

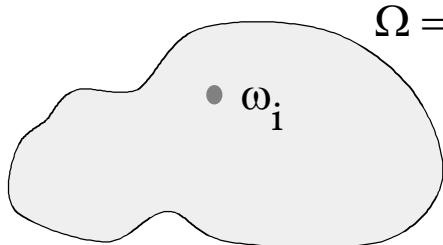
1 Statistische Grundlagen (Teil 1)

Zentraler Punkt der Mathematik ist die zahlenmäßige (quantitative) Erfassung der Welt. Während der Übergang von körperlich vorhandenen Dingen zu abstrakten Größen wie Länge, Masse oder Zeit zügig gemeistert wurde, blieb die quantitative Erfassung von Wahrscheinlichkeiten lange Zeit problematisch. Der Begriff "Wahrscheinlichkeit" wird seit 1933 durch die Axiome von Kolmogorow exakt implizit definiert. Damit wurde jedoch nicht geklärt, was Wahrscheinlichkeit ist, sondern nur herausgearbeitet, welche Eigenschaften strukturell Wahrscheinlichkeiten auszeichnen. Die Interpretation der Axiome bleibt eine offene Frage. Hier bestehen weiterhin unterschiedliche Auffassungen.

Quelle: https://de.wikipedia.org/wiki/Geschichte_der_Wahrscheinlichkeitsrechnung

1.1 Zufallsereignis und Wahrscheinlichkeit

1.1.1 Zufallsereignisse



$$\Omega = \{\omega_1, \omega_2, \dots\}$$

Ω : Ereignisraum

ω_i : Elementarereignis

Bsp.: Würfel $\Omega = \{1, 2, 3, 4, 5, 6\}$

1.1.2 Wahrscheinlichkeitsbegriff

a) Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit

n : Experimente,

A_i : ausgewähltes Merkmal, Menge von elementaren Ereignissen

n_i : Anzahl der Experimente, bei denen A_i festgestellt wird

relative Häufigkeit für A_i

$$h(A_i) = \frac{n_i}{n}$$

Wahrscheinlichkeit A_i

$$P(A_i) = \lim_{n \rightarrow \infty} \frac{n_i}{n}$$

(2.1)

Eigenschaften:

- $0 \leq h(A_i) \leq 1 \quad \Rightarrow \quad 0 \leq P(A_i) \leq 1$
- disjunkte Merkmale A_i, A_j ($A_i \cap A_j = \emptyset$)

$$h(A_i \cup A_j) = h(A_i) + h(A_j) \quad \Rightarrow \quad P(A_i \cup A_j) = P(A_i) + P(A_j)$$
- Vollständigkeitsrelation

$$\sum_i h(\omega_i) = 1 \quad \Rightarrow \quad \sum_i P(\omega_i) = 1$$

b) Laplace-Wahrscheinlichkeit

„A priori“-Wahrscheinlichkeit aufgrund symmetrischer Eigenschaften, z.B. Münze, Würfel, etc.

Experiment mit m gleichwertigen Alternativen

(m : Anzahl aller Alternativen)

m_i : Anzahl der Alternativen für A_i

Wahrscheinlichkeit für A_i wird definiert als:

$$P(A_i) = \frac{m_i}{m} \quad (2.2)$$

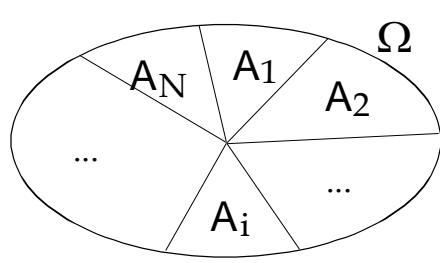
Beispiel Würfel :

- $m=6$ Alternativen (gleichwertig bei vollkommen symmetrischen Würfeln)
- A_i : Augenzahl ungerade

$$m_i = 3 \Rightarrow P(A_i) = \frac{3}{6} = 0,5$$

1.1.3 Wichtige Begriffe und Gesetze

a) Vollständiges Ereignissystem



Gegeben:

$$\begin{aligned} M &= \{A_1, A_2, \dots, A_N\}, \\ A_i \cap A_j &= \emptyset \quad \forall i \neq j \\ \Rightarrow P(M) &= \sum_{i=1}^N P(A_i) \end{aligned} \quad (2.3)$$

Die Menge von disjunkten Ereignissen $M = \{A_1, \dots, A_N\}$ bildet ein *vollständiges Ereignissystem* bzw. eine Partition, falls deren Vereinigung den gesamten Ereignisraum abdeckt, d.h. $\Omega = \bigcup_{1 \leq i \leq N} A_i$

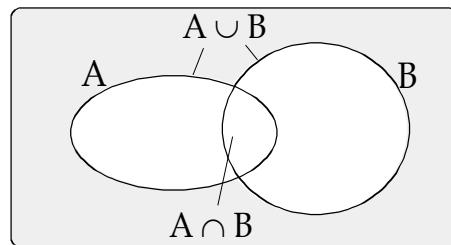
b) Verbundereignis und Verbundwahrscheinlichkeit

- A, B : zwei (nicht notwendigerweise disjunkte) Ereignisse

Verbundereignis : $(A \cap B) = (A, B)$

Verbundwahrscheinlichkeit : $P(A \cap B) = P(A, B) = P(B, A)$

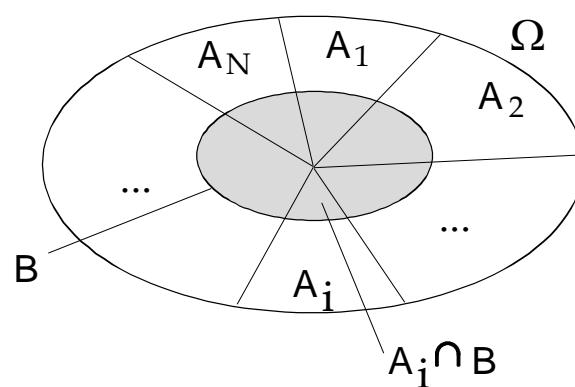
Weiter gilt: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (2.4)



Anwendungsbeispiele:

- A: „Augenzahl maximal 3“
- B: „Augenzahl gerade“
- Falls $M = \{A_1, \dots, A_N\}$ ein vollständiges Ereignissystem bildet, dann gilt:

$$P(B) = \sum_{i=1}^N P(A_i, B) \quad (2.5)$$



c) Bedingte Wahrscheinlichkeit

- Bedingtes Ereignis ($A | B$)

Eintritt des Ereignisses A unter der Bedingung, dass Ereignis B eintritt;

Voraussetzung ist $P(B) > 0$.

- Bedingte Wahrscheinlichkeit

$$P(A | B) = \frac{P(A, B)}{P(B)} \quad , \quad \text{d.h.} \quad P(A | B) \geq P(A, B) \quad (2.6)$$

- Simpson Paradoxon

Eigenschaften einer Gesamtmenge stehen mindestens teilweise im Widerspruch zu den Eigenschaften der Teilmengen.

Beispiel:

<http://www.spiegel.de/wissenschaft/mensch/simpsons-paradoxon-diese-statistik-kann-nicht-stimmen-oder-doch-a-1068204.html>

- Bewerbungen an einer Hochschule: 500 Frauen und 500 Männer
- Zulassungen: 240 Frauen (48% Quote) und 300 Männer (60% Quote)

⇒ In der Gesamtmenge wurden mehr Männer zugelassen
- 2 Studiengänge
 1. Studiengang: 360 Studienplätze
Bewerbungen: 100 Frauen und 400 Männer
Zulassungen: 80 Frauen (80% Quote) und 280 Männer (70% Quote)
 2. Studiengang: 180 Studienplätze
Bewerbungen: 400 Frauen und 100 Männer
Zulassungen: 160 Frauen (40% Quote) und 20 Männer (20% Quote)

⇒ In beiden Studiengängen wurden mehr Frauen zugelassen.

d) Statistische Unabhängigkeit

Ereignisse A und B sind statistisch unabhängig, wenn gilt:

$$P(A | B) = P(A) \quad \text{oder} \quad P(A, B) = P(A) \cdot P(B) \quad (2.7)$$

e) Gesetz der totalen Wahrscheinlichkeit

gilt nur bei Unabhängigkeit

Es sei $\{A_i, i = 1, \dots, N\}$ ein vollständiges Ereignissystem und $B \subset \Omega$. Dann gilt:

$$P(B) = \sum_{i=1}^N P(A_i, B) = \sum_{i=1}^N P(B | A_i) \cdot P(A_i) \quad \text{aus 2.6}$$

Aus der zweimaligen Anwendung von (2.6) erhalten wir

$$P(A_i, B) = P(B | A_i) \cdot P(A_i) = P(A_i | B) \cdot P(B)$$

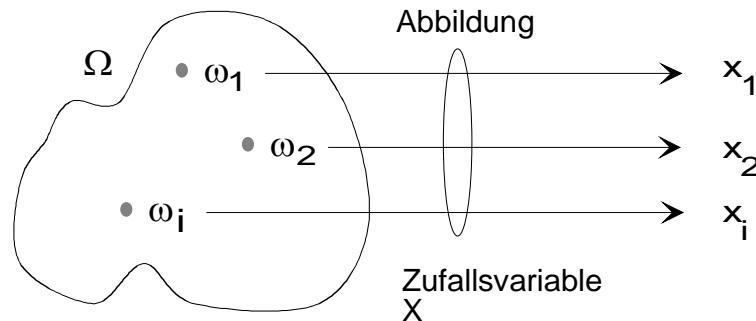
woraus sich die Bayes-Formel herleitet:

$$P(A_i | B) = \frac{P(B | A_i) \cdot P(A_i)}{P(B)} = \frac{P(B | A_i) \cdot P(A_i)}{\sum_{j=1}^N P(B | A_j) \cdot P(A_j)} \quad (2.8)$$

Sie hilft $P(A_i | B)$ zu berechnen, wenn nur $P(B | A_i)$ und $P(A_i)$ bekannt sind.

1.2 Zufallsvariable, Verteilung und Verteilungsfunktion

1.2.1 Zufallsvariable



- Zufallsvariable (ZV)
 - Offiziell: Funktion, die jedem Elementarereignis ω_i eine reelle Zahl zuordnet.
 - Intuitiv: Variable, deren Wert (eventuell mehrmals) zufällig nach einer gegebenen (Wahrscheinlichkeits-)Verteilung(sfunktion) ausgewürfelt wird.
- Diskrete Zufallsvariable
 - ZV X hat diskreten Wertebereich (meist nur ganzzahlige Werte)
 - Beispiel: Wurf mit zwei Würfeln, $X \in \{2, 3, \dots, 12\}$
- Kontinuierliche Zufallsvariable
 - ZV A hat reellwertigen Wertebereich
 - Beispiel: Durchlaufzeit eines Datenpakets im Rechnernetz, $A \in (150\text{ms}, 700\text{ms}]$

1.2.2 Verteilung

a) Definition

- X : diskrete Zufallsvariable

$$x(i) = P(X = i), \quad i = 0, 1, \dots, X_{\max} \quad (\text{Verteilung}) \quad (2.9)$$

$$\sum_{i=0}^{X_{\max}} x(i) = 1 \quad (\text{Vollständigkeitsrelation}) \quad (2.10)$$

b) Beispiele einfacher Verteilungen

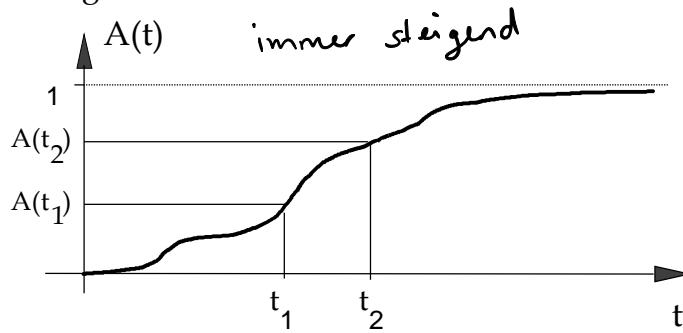
- Bernoulli-Verteilung
 - Bernoulli-Versuch: Experiment mit zwei möglichen Ergebnissen (z.B. Münze)
 - X: ZV für den Ausgang eines Bernoulli-Versuchs
 - Verteilung: $x(i) = \begin{cases} 1-p & i = 0 \text{ (Misserfolg)} \\ p & i = 1 \text{ (Erfolg)} \end{cases}$ (2.11)
 - Wertebereich: $i \in \{0,1\}$
 - Beispiel: Übertragung eines Bits als Bernoulli-Versuch
 - Bitfehlerwahrscheinlichkeit $p_b \Rightarrow p = 1 - p_b$
- Binomial-Verteilung
 - X: ZV für die Anzahl von Erfolgen bei n Bernoulli-Versuchen mit Erfolgswahrscheinlichkeit p
 - Verteilung: $x(i) = \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i}$ (2.12)
 - $\binom{n}{i} = \frac{n!}{i!(n-i)!}$ verschiedene Muster mit i Erfolgen
 - Jedes Muster mit i Erfolgen tritt auf mit Wahrscheinlichkeit $p^i \cdot (1-p)^{n-i}$
 - Wertebereich: $0 \leq i \leq n$
 - Beispiel: Fehlerwahrscheinlichkeit p_p eines Datenpakets mit n Bits
 - Gegeben: Bitfehlerwahrscheinlichkeit p_b
 - Wahrscheinlichkeit für i Bitfehler im Paket wird durch Binomial-Verteilung beschrieben
 - Paketfehlerwahrscheinlichkeit $\underbrace{p_p}_{\text{Bit korrekt}} = \sum_{1 \leq i \leq n} x(i) = 1 - x(0) = 1 - (1 - p_b)^n \neq n \cdot p_b$
 - Für $p_b \ll 1$ gilt: $(1 - p_b)^n \approx 1 - n \cdot p_b \Rightarrow p_p \approx n \cdot p_b$
- Geometrische Verteilung
 - X: ZV für die Anzahl von Fehlversuchen bei wiederholten Bernoulli-Versuchen mit Erfolgswahrscheinlichkeit p bis zum ersten Erfolg
 - Verteilung: $x(i) = (1-p)^i \cdot p$ (2.13)
 - Wertebereich: $0 \leq i < \infty$
 - Beispiel: Anzahl von fehlerhaften Übertragungsversuchen Y beim Versenden eines Datenpaketes mit Fehlerwahrscheinlichkeit p_p
 - $y(i) = p_p^i \cdot (1 - p_p)$ $\gamma(i)$ aus 2.9

1.2.3 Verteilungsfunktion und -dichtefunktion

A : beliebige (diskrete oder kontinuierliche) Zufallsvariable
 ↘ b)

a) Verteilungsfunktion (VF)

- $A(t) = P(A \leq t)$ (Verteilungsfunktion)
- Visualisierung



- Eigenschaften
 - $t_1 < t_2 \Rightarrow A(t_1) \leq A(t_2)$ (Monotonieeigenschaft)
 - $t_1 < t_2 \Rightarrow P(t_1 < A \leq t_2) = A(t_2) - A(t_1)$
 - $A(-\infty) = 0$
 - $A(\infty) = 1$
- $A^c(t) = P(A > t) = 1 - A(t)$ (komplementäre Verteilungsfunktion)
symmetrisch

b) Verteilungsdichtefunktion (VDF)

- Nur bei kontinuierlichen ZV anwendbar
- Pendant zur Verteilung von diskreten ZV

$$\bullet \quad a(t) = \frac{d}{dt} A(t) \quad (\text{Verteilungsdichtefunktion}) \quad (2.15)$$

$$\bullet \quad \int_{-\infty}^{\infty} a(t) dt = 1 \quad (\text{Vollständigkeitsrelation}) \quad (2.16)$$

$$\bullet \quad A(t) = \int_{-\infty}^t a(\tau) d\tau \quad (\text{Zusammenhang zw. VDF und VF})$$

- Beispiel: Exponentielle VF

- VF: $A(t) = 1 - e^{-\lambda \cdot t}$
- VDF: $a(t) = \lambda \cdot e^{-\lambda \cdot t}$

\hookrightarrow Ableitung

Fläche unter Kurve ist 1

VF gibt Wahrscheinlichkeiten an

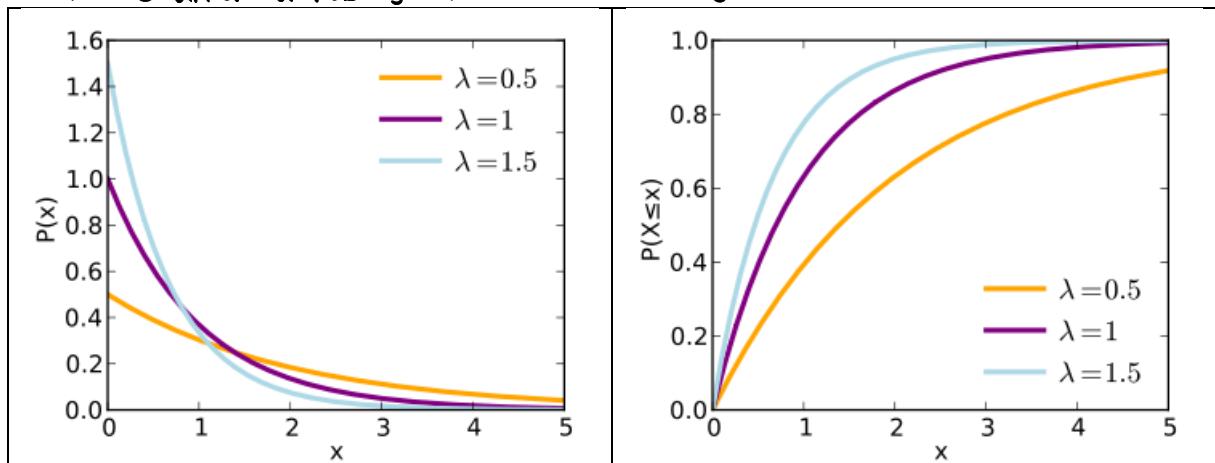
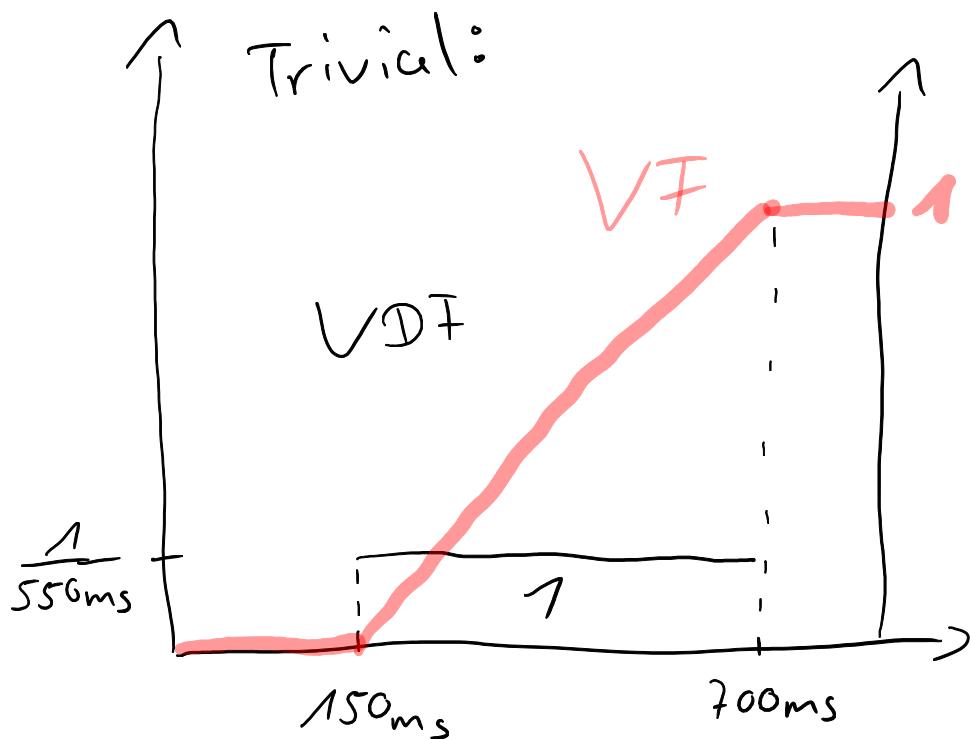


Figure 1: VDF und VF der exponentiellen Verteilung (Quelle: Wikipedia)

je höher Dichte, desto häufiger treten die Werte auf



1.2.4 Special Operations on Distributions

Applicable to independent random variables.

Examples: X_i = number of pips, A_i = exponentially distributed duration

1.2.4.1 Sum of Two Random Variables (Convolution, „Faltung“)

1.2.4.1.1 Discrete Distributions (X, Y)

- X_1, X_2 : random variables
- Wanted: distribution of $X = X_1 + X_2$
- $P(X = i) = x(i) = \sum_j x_1(j) \cdot x_2(i - j) = (x_1 * x_2)(i)$

1.2.4.1.2 Continuous Distributions (A, B)

- A_1, A_2 : random variables
- Wanted: probability density function of $A = A_1 + A_2$
- $a(t) = \int_{-\infty}^{\infty} a_1(\tau) \cdot a_2(t - \tau) d\tau = (a_1 * a_2)(t)$

1.2.4.2 Difference of Two Random Variables

1.2.4.2.1 Discrete Distributions

- X_1, X_2 : random variables
- Wanted: distribution of $X = X_1 - X_2$
- $P(X = i) = x(i) = \sum_j x_1(i + j) \cdot x_2(j) \quad (i+j) - j = i$

1.2.4.2.2 Continuous Distributions

- A_1, A_2 : random variables
- Wanted: probability density function of $A = A_1 - A_2$
- $a(t) = \int_{-\infty}^{\infty} a_1(t + \tau) \cdot a_2(\tau) d\tau \quad //$

1.2.4.3 Maximum of Several Random Variables

- A_i : random variables
- Wanted: distribution function of $A = \max(A_0, \dots, A_{k-1})$
- $k = 2$: $A(t) = A_1(t) \cdot A_2(t)$
Proof: $A(t) = P(\max(A_1, A_2) \leq t) = P(A_1 \leq t) \cdot P(A_2 \leq t) = A_1(t) \cdot A_2(t)$
- $k=2$: $a(t) = a_1(t) \cdot A_2(t) + a_2(t) \cdot A_1(t) \quad \underbrace{\text{Abgeleitet (Produktregel)}}$
Proof: $a(t) = \frac{dA(t)}{dt}$
- $A(t) = \prod_{0 \leq i < k} A_i(t)$

$$\begin{aligned}
 P(\max(A_1, A_2) \leq 4) &= \\
 P(A_1 \leq 4, A_2 \leq 4) &= \\
 P(A_1 \leq 4) \cdot P(A_2 \leq 4) &= \\
 \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}
 \end{aligned}$$

1.2.4.4 Minimum of Several Random Variables

- A_i : random variables
- Wanted: distribution function of $A = \min(A_0, \dots, A_{k-1})$
- $k = 2$: $A^c(t) = A_1^c(t) \cdot A_2^c(t)$, $a(t) = a_1(t) \cdot A_2^c(t) + a_2(t) \cdot A_1^c(t)$
- $A^c(t) = \prod_{0 \leq i < k} A_i^c(t)$
- In other words: $A(t) = 1 - \prod_{0 \leq i < k} (1 - A_i(t))$

1.3 Statistische Größen von Verteilungen

1.3.1 Erwartungswert

- o ZV A mit VDF $a(t)$
- o Reellwertige Funktion $g(A)$ der ZV A

$$E[g(A)] = \int_{-\infty}^{+\infty} g(t) \cdot a(t) dt \quad (\text{Erwartungswert von } g(A)) \quad (2.17)$$

t: Element aus Wertebereich

- o Analog für diskrete ZV X mit Verteilung $x(i)$: $E[g(X)] = \sum_i g(i) \cdot x(i)$

1.3.2 Erwartungswert (erstes Moment) einer Zufallsvariablen

$$g(A) = A \Rightarrow m_1 = E[A] = \int_{-\infty}^{+\infty} t \cdot a(t) dt \quad (2.18)$$

*Würfel: i = Augenz.
x(i) = Wahrsch.
für i*

1.3.3 Verallgemeinerung: Gewöhnliche Momente

- o k-tes gewöhnliches Moment von A

$$g(A) = A^k \Rightarrow m_k = E[A^k] = \int_{-\infty}^{+\infty} t^k \cdot a(t) dt \quad (2.19)$$

- o Analog für diskrete ZV X mit Verteilung $x(i)$: $m_k = E[X^k] = \sum_i i^k \cdot x(i)$
- o Das erste gewöhnliche Moment heißt Erwartungswert $E[A]$ von A (mean, expected value).

1.3.4 Zentrale Momente

- o **k-tes zentrales Moment von A**

$$g(A) = (A - m_1)^k \Rightarrow \mu_k = E[(A - m_1)^k] = \int_{-\infty}^{+\infty} (t - m_1)^k \cdot a(t) dt \quad (2.20)$$

- o Analog für diskrete ZV X mit Verteilung $x(i)$:

$$\mu_k = E[(X - m_1)^k] = \sum_i (i - m_1)^k \cdot x(i)$$

- **Varianz (k=2) (variance)** Abweichung von Mittelw.

$$\begin{aligned} \text{VAR}[A] &= \mu_2 = E[(A - m_1)^2] = E[A^2 - 2Am_1 + m_1^2] \\ &= \underbrace{E[A^2]}_{m_2} - \underbrace{2m_1 E[A]}_{2m_1 z} + m_1^2 \\ &= m_2 - m_1^2 \end{aligned} \quad (2.21)$$

- **Standardabweichung (standard deviation)**

$$\sigma_A = \sqrt{\text{VAR}[A]} \quad (2.22)$$

- **Variationskoeffizient (coefficient of variation)** relative Schwankung bezogen auf Mittelw.

$$c_{\text{var}}[A] = c_A = \frac{\sigma_A}{m_1} = \frac{\sqrt{\text{VAR}[A]}}{E[A]} \quad (2.23)$$

Der Variationskoeffizient macht nur für positive A Sinn, denn sonst kann der Mittelwert $E[A]$ null oder negativ werden.

- **Schiefe (skewness)**

nice to know

$$v_A = \frac{E[(A - E[A])^3]}{\sigma_A^3} \quad (2.24)$$

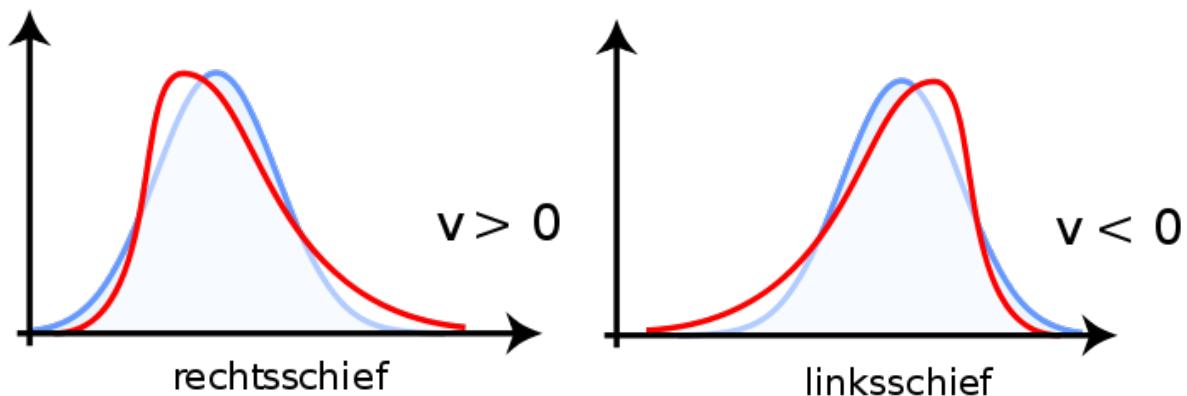
symmetrische Verteilung: $v_A = 0$

$v_A > 0$: rechtsschief

$v_A < 0$: linksschief

x0	p(x0)	x1	p(x1)	E[A]	E[A ²]	E[A ³]	VAR[A]	stddev[A]	E[(A-E[A]) ³]	skewness[A]
-1	0,1	1	0,9	0,8	1	0,8	0,36	0,6	-0,576	-2,666666667
-1	0,5	1	0,5	0	1	0	1	1	0	0
-1	0,9	1	0,1	-0,8	1	-0,8	0,36	0,6	0,576	2,666666667

Beispiele (Quelle: [http://de.wikipedia.org/wiki/Schiefe_\(Statistik\)](http://de.wikipedia.org/wiki/Schiefe_(Statistik)))



1.3.5 Beispiele ✓ = Einfach, Herleiten

1.3.5.1 Bernoulli-Verteilung

$$\checkmark E[X] = \sum_i i \cdot x(i) = p$$

$$\checkmark E[X^2] = \sum_i i^2 \cdot x(i) = p$$

$$\checkmark VAR[X] = E[X^2] - E[X]^2 = p - p^2 = p \cdot (1 - p)$$

$$\checkmark \sigma[X] = \sqrt{p \cdot (1 - p)}$$

$$\checkmark c_{var}[X] = \frac{\sqrt{p \cdot (1 - p)}}{p} = \sqrt{\frac{(1 - p)}{p}}$$

1.3.5.2 Geometrische Verteilung

$$\begin{aligned} \cancel{\times} E[X] &= \sum_{i=0}^{\infty} i \cdot x(i) = \sum_{i=0}^{\infty} i \cdot (1 - p)^i \cdot p = p \cdot (1 - p) \cdot \sum_{i=1}^{\infty} i \cdot (1 - p)^{i-1} \\ &= p \cdot (1 - p) \cdot \left(- \sum_{i=0}^{\infty} (1 - p)^i \right)' = p \cdot (1 - p) \cdot \left(\frac{-1}{p} \right)' = p \cdot (1 - p) \cdot \frac{1}{p^2} = \frac{(1 - p)}{p} \end{aligned}$$

$$\begin{aligned} \cancel{\times} E[X^2] &= \sum_{i=0}^{\infty} i^2 \cdot x(i) = \sum_{i=0}^{\infty} i^2 \cdot (1 - p)^i \cdot p = \sum_{i=0}^{\infty} i^2 \cdot q^i \cdot (1 - q) \\ &= (1 - p)^2 \cdot p \cdot \left(\sum_{i=0}^{\infty} i^2 \cdot (1 - p)^{i-2} \right) \\ &= (1 - p)^2 \cdot p \cdot \left(\sum_{i=0}^{\infty} (i - 1) \cdot i \cdot (1 - p)^{i-2} + \sum_{i=0}^{\infty} i \cdot (1 - p)^{i-2} \right) \\ &= (1 - p)^2 \cdot p \cdot \left(\sum_{i=0}^{\infty} (1 - p)^i \right)'' + \sum_{i=0}^{\infty} i \cdot (1 - p)^i \cdot p = (1 - p)^2 \cdot p \cdot \left(\frac{1}{p} \right)'' + \frac{(1 - p)}{p} \\ &= (1 - p)^2 \cdot p \cdot \frac{2}{p^3} + \frac{(1 - p)}{p} = \frac{2 \cdot (1 - p)^2 + p \cdot (1 - p)}{p^2} = \frac{(1 - p)^2 + (1 - p)}{p^2} \end{aligned}$$

$$\cancel{\times} VAR[X] = E[X^2] - E[X]^2 = \frac{(1 - p)^2 + (1 - p)}{p^2} - \left(\frac{(1 - p)}{p} \right)^2 = \frac{(1 - p)}{p^2}$$

$$\cancel{\times} \sigma[X] = \frac{\sqrt{(1 - p)}}{p}$$

$$\cancel{\times} c_{var}[X] = \frac{\sigma[X]}{E[X]} = \frac{\sqrt{(1 - p)}}{p} \cdot \frac{p}{1 - p} = \frac{1}{\sqrt{1 - p}}$$

1.3.5.3 Exponentielle Verteilung

$$\cancel{X} E[A] = \int_0^\infty t \cdot a(t) dt = \int_0^\infty t \cdot \lambda \cdot e^{-\lambda \cdot t} dt = [t \cdot (-e^{-\lambda \cdot t})]_0^\infty - \int_0^\infty (-e^{-\lambda \cdot t}) dt =$$

$$0 - 0 - \left(\frac{1}{\lambda} [e^{-\lambda \cdot t}]_0^\infty \right) = \frac{1}{\lambda} \text{ Merken}$$

$$\cancel{X} E[A^2] = \int_0^\infty t^2 \cdot a(t) dt = \int_0^\infty t^2 \cdot \lambda \cdot e^{-\lambda \cdot t} dt = \dots = \frac{2}{\lambda^2}$$

$$\cancel{X} VAR[A] = E[A^2] - E[A]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

$$\cancel{X} \sigma[A] = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda} \quad \leftarrow \quad \left. \begin{array}{l} \text{Herleiten} \\ \text{Merken} \end{array} \right\}$$

$$\cancel{X} c_{var}[A] = \frac{\sigma[A]}{E[A]} = \frac{1}{\lambda} / \frac{1}{\lambda} = 1 \quad \leftarrow \quad \left. \begin{array}{l} \text{Merken} \end{array} \right\}$$

1.3.6 Skalierungseigenschaften

- $C = b \cdot A$

- $\circ \quad E[C] = b \cdot E[A]$ (2.27a)

- $\circ \quad VAR[C] = b^2 \cdot VAR[A]$ (2.27b)

- $C = A + B$

- $\circ \quad E[C] = E[A] + E[B]$ (immer) (2.27c)

- $\circ \quad VAR[C] = VAR[A] + VAR[B]$ (falls A und B unabhängig) (2.27d)

Beweis siehe Tran-Gia, Kap. 1.2.5 c)

Gegen bsp.

$$C = A + B$$

0	-1	1
0	1	-1

1.3.7 p-Quantil t_p (im Englischen auch p-percentile)

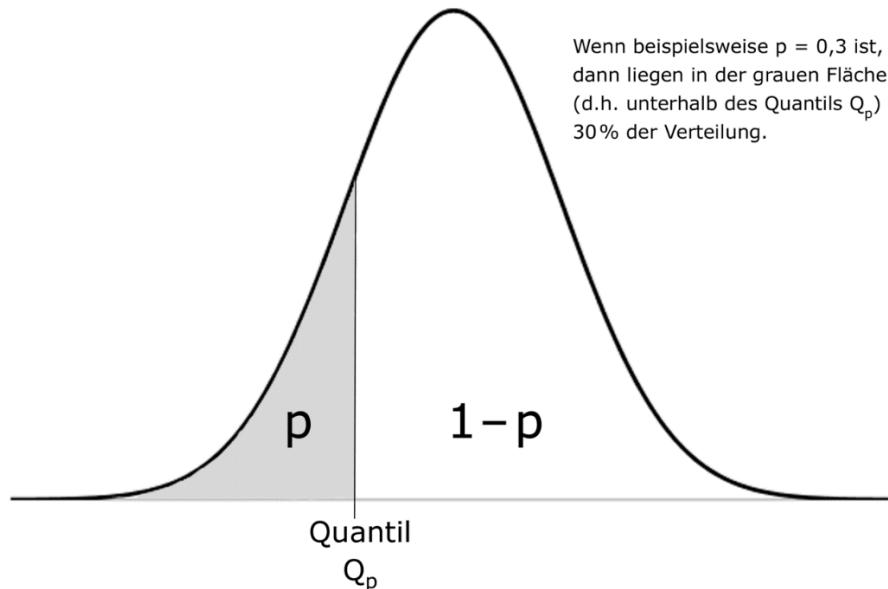
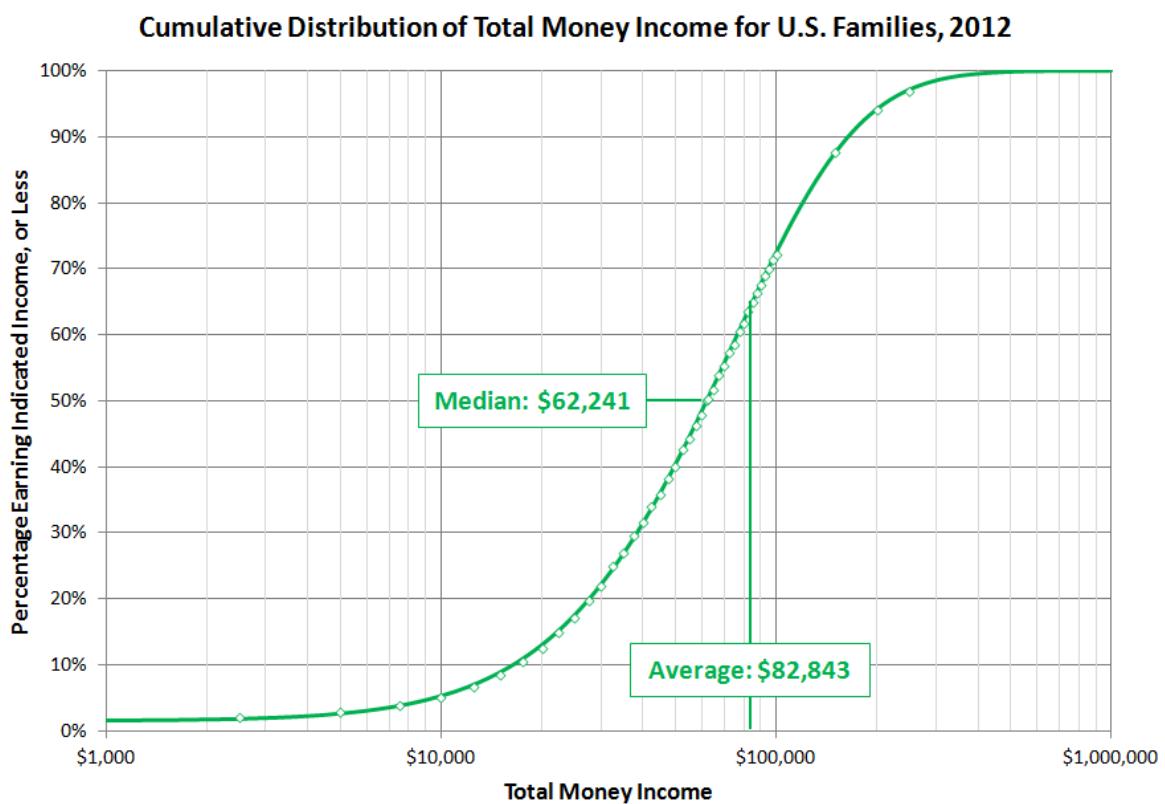


Abbildung 1: Das p-Quantil (t_p oder Q_p) visualisiert anhand der Verteilungsdichtefunktion.
Quelle: <https://upload.wikimedia.org/wikipedia/commons/b/ba/Normalverteilung.png>

- $t_p = \inf_{t \in R} (A(t) \geq p) =: A^{-1}(p)$ für p ($0 < p < 1$)
 - Wird auch Q_p genannt
- Bei kontinuierlichen ZVs ist eine echte Umkehrung möglich, da $A(t)$ bijektiv ist.
- Spezielle Quantile
 - Das 0.5-Quantil heißt Median.
 - Median und Erwartungswert stimmen für symmetrische Verteilungen überein.
 - Der Median kann aussagekräftiger sein als der Mittelwert, wenn dieser durch Ausreißer stark beeinflusst wird.
 - Das 0.25 bzw 0.75-Quantil heißt unteres bzw. oberes Quartil.
- Einhaltung von Grenzwerten
 - Mittelwert nicht aussagekräftig
 - Harte Obergrenze möglicherweise schwer einzuhalten
 - Grenzwert darf mit 0.1% überschritten werden: Das 99.9% Quantil muss kleiner als der Grenzwert sein.
Vorteil: Grenzwert wird fast immer eingehalten, das technische System ist aber i.d.R. aber wesentlich einfacher als eines, welche strikte Obergrenzen einhalten muss.



Source: U.S. Census, Current Population Survey, Annual Social and Economic Supplement, 2013

© Political Calculations 2013

Abbildung 2: Das 50%-Quantil (Median) anhand einer Verteilungsfunktion.
Quelle: <http://politicalcalculations.blogspot.de/2013/09/what-is-your-us-income-percentile.html>

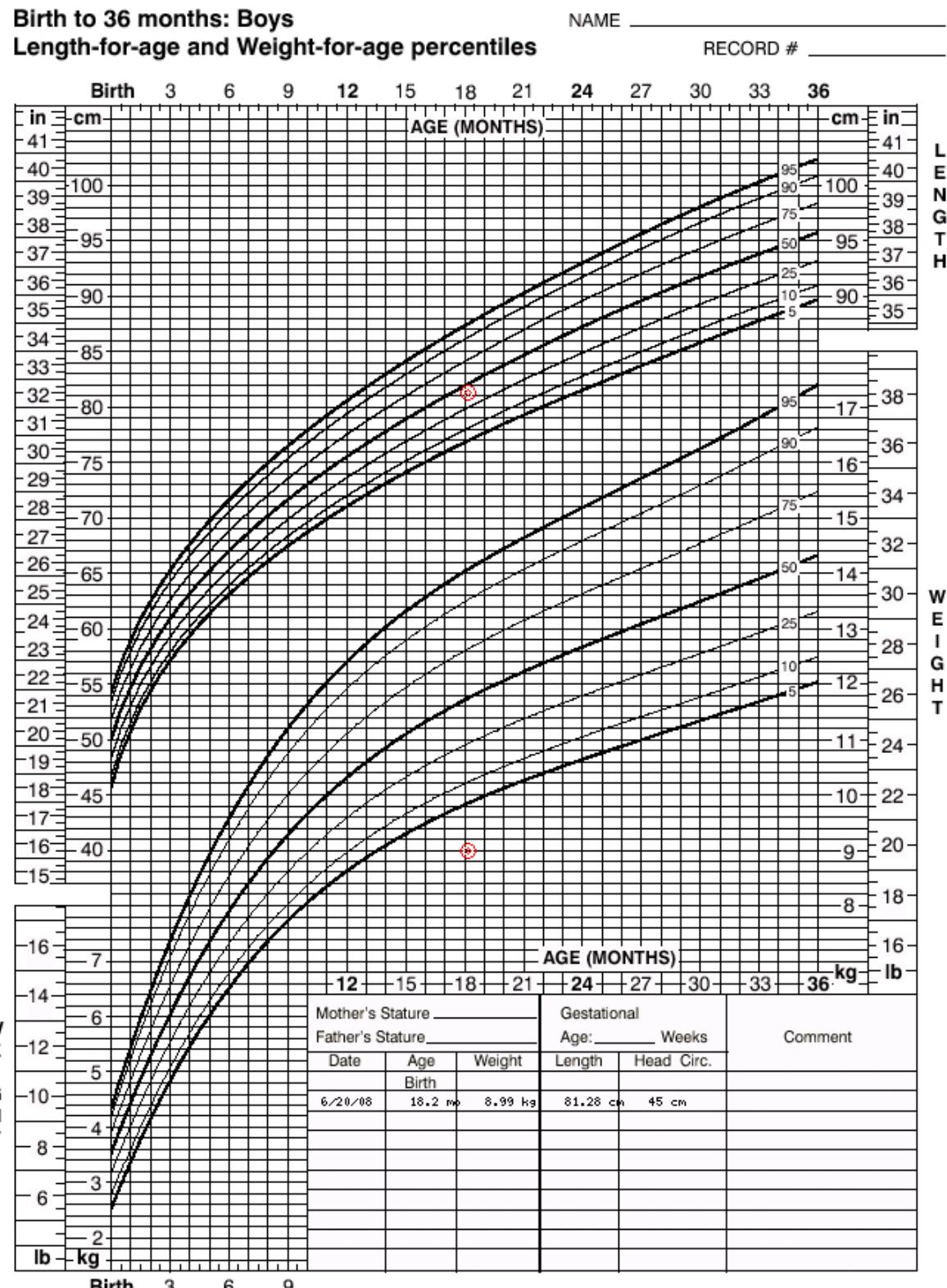


Abbildung 3: 5%, 10%, 25%, 50%, 75%, 90% und 95% Quantile von Körperlänge und -gewicht bei Jungen im Alter zwischen 0 und 36 Monaten.

Quelle: <https://higherhighslowerlows.files.wordpress.com/2010/09/growth-chart.png>

1.4 Statistische Größen von Stichproben (LK 4.3, 4.4, 6.4.1)

Definition:

Ein stochastischer Prozess ist eine geordnete Menge von Tupeln $\{X(t), t\}$, $X(t) \in \Xi$, $t \in \Gamma$ mit Zustandsraum Ξ und Indexmenge Γ .

Der Prozess ist zeitdiskret falls $\Gamma = N_0$ (=Betrachtungszeitpunkte) ist und zeitkontinuierlich falls $\Gamma = R_0$ (=zeitlicher Verlauf) ist.

Problem: Keine Verteilung gegeben sondern nur eine Stichprobenmenge X_0, X_1, \dots, X_{n-1} mit $X_i = X(t_i)$ und $t_i \in [0, T]$.
Statistische Größen müssen anders berechnet werden.

1.4.1 Allgemeine Empirische Momente

$$\overline{X^k} = \begin{cases} \overline{X^k}(n) = \frac{1}{n} \cdot \sum_{0 \leq i < n} X_i^k & \text{für zeitdiskreten Prozess} \\ \overline{X^k}(T) = \frac{1}{T} \cdot \int_0^T X(t)^k dt & \text{für zeitkontinuierlichen Prozess} \end{cases} \quad (2.28)$$

wobei n bzw. T die Größe der Stichprobe widerspiegelt.

1.4.2 Empirischer Mittelwert (sample mean)

$$\bar{X} = \overline{X^1} \quad (2.29)$$

1.4.3 Empirische Varianz (sample variance)

$$S^2 = \begin{cases} S^2(n) = \frac{1}{n-1} \cdot \sum_{0 \leq i < n} (X_i - \bar{X})^2 = \frac{n}{n-1} \cdot \left(\overline{X^2} - \overline{X^1}^2 \right) & \text{für zeitdiskreten Prozess} \\ S^2(T) = \frac{1}{T} \int_0^T (X(t) - \bar{X})^2 dt = \overline{X^2} - \overline{X^1}^2 & \text{für zeitkontinuierlichen Prozess} \end{cases} \quad (2.30)$$

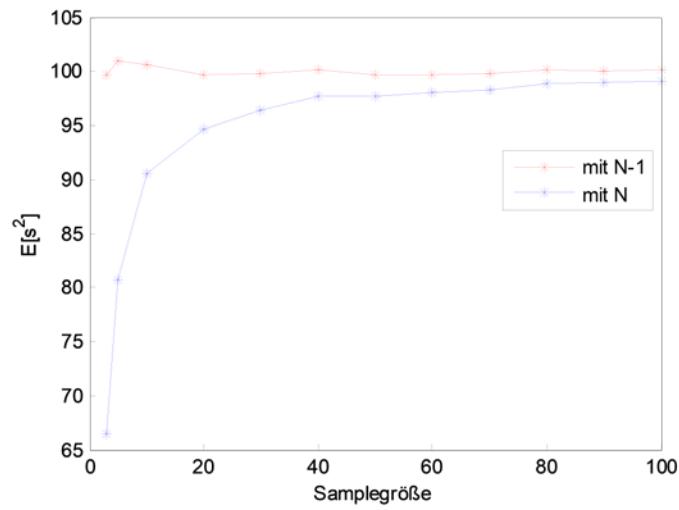


Abbildung 4: Einfluss der Erwartungstreue der Varianzschätzer auf die empirische Varianz.

1.4.4 Empirische Schiefe (sample skewness)

$$\hat{\nu} = \frac{\overline{(X-\bar{X})^3}}{(S^2)^{\frac{3}{2}}} = \frac{n}{(n-1)\cdot(n-2)} \cdot \sum_{i=0}^n \left(\frac{(X_i - \bar{X})}{\sqrt{S^2}} \right)^3 \quad (2.31)$$

Siehe: [http://de.wikipedia.org/wiki/Schiefe_\(Statistik\)](http://de.wikipedia.org/wiki/Schiefe_(Statistik))

1.5 Messobjekte in Simulationen

In Simulationsprogrammen werden Statistikobjekte angelegt, denen im Laufe einer Simulation Stichproben mit Zeitstempel zur Auswertung übergeben werden. Es gibt diskrete und zeitkontinuierliche Messwerte. Diese müssen bei der Datenerfassung unterschiedlich behandelt werden.

- <http://simul.iro.umontreal.ca/ssj/doc/html/umontreal/iro/lecuyer/stat/StatProbe.html>
- <http://desmoj.sourceforge.net/tutorial/statistics/0.html>

1.5.1 Erfassungsobjekte für Momente von diskreten Messwerten

- „Discrete counters (DC)“, „tally“ (Strichliste)
- Beispiele: Wartezeiten, Paketgrößen, Zwischenankunftszeiten
- Übergabe der Stichproben X_i an das Statistikobjekt
- Interne Datenhaltung: $\sum_{0 \leq i < n} X_i^k$ und n
- Berechnung der statistischen Größen nach 2.28 und 2.30

1.5.2 Erfassungsobjekte für Verteilungen von diskreten Messwerten (LK 6.4.2)

- Histogramm mit Gesamtwertebereich: $[a_0, b_{m-1}]$
- Enthält m (z.B. $m=2^k$) Zielintervalle mit je einer Zählvariable (bin)
- Array von bins entsprechend dem Wertebereich $\text{Range}(j) = [a_j, b_j]$
 - Breite äquidistanter Intervalle: $\Delta = \frac{b_{m-1} - a_0}{m}$
 - $a_j = a_0 + j * \Delta$
 - $b_j = a_0 + (j+1) * \Delta$
- Anzahl der bisherigen Messwerte n wird gespeichert
- Aktionen bei Übergabe der Stichprobe X_i
 - falls $X_i \in \text{Range}(j)$, inkrementiere $\text{bin}(j)$ um 1,
 - inkrementiere n um 1
- Ausgabe als Liste mit Tupeln
 - $(\text{Range}(j), h_j)$ mit relativer Häufigkeit: $h_j = \text{bin}(j)/n$
 - Diskrete empirische Verteilung: $((b_j - a_j)/2, h_j)$
 - Kontinuierliche empirische Verteilungsdichtefunktion:

$$h(x) = \begin{cases} 0 & \text{for } x < a_0 \\ \frac{h_j}{b_j - a_j} & \text{for } j \in \text{Range}(j) \\ 0 & \text{for } b_{m-1} \leq x \end{cases}$$
 - Fläche unter Histogramm ist 1.
- Problem: Wahl von m
 - m klein: Informationsverlust, m groß: Form unregelmäßig
 - Lösung
 - Wähle m anfangs groß
 - Aggregiere benachbarte bins um die bin-Werte für $m/2$ Intervalle zu erhalten
 - Führe Verfahren solange durch bis ein „vernünftiges“ Bild eintritt
 - Zähler kann durch geeignete Vererbung für Erfassung von Momenten erweitert werden

1.5.3 Erfassungsobjekte für Momente von zeitkontinuierlichen Messwerten

- „Time-weighting counters“, „accumulate“

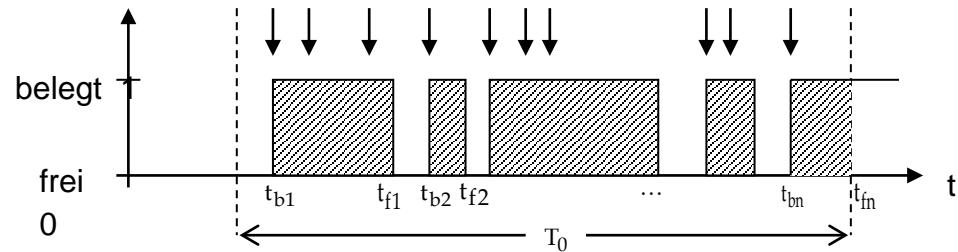
1.5.3.1 Motivation

- Beobachtung zeitkontinuierlicher Prozesse

- Mittelwert: $\bar{Z} = \frac{1}{T_0} \int_{T_0} Z(t) dt$

- Beispiel: Prozessoreinheit

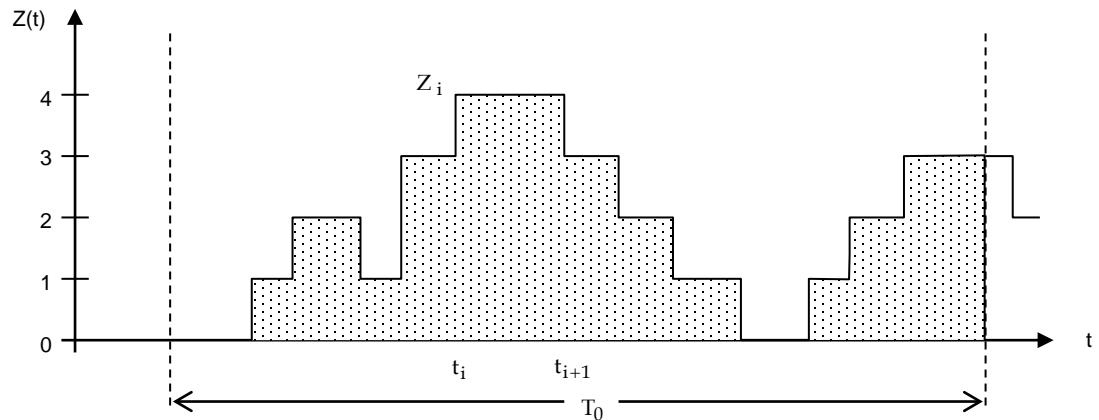
- Gesuchte Größe: Auslastung $\frac{1}{T_0} \sum_{i=1}^n (t_{fi} - t_{bi})$



- Beispiel: Warteschlange

- Gesuchte Größe: Mittlere Warteschlangenlänge

$$\bar{Z} = \frac{1}{T_0} \cdot \sum_{0 \leq i < n} Z_i \cdot (t_{i+1} - t_i)$$



1.5.3.2 Implementation of Time-Weighting Counters (TWCs)

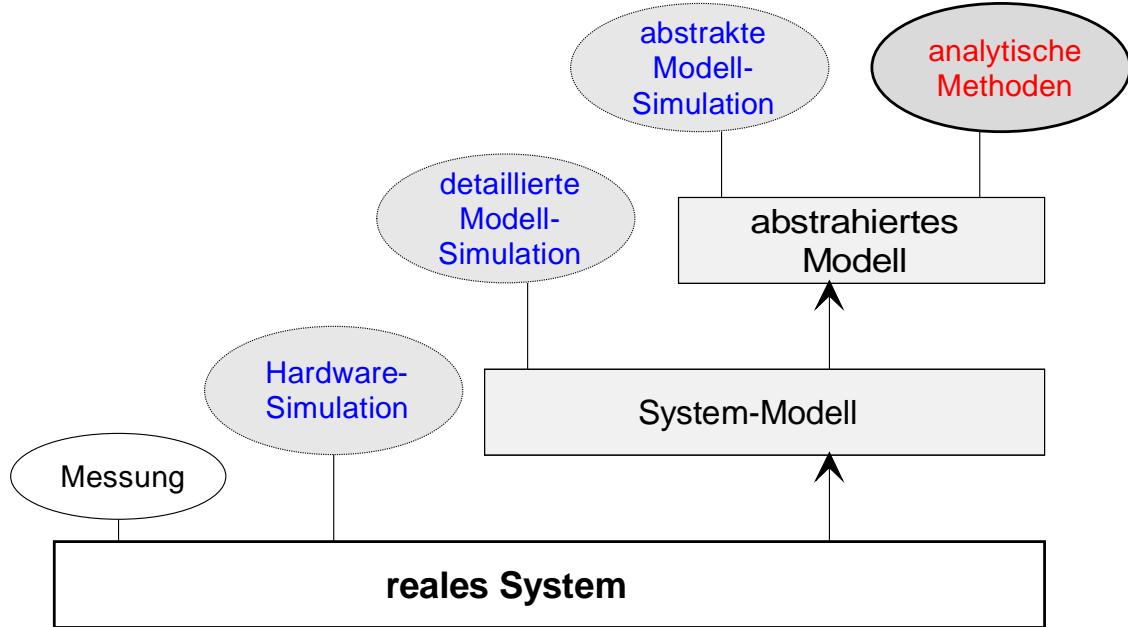
- Datenhaltung
 - Summe Σ^k der k-ten Potenzen der beobachteten Messwerte gewichtet mit der Zeit, zum Zeitpunkt $t_n \leq t < t_{n+1}$:
$$\Sigma^k = \sum_{0 \leq i < n} (X_i)^k \cdot (t_{i+1} - t_i)$$
 - Beginn der Datenerfassung t_{start} ,
 - Zeit der letzten Stichprobenübergabe t_{last}
 - Letzte übergebene Stichprobe X_{last}
 - Ende der Datenerfassung t_{stop}
- Aktionen zu Beginn der Statistikerhebung zum Zeitpunkt t_0
 - $X_{last} = X_0, t_{start} = t_0, t_{last} = t_0$
- Aktionen bei Übergabe der Stichprobe X_i zum Änderungszeitpunkt t_i
 - $\Sigma^k = \Sigma^k + (X_{last})^k \cdot (t_i - t_{last})$
 - $X_{last} = X_i, t_{last} = t_i$
- Aktionen bei Ende der Statistikerhebung zum Zeitpunkt t
 - $t_{stop} = t$
 - $\Sigma^k = \Sigma^k + (X_{last})^k \cdot (t_{stop} - t_{last})$
- Berechnung der Momente
 - $\bar{X}^k = \frac{\Sigma^k}{t_{stop} - t_{start}}$
 - Darauf basierend Berechnung weiterer statistischen Größen, z.B. der empirischen Varianz nach Gleichung 2.30

1.5.4 Erfassungsobjekte für Verteilungen von zeitkontinuierlichen Messwerten

- Datenhaltung
 - Wie bei Histogramm für diskrete Messwerte: Zähler $bin(j)$ für Intervalle mit $Range(j)$,
 - Wie bei Erfassung von Momenten zeitkontinuierlicher Messwerte:
 $t_{start}, t_{last}, X_{last}, t_{stop}$
- Aktionen zu Beginn der Statistikerhebung zum Zeitpunkt t_0
 - $X_{last} = X_0, t_{start} = t_0, t_{last} = t_0$
- Aktionen bei Übergabe der Stichprobe X_i zum Änderungszeitpunkt t_i
 - If $X_{last} \in Range(j)$, increment $bin(j)$ by $(t_i - t_{last})$
 - $X_{last} = X_i, t_{last} = t_i$
- Aktionen bei Ende der Statistikerhebung zum Zeitpunkt t_{stop}
 - If $X_{last} \in Range(j)$, increment $bin(j)$ by $(t - t_{last})$
 - $t_{stop} = t$
- Berechnung der relativen Häufigkeiten
 - $$h_j = \frac{bin(j)}{t_{stop} - t_{start}}$$

2 Einführung in die Simulationstechnik

2.1 Leistungsbewertungsmethoden und Simulation (LK 1.2)

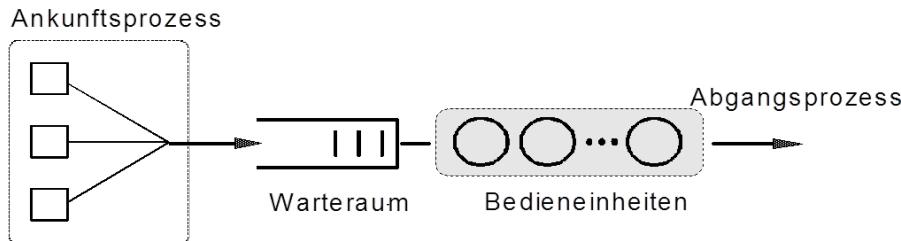


- Messung am realen System: nur möglich, wenn dieses schon existiert und zugänglich ist
- Hardware-Simulation: die Hardware eines realen Systems wird mit Hilfe von Software simuliert
- Detailliertes System-Modell: Abstraktion des realen Systems, Implementierungsdetails wie Hardware-Schaltungen spielen keine Rolle
- Detaillierte Modell-Simulation: Simulation des detaillierten System-Modells mit sehr vielen Details, die möglicherweise nicht alle einen entscheidenden Einfluss auf das Modell-Verhalten haben
- Abstrahiertes System-Modell: Abstraktion des realen Systems, nur noch die für das Verhalten des System wichtigen Eigenschaften werden z.B. als Warteschlange modelliert
- Abstrakte Modell-Simulation: Simulation des abstrahierten System-Modells
- Analytische Methoden: Direkte Quantifizierung des Verhaltens des abstrahierten System-Modells mit Hilfe von Mathematik und ohne Simulation

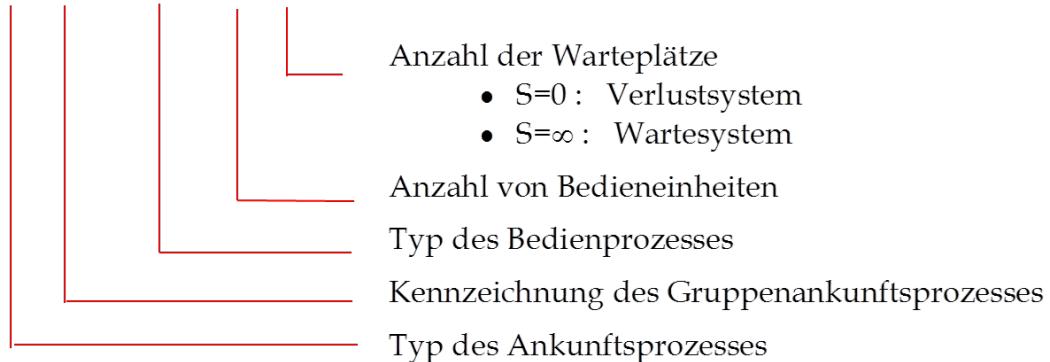
2.2 Ereignistreue Simulation und ihre Anwendungen

- Simulation: Nachbildung der Struktur und des Zustandes des Systems sowie seiner dynamischen Entwicklung in einem Programm
- Zeittreue bzw. ereignistreue Simulation: Simulationsverfahren, bei dem sich der Systemzustand nur zu bestimmten Ereignissen verändert; diese Ereignisse sind mit einem Zeitpunkt assoziiert, der eine Entsprechung im Realsystem hat.
- Hauptaufgaben:
 - Bei Untersuchung eines Systems, dessen Verhalten relativ unbekannt ist, nutzt man ereignistreue Simulation um das Verhalten des Systems zu erforschen.
 - Simulation ist notwendig, solange das System nicht mit analytischen Methoden modelliert und untersucht werden kann.
 - Simulationsergebnisse helfen analytische Modelle zu finden, denn sie geben Hinweise auf das Verhalten eines Systems
 - Simulationen dienen auch zur Validierung von Ergebnissen, die auf Basis analytischer Modelle und Methoden gewonnen wurden
 - Simulationsergebnisse zu produzieren ist bei komplexen Systemen oft einfacher, aber analytische Modelle haben den Charme, dass sie ein gewisses Systemverständnis ausdrücken.
 - Simulation wird auch zur Veranschaulichung bzw. Animation von Systemabläufen genutzt.

2.3 Notation für einstufige Warteschlangenmodelle



$GI^{[X]}/GI/n - S$ (Kendall-Notation)

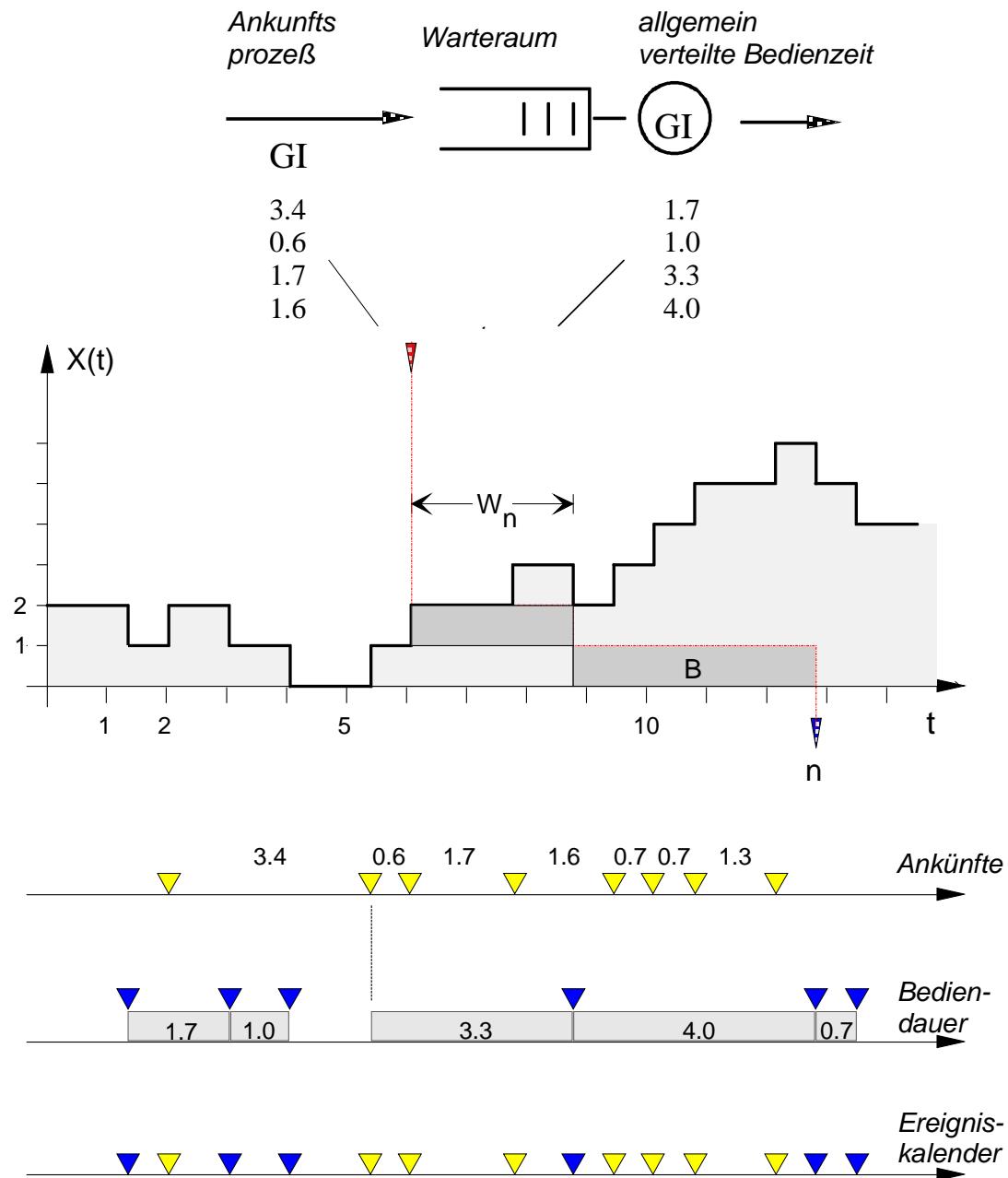


- GI : (general independent)
 - Ankunftsprozesse mit Erneuerungseigenschaft (Erneuerungsprozess)
 - Bedienprozesse, die jeweils mit Hilfe einer Zufallsvariablen beschrieben werden können;
 - Realisierungen dieser Zufallsvariablen sind statistisch unabhängig voneinander
- D : Deterministisch
- M : Markov
 - Die zugehörige Zufallsvariable ist exponentiell verteilt.
 - Ein M-Ankunftsprozess ist dementsprechend ein Poisson-Prozess
- E_k : Erlang-k-verteilt
- H_k : Hyperexponentiell verteilt, k-ter Ordnung

2.4 Komponenten und Organisation ereignistreuer Simulationen

2.4.1 Einführendes Beispiel: Simulation eines einstufigen GI/GI/1-S Wartesystems

- Anwendungsbeispiel: Ausgehendes Interface eines Routers
- Zeitdiskrete Simulation des theoretischen Models
- Systemzustand $X(t)$: Anzahl der Anfragen im System
- Änderungen des Systemzustands durch Ankunfts- und Bedienendeereignisse



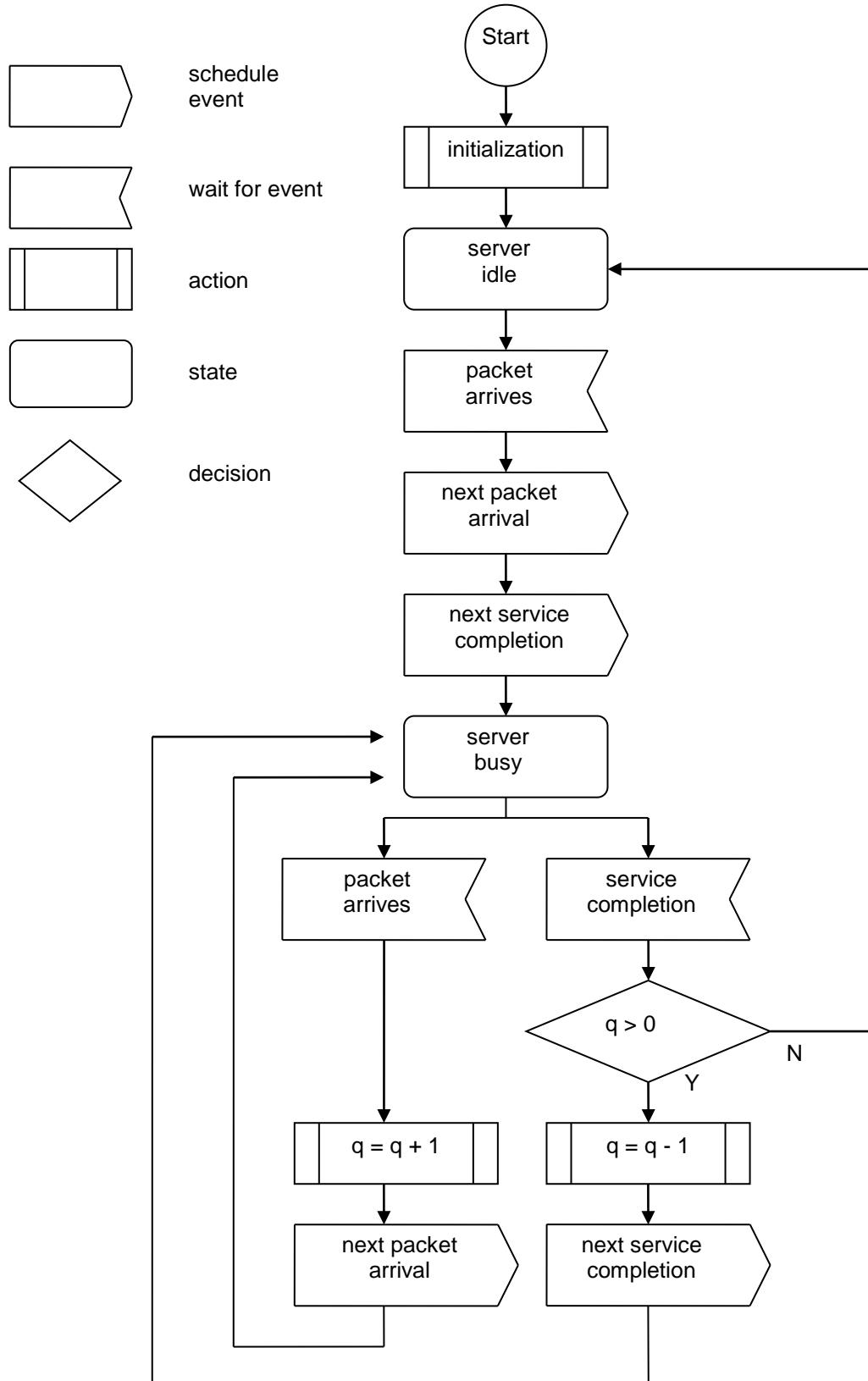
2.4.2 Aufbau einer „Discrete Event Simulation (DES)“ (LK 1.3.2)

- Simulation Clock (Simulationsuhr)
 - Wiedergabe der aktuellen Systemzeit
 - Implementierung der Simulationszeit üblicherweise als *long int* → Warum?
 - Umrechnung von Echtzeit auf Simulationszeit notwendig
- Event (Ereignis)
 - Mit einem diskreten Simulationszeitpunkt assoziiert
 - Beispiele: nächste Paketankunft, nächstes Bedienende
 - Führt zu diesem Zeitpunkt eine Event Routine aus, die eine Aktion im System verursachen, z.B. den System State verändern kann.
 - Verwaltung in der Event Chain (Ereignisliste)
- Statistic Counters
 - Variablen zur Speicherung von statistischen Informationen über das Systemverhalten
 - Discrete counters, z.B. Paketgrößen
 - Time-weighting counters, z.B. Warteschlangenlängen
 - Momente, Konfidenzintervalle, Output-Generierung
- Hauptprogramm
 - Definition und Initialisierung des System State (Systemzustand: Inhalt von Variablen, Listen, ...)
 - Initialisierung der Event Chain
 - Kontrolle und Abarbeitung der Event Chain
 - Bestimmung des nächsten Ereignisses in der Event Chain
 - Vorstellen der Simulationsuhr auf den Zeitpunkt dieses Ereignisses
 - Ausführen der Event Routine dieses Ereignisses
 - Kontrolle der Simulationsdauer
 - Anzahl bestimmter Events
 - Vorgegebene Simulationszeit

- Statistische Signifikanz
- Report Generator (Ergebnisroutine):
 - Verständliche Ausgabe der Simulationsergebnisse am Ende der Simulation
 - Leistungsmaße basierend auf den Statistic Counters
- Bibliotheksrouterien (*library routines*) zur
 - Erzeugung von Zufallsgrößen, Erhebung von Statistiken, ...

2.4.3 Hauptmodellierungsansätze (LK 3.3)

- Ereignisorientiert (*event-scheduling approach*)
 - Sequentielle Abarbeitung der Ereignisse nach obigem Konzept (aufsteigende Simulationszeit)
 - Kein Fortschreiten der Simulationszeit während der Bearbeitung eines Ereignisses, Beispiele:
 - Inkrementieren des Warteschlangenfüllstandes bei Paketankunft
 - Dekrementieren des Warteschlangenfüllstandes bei Bedienende
 - Übliches Konzept für einfache Eigenimplementierungen (siehe Übung!)
- Prozessorientiert
 - Fokus der Betrachtung: Objekt (z.B. Warteschlange)
 - Definition des dazugehörigen Prozesses durch ein Ablaufdiagramm, das auf Ereignisse reagiert ⇒ Ein Prozess ist eine zeitlich geordnete Folge von zusammenhängenden Ereignissen, die ein Objekt (d.h. Kunde, Job, Teil, etc.) erfährt.
 - Fortschreiten der Simulationszeit während eines Prozesses
 - Im Allgemeinen mehrere Eintrittspunkte für einen Prozess
 - Wegen Einfachheit Verwendung zur Modellbeschreibung in professionellen Simulationssprachen (z.B. OpNet)
 - Interne Umwandlung in Ereignisse
 - Beispiel: Prozessorientierte Modellbeschreibung eines GI/GI/1-∞ Warteschlangensystems



2.5 Klassifikation von Simulationsmodellen (LK 1.2)

- Deterministisch vs. stochastisch
 - *Deterministisch:* System enthält keine zufallsabhängigen Komponenten (z.B.: Differentialgleichungen zur Beschreibung chemischer Reaktion)
 - *Stochastisch:* Systemverhalten wird durch zufällige Ereignisse beeinflusst (z.B.: Warteschlangen- oder Lagerhaltungssysteme)
- Kontinuierlich vs. diskret
 - *Kontinuierlich:* Systemzustände ändern sich kontinuierlich (z.B. chemische Reaktionen, Differentialgleichungssysteme)
 - *Diskret:* Systemzustände ändern sich an diskreten Zeitpunkten (z.B. Lagerhaltungssysteme)
- Statisch vs. dynamisch
 - *Statisch:* System nur an einem Zeitpunkt betrachtet oder Zeit spielt keine Rolle (z.B.: Monte-Carlo-Simulation, LK 1.8.3, Auswürfeln von Snapshots mit nachfolgender deterministischer Auswertung)
 - *Dynamisch:* Modell repräsentiert das zeitliche Verhalten des Systems (z.B.: Simulation von Warteschlangen)

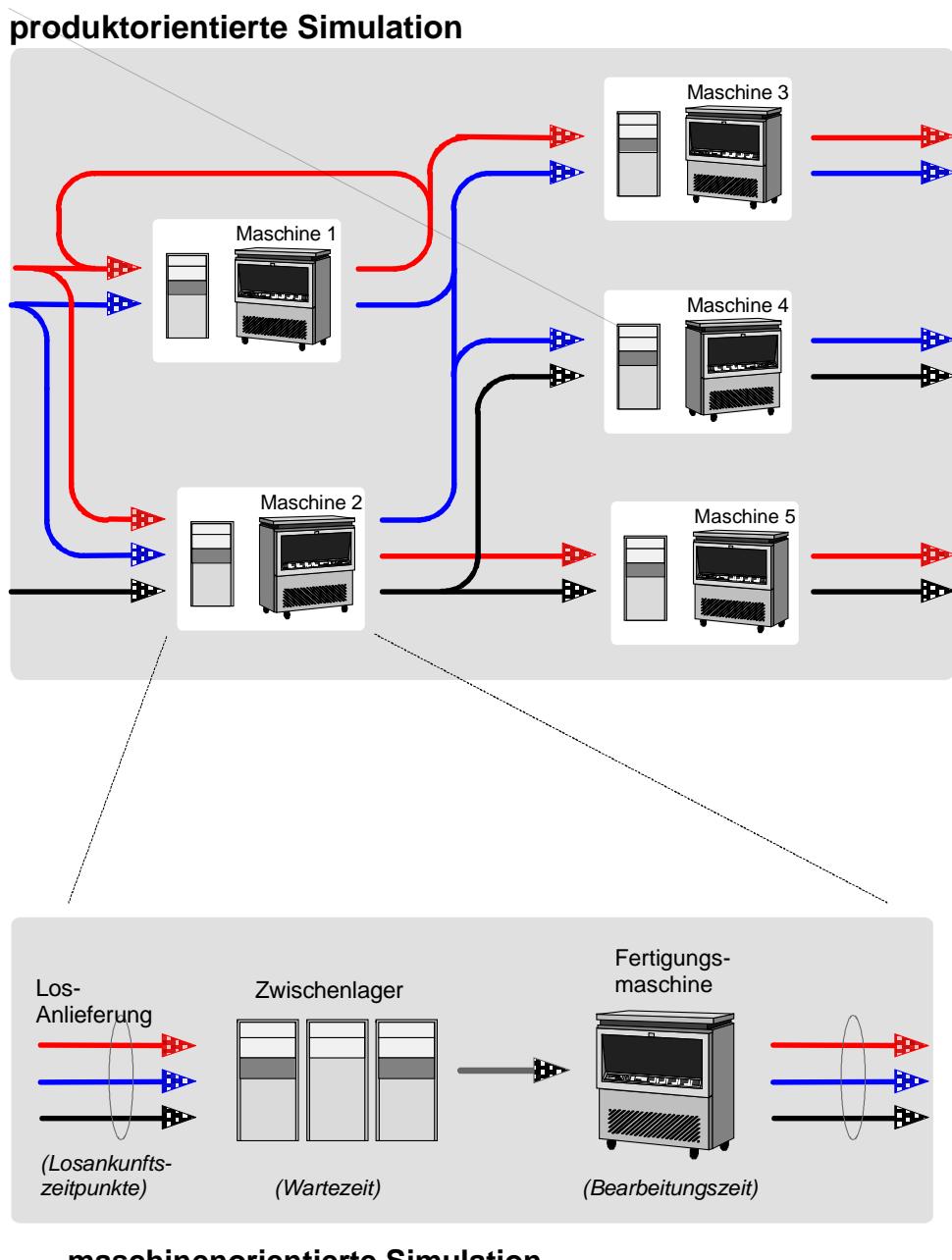
Definition: Eine *ereignistreue Simulation* (*discrete event simulation*, kurz: DES) ist die Modellierung der zeitlichen Entwicklung eines Systems durch eine Darstellung, bei der sich die Zustandsvariablen des Systems an diskreten Zeitpunkten augenblicklich ändern. Diese Zeitpunkte werden als *Ereignisse* bezeichnet.

Bemerkung: Es gibt aber auch Ereignisse, die nicht zu Zustandsänderungen führen, z.B. Simulationsende, zeitlich festgelegte Erfassung von Statistiken, etc.

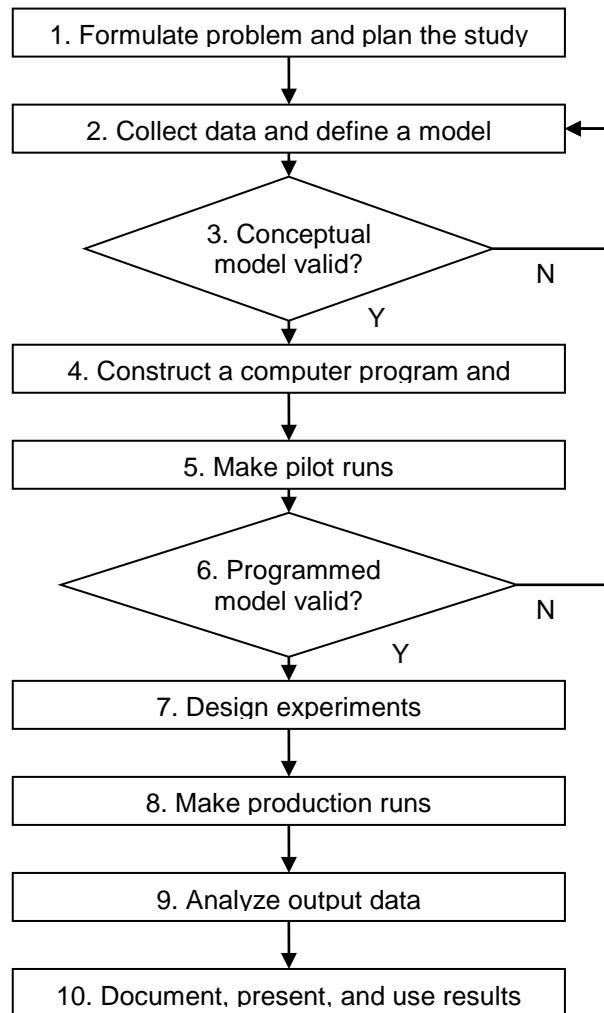
2.6 Fokus von Simulationen

Unterschiedliche Fragestellungen bedürfen unterschiedlicher Modelle und Simulationen

- Bestimmung der Durchlaufzeit eines Produktes in einer Fabrik bei einem bestimmten Produktmix
- Vergleich unterschiedlicher Scheduling-Mechanismen in einer Maschine



2.7 Bestandteile einer Simulationsstudie (LK 1.7)



1. Problemformulierung und Planung der Studie
 - Kickoff-Workshop, Interview mit Sachexperten
 - Festlegen der Fragestellung und der Leistungsmaße
2. Datensammlung und Modelldefinition
 - Herausfinden der wesentlichen Eigenschaften des Systems in Bezug auf die Fragestellung
 - Festlegen eines geeigneten Abstraktionsmodells
 - Erhebung statistischer Größen über das zu untersuchende System (z.B. Anzahl der Maschinen, Verteilungen von Zwischenankunftszeiten, Annahmen über nicht messbare Größen)
3. Validierung des Konzeptes
 - Rücksprache mit den Sachexperten

4. Erstellung und Verifikation eines Simulationsprogramms
 - Wahl einer Programmiersprache (C++, Java, ...) oder eines Simulationstools (ns3, OMNeT++, OpNet, ...)
 - Debugging
5. Pilotläufe
6. Validierung des Simulationsprogramms
 - Vergleich der Ergebnisse mit originalen Messdaten (falls möglich)
 - Plausibilitätsprüfung der Ergebnisse zusammen mit Sachexperten
 - Durchführung von Sensitivitätsanalysen zur Bestimmung signifikanter Einflüsse auf die Systemleistung mit anschließender Modellverfeinerung
7. Design von Experimenten
 - Erschöpfende Systemsimulation wegen Begrenztheit von Simulationsressourcen nicht möglich
 - Nur aussagekräftige Experimente interessant
 - Leistungsbewertung unterschiedlicher Systemkonfigurationen bei gleichen Systemmechanismen
 - Leistungsbewertung unterschiedlicher Systemmechanismen bei gleicher Systemkonfiguration
 - Verfahren zur Erhöhung der Zuverlässigkeit der Ergebnisse
 - Bestimmung geeigneter Aufwärmphasen
 - Bestimmung der Simulationslängen
8. Ergebnisläufe
9. Analyse der Simulationsergebnisse
 - Testen der statistischen Zuverlässigkeit der Ergebnisse (Konfidenzintervalle)
 - Bewertung konkurrierender Systementwürfe
10. Dokumentation, Präsentation und Nutzung der Ergebnisse
 - Dokumentation des Modells, der Annahmen, der Experimente, der Ergebnisse
 - Graphische Aufbereitung zur Vermittlung der Ergebnisse, Erklärung der Ergebnisse, Vertrauen in die Ergebnisse durch Diskussion des Modells und seiner Validierung
 - Nutzung der Ergebnisse in Entscheidungsprozessen (durch den Auftraggeber)

Bemerkungen

- Nicht alle Simulationsstudien enthalten notwendigerweise jeden der obigen Schritte.
- Oft mehrere Zyklen in der Praxis
 - Systemkenntnis wächst mit dem Studienfortschritt
 - Modelländerungen auch in späteren Schritten möglich

3 Verteilungsfunktionen (Teil 1)

3.1 Erzeugung von Zufallsgrößen (LK 8.2)

3.1.1 Problemstellung

Messung

Gegeben: Stichproben (=Realisierungen) X_i der ZV X

Gesucht: VF F(x)

Simulation

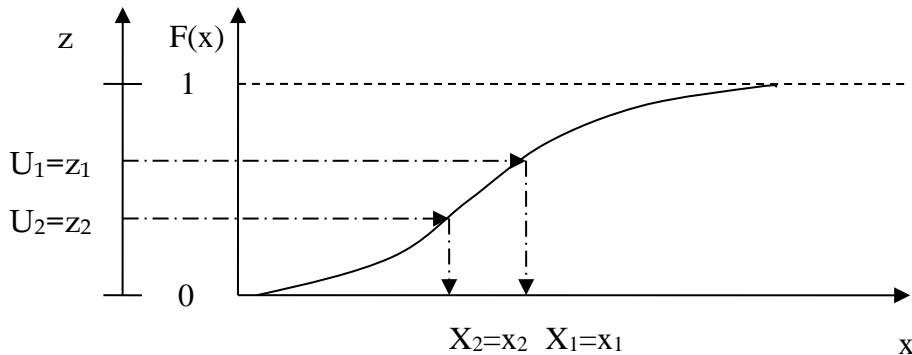
Gegeben: VF F(x)

Gesucht: Erzeugung von Stichproben X_i der ZV X

3.1.2 Prinzipielle Vorgehensweise

- Erzeugung von U(0,1) verteilten Zufallszahlen durch einen Zufallszahlengenerator (z.B. Pseudozufallszahlen durch Kongruenzmethoden, siehe Kapitel 17)
- Transformation dieser Zufallszahlen in gewünschte Stichprobe mit VF F(x)

3.1.3 Inversionsmethode (LK 8.2.1)



- Realisierung einer Zufallszahl $z_i \sim U(0,1)$
- Transformation von z_i in eine Stichprobe X_i gemäß der VF $F(x)$:

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

- Beispiel: Exponentielle VF

- $z_i = F(x_i) = 1 - e^{-\lambda \cdot x_i} \rightarrow x_i = -\frac{1}{\lambda} \ln(1 - z_i) = -\frac{1}{\lambda} \ln(z_i^*)$

- Aus Symmetriegründen $z_i^* = z_i$ möglich

- Problem: $F^{-1}(z)$ nicht immer analytisch ermittelbar

Lösung durch andere direkte Methoden: Ersatzverteilung durch

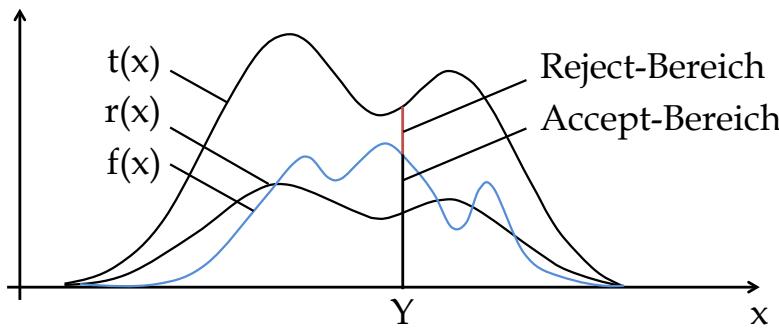
- Komposition aus ZV mit invertierbaren VF (LK 8.2.2)
 - Erlang-verteilte ZV: Summe (LK 8.2.3) von exponentiellen ZV
 - Hyperexponentiell verteilte ZV: Zufällige Wahl von exponentiellen ZV mit unterschiedlichem Parameter

- invertierbare Approximation einer Verteilungsfunktion (z.B. rationale Funktion)
 - Abramowitz & Stegun: Handbook of Mathematical Functions

Bei manchen Verteilungen nicht oder nicht effizient (Rechenzeit) anwendbar

3.1.4 Accept-Reject-Methode (LK 8.2.4)

- Indirekter Ansatz
 - Erzeuge eine Zufallszahl und generiere eine Stichprobe Y
 - Erzeuge eine weitere Zufallszahl und ermittle dadurch, ob Y als gesuchte Stichprobe X genommen werden kann
- Voraussetzung
 - Majorante $t(x)$ von $f(x)$, d.h. $\forall x: t(x) \geq f(x)$
 - Die Integralfunktion von $t(x)$ soll leicht invertierbar sein
 - $t(x)$ ist i.A. keine Dichtefunktion wegen
$$c = \int_{-\infty}^{\infty} t(x) dx \geq \int_{-\infty}^{\infty} f(x) dx = 1$$
 - Dichtefunktion $r(x)$ durch Normierung: $r(x) = \frac{1}{c} \cdot t(x)$
 - Invertierbare Verteilungsfunktion: $R(x) = \int_{-\infty}^x r(t) dt$



- Algorithmus zur Erzeugung der ZV X
 - Erzeuge Y gemäß Verteilungsfunktion $R(x)$
 - Erzeuge $U \sim U(0,1)$, unabhängig von Y
 - Entscheide
 - $X=Y$ falls $U \leq \frac{f(Y)}{t(Y)}$ (Accept)
 - Ansonsten verwirfe Y (Reject) und erneuter Versuch
- Y wird akzeptiert, falls Punkt $(Y, U \cdot t(Y))$ unter der Dichtekurve $f(x)$ liegt.
- Wichtig für gute Effizienz: $t(x)$ eng an $f(x)$ anliegend

- Beweis der Gültigkeit des Algorithmus: LK Appendix 8A

- $X \sim f(x)$

- Majorante: $t(x) \geq f(x)$

- $Y \sim r(x) = \frac{t(x)}{c} = \frac{t(x)}{\int_{-\infty}^{\infty} t(x) dx}$

- Ereignis A: $Y=y$ wird angenommen $\Leftrightarrow U \leq \frac{f(y)}{t(y)}$ mit $U \sim U(0,1)$

$$\Rightarrow P(A | Y = y) = \frac{f(y)}{t(y)}$$

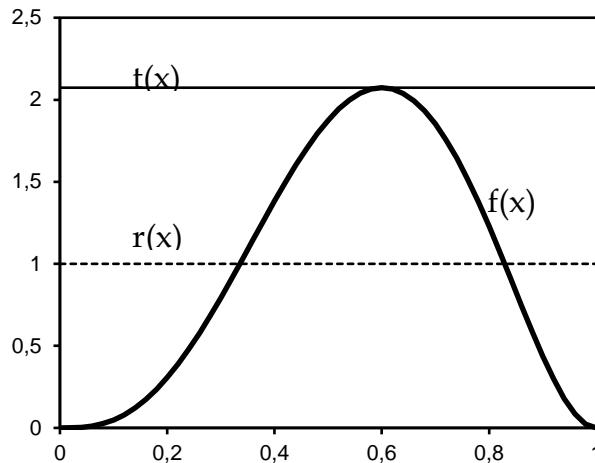
- $P(X \leq x) = P(Y \leq x | A) = \frac{P(A, Y \leq x)}{P(A)} =$

$$\frac{\int_{-\infty}^x P(A | Y = y) \cdot r(y) dy}{\int_{-\infty}^{\infty} P(A | Y = y) \cdot r(y) dy} =$$

$$= \frac{\int_{-\infty}^x P(A | Y = y) \cdot \frac{t(y)}{c} dy}{\int_{-\infty}^{\infty} \frac{t(y)}{c} dy} = \frac{\int_{-\infty}^x \frac{f(y)}{t(y)} \cdot \frac{t(y)}{c} dy}{\int_{-\infty}^{\infty} \frac{f(y)}{t(y)} dy} =$$

$$= \frac{\frac{1}{c} \int_{-\infty}^x f(y) dy}{\frac{1}{c}} = \int_{-\infty}^x f(y) dy = F(x) \quad \text{q.e.d.}$$

- Beispiel: Erzeugung von ZV gemäß der beta(3,4) Verteilungsfunktion
- $f(x) = \begin{cases} 60x^3(1-x)^2 & 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$



- Verteilungsfunktion ist Polynom höherer Ordnung und schwer zu invertieren.
- Maximum: (0.6 ; 2.0736)
- Majorante: $t(x) = \begin{cases} 2.0736 & 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$
- $\Rightarrow c = \int_0^1 t(x) dx = 2.0736$
- $\Rightarrow r(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases} \sim U(0,1)$
- Erzeugung der Zufallszahlen:
 - $Y \sim U(0,1)$
 - $U \sim U(0,1)$
 - $U \leq \frac{60Y^3(1-Y)^2}{2.0736} ?$

3.2 Kontinuierliche Verteilungsfunktionen (LK 6.2.2, 6.2.4 und 8.3)

3.2.1 Gleichverteilung: ZV $X \sim U(a,b)$ (LK 8.3.1)

- VDF: $f(x)=1/(b-a)$, $X \in [a; b]$
- Range $[a, b]$
- VF: $F(x)=(x-a)/(b-a)$
- $E[X]=(a+b)/2$
- $VAR[X]=(b-a)^2/12$
- Mode: --
- Erzeugung durch Inversion: $U \sim U(0,1)$, $X = a + (b - a)U$.

3.2.2 Exponentialverteilung: ZV $X \sim \text{expo}(\lambda)$ (LK 8.3.2)

- VDF: $f(x)=\lambda \cdot e^{-\lambda \cdot x}$ für $x \geq 0$
- VF: $F(x)=1-e^{-\lambda \cdot x}$
- Range: $[0, \infty)$
- $E[X]=1/\lambda$
- $VAR[X]=1/\lambda^2$
- $c_{\text{var}}=1$
- Mode: 0
- Erzeugung durch Inversion: $U \sim U(0,1)$, $X = -\ln(U)/\lambda$
- Exponentiell verteilte Zwischenankunftszeiten ergeben einen **Poisson-Prozess**

3.2.2.1 Definition: Rate

- Wahrscheinlichkeit, dass ein Ereignis innerhalb eines Intervalls dt eintritt: $P(A \leq dt)$
- Rate=Wahrscheinlichkeit/Intervalllänge dt , wenn die Intervalllänge dt infinitesimal klein wird: $\lim_{dt \rightarrow 0} \frac{P(A \leq dt)}{dt} = \lambda$
- Die Rate der exponentiellen Verteilungsfunktion erfüllt diese Definition:

$$\lim_{dt \rightarrow 0} \frac{1-e^{-\lambda \cdot dt}}{dt} = \lim_{dt \rightarrow 0} \frac{1-\left(1-\frac{\lambda \cdot dt}{1!}+\frac{(\lambda \cdot dt)^2}{2!}-\dots\right)}{dt} = \lambda$$

(hint: <http://de.wikipedia.org/wiki/Taylorreihe>)

3.2.2.2 Anwendung: Abstand von zufälligen (Beobachtungs)Punkten

Problem: Punkte sollen zufällig mit einem mittleren Abstand Δ und mit gleicher Dichte über eine beliebig lange Strecke verteilt sein. Welcher Verteilungsfunktion folgt der Abstand zwischen den Punkten?

Lösung

- Gegeben sei ein Kreis der Länge $n \cdot \Delta$ mit n zufälligen Punkten P_i , welche zufällig auf dem Kreis verteilt sind.

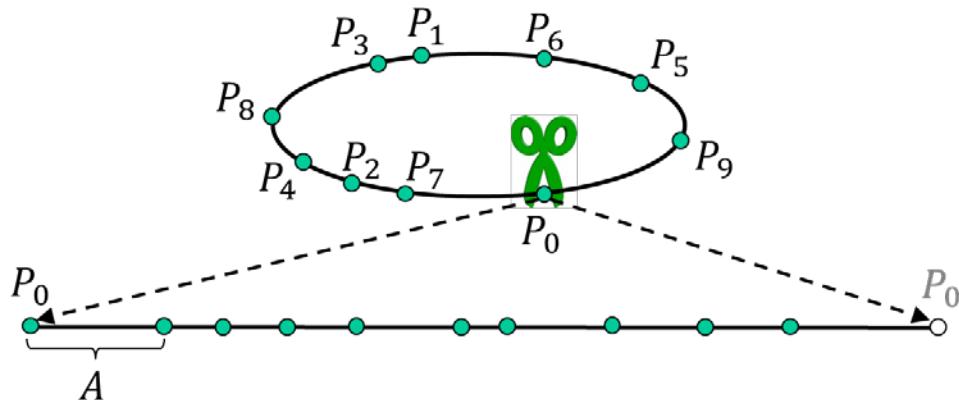


Abbildung 1: Der Abstand zwischen zufälligen Punkten auf einem Kreis ist gemäß Konstruktion identisch verteilt und entspricht der Zufallsvariablen A .

- Wir schneiden den Kreis o.B.d.A. an Punkt P_0 auf.
 - Die Zufallsvariable D_i (für $0 < i < n$) bezeichne die Entfernung von P_i zu P_0 im Uhrzeigersinn.
 - Gemäß Konstruktion ist sie zufällig und uniform zwischen 0 und $n \cdot \Delta$ verteilt und kann somit mit $D_i(t) = \frac{t}{n \cdot \Delta}$ angegeben werden.
 - Die Zufallsvariable A bezeichne die Entfernung zwischen Punkt P_0 und seinem Nachbarn im Uhrzeigersinn.
 - Die Verteilung von $A = \min_{0 < i < n} (D_i)$ berechnet sich berechnet als
- $$A(t) = 1 - \prod_{0 < i < n} (1 - D_i(t)) = 1 - \prod_{0 < i < n} \left(1 - \frac{t}{n \cdot \Delta}\right) = 1 - \left(1 + \frac{-t}{n \cdot \Delta}\right)^{n-1}.$$
- Für $n \rightarrow \infty$ erhalten wir $A(t) = \lim_{n \rightarrow \infty} 1 - \frac{\left(1 + \frac{-t}{n \cdot \Delta}\right)^n}{\left(1 + \frac{-t}{n \cdot \Delta}\right)} = 1 - e^{\left(\frac{1}{\Delta}t\right)}$ für ein festes t .
 - Die Entfernung zwischen zufällig verteilten Punkten mit durchschnittlichem Abstand Δ folgt einer exponentiell Verteilungsfunktion mit Rate $\frac{1}{\Delta}$.

Folgerung 1: Zufällig und mit gleicher Dichte über eine beliebig lange Strecke verteilte Punkte können durch einen Poisson-Prozess realisiert werden.

Besonderheit: Gedächtnislosigkeit (Vorwärtsrekurrenzzeit $\sim \text{expo}(\lambda)$)

Folgerung 2: Bei einem Poisson-Prozess ist zu jedem Zeitpunkt die Dauer bis zum nächsten Ereignis unabhängig von der Zeit seit dem letzten Ereignis.

3.2.3 Erlang-Verteilung: ZV $X \sim k\text{-Erlang}(\lambda)$ (LK 8.3.3)

- ZV ist Summe von k $\text{expo}(\lambda)$ verteilten ZVs (Phasen)

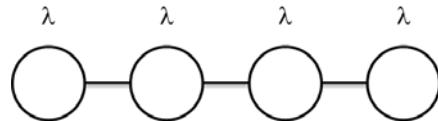


Abbildung 3-2: Phasendiagramm einer Erlang-4-Verteilung als Serienschaltung von 4 exponentiell verteilten Phasen.

- VDF: $f(x) = \begin{cases} \frac{(\lambda x)^{k-1}}{(k-1)!} \cdot \lambda \cdot e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$

- VF: $F(x) = \begin{cases} 1 - e^{-\lambda x} \cdot \sum_{0 \leq i < k} \frac{(\lambda x)^i}{i!} & x \geq 0 \\ 0 & x < 0 \end{cases}$

- Range: $[0, \infty)$
- $E[X] = k/\lambda$
- $\text{VAR}[X] = k/\lambda^2$
- $c_{\text{var}}[X] = 1/\sqrt{k}$
- Mode: $\frac{k-1}{\lambda}$
- Erzeugung durch Inversion:

$$U_i \sim U(0,1), X = \sum_{0 \leq i < k} -\frac{1}{\lambda} \cdot \ln(U_i) = -\frac{1}{\lambda} \cdot \ln(\prod_{0 \leq i < k} U_i)$$

3.2.4 Hyperexponentialverteilung: ZV $X \sim H(\lambda_i, p_i, 0 \leq i < k)$

- ZV ist Kombination von k $\text{expo}(\lambda)$ verteilten ZVs (Phasen)

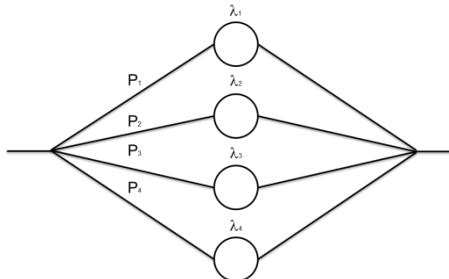


Abbildung 3-3: Phasendiagramm einer Hypoexponentiellen Verteilung: Parallelschaltung von 4 exponentiell verteilten Phasen, von denen genau eine ausgeführt wird. Vorgegebene Wahrscheinlichkeiten beeinflussen die zufällige Auswahl dieser Phase.

- VDF: $f(x) = \sum_{0 \leq i < k} p_i \cdot \lambda_i e^{-\lambda_i \cdot x}$
- VF: $F(x) = 1 - \sum_{0 \leq i < k} p_i e^{-\lambda_i \cdot x}$
- Range: $[0, \infty)$
- $E[X] = \sum_{0 \leq i < k} (p_i / \lambda_i)$
- $c_{var}[X] = \sqrt{\frac{2 \cdot \sum_{0 \leq i < k} (p_i / \lambda_i^2)}{(\sum_{0 \leq i < k} (p_i / \lambda_i))^2} - 1}$ (aus Tran-Gia, Kapitel 1.3)
- Erzeugung durch $X = \begin{cases} X_0 \sim \text{expo}(\lambda_0) & \text{with probability } p_0 \\ \dots & \dots \\ X_{k-1} \sim \text{expo}(\lambda_{k-1}) & \text{with probability } p_{k-1} \end{cases}$ mit $\sum_{0 \leq i < k} p_i = 1$
 - Erzeuge Zufallszahlen U_1, U_2
 - $i = \min(m : \sum_{0 \leq j \leq m} p_j \geq U_1) //$ bestimme den Phasenindex
 - Benutze U_2 zur Erzeugung von $X \sim \text{expo}(\lambda_i)$
- Spezialfall: Hyperexponentielle Verteilung 2. Ordnung (H_2) zur Generierung einer ZV mit vorgegebenem $E[X]$ und $c_{var}[X]$
 - Symmetriebedingung: $\frac{p_1}{\lambda_1} = \frac{p_2}{\lambda_2}$

$$\text{Geeignete Parameterwahl: } \lambda_{1,2} = \frac{1}{E[X]} \cdot \left(1 \pm \sqrt{\frac{c_{var}^2 - 1}{c_{var}^2 + 1}} \right)$$

3.2.5 Normalverteilung: ZV $X \sim N(\mu, \sigma^2)$ (LK 8.3.6)

- $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- Range: $(-\infty, \infty)$
- $E[X] = \mu$, $\text{VAR}[X] = \sigma^2$
- Skalierbarkeit: $X \sim N(0,1) \Rightarrow (\mu + \sigma X) \sim N(\mu, \sigma^2)$.
- Beispiel

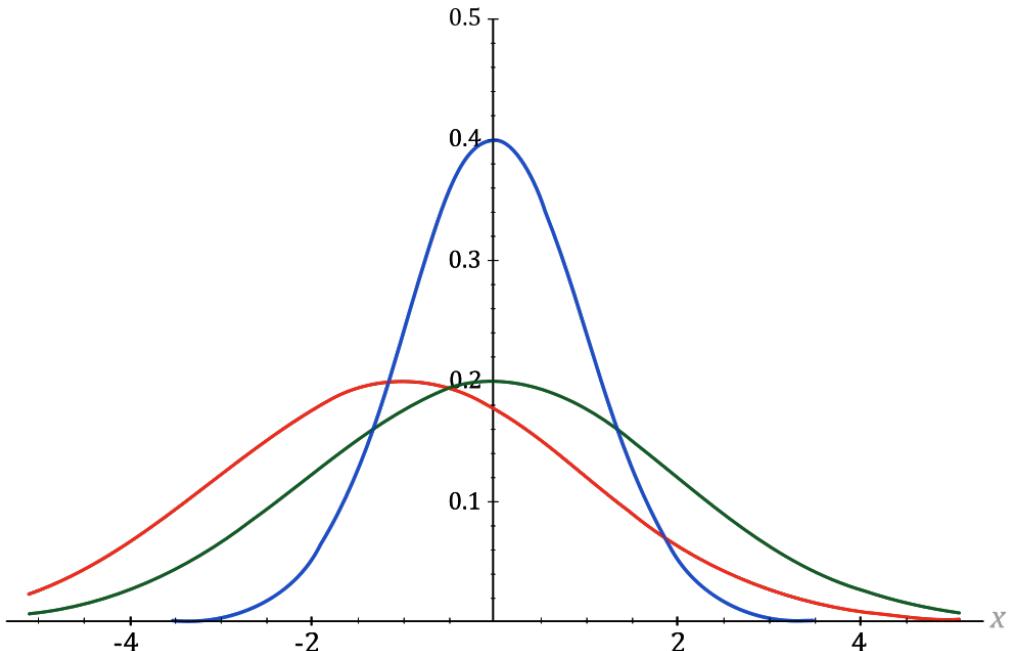


Abbildung 4: Dichtefunktionen der Normalverteilungen $N(0,1)$, $N(0,2)$, $N(-1,2)$.

- Erzeugung durch Accept-Reject Methode
 $U_1, U_2 \sim U(0,1)$, wobei U_1 und U_2 unabhängig sein müssen.

$$V_i = 2U_i - 1, W = V_1^2 + V_2^2.$$

falls $W \leq 1$

dann Accept: $Y = \sqrt{\frac{-2 \ln W}{W}}$, $X_1 = V_1 \cdot Y$, $X_2 = V_2 \cdot Y$ (2 Zufallsvariablen)

sonst Reject

3.3 Diskrete Verteilungen (LK 6.2.3 und 8.4)

3.3.1 Deterministische Verteilung mit Parameter γ

- Verteilung: $x(i) = \begin{cases} 1 & \text{für } i = \gamma \\ 0 & \text{sonst} \end{cases}$
- $E[X] = \gamma$
- $VAR[X] = 0$
- Range: γ
- Mode: γ

3.3.2 Diskrete Gleichverteilung: $ZV \sim DU(i,j)$ (LK 8.4.2)

- Verteilung: $p(k) = \begin{cases} \frac{1}{j-i+1} & \text{if } k \in \{i, i+1, \dots, j\} \\ 0 & \text{otherwise} \end{cases}$

- Range: $i \leq k \leq j$
- $E[X] = (i+j)/2$
- $VAR[X] = \frac{(j-i+1)^2 - 1}{12}$
- Mode: --
- Erzeugung durch Inversion:
 $U \sim U(0, 1)$
 $X = i + \lfloor (j - i + 1) \cdot U \rfloor$

3.3.3 Beliebige diskrete Verteilung (LK 8.4.3)

- $p(x) = \begin{cases} p_k & \text{if } x = x_k, 0 \leq k < n \\ 0 & \text{otherwise} \end{cases}$

- Range: beliebig aber diskret
 - Erzeugung durch Inversion:
 $U \sim U(0,1).$

$$X = x_k \text{ mit } k = \min(i : \sum_{0 \leq j \leq i} p_j \geq U)$$

Hoher Suchaufwand durch intelligente Datenstruktur vermeidbar

3.3.4 Bernoulli-Verteilung: ZV ~ Bernoulli(p) (LK 8.4.1)

- „Einfacher Münzwurf“
- Verteilung: $p(k) = \begin{cases} 1-p & \text{if } k=0 \\ p & \text{if } k=1 \\ 0 & \text{else} \end{cases}$

- $E[X]=p$
- $VAR[X]=p \cdot (1-p)$
- $c_{\text{var}} = \sqrt{\frac{1-p}{p}}$
- Mode: 0 bzw. 1
- Erzeugung durch Inversion:
 $U \sim U(0, 1)$

Falls $U < p$, dann $X = 1$, sonst $X = 0$.

3.3.5 Binomial-Verteilung: ZV ~ Binom(n,p) (LK 8.4.4)

- $p(k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$ mit $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ und $0 \leq k \leq n$, $k \in \mathbb{N}_0$
- $E[X]=n \cdot p$
- $VAR[X]=n \cdot p \cdot (1-p)$
- $c_{\text{var}}[X] = \sqrt{\frac{1-p}{n \cdot p}}$
- Besonderheiten:
 - Für $n \rightarrow \infty$ ist $\text{Binom}(n,p)$ durch $N(E[X], VAR[X])$ approximierbar
 - Für $p \ll 1$ und n sehr groß ist $\text{Binom}(n,p)$ durch $\text{Poisson}(E[X])$ approximierbar
- Erzeugung durch Komposition wegen $\text{Binom}(n, p) \sim \sum_{0 \leq i < n} \text{Bernoulli}(p)$:

Erzeuge $Y_i \sim \text{Bernoulli}(p)$, $X = \sum_{0 \leq i < n} Y_i$

- Visualisierung

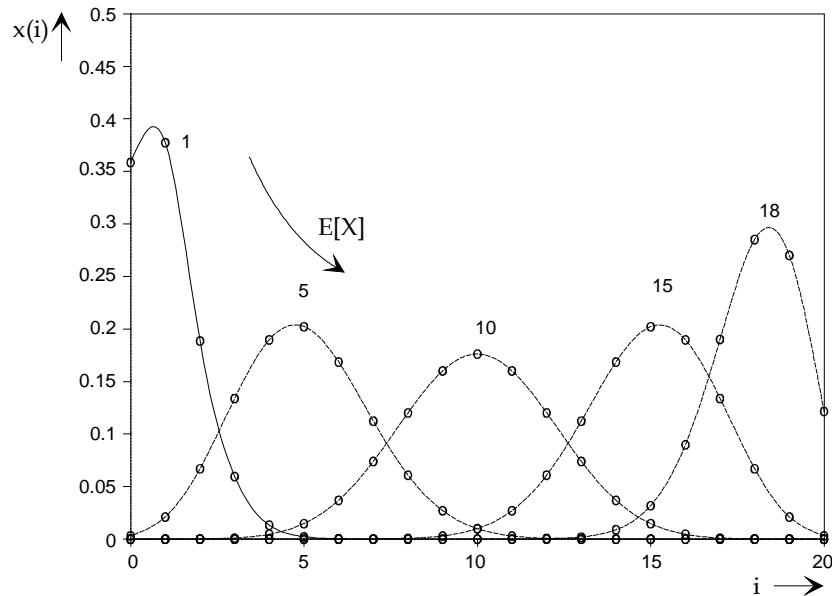


Abbildung 5: Binomialverteilungen.

3.3.6 Geometrische Verteilung: ZV $\sim \text{Geom}(p)$ (LK 8.4.5)

- „Anzahl der Fehlversuche bis zum ersten Erfolg eines Bernoulli-Experiments mit Erfolgswahrscheinlichkeit p “
- $p(k) = (1 - p)^k \cdot p, k \geq 0$
- VF: $F(x) = 1 - (1 - p)^{\lfloor x \rfloor + 1}$

Beweis: Geometrische Reihe: $\sum_{i=0}^k q^i = \frac{1 - q^{k+1}}{1 - q}, \sum_{i=0}^{\infty} q^i = \frac{1}{1 - q}$; setze $q = 1 - p$;

$$\begin{aligned} F(x) &= P(X \leq (k = \lfloor x \rfloor)) = p \cdot \sum_{i=0}^k (1 - p)^i = p \cdot \frac{1 - (1 - p)^{k+1}}{1 - (1 - p)} \\ &= p \cdot \frac{1 - (1 - p)^{k+1}}{p} = 1 - (1 - p)^{\lfloor x \rfloor + 1} \end{aligned}$$

- $E[X] = (1 - p)/p$
- $\text{VAR}[X] = (1 - p)/p^2$
- $c_{\text{var}}[X] = \frac{1}{\sqrt{(1 - p)}}$
- Mode 0

- Erzeugung durch Inversion:

$$U \sim U(0,1)$$

$$X = \left\lfloor \frac{\ln(U)}{\ln(1-p)} \right\rfloor$$

- Visualisierung

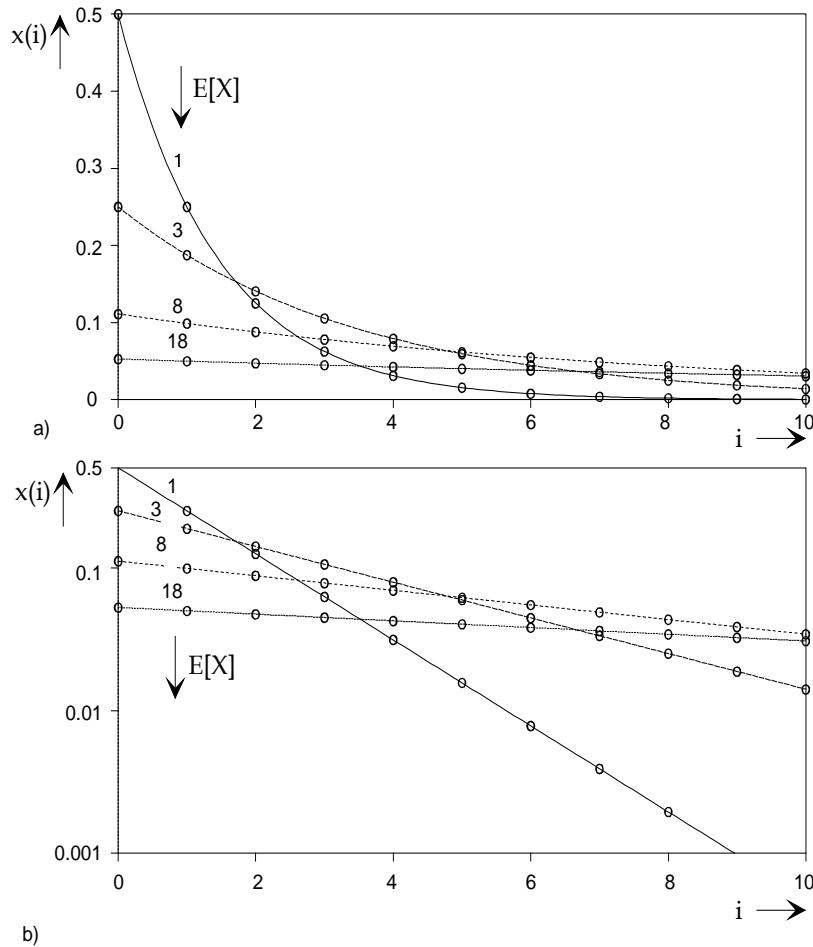


Abbildung 6: Geometrische Verteilungen.

3.3.7 „Um 1 Verschobene“ Geometrische Verteilung

- „Anzahl aller Versuche bis das Bernoulli-Experiment mit Erfolgswahrscheinlichkeit p einen Erfolg zeigt“
- $p(k) = (1-p)^{k-1} \cdot p, k \geq 1$
- $E[X] = (1-p)/p + 1 = 1/p$
- $VAR[X] = (1-p)/p^2$ (dieselbe wie bei der nicht verschobenen Geom)
- $c_{var}[X] = \sqrt{1-p}$

3.3.8 Negativ-binomiale Verteilung: ZV $\sim \text{NegBin}(s,p)$ (LK 8.4.6)

http://en.wikipedia.org/wiki/Negative_binomial_distribution

- „Anzahl der Fehlversuche bis sich s Erfolge bei Bernoulli-Experimenten eingestellt haben. Die s erfolgreichen gehören wie bei der Geom(p)-Verteilung nicht dazu!“
- $\text{NegBin}(s,p) \sim \sum_{0 \leq i < s} \text{Geom}(p)$
- $p(k) = \binom{s+k-1}{k} \cdot p^s \cdot (1-p)^k = \binom{-s}{k} \cdot p^s \cdot (-1-p)^k$

Zur Umformung : <http://de.wikipedia.org/wiki/Binomialkoeffizient>

- $E[X] = s \cdot (1-p)/p$
- $\text{VAR}[X] = s \cdot (1-p)/p^2$
- $c_{\text{var}}[X] = \frac{1}{\sqrt{s \cdot (1-p)}}$
- Mode: $\begin{cases} y \text{ und } y+1 & \text{falls } y \text{ ganzzahlig} \\ \lfloor y \rfloor + 1 & \text{sonst} \end{cases}$ mit $y = [s^*(1-p)-1]/p$,
- Erzeugung durch Komposition:

$Y_1, Y_2, \dots, Y_s \sim \text{geom}(p)$ und unabhängig.

$X = Y_1 + Y_2 + \dots + Y_s$.

- Erweiterung auf reelle s möglich:

$$p(k) = \frac{\Gamma(s+k)}{\Gamma(s) \cdot k!} \cdot p^s \cdot (1-p)^k \quad \text{im Spezialfall } s \text{ ganzzahlig: } \Gamma(s) = (s-1)! \text{ (siehe später!)}$$

- Der Grenzwert der Negbin für sehr große p und s nähert sich einer Poisson-Verteilung an:

$$\text{Substituiere } p = \frac{\omega}{\lambda + \omega} \text{ und } s = \omega$$

$$\text{Dann gilt: } p(k) = \frac{y^k}{k!} \cdot \frac{\Gamma(\omega+k)}{\Gamma(\omega) \cdot (y+\omega)^k} \cdot \frac{1}{\left(1 + \frac{y}{\omega}\right)^\omega}$$

und im Grenzfall bekommen wir $\lim_{\omega \rightarrow \infty} p(k) = \frac{y^k}{k!} \cdot 1 \cdot \exp(-y)$ (Poissonverteilung)

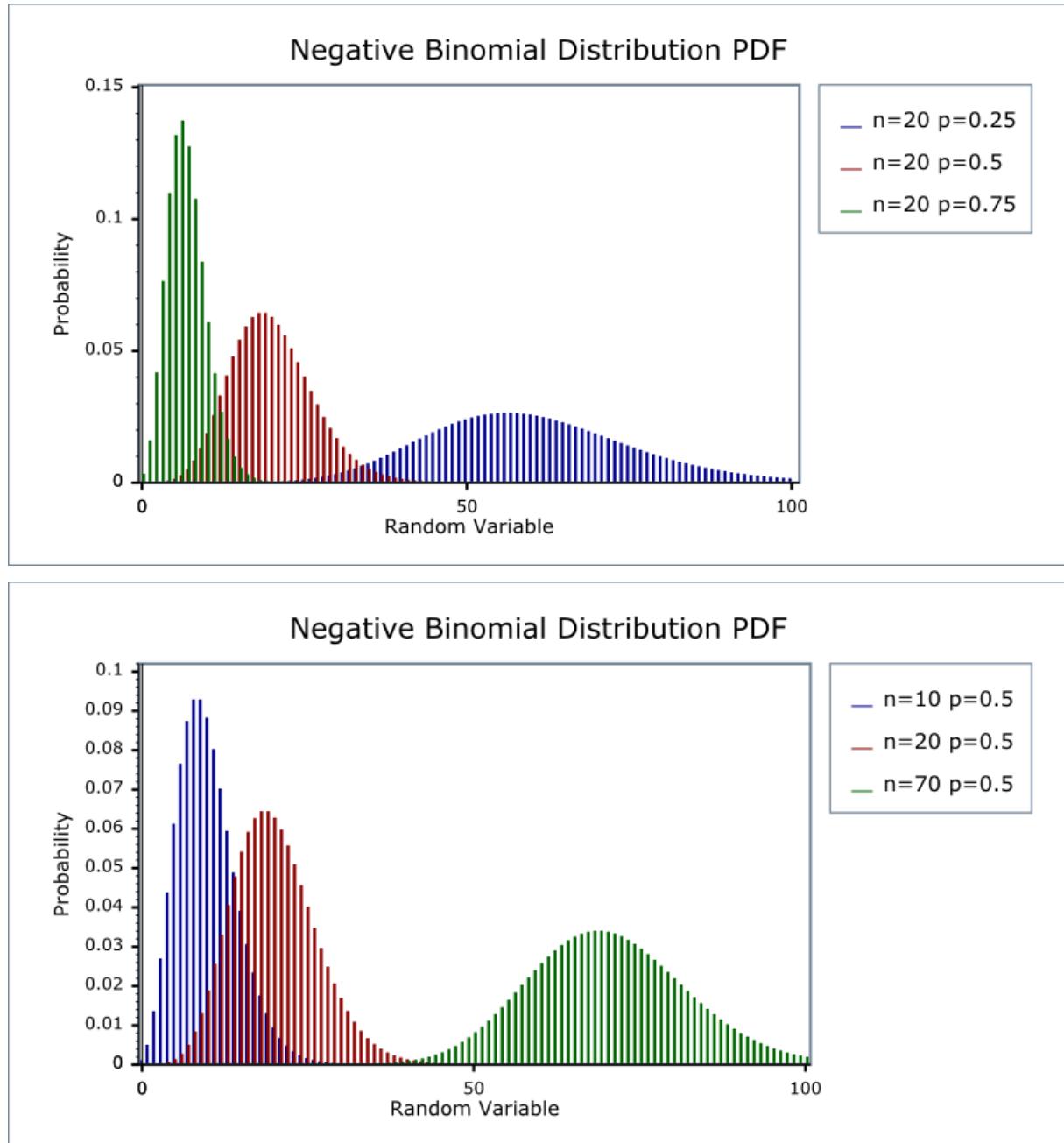


Abbildung 7: Negativ-binomiale Verteilung: Einfluss der Parameter.

Quelle:

http://www.boost.org/doc/libs/1_51_0/libs/math/doc/sf_and_dist/html/math_toolkit/dist/dist_ref/dists/negative_binomial_dist.html

3.3.9 Poisson-Verteilung: ZV \sim Poisson(y) (LK 8.4.7)

- Anzahl der Ankünfte eines Poisson-Prozesses (= exponentielle Zwischenankunftszeiten) mit Rate λ in einem Intervall der Länge τ und $y = \lambda * \tau$

- $p(k) = \frac{y^k}{k!} \cdot e^{-y}$

Mit $e^y = \sum_{0 \leq k < \infty} \frac{y^k}{k!}$ ist klar, dass es sich um eine Verteilung handelt.

- $E[X] = y$

Beweis: $E[X] = \sum_{0 \leq k < \infty} k \cdot \frac{y^k}{k!} \cdot e^{-y} = y \cdot e^{-y} \cdot \sum_{1 \leq k < \infty} \frac{y^{k-1}}{(k-1)!} = y \cdot e^{-y} \cdot e^y = y$

- $VAR[X] = y$

- $c_{\text{var}}[X] = \frac{1}{\sqrt{y}}$

- Mode $\begin{cases} y \text{ und } y-1 & \text{falls } y \text{ ganzzahlig} \\ \lfloor y \rfloor & \text{sonst} \end{cases}$

- Beispiel

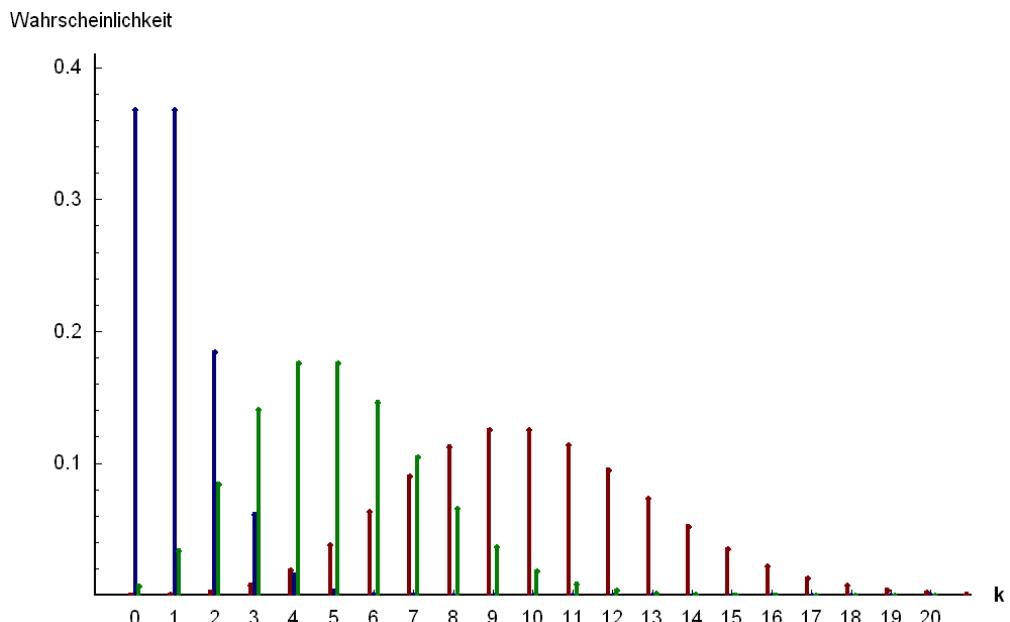


Abbildung 8: Wahrscheinlichkeitsfunktion der Poisson-Verteilung für $\lambda = 1$ (blau), $\lambda = 5$ (grün) und $\lambda = 10$ (rot); Quelle: <http://de.wikipedia.org/wiki/Poisson-Verteilung>

- Erzeugung durch Quasi-Komposition: Idee $\tau=1 \Rightarrow \lambda=y$

$$Y_j \sim \text{expo}(y) \Rightarrow X = \max \left\{ k : \sum_{j=1}^k Y_j \leq 1 \right\} \sim \text{Poisson}(y).$$

$$\begin{aligned} \sum_{0 \leq j < k} Y_j &\leq 1 < \sum_{0 \leq j \leq k} Y_j \\ \sum_{0 \leq j < k} \left(-\frac{1}{y} \ln(U_j) \right) &\leq 1 < \sum_{0 \leq j \leq k} \left(-\frac{1}{y} \ln(U_j) \right) \\ \ln \left(\prod_{0 \leq j < k} U_j \right) &\geq -y > \ln \left(\prod_{0 \leq j \leq k} U_j \right) \\ \prod_{0 \leq j < k} U_j &\geq e^{-y} > \prod_{0 \leq j \leq k} U_j \end{aligned}$$

Algorithmus:

- $a = e^{-y}, b = 1, k = -1.$
- Wiederhole
 - $k = k + 1, U_i \sim U(0,1), b = b \cdot U_k$
 - bis $b \leq a$
 - $X=k$
- Achtung: e^{-y} kann sehr klein werden. Beispiel: $y=1000 \Rightarrow e^{-y}=5 \cdot 10^{-435}$
 Dann ergeben sich numerische Fehler, die zu falschen Mittelwerten und Varianzen führen. Abhilfe durch clevere Implementierung, die große Zahlen vermeidet.
 - $\prod_{0 \leq j < k} U_j \geq e^{-y} > \prod_{0 \leq j \leq k} U_j$
 - $e^y \cdot \prod_{0 \leq j < k} U_j \geq 1 > e^y \cdot \prod_{0 \leq j \leq k} U_j$
 - Idee
 - $e^y = (e^{10})^{(y \text{ div } 10)} \cdot e^{(y \bmod 10)}$
 - Multipliziere abwechselnd Zufallszahlen U_j und $(y \text{ div } 10)$ Mal e^{10} bzw. $e^{(y \bmod 10)}$, so dass das Produkt nie kleiner als eine gewisse Schranke wird.
 - Implementierung siehe Übung!

3.4 Verwandtschaft kontinuierlicher und diskreter Verteilungen



Abbildung 9: Zeitkontinuierlicher Ankunftsprozess.

Wir nehmen an, die Zeit ist kontinuierlich und zu jedem Zeitpunkt ist mit derselben Wahrscheinlichkeitsdichte eine Ankunft möglich. Dies wird in Abbildung 9 illustriert.

- Die Zeit zwischen Ankünften ist exponentiell verteilt mit Rate λ ($\text{Expo}(\lambda)$), durchschnittliche Zwischenankunftszeit ist $1/\lambda$.
- Die Zeit, bis sich k Ankünfte ereignen, ist Erlang- k verteilt ($\text{Erlang}(k, \lambda)$).
- Die Anzahl der Ankünfte innerhalb eines Intervalls fester Länge τ ist Poisson-verteilt mit Parameter $y=\tau^*\lambda$ ($\text{Poisson}(y)$).



Abbildung 10: Zeitdiskreter Ankunftsprozess, bei dem rote Slots Ankünfte darstellen.

Wir nehmen an, die Zeit ist in Slots eingeteilt und in jedem Slot ist mit Wahrscheinlichkeit p eine Ankunft möglich wie Abbildung 10 illustriert.

- Die Anzahl der Slots zwischen Ankünften ist geometrisch $\text{Geom}(p)$ verteilt mit $p=1/(E[X]+1)$.
- Die Anzahl der freien Slots, bis s Ankünfte geschehen sind, ist negativ-binomial verteilt ($\text{NegBin}(s,p)$).
- Die Anzahl der Ankünfte innerhalb eines Intervalls fester Länge n ist binomial verteilt ($\text{Binom}(n,p)$).

Somit begründen wir die Verwandtschaft der entsprechenden Verteilungen.

3.5 Übersicht behandelter Verteilungen

3.5.1 Kontinuierliche Verteilungen

1. Gleichverteilung
2. Exponentielle Verteilung (λ) (cvar=1)
 - o „zufällige“ Zwischenankunftszeiten (Anrufe)
3. Erlang-k-Verteilung (k, λ) (cvar<1)
 - o Summe von k exponentiellen Phasen
4. Hyperexponentielle Verteilung (cvar>1)
5. Normalverteilung (μ, σ)
 - o Summe sehr vieler iid Zufallsvariablen

3.5.2 Diskrete Verteilungen

1. Diskrete Gleichverteilung
2. Beliebige diskrete Verteilung
3. Bernoulli-Verteilung Bernoulli(p)
 - o Münzwurf
4. Binomial-Verteilung Binom(n, p)
 - o Wiederholter Münzwurf
 - o Ziehen von Kugeln mit Zurücklegen
 - o Ankünfte mit geometrisch verteilten Zwischenankunftszeiten Geom(p) in einem Intervall ganzzahliger Länge n
5. Geometrische Verteilung Geom(p)
 - o Wiederholter Münzwurf bis 1 Mal gewonnen wurde
 - o ganzzahliges Gegenstück zur exponentiellen Verteilung
6. Negativ-binomiale Verteilung NegBin(s, p)
 - o Wiederholter Münzwurf: Anzahl von Fehlversuchen bis s Mal gewonnen wurde
7. Poisson-Verteilung Poisson(y)
 - o Anzahl von Ankünften mit exponentiell verteilten Zwischenankunftszeiten in einem Intervall fester Länge

Seed: Zahl mit der ZG-Generator initialisiert wird.

4 Aufbereitung von Stichproben

Eine Stichprobe kann sich aus wiederholten Simulationsergebnissen mit unterschiedlicher Seed oder Messwerten zusammensetzen. Konfidenzintervalle geben an, wie genau die Stichprobe den wahren Mittelwert schätzt.

Standardabweichungen und Quantile geben an, wie stark eine Stichprobe streut. Eine Reihe von Stichprobenmittelwerten, die von einem Parameter abhängen, kann in Abhängigkeit von diesem aufgetragen werden. Durch geschickte Achsenkalierung kann man evtl. die Art der Abhängigkeit ablesen. Bei Stichproben, die von mehreren Parametern abhängen, können Pivot-Tables bei der Ergebnisaufbereitung helfen.

4.1 Wiederholung empirischer Schätzer (LK 4.4)

Gegeben: n statistisch unabhängige und identisch verteilte (independent and identically distributed, IID) Zufallsvariablen X_1, X_2, \dots, X_n mit Mittelwert $E[X]$ und Varianz $\text{VAR}[X]$

- Schätzer für $E[X]$: empirischer Mittelwert: $\bar{X}(n) = \frac{\sum_{i=1}^n X_i}{n}$
- Schätzer für $\text{VAR}[X]$: empirische Varianz: $S^2(n) = \frac{\sum_{i=1}^n [X_i - \bar{X}]^2}{n-1}$

4.2 Güte der Mittelwertwertschätzung (LK 4.6)

- Starkes Gesetz der großen Zahlen (aus der Wahrscheinlichkeitstheorie):

$$\bar{X}(n) \xrightarrow{n \rightarrow \infty} E[X] \text{ mit Wahrscheinlichkeit 1.}$$

- Problem: $n \rightarrow \infty$ nicht simulierbar.
- Wie stark schwankt der Schätzer für endliches n ?

$$\text{VAR}[\bar{X}(n)] = \text{VAR}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \text{VAR}\left[\sum_{i=1}^n X_i\right] \stackrel{\text{IID}}{\downarrow} \frac{1}{n^2} \sum_{i=1}^n \text{VAR}[X] = \frac{1}{n} \text{VAR}[X] \quad (4.1)$$

$$\text{Schätzwert für } \text{VAR}[\bar{X}(n)] = \frac{S^2(n)}{n} = \frac{\sum_{i=1}^n [X_i - \bar{X}]^2}{n \cdot (n-1)} \quad (4.2)$$

4.3 Standard-Normalverteilung (LK 4.5)

- Die Verteilungsfunktion der **Standard-Normal-Verteilung** mit Mittelwert 0 und Standardabweichung 1 ist gegeben durch:

$$\Phi(z) = \frac{1}{\sqrt{2 \cdot \pi}} \int_{-\infty}^z e^{-y^2/2} dy \quad (4.3)$$

Sie ist analytisch nicht darstellbar und kann nur numerisch gelöst werden.
→ Werte aus Tafelwerken

- Transformation einer normal verteilten Zufallsvariable Y mit $E[Y]$ und $VAR[Y]$ in eine standard-normal verteilte Zufallsvariable Z: $Z = \frac{Y - E[Y]}{\sqrt{VAR[Y]}}$

(4.4)

- Zentraler Grenzwertsatz:**

Seien X_1, \dots, X_n IID beliebig verteilte Zufallsvariablen mit Mittelwert $E[X]$ und $VAR[X] > 0$. Wir definieren $\bar{Y} = \bar{X}(n)$ als ihren Mittelwert und führen eine neue

Zufallsvariable $Z_n = \frac{\bar{X}(n) - E[X]}{\sqrt{VAR[X]/n}}$, die einer Transformation von Y

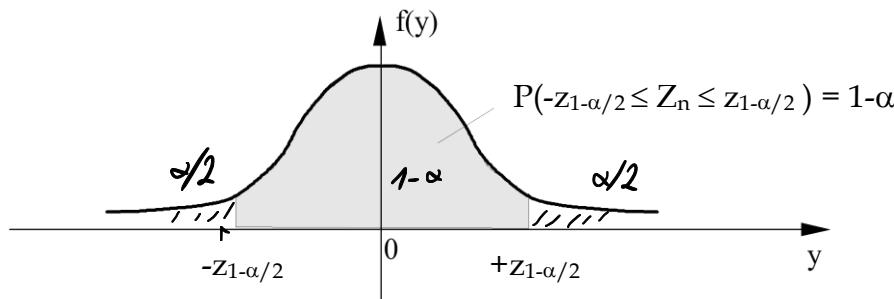
entspricht. Die Verteilungsfunktion der neuen Zufallsvariablen sei mit $F_n(z)$ bezeichnet. Dann gilt:

$$F_n(z) \xrightarrow{n \rightarrow \infty} \Phi(z). \quad (4.5)$$

Andere Formulierung: Für hinreichend großes n ist der Mittelwert von n IID Zufallsvariablen normal verteilt mit Mittelwert $E[X]$ und Varianz $VAR[X]/n$.

4.4 Konfidenzintervalle für Mittelwerte (LK 4.5)

- Es gilt $\text{VAR}[X] \approx S^2(n)$ und somit $Z_n \approx \frac{\bar{X}(n) - E[X]}{\sqrt{S^2(n)/n}}$
- Für hinreichend großes n gilt für den transformierten Mittelwert Z_n
 - $P(-z_{1-\alpha/2} \leq Z_n \leq z_{1-\alpha/2}) \approx 1 - \alpha$ (4.6)
 - wobei $z_{1-\alpha/2}$ das $1-\alpha/2$ -Quantil der Standard-Normalverteilung ist.



- Aussage: Mit einer *statistischen Sicherheit* (=Wahrscheinlichkeit) von $1-\alpha$ liegt eine Stichprobe Z_n im Intervall $[-z_{1-\alpha/2}; +z_{1-\alpha/2}]$.
- Problemstellung ist aber i. d. R. umgekehrt: Stichproben sind gegeben und Aussage über $E[X]$ wird gesucht.

- Durch Umformung

$$P\left(-z_{1-\alpha/2} \leq \frac{\bar{X}(n) - E[X]}{\sqrt{S^2(n)/n}} \leq z_{1-\alpha/2}\right) \approx 1 - \alpha \quad \text{nur Umformung} \quad (4.7)$$

$$P\left(\bar{X}(n) - z_{1-\alpha/2} \cdot \sqrt{S^2(n)/n} \leq E[X] \leq \bar{X}(n) + z_{1-\alpha/2} \cdot \sqrt{S^2(n)/n}\right) \approx 1 - \alpha$$

lower $l(n, \alpha) = \bar{X}(n) - z_{1-\alpha/2} \cdot \sqrt{S^2(n)/n}$ upper $u(n, \alpha) = \bar{X}(n) + z_{1-\alpha/2} \cdot \sqrt{S^2(n)/n}$ (4.8)

- [$l(n, \alpha); u(n, \alpha)$] wird **Konfidenzintervall** oder Vertrauensintervall genannt.
- $1-\alpha$ ist das **Konfidenzniveau** (confidence level) und α das **Signifikanzniveau** (level of significance)
- Aussage:** $1-\alpha$ aller empirisch gewonnenen Konfidenzintervalle enthalten den wahren Wert $E[X]$.

4.5 Konfidenzintervalle für Mittelwerte bei kleinem Stichprobenumfang (LK 4.5)

- Für einen kleinen Stichprobenumfang mit n Elementen nimmt man das $1-\alpha/2$ -Quantil $t_{n-1, 1-\alpha/2}$ der **Student-t Verteilungsfunktion** (Gosset, 1908) mit „ $n-1$ Freiheitsgraden“ zur Berechnung des Konfidenzintervalls
 - Verteilungsdichtefunktion

$$f_{n-1}(t) = \frac{1}{\sqrt{(n-1)\pi}} \cdot \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}} \quad (4.9)$$

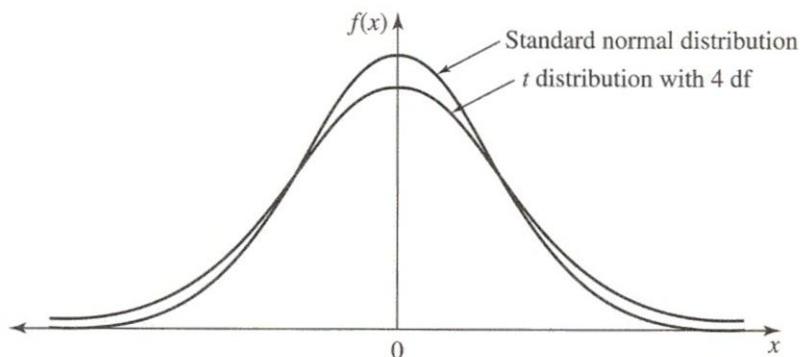


Abbildung 4.1: Die Standard-Normalverteilung hat weniger Varianz als die Student-t-Verteilung mit 4 Freiheitsgraden. (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 256)

- Die Student-t-Verteilungen konvergieren gegen die Standard-Normalverteilung für $n \rightarrow \infty$.
- Sie sind flacher und haben eine größere Varianz.
- Falls die X_1, \dots, X_n **normal verteilt** sind, ist das Konfidenzintervall exakt.
- Je schiefer die Verteilung der Stichprobenwerte ist, desto kleiner ist der Prozentsatz der Konfidenzintervalle, die den wahren Wert $E[X]$ tatsächlich beinhalten (LK Table 4.1, S. 257).
- Das Konfidenzintervall wird als *Student-t-Vertrauensintervall* bezeichnet.

Tabelle 1: 97,5%-Quantile der Student-t Verteilungen für Konfidenzintervalle mit einem Signifikanzniveau von $\alpha=5\%$.

Stichprobenumfang n	2	3	4	5	6	7	8	9	10	20	∞
$t_{n-1, 1-\alpha/2}$	12,706	4,303	3,182	2,776	2,571	2,447	2,365	2,306	2,262	2,093	1,960

	<i>One Sided</i>	75%	80%	85%	90%	95%	97.5%	99%	99.5%	99.75%	99.9%	99.95%
	<i>Two Sided</i>	50%	60%	70%	80%	90%	95%	98%	99%	99.5%	99.8%	99.9%
df=1	1.000	1.376	1.963	3.078	6.314	12.71	31.82	63.66	127.3	318.3	636.6	
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	14.09	22.33	31.60	
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	7.453	10.21	12.92	
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610	
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869	
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959	
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408	
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041	
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781	
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587	
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437	
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318	
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221	
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140	
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073	
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015	
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965	
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922	
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883	
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850	
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819	
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792	
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767	
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745	
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725	
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707	
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690	
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674	
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659	
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646	
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551	
50	0.679	0.849	1.047	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496	
60	0.679	0.848	1.045	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460	
80	0.678	0.846	1.043	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416	
100	0.677	0.845	1.042	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390	
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373	
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.291	

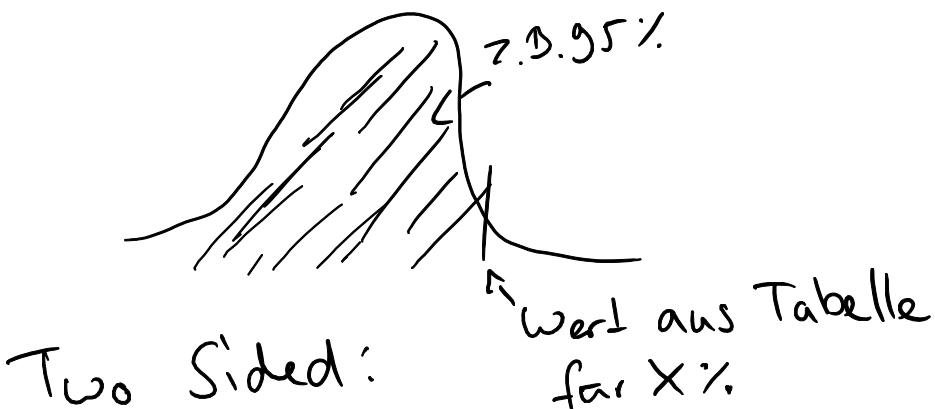
- Beispiel zur Illustration der Genauigkeit von Konfidenzintervallen

Estimated coverages based on 500 experiments

Distribution	Skewness ν	$n = 5$	$n = 10$	$n = 20$	$n = 40$
Normal	0.00	0.910	0.902	0.898	0.900
Exponential	2.00	0.854	0.878	0.870	0.890
Chi square	2.83	0.810	0.830	0.848	0.890
Lognormal	6.18	0.758	0.768	0.842	0.852
Hyperexponential	6.43	0.584	0.586	0.682	0.774

Abbildung 4.2: Konfidenzintervalle wurden 500 Mal für eine endliche Stichprobengröße n berechnet. Die Daten zeigen den Anteil von ihnen, der den wahren Mittelwert tatsächlich enthält. Konfidenzintervalle sind nur für Normal-verteilte Zufallsvariablen exakt. Mit zunehmender Stichprobengröße verlieren sich aber Ungenauigkeiten auf Grund der Schiefe anderer Verteilungen. (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 257)

One Sided:



4.6 Konfidenzintervalle für Wahrscheinlichkeiten (LK 9.4.1)

- Schätzwert \hat{p} für Wahrscheinlichkeit p ($0 \leq p \leq 1$) eines Merkmals A soll simulativ ermittelt werden: n Ereignisse insgesamt, davon k Ereignisse für Merkmal A
- Empirische Wahrscheinlichkeit: $\hat{p} = \frac{k}{n}$ (4.10)
- Man kann Konfidenzintervalle für diese gesuchte Wahrscheinlichkeit berechnen, indem man Konfidenzintervalle zum empirischen Mittelwert von n Bernoulli-Versuchen mit Hilfe der empirischen Varianz $S^2(n)$ nach Gleichung (4.8) berechnet.
- Ziel im Folgenden: Schätzen von Wahrscheinlichkeiten ohne die Berechnung der empirischen Varianz $S^2(n)$
 - Idee
 - Die Anzahl der Ereignisse für das Merkmal A ist binomial verteilt: $k = \text{Binom}(n, p)$
 - $Y = \hat{p} = \frac{k}{n} = \frac{\text{Binom}(n, p)}{n}$
 - $E[Y] = p$: Der Gesuchte Erwartungswert ist der Parameter p einer Binomialverteilung
 - Da Y auf eine Binomialverteilung zurückgeht, gilt:
 - $\text{VAR}[Y] = \frac{1}{n^2} \cdot \text{VAR}[\text{Binom}(n, p)] = \frac{p \cdot (1-p)}{n}$
 - Y ist annähernd normalverteilt für große n
 - \hat{p} ist der Schätzwert für p
 - $\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$ ist approximativ $N(0,1)$ -verteilt (standardnormalverteilt):
 - $P\left(-z_{1-\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq z_{1-\alpha/2}\right) \approx 1 - \alpha$
 - $P\left(\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \approx 1 - \alpha$
 - $\left[\hat{p} - z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{1-\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$ (4.11)

ist das approximative Konfidenzintervall mit der statistischen Sicherheit $1 - \alpha$, wenn $z_{1-\alpha/2}$ das $1 - \alpha/2$ -Quantil der Standardnormalverteilung $N(0,1)$ ist.

4.7 Interpretation von Fehlerbalken

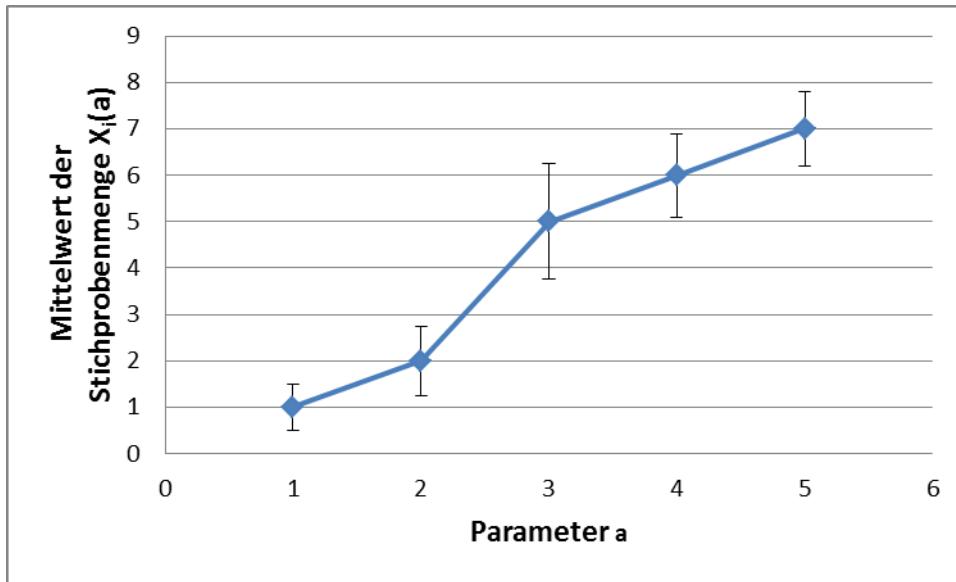


Abbildung 1: In einer Graphik, die Kurven für Mittelwerte $E[X(a)]$ (y-Achse) in Abhängigkeit eines Parameters a (x-Achse) zeigt, werden häufig „Fehlerbalken“ (Error Bars) angegeben.

In Graphiken, die beispielsweise Kurven für Mittelwerte $\bar{X}(a)$ in Abhängigkeit eines Parameters a zeigen, werden häufig Fehlerbalken (Error Bars) als Abweichung nach oben und unten angegeben. Diese können unterschiedliche Semantik haben.

- **Minimum und Maximum**
- **Konfidenzintervalle**
 - Aussage über die Genauigkeit des Mittelwertes
 - Keine Aussage über die Schwankung der betrachteten Größe $X(a)$
 - Mit zunehmender Stichprobengröße kann das Konfidenzintervall verkleinert werden.
- **Standardabweichung**
 - Aussage über die Schwankung der betrachteten Größe $X(a)$
 - Keine Aussage über die Genauigkeit des Mittelwertes $X(a)$
 - Standardabweichung hängt von der Verteilung von $X(a)$ ab und kann nicht verkleinert werden.

4.8 Boxplots

Quelle <https://de.wikipedia.org/wiki/Boxplot>

Boxplots zeigen die Streuung von Stichprobenwerten an.

- Unteres Quartil und oberes Quartil beschreiben eine Box; Höhe der Box ist wird Interquartilsabstand (interquartile range (IQR)) genannt
- Median wird in der Box als Querstrich eingezeichnet
- Antennen (Whiskers) verlängern die Box, sie können unterschiedliche Bedeutungen haben
 - Länge der einzelnen Whiskers ist maximal $1.5 \cdot \text{IQR}$, die Whiskers enden aber mit dem letzten in diesem Bereich beobachteten Wert; Werte zwischen $1.5 \cdot \text{IQR}$ und $3 \cdot \text{IQR}$ werden milde Ausreißer genannt und mit ◦ markiert; Werte außerhalb werden extreme Ausreißer genannt und mit * markiert.
 - Die beiden Whiskers kennzeichnen das 2.5% und 97.5% Quantil
 - Die beiden Whiskers kennzeichnen Minimum und Maximum
- Das arithmetische Mittel (Mittelwert) kann als Stern in den Boxplot aufgenommen werden, ist aber unüblich.

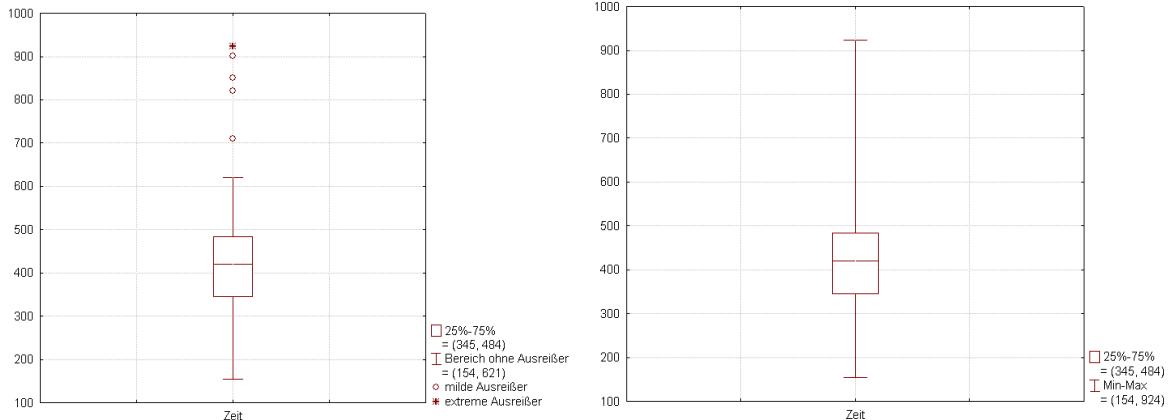


Abbildung 2: Zwei Arten von Boxplots

Quelle: „Box-Plot mit Interquartilsabstand“ von Schlurcher - Eigenes Werk. Lizenziert unter CC BY 3.0 über Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Box-Plot_mit_Interquartilsabstand.png#/media/File:Box-Plot_mit_Interquartilsabstand.png; „Box-Plot mit Min-Max Abstand“ von Schlurcher - Eigenes Werk. Lizenziert unter CC BY 3.0 über Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Box-Plot_mit_Min-Max_Abstand.png#/media/File:Box-Plot_mit_Min-Max_Abstand.png

4.9 Quantifizieren und Darstellung des Einflusses von Faktoren

Faktoren haben i. A. nichtlinearen Einfluss auf die Systemantwort.

Problem: Wie können logarithmische, polynomiale oder exponentielle Zusammenhänge erkannt und dargestellt werden?

- Idee: Koordinatentransformation durch Skalierung der x- bzw. y-Achse
- Platzierung der Koordinaten
 - Lineare Skalierung: $\hat{x}(x) = x$ bzw. $\hat{y}(y) = y$
 - Logarithmische Skalierung: $\hat{x}(x) = \ln(x)$ bzw. $\hat{y}(y) = \ln(y)$

4.9.1 Skalierung zur Darstellung exponentiellen Wachstums

- Faktor x beeinflusst Systemantwort exponentiell: $y(x) = \exp(\lambda \cdot x)$
- Linear skalierte x-Achse: $\hat{x}(x) = x$
- Logarithmisch skalierte y-Achse: $\hat{y}(x) = \ln(y(x)) = \ln(\exp(\lambda \cdot x)) = \lambda \cdot x = \lambda \cdot \hat{x}(x)$
- Punkte im Koordinatensystem: $(\hat{x}(x), \lambda \cdot \hat{x}(x))$ ergeben eine Gerade mit Steigung λ

4.9.2 Skalierung zur Darstellung logarithmischen Wachstums

- Faktor x beeinflusst Systemantwort logarithmisch: $y(x) = \ln(\lambda \cdot x^\alpha)$ $\alpha = \text{Steigung der Geraden}$
- Logarithmisch skalierte x-Achse: $\hat{x}(x) = \ln(x)$
- Linear skalierte y-Achse:

$$\hat{y}(x) = y(x) = \ln(\lambda \cdot x^\alpha) = \ln(\lambda) + \alpha \cdot \ln(x) = \ln(\lambda) + \alpha \cdot \hat{x}(x)$$
- Punkte im Koordinatensystem: $(\hat{x}(x), \ln(\lambda) + \alpha \cdot \hat{x}(x))$ ergeben eine Gerade, die die y-Achse bei $\ln(\lambda)$ schneidet
- Problem: Messpunkte sollen auch bei logarithmischer Skalierung der x-Achse in der Graphik gleichen Abstand voneinander haben, das erfordert einen exponentiellen Abstand der Eingangsparameter im Wertebereich.

- Beispiel
 - Gewünscht: 10 Messwerte im Parameterbereich 1 bis 1000
 - Abstand der Messwerte im logarithmischen Bereich: $\frac{3-0}{10-1} = \frac{3}{9} = \frac{1}{3}$
 - Wähle Messwerte $x_k = 10^{\left(\frac{k}{3}\right)}$ mit $k \in [0;9]$

4.9.3 Skalierung zur Darstellung polynomiellen Wachstums (Potenzen)

- Faktor x beeinflusst Systemantwort polynomiell: $y(x) = x^\lambda$
- Logarithmisch skalierte x -Achse: $\hat{x}(x) = \ln(x)$
- Logarithmisch skalierte y -Achse: $\hat{y}(x) = \ln(y(x)) = \ln(x^\lambda) = \lambda \cdot \ln(x) = \lambda \cdot \hat{x}(x)$
- Punkte im Koordinatensystem: $(\hat{x}(x), \lambda \cdot \hat{x}(x))$ ergeben eine Gerade mit Steigung λ

4.9.4 Skalierung zur Darstellung linearen Wachstums

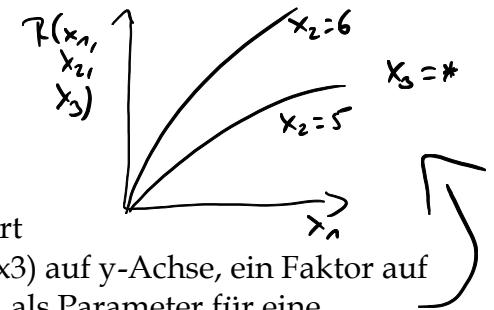
- Faktor x beeinflusst Systemantwort linear: $y(x) = \lambda \cdot x$
- Linear skalierte x - und y -Achse
 - $\hat{x}(x) = x$, $\hat{y}(x) = y(x) = \lambda \cdot x = \lambda \cdot \hat{x}(x)$
 - Punkte im Koordinatensystem: $(\hat{x}(x), \lambda \cdot \hat{x}(x))$ ergeben eine Gerade mit Steigung λ .
 - Evtl. Abweichungen im Bereich kleiner Parameter sind kaum sichtbar, da bei äquidistantem Abtasten der x -Achse nur große Parameter betrachtet werden.
- Logarithmisch skalierte x - und y -Achse
 - $\hat{x}(x) = \ln(x)$, $\hat{y}(x) = \ln(y(x)) = \ln(\lambda \cdot x) = \ln(\lambda) + \ln(x) = \ln(\lambda) + \hat{x}(x)$
 - Punkte im Koordinatensystem: $(\hat{x}(x), \ln(\lambda) + \hat{x}(x))$ ergeben Gerade mit Steigung 1, e^{Offset} zur Winkelhalbierenden auf der y -Achse gibt Faktor λ an.
 - Evtl. Abweichungen im Bereich kleiner Parameter sind gut sichtbar, da alle Größenordnungen bei äquidistantem Abtasten der x -Achse untersucht werden.

4.10 Organisation von Ergebnissen in PivotTables zur flexiblen Darstellung

Problembeschreibung

- Man hat es oft mit sehr vielen Systemfaktoren zu tun, z.B. mit $x_1, x_2, x_3, x_4, \dots$ und hat für deren Kombinationen Systemantworten (Ergebnisse) ermittelt.
- Die Daten liegen oft in Form einer Ergebnistabelle vor:

x_1	x_2	x_3	$R(x_1, x_2, x_3)$
1	4	6	24
1	4	7	28
1	5	6	30
1	5	7	35
2	4	6	48
2	4	7	56
2	5	6	60
2	5	7	70
3	4	6	72
3	4	7	84
3	5	6	90
3	5	7	105



- Darstellungsmöglichkeiten für die Systemantwort
 - Zweidimensionales Diagramm: $R(x_1, x_2, x_3)$ auf y-Achse, ein Faktor auf x-Achse, z.B. x_1 , ein anderer, z.B. x_2 , wird als Parameter für eine Kurvenschar benutzt.
 - Dreidimensionales Diagramm: $R(x_1, x_2, x_3, \dots)$ auf y-Achse, zwei Faktoren, z.B. x_1 und x_2 , werden auf den zwei x-Achsen dargestellt.
 - Für die nicht berücksichtigten Faktoren kann dabei
 - entweder ein fester Wert angenommen werden, z.B. $x_3=6$,
 - oder die Systemantwort wird über alle Werte von x_3 , also $x_3=6$ und $x_3=7$, gemittelt.
 - Problem: oft ist der Haupteinflussfaktor für die Systemantwort a priori noch nicht bekannt und muss noch gefunden werden.
 - Probiere x_1, x_2 , und x_3 jeweils als x-Achse für die Darstellung aus!
 - Randbedingung vieler Darstellungsprogramme wie z.B. Excel
 - x- und y-Werte müssen zum Plotten als Vektoren gegeben sein, indem sie in einer Spalte oder Zeile aufeinander folgen.
 - Problem: das ist in der obigen Ergebniszusammenstellung nur für x_3 mit festem x_1 und x_2 gegeben. Beispiel: $x_3=[6,7]$ mit $R(1,4,[6,7])=[24,28]$.
 - Für andere Darstellungen muss die Ergebnistabelle umorganisiert werden, das wird durch Pivot-Tables erleichtert.

Benutzung einer Pivot-Tabelle in Excel (Beschreibung für Office 2010)

- Organisation der Daten wie in der Ergebnistabelle
 - Oberste Zeile: Parameterbezeichnungen
 - Restliche Zeilen: Ergebniswerte
- Durchführung des PivotTable-Assistenten (Einfügen → PivotTable)
 - Einlesen der gesamten Tabelle inklusive Überschrift
 - Konfiguration des PivotTables
 - Es erscheint eine PivotTable-Feldliste
 - Ziehe den Faktor für die x-Achse (im Beispiel x1) in das Zeilenbeschriftungen-Feld
 - Ziehe den Faktor für die Kurvenschaar (im Beispiel x2) in das Spaltenbeschriftungen-Feld
 - Ziehe die restlichen Faktoren (im Beispiel x3) in das Berichtsfilter-Feld
 - Ziehe die Systemantwort (im Beispiel R(x1, x2, x3)) in das Wertefeld
 - Linksklick auf Wertfeldeinstellungen und stelle von Summenwert auf den Mittelwert von R(x1, x2, x3) um!

x3	(Alle)
----	--------

Mittelwert von R(x1,x2,x3)	x2			Gesamtergebnis
x1	4	5		
1	26	32,5		29,25
2	52	65		58,5
3	78	97,5		87,75
Gesamtergebnis	52	65		58,5

- In den Spaltenbeschriftungen stehen keine Einzelwerte sondern die Durchschnitte über alle sonstigen Faktoren in den Seitenfeldern, man kann aber auch auf einen bestimmten Wert umstellen. (Das geht über die Spaltenbeschriftungsauswahl)
- Parameter im Berichtsfilter, Zeilenbeschriftungen, Spaltenbeschriftungen können beliebig verschoben werden und somit ist es einfach Vektoren im Wertefeld gegen unterschiedliche Vektoren im Zeilenfeld zu plotten.
- Zum Plotten selber: kopiere die Vektoren aus dem PivotTable und erstelle mit Hilfe des Diagramm-Assistenten eine XY-Graphik! Wähle aus, ob Spalten oder Zeilen gegeneinander geplottet werden sollen.

5 Statistische Grundlagen (Teil 2)

Die Kovarianz zweier Zufallsvariablen gibt an, ob sich diese tendenziell ähnlich verhalten bezüglich ihrer Abweichungen von ihrem jeweiligen Mittelwert. Die Korrelation ist die normierte Kovarianz. Kovarianzen können für Zufallsvariablen und für geordnete Stichproben (Zeitreihen) berechnet werden.

Die Autokovarianz einer Zeitreihe zum Abstand j gibt an, ob sich mit Abstand j aufeinanderfolgende Werte einer Zeitreihe tendenziell ähnlich verhalten. Die normierte Autokovarianz nennt sich Autokorrelation. Sie kann durch Messobjekte gemessen werden.

5.1 Kovarianz zweier Zufallsvariablen (LK 4.2)

- Definition: **Kovarianzfunktion**

$$\text{Cov}[X, Y] = E[(X - E[X]) \cdot (Y - E[Y])] = E[X \cdot Y] - E[X] \cdot E[Y]$$

- X, Y unabhängig $\Rightarrow \text{Cov}[X, Y] = 0$

- Beweis (LK Problem 4.8)

$$\text{Cov}[X, Y] = E[X \cdot Y] - E[X] \cdot E[Y] =$$

$$\int_{-\infty}^{\infty} x \cdot y \cdot f(x, y) dx dy - E[X] \cdot E[Y] \quad \text{Unabhängigkeit} =$$

$$\int_{-\infty}^{\infty} x \cdot y \cdot f(x) \cdot f(y) dx dy - E[X] \cdot E[Y] =$$

$$E[X] \cdot E[Y] - E[X] \cdot E[Y] = 0$$

- ABER nicht: $\text{Cov}[X, Y] = 0 \Rightarrow X, Y$ unabhängig

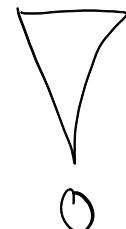
- Gegenbeispiel (LK Problem 4.9)

Sei X diskrete ZV mit gleichwahrscheinlichen Ereignissen -2, -1, 1 und 2. Sei $Y = X^2$. Offensichtlich sind X und Y nicht unabhängig, aber sie sind unkorreliert:

$$E[X] = 0$$

$$\text{Cov}[X, Y] = E[X \cdot Y] - E[X] \cdot E[Y] =$$

$$\frac{1}{4}(-2 \cdot 4 - 1 \cdot 1 + 1 \cdot 1 + 2 \cdot 4) - 0 \cdot E[Y] = 0$$



(2.25)

5.2 Korrelation zweier Zufallsvariablen (LK 4.2)

- Definition: **Korrelationsfunktion**

$$\text{Cor}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{VAR}[X] \cdot \text{VAR}[Y]}} \quad (2.26)$$

- Spezialfälle

- Maximale positive Korrelation

$$Y=X \Rightarrow \text{Cor}[X, Y]=1$$

- Maximale negative Korrelation

$$Y=-X \Rightarrow \text{Cor}[X, Y]=-1$$

- Die Korrelationsfunktion beschreibt auf normierte Weise, wie stark zwei Zufallsvariablen gleichzeitig von ihrem Mittelwert abweichen.
 - $\text{Cor}[X, Y]>0$: gleichzeitige Differenzen der beiden ZV X und Y von ihrem jeweiligen Mittelwert tendieren gleiches Vorzeichen zu haben.
 - $\text{Cor}[X, Y]<0$: gleichzeitige Differenzen der beiden ZV X und Y von ihrem jeweiligen Mittelwert tendieren unterschiedliches Vorzeichen zu haben.

5.3 Empirische Kovarianz und Korrelation zweier Zeitreihen

Das Konzept der Kovarianz (Korrelation) $\text{Cov}[X, Y]$ wurde für zwei Zufallsvariablen definiert und die Formeln entsprechen wie bei den Momenten auch einer Auswertung über den Wertebereich ihres Verbundereignisses (X, Y). Bei der empirischen Kovarianz wird anstelle des Wertebereiches die Zeit als Grundlage für die Auswertung genommen.

- a) Zeitdiskreter Fall

$$\overline{\text{COV}}[X, Y] = \frac{1}{n-1} \cdot \sum_{0 \leq i < n} (X_i - \bar{X}(n)) \cdot (Y_i - \bar{Y}(n)) \quad (2.32)$$

Mittelwert

- b) Zeitkontinuierlicher Fall

$$\overline{\text{COV}}[X, Y] = \frac{1}{T} \cdot \int_0^T (X(t) - \bar{X}) \cdot (Y(t) - \bar{Y}) dt \quad (2.33)$$

Beweis

Idee des Beweises: wir berechnen den Erwartungswert der empirischen Kovarianz

$E[\overline{\text{COV}}[X, Y]] = \frac{1}{n-1} \cdot \sum_{0 \leq i < n} E[(X_i - \bar{X}(n)) \cdot (Y_i - \bar{Y}(n))]$, die sich als Funktion der Zufallsvariablen X_i und Y_j berechnet, und zeigen, dass dieser gleich der Kovarianz von X und Y ist, wenn alle X_i und Y_j nach X und Y verteilt sind.

Es gilt:

$$\begin{aligned} E[X_i \cdot \bar{Y}(n)] &= E\left[X_i \cdot \frac{1}{n} \cdot \sum_{0 \leq j < n} Y_j\right] = E\left[\frac{1}{n} \cdot X_i \cdot Y_i + \frac{1}{n} \cdot X_i \cdot \sum_{0 \leq j < n, j \neq i} Y_j\right] = \\ &= \frac{1}{n} \cdot E[X_i \cdot Y_i] + \frac{1}{n} \cdot E\left[E[X_i] \cdot \sum_{0 \leq j < n, j \neq i} E[Y_j]\right] = \frac{1}{n} \cdot E[X \cdot Y] + \frac{n-1}{n} \cdot E[X] \cdot E[Y] \end{aligned}$$

Hinweis: Bei dieser Rechnung wurden folgende Rechenregeln verwendet:

1. X_i und Y_j werden nicht gleichzeitig realisiert und sind deswegen unabhängige Zufallsvariablen. Darum darf beim nachfolgenden Ausdruck der Erwartungswert auch auf die Zufallsvariablen der Stichproben ein zweites Mal angewendet werden:

$$E\left[X_i \cdot \sum_{j \neq i} Y_j\right] = E\left[E[X_i] \cdot \sum_{j \neq i} E[Y_j]\right]$$

2. Die Stichprobenwerte X_i (auch Y_i) sind identisch verteilt: $E[X_i] = E[X]$
3. Die Erwartungswerte von Erwartungswerten sind wieder die Erwartungswerte: $E[E[X]] = E[X]$

Weiter gilt:

$$\begin{aligned} E[\bar{X}(n) \cdot \bar{Y}(n)] &= E\left[\left(\frac{1}{n} \cdot \sum_{0 \leq i < n} X_i\right) \cdot \left(\frac{1}{n} \cdot \sum_{0 \leq i < n} Y_i\right)\right] = \\ &= \frac{1}{n^2} \cdot E\left[\sum_{0 \leq i < n} X_i \cdot Y_i + \sum_{0 \leq i, j < n, i \neq j} X_i \cdot Y_j\right] = \frac{1}{n} \cdot E[X \cdot Y] + \frac{n-1}{n} \cdot E[X] \cdot E[Y] \end{aligned}$$

Damit ist der Erwartungswert der empirischen Varianz erwartungstreu:

$$\begin{aligned} E[\overline{\text{COV}}[X, Y]] &= \frac{1}{n-1} \cdot \sum_{0 \leq i < n} E[(X_i - \bar{X}(n)) \cdot (Y_i - \bar{Y}(n))] = \\ &= \frac{1}{n-1} \cdot \sum_{0 \leq i < n} (E[X_i \cdot Y_i] - E[X_i \cdot \bar{Y}(n)] - E[Y_i \cdot \bar{X}(n)] + E[\bar{X}(n) \cdot \bar{Y}(n)]) = \\ &= \frac{1}{n-1} \cdot \sum_{0 \leq i < n} \left(E[X \cdot Y] - \frac{1}{n} \cdot E[X \cdot Y] - \frac{n-1}{n} \cdot E[X] \cdot E[Y] - \frac{1}{n} \cdot E[X \cdot Y] - \right. \\ &\quad \left. - \frac{n-1}{n} \cdot E[X] \cdot E[Y] + \frac{1}{n} \cdot E[X \cdot Y] + \frac{n-1}{n} \cdot E[X] \cdot E[Y] \right) = \\ &= \frac{1}{n-1} \cdot \sum_{0 \leq i < n} \left(\frac{n-1}{n} \cdot E[X \cdot Y] - \frac{n-1}{n} \cdot E[X] \cdot E[Y] \right) = \\ &= \frac{1}{n} \cdot \sum_{0 \leq i < n} (E[X \cdot Y] - E[X] \cdot E[Y]) = E[X \cdot Y] - E[X] \cdot E[Y] = \text{COV}[X, Y] \end{aligned}$$

Insbesondere ist damit auch die Gültigkeit von (2.30) gezeigt.

Die empirische Korrelation zweier Zeitreihen erhält man durch Normierung ihrer Kovarianz mit der empirischen Varianz:

$$\overline{\text{COR}}[X, Y] = \frac{\overline{\text{COV}}[X, Y]}{\sqrt{S^2[X] \cdot S^2[Y]}}$$

5.4 Autokovarianz und Autokorrelation einer Zeitreihe (LK Gl. 4.9)

Wir betrachten zwei Zeitreihen:

$$X_i : 1, 2, 2, 3, 4, 4, 4, 5, 5, 4, 3, 3, 3, 2, 2, 3, 3, 2, 2, 1, 1, 1, 2, 2, 3$$

sowie

$$Y_i : 1, 4, 2, 5, 3, 4, 2, 1, 5, 2, 4, 1, 4, 2, 5, 4, 1, 3, 5, 4, 1, 3, 4, 1, 5$$

Bei Zeitreihe X_i sind sich aufeinanderfolgende Zahlen ähnlich, während sich bei Y_i aufeinanderfolgende Zahlen eher unähnlich sind. Die Autokorrelation ist ein Maß für die Ähnlichkeit von innerhalb einer Zeitreihe aufeinanderfolgender Zufallsvariablen. Der Abstand der betrachteten Zufallsvariablen kann vorgegeben werden und wird im Englischen als „lag“ bezeichnet.

Die Autokorrelation ist die normierte Autokovarianz.

Die Autokovarianz berechnet sich als die Kovarianz der betrachteten Zeitreihe und der um das lag verschobenen betrachteten Zeitreihe.

Definition:

Die Autokovarianz C_j eines stochastischen Prozesses X_i mit Abstand j ist definiert durch $\text{COV}[X, Y]$ mit $Y_i = X_{i+j}$.

Berechnung der empirischen Autokovarianz:

$$\begin{aligned} \hat{C}_j(n) &= \frac{1}{(n-j)} \sum_{0 \leq i < n-j} [X_i - \bar{X}] [X_{i+j} - \bar{X}] = \\ &= \frac{1}{n-j} \cdot \left(\sum_{j \leq i < n} (X_i \cdot X_{i-j}) - \bar{X}(n) \cdot \left(2 \cdot \sum_{0 \leq i < n} X_i - \sum_{0 \leq i < j} X_i - \sum_{n-j \leq i < n} X_i \right) \right) + (\bar{X}(n))^2 \end{aligned} \quad (2.34)$$

Es gibt auch eine ähnliche Definition mit einem anderen Vorfaktor, die bessere analytische Eigenschaften hat.

$$\hat{C}_j(n) = \frac{1}{n} \sum_{0 \leq i < n-j} [X_i - \bar{X}] [X_{i+j} - \bar{X}] \quad (2.35)$$

Eine analoge Definition gibt es auch für den zeitkontinuierlichen Fall:

$$\hat{C}_u(T) = \frac{1}{T} \cdot \int_0^{T-u} (X(t+u) - \bar{X}) \cdot (X(t) - \bar{X}) dt \quad (2.36)$$

Keiner der Schätzer 2.34, 2.35 und 2.36 ist erwartungstreu (LK p.253). Die Schätzergebnisse werden schlechter, je größer der betrachtete Abstand ist. Außerdem haben sie eine hohe Varianz v.a. bei kurzen Messreihen und sind untereinander korreliert, d.h. $\text{COV}(\hat{C}_j, \hat{C}_k) \neq 0$. Für unsere Zwecke nutzen wir den Schätzer 2.34.

5.5 Empirische Autokorrelation einer Zeitreihe (LK Gl. 4.9)

Die Autokovarianz wird durch Normierung zur Autokorrelation:

$$\rho_j(n) = \frac{\hat{C}_j(n)}{S^2(n)} \quad (2.37)$$

Auch der Schätzer für die Autokorrelation ist nicht gut, insbesondere für große Abstände j . Siehe Beispiel in Kapitel 5.6!

5.6 Erfassungsobjekte für Autokovarianzen und Autokorrelationen von zeitdiskreten Messwerten

- Erweiterung des „discrete counters“
- Spezifikation des max. Abstandes j für die Autokovarianz bzw. Autokorrelation
- Interne Datenhaltung
 - j Variablen um die ersten Werte X_0, \dots, X_{j-1} zu speichern
 - j Variablen um die letzten Werte zu X_{n-j}, \dots, X_{n-1} zu speichern
 - 1 Zähler für die Summe $X_{\Sigma} = \sum_{0 \leq i < n} X_i$
 - $j+1$ Zähler für die Summen $X_{\Sigma}^2(k) = \sum_{k \leq i < n} X_{i-k} \cdot X_i$ für $0 \leq k \leq j$
- Berechnung der Autokorrelations- und Autokovarianzfunktion gemäß Gleichung (2.34) und (2.37).
- Achtung: die Werte sind nur gut für $j < n$! Beispiele:
 - $X=0, 1, 1, 2 \Rightarrow \rho_0(4)=0.75, \rho_1(4)=0.00, \rho_2(4)=0.00, \rho_3(4)=-1.50$
 - $X=0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \Rightarrow \rho_0(10)=0.90, \rho_1(10)=0.70, \rho_2(10)=0.46, \rho_3(10)=0.19, \rho_4(10)=-0.12, \rho_5(10)=-0.46, \rho_6(10)=-0.85, \rho_7(10)=-1.26, \rho_8(10)=-1.72, \rho_9(10)=-2.21$

Betragswerte größer 1 sind offensichtlich keine gültigen Korrelationswerte.

5.7 Wichtige Begriffe auf Deutsch und Englisch

Deutscher Begriff	Englischer Begriff
Verteilung	distribution
Verteilungsfunktion	distribution function cumulative distribution function (cdf)
Verteilungsdichtefunktion	probability density function (pdf)
p-Quantil	p-percentile, p-quantile
Faltung	convolution
Erwartungswert	mean
Moment	moment
Zentrales Moment	central moment
Varianz	variance
Standardabweichung	standard deviation
Variationskoeffizient	coefficient of variation
Schiefe	skewness
Kovarianz	covariance
Korrelation	correlation
Empirisches Moment	sample moment
Stichprobenmoment	
Empirischer Mittelwert	sample mean
Stichprobenmittelwert	
Empirische Varianz	sample variance
Stichprobenvarianz	
Empirische Schiefe	sample skewness
Stichprobenschiefen	
Empirische Kovarianz	sample covariance
Stichprobenkovarianz	
Empirische Autokovarianz	sample autocovariance
Stichprobenautokovarianz	
Empirische Autokorrelation	sample autocorrelation
Stichprobenautokorrelation	

6 Statistische Auswertung von Simulationen

6.1 Dauer von Simulationen: Benötigte Stichprobengröße (LK 9.4.1)

Durch Simulationen gewonnene Mittelwerte sind nur Schätzungen und insofern mit Fehlern behaftet. Der wahre Wert befindet sich mit Wahrscheinlichkeit $1-\alpha$ im Konfidenzintervall um den empirischen Mittelwert. Je kleiner das Konfidenzintervall ist, desto genauer ist also die Schätzung. Für kleine Konfidenzintervalle braucht man allerdings viele Stichproben, d.h. man muss lange simulieren. Das führt zur Frage, wie lange man simulieren sollte um hinreichend genaue Schätzungen für den Mittelwert zu erhalten. Herangehensweise: Kontrolle der Simulationsdauer durch den absoluten bzw. relativen Fehler des Konfidenzintervalls.

6.1.1 Absoluter Fehler

- $\bar{X}(n)$ ist der geschätzte und μ ist der tatsächliche Mittelwert
- Absoluter Fehler der Schätzung $|\bar{X}(n) - \mu|$ (5.1)
- Mit Wahrscheinlichkeit von mindestens $1-\alpha$ gilt:
 - Der wahre Mittelwert μ liegt im Konfidenzintervall $[\bar{X}(n) - \delta(n, \alpha); \bar{X}(n) + \delta(n, \alpha)]$
 - Der absolute Fehler ist kleiner als die Halblänge des Konfidenzintervalls $\delta(n, \alpha)$
 - Algorithmus für die Simulation eines maximalen absoluten Fehlers δ^*
 - Inkrementiere die Stichprobenmenge der Größe n
 - Berechne neu: $\bar{X}(n), S^2(n), \delta(n, \alpha)$
 - Solange bis $\delta(n, \alpha) \leq \delta^*$
 - Abschätzung der Anzahl benötigter Stichproben während der Laufzeit
 - Nutzen: Manche Simulationen brauchen enorm viel Rechenzeit um eine Stichprobe zu berechnen. Für sehr lange Simulationen ist eine ungefähre Abschätzung der Gesamtdauer der Simulation von Interesse.
 - Gegeben: Stichprobenmenge mit n Werten, $S^2(n), \delta(n, \alpha)$

$\Rightarrow \bar{X}(n)$ hat einen *absoluten Fehler* von höchstens

$$\delta(n, \alpha) = t_{n-1, 1-\alpha/2} \cdot \sqrt{\frac{S^2(n)}{n}}$$

- Frage: Wie viele Stichproben n_{\min} werden etwa benötigt, damit ein absoluter Fehler von höchstens δ^* (mit Wahrscheinlichkeit $1-\alpha$) nicht überschritten wird?
- Antwort
 - Annahme: $S^2(m) \approx S^2(n)$
 - $\Rightarrow n_{\min} = \min_{m \geq n} \left(t_{m-1, 1-\alpha/2} \cdot \sqrt{\frac{S^2(n)}{m}} \leq \delta^* \right)$ (5.2)

6.1.2 Relativer Fehler

- Relativer Fehler: $\gamma = \frac{|\bar{X}(n) - \mu|}{|\mu|} \cdot 100\%$ (oder auch Prozentfehler) (5.3)
- Mit Wahrscheinlichkeit von mindestens $1-\alpha$ gilt:
 - $|\bar{X}(n) - \mu| \leq \delta(n, \alpha)$ und $|\mu| \geq |\bar{X}(n) - \delta(n, \alpha)|$
 - Falls das Konfidenzintervall nicht die 0 enthält, ist der maximale relative Fehler $\gamma_{\max} = \frac{\delta(n, \alpha)}{|\bar{X}(n) - \delta(n, \alpha)|} \geq \frac{|\bar{X}(n) - \mu|}{|\mu|} = \gamma$
- Achtung: Der maximale relative Fehler ist nicht $\frac{\delta(n, \alpha)}{|\bar{X}(n)|}$. Vielmehr gilt $\frac{\delta(n, \alpha)}{|\bar{X}(n)|} \leq \frac{\gamma_{\max}}{1 + \gamma_{\max}}$, d.h. der intuitive Ausdruck unterschätzt den maximalen relativen Fehler.
- Algorithmus für die Simulation eines maximalen relativen Fehlers γ_{\max}
 - Inkrementiere die Stichprobenmenge der Größe n
 - Berechne neu: $S^2(n)$, $\delta(n, \alpha)$
 - Solange bis $\frac{\delta(n, \alpha)}{|\bar{X}(n) - \delta(n, \alpha)|} \leq \gamma_{\max}$
- Abschätzung der Anzahl benötigter Stichproben während der Laufzeit
 - Gegeben: Stichprobenmenge mit n Werten, Stichprobenvarianz $S^2(n)$
 $\Rightarrow \bar{X}(n)$ hat mit Wahrscheinlichkeit $1-\alpha$ einen *absoluten Fehler* von höchstens $\delta(n, \alpha) = t_{n-1, 1-\alpha/2} \cdot \sqrt{\frac{S^2(n)}{n}}$
 - Frage: Wie viele Stichproben n_{\min} werden etwa benötigt, damit ein relativer Fehler von höchstens γ_{\max} (mit Wahrscheinlichkeit $1-\alpha$) nicht überschritten wird?
 - Antwort
 - Annahme: $S^2(m) \approx S^2(n)$
 - $\Rightarrow n_{\min} = \min_{m \geq n} \left(\frac{t_{m-1, 1-\alpha/2} \cdot \sqrt{\frac{S^2(n)}{m}}}{|\bar{X}(n)|} \leq \frac{\gamma_{\max}}{1 + \gamma_{\max}} \right)$ (5.5)

6.2 Arten von Simulationen

Die statistische Datenerhebung ist von der Art der Simulation abhängig. Es gibt

- Terminierende Systeme / Simulationen
- Nicht-terminierende Systeme / Simulationen

6.2.1 Terminierende Simulationen (LK 9.4)

- Vorgegebene natürliche Simulationsdauer
- Beispiele
 - Begrenzte Simulationszeit auf Grund von Modellannahmen (z.B. 8-Stunden-Tag)
 - Simulation der Zeit bis zum Eintreffen eines bestimmten Ereignisses („first passage time“, z.B. beim „random walk“)

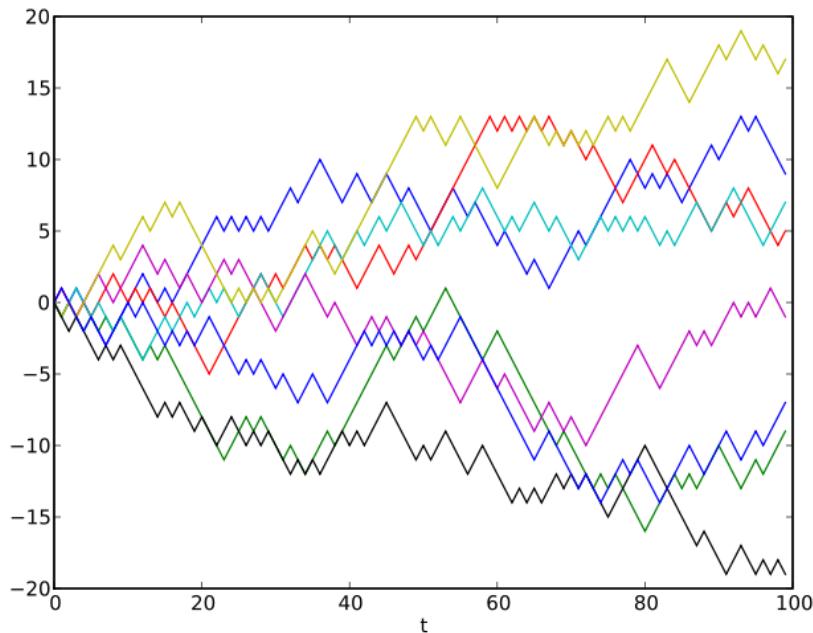


Abbildung 1: Beim „random walk“ startet ein Prozess in Zustand 0 und verändert in jedem Zeitschritt seinen Zustand um $+/-1$ mit Wahrscheinlichkeit 50%. Varianten existieren. Die Zeit bis zum ersten Erreichen einer Schwelle wird „first passage time“ genannt. (Bildquelle: http://de.wikipedia.org/wiki/Random_Walk)

- Art der Aussagen, die in einer Simulation gewonnen werden
 - Maße bezogen auf die Dauer eines gesamten Simulationslaufes (z.B. Anzahl bedienter Kunden pro Tag)
 - First passage time
 - Systemgrößen, die während der Simulation beobachtet werden (z.B. Warteschlangenlängen, Server-Auslastungen ...)

- Durchführung
 - Sinnvolle Wahl für die Initialisierung des Systemzustands der Simulation
 - Mehrmalige Durchführung von Simulationsläufen zur Generierung von Stichproben X_1, \dots, X_n
 - Unabhängigkeit der Stichproben durch unterschiedliche Zufallszahlen für die Simulationsläufe
 - Konstruktion von Konfidenzintervallen
 - Feste Anzahl n von Simulationsläufen
 - Simulation der Stichproben X_1, \dots, X_n
 - Wahl eines Signifikanzniveaus (z.B. $\alpha=5\%$)
 - Berechnung des Konfidenzintervalls durch einmalige Anwendung von Formel (4.8) auf Basis der Normal- bzw. der Student-t-Verteilung
 - Vorgegebener maximaler relativer oder absoluter Fehler
 - Sukzessive Simulation von Stichproben und Berechnung von Konfidenzintervallen auf Grund der bisherigen Stichprobenmenge bis relativer oder absoluter Fehler klein genug ist.

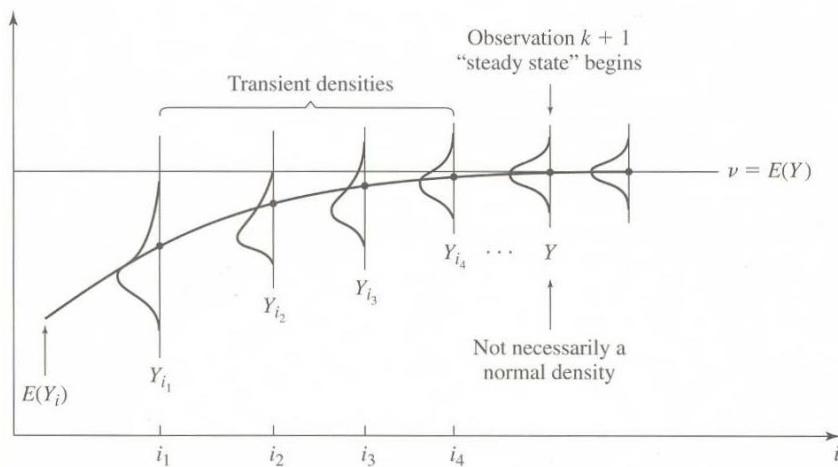
6.2.2 Nicht-terminierende Simulationen (LK 9.5)

- Technische Systeme, bei denen keine charakteristischen Zeiteinheiten zu erkennen sind, die sich auch in der Praxis als abgeschlossene Operation wiederholen und keinen Einfluss auf das zukünftige Geschehen des Systems haben.
 - Beispiele: Langzeitverhalten von
 - Fabrikanlagen
 - Rechnernetzen
 - Art der Aussagen
 - Verteilungen über das Langzeitverhalten von Systemgrößen wie Serverbelegungen, Durchsatz, Warteschlangenlängen, Wartezeiten, Verlustwahrscheinlichkeiten, etc.
 - Anfangsphase nicht relevant, da untypisch
 - Beschreibung durch die Verteilung der beobachteten Größen im „eingeschwungenen“ System (steady state distribution)
 - Probleme bei der Auswertung
 - Anfangsphase liefert untypische Ergebnisse
 - Beispiel
 - Leere Fabrik zu Beginn
 - Wartezeiten erster fertiger Produkte sind gering
 - Wartezeiten von Produkten steigen anfangs mit zunehmender Simulationszeit bis Fabrik einen typischen Betriebszustand erreicht hat
 - Aufeinanderfolgende Stichproben sind oft positiv korreliert, bisherige Stichprobentheorie verlangt aber unabhängige Stichproben
 - Beispiel
 - Typische Zeitreihe von Wartezeiten:
0, 0, 1, 1, 2, 2, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6, 6, 5, 5, 4, 3, 3, 2, 1, 1, ...
 - Autokorrelation
 - Berechnung von Konfidenzintervall für Mittelwert mit konventioneller Stichprobentheorie nicht möglich
- Diese Probleme werden in den nächsten beiden Abschnitten behandelt.

6.3 Methoden zur Bestimmung der transienten Phase

6.3.1 Transiente und stationäre Phase eines Systems

- Beispiel: M/M/1- ∞ Wartesystem
 - Beobachtete Zufallsvariablen: Wartezeit Y_i des Kunden i
 - Wartezeiten zu Beginn hängen vom Anfangszustand des Systems ab
 - Bei leerem Anfangszustand hat erster Kunde keine Wartezeit
 - Zweiter Kunde hat geringe durchschnittliche Wartezeit, da er höchstens auf einen anderen Kunden warten muss
 - Mit zunehmender Kundenzahl verschwindet diese Abhängigkeit vom Anfangszustand des Systems
- Betrachtung der simulierten Zufallsvariablen als zeitdiskreten stochastischen Prozess
 - Aufeinanderfolgende Zufallsvariablen Y_i einer konkreten Simulation sind voneinander abhängig und eine Ausprägung dieses Prozesses.
 - Betrachte viele solcher Ausprägungen; dann kann für jedes Y_i eine Verteilung angegeben werden:



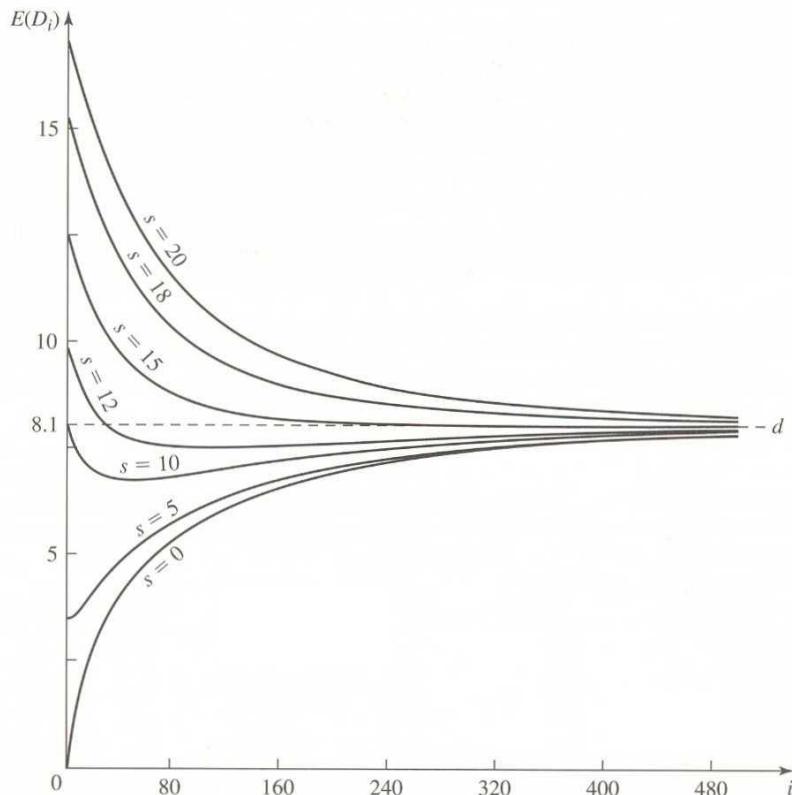
Transient and steady-state density functions for a particular stochastic process Y_1, Y_2, \dots and initial conditions I .

Abbildung 5.1: Die Verteilung der Zufallsvariablen Y_i (z.B. Wartezeit bei M/M/1- ∞) konvergiert in der transienten Phase gegen die Verteilung der Zufallsvariablen im stationären Zustand, in dem sich die Verteilungen aufeinanderfolgender Zufallsvariablen nicht mehr ändern. (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 499)

- Verteilung von Y_i ist stationär, falls sie der Verteilung von Y_k gleicht für alle $k > i$
- Gesucht: stationäre Verteilung von Y um Aussagen über das Langzeitverhalten des Systems zu machen
 - Die untypische Anfangsphase vor dem „stationären bzw. eingeschwungenen Zustand“ bzw. dem Erreichen des statistischen Gleichgewichts der Simulation wird transiente Phase, Aufwärmphase

oder Einschwingphase (warm-up period, initial transient, startup problem) genannt.

- Datenerhebung für statistische Aussagen über Y nur im eingeschwungenen Zustand
- Lösungen
 - Verwerfen der Stichproben in der Aufwärmphase der Simulation: Wann ist die zu Ende?
 - Solange simulieren bis der Einfluss der Aufwärmphase verschwindet: Wie lange?
 - Wahl eines geschickten Anfangszustandes (z.B. eines sehr wahrscheinlichen bzw. typischen Zustands): Welcher ist das?
- Probleme beim Erkennen des Endes der transientes Phase
 - Gleichgewichtszustand durch eine Verteilung gekennzeichnet, nicht durch einen konkreten Wert
 - Länge der transientes Phase stark von Bedingungen bei Simulationsbeginn abhängig



$E(D_i)$ as a function of i and the number in system at time 0, s , for the $M/M/1$ queue with $\rho = 0.9$.

Abbildung 5.2: Der Mittelwert einer Zufallsvariablen D_i (z.B. Wartezeit bei $M/M/1-\infty$) konvergiert in der transientes Phase gegen den Mittelwert d der stationären Verteilung. Dabei hängt er vom Anfangszustand s (z.B. Anzahl der Kunden in der Warteschlange zu Simulationsbeginn) ab. (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 500)

6.3.2 Naive Methode zur Erkennung des eingeschwungenen Systems

- Direkter Erkennungsversuch über Verteilungen für Y_i
- Parallele Simulationen erzeugen Ausprägungen der Y_i
- Erstellung von Histogrammen für alle Y_i
- Statistisches Gleichgewicht ist erreicht, wenn sich das Histogramm von Y_i von dem seiner Nachfolger nicht mehr unterscheidet
- Wegen des enorm hohen Simulationsaufwandes nicht durchführbar; Zufallsvariablen können auch mehrdimensional sein!
- Sobald man die stationäre Phase erkannt hätte, bräuchte man möglicherweise keine weiteren Simulationen mehr, da die stationäre Zustandsverteilung oft schon das gesuchte Ziel ist.

6.3.3 Graphisches Verfahren von Welch (LK 9.5.1)

- Heuristisches, graphisches Verfahren zur Bestimmung der Länge der transienten Phase
 - Reduktion des naiven Ansatzes auf die Mittelwertbetrachtung
 - Einfachere Betrachtung als naive Methode
 - Schnellere Konvergenz als naive Methode
- Messwerte aus n relativ langen Testläufen ($n=10$ oder 20 , X_i^j ist Messwert i im Lauf j)
 - Reihe der gemittelten Messreihen bilden $\bar{X}_i = \frac{\sum_{j=1}^n X_i^j}{n}$ (5.5)
 - Bestimme das i , ab dem die gemittelte Messreihe konvergiert und setze $l=i$. Index l markiert das Ende der initialen Phase.
 - Bei den „richtigen“ Simulationsläufen werden dann die ersten l Werte nicht berücksichtigt, d.h. bei den neuen Messwerten X_0, X_1, \dots, X_m schätzt man den Mittelwert durch $\bar{X}(m,l) = \frac{\sum_{i=l+1}^m X_i}{m-l}$ (5.6)
- Beispiel 1: Durchlaufzeit der Kunden in einem $M/M/1-\infty$ Wartesystem

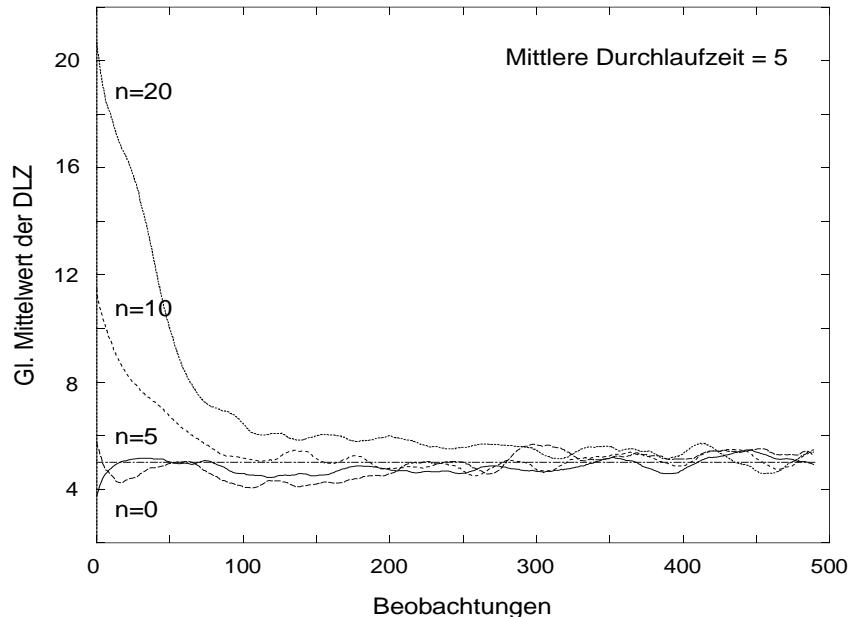
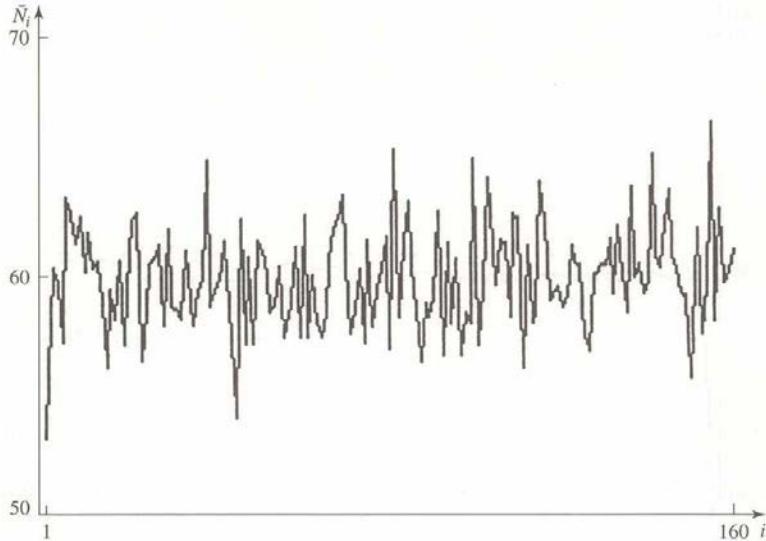


Abbildung 5.3: $M/M/1$ -Wartesystem mit 80% Auslastung und $E[B]=1$:
Mittelwert der Durchlaufzeiten bei 100 Replikationen ohne Glättung durch
gleitenden Durchschnitt.

- Beispiel 2: Stündlicher Durchsatz einer kleinen Fabrik



Averaged process for hourly throughputs, small factory.

Abbildung 5.4: Das Ende der transienten Phase ist aus der gemittelten Reihe mehrerer Simulationsläufe noch nicht zu erkennen. (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 523)

- An Beispiel 2 sehen wir, dass die Reihe der gemittelten Zeitreihen nicht immer eine initiale Phase erkennen lässt.
- Darum wende anschließend einen gleitenden Mittelwert (moving average) mit halber Fensterbreite w zur Beseitigung hochfrequenter Schwingungen an:

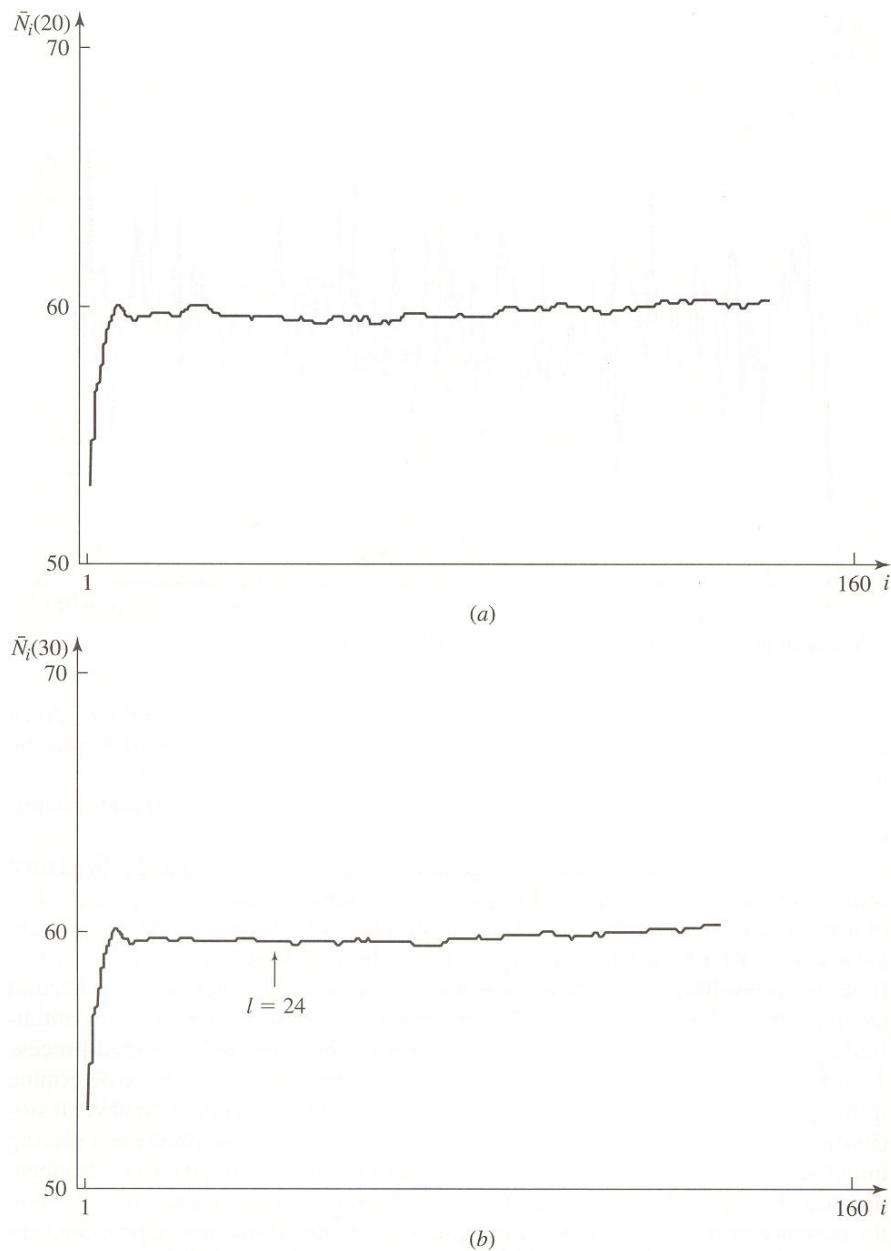
$$\bar{X}_i(w) = \begin{cases} \frac{\sum_{s=-w}^w \bar{X}_{i+s}}{2w+1} & \text{falls } i > w \\ \frac{\sum_{s=-(i-1)}^{i-1} \bar{X}_{i+s}}{2i+1} & \text{falls } 1 \leq i \leq w \end{cases} \quad (5.7)$$

- Beispiel:

$$\begin{array}{ll} X_i: & 1.0 \ 2.0 \ 2.0 \ 1.0 \ 2.0 \ 3.0 \ 3.0 \ 4.0 \ 5.0 \ 2.0 \ 6.0 \ 5.0 \ 2.0 \ 3.0 \ 3.0 \ 1.0 \ 2.0 \\ \bar{X}_i: & 1.7 \ 1.7 \ 1.7 \ 2.0 \ 2.7 \ 3.3 \ 4.0 \ 3.7 \ 4.3 \ 4.3 \ 4.3 \ 3.3 \ 2.7 \ 2.3 \ 2.0 \end{array}$$

Abbildung 2: Originale und gemittelte Zeitreihe (halbe Fensterbreite $w=1$).

- Verschiedene Werte w testen, geglättete Mittelwertreihe $\bar{X}_i(w)$ mit geeignetem Parameter w wählen, bei der sich eine Konvergenz der Reihe deutlich feststellen lässt, um Index 1 zu bestimmen.
- Falls keine brauchbaren Kurven zu finden, 10 bis 20 weitere Läufe durchführen und Prozedur wiederholen.



Moving averages for hourly throughputs, small factory: (a) $w = 20$; (b) $w = 30$.

Abbildung 5.5: Das Einschwingen des Systems ist zu Beginn des Prozesses gut durch die Glättung der gemittelten Messreihe zu erkennen. (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 524)

Als Ende der Einschwingphase wird $l=24$ Stunden gewählt. Dieser Wert ist zur Sicherheit deutlich größer angenommen als der Wert, den man in der Graphik als Grenze erkennt (z.B. $l=6$ Stunden).

6.3.5 Elegante Berechnung des Endes der Transienten Phase

- Beispiel: M/M/n Verlustsystem
 - System mit n parallelen Servern ohne Warteraum
 - Modell zum Dimensionieren von Leitungskapazitäten
 - Angenommen, zu Beginn der Simulation wäre schon ein Kunde im System gewesen.
 - Die Wahrscheinlichkeit, dass er sich nach t Zeit noch immer im System befindet ist $\exp(-\mu \cdot t)$, wobei $1/\mu$ die mittlere Bedienzeit ist.
 - Falls das System zu Beginn der Simulation voll besetzt gewesen wäre, ist die Anzahl der zur Zeit t noch aktiven Nutzer vom Beginn der Simulation nach Binom(n, $\exp(-\mu \cdot t)$) verteilt.
 - Nach welcher Zeit t ist die Wahrscheinlichkeit, dass kein Benutzer vom Beginn der Simulation noch aktiv ist, größer als p?

$$(1 - \exp(-\mu \cdot t))^n \geq p$$

$$t \geq \frac{\ln(1 - \sqrt[n]{p})}{-\mu} = -\ln(1 - \sqrt[n]{p}) \cdot \frac{1}{\mu} \quad (5.10)$$
 - Zahlenbeispiel: p=0.99, n=100:
Nach 9.21-facher mittlerer Bedienzeit ist die Wahrscheinlichkeit, dass kein Benutzer seit Systemstart noch aktiv ist, größer als 99%
- Anwendbarkeit stark modellabhängig

6.3.6 Sinnvolle Initialisierung des Systems

- Beispiel: M/M/n-∞ Wartesystem
 - System mit n parallelen Servern mit Warteraum
 - Mittlere Anzahl von Kunden im System: $a = \lambda / \mu$ wobei
 - die mittlere Zwischenankunftszeit $1/\lambda$ und
 - die mittlere Bedienzeit $1/\mu$ ist.
 - Initialisiere die Simulation mit a Kunden und ermittle ihre Restarbeitszeit
 - Wegen Gedächtnislosigkeit der Exponentialverteilung kann einfach eine volle zufällige Bedienzeit dafür genommen werden.
- Anwendbarkeit ebenfalls stark modellabhängig

6.3.7 Test auf Schätzwertverfälschung (Bias) durch transiente Phase

Siehe später bei Signifikanztests von Hypothesen!

6.4 Stichprobentheorie für korrelierte Zeitreihen (LK 4.4)

- Ziel: Konstruktion von Konfidenzintervallen für den Mittelwert einer Zeitreihe X_i
- Problem: Bisherige Stichprobentheorie nicht anwendbar
 - Theorie fordert unabhängige Stichproben
 - Werte einer Zeitreihe (aus Simulation) sind in der Regel stark korreliert

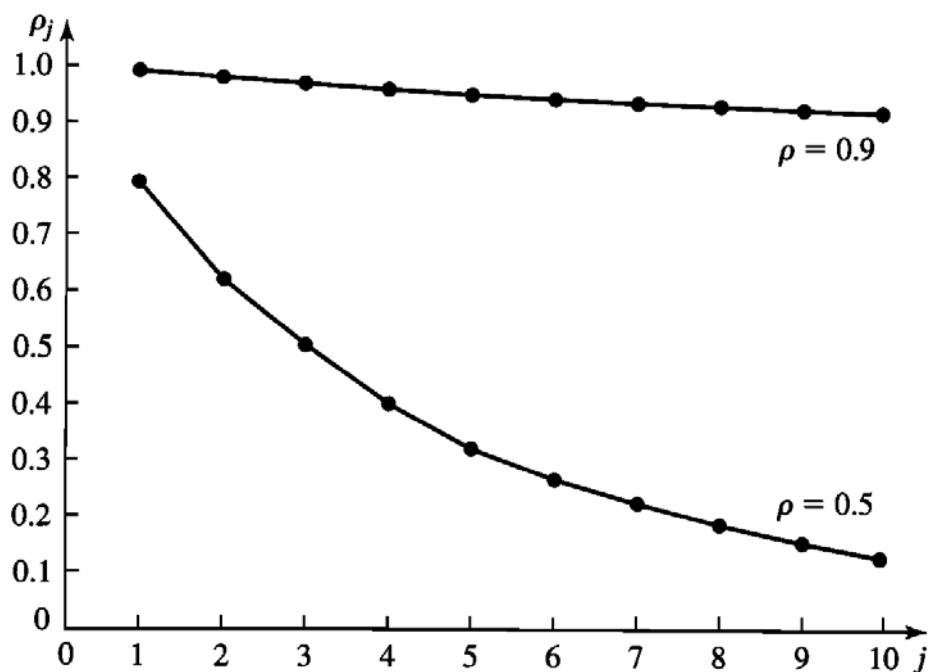


Abbildung 5.6: Korrelation aufeinanderfolgender Stichproben bei einem M/M/1-Wartesystem mit Auslastung ρ . (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 248)

6.4.1 Stichprobentheorie für korrelierte Zeitreihen (LK 4.4)

- Es sei $c = 1 + 2 \cdot \sum_{0 < j < n} (1 - j/n) \cdot \rho_j$ (5.11)
 - n ist die Größe der Stichprobe X_i und es gilt $0 \leq i \leq n$
 - ρ_j sind die Korrelationskoeffizienten der Zeitreihe X_i und es gilt $0 < j < n$
 - Bei positiv korrelierten Zeitreihen gilt $c > 1$.
 - Problem: bei korrelierten Stichproben schätzt $S^2(n)$ die Varianz σ^2 einer Stichprobenmenge falsch, vielmehr gilt:

$$E[S^2(n)] = \text{VAR}[X] \cdot \frac{n-c}{n-1}$$
 (5.12)
- Folge: Stichprobenvarianz wird bei positiver Korrelation unterschätzt.

- Zusätzlich berechnet sich die Varianz des Mittelwertes anders als bei unabhängigen Stichprobenwerten (siehe Gl. 4.1):

$$\text{VAR}[\bar{X}(n)] = \frac{\text{VAR}[X]}{n} \cdot c \quad (5.13)$$

Folge: auch sie wird mit Gl. 4.1 unterschätzt, so dass davon abgeleitete Konfidenzintervalle zu klein wären.

- Aus Kombination von (5.12) und (5.13) ergibt sich

$$\text{VAR}[\bar{X}(n)] = \frac{E[S^2(n)]}{n} \cdot \frac{n-1}{n-c} \cdot c \quad (5.14)$$

Für unabhängige Stichproben erhalten wir mit $c=1$ genau Gl. 4.1

- Beispiel aus Abbildung 5.6:

- Annahme: exponentieller Abfall der Korrelationswerte: $\rho_n = \rho_1 \cdot p^{n-1}$

(Wert für p kann mit $p = \sqrt[9]{\frac{\rho_{10}}{\rho_1}}$ aus Abbildung 5.6 ermittelt werden.)

- Sei $c(n)$ der für c errechnete Wert bei einer Stichprobengröße von n .

Dann bezeichnen wir mit $bias(n) = \frac{n-1}{n-c(n)} \cdot c(n)$ den multiplikativen Bias bezüglich der konventionell berechneten Varianz.

- Beispiel 1: Korrelationen $\rho_1 = 0.88, \rho_{10} = 0.22 \Rightarrow p=0.8572$

n	c(n)	bias(n)
10	2.95	3.78
100	5.58	5.86
1000	5.96	5.99
10000	6.00	6.00

- Beispiel 2: Korrelationen $\rho_1 = 0.99, \rho_{10} = 0.9 \Rightarrow p=0.9895$

n	c(n)	bias(n)
10	5.24	9.91
100	36.04	55.78
1000	85.01	92.81
10000	93.04	93.90
100000	93.84	93.93
1000000	93.93	93.93

- Konventionelle Konfidenzintervallhalblänge: $\delta(n, \alpha) = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s^2(n)}{n}}$

Korrektur: $\delta^*(n, \alpha) = z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s^2(n)}{n} \cdot bias(n)} = \delta(n, \alpha) \cdot \sqrt{bias(n)}$

6.5 Robuste Strategien für die Berechnung von Konfidenzintervallen für korrelierte Zeitreihen

Wir entwickeln zwei Strategien zur Berechnung eines Konfidenzintervalls für den Mittelwert einer korrelierten Zeitreihe.

- Naiver, direkter Ansatz
 - Aufwändig, weil zur Berechnung von c (5.11) sehr viele Autokorrelationen ρ_j benötigt werden
 - Schlecht, weil es keine guten Schätzer für die Autokorrelationen ρ_j gibt
- Indirekter Ansatz basierend auf Mittelwertstichproben
 - Problem: Einzelmesswerte sind korreliert und allgemeine Stichprobentheorie ist nicht anwendbar zur Berechnung von Konfidenzintervallen
 - Idee: Bilde Mittelwerte $\bar{X}(m)_j$ aus m aufeinanderfolgenden Einzelmesswerten
 - Die Reihe der Mittelwerte $\bar{X}(m)_j$ hat denselben Mittelwert wie die Reihe der Einzelmesswerte.
 - Mittelwerte sind i.d.R. normalverteilt
 - Unkorreliertheit der Reihe der Mittelwerte $\bar{X}(m)_j$ kann empirisch überprüft werden.
 - Keine oder nur schwache Korrelation dieser Mittelwerte, falls m groß genug, so dass annähernd Unabhängigkeit angenommen werden kann.
 - Berechnung von approximativen Konfidenzintervallen auf Basis dieser Mittelwerte möglich

Neben dem indirekten Ansatz basierend auf Mittelwertstichproben gibt es noch weitere Verfahren zur Berechnung von Konfidenzintervallen für Mittelwerte von Einzelmesswerten. Sie alle versuchen die Varianz der Mittelwerte $\text{VAR}[\bar{X}(n)]$ auf unterschiedliche Arten zu schätzen (siehe LK 9.5.3).

Es gibt zwei vorrangige Verfahren zur Erzeugung der Mittelwertstichproben: Batch-Means und Replicate-Delete.

6.5.1 Batch-Means Verfahren (LK 9.5.3)

- Variante A: die Stichproben liegen schon vor
 - Entfernung der transienten Phase
 - Einteilung der Messwerte in n gleichgroße Intervalle (Batches) der Größe m ; vermeide dabei Korrelationen zwischen benachbarten Mittelwerten; Verfahren für die Wahl von m :
 - Bestimmung der empirischen Autokorrelationskoeffizienten $\hat{\rho}_k$ der Messwerte (ohne jene der transienten Phase)
 - Wahl von k_{\max} so, dass die $\hat{\rho}_k$ für $k > k_{\max}$ praktisch verschwinden
 - Randbedingungen
 - $m \geq 10 \cdot k_{\max}$ (m : Größe des Batches)
 - $n \geq 10$ (n : Anzahl der Batches)
 - Berechnung der Konfidenzintervalle basierend auf den Mittelwerten der Messwerte in den n Batches

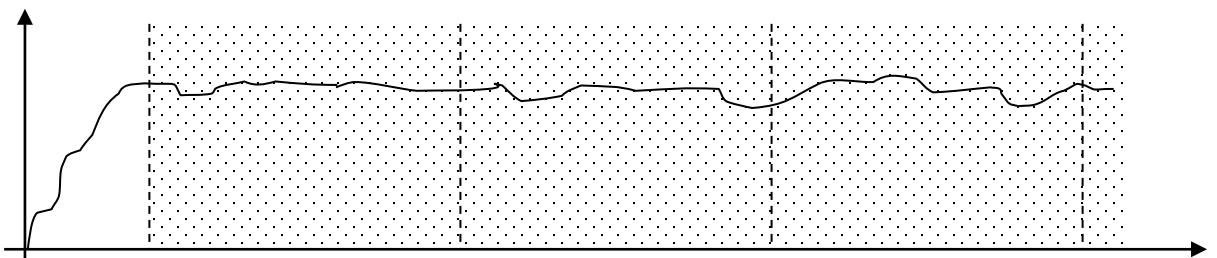


Abbildung 5.8: Beim Batch-Means Verfahren wird ein sehr langer Simulationslauf gestartet, dessen transiente Phase wird verworfen und der Restlauf wird in n Teile zerlegt, deren Mittelwerte unabhängige Stichproben zur Ermittlung von Konfidenzintervallen ergeben.

- Variante B: die Stichproben müssen noch erzeugt werden
 - Ziel: Simulation von Stichproben bis Konfidenzintervalle klein genug sind hinsichtlich des relativen oder absoluten Fehlers (siehe 5.1)
 - Bestimmen der transienten Phase
 - Wahl eines geeigneten m
 - Simulation so vieler Batches bis Konfidenzintervalle klein genug sind
 - Überprüfung der Mittelwertstichproben auf Korrelationen um Fehler durch zu kleine Batches auszuschließen
- Vorteile
 - Geringer Aufwand durch einmalige transiente Phase
 - Fehler bei der Schätzung der transienten Phase haben geringen Einfluss

- Nachteile
 - Bestimmung von m (bzw. von n und m) schwierig
 - Korrelationen möglich \Rightarrow Berechnung der Korrelationskoeffizienten zur Vermeidung korrelierter Intervallmittelwerte notwendig

6.5.2 Replicate/DeleteVerfahren (LK 9.5.2)

- Durchführung
 - Transiente Phase bestimmen
 - Läufe durchführen (Replicate), Länge: Vielfaches der transienten Phase
 - Messwerte der transienten Phase abschneiden (Delete)
- Berechnung der Konfidenzintervalle basierend auf den Mittelwerten der Einzelläufe bzw. Anwendung der Verfahren aus 5.1

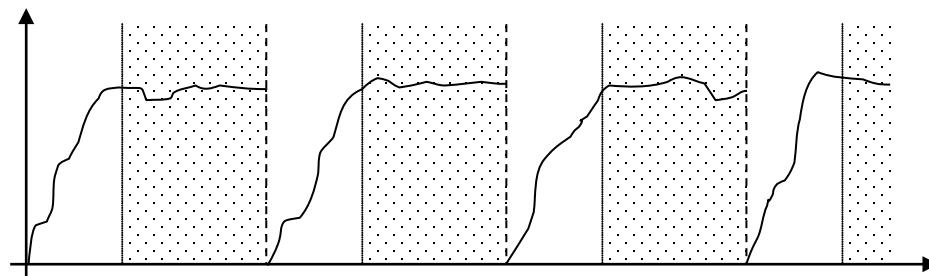


Abbildung 5.7: Beim Replicate/Delete Verfahren werden n Läufe hintereinander gestartet, die transiente Phase wird verworfen und die Mittelwerte der Restläufe ergeben unabhängige Stichproben zur Ermittlung von Konfidenzintervallen.

- Vorteile
 - Einfach
 - Nicht anfällig gegen Korrelationen
 - Von allen gängigen Simulationspaketen unterstützt
- Nachteile
 - Korrekte Bestimmung des Endes der transienten Phase hat großen Einfluss auf die Güte der Ergebnisse
 - Hoher Aufwand durch die Simulation vieler ungenutzter transienter Phasen

6.5.3 Konfidenzintervalle für andere Leistungskriterien als Mittelwerte (LK 9.5.4)

- Wahrscheinlichkeiten
 - Prinzipiell wie bei Mittelwerten (vgl. 4.1.5)
 - Verteilungswerte $P(X \leq x)$ sind auch Wahrscheinlichkeiten
 - Falls solche Wahrscheinlichkeiten für mehrere Werte von x berechnet werden, kann man eine Verteilungsfunktion mit Konfidenzintervallen approximieren
- Quantile
 - Wiederholung der Definition des Quantils: $Q_p = \inf\{x \in W(X) : P(X \leq x) \geq p\}$ ist ein Wert aus dem Wertebereich der Stichprobenmenge, evtl. mit Einheiten behaftet
 - p ist gegeben, Q_p ist gesucht
 - Konfidenzintervalle sind schwierig anzugeben, kein allgemein anerkanntes Verfahren

- Quotienten
 - Beobachtungsgrößen X_i und Y_i (iid)
 - Gesucht $\phi = \frac{E[X]}{E[Y]}$, Achtung: nicht $\phi = E\left[\frac{X}{Y}\right]!$
 - Problem: einfacher Schätzer $\hat{\phi}_C(n) = \frac{\bar{X}(n)}{\bar{Y}(n)}$ ist nicht erwartungstreu und hat keine Konfidenzintervalle
 - Lösung: Jackknife Estimator $\hat{\phi}_J(n)$
 - zwar kein erwartungstreuer Schätzer für $\phi = \frac{E[X]}{E[Y]}$
 - aber ein besserer als $\hat{\phi}_C(n)$.
 - Idee: Konstruiere aus den X_i und Y_i neue Zufallsvariablen θ_i um durch diese neuen Stichproben Konfidenzintervalle zu erzeugen:

- $\theta_i = n \cdot \hat{\phi}_C(n) - (n-1) \cdot \frac{\sum_{\substack{0 \leq j < n, j \neq i}} Y_j}{\sum_{\substack{0 \leq j < n, j \neq i}} X_j}$ für $0 \leq i < n$
- Mittelwert: $\hat{\phi}_J(n) = \frac{1}{n} \cdot \sum_{0 \leq i < n} \theta_i$
- Varianz: $\hat{\sigma}_J^2(n) = \frac{\sum_{0 \leq g < n} (\theta_g - \hat{\phi}_J(n))^2}{n-1}$
- Es gilt: $\frac{\hat{\phi}_J(n) - \phi}{\sqrt{\hat{\sigma}_J^2(n)/n}} \xrightarrow{\mathcal{D}} N(0,1)$ (Konvergenz der Verteilung)
- Konfidenzintervall für das Konfidenzniveau $1-\alpha$:

$$\hat{\phi}_J(n) \pm z_{1-\alpha/2} \cdot \sqrt{\hat{\sigma}_J^2(n)/n}$$
- Nachteil: für große Stichproben aufwändig

6.6 Simulation nicht-ergodischer Prozesse

6.6.1 Ergodische und nicht-ergodische Prozesse

6.6.1.1 Definitionen

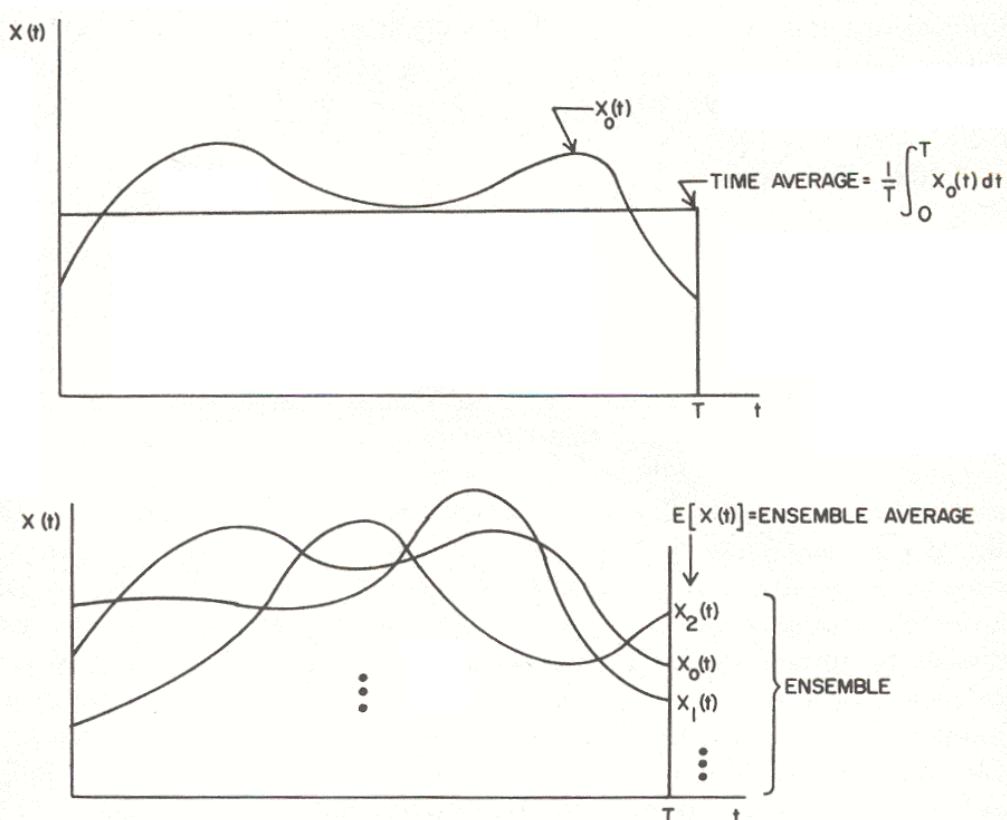


Abbildung 5.9: Zeitmittel und Ensemble-Mittel eines stochastischen Prozesses. (aus: Fundamentals of Queuing Theory, Gross/Harris, 1985, 2nd edition)

Definition: Zeit-Mittel eines stochastischen Prozesses über den Zeitraum T:

$$\overline{X_T^k} = \frac{1}{T} \cdot \int_0^T [X_0(t)]^k dt, \text{ wobei mit } X_0(t) \text{ nur eine einzige Realisierung des}$$

stochastischen Prozesses beobachtet wird.

Definition: Ensemble-Mittel eines stochastischen Prozesses zum Zeitpunkt t:

$$E[X^k(t)] = \lim_{n \rightarrow \infty} \frac{\sum_{0 \leq i < n} [X_i(t)]^k}{n} = m_k(t), \text{ wobei } X_i(t) \text{ unterschiedliche Realisierungen}$$

des stochastischen Prozesses zum Zeitpunkt t sind.

Definition: Ein stochastischer Prozess ist ergodisch für das k-te Moment, falls gilt:

$$\lim_{t \rightarrow \infty} m_k(t) = \lim_{T \rightarrow \infty} \overline{X_T^k} = \overline{X^k}$$

Definition: Ein stochastischer Prozess ist ergodisch, falls er für alle Momente ergodisch ist.

Folgerung: Falls ein stochastischer Prozess Mittelwert-ergodisch ist, kann man ihn mit der Batch-Means-Methode simulieren.

6.6.1.2 Beispiel eines Nicht-Ergodischen Prozesses

Zu Beginn des Prozesses wird eine Zahl gewürfelt, die der Prozess für immer als Zustand annimmt.

- Die Zeitmittelwerte $\overline{X_T^1}$
 - haben den Wertebereich $\{1, 2, 3, 4, 5, 6\}$.
 - hängen zwar nicht von T ab aber von der Realisierung des Prozesses.
 - sind somit nicht eindeutig.
- Der Ensemble-Mittelwert $m_1(t)$ für beliebige Zeitpunkte ist 3,5.
- Wegen $m_1(t) \neq \overline{X_T^1}$ ist dieser Prozess nicht Mittelwert-ergodisch.

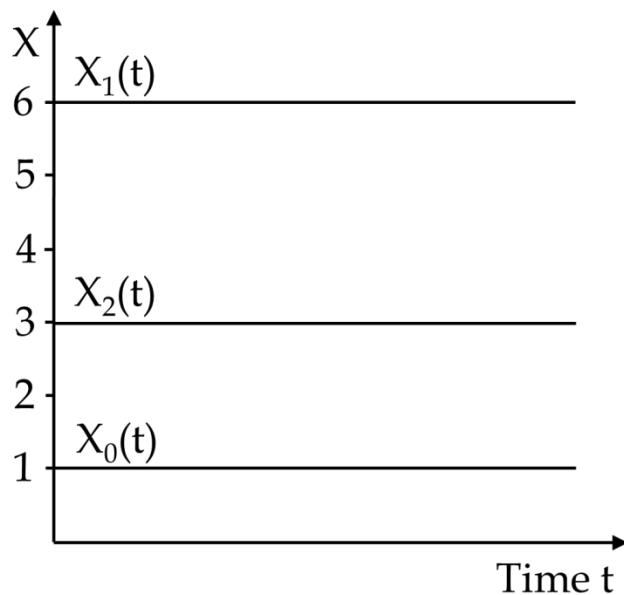
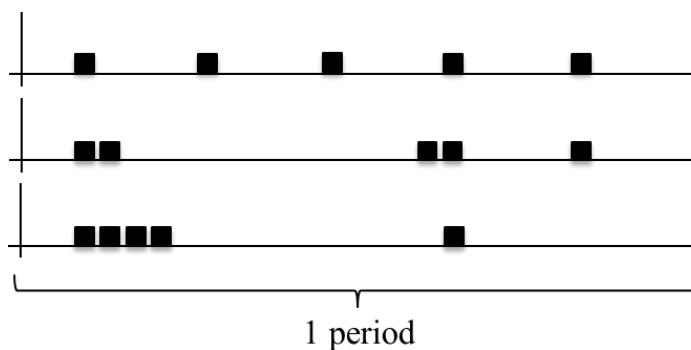


Abbildung 3: Drei Realisierungen eines einfachen nicht-ergodischen Prozesses.

6.6.1.3 Weitere Beispiele für Ergodische und Nicht-Ergodische Prozesse

- Ergodische Prozesse
 - Wartesysteme (GI/GI/n-∞)
 - Warteverlustsysteme (GI/GI/n-S)
 - Charakteristisch
 - Zufällige und unabhängige Ankunfts- und Bedienzeiten

- Nicht-ergodische Prozesse
 - Wartesysteme ($n \cdot D/D/m - \infty$)
 - Warteverlustsysteme ($n \cdot D/D/m - S$)
 - Charakteristisch
 - n verschiedene Quellen senden Kunden in die Warteschlange
 - Die Zwischenankunftszeit zwischen Kunden einer Quelle ist konstant \Rightarrow periodische Ankunftszeiten bzw. -muster
 - Kunden erfahren konstante Bedienzeit



- Abbildung 4: Unterschiedliche Sendemuster (Phasen der Sendezeitpunkte innerhalb einer Periode).
 - Dieselben Systeme mit Varianten bezüglich Paketgrößen
 - On/off-Modulation (On/off-Quellen); erzeugt von Sprach-Codecs mit Silence Detection
 - Variable Paketgrößen gemäß einer Verteilung, wobei aufeinanderfolgende Paketgrößen i.d.R. korreliert sind; erzeugt von Sprach-Codecs für den Mobilfunk oder Skype

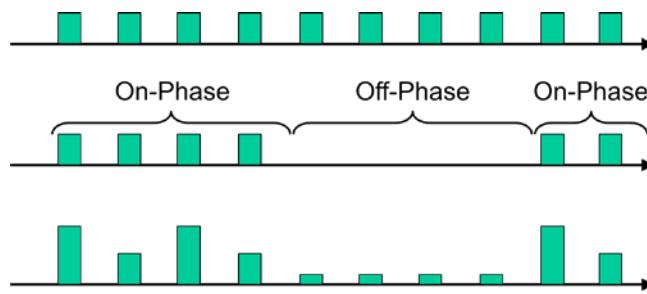


Abbildung 5: Konstante Paketgrößen, on/off-modulierte Paketgrößen und korrelierte, unterschiedliche Paketgrößen. Pakete mit unkomprimierten Sprachproben, die alle 20 ms gesendet werden, enthalten $\frac{1 \text{ byte}}{\text{sample}} \cdot \frac{8000 \text{ samples}}{\text{s}} \cdot 20 \text{ ms} = 160 \text{ bytes}$ Nutzlast.

6.6.2 Simulation von periodischen Systemen

- Anwendungsbeispiel: Echtzeitverkehr
 - Datenquelle generiert mit einer festen Periode A Sprachproben, Videoframes, etc.
 - Datenpakete
 - (i) können gleich groß sein,
 - (ii) können aber auch gemäß einer empirischen Verteilung iid sein oder on/off-moduliert sein
 - Wenn die Daten von n solchen Quellen mit gleicher Periode A auf ein gemeinsames Übertragungsmedium gemultiplext werden, dann wiederholen sich die Ankunfts muster in einem Zeitfenster der Länge A
 - Die Paketverzögerungen sind für jedes Intervall der Länge A für die Verbindungen einer Quelle
 - (i) gleich bzw.
 - (ii) ähnlich,
da sie hauptsächlich durch die Phasenlage der Quellen innerhalb einer Periode (Ankunfts muster) bestimmt sind.
- Folgerung für Simulationen
 - Problem: Wartezeit hängt von Ankunfts muster ab
 - Lösung
 - Simuliere viele Ankunfts muster und
 - (i) Berechne die Wartezeit deterministisch für eine Periode
 - (ii) Simuliere die Wartezeit stochastisch für mehrere Perioden
 - Die Mittelwerte der Wartezeit für unterschiedliche Ankunfts muster bilden die Stichprobenmenge für
 - die Ensemble-Mittelung für die Gesamtstatistik
 - die Berechnung von Konfidenzintervallen
 - Im Fall (i) handelt es sich um eine typische Monte-Carlo-Simulation
 - Unabhängige Muster werden ausgewürfelt.
 - Auf Basis eines Musters wird eine Stichprobe rein deterministisch berechnet, was durch Discrete Event Simulation geschehen kann, aber andere Verfahren sind auch möglich.
 - Auf Basis einer Stichprobenmenge wird eine Statistik erstellt.
 - Im Fall (ii) benutzt man Discrete Event Simulation (dynamisch) und die Replicate/Delete Methode
 - Batch-Means Methode nicht anwendbar wegen fehlender Mittelwert-Ergodizität des Systems

nicht im Detail (keine Formeln)

6.7 Vergleich alternativer Systemkonfigurationen (LK 10)

- Testkriterium: Der Unterschied zwischen zwei Systemen ist statistisch signifikant, falls 0 nicht mehr im Konfidenzintervall der Differenz liegt.
 - Zwei Methoden zur Bestimmung der Konfidenzintervalle
 - t-Confidence Interval (t-C.I.) für Paare
- Student-t*
- erfordert gleiche Anzahl von Stichproben der zu vergleichenden Systeme
 - Modifizierte t-C.I. für zwei Stichproben
 - erfordert Unabhängigkeit der Stichproben

6.7.1 t-Konfidenzintervall für Paare (Paired-t C.I., LK 10.2.1)

- Gegeben
 - (Mittelwert)stichproben zweier Systeme A und B: X_i^A und X_i^B , $0 \leq i < n$
- Berechne
 - Folge der Differenzen $Z_i = X_i^A - X_i^B$
 - Konfidenzintervall mit Signifikanz α auf Basis der Student-t-Verteilung für den empirischen Mittelwert \bar{Z} (t-Konfidenzintervall)
 - Falls aufeinanderfolgende Differenzen Z_i korreliert sind, dürfen Konfidenzintervalle nur auf Basis von Mittelwertstichproben berechnet werden.
- Falls das t-Konfidenzintervall die Null nicht enthält, dann sind die wahren Mittelwerte für X^A und X^B mit Wahrscheinlichkeit $1-\alpha$ unterschiedlich.
- Vorteil: Annahmen über Unabhängigkeit beider Messreihen nicht notwendig
- **Nachteil: Beide Messreihen müssen gleich lang sein**
 - Wenn das nicht der Fall ist, können nicht alle Messwerte genutzt werden.
 - Könnte nicht erfüllt sein, wenn z.B. Simulationsdaten und reale Messdaten verglichen werden.

nicht im Detail (keine Formeln)

6.7.2 Modifiziertes t-Konfidenzintervall für zwei Stichproben (LK 10.2.2)

- Problem beim paired t-confidence interval: gleichgroße Stichprobenmengen
- Lösung: Welch-Konfidenzintervall (Welch, 1938)
 - Annahme
 - Unabhängige Stichprobenreihen X_{1j} und X_{2j}
 - Normalverteilte Stichproben
 - Berechne $\bar{X}_i(n_i) = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}$ und $S_i^2(n_i) = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} [X_{ij} - \bar{X}_i(n_i)]^2$
für $i = 1, 2$.
 - Schätzung der Freiheitsgrade: der Student-t-Verteilung

$$\hat{f} = \frac{\left[S_1^2(n_1) / n_1 + S_2^2(n_2) / n_2 \right]^2}{\left[S_1^2(n_1) / n_1 \right]^2 / (n_1 - 1) + \left[S_2^2(n_2) / n_2 \right]^2 / (n_2 - 1)}$$
 - Approx. $100(1-\alpha)\%$ C.I.: $\bar{X}_1(n_1) - \bar{X}_2(n_2) \pm t_{\lfloor \hat{f} \rfloor, 1-\alpha} \sqrt{\frac{S_1^2(n_1)}{n_1} + \frac{S_2^2(n_2)}{n_2}}$
- Vorteil: $n_1 = n_2$ nicht nötig
 - Gute Ausnutzung aller vorhandener Daten
- Nachteile: Vorausgesetzte Annahmen oft nicht erfüllt

6.7.3 Gleichzeitiges Schätzen von mehreren Konfidenzintervallen (LK 10.3)

- Gegeben
 - System, bei dem n unterschiedliche Parameter, z.B. Mittelwerte $\mu_i, 0 \leq i < n$, geschätzt werden mit Konfidenzintervallen für ein Signifikanzniveau von α_i
- Frage: Wie groß ist die Wahrscheinlichkeit, dass die wahren Werte für die Parameter alle in den Konfidenzintervallen enthalten sind?
 - Antwort: $\prod_{0 \leq i < n} (1 - \alpha_i)$ z.B. $n=4, \alpha_i = 10\%$
 $0,9 \cdot 0,9 \cdot 0,9 \cdot 0,9 \approx 0,65$

6.7.3.1 Bonferroni-Ungleichung

- Bonferroni-Ungleichung: $\prod_{0 \leq i < n} (1 - \alpha_i) \geq \max(1 - \sum_{0 \leq i < n} \alpha_i, 0)$
 $1 - (0,1 + 0,1 + 0,1 + 0,1) = 0,6$
- Beispiel 1
 - $n=10$, gewünschtes Signifikanzniveau für Gesamtvergleich: $\alpha=10\%$
 - Wie groß muss Signifikanzniveau für Einzelintervalle sein?
 - Mit Bonferroni: $\alpha_i=1,000\%$
 - Exakt: 1,048%
- Beispiel 2
 - $n=10$, Signifikanzniveau für Einzelintervalle: $\alpha_i=0,1$
 - Signifikanzniveau für Gesamtvergleich (Wahrscheinlichkeit, dass **nicht** alle 10 Parameter in ihrem Konfidenzintervall enthalten sind)
 - Mit Bonferroni: 100%
 - Exakt: gut 65%
 - Keine hilfreiche Aussage

6.7.3.2 Vergleich mit einem Referenzsystem (LK 10.3.1)

- k Systeme, eines davon Referenzsystem $\Rightarrow k-1$ Vergleiche
- Mögliches Ziel: Referenzsystem ist besser als die $k-1$ Vergleichssysteme
- Gewünschtes Signifikanzniveau für Gesamtvergleich α
- Mit Bonferroni: $\alpha_i = \frac{\alpha}{k-1}$

6.7.3.3 Paarweise Vergleiche (LK 10.3.2)

- k Systeme $\Rightarrow k(k - 1)/2$ Vergleiche
- Mögliches Ziel: Sortierung von k Systemen nach Leistung
- Gewünschtes Signifikanzniveau für Gesamtvergleich α
- Mit Bonferroni: $\alpha_i = \frac{2 \cdot \alpha}{k \cdot (k-1)}$

6.8 Simulationsbeschleuniger

Zwei Beispiele

- Geeignete Modellierung von Systemen mit abhängigen Zufallsvariablen
- Vergleich von Systemen unter vergleichbaren Bedingungen

6.8.1 Geeignete Modellierung von Systemen mit abhängigen Zufallsvariablen (LK 11.6 Conditioning)

- Gegeben
 - Zwei Zufallsvariablen $X(t)$ und $Y(t)$ in einer Simulation
 - Beispiel
 - X : Örtliche Verteilung von Nutzern in einer Mobilfunkzelle
 - Y : Interferenz in einer Mobilfunkzelle in Abhängigkeit von X ; für diese kann ein Mittelwert $\bar{Y}(X)$ analytisch berechnet werden.
- Gesucht: \bar{Y}
- Naiver Ansatz
 - X kann simuliert werden in Abhängigkeit der Zeit
 - Batch-Means oder Replicate/Delete Methode für Erzeugung von Konfidenzintervallen
 - Hoher Aufwand um den Wertebereich von X hinreichend gut abzudecken, wenn die Werte von X zeitlich korreliert sind (=sich nur langsam verändern)
- Cleverer Ansatz
 - Erzeuge Samples (Snapshots) X_i mit Hilfe einer Monte-Carlo-Simulation
 - Berechne Mittelwert und Konfidenzintervall von Y auf Basis der Stichproben $\bar{Y}(X_i)$

6.8.2 Vergleich von Systemen unter vergleichbaren Bedingungen

(LK 11.2 Common Random Numbers)

6.8.2.1 Herstellung von vergleichbaren Bedingungen

- Beispiel: M/M/n-Wartesystems
 - Vergleiche zwei Systemalternativen, welches führt zu kürzeren Wartezeiten?
 - M/M/1 mit einfacher Bedienzeit
 - M/M/2 mit doppelter Bedienzeit
- Verwende bei der Simulation aller Systeme dieselben Zufallszahlen zur Erzeugung von Zufallsvariablen
 - aufeinanderfolgender Zwischenankunftszeiten
 - aufeinanderfolgender Bedienzeiten
- Wichtig: entsprechende Zufallszahlen müssen für korrespondierende Zwecke verwendet werden
 - Nutze separate Zufallszahlenströme für Zwischenankunftszeiten und Bedienzeiten
 - Achtung bei Vergleich von Systemen mit unterschiedlichen Verteilungen: nicht anwendbar, wenn variable Anzahl von Zufallszahlen zur Erzeugung entsprechender Zufallsvariablen benötigt wird

6.8.2.2 Beispiel

Führt ein M/M/1 oder ein vergleichbares M/M/2-Wartesystem zu kürzeren Wartezeiten?

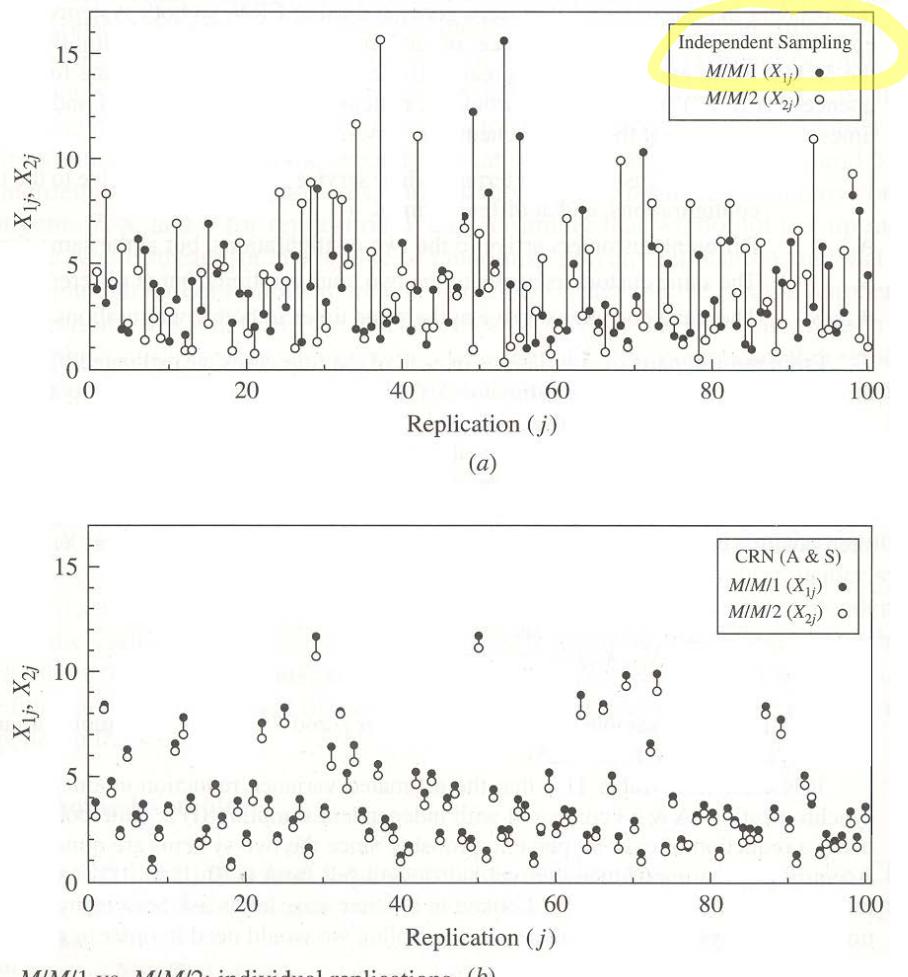


Abbildung 5.10: Differenz der Wartezeiten in vergleichbaren M/M/1 und M/M/2 Systemen. Der Vorteil des M/M/2 Systems ist bei der Nutzung gemeinsamer Zufallszahlen viel leichter ersichtlich. (aus Law/Kelton: "Simulation Modeling and Analysis", 3rd Edition, S. 592)

6.8.2.3 Mathematische Erklärung der Simulationsbeschleunigung durch vergleichbare Bedingungen

Systemvergleiche unter vergleichbaren Bedingungen scheinen vernünftig – aber warum muss man weniger simulieren?

- Unter vergleichbaren Bedingungen führen ähnliche Systeme zu tendenziell größeren bzw. kleineren Ergebnissen X_i^A und X_i^B (z.B. mittlere Wartezeiten beim i-ten Simulationsdurchlauf), d.h. sie sind positiv korreliert, so dass $\text{COV}(X^A, X^B) > 0$
- Achtung: durch die Korrelation von X_i^A und X_i^B wird nicht die Korrelation aufeinander folgender Zufallsvariablen $Z_i = X_i^A - X_i^B$ erhöht! Also kann die übliche Stichprobentheorie noch angewandt werden. Im obigen Beispiel sind die X_i^A und X_i^B Mittelwertstichproben einzelner Simulationsläufe.
- Allgemein gilt: $\text{VAR}[a \cdot X + b \cdot Y] = a^2 \cdot \text{VAR}[X] + 2 \cdot a \cdot b \cdot \text{COV}[X, Y] + b^2 \cdot \text{VAR}[Y]$

Beweis:

Es gilt:

$$\begin{aligned} E[a \cdot X + b \cdot Y] &= a \cdot E[X] + b \cdot E[Y] \\ E[(a \cdot X + b \cdot Y)^2] &= E[a^2 \cdot X^2 + 2 \cdot a \cdot b \cdot E[X \cdot Y] + b^2 \cdot Y^2] = \\ &= a^2 \cdot E[X^2] + 2 \cdot a \cdot b \cdot E[X \cdot Y] + b^2 \cdot E[Y^2] \end{aligned}$$

Daraus folgt:

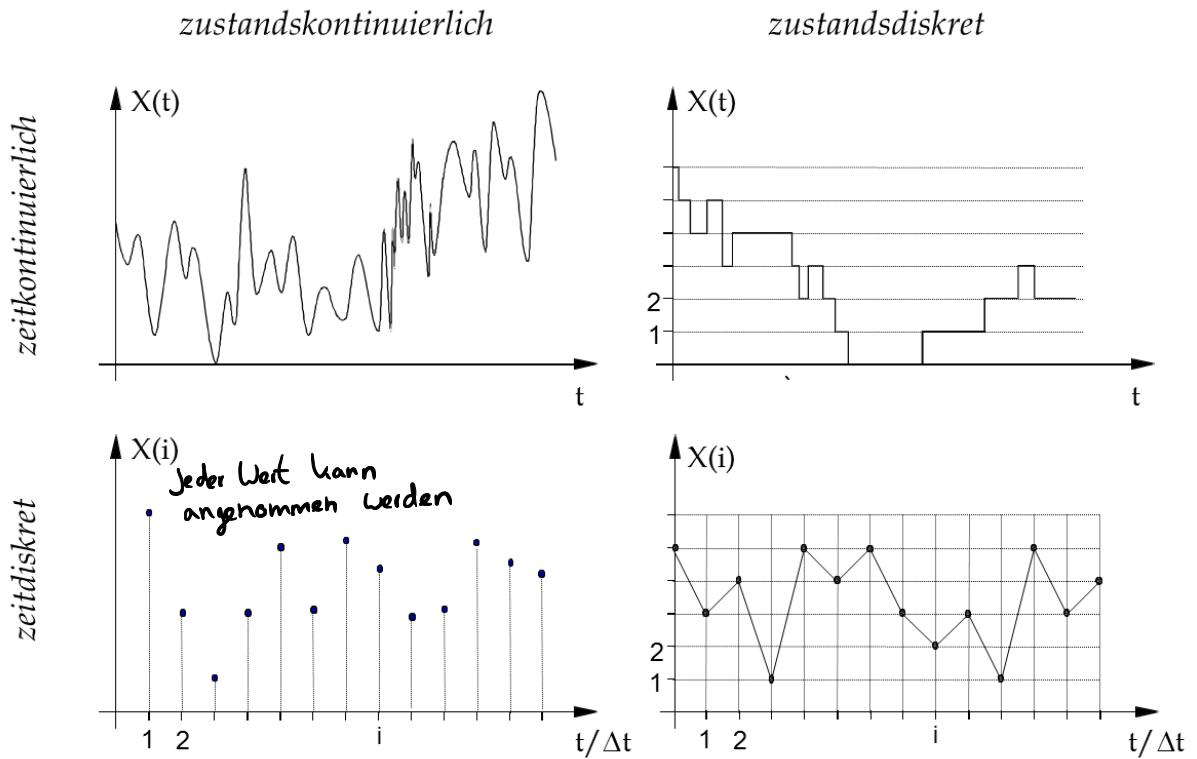
$$\begin{aligned} \text{VAR}[a \cdot X + b \cdot Y] &= E[(a \cdot X + b \cdot Y)^2] - E[(a \cdot X + b \cdot Y)]^2 = \\ &= a^2 \cdot E[X^2] + 2 \cdot a \cdot b \cdot E[X \cdot Y] + b^2 \cdot E[Y^2] - (a \cdot E[X] + b \cdot E[Y])^2 = \\ &= a^2 \cdot E[X^2] + 2 \cdot a \cdot b \cdot E[X \cdot Y] + b^2 \cdot E[Y^2] - a^2 \cdot E[X]^2 - 2 \cdot a \cdot b \cdot E[X] \cdot E[Y] - b^2 \cdot E[Y]^2 = \\ &= a^2 \cdot \{E[X^2] - E[X]^2\} + 2 \cdot a \cdot b \cdot \{E[X \cdot Y] - E[X] \cdot E[Y]\} + b^2 \cdot \{E[Y^2] - E[Y]^2\} = \\ &= a^2 \cdot \text{VAR}[X] + 2 \cdot a \cdot b \cdot \text{COV}[X, Y] + b^2 \cdot \text{VAR}[Y] \end{aligned}$$

- Anwendung auf Vergleich zweier Systeme mittels $Z = X - Y$:
 $\text{VAR}[Z] = \text{VAR}[X] - 2 \cdot \text{COV}[X, Y] + \text{VAR}[Y]$
 - Verringerung der Varianz von Z durch positive Korrelation von X und Y
- Berechnung des Konfidenzintervalls: $Z \pm t_{n-1, 1-\frac{\alpha}{2}} \cdot \sqrt{\frac{S^2(n)[Z]}{n}}$
 - Reduktion der empirischen Varianz $S^2(n)[Z]$ von Z führt zu kleinerer Anzahl benötigter Stichproben n um Konfidenzintervalle mit gewünschter Maximalgröße zu erreichen
 - Daraus ergibt sich die Simulationsbeschleunigung

7 Stochastische Prozesse

7.1 Definition

- $\{t, X_t\}$ oder $\{t, X(t)\}$ ist ein stochastischer Prozess, wobei $t \in \mathcal{T}$ (Indexmenge (Zeit), i.d.R. $\mathbb{N}_0, \mathbb{R}^+$) und $X(t) \in \mathcal{S}$ (Zustandsmenge).
 - Beispiel: $(t_0, x_0), (t_1, x_1), (t_2, x_2), \dots$ mit $x_i \in \mathcal{S}$
- Die Indexmenge kann sein
 - zeitkontinuierlich oder zeitdiskret
- Der Zustandsraum kann sein
 - wertkontinuierlich oder wertdiskret
 - endlich oder unendlich
 - ein- oder mehrdimensional
- Das Verhalten des Prozesses kann evtl. durch stochastische Gleichungen beschrieben werden.
- Beispiele



7.2 Beispiele von Prozessen

7.2.1 Autoregressiver Prozess (AR)

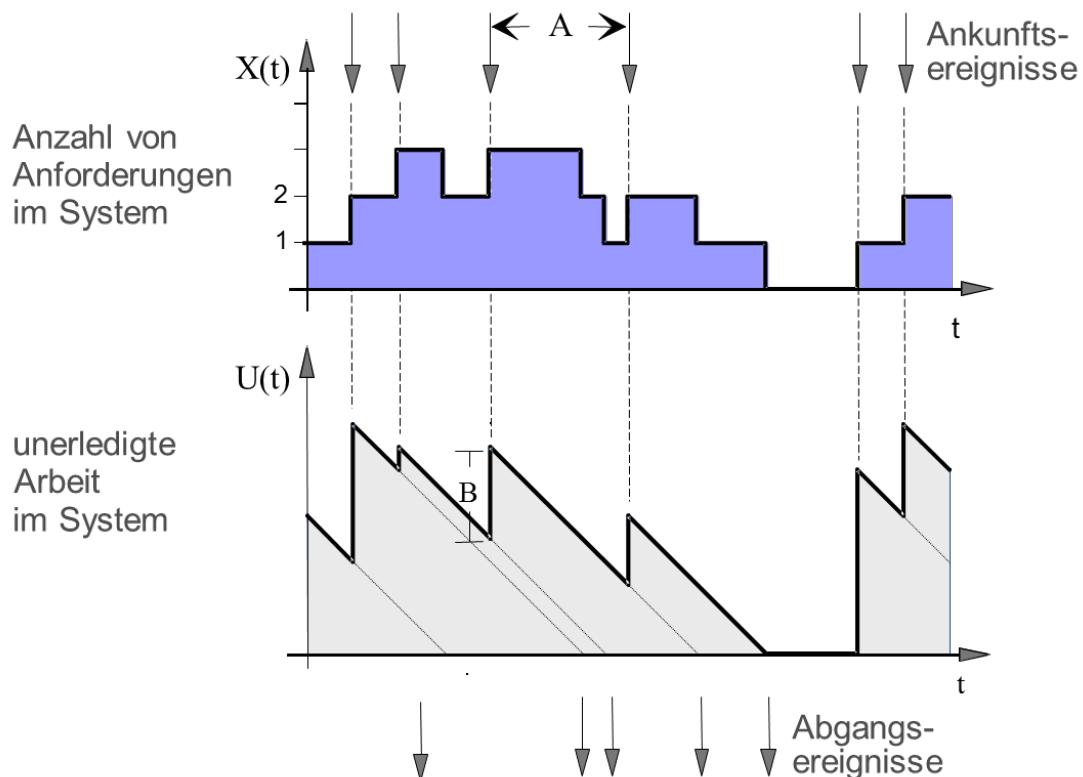
$$\text{Erzeugungsvorschrift: } X_i = \mu + \left(\sum_{0 < k \leq p} \phi_k \cdot (X_{i-k} - \mu) \right) + \varepsilon_i$$

- p : Ordnung des Prozesses (=Anzahl berücksichtigter Vorgänger-ZVs)
- μ : Mittelwert der X_i
- ε_i : normalverteilte Zufallsvariable mit Mittelwert 0
- ϕ_k : Gewichtungskoeffizienten mit $\sum_{0 < k \leq p} \phi_k = 1$

Es gibt noch viele andere, ähnliche stochastische Prozesse. Literatur dazu findet man unter dem Stichwort „Zeitreihen“ (R. Schlittgen, B. Streitberg: Zeitreihenanalyse, Oldenbourg, 1995)

7.2.2 Zustandsprozess

- Zustand des Prozesses beschreibt Zustand eines realen Systems
- Weitere Entwicklung des Prozesses evtl. vom Systemzustand abhängig
- Oft unterschiedliche Beschreibungen des gleichen Systems mit unterschiedlichen Eigenschaften (siehe Markov-Prozess).



7.2.3 Punktprozess

- Endliche oder abzählbar unendliche Folge von zufälligen Zeitpunkten bzw. Ereignissen auf der reellen Zeitachse
- Abstand zwischen t_i und t_{i+1} ist beschrieben durch $A_i = t_{i+1} - t_i$ (Zwischenankunftszeit)
- A_i müssen nicht durch eine einzige Verteilungsfunktion $A(t)$ beschrieben werden können
- Zeitkontinuierlich
 - Einzelankünfte
- Zeitdiskret
 - Einzelankünfte falls $P(A=0)=0$
 - Gruppenankünfte falls $P(A=0)>0$

7.2.4 Erneuerungsprozess

- Spezialfall eines Punktprozesses
- Unabhängig und identisch verteilte Zwischenankunftszeiten (independently and identically distributed, iid)
- Falls $P(A=0)>0$, dann geometrisch verteilte Gruppengröße; warum?

7.2.5 Ankunftsprozess

- Markiert Zeitpunkte von Ankünften (z.B. Anforderungen in einem System)
- Es kann Einzel- oder auch Gruppenankünfte geben
- Die Größe der Gruppe kann zufällig sein

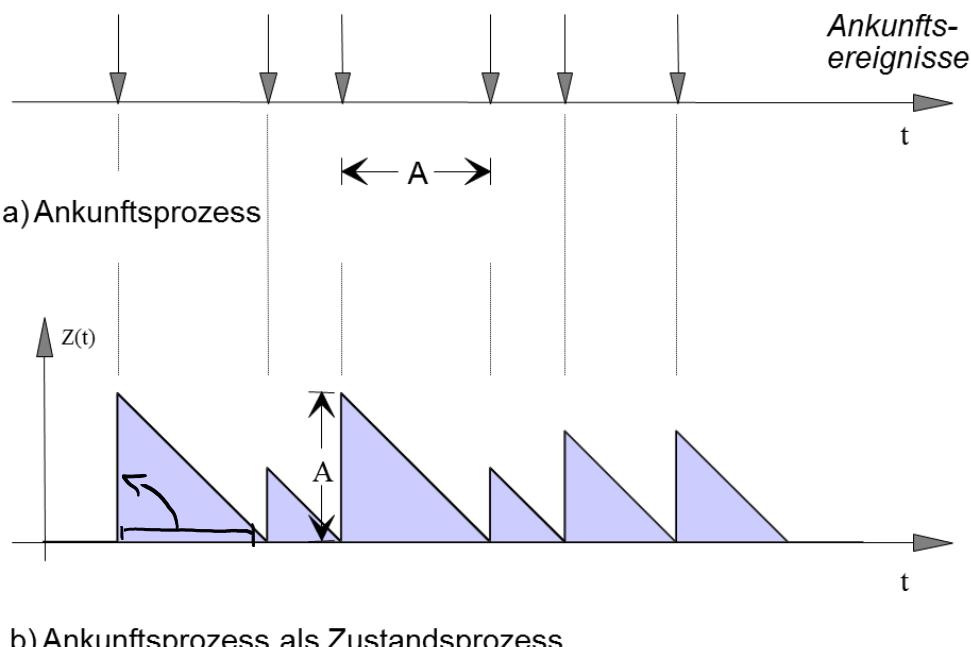
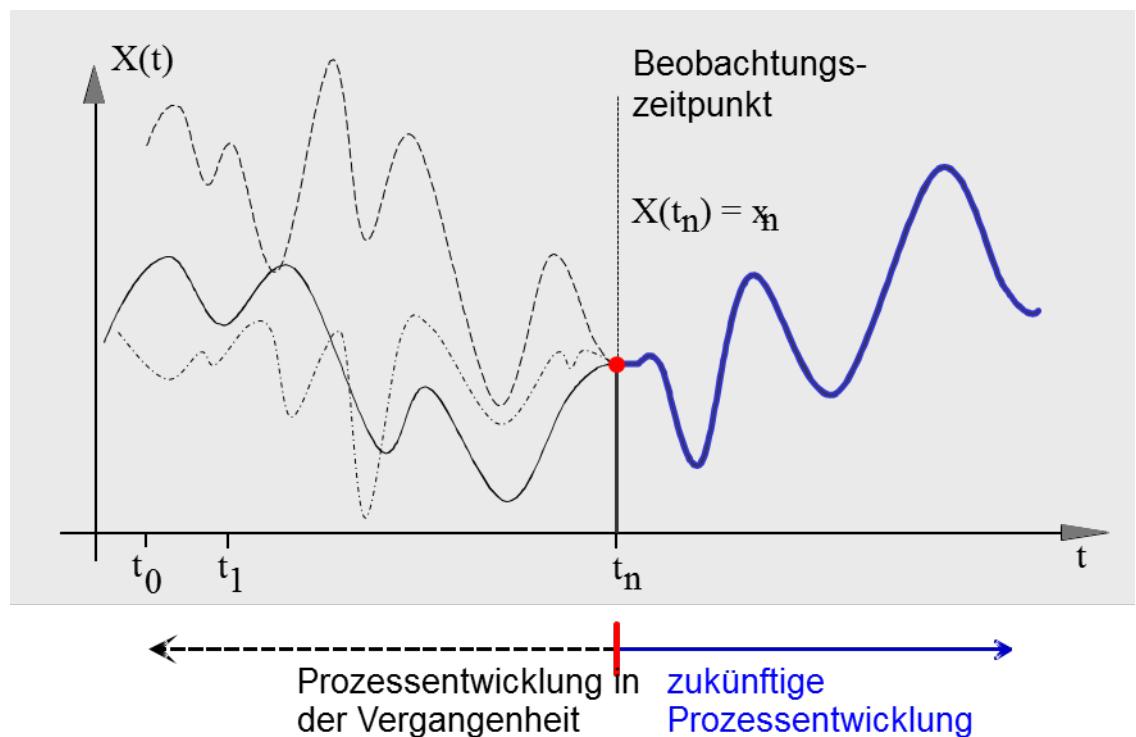


Abbildung 1: Ankunftsprozess als Punktprozess und als Zustandsprozess.

7.2.6 Markov-Prozess

- Markov-Eigenschaft (Gedächtnislosigkeit): das zukünftige Verhalten des Prozesses hängt nur von seinem Zustand ab und insbesondere nicht davon, wie er sich in der Vergangenheit entwickelt hat.
- In Formeln
 - $P(X(t_{n+1}) = s_{n+1} | X(t_n) = s_n, X(t_{n-1}) = s_{n-1}, \dots, X(t_0) = s_0) = P(X(t_{n+1}) = s_{n+1} | X(t_n) = s_n)$ für zeitdiskrete Prozesse oder
 - $P(X(t + \Delta) = s_{t+\Delta} | X(t) = s_t, X(t - \theta) = s_{t-\theta}, \dots) = P(X(t + \Delta) = s_{t+\Delta} | X(t) = s_t)$ für zeitkontinuierliche Prozesse



- Beispiele für Prozesse ohne Markov-Eigenschaft
 - Autoregressiver Prozess
 - Ankunftsprozess ohne Prozesszustand (Ausnahme: Poisson-Prozess)
 - Zustandsprozess, der nur Anzahl der Anforderungen in einem System charakterisiert; die Restarbeit der Anforderung in der Bedienung als zusätzliche Information (zweidimensionaler Systemzustand!) würde das System wieder gedächtnislos machen.
- Beispiele für Markov-Prozesse
 - Ankunftsprozess als Zustandsprozess
 - Zustandsprozess, der unerledigte Arbeit im System charakterisiert

7.3 Rekurrenzzeit (TR 2.2.2)

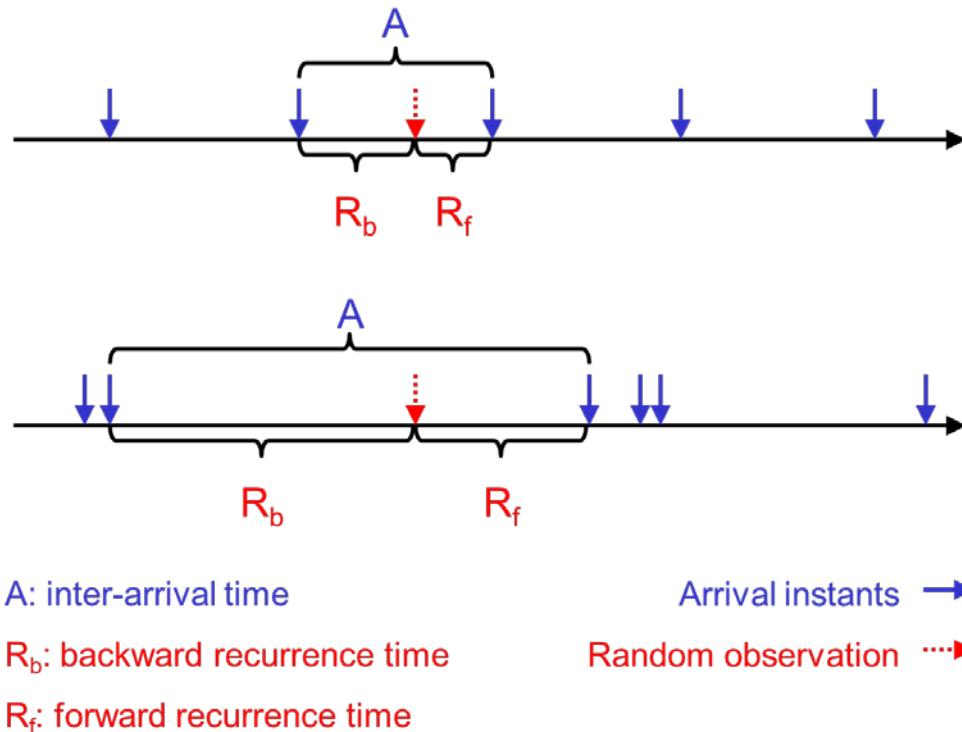


Abbildung 2: Vorwärts- und Rückwärtsrekurrenzzeit für deterministische und Pareto-verteilte Zwischenankunftszeiten

- Betrachte Erneuerungsprozess
- Wähle zufälligen Zeitpunkt (gleichwahrscheinlich auf jeder Position der Zeitachse)
- Vorwärtsrekurrenzzeit R_v : Zeitdauer bis zum Auftreten des nächsten Prozesspunktes
- Rückwärtsrekurrenzzeit R_r : Zeitdauer bis zum Auftreten des vorherigen Prozesspunktes

7.3.1 Rekurrenzzeit für zeitkontinuierliche Erneuerungsprozesse

- Feststellung: Vorwärts- und Rückwärtsrekurrenzzeit haben dieselben statistischen Eigenschaften bei zeitkontinuierlichen Erneuerungsprozessen. Darum sprechen wir nur noch von der Rekurrenzzeit.
- Wahrscheinlichkeitsdichte der Rekurrenzzeit: $r(t) = \frac{1}{E[A]} \cdot (1 - A(t))$
- Herleitung
 - Wahrscheinlichkeitsdichte für das Antreffen eines Intervalls der Länge τ : $q(\tau) = \frac{a(\tau) \cdot \tau}{E[A]}$
 - Proportional zur Länge des Intervalls
 - Proportional zur Wahrscheinlichkeitsdichte der Intervalllänge
 - Normierung mit $\frac{1}{E[A]}$, damit $\int_{-\infty}^{\infty} q(\tau) d\tau = 1$ erfüllt ist
 - Wahrscheinlichkeitsdichte für das Auftreten der Rückwärtsrekurrenzzeit t unter der Voraussetzung, dass das betrachtete Zwischenankunftsintervall τ lang ist: $r(t|A = \tau) = \begin{cases} 0 & t < 0 \\ \frac{1}{\tau} & 0 \leq t \leq \tau \\ 0 & t > \tau \end{cases}$
 - Anwendung des Gesetzes der totalen Wahrscheinlichkeit ergibt:

Sollte man verstehen, "nur für die Klausur"

$$r(t) = \int_{\tau=0}^{\infty} r(t|A = \tau) \cdot q(\tau) d\tau =$$

$$\begin{cases} \int_{\tau=0}^t 0 \cdot \frac{a(\tau) \cdot \tau}{E[A]} d\tau + & \text{für } \tau < t \\ \int_{\tau=t}^{\infty} \frac{1}{\tau} \cdot \frac{a(\tau) \cdot \tau}{E[A]} d\tau & \text{für } \tau \geq t \end{cases} = \frac{1}{E[A]} \cdot (1 - A(t))$$

- Momente der Rekurrenzzeit: $E[R^k] = \frac{E[A^{k+1}]}{(k+1) \cdot E[A]}$
- Mittelwert $E[R] = \frac{E[A^2]}{2 \cdot E[A]} = \frac{E[A]}{2} \cdot ((c_{var}[A])^2 + 1)$
- Daraus folgt für $c_{var}[A] < 1$: $E[R] < E[A]$
- $c_{var}[A] > 1$: $E[R] > E[A]$
 - Das zweite ist schwer vorstellbar.
 - ABER: weil besonders lange Zwischenankunftsintervalle mit besonders hoher Wahrscheinlichkeit angetroffen werden, bekommt man häufig Rekurrenzzeiten, die größer sind als der Mittelwert der Zwischenankunftszeit.
- Die Exponentialverteilung ist die einzige zeitkontinuierliche Verteilung mit $R(t) = A(t)$: $r(t) = \frac{1}{E[A]} \cdot (1 - A(t)) = \lambda \cdot (1 - (1 - e^{-\lambda \cdot t})) = \lambda \cdot e^{-\lambda \cdot t} = a(t)$

7.3.2 Poisson-Prozess

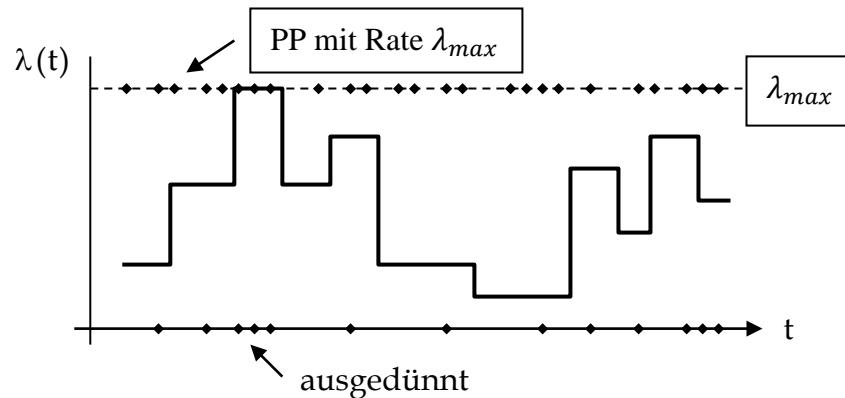
- Erneuerungsprozess mit exponentiell verteilten Zwischenankunftszeiten

$$(A(t) = 1 - e^{-\lambda \cdot t})$$
- Erzeugung: $U \sim U(0,1)$, $t_i \leftarrow t_{i-1} - \ln U / \lambda$
- Wegen des obigen Phänomens bedeutet das, dass beim Poisson-Prozess zu jedem Zeitpunkt die Zeit bis zur nächsten „Ankunft“ wieder exponentiell und sogar identisch verteilt ist.
 - Dafür ist es insbesondere egal, wann die letzte Ankunft stattfand.
 - Darum ist der Poisson-Prozess gedächtnislos und besitzt somit die Markov-Eigenschaft.
 - Aus diesem Grund wird die exponentielle Verteilung auch oft mit „M“ abgekürzt.
 - **Der Poisson-Prozess ist der einzige zeitkontinuierliche Erneuerungsprozess mit der Markov-Eigenschaft.**
- Anwendungsbeispiel: zu Simulationsbeginn seien n Server belegt und die Bedienzeiten exponentiell verteilt. Die verbleibende Zeit bis zum jeweiligen Bedienende kann wegen der $R=A$ als volle Bedienzeit initialisiert werden.

7.3.3 Sonderfall: Instationärer Poisson-Prozess (mit zeitabhängiger Rate)

- Der „normale“ Poisson-Prozess wird auch als stationär / homogen bezeichnet.
- Die Rate $\lambda(t)$ der exponentiellen Verteilungen, durch welche die Zwischenankunftszeiten des instationären Poisson-Prozesses erzeugt werden, ist von der Zeit abhängig.
 - \Rightarrow die Zwischenankunftszeiten des Prozesses sind nicht iid
 - \Rightarrow der Prozess ist kein Erneuerungsprozess
 - \Rightarrow der Prozess ist kein Poisson-Prozess im eigentlichen Sinne

- Realisierung durch Accept/Reject Methode
 - Auch „Ausdünnen“ oder Thinning genannt
 - Notwendige Annahme: $\lambda_{max} = \max_t \lambda(t) < \infty$
 - Erzeuge nächsten Ankunftszeitpunkt t mit Rate λ_{max} und akzeptiere ihn mit Wahrscheinlichkeit $\lambda(t)/\lambda_{max}$



- Algorithmus

```

 $t \leftarrow t_{i-1}$ 
Repeat
  Erzeuge  $U_1, U_2 \sim U(0,1)$ 
   $t \leftarrow t - \frac{\ln(U_1)}{\lambda_{max}}$ 
Until  $U_2 \leq \frac{\lambda(t)}{\lambda_{max}}$ 
 $t_i \leftarrow t$ 
  
```

- Der Algorithmus ist wegen der Gedächtnislosigkeit der exponentiellen Verteilung korrekt.
- Vorteil: einfaches Verfahren
- Nachteil: ineffizient, falls $\lambda(t)$ deutlich kleiner als λ_{max}

7.3.4 Rekurrenzzeit für zeitdiskrete Erneuerungsprozesse

- Im Gegensatz zu zeitkontinuierlichen Prozessen beschränkt man sich hier auf „ganzzahlige“ zufällige Zeitpunkte.
- Betrachtung der Rekurrenzzeit
 - Kurz vor möglichen diskreten Zeitpunkten
 - Kurz nach möglichen diskreten Zeitpunkten
- Was ist die Auswirkung auf die Rekurrenzzeit?

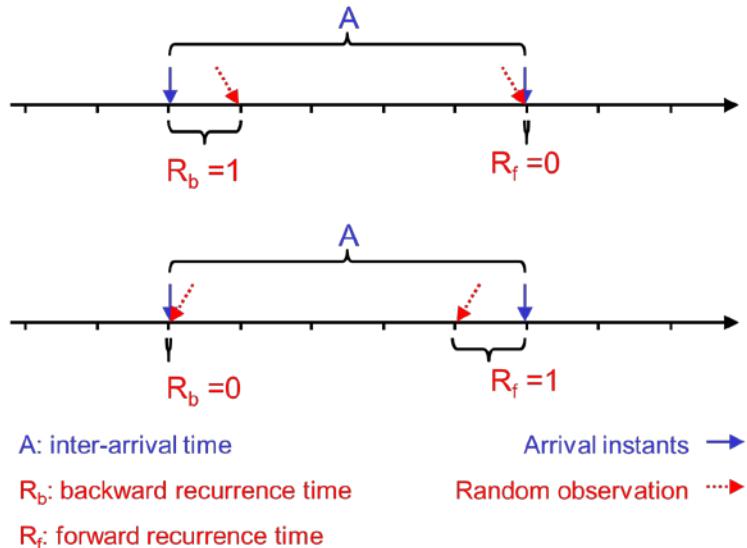


Abbildung 3: Vorwärts- und Rückwärtrekurrenzzeit für Beobachtungen kurz vor und kurz nach diskreten Zeitpunkten.

- Beobachtung kurz vor möglichen diskreten Zeitpunkten
 - Mittelwerte der Vorwärtsrekurrenzzeit R_v
 - Einziger zeitdiskreter Erneuerungsprozess mit Gedächtnislosigkeit ($r_v(k) = a(k)$)
 - Zwischenankunftszeiten, die einer nicht verschobenen Geometrischen Verteilung $a(k) = q^k \cdot (1 - q), k \geq 0$ folgen
 - $P(A=0)>0$
 - Gruppenankünfte möglich
- Beobachtung kurz nach möglichen diskreten Zeitpunkten
 - Mittelwerte der Vorwärtsrekurrenzzeit R_v
 - Einziger zeitdiskreter Erneuerungsprozess mit Gedächtnislosigkeit ($r_v(k) = a(k)$)
 - Zwischenankunftszeiten, die einer um 1 verschobenen Geometrischen Verteilung $a(k) = q^{k-1} \cdot (1 - q), k \geq 1$ folgen
 - $P(A=0)=0$
 - Nur Einzelankünfte

8 Discrete-Time Markov Chains (DTMCs) and Semi-Markov-Processes

Notation

\mathcal{X} : Wertebereich der ZV X

$P(X = i) = x[i]$: Wahrscheinlichkeit, dass die ZV X den Wert i annimmt.

$x = (x[0], \dots, x[n - 1])$: Wahrscheinlichkeiten für alle möglichen Werte der ZV X als Vektor

p_{ij} : Übergangswahrscheinlichkeit von Zustand i zu Zustand j einer Markovkette

X_i : i -te ZV einer Zeitreihe (in diesem Kapitel durch eine Markovkette erzeugt)

$x_i = (x_i[0], \dots, x_i[n - 1])$: Zustandsverteilung der Markovkette nach dem i -ten Übergang

8.1 Basics about DTMCs

8.1.1 Definition

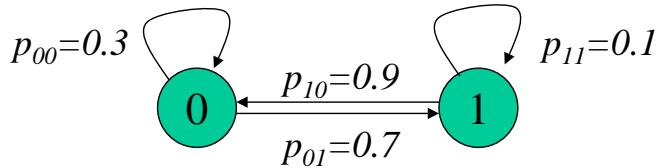


Abbildung 1: State transition graph or state transition diagram.

- Stochastic process
- States (e.g. 0 and 1)
- Start state: X_0
- State transition at certain times $t_i, i = 0, 1, 2, \dots$
- State after transition i : X_i
- Series of X_i forms a Markov chain
- Probability for transition from state i to state j : p_{ij}

8.1.2 Markov Chain Simulation

MCs can be simulated to calculate their average state distribution x .

- The simulation state is X
- Simulation calculates series X_n to find x
- Successor state $X_{n+1} = k$ for $X_n = i$ needs to be simulated
 - Random number $U \in (0; 1)$
 - Given $X_n = i$, $X_{n+1} = \min(k : \sum_{0 \leq j \leq k} p_{ij} \geq U)$

8.1.3 Sojourn time

- How long does the system remain in state j ?
 - $P(k \text{ steps}) = p_{jj}^{k-1} \cdot (1 - p_{jj})$
 - Distributed according to the *shifted geometric distribution*

8.1.4 Calculation of consecutive state distributions

- State transition matrix: $P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$
 - Stochastic matrix: sum over row is 1.
- State distribution after transition i : $x_i = (P(X_i = 0), P(X_i = 1))$
- $x_{i+1} = x_i \cdot P$ *
- Example
 - $P = \begin{pmatrix} 0.3 & 0.7 \\ 0.9 & 0.1 \end{pmatrix}$
 - $x_0 = (1, 0)$, $x_1 = (0.3, 0.7)$, $x_2 = (0.72, 0.28)$, $x_3 = (0.468, 0.532)$, $x_4 = (0.6192, 0.3808)$, $x_5 = (0.52848, 0.47152)$, ..., $x_{14} = (0.56284, 0.43716)$, $x_{15} = (0.56229, 0.43771)$, $x_{16} = (0.56262, 0.43738)$, ...

8.1.5 Average state distribution (Cesàro Limit)

- Average state distribution $x_s = \lim_{n \rightarrow \infty} \left(\frac{1}{n+1} \cdot \sum_{i=0}^n x_i \right)$
- Also called *stationary state distribution*
- Useful to compute performance measures
- x_s is left-hand eigenvector of P : $x_s = x_s \cdot P$
- If series of x_i converges, the resulting limit is equal to the average state distribution.

8.1.6 k -transition probabilities

- Probability that system changes after k steps from state i to state j : p_{ij}^k
- Corresponding transition matrix $P^{(k)}$ satisfies $x_{i+k} = x_i \cdot P^{(k)} \Rightarrow P^{(k)} = P^k$

8.1.7 Matrix powering

- $P = \begin{pmatrix} 0.3 & 0.7 \\ 0.9 & 0.1 \end{pmatrix}$, $P^2 = \begin{pmatrix} 0.72 & 0.28 \\ 0.36 & 0.64 \end{pmatrix}$, $P^4 = \begin{pmatrix} 0.6192 & 0.3808 \\ 0.4896 & 0.5104 \end{pmatrix}$, $P^8 = \begin{pmatrix} 0.56985 & 0.43015 \\ 0.55305 & 0.44695 \end{pmatrix}$, $P^{16} = \begin{pmatrix} 0.56262 & 0.43738 \\ 0.56234 & 0.43766 \end{pmatrix}$, $P^{32} = \begin{pmatrix} 0.56250 & 0.43750 \\ 0.56250 & 0.43750 \end{pmatrix}$
- Matrix powering may quickly approximate $P^\infty = \begin{pmatrix} x_s \\ x_s \end{pmatrix}$
 - Efficient for small matrices

* Disp.:

$$\underbrace{P(X_i=0), P(X_i=1)}_{\text{Zustände}} \cdot \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} \text{Zustand } 0 \\ \text{nach } 0 \end{pmatrix} = \underbrace{(P(X_i=0) \cdot p_{00} + P(X_i=1) \cdot p_{10})}_{\dots} + \underbrace{\dots}_{\text{nach } 1}$$

8.2 Examples for Discrete-Time Markov Chains

8.2.1 The Weather in Belfast

- The description in [Stewart] of the daily weather changes in Belfast (Northern Ireland) illustrates nicely the concept of a MC.
- [Stewart] Stewart, W.J.: Introduction to the Numerical Solution of Markov Chains. 1 ed. Princeton University Press, Princeton (1994)
- The state is given by the weather: rainy (1), cloudy (2), and sunny (3). The state transition probabilities can be retrieved from empirical data and are given by the following state transition matrix:

$$P = \begin{pmatrix} 0.8 & 0.15 & 0.05 \\ 0.7 & 0.2 & 0.1 \\ 0.5 & 0.3 & 0.2 \end{pmatrix}.$$

On a rainy day, the probability that tomorrow is rainy again is 80%.

- The average state distribution can be computed:

$$P^\infty = \begin{pmatrix} 0.7625 & 0.1688 & 0.0688 \\ 0.7625 & 0.1688 & 0.0688 \\ 0.7625 & 0.1688 & 0.0688 \end{pmatrix}$$

It reveals that the probability for a rainy day in Belfast is 76.25%, for a cloudy day it is 16.88%, and the sun shines with a probability of 6.88%.

8.2.2 Analysis of Monopoly Using Markov Chains

Question: what is the visiting probability for fields in Monopoly?

Model

- Each field corresponds to a state
- Transition probability to other fields is influenced by two dices and event cards
- Set up state transition matrix
- Stationary state distribution may be taken as visiting probability for fields under the assumption that sufficiently many rounds are played.

See: Jörg Bewersdorff: Glück, Logik und Bluff: Mathematik im Spiel – Methoden, Ergebnisse und Grenzen, Vieweg Verlag, 3. Auflage, 2003, Kapitel 1.16

8.2.3 Analysis of Web Pages: Google's PageRank Algorithm

Question: how important is a web site?

Answer: depends on how often the web page is referenced by other pages and how important the referencing web pages are!

Model

- Each web page w is a state
- Let n_w be the number of links of a web page w ; then the state transition probability to their referenced web pages is set to $1/n_w$
- The stationary state distribution yields an estimate of the importance of the web pages.

This idea served in the early days of Google (~ 1999) to return most important web pages first. This idea led to the foundation of Google. Nowadays Google uses more complex, proprietary algorithms. Ideas of PageRank may still be a part of them.

See:

- <https://de.wikipedia.org/wiki/PageRank>
- Sergei Brin, Lawrence Page: *The Anatomy of a Large-Scale Hypertextual Web Search Engine*. In: Computer Networks and ISDN Systems, Band 30, 1998, S. 107–117, <http://infolab.stanford.edu/~backrub/google.html>

8.3 Where do we find Markov chains?

8.3.1 Discrete-Time Markov Chains (DTMCs)

- Discrete-time stochastic process
- Transitions according to a transition matrix
 - Sequence X_i forms a Markov chain
- Transition points only at integer values, i.e., $t_i = i$
 - This leads to geometrically distributed sojourn times (in terms of time, not only in terms of transition steps) for every state j . More specifically: distributed according to the *shifted* geometric distribution
 - A DTMC has the Markov property, i.e., it is “memoryless”.
 - Its evolution depends only on its current state, but not on its past, at any discrete observation point.

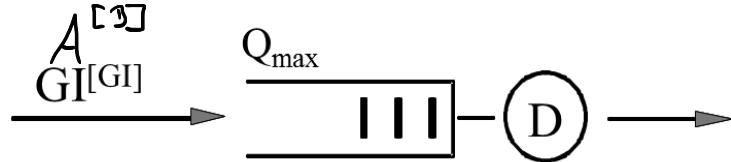
8.3.2 Semi-Markov Processes (SMPs) with Embedded Markov Chains

- Continuous- or discrete-time stochastic process
- Discrete states
- State transitions according a state transition matrix
- Time between state transitions can be general
 - Process is memoryless only shortly before or after transitions, but not between transitions.
 - These are embedded points whose system states form a Markov chain.
 - Transitions from a state to itself are still possible.
- Sojourn times in states do not need to be geometrically distributed because time between transitions may vary.
 - But duration in a state in terms of transition steps follows again a *shifted* geometric distribution.

8.4 Specification and Analysis of Embedded Markov Chains Using Recursive Stochastic Equations

Objective: alternative description of Markov chains and alternative calculation of consecutive state distributions x_i

8.4.1 Example: $GI^{[GI]}/D/1 - Q_{max}$ Queue



Description

- Queue with single server and deterministic service time D
- Finite buffer for at most Q_{max} waiting customers (bytes)
- Inter-arrival time A between customer arrivals
- Batches of B customers arrive *1 oder mehrere auf einmal.*
- A and B are iid random variables

Objective of analysis

- Blocking probability for customers
- Waiting time distribution for customers

Analysis

- Time discretized in time units \Rightarrow discrete-time stochastic process
 - A and D given in multiples of time units
- System state U : unfinished work (time units) in server and queue
 - $U_{max} = D \cdot (1 + Q_{max})$, $0 \leq U \leq U_{max}$ *1 in D + Q_{max}*
- Wanted performance measure
 - Distribution of unfinished work U shortly before arrivals
 - \Rightarrow Embedding of observation points t_i for U_i shortly before arrivals
- Series of U_i yields an embedded Markov chain, process is a SMP
- When batch of B customers arrives, clients that still fit into the queue are accepted, others are rejected.
- The following is an approximation since fractional users are accepted:
 - New unfinished work: $U' = \min(U_i + D \cdot B, U_{max})$
- During an inter-arrival time A , unfinished work is reduced by A time units, but it cannot fall below zero:
 - New unfinished work:

$$U_{i+1} = \max(U' - A, 0) = \max(\min(U_i + D \cdot B, U_{max}) - A, 0)$$
- Full model description requires concrete values
 - System starts with no unfinished work ($U_0 = 0$)
 - $Q_{max} = 5$
 - Distributions for random variables A and B

A=i	2	3	4		B=i	1	2	3
P(A=i) (a[i])	0.2	0.6	0.2		P(B=i) (b[i])	0.6	0.3	0.1

8.4.2 Formalization of the Functional Description

- Embedded points \mathcal{T} (for the SMP)
 - Example: time instants t_i shortly before customer arrivals
 - Choice determines the model
- State X
 - Example: unfinished work $X = (U)$
 - State space $\mathcal{X} = [0; U_{max}]$
- Initial state distribution x_0
 - Example: start state $X_0 = 0$
 - $x[0] = 1, x[i \neq 0] = 0$
- Factor Y and its distribution
 - Influences system behavior
 - Not part of system state
 - Factor space \mathcal{Y} and distribution y
 - Example: $Y = (B, A)$
 - Size of arriving batches B and inter-arrival times A
 - $\mathcal{Y} = (\mathcal{B} \times \mathcal{A})$
 - B and A are both iid $\Rightarrow P(Y = (i, j)) = b[i] \cdot a[j]$ determines y
- State transition function $f: (X \times Y) \rightarrow X$
 - Recursive stochastic equation
 - Describes system behavior
 - Example: $U_{i+1} = f(U_i, (B, A)) = \max(\min(U_i + B, U_{max}) - A, 0)$
- Full specification of MC given by $\mathcal{D} = (\mathcal{T}, \mathcal{X}, x_0, \mathcal{Y}, y, f)$
 - Both state space \mathcal{X} and factor space \mathcal{Y} may be multi-dimensional.
 - Intuitive description of system behavior at embedded points

See: M. Menth: "Description and Analysis of Markov Chains Based on Recursive Stochastic Equations and Factor Distributions"

<http://www.worldacademicunion.com/journal/1746-7233WJMS/wjmsvol07no01paper01.pdf>

8.4.3 Algorithmic Calculation of Consecutive State Distributions

8.4.3.1 Backward Algorithm for Computation of Successor Distribution x_{n+1}

- The recursive stochastic equation and the distribution of the factors allow calculation of the state probabilities at consecutive embedded points by applying the law of total probability:

$$x_{n+1}[k] = \sum_{i \in \mathcal{X}, j \in \mathcal{Y}} P(X_{n+1} = k | X_n = i \wedge Y = j) \cdot x_n[i] \cdot y[j]$$

- The conditional probability can be computed using the state transition function:

$$P(X_{n+1} = k | X_n = i \wedge Y = j) = \begin{cases} 0 & \text{if } f(i, j) \neq k \\ 1 & \text{if } f(i, j) = k \end{cases}$$

- Together with the preimage $Z(k) = \{(i, j) : f(i, j) = k, i \in \mathcal{X}, j \in \mathcal{Y}\}$ of state k with respect to f , this equation simplifies the above expression to

$$x_{n+1}[k] = \sum_{(i,j) \in Z(k)} x_n[i] \cdot y[j] \quad (\text{backward equation})$$

- Backward algorithm requires calculation of preimage
 - Difficult
 - Limits tractability of complex models

8.4.3.2 Forward Algorithm for Computation of Successor Distribution x_{n+1}

Rearrangement of computation steps leads to “forward algorithm”

- Numerically equivalent to “backward approach”

input: state distribution x_n and factor distribution y

 initialize x_{n+1} with zeros

for all $i \in \mathcal{X}$ **do**

for all $j \in \mathcal{Y}$ **do**

$x_{n+1}[f(i, j)] = x_{n+1}[f(i, j)] + x_n[i] \cdot y[j]$

end for

end for

output: x_{n+1}

Advantages of “forward approach” vs. “backward approach”

- Does not require calculation of preimage
- Program for calculation of x_{n+1} can be syntactically derived from functional description \mathcal{D} .

Advantages of “forward approach” vs. vector-matrix-multiplication $x_{i+1} = x_i \cdot P$

- State transition matrix P not needed
 - Can be very large for large state space in spite of sparse matrix representation, storage problem
- Multiple optimization possibilities speed up computation time
 - E.g. decomposition of state transition function: $f = f_0 \circ f_1$ with additional embedded points, states, and different factors
 - See paper

8.4.4 Derivation of the State Transition Matrix P

8.4.4.1 Backward Algorithm for Computation of State Transition Matrix P

The state transition matrix P for a Markov model can be computed by

$$p_{ik} = \sum_{\{j \in \mathcal{Y}: f(i,j)=k\}} y[j]$$

- Equation corresponds to the backward algorithm
- Rearrangement to forward algorithm possible

8.4.4.2 Forward Algorithm for Computation of State Transition Matrix P

```

input: factor distribution  $y$ 
      initialize  $P$  with zeros
      for all  $i \in \mathcal{X}$  do
        for all  $j \in \mathcal{Y}$  do
           $p_{i,f(i,j)} = p_{i,f(i,j)} + y[j]$ 
        end for
      end for
output:  $P$ 

```

One can show that the functional description \mathcal{D} and the state transition matrix P have the same expressiveness (see paper).

8.4.5 MC Simulation Based on the Functional Description \mathcal{D}

- Use random number U to realize random variable for factor Y
- $X_{n+1} = f(X_n, Y)$

8.5 Use of $GI^{[GI]}/D/1 - Q_{max}$ for Approximation of Waiting Time in $GI/GI/1 - \infty$

- Difference between $GI/GI/1 - \infty$ and $GI^{[GI]}/D/1 - Q_{max}$
 - Arrival process: single vs. batch arrivals
 - Service process: GI process vs. D process
 - Queue: infinite vs. limited
- Why approximation works
 - Service time for batches of customers in $GI^{[GI]}/D/1 - Q_{max}$ equals service time of single customer in $GI/GI/1 - \infty$
 \Rightarrow Queue length distribution of $GI^{[GI]}/D/1 - Q_{max}$ approximates waiting time distribution of $GI/GI/1 - \infty$
 - Loss occurs only with limited queue
 - Prerequisite for approximation: large Q_{max} to keep loss small in $GI^{[GI]}/D/1 - Q_{max}$
 - Problem: what to do when arriving batch does not fully fit in queue?
 - Fill queue up to Q_{max} ; this best approximates the evolution of the queue occupancy of an infinite queue.
 - This is also in line with the state transition function for $GI^{[GI]}/D/1 - Q_{max}$ presented in 8.4.1.
- Configuration of $GI^{[GI]}/D/1 - Q_{max}$ for approximation
 - Choose discrete time unit u as deterministic service time
 - Discretize distributions for inter-arrival times and service times for batches in multiples of u
 - Possibly truncate and normalize the distributions
 - Make sure that mean and coefficient of variation do not change significantly
 - Choose Q_{max} large enough so that stationary probabilities for states near Q_{max} are sufficiently small.
 - Results of the analysis
 - Stationary distribution of the queue occupation Q which translates into
 - Waiting time of arriving batches in $GI^{[GI]}/D/1 - Q_{max}$
 - Waiting time of customers in $GI/GI/1 - \infty$
 - Distribution may be used to calculate
 - Distribution function
 - Mean value
 - Variance
 - Standard deviation
 - Coefficient of variation

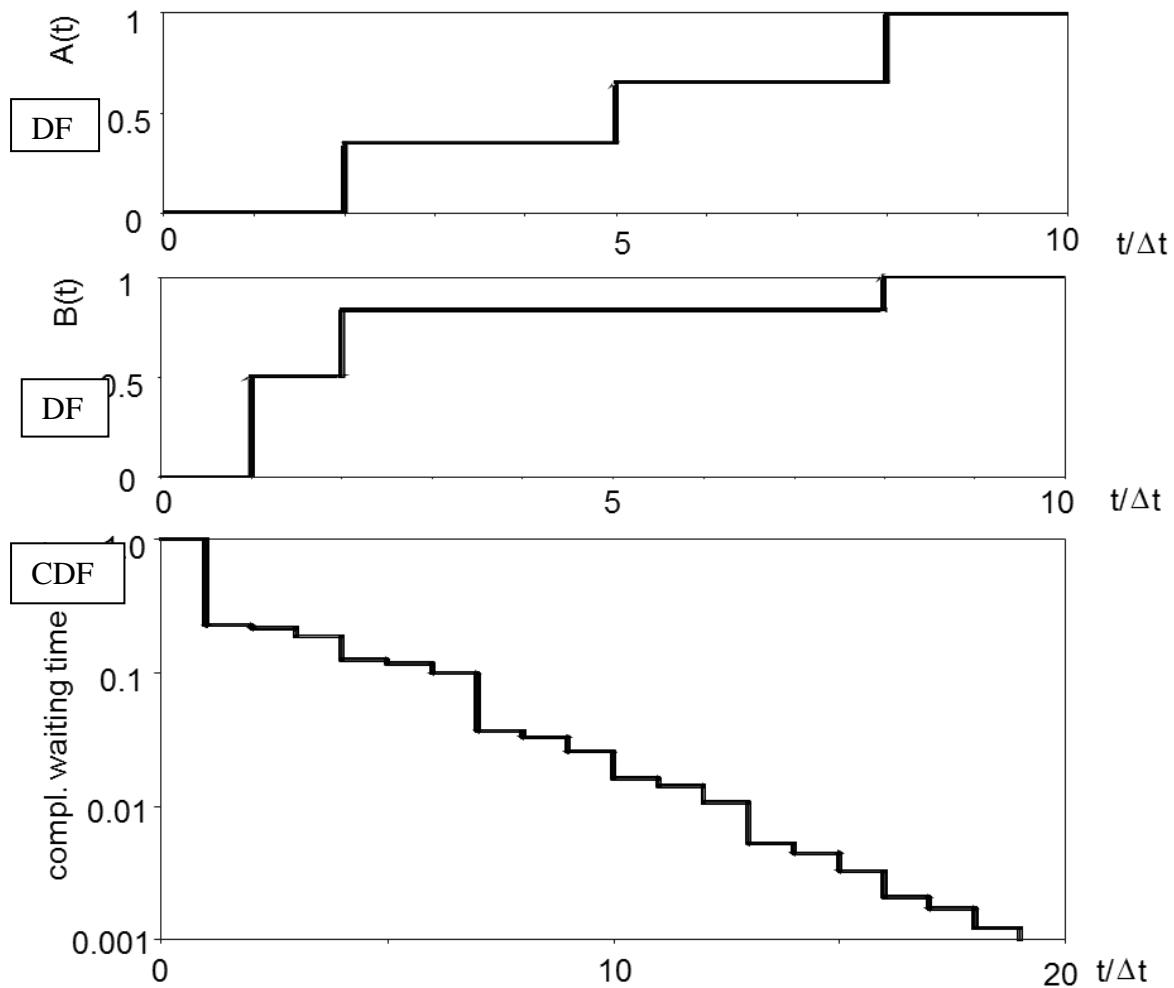


Abbildung 2: Distribution functions $A(t)$ and $B(t)$ are discretized. Therefore, the resulting complementary waiting time distribution function is also a step function. When the discretization is fine enough, steps are no longer visible.

Note: the complementary distribution function is more appropriate for presentation than the distribution function.

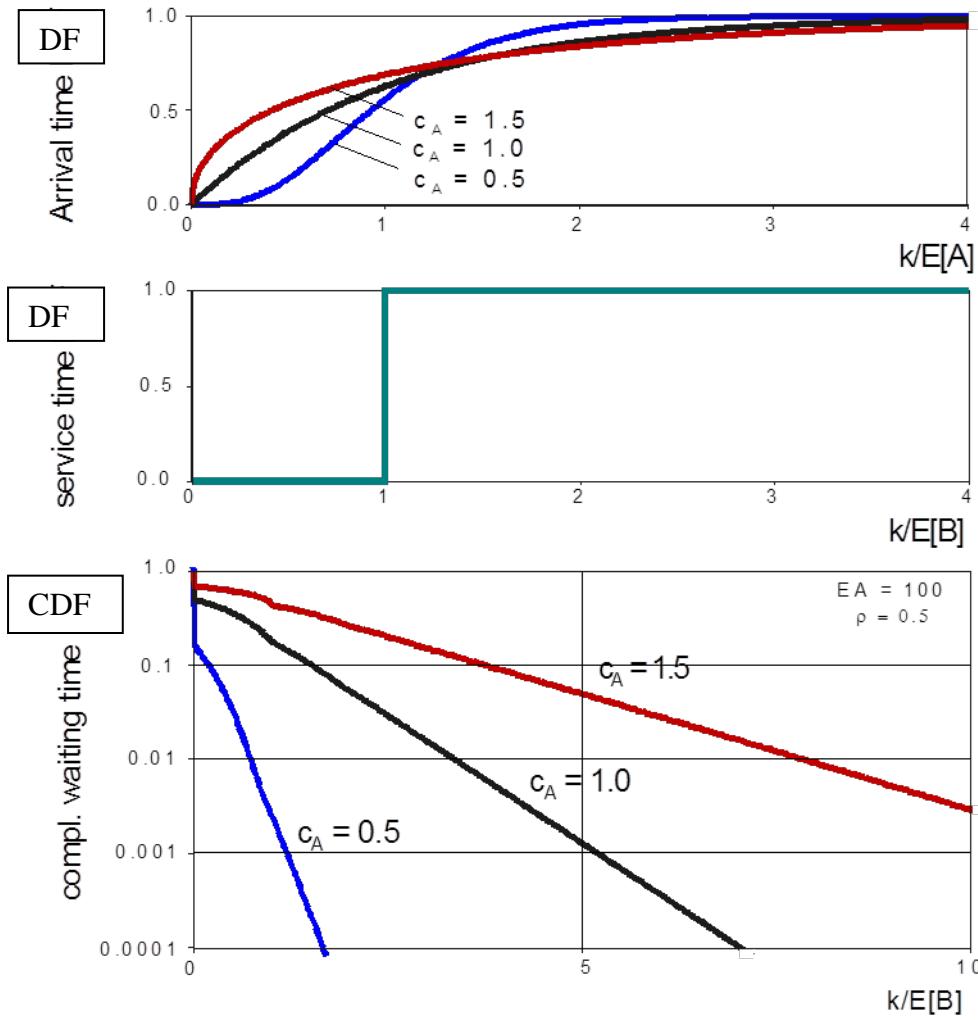


Abbildung 3: Analysis of a $G/D/1-\infty$ waiting system. More variable inter-arrival times lead to larger waiting probabilities and longer waiting times.

Note: the complementary distribution function is more appropriate for presentation than the distribution function.

8.6 Traffic Descriptors, Token Bucket, Markers, Policers, and Shapers

8.6.1 Traffic Descriptor

- Describes (upper bound of) intensity and variability of traffic: traffic rates, maximum packet sizes, maximum burst sizes, possibly on different time scales.
- Token bucket is a simple but widespread traffic descriptor.
- Combinations and variations of token buckets and other approaches exist.

8.6.2 Token Bucket

- Has a bucket that may hold at most S tokens (e.g., bytes)
- Current fill state of bucket is F .
- Bucket is continuously filled with rate R (e.g. bytes/s).

UpdateBucket()

- $t = \text{now} - \text{lastFill}$
- $F = \min(F + t * R, S)$
- $\text{lastFill} = \text{now}$

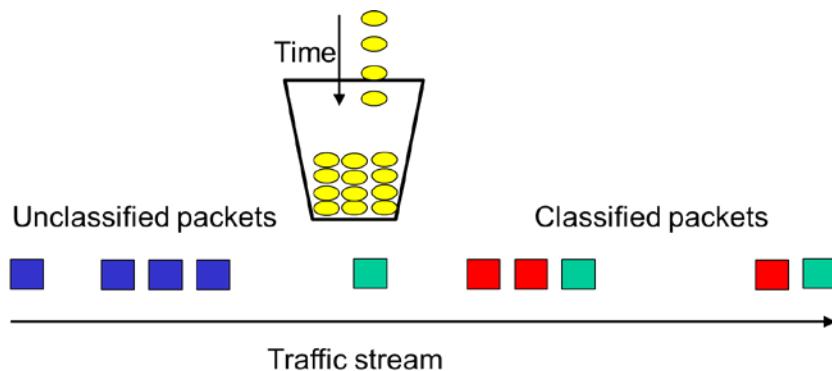


Abbildung 4: Token bucket based traffic meter and classifier.

- Packets of a traffic stream are passed to a token bucket meter that decides whether a packet of size B is in/out-of-profile.

Check(PacketSize B)

- UpdateBucket()
- If ($B \leq F$) then
 - $F = F - B$
 - Return in-profile
- Else
 - Return out-of-profile

8.6.3 *Traffic Marker*

- Some nodes in the Internet mark some packets when overloaded to request rate reduction from source (explicit congestion notification, ECN) [RFC3168].
- Threshold marking marks all packets of a traffic stream as long as the traffic stream exceeds a token bucket descriptor [RFC5670]; marked packets are used to detect “pre-congestion” and to stop admission of new flows (pre-congestion notification, PCN).
- Excess traffic marking marks some packets of a traffic stream, namely those that exceed a token bucket traffic descriptor [RFC5670]; unmarked and marked packets are used to quantify overload and determine the amount of traffic that may be removed from the network.

8.6.4 *Policer*

- Drops packets of a traffic stream that are out-of-profile with regard to a traffic descriptor.

8.6.5 *Spacer*

- Delays packets of a traffic stream that are out-of-profile with regard to a traffic descriptor.
- Uses a buffer to delay packets.
- Packets may be lost if buffer overflows.

8.7 Analysis of a Spacer Using $GI^{[GI]}/D/1-Q_{max}$

We assume a spacer that enforces a maximum *rate of R (bytes/s)* in communication networks by buffering packets between a station and a network. The *buffer is S bytes large*. The spacer makes sure that if a packet of size B (bytes) is sent, the time until the next packet is sent is at least B/R.

- Model of the spacer
 - The bucket of a spacer is modelled by unfinished work U
 - Measured in time
 - Describes time to wait until a newly arrived packet can be sent
 - $U_{max} = S/R$
 - Change of unfinished work at packet arrivals
 - If $(U+B/R \leq U_{max})$
 - $U' = U + B/R$ // accept packet, packet will be sent after U
 - Else
 - $U' = U$ // drop packet
 - Endif
 - Change of unfinished work during a packet inter-arrival time A (time)
 - $U'' = \max(U' - A, 0)$
- This can be modeled by a modified discrete-time $GI^{[GI]}/D/1 - Q_{max}$
 - The queue models the bucket
 - An arriving batch corresponds to an arriving packet of B bytes
 - All bytes of a packet are accepted or dropped
 - Modification of the state transition function presented for $GI^{[GI]}/D/1 - Q_{max}$ in 8.4.1 is needed:
 - $U_{max} = S/R$
 - State transition function


```
If  $U_n + \frac{B}{R} \leq U_{max}$ 
                   $U' = U_n + B/R$ 
              Else
                   $U' = U_n$ 
              Endif
               $U_{n+1} = \max(U' - A, 0)$ 
```
- The resulting stationary state distribution indicates the fill state of the spacer at packet arrivals so that the spacing delay and the packet loss due to the spacer can be predicted.

- Example
 - An ATM Virtual Connection (VC) requires a minimum time T between consecutive ATM cells (each of size 53 bytes).
 - The cell stream needs to be spaced before entering the network.
 - The spacer has a buffer for τ_{max} cells (buffer size S) and enforces a maximum rate of 1 cell within T time, i.e., $R=1 \text{ cell}/T$.
 - The distribution of inter-arrival times between cells is different before entering and after leaving the spacer
 - The time between cells in the departure process starts only with T .
 - The spacing result depends on the buffer size, analysis for
 - $\tau_{max} = 60$ cells
 - $\tau_{max} = 0$ cells
 - The spacer buffer may overflow \Rightarrow ATM cells may be lost, especially for small buffers

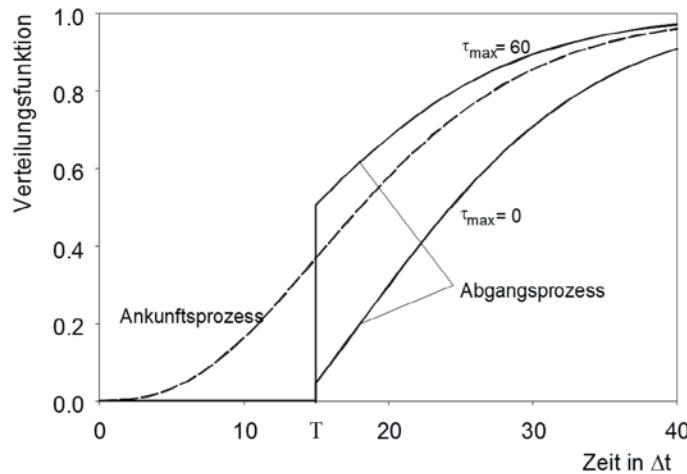


Abbildung 5: Distribution of inter-arrival time between cells before entering and after leaving a spacer.

8.8 Markov Chain Simulation

MCs can be simulated to calculate their average state distribution x .

- The simulation state is X
- Simulation calculates series X_n to find x
- Successor state $X_{n+1} = k$ for $X_n = i$ needs to be simulated

8.8.1 MC Simulation Based on the State Transition Matrix

- Random number $U \in (0; 1)$
- Given $X_n = i$, $X_{n+1} = \min(k : \sum_{0 \leq j \leq k} p_{ij} \geq U)$

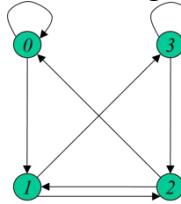
8.8.2 MC Simulation Based on the Functional Description \mathcal{D}

- Use random number U to realize random variable for factor Y
- $X_{n+1} = f(X_n, Y)$

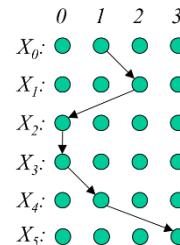
8.9 Comparison of Simulation and Calculation of Subsequent State Distributions for MC Analysis

8.9.1 Comparison of Operation

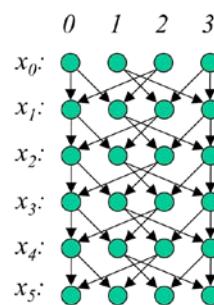
- Consider the following state transition graph



- MC simulation takes random walk through the transition graph and constructs a series $(X_n)_{0 \leq n \leq n_{max}}$
 - Single iteration step is quite cheap, but it reflects only a single transition.
 - Millions or billions of such simulation steps are usually needed to calculate sufficiently accurate average state probabilities x .



- Mathematical/algorithmic calculation of consecutive state distributions x_n
 - Takes into account all possible transition steps in a single iteration step
 - Probability mass of state distribution x_n is propagated from all states over all possible transitions to potential successor states which leads to the new state distribution x_{n+1} .
 - Here, a single iteration step requires a lot of computation effort, but usually a few tens or hundreds iteration steps are sufficient until x_n have converged with sufficient accuracy.



8.9.2 Comparison of Efficiency

Consider a convergence accuracy of $\|x_n - x\| \leq \varepsilon$. How many samples or iteration steps N are needed for sufficient accuracy?

Efficiency of MC simulation

- An accuracy of $\|x_n - x\| \leq \varepsilon$ requires more than $N > \frac{1}{\varepsilon}$ samples
- If ε decreases exponentially, then N increases exponentially.

Efficiency of calculation of consecutive state distributions x_n

- λ eigenvalue of P if $x \cdot P = \lambda \cdot x$
- P stochastic $\Rightarrow |\lambda| \leq 1$
- Convergence speed of x_n towards x : $\|x_n - x\| \leq (\lambda_s)^n$
 - λ_s largest eigenvalue with $|\lambda_s| < 1$ (subdominant eigenvalue)
- $(\lambda_s)^n \leq \varepsilon \Rightarrow n \geq \ln(\varepsilon)/\ln(\lambda_s)$
- If ε decreases exponentially, then n increases linearly.



Conclusion

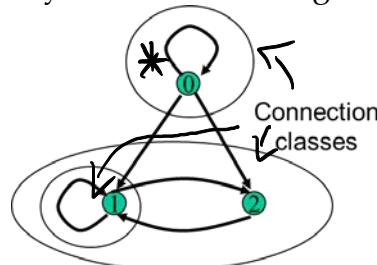
6

- For small values of ε , MC simulation is less efficient than calculation of consecutive state distributions x_n .
- But may be efficient for large ε .
- Combine simulation and analysis for speedup (optimization method):
 - First simulate x rather quickly using MC simulation for little accuracy
 - Then use x as start vector for calculation of consecutive state distributions x_n

8.10 Markov Chain Classification

Example 1:

- $P = \begin{pmatrix} 0.1 & 0.45 & 0.45 \\ 0 & 0.4 & 0.6 \\ 0 & 1 & 0 \end{pmatrix}$
- $x_0 = (1,0,0), x_1 = (0.1,0.45,0.45), x_2 = (0.01,0.675,0.315), x_3 = (0.0015895,0.4095), x_4 = (0.000164575,0.35415), x_5 = (0.000001,0.6125,0.3875), x_6 = (0.000001,0.6325,0.3675), x_7 = (0.0000001,0.6205,0.3675), x_8 = (0.00000001,0.6277,0.3723), \dots, x_{19} = (0.0,0.62499,0.37501), x_{20} = (0,0.62501,0.374993), \dots$
- Observation: probability for state 0 converges to zero.

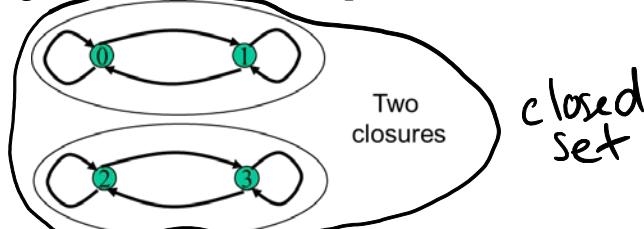


State transition graph:

- *Transient state*: will not be reached again with 100% probability ~~state 0~~
- *Recurrent state*: will be reached again with 100% probability ~~state 1, 2~~
- *Connection class*: set of states where each state reaches all other states in that set

Example 2:

- $P = \begin{pmatrix} 0.3 & 0.7 & 0 & 0 \\ 0.6 & 0.4 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.1 & 0.9 \end{pmatrix}$
- $x_0 = (1,0,0,0), x_1 = (0.3,0.7,0,0), x_2 = (0.51,0.49,0,0), x_3 = (0.447,0.553,0,0), x_4 = (0.4659,0.5341,0,0), x_5 = (0.4602,0.5398,0,0), x_6 = (0.4619,0.5381,0,0), x_7 = (0.4614,0.5386,0,0), x_8 = (0.4616,0.5384,0,0), \dots$
- $x_0 = (0,0,0.5,0.5), x_1 = (0,0,0.3,0.7), x_2 = (0,0,0.22,0.78), x_3 = (0,0,0.188,0.812), x_4 = (0,0,0.1752,0.8248), x_5 = (0,0,0.1701,0.8299), x_6 = (0,0,0.1680,0.8320), x_7 = (0,0,0.1672,0.8328), x_8 = (0,0,0.1669,0.8331), \dots$
- Observation: average state distribution depends on start vector x_0

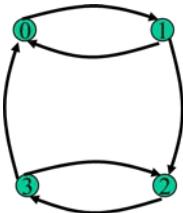


State transition graph:

- A subset of states is *closed* if no other state can be reached from it.
- *Closure*: a minimum closed set.
 - A maximum connection class is either a closure or completely “absorbed” ~~* State 0~~
- A MC is *irreducible* if the set of all states is the only closure, otherwise it is *reducible*.

Example 3:

- $P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 1 \\ 0.9 & 0 & 0.1 & 0 \end{pmatrix}$
- $x_0 = (1,0,0,0), x_1 = (0,1,0,0), x_2 = (0.4,0,0.6,0), x_3 = (0,0.4,0,0.6), x_4 = (0.7,0,0.3,0), x_5 = (0,0.7,0,0.3), x_6 = (0.55,0,0.45,0), x_7 = (0,0.55,0,0.45), x_8 = (0.625,0,0.375,0), x_9 = (0,0.625,0,0.375), x_{10} = (0.5875,0,0.4125,0), x_{11} = (0,0.5875,0,0.4125), x_{12} = (0.6063,0,0.3938,0), x_{13} = (0,0.6063,0,0.3938), x_{14} = (0.5969,0,0.4031,0), x_{15} = (0,0.5969,0,0.4031), \dots$
- Observation
 - Oscillation with period $p = 2 \Rightarrow x_{12} = x_{14}; x_{13} = x_{15} \dots$
 - Consecutive state distributions do not converge!
 - But series $(x_{p \cdot n + k})_{0 \leq n < \infty}$ converges



State transition graph:

- A state i has *period* p if it can be reached again only after multiples of p transition steps.
- Period p is greatest common divisor of all cycle lengths in state transition graph
- All states of a closure have the same period.
- A closure with period 1 is called *aperiodic* and a closure with period $p > 1$ is called *p-cyclic*.
- If all closures of a MC are aperiodic, the MC is also called *aperiodic*, otherwise it is called *periodic*.

Optimization for calculation of average state distribution for periodic MCs

- Recall average state distribution: $x = \lim_{n \rightarrow \infty} \left(\frac{1}{n+1} \cdot \sum_{i=0}^n x_i \right)$
 - Slow convergence
- Limit $x = \lim_{n \rightarrow \infty} x_n$ converges faster
 - But exists only for aperiodic MCs
 - $x = \lim_{n \rightarrow \infty} x_n$ does not exist for periodic MCs
- Use $x = \lim_{n \rightarrow \infty} \left(\frac{1}{p} \cdot \sum_{n \leq i < n+p} x_i \right)$ for fast calculation of avg. state distribution
 - Period p needs to be computed before

8.11 Analysis of Markov Chains with Stopping Times

Principle

- Markov chain runs until it is stopped by a defined condition, e.g., by the fact that a stop state is reached.

Example: game of dice

- A dice is subsequently thrown until the sum of the pips is at least ten. What is the distribution of the required number of throws and the number of pips at the end of the game?

Model

- State space $\mathcal{X} = \{0, \dots, 15\}$: sum of pips
 - Set of stopping states: $\mathcal{X}_s = \{10, \dots, 15\}$ (end of the game)
 - Set of transient states: $(\mathcal{X} \setminus \mathcal{X}_s) = \{0, \dots, 9\}$ (during the game)
 - Start state $X_0 = 0 \Rightarrow$ start state distribution x_0
- Factor space $\mathcal{Y} = \{1, \dots, 6\}$: pips per throw
 - Factor distribution y : uniform
- State transition function: $f(X, Y) = X + Y$

Wanted

- Distribution of (n, i)
 - Process stops after n transitions in stopping state $i \in \mathcal{X}_s$.
 - (Game is over after n throws with a sum of i pips.)
- Useful for performance analysis
 - Distribution of number of transitions until Markov chain stops (stopping time)
 - Distribution of pips at the end of the game

Idea for calculation

- Perform all possible state transitions until resulting state is stopping state
- Record probabilities for (n, i)

Algorithm

- Input: x_0
 - $m = 0$ // number of transitions
 - Repeat
 - For all $i \in \mathcal{X}$ // init state probabilities
 - $x_{m+1}(i) = 0$
 - Endfor
 - For all $i \in \mathcal{X} \setminus \mathcal{X}_s$
 - For all $j \in \mathcal{Y}$
 - $k = f(i, j)$
 - $x_{m+1}(k) = x_{m+1}(k) + x_m(i) \cdot y(j)$
 - Endfor
 - Endfor
 - $m = m + 1$
 - Until $\sum_{i \in \mathcal{X} \setminus \mathcal{X}_s} x_m(i) = 0$
- Output: $m, x_n(i), 0 \leq n \leq m, i \in \mathcal{X}_s$

Comments

- $x_n(i), 0 \leq n \leq m, i \in \mathcal{X}_s$ hold probabilities for (n, i) and yield the wanted distribution.
- x_n is not a distribution for $n > 0$ as $\sum_{i \in \mathcal{X}} x_n(i) = 1$ does not hold in general.
- This algorithm terminates only if the underlying Markov chain is stopped within a finite number of transition steps.
- Example calculations
 - Distribution of number of transitions until Markov chain stops (stopping time): $P(N = n) = \sum_{i \in \mathcal{X}_s} x_n(i)$
 - Distribution of pips at end of the game: $P(X = i) = \sum_{0 \leq n \leq m} x_n(i)$
- More on this topic: look for “stopping time” in mathematical literature

Other examples

- M. Menth and F. Lehrieder: “PCN-Based Measured Rate Termination”, Computer Networks, vol. 54, no. 13, Sept. 2010, p. 2099 – 2116
- A. Clark and S. Gilmore: “Terminating Passage-Time Calculations on Uniformised Markov Chains”, 24th UK Performance Engineering Workshop (UKPEW), London, UK, July 2008
- Chuan Heng Foh and Moshe Zukerman: “Performance Evaluation of IEEE 802.11”, IEEE Semiannual Vehicular Technology Conference (VTC, Fall), 2001
- Einführendes Beispiel in Modellierung & Simulation: “Tempo, kleine Fische!” (Übung -1)

8.12 Application Example: Analysis of a GI/GI/1- Q_{\max} Queue

Use MCs with stopping times to calculate

- the number of customers C that are served within an interval of size T with an i.i.d. service time B
- the remaining service time D of the following customer that could not be fully served

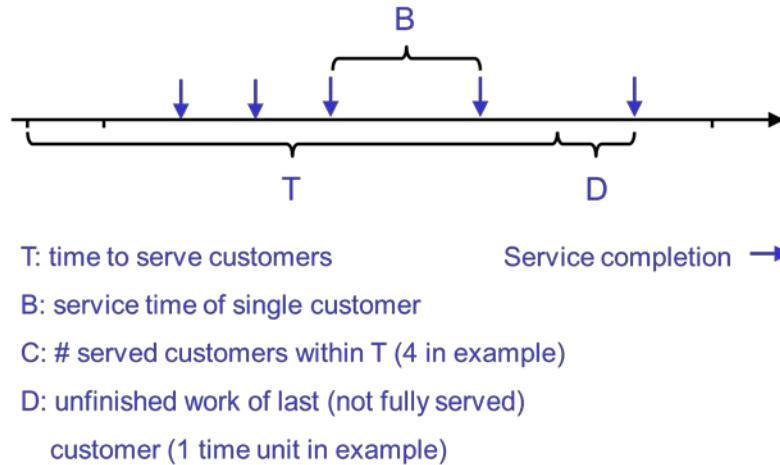


Abbildung 6: C customers are served within time T and a service time D remains from another customer.

- How are $(C,D)(T)$ computed?
 - Use Markov chain with stopping time (similar to game of dice)
 - State: X (cumulated service times), initial state $X=0$
 - Factor: B (service time)
 - State transition function: $f(X,B)=X+B$
 - Process stops when $X+B > T$
 - $C = \text{number of transitions when process stops} - 1$
 - $D = X-T > 0$

Analyze GI/GI/1- Q_{\max}

- Difference between GI/GI/1- Q_{\max} and GI/GI/1- ∞ (previously considered system): queue holds customers, not unfinished work
- State: (Q,U)
 - Q: number of waiting customers in queue (max. size Q_{\max})
 - U: unfinished work in server unit (in time)

- State evolution depends on inter-arrival time A and service time B
- Embedded points: shortly *before* customer arrivals
- What happens between embedded points?
 - New customer arrives; if $Q = Q_{\max}$, new customer is dropped
 - Service of customer in service unit possibly completes so that T time is left until next arrival
 - Possibly C additional customers are fully served
 - Possibly another customer is partially served so that D unfinished work remains in server unit
 - Conditional factor $(C, D)(T)$, $T \geq 0$
 - $C(T)$: number of fully served customers within time T ($C \geq 0$)
 - $D(T)$: unfinished work of $(C+1)^{\text{st}}$ customer within time T ($D > 0$)
- State transition function
 - $Q' = \min(Q_n + 1, Q_{\max})$
 - If $(U_n > A) //$ customer in service does not finish
 - $U_{n+1} = U_n - A$
 - $Q_{n+1} = Q'$
 - Else // customer in service finishes, T time remains, within which further C cust may finish and additional unfinished work D may be left
 - $Q'' = \max(Q' - 1, 0)$
 - $T = \max(A - U_n, 0)$
 - If $Q'' \leq C(T) //$ all customers in queue can be served
 - $Q_{n+1} = 0$
 - $U_{n+1} = 0$
 - Else
 - $Q_{n+1} = Q'' - C(T)$
 - $U_{n+1} = D(T)$
 - Endif
 - Endif

8.13 Memory Markov Chain (MMC) (see Paper O. Rose, ITC, 1999)

8.13.1 Struktur der MMC

Literatur:

- O. Rose, „A Memory Markov Chain for VBR Traffic With Strong Positive Correlations“, 16th International Teletraffic Congress, Edinburgh
- M. Menth, A. Binzenhöfer, and S. Mühleck: “Source Models for Speech Traffic Revisited”, in IEEE/ACM Transactions on Networking, vol. 17, no. 4, August 2009, IEEE

Problem

- Korrelierte Zufallsvariablen, deren Korrelationen sich nicht oder nur schwer mittels einfacher stochastischer Prozesse beschreiben lassen.
- Beispiel: Folge von Paketgrößen komprimierter Sprachproben
31, 31, 31, 31, 11, 4, 4, 4, 12, 4, 4, 31, 31, 31, 11, 4, 4, 4, 12, 4, 4, 31, 31, 11, 4, 4, 4, 12, 4, 31, ...

Idee: Hidden Markov Model

- Eine Variable S charakterisiert den Zustand eines Prozesses
- Eine bedingte ZV $U(S)$ realisiert eine zufällige aber zustandsabhängige Ausgabe
- Nach jeder Ausgabe kann sich der Zustand S ändern.
- Nur U ist von außen beobachtbar
- S ist von außen nicht beobachtbar, die Markov-Kette ist also versteckt.

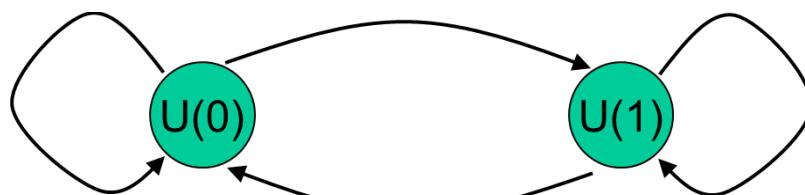


Abbildung 7: Markovkette mit zwei Zuständen und zustandsabhängigen Zufallsvariablen $U(S)$ als Ausgabe.

Simulation

- Realisierung der Ausgabe gemäß der bedingten Zufallsvariable $U(S)$
 - Benutzt bedingte Verteilungen $u(S)$
- Berechnung des nächsten Zustandes S der Markov-Kette
 - Benutzt eine Zustandsübergangsmatrix P

Beispiel

- Folge von Paketgrößen komprimierter Sprachproben (z.B. 31, 31, 31, 31, 11, 4, 4, 12, 4, 4, 31, 31, 31, 11, 4, 4, 4, 12, 4, 4, 31, 31, 11, 4, 4, 4, 12, 4, 31, ...) soll durch Markov-Kette mit zustandsspezifischer Ausgabe erzeugt werden können.
- Lösung:

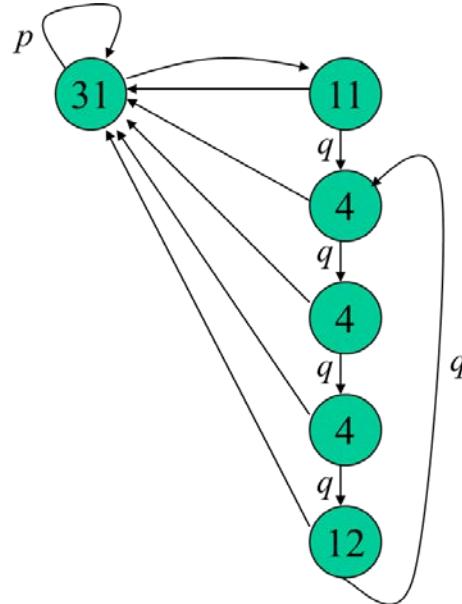


Abbildung 8: Zustandsübergangsdiagramm für die steuernde Markov-Kette einer MMC. Die Aktiv- und Passivzeiten des Sprechers sind geometrisch verteilt mit Parameter p und q . Die Zahl im Kreis entspricht einer Ausgabefunktion für die deterministische ZV in diesem Zustand.

- Zustände: 0, 1, ..., 5
- Zustandsspezifische Ausgabe: $U(0)=31, U(1)=11, U(k)=4$ mit $k \in \{2,3,4\}$, $U(5)=12$

$$\bullet \text{ Übergangsmatrix } P = \begin{bmatrix} p & 1-p & 0 & 0 & 0 & 0 \\ 1-q & 0 & q & 0 & 0 & 0 \\ 1-q & 0 & 0 & q & 0 & 0 \\ 1-q & 0 & 0 & 0 & q & 0 \\ 1-q & 0 & 0 & 0 & 0 & q \\ 1-q & 0 & q & 0 & 0 & 0 \end{bmatrix}$$

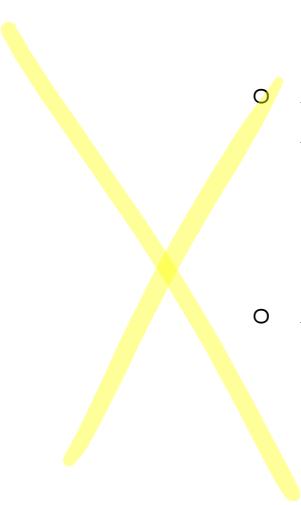
- Die Verallgemeinerung dieser Struktur ist eine „Memory Markov Chain“ (MMC)

8.13.2 Parametrisierung der MMC

Gegeben ist eine Zeitreihe x_1, \dots, x_N und gesucht ist eine MMC, die ähnliche Zeitreihen erzeugt.

- MMC(W, M_s, M_a)
- Parameter: Fensterlänge W , Stichprobenzustandsgröße M_s , Durchschnittszustandsgröße M_a
- $\bar{x}_i = \frac{1}{W} \cdot \sum_{i-W < k < i} x_k$ für $i=W+1, \dots, N$
- Generierung von Paaren (sample, average) = (x_i, \bar{x}_i)
- Reduktion der Paare auf einen geeigneten Zustandsraum
 - Zustandsraum $(v, w) \in \{1, \dots, M_s\} \times \{1, \dots, M_a\}$; alternativ: M_s bzw. M_a diskrete Werte, die keine natürlichen Zahlen darstellen müssen
 - Zuordnung $(x_i, \bar{x}_i) \sim (y_i, \bar{y}_i) \in \{1, \dots, M_s\} \times \{1, \dots, M_a\}$
 - $y_i = \begin{cases} 1 & x_i = \min_j(x_j) \\ \frac{x_i - \min_j(x_j)}{\max_j(x_j) - \min_j(x_j)} \cdot M_s & \text{otherwise} \end{cases}$
 - $\bar{y}_i = \begin{cases} 1 & \bar{x}_i = \min_j(\bar{x}_j) \\ \frac{\bar{x}_i - \min_j(\bar{x}_j)}{\max_j(\bar{x}_j) - \min_j(\bar{x}_j)} \cdot M_a & \text{otherwise} \end{cases}$
 - Aufzählung der Zustände (v, w) aufsteigend nach absoluter Ordnung
 - Empirische Bestimmung der Übergangswahrscheinlichkeitsmatrix P anhand der Zeitreihe (y_i, \bar{y}_i)
 - Bestimmung der stationären Zustandsverteilung π für die Zustände $(v, w) \in \{1, \dots, M_s\} \times \{1, \dots, M_a\}$
 - Assoziation der Zustandskomponenten $v \in \{1, \dots, M_s\}$ mit einer Ausgabefunktion $u(v)$
 - Wenn $x_i = y_i$ gilt, dann kann für $u(v) = v$ gewählt werden.
 - Ansonsten:
 - $\delta(y_i, v) = \begin{cases} 0 & y_i \neq v \\ 1 & y_i = v \end{cases}$
 - $u(v) = \frac{\sum_{W < i \leq N} \delta(y_i, v) \cdot x_i}{\sum_{W < i \leq N} \delta(y_i, v)}$

- Vergleich von empirischer und theoretischer Korrelation
 - Empirische (Auto-)Kovarianz (siehe Kapitel 2)
 - $\overline{\text{COV}}[X, Y] = \frac{1}{n-1} \cdot \sum_{0 \leq i < n} E[X_i - \bar{X}(n)] \cdot [Y_i - \bar{Y}(n)]$
 - $\hat{C}_j(n) = \frac{1}{n} \sum_{0 \leq i < n-j} [X_i - \bar{X}] [X_{i+j} - \bar{X}]$

 Zur Erinnerung: theoretische Kovarianz bzw. Korrelation (siehe Kapitel 2)

- Analog theoretische Autokovarianz bzw. Autokorrelation für MMC:
 - Vektor u enthält die zu (v, w) entsprechenden Ausgabewerte $U(v)$.
 - $diag(u)$ ist die zu u entsprechende Diagonalmatrix.
 - Berechnung der Momente: $E[X^k] = \pi \cdot diag(u)^k \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$
 - Autokovarianz zum Lag k:

$$Cov(k) = \pi \cdot diag(u) \cdot P^k \cdot diag(u) \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - E[X]^2$$

$$\text{▪ Autokorrelation zum Lag k: } Cor(k) = \frac{Cov(k)}{E[X^2] - E[X]^2}$$

- Falls empirische und theoretische Korrelation nicht gut übereinstimmen, wähle andere Parameter W, M_s, M_a !

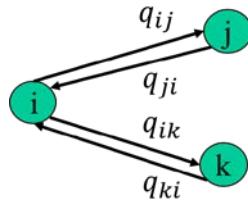
9 Continuous-Time Markov Chains (CTMCs)

See: P. Tran-Gia: "Einführung in die Leistungsbewertung und Verkehrstheorie", 2. Auflage, 2005, Oldenbourg

9.1 Basics about CTMCs

9.1.1 Definition

- Stochastic process with transitions from state i to another state j after exponentially distributed time ($A_{ij}(t) = 1 - e^{-q_{ij} \cdot t}$)
 - Process has memoryless property at any time instant.
 - "Transitions happen at rate q_{ij} "



9.1.2 Simulation

Input: start state X, start time t

While(true)

```

 $A_{min} = \infty; \quad next = -1$ 
For all  $q_{XY} > 0$ :
 $A = \frac{-\ln(U)}{q_{XY}}$ 
If  $A < A_{min}$ :
 $A_{min} = A; \quad next = Y$ 
 $X = next; \quad t += A_{min}$ 

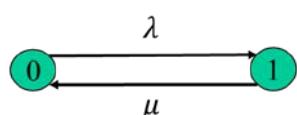
```

`TWH.count(X,t)` // time-weighting histogram with 1 bin per state

9.1.3 Application Example

- Single server **M/M/1-0 System**
- Exponentially distributed inter-arrival time $A(t) = 1 - e^{-\lambda \cdot t}$
- Exponentially distributed service time $B(t) = 1 - e^{-\mu \cdot t}$
- Two states
 - 0: server idle
 - 1: server busy
- State transition diagram
 - Transitions marked with the corresponding exponential rates

allgemein:
 λ =Ankunftsrate
 μ =Dienratenrate



9.1.4 Modelling „Aggregate“ Rates

- Consider M/M/3 loss model
 - 3 servers
 - State indicates number of occupied service units
 - State transition diagram:
- Transition with rate $k \cdot \mu$ from k to $k - 1$ occupied service units – to be shown in the following
- k occupied service units, each with exponentially distributed service time
- $P(B_{single}^i < t) = 1 - e^{-\mu \cdot t}$
- What is the distribution of time B until a first service unit completes its service?
- Answer:

$$1 - B(t) = P(\min_{0 \leq i \leq k} (B_{single}^i) > t) = \prod_{0 \leq i \leq k} P(B_{single}^i > t) =$$

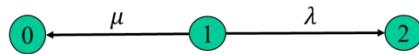
$$\prod_{0 \leq i \leq k} e^{-\mu \cdot t} = e^{-k \cdot \mu \cdot t}, \text{ i.e., } B(t) = 1 - e^{-k \cdot \mu \cdot t}$$

9.1.5 Distribution of sojourn time S_i in State i

- $1 - S_i(t) = P(\min_{j \neq i} (A_{ij}) > t) = P(A_{ij} > t, j \neq i) = \prod_{j \neq i} P(A_{ij} > t) =$
 $\prod_{j \neq i} e^{-q_{ij} \cdot t} = e^{-(\sum_{j \neq i} q_{ij}) \cdot t}$ (13.2)
- Define rate to leave state i is $q_i = \sum_{j \neq i} q_{ij}$
so that we get $1 - S_i(t) = e^{-q_i \cdot t}$ (13.3)

9.1.6 Transition Probabilities

- Problem: process goes from state 1 to state 2 after A time (arrival event, $A(t) = 1 - e^{-\lambda \cdot t}$) or to state 0 after B time (service event, $B(t) = 1 - e^{-\mu \cdot t}$)



- What is the probability that the next state is 2 or 0, respectively?
- Solution

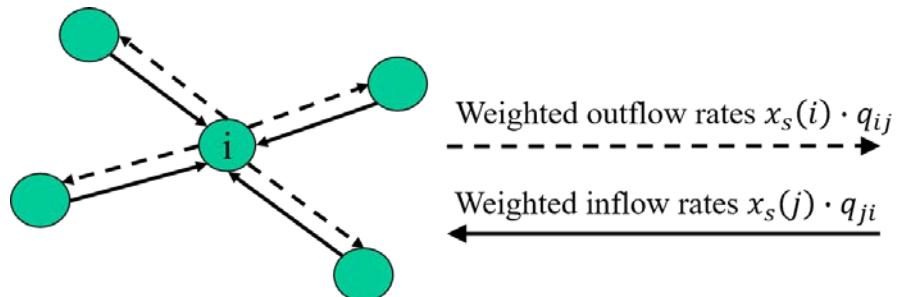
- Next state is 2 with probability $\int_0^\infty a(t) \cdot P(B > t) dt = \int_0^\infty \lambda \cdot e^{-\lambda \cdot t} \cdot e^{-\mu \cdot t} dt = \frac{-\lambda}{\lambda + \mu} \cdot \int_0^\infty -(\lambda + \mu) \cdot e^{-(\lambda + \mu) \cdot t} dt = \frac{-\lambda}{\lambda + \mu} \cdot [e^{-(\lambda + \mu) \cdot t}]_0^\infty = \frac{\lambda}{\lambda + \mu}$
- Next state is 0 with probability $\frac{\mu}{\lambda + \mu}$

- Generally
 - Process in state i
 - Next state is j with probability $\frac{q_{ij}}{\sum_{k \neq i} q_{ik}}$

9.2 Analysis of CTMCs

- Given
 - $n+1$ states, numbered $0 \leq i \leq n$
 - State transition rates q_{ij}
- Wanted: average state distribution
- Hint: stationary state distribution exists and equals average state distribution
- Property of the stationary state distribution $x_s = (x_s(0), \dots, x_s(n))$:
 - It's a distribution: $\sum_i x_s(i) = 1$ (13.5)
 - For any state i holds: $x_s(i) \cdot \sum_{j \neq i} q_{ij} = x_s(i) \cdot q_i = \sum_{j \neq i} x_s(j) \cdot q_{ji}$ (13.6)

- Illustration



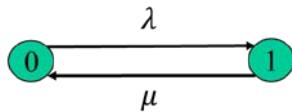
- Background: for any stationary state distribution x_s holds: sum of weighted outflow rates equals sum of weighted inflow rates
- $n + 2$ equations for $n + 1$ free variables $x_s(i), 0 \leq i \leq n$, but equations are linearly dependent and one of them can be eliminated

- Find the stationary state distribution x_s by solving the equation system above.
- How to express the Conditions (13.5) and (13.6) as a vector-matrix operation?
 - Reformulation of (13.6): $\sum_{j \neq i} q_{ji} \cdot x_s(j) - q_i \cdot x_s(i) = 0$
 - Define $q_{ii} = -q_i = -\sum_{j \neq i} q_{ij}$ (13.7)
 - q_{ii} is similar to rate q_i for leaving state i , but it is negative!
 - Define the **transition rate matrix** $Q = \begin{pmatrix} q_{00} & \cdots & q_{0,n} \\ \vdots & \ddots & \vdots \\ q_{n,0} & \cdots & q_{n,n} \end{pmatrix}$ (13.8)
 - Note: the sum of the entries in a line is 0.
 - The diagonal has negative entries.
 - Don't confuse with state transition probability matrix P where sum of entries in a line is 1!
 - Then the stationary state distribution x_s is defined by
 - $x_s \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = 1$ according to (13.5) (13.9)
 - $x_s \cdot Q = (0 \quad \cdots \quad 0)$ according to (13.6) (13.10)

9.3 Comparative Calculation of Stationary State Distributions

9.3.1 CTMC

- State transition diagram



- Rate matrix $Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$
- Conditions for stationary state distribution
 - (I) $x(0) + x(1) = 1 \rightarrow x(0) = 1 - x(1)$
 - (II) $x \cdot Q = (0, \dots, 0)$
 - (IIa) $-\lambda \cdot x(0) + \mu \cdot x(1) = 0$
- Solve the equation system

(I) in (IIa)

$$-\lambda \cdot [1 - x(1)] + \mu \cdot x(1) = 0$$

$$-\lambda + \lambda \cdot x(1) + \mu \cdot x(1) = 0$$

$$(\mu + \lambda) \cdot x(1) = \lambda$$

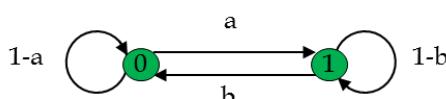
$$x(1) = \frac{\lambda}{\mu + \lambda} \rightarrow x(0) = \frac{\mu}{\mu + \lambda}$$

- Numerical example: $\lambda = \frac{2}{s}$, $\mu = \frac{3}{s} \rightarrow x(0) = 0.6, x(1) = 0.4$

Note: (II) does not need to be given in vector-matrix form, solving (13.5) and (13.6) is sufficient.

9.3.2 DTMC

- State transition diagram



- State transition matrix $P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$
- Conditions for stationary state distribution
 - (I) $x(0) + x(1) = 1 \rightarrow x(0) = 1 - x(1)$
 - (II) $x \cdot P = x$
 - (IIb) $a \cdot x(0) + (1-b) \cdot x(1) = x(1)$
- Solve the equation system

(I) in (IIb)

$$a \cdot [1 - x(1)] + (1-b) \cdot x(1) = x(1)$$

$$a - a \cdot x(1) + x(1) - b \cdot x(1) = x(1)$$

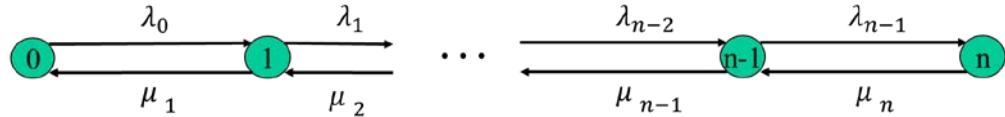
$$a = x(1) \cdot (1 + a - 1 + b)$$

$$x(1) = \frac{a}{a+b} \rightarrow x(0) = \frac{b}{a+b}$$

- Numerical example: $a = 0.2, b = 0.3 \rightarrow x(0) = 0.6, x(1) = 0.4$

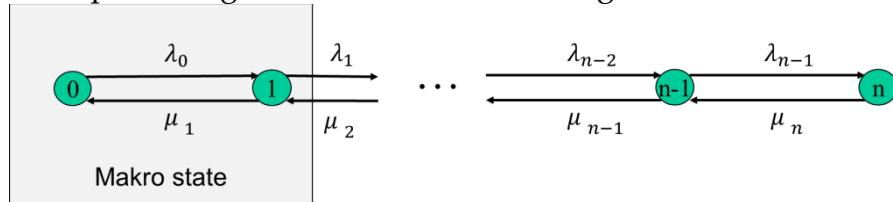
9.4 Birth-Death Processes

- Markov process
- Often (but not necessarily) one-dimensional state space, e.g., number of customers in the system
- State transitions only between “neighboring” states
- State transition diagram of a typical birth-death process with $n+1$ states



- This structure can be used to model M/M/s-K systems
 - M/M/n-0 ($s=n, K=0$)
 - Upward transition rates $\lambda_i = \lambda$
 - Downward transition rates $\mu_i = i \cdot \mu$
 - M/M/1-K ($s=1, K=n-1$)
 - Upward transition rates $\lambda_i = \lambda$
 - Downward transition rates $\mu_i = \mu$
 - M/M/s-K ($K=n-s$)
 - Upward transition rates $\lambda_i = \lambda$
 - Downward transition rates $\mu_i = \begin{cases} i \cdot \mu & \text{for } i \leq s \\ s \cdot \mu & \text{for } i > s \end{cases}$
- Equation system
 - $\sum_i x(i) = 1$
 - $\lambda_0 \cdot x(0) = \mu_1 \cdot x(1)$
 - $(\lambda_i + \mu_i) \cdot x(i) = \lambda_{i-1} \cdot x(i-1) + \mu_{i+1} \cdot x(i+1), i = 1, 2, \dots, n-1$
 - $\lambda_{n-1} \cdot x(n-1) = \mu_n \cdot x(n)$

- Simplification of the calculation of the stationary state distribution
 - Macro state: set of neighboring states
 - Example: contiguous sets of nodes starting with state 0



- Principle for new equations: sum of weighted inflow rates into macro state equals sum of weighted outflow rates out of macro state
- Simplified but equivalent equation system
 - $\sum_i x(i) = 1$
 - $\lambda_{i-1} \cdot x(i-1) = \mu_i \cdot x(i), i = 1, 2, \dots, n$
- Solution
 - $x(i) = x(0) \cdot \frac{\prod_{0 < k \leq i} \lambda_{k-1}}{\prod_{0 < k \leq i} \mu_k}, i = 1, 2, \dots, n \text{ and}$
 - $x(0) = \left(1 + \sum_{0 < i \leq n} \frac{\prod_{0 < k \leq i} \lambda_{k-1}}{\prod_{0 < k \leq i} \mu_k} \right)^{-1}$

9.5 Loss System M/M/n-0

9.5.1 Model and Performance Metrics

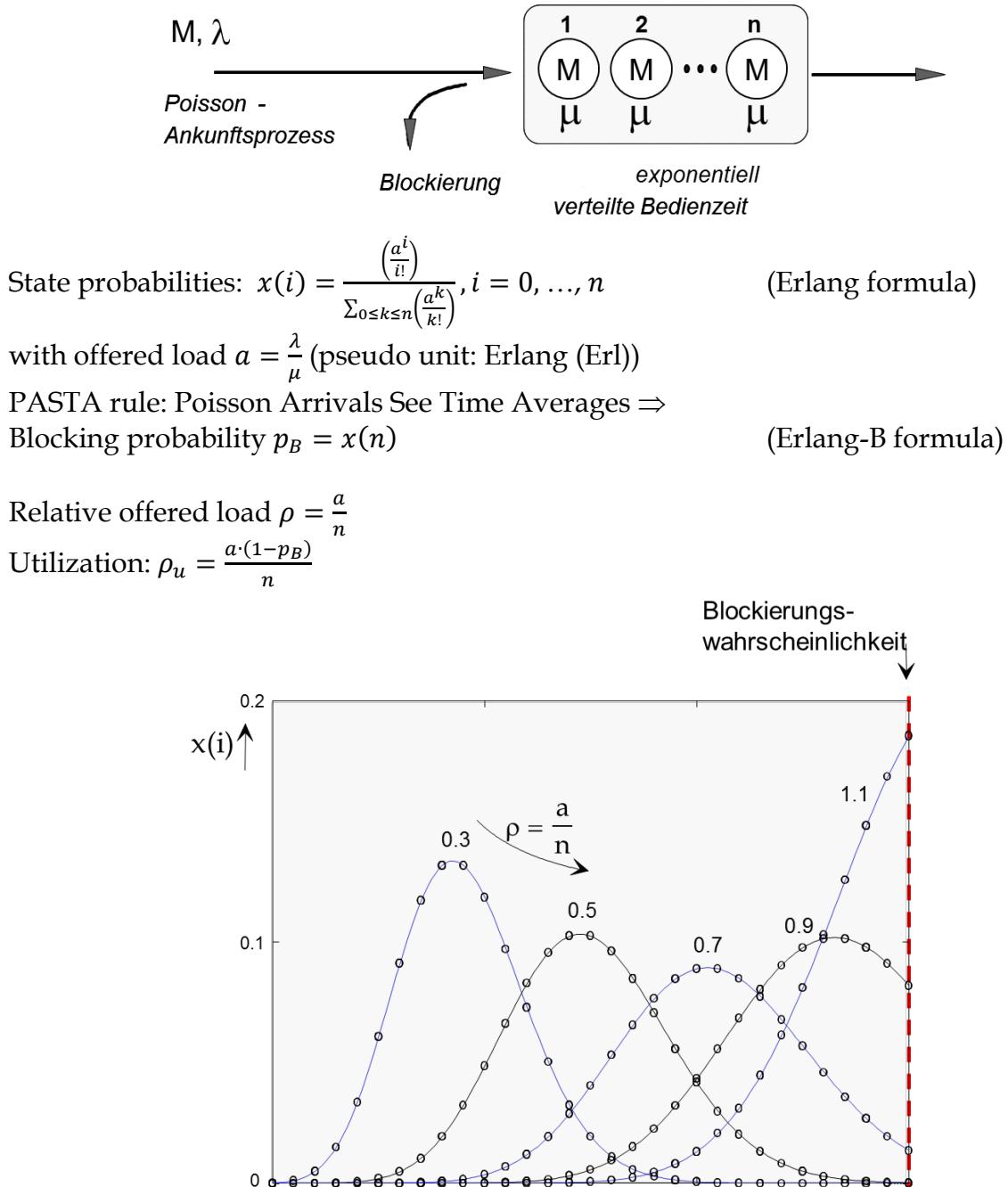


Abbildung 1: Stationary state distribution of a M/M/30-0 loss system for different relative offered load ρ .

9.5.2 When is PASTA not applicable?

Assume a M/M/1 loss system where requests arrive with rate λ in the presence of a free service unit and $\frac{\lambda}{2}$ in the presence of an occupied service unit.

Thus, the arrival process depends on the system state. Therefore, it is not homogeneous and it is not a Poisson arrival process.

The system can be modelled with two states (0 free, 1 occupied).

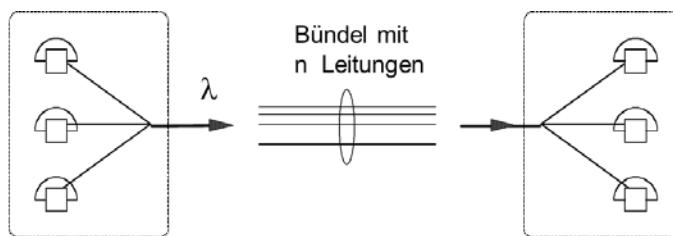
The PASTA rule cannot be applied and the blocking probability is NOT $p_B = x(1)$.

The blocking probability is rather $p_B = \frac{\frac{\lambda}{2}x(1)}{\lambda \cdot x(0) + \frac{\lambda}{2}x(1)}$.

9.5.3 Application Example

Beispiel: Dimensionierung einer Ortsvermittlungsstelle (Telefon)

- Assumptions about
 - Arrival rates
 - One call per second arrives
 - Exponential inter-arrival time
 - $\Rightarrow \lambda = \frac{1}{s}$
 - Service rates
 - Calls take on average 90 s
 - Exponential call-holding time
 - $\Rightarrow \mu = \frac{1}{90s}$
 - \Rightarrow Offered load $a = \frac{\lambda}{\mu} = 90$ Erl
- Needed: number of outgoing circuits so that the blocking probability of an incoming call is at most $p_B \leq 10^{-3}$
 - n=117 circuits are needed



n : Anzahl der Leitungen

B : Gesprächsdauer ($E[B] = 90$ sec)
 λ : Anrufe/sec (Poisson Prozess) ($\lambda=1$ Anruf/sec)

→ Angebot
 $a=90$

Dimensionierung mit
vorgegebener Dienstgüte $p_B \leq 10^{-3}$
(Quality of Service QoS)

→ n=117 Leitungen

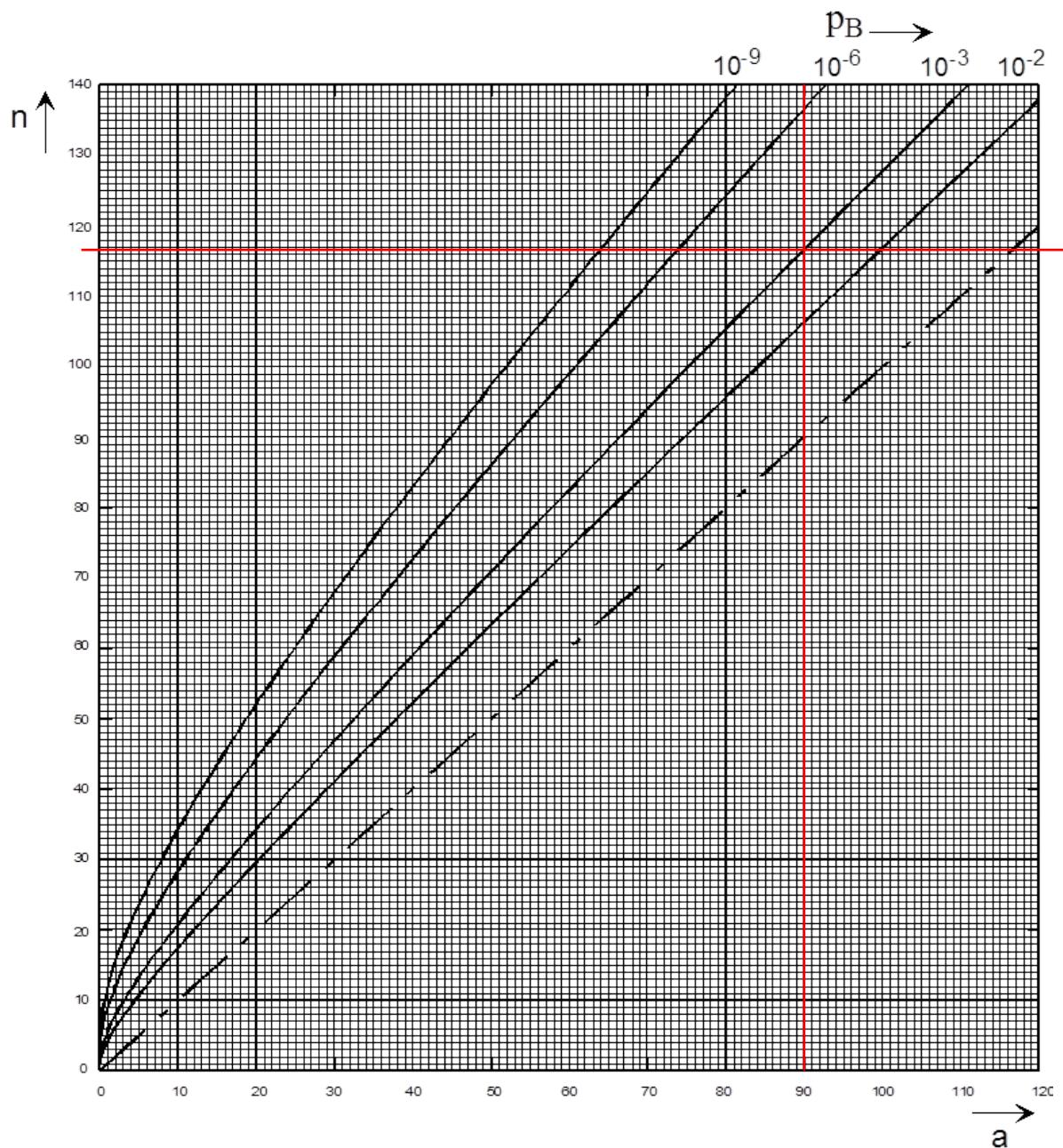


Abbildung 2: Benötigte Anzahl von Bedieneinheiten n bei einem $M/M/n$ Verlustsystem um eine gewisse Blockierwahrscheinlichkeit p_B bei vorgegebener Last a nicht zu überschreiten. Für geringere Blockierwahrscheinlichkeiten werden mehr Bedieneinheiten benötigt.

9.5.4 Economy of Scale (Bündelungsgewinn) with $M/M/n$

- Bei größerem Angebot werden relativ weniger Leitungen benötigt um die gleiche Blockierwahrscheinlichkeit zu erreichen
- Bei größerem Angebot bessere Auslastung von entsprechend dimensionierten Systemen möglich

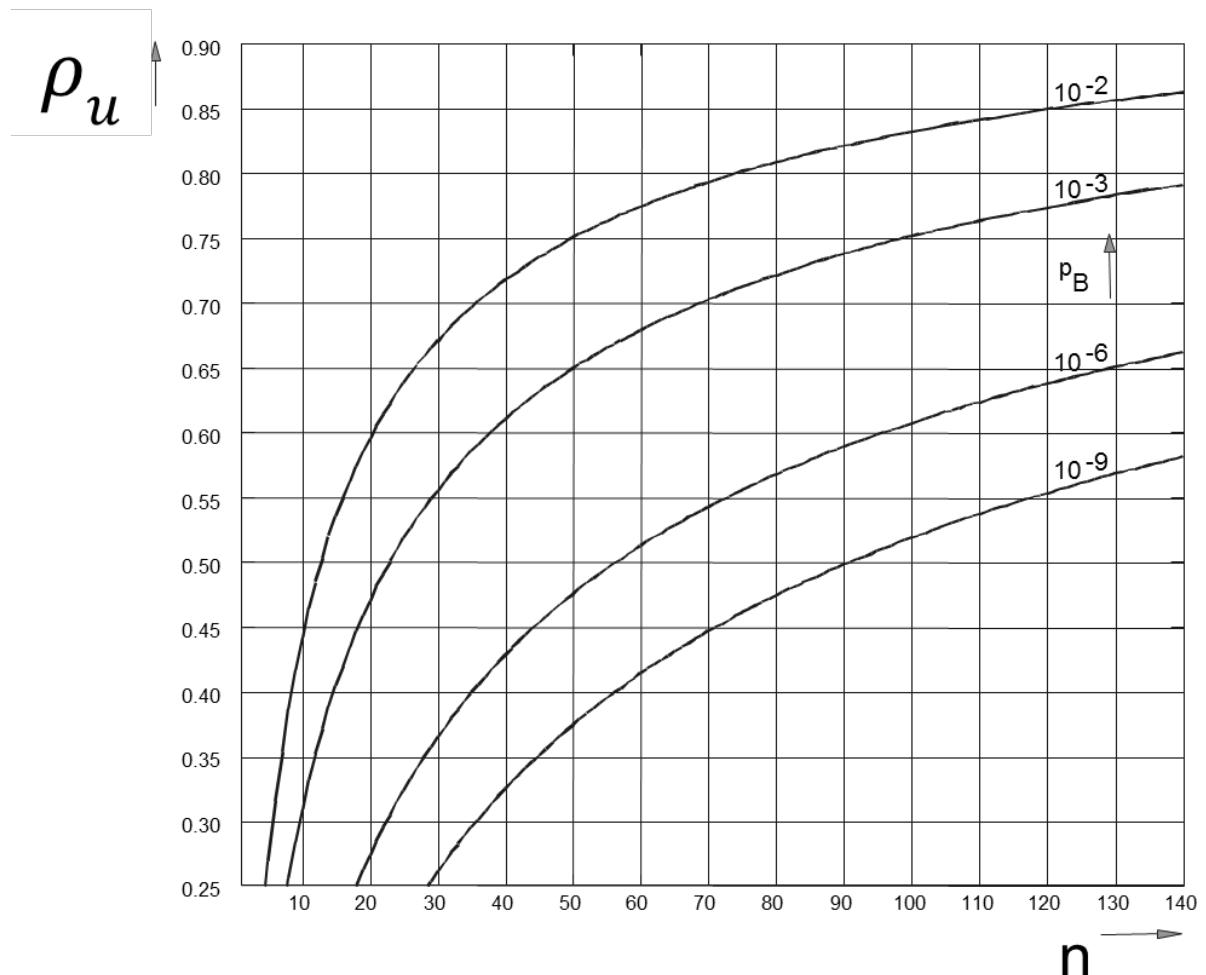


Abbildung 3: Auslastung von Bedieneinheiten, die so stark belastet sind, dass eine vorgegebene Blockierwahrscheinlichkeit eingehalten wird.

- Bei gleicher Blockierungswahrscheinlichkeit steigt die Auslastung der Leitungen ρ_u mit der Bündelgröße n
- Größere Leitungsbündel sind wirtschaftlicher
- Bündelungsgewinn lässt sich nicht beliebig steigern

- Andere Sichtweise:

Wenn zwei kleine M/M/n Systeme zu einem großen M/M/2n System zusammengefasst werden, ist die Blockierwahrscheinlichkeit des großen Systems geringer als die der kleinen Systeme.

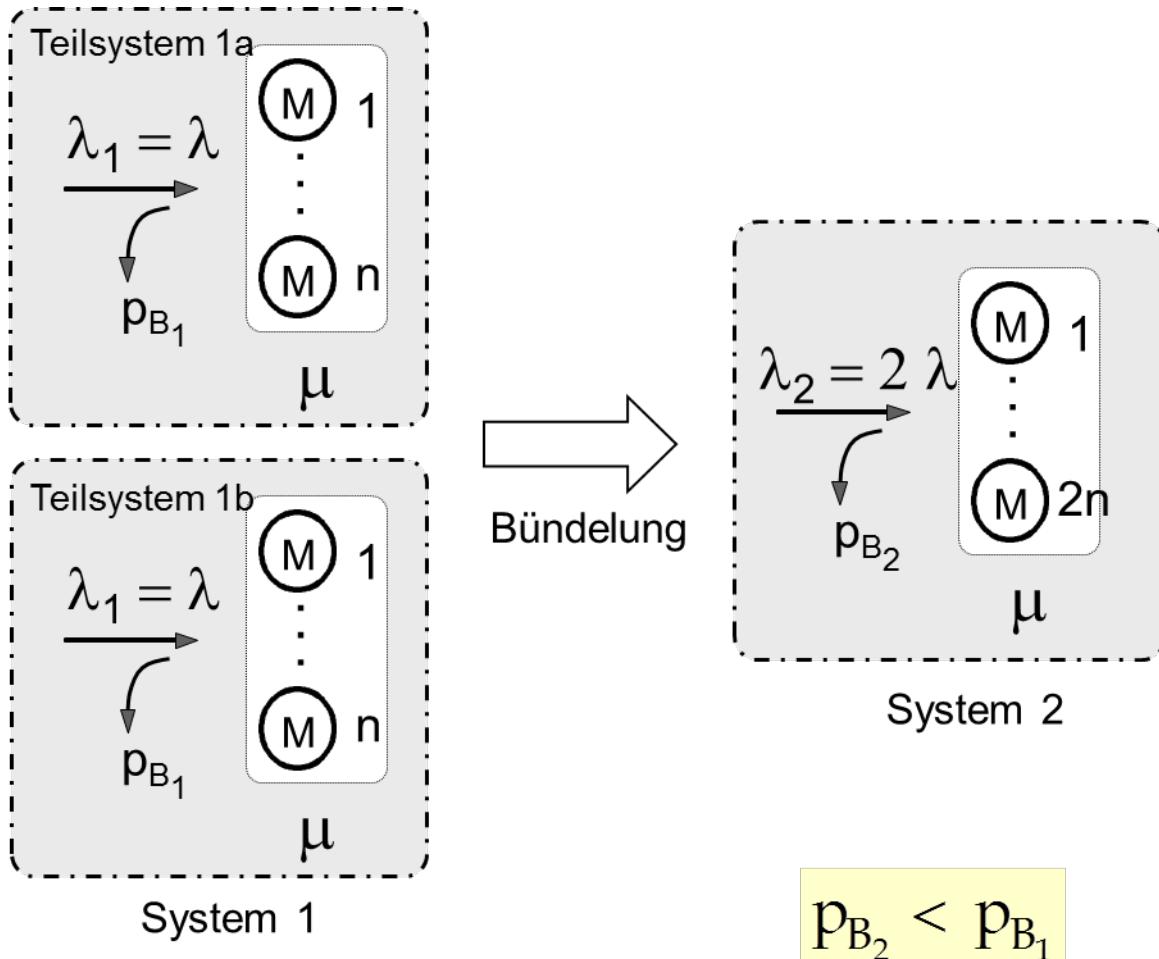
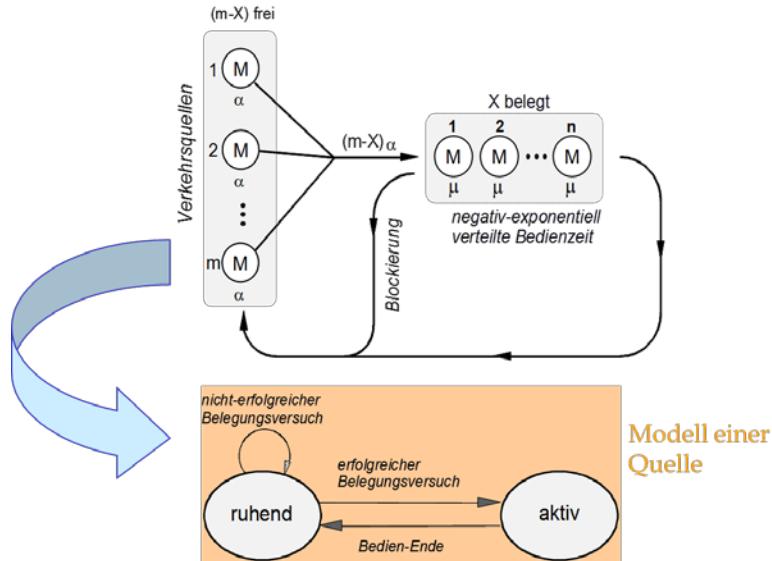


Abbildung 4: Combining customer arrivals and service unites of separate M/M/n-0 loss systems into a single one leads to lower blocking probabilities.

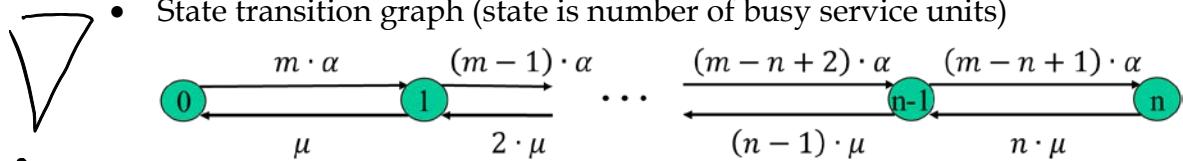
9.6 Loss System M/GI/n-0

Same formulae as for M/M/n-0

9.7 Loss System with a Finite Number of Sources



- Number of service units: n
- Number of sources: $m \geq n$
- Source can be inactive
 - Distribution function of inactive time is $I(t) = 1 - e^{-\alpha \cdot t}$
- At the end of an inactivity phase, the source makes a service request
 - Free service unit available
 - Source is serviced
 - Distribution function of service time is $B(t) = 1 - e^{-\mu \cdot t}$
 - Source becomes again inactive after service completion.
 - No free service unit available
 - Source is blocked
 - Source becomes inactive again
- State transition graph (state is number of busy service units)



- Distribution of busy service units
 - State probabilities: $x(i) = \frac{\binom{m}{i} \cdot (a^*)^i}{\sum_{k=0}^{m-1} \binom{m}{k} \cdot (a^*)^k}, i = 0, \dots, n$
 - With $a^* = \frac{\alpha}{\mu}$ (offered load of an inactive source)
- Blocking probability
 - $p_B = \frac{\binom{m-1}{n} \cdot (a^*)^n}{\sum_{k=0}^{m-1} \binom{m-1}{k} \cdot (a^*)^k}$ (Engset formula) *entweder Name oder Formel*
 - Note: blocking probability is based on state probability for n busy service units in the presence of $m - 1$ customers, that's what the requesting customer sees.
- Formulae also hold for GI distributed service time.

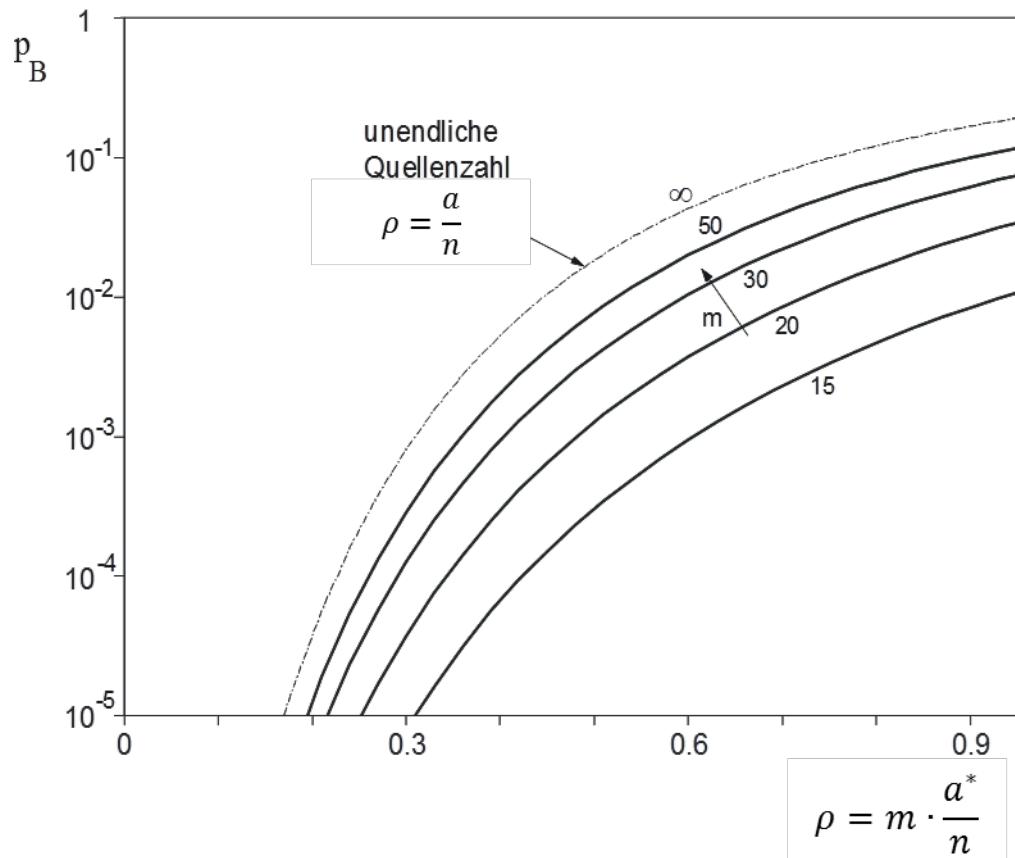
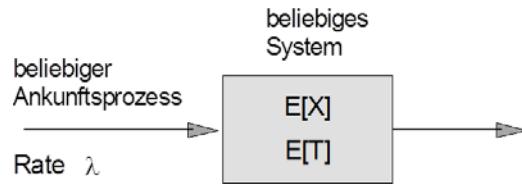


Abbildung 5: Blocking probabilities according to the Engset formula (The number of servers n is fixed but unknown, in any case $n < 15$). For an increasing number of customers m the blocking probability converges to the one of an M/M/n-0 loss system.

9.8 Theorem von Little



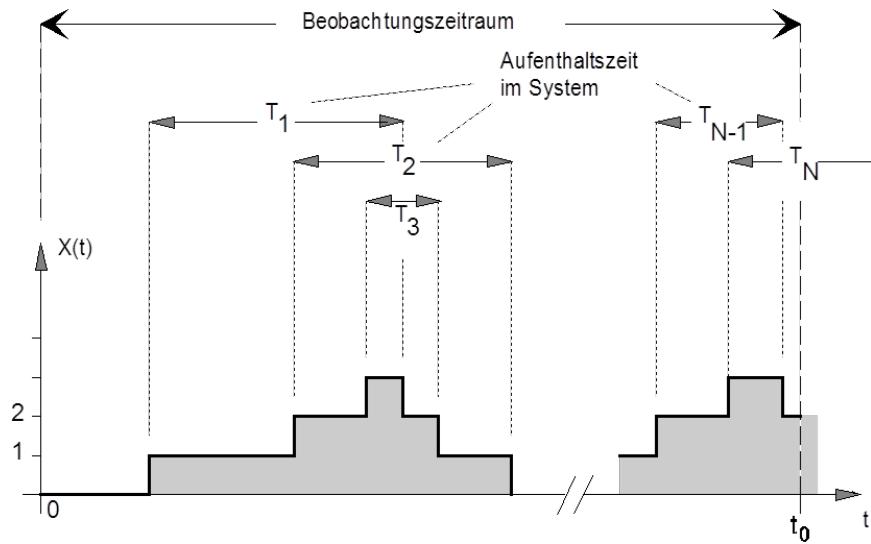
Nomenklatur (allgemein und im endlichen Beispiel für den Beweis)

- $\lambda, \bar{\lambda}$: mittlere Ankunftsrate des Ankunftsprozesses
- $E[T], \bar{T}$: mittlere Aufenthaltszeit im System
- $E[X], \bar{X}$: mittlere Anzahl von Anforderungen im System

$\left. \begin{array}{c} \bar{\lambda} \\ = \\ \frac{1}{\bar{T}} \\ \bar{X} \end{array} \right\}$ in definiertem Zeitraum

Es gilt: $\lambda \cdot E[T] = E[X]$ (Theorem von Little)

Beweis



Wir betrachten endlichen Prozessausschnitt

- t_0 : Dauer des betrachteten Prozesses
- N : Anzahl Ankünfte
- Es gilt
 - (1) $\bar{\lambda} = \frac{N}{t_0} \rightarrow t_0 = \frac{N}{\bar{\lambda}}$
 - (2) $\bar{X} = \frac{1}{t_0} \cdot \int_0^{t_0} X(t) dt$
 - (3) $\bar{T} \approx \frac{1}{N} \cdot \int_0^{t_0} X(t) dt \rightarrow \int_0^{t_0} X(t) dt \approx N \cdot \bar{T}$
- Nach Einsetzen von (1) und (3) in (2) folgt das Little-Theorem $\bar{X} \approx \bar{\lambda} \cdot \bar{T}$

9.9 Waiting System M/M/n-∞

- Can be modeled as birth-death process with infinite state space
- State indicates number of customers in the system
- State probabilities for $a < n$ (stability condition):

$$x(i) = \begin{cases} x(0) \cdot \frac{a^i}{i!} & i = 0, 1, \dots, n \\ x(n) \cdot \rho^{i-n} & i > n \end{cases} \text{ with relative offered load } \rho = \frac{a}{n} \text{ and}$$

$$x(0) = \left(\sum_{0 \leq k < n} \left(\frac{a^k}{k!} \right) + \frac{a^n}{n!} \cdot \frac{1}{1-\rho} \right)^{-1}$$

- For $n = 1$
 - $x(0) = \left(1 + a \cdot \frac{1}{1-\rho} \right)^{-1} = \left(\frac{1-\rho}{1-\rho} + \frac{\rho}{1-\rho} \right)^{-1} = 1 - \rho$
 - $x(1) = x(0) \cdot a = (1 - \rho) \cdot \rho; x(i) = (1 - \rho) \cdot \rho^i$ sogar für alle $i \geq 0$
- Utilization ρ_u equals relative offered load ρ since $p_B = 0$.

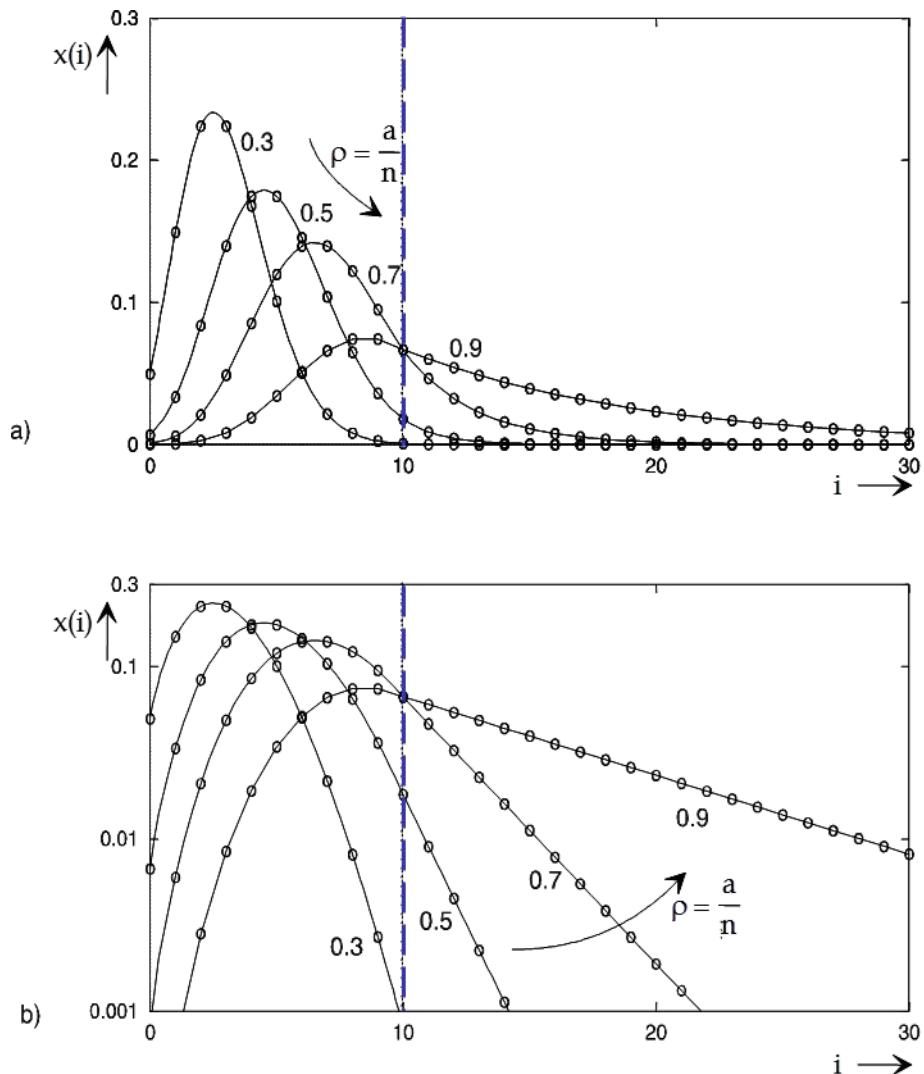


Abbildung 6: Stationary state distribution of an M/M/10-∞ waiting system for different relative offered load on a linearly and logarithmically scaled y-axis.

Waiting probability $p_W = \sum_{n \leq i < \infty} x(i) = x(n) \cdot \frac{1}{1-\rho}$ (Erlang waiting formula)

- For $n = 1$: $p_W = 1 - x(0) = \rho$

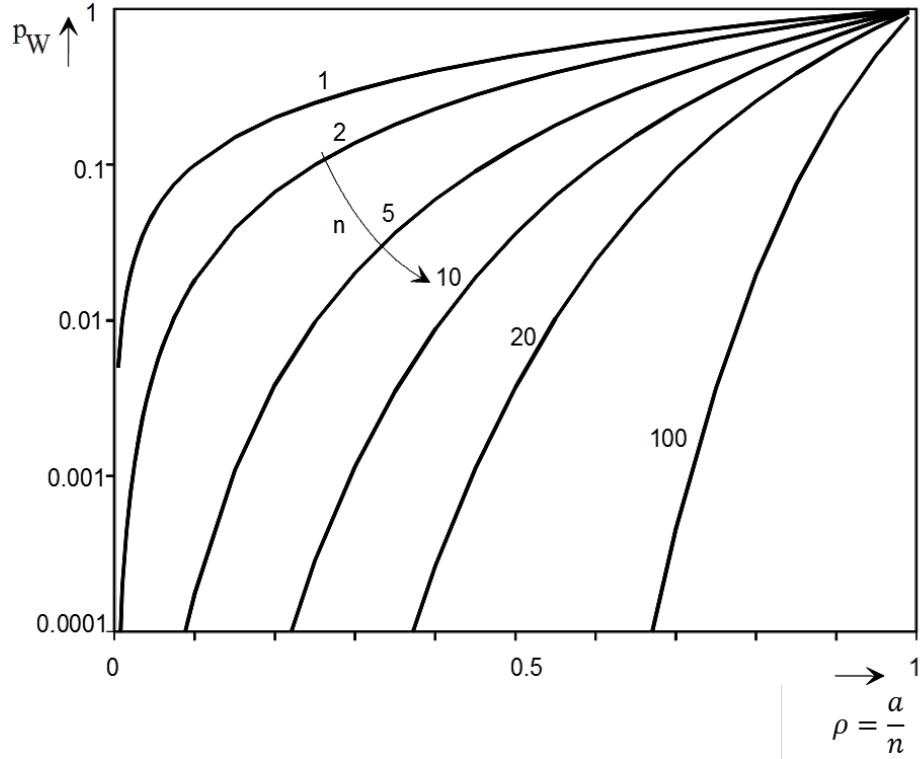


Abbildung 7: Waiting probabilities in a M/M/n-∞ waiting system. Economy of scale: systems with more service units reveal lower waiting probabilities for the same relative load.

X_w : number of waiting customers

Mean queue length:

$$E[X_w] = \sum_{n < i < \infty} (i - n) \cdot x(i) = x(n) \cdot \sum_{0 < i < \infty} i \cdot \rho^i = x(n) \cdot \frac{\rho}{(1 - \rho)^2} = p_W \cdot \frac{\rho}{1 - \rho}$$

Mean waiting time of all customers $E[W]$

Consider system I: queue is traversed by all customers, some have waiting time $W = 0$

- $\lambda_I = \lambda$
- $E[X_I] = E[X_W]$

$$\Rightarrow E[W] = E[T_I] = \frac{E[X_I]}{\lambda_I} = \frac{E[X_W]}{\lambda} = \frac{p_W \cdot \rho}{\lambda \cdot (1 - \rho)} \text{ (Little)}$$

Mean waiting time of waiting customers $E[W_W]$

Consider system II: queue is traversed only by waiting customers

- $\lambda_{II} = \lambda \cdot p_W$
- $E[X_{II}] = E[X_W]$

$$\Rightarrow E[W_W] = E[T_{II}] = \frac{E[X_{II}]}{\lambda_{II}} = \frac{p_W \cdot \frac{\rho}{1 - \rho}}{\lambda \cdot p_W} = \frac{1}{\lambda} \cdot \frac{\rho}{1 - \rho} \text{ (Little)}$$

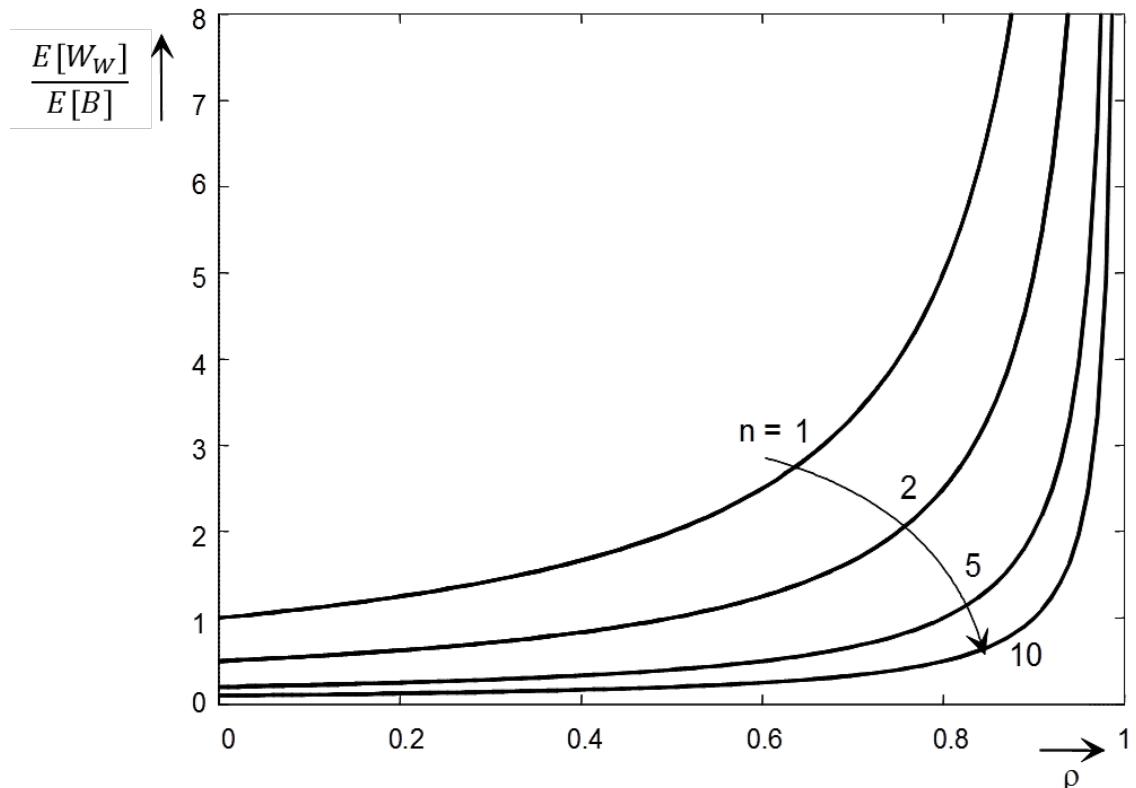


Abbildung 8: Mean waiting time of waiting customers in a $M/M/n-\infty$ queuing system (normalized by $E[B] = 1/\mu$).

Why do we get a value of $1/n$ for $\rho \rightarrow 0$ in Abbildung 8?

- It's mean waiting time only for waiting customers \Rightarrow must be larger than zero, for $\rho \rightarrow 0$ we find waiting customers just very rarely
- Service unit in a blocking $M/M/n-\infty$ system becomes available with rate $n \cdot \mu \Rightarrow$ mean time until then is $\frac{E[B]}{n}$.

Distribution function of the waiting time of

- Non-waiting customers: $P(W_{\bar{W}} \leq t) = 1$
- Waiting customers: $P(W_W \leq t) = 1 - e^{-(1-\rho) \cdot n \cdot \mu \cdot t}$
- All customers: $P(W \leq t) =$
 $(1 - p_W) \cdot P(W_{\bar{W}} \leq t) + p_W \cdot P(W_W \leq t) =$
 $(1 - p_W) \cdot 1 + p_W \cdot (1 - e^{-(1-\rho) \cdot n \cdot \mu \cdot t}) =$
 $1 - p_W \cdot e^{-(1-\rho) \cdot n \cdot \mu \cdot t}$

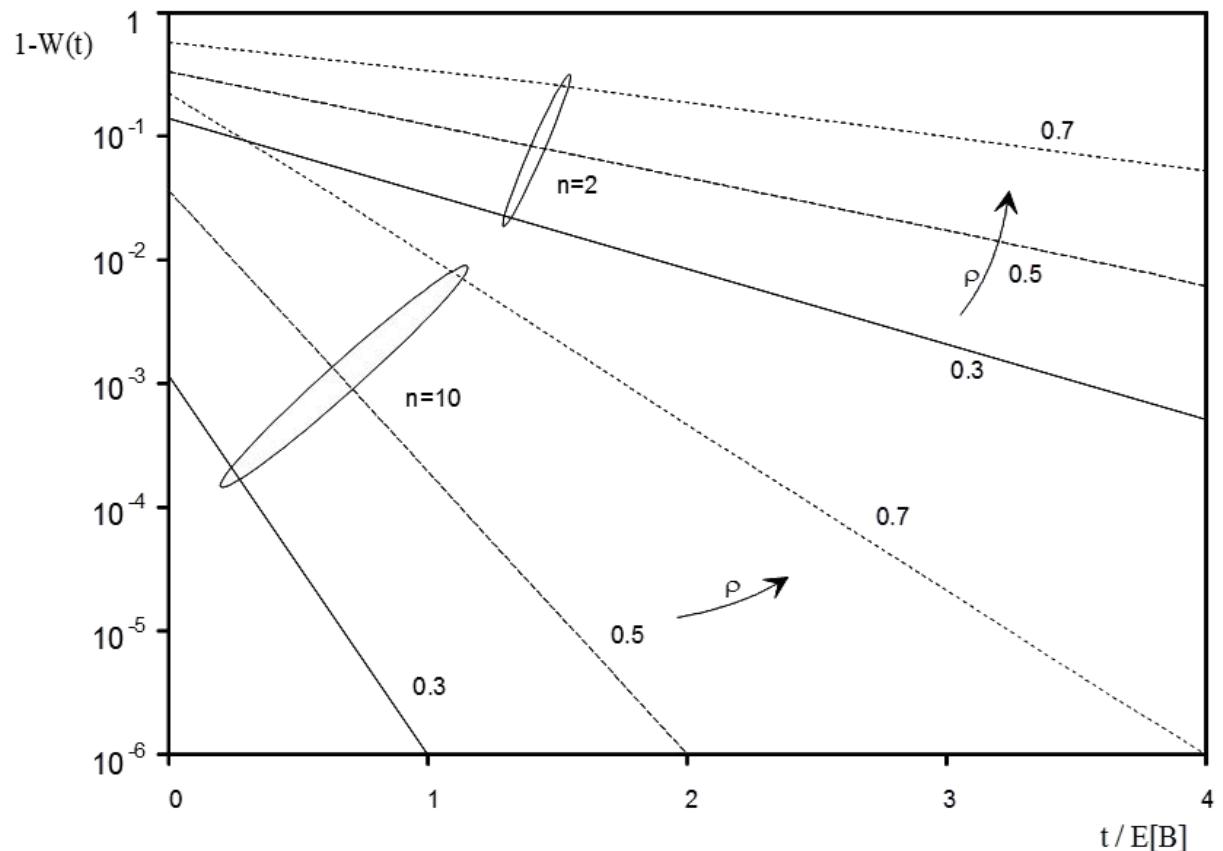


Abbildung 9: Complementary distribution function of the waiting time of all customers in an M/M/n-∞ waiting system.

Bündelungsgewinn in M/M/n-∞

- Systeme, die auf eine maximale Wartewahrscheinlichkeit bzw. Wartezeit dimensioniert sind, können für höhere angebotene Lasten wirtschaftlicher arbeiten (höhere Auslastung bei gleicher p_w bzw. $E[W]$ möglich)
- Werden die Ressourcen zweier M/M/n-∞ Systeme zu einem großen M/M/2n-∞ zusammengefasst, dann weist dieses eine kleinere Wartewahrscheinlichkeit bzw. kleinere Wartezeiten auf.

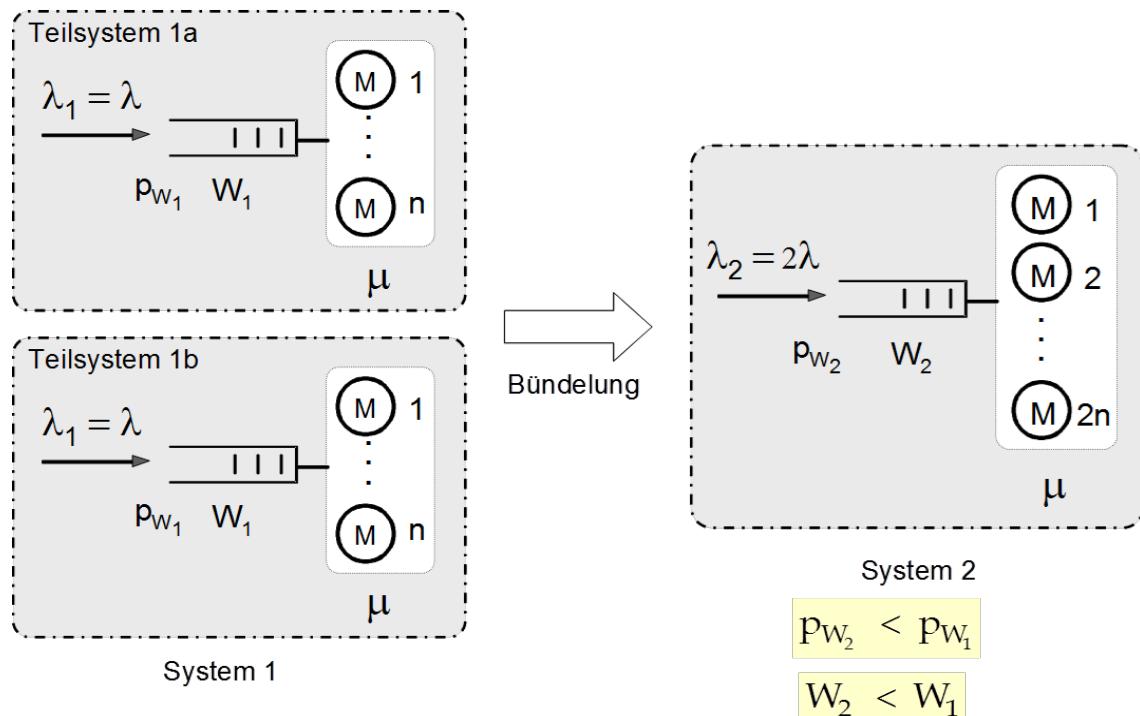


Abbildung 10: M/M/n-∞ waiting systems with more server units and the same relative load lead to lower waiting probabilities and shorter waiting times.

9.10 Waiting System M/GI/1-∞

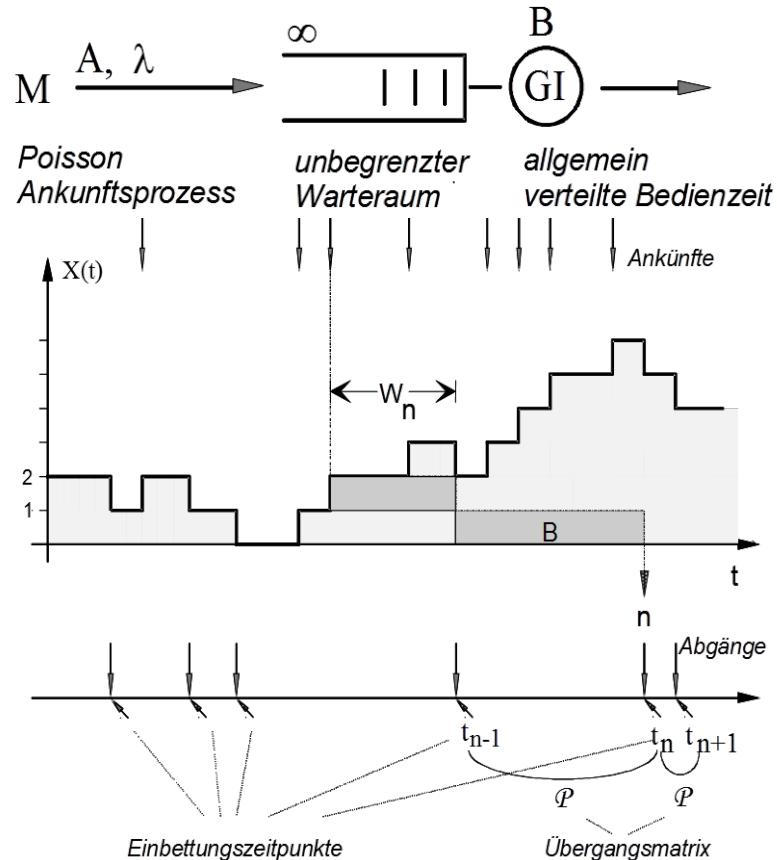


Abbildung 11: Embedded Markov chain with embedded points at service completion instants used for analysis.

Waiting probability: $p_W = \rho$

Mean waiting time of all customers: $E[W] = E[B] \cdot \frac{\rho \cdot (1 + (c_B)^2)}{2 \cdot (1 - \rho)} = \frac{\lambda \cdot E[B^2]}{2 \cdot (1 - \rho)}$

Mean waiting time of waiting customers: $E[W_W] = \frac{E[W]}{p_W} = E[B] \cdot \frac{(1 + (c_B)^2)}{2 \cdot (1 - \rho)}$ with coefficient of variation c_B of service time B

Higher moments of waiting time of all customers (Takács recursion formula):

$$E[W^0] = 1$$

$$E[W^k] = \frac{\lambda}{(1 - \rho)} \cdot \sum_{0 < i \leq k} \binom{k}{i} \cdot \frac{E[B^{i+1}]}{i + 1} \cdot E[W^{k-i}]$$

$$k = 1: \quad E[W] = \frac{\lambda \cdot E[B^2]}{2 \cdot (1 - \rho)}$$

$$k = 2: \quad E[W^2] = 2 \cdot E[W]^2 + \frac{\lambda \cdot E[B^3]}{3 \cdot (1 - \rho)}$$

ohne Formeln

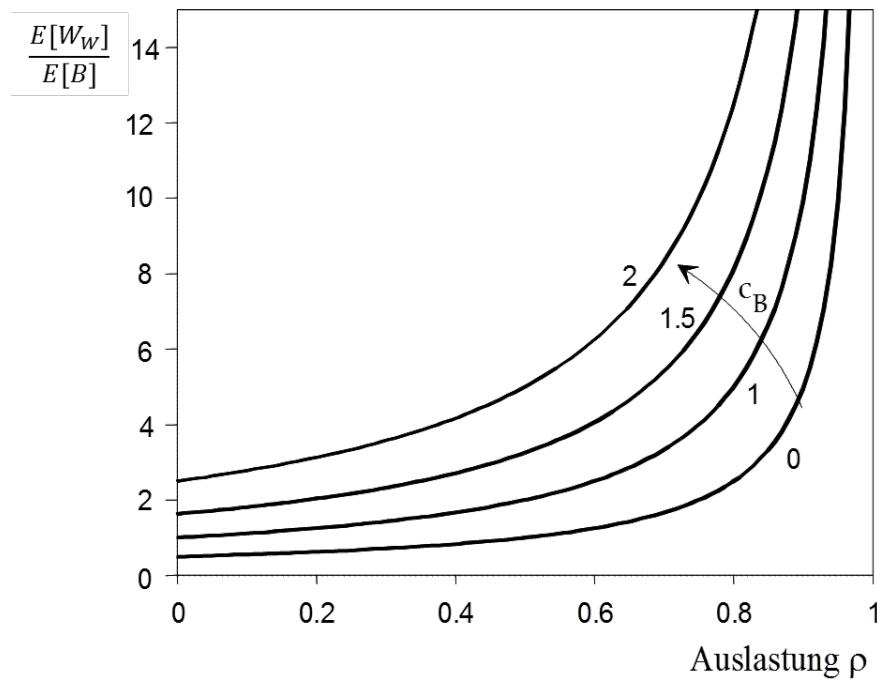


Abbildung 12: Mean waiting time of waiting customers for a $M/G/1-\infty$ waiting system normalized by $E[B]$.

The distribution function of the waiting time may be approximated by a Gamma distribution (M. Menth, R. Henjes, C. Zepfel, P. Tran-Gia: "Gamma-Approximation for the Waiting Time Distribution Function of the $M/GI/1-\infty$ Queue", in 2nd Conference on Next Generation Internet Networks (NGI), April 2006, Valencia, Spain).

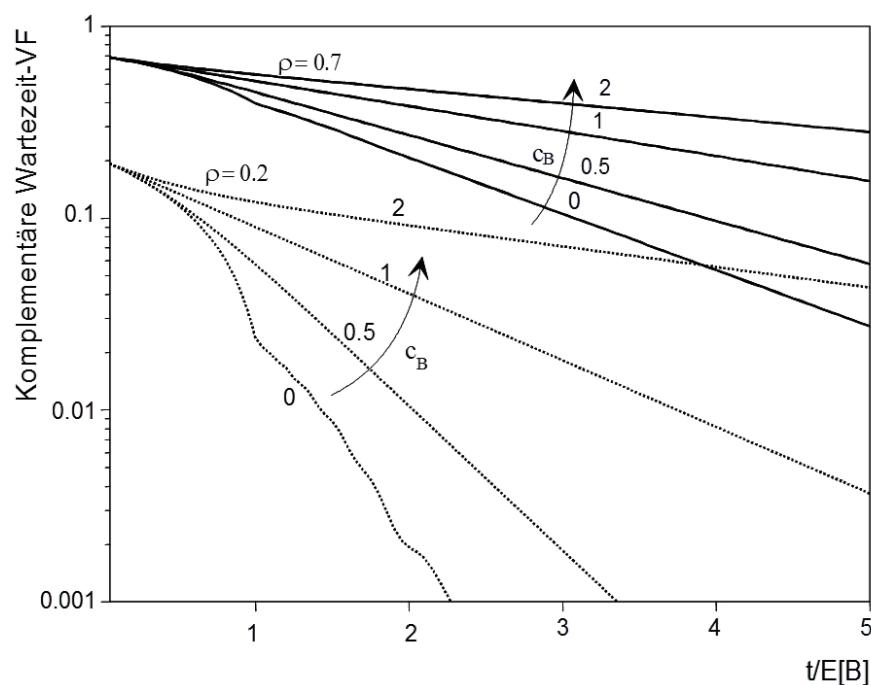


Abbildung 13: Complementary distribution function of the waiting time of all customers for a $M/G/1-\infty$ waiting system normalized by $E[B]$ (approximated by a discrete-time analysis method).

9.11 Waiting System M/D/1-∞

See: [VBI15] V.B. Iversen, "TELETRAFFIC ENGINEERING and NETWORK PLANNING", 2015,
http://orbit.dtu.dk/files/118473571/Teletraffic_34342_V_B_Iversen_2015.pdf

9.12 Waiting System n·D/D/1-∞

Approximationsformel für die Wartezeitverteilung von $n \cdot D/D/1-\infty$

- n: Anzahl der periodischen Flüsse
- A: konstante Zwischenankunftszeit ~~einer Quellen~~
aller
- B: konstante Kundenbedienzeit
- Daraus ergibt sich Systemauslastung $\rho = \frac{n \cdot B}{A}$
- $P(W > t) \approx \exp\left(\left(\frac{-2 \cdot t}{B}\right) \cdot \left(\frac{t}{n \cdot B} + 1 - \rho\right)\right)$ (aus COST 242)
- Randbedingung für die Gültigkeit der Formel
 - $\rho < 1$
 - Die Wartezeit kann nicht länger dauern als A
 ⇒ $t > A$ zu verwenden ist nicht sinnvoll
 - ⇒ relativ geringe mittlere Wartezeit sogar für sehr hohe Systemauslastungen von knapp unter 100%
- Illustration des Wartezeitverhaltens von $n \cdot D/D/1-\infty$ Wartesystemen in Abhängigkeit von Periode und Anzahl gemultiplexter Flüsse
 (Quelle: Michael Menth und Stefan Mühleck: „Packet Waiting Time for Multiplexed Periodic On/Off Streams in the Presence of Overbooking“, International Journal of Communication Networks and Distributed Systems (IJCNDS), vol. 4, no. 2, pages 207 – 229, 2010)

Das $n \cdot D/D/1-\infty$ System ist periodisch, so dass sich die Wartezeiten innerhalb einer Prozessrealisierung nach jeder Periode wiederholen. Allerdings unterscheiden sich die Wartezeiten in unterschiedlichen Prozessrealisierungen, weil sich die Phasen der Sendezeitpunkte innerhalb einer Periode (Sendemuster) unterscheiden.

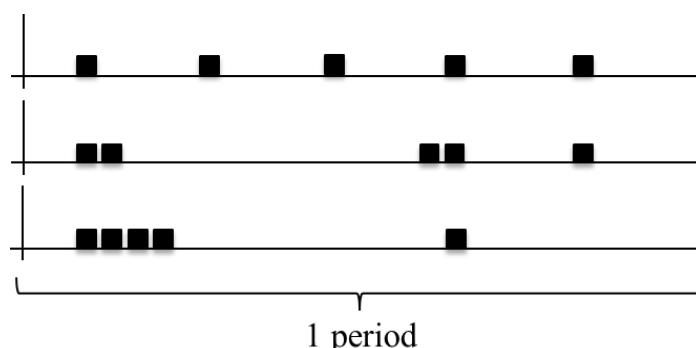


Abbildung 14: Unterschiedliche Sendemuster (Phasen der Sendezeitpunkte innerhalb einer Periode).

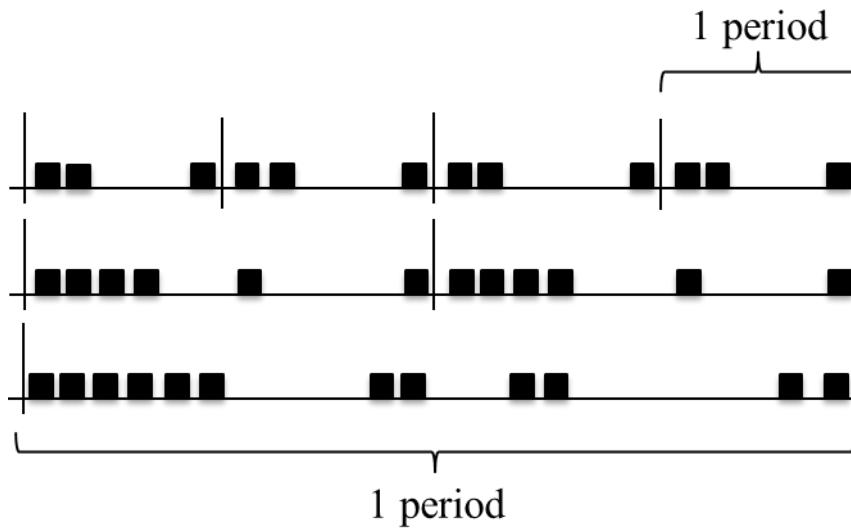


Abbildung 15: Wenn im $nD/D/1-\infty$ System die Perioden bei gleicher Last länger werden, dann senden mehr Quellen pro Periode. Es können sich größere Bursts bilden und somit längere Wartezeiten auftreten.

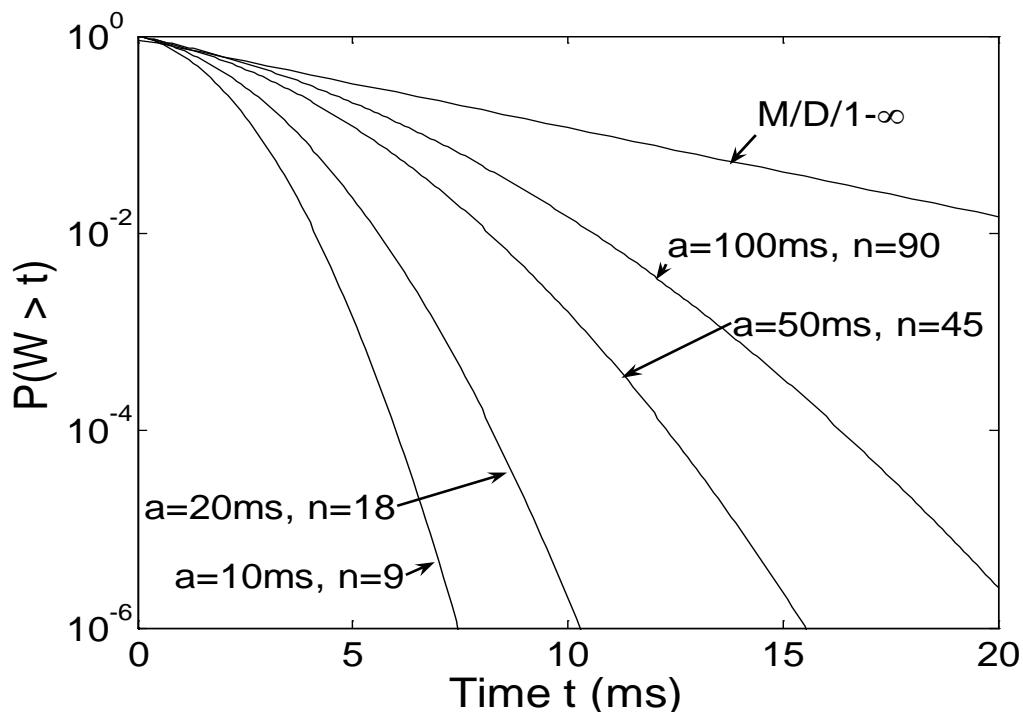


Abbildung 16: Complementary cumulative distribution function (ccdf) of the waiting time of an $n*D/D/1-\infty$ waiting system. The number of periodic sources is n . Utilization $\rho = 0.9$ and packet size $B \sim 1 \text{ ms}$ are fixed. The period a and the number of multiplexed flows vary. With increasing period, the waiting time converges to the one of an $M/D/1-\infty$ waiting system with utilization $\rho = 0.9$.

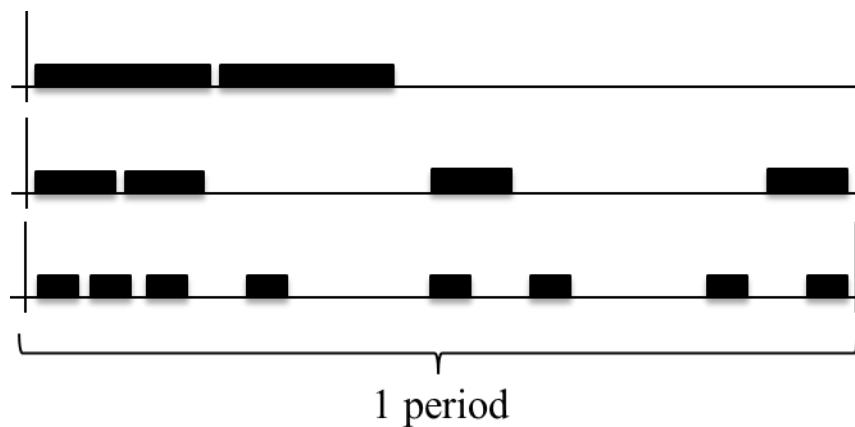


Abbildung 17: Wenn im $nD/D/1-\infty$ System die Anzahl der Sender bei gleicher Last und Periodenlänge zunimmt, werden die Pakete kürzer. Die Last verteilt sich besser über die Periode und Wartezeiten werden kürzer.

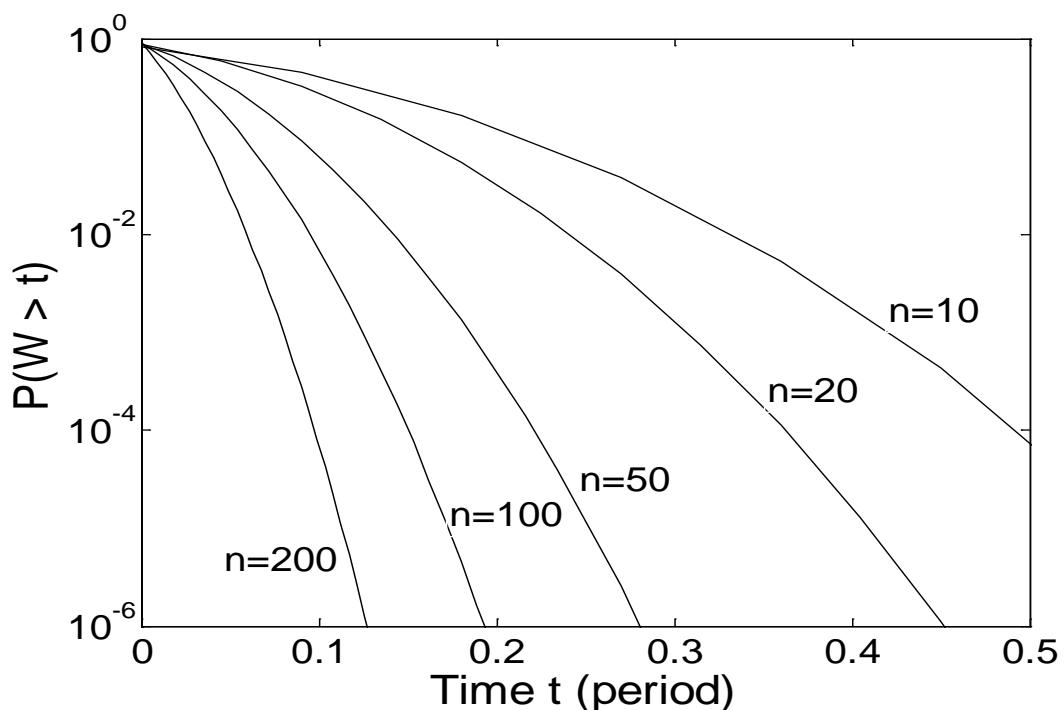


Abbildung 18: Complementary cumulative distribution function (ccdf) of the waiting time of an $n^*D/D/1-\infty$ waiting system. Utilization $\rho = 0.9$ and period a are fixed. The number of periodic sources n and the packet size B vary. With increasing number of sources, the waiting time decreases.

9.13 Multi-Rate M/GI/ ∞ Queue

- K. W. Ross and D. H. K. Tsang. The Stochastic Knapsack Problem. IEEE/ACM Transactions on Networking, 37(7):740–747, 1989
- Service class $s, 1 \leq s \leq S$ has
 - Markovian arrival rate λ_s
 - Markovian service rate μ_s
 - Offered load $a_s = \frac{\lambda_s}{\mu_s}$
 - Requires c_s service units
- S -dimensional state $x = (x_1, \dots, x_S)$ with x_s indicating the number of requests in service of type s
- Link capacity required for state x

$$c(x) = \sum_{1 \leq s \leq S} c_s \cdot x_s$$

- Computation of state probabilities

$$p(x) = \prod_{1 \leq s \leq S} \frac{(a_s)^{x_s}}{x_s!} \cdot e^{-a_s}$$

- Formulae exist for probability that less than c capacity is required $P(C \leq c)$
- Application
 - Traffic mix of small and large flows
 - 64 kb/s voice, 256 kb/s low-bitrate video, 2048 kb/s high-bitrate video
 - What's the probability for certain traffic rates on a link without admission control (blocking)?
 - M. Menth, R. Martin, J. Charzinski: "Capacity Overprovisioning for Networks with Resilience Requirements", ACM SIGCOMM, Sept. 2006, Pisa, Italy

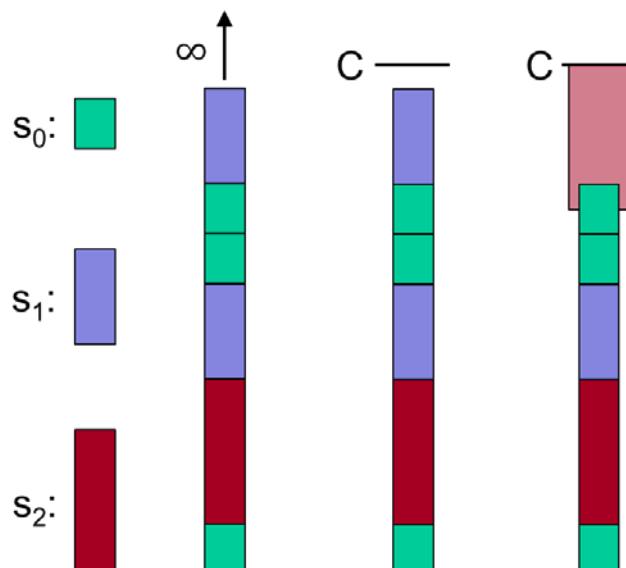


Abbildung 19: Multirate M/M/ ∞ , M/M/C, and M/M/C with trunk reservation. The rectangles represent jobs/customers and their sizes are proportional to their required number of servers.

9.14 Kaufman & Roberts Formula for Blocking Probabilities in a Multi-Dimensional Loss System ("Multirate M/M/C-0")

Problem

- Server with C service units
- S different service classes
- Service class s , $1 \leq s \leq S$ has
 - Markovian arrival rate λ_s
 - Markovian service rate μ_s
 - Offered load $a_s = \frac{\lambda_s}{\mu_s}$
 - Requires c_s service units
- New request for service s blocked if less than c_s free service units available
- What are the probabilities $x(c)$ for c busy service units?

Solution

- $x(c) = \frac{\tilde{x}(c)}{\sum_{0 \leq c \leq C} \tilde{x}(c)}$
- $\tilde{x}(c) = \begin{cases} 0 & \text{for } c < 0 \\ 1 & \text{for } c = 0 \\ \sum_{1 \leq s \leq S} a_s \cdot \frac{c_s}{c} \cdot \tilde{x}(c - c_s) & \text{for } 0 < c \leq C \end{cases}$

Application

- Flows of different rates demand for admission to a link and are blocked if remaining capacity does not suffice
- Computation of service-specific blocking probabilities $p_{B,s} = \sum_{c=c_s < c \leq C} x(c)$
- May be used for link dimensioning: how much capacity C is needed to keep blocking probabilities below a certain value?
- M. Menth: "Efficient Admission Control and Routing in Resilient Communication Networks", Doktorarbeit, Universität Würzburg, Am Hubland, 97074 Würzburg, 2004, <http://www.opus-bayern.de/uni-wuerzburg/volltexte/2004/994/pdf/Menth04.pdf>

Observation: services with larger c_s experience larger $p_{B,s}$

9.15 Multirate M/M/C-0Twith Trunk Reservation

- Improvement for more fairness: equal $p_{B,s}$ for all services
- Block new request if less than $c_{max} = \max_{1 \leq s \leq S} (c_s)$ free service units available
- Modifications of equations above lead to sufficiently accurate approximation
 - $\tilde{x}(c) = \begin{cases} 0 & \text{for } c < 0 \\ 1 & \text{for } c = 0 \\ \sum_{1 \leq s \leq S} a_s \cdot \frac{c_{TR}(C, c, c_s)}{c} \cdot \tilde{x}(c - c_s) & \text{for } 0 < c \leq C \end{cases}$
 - with $c_{TR}(C, c, c_s) = \begin{cases} c_s & \text{for } c - c_s \leq C - c_{max} \\ 0 & \text{for } c - c_s > C - c_{max} \end{cases}$
 - $p_{B,s} = \sum_{c=c_{max} < c \leq C} x(c)$