange kein Kunde an $\Rightarrow$ die Queue leert sich
- Augenscheinvalidität
  - Experten fragen, ob Modellstruktur und Ausgabe Sinn ergeben
- Historische Datenvalidation
  - trace-driven Simulation nutzen, um mit echtem System zu vergleichen
- Parametervariabilitäts-Sensibilitäts Analyse
  - Parameter variieren, um Effekt auf Ausgabe zu bestimmen
  - selbe Effekte sollten bei echtem System auftreten
  - für sensible Parameter: auf ausreichende Genauigkeit achten

4.3.3 *Selten.*

- Funktionelle Grafiken
  - Graphen von Leistungsmessungen während der Modellläufe anzeigen
- Voraussagende Validierung
  - Modell nutzen, um Systemverhalten vorauszusagen und dann vergleichen (Feldtest)
- Turing-Test
  - Experten fragen, ob er zwischen System- und Modellausgaben unterscheiden kann

4.3.4 *Untypisch.*

- Interne Validität
  - mehrere Replikationen erstellen, um stochastische Varianz im Modell zu bestimmen
- Historische Methoden
  - Rationalismus: Annahmen als wahr annehmen $\longrightarrow$ Modell durch logische Schlüsse bilden
  - Empirismus: Annahmen und Ergebnisse empirisch validieren
- Positive Ökonomie
  - Modell muss Zukunft voraussagen können
  - Annahmen und Modellstruktur nicht relevant
- Multistate Validierung
  - Kombination von Rationalismus, Empirismus und Positiver Ökonomie

4.4 Empfohlenes Vorgehen

- Animationen nutzen, um Systemzustand zu visualisieren
- Live-Graphen in GUI nutzen
- Degenerate Tests (z. B. Overload)
- Parametersensibilität testen
- Vergleich mit anderen Implementationen
- Augenscheinvalidierung des konzeptuellen Modells
- Modulweises Debugging