4.5 Stochastische Modelle: Analysegrundlage "Stochastische Prozesse"

Familie von ZV

Hatten wir es in Abschn. 4.4 jeweils mit einzelnen ZV zu tun, so geht es hier, bei einem stochastischen Prozeß, um eine Menge zusammengehöriger ZV, um eine "Familie" von ZV. Wir unterscheiden die Mitglieder der Familie durch einen Index t, der in einer geordneten Indexmenge T variiert. Sei also

$$(X_t)_{t \in T} \quad \text{oder auch:} \quad (X(t); t \in T)$$

diese Familie von ZV, alle als Abbildungen aus derselben Menge von Elementarereignissen auf denselben Wertebereich Ω definiert. Für jede ZV X_t, $t \in T$, ist auch ihre Verteilung (ihr Wahrscheinlichkeits-Maß) P_t erklärt. Man nennt T i.allg. die "Parametermenge" des Prozesses und Ω seinen "Zustandsraum".

Veranschaulichung

Zur Veranschaulichung:
Die "Zusammengehörigkeit" der Familie von ZV (X_t) kommt darin zum Ausdruck, daß ihre Realisierungen (x_t) als "simultan entstehend" zu verstehen sind, aus einem einzigen Elementarereignis resultieren, so daß jedes Elementarereignis σ eine Abbildung $t \rightarrow x_t(\sigma)$, eine Abbildung von T in Ω, impliziert. Dies klingt weit komplizierter als es tatsächlich ist. Machen wir uns die Zusammenhänge an einem kleinen Beispiel klar, in dem wir $\Omega = \mathbf{R}^+$, $T = \mathbf{R}^+$ wählen und uns mögliche Realisierungen des Prozesses (Realisierungen aller ZV der Familie) für zwei Elementarereignisse σ_1, σ_2 ansehen (Abb. 4.5.1). Zu jedem Elementarereignis (hier zu sowohl σ_1 als auch σ_2) gehört je eine Realisierung des stochastischen Prozesses (hier die Realisierungen $x_t(\sigma_1)$ und $x_t(\sigma_2)$). Eine Realisierung eines stochastischen Prozesses ist die gemeinsame Realisierung aller beteiligten ZV, ist eine Abbildung $T \rightarrow \Omega$ (hier: $\mathbf{R}^+ \rightarrow \mathbf{R}^+$). Die häufigste Interpretation der Parametermenge T ist "Zeit" - wir werden ausschließlich diese Interpretation verwenden. In dieser Interpretation ist eine Realisierung eines stochastischen Prozesses der Verlauf seines Zustands (hier: Zustand aus $\mathbf{R}^+$) über der Zeit, auch "Trajektorie" genannt.

Trajektorien

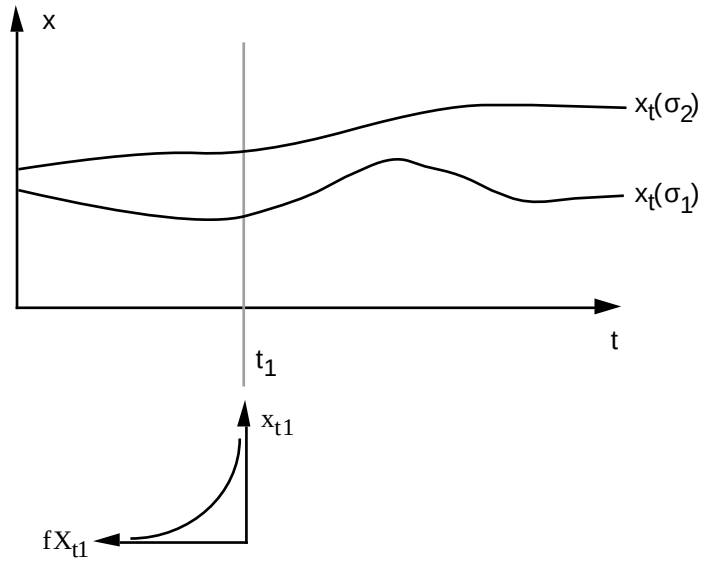


Abbildung 4.5.1: Trajektorien eines stochastischen Prozesses
und Zustandsverteilung zu festem Zeitpunkt

Für festen Beobachtungszeitpunkt t (in Abb. 4.5.1 für t_1 markiert) erhalten wir (sozusagen "quer" über alle Trajektorien, d.h. als Eindruck über alle möglichen

Zustandsverläufe, deren jeder ja genau einem Elementarereignis zugeordnet ist) die Zufallsvariable X_{t1}, deren mögliche Verteilung durch eine Skizze ihrer Dichtefunktion charakterisiert ist. Für einen anderen Zeitpunkt $t_2 \neq t_1$ ist dementsprechend eine andere ZV, nämlich X_{t2} maßgebend, mit einer (i. allg.) anderen Verteilung. X_{t1} und X_{t2} sollten allerdings (i. allg.) voneinander abhängig sein: Einem Elementarereignis ist ja eine gesamte Trajektorie zugeordnet, insbesondere also "zusammengehörige" Zustandswerte für t_1 und t_2.

Klassifizierungen stochastischer Prozesse

In noch stärkerem Maße als bei den ZV unterdrückt man bei praktischen Anwendungen mit Vorliebe den Bezug auf den Raum der Elementarereignisse und bewegt sich völlig im Zustandsraum Ω. Klassifizierungen von stochastischen Prozessen erfolgen bezüglich

(a) der Natur des Zustandsraums Ω,
(b) der Natur der Parametermenge T,
(c) der Art der gegenseitigen Abhängigkeiten der Zufallsvariablen $(X_t)_{t \in T}$

Zustandsraum

zu (a): Von uns verwendete Zustandsräume sind ausschließlich numerisch, oft eindimensional (gelegentlich mehrdimensional). Ist Ω eine diskrete, (endliche oder abzählbare) Menge, dann heißt der Prozeß "zustandsdiskret", sonst "zustandskontinuierlich". Häufig auftretende Fälle sind

$\Omega = \mathbf{N_0}$ (zustandsdiskret)
$\Omega = \mathbf{R}^+$ (zustandskontinuierlich)

Stochastische Ketten

Zustandsdiskrete stochastische Prozesse heißen auch "stochastische Ketten".

Parametermenge

zu (b): Bei diskreter Parametermenge (Interpretation: Zeit) nennt man den Prozeß zeitdiskret, sonst zeitkontinuierlich. Häufig auftretende Fälle sind wieder

$T = \mathbf{N_0}$ (zeitdiskret)
$T = \mathbf{R}^+$ (zeitkontinuierlich)

Stochastische Folgen

Zeitdiskrete stochastische Prozesse nennt man auch "stochastische Folgen".

Testfrage

Testfrage 4.5.2: Skizzieren Sie (analog Abb. 4.5.1) das prinzipielle Aussehen von stochastischen Prozessen in den vier interessanten Sonderfällen

$(\Omega,T) \quad \{(\mathbf{N_0},\mathbf{N_0}),(\mathbf{N_0},\mathbf{R}^+),(\mathbf{R}^+,\mathbf{N_0}),(\mathbf{R}^+,\mathbf{R}^+)\}$

Abhängigkeiten

zu (c): Die gegenseitigen Abhängigkeiten einer Menge von ZV sind ja durch ihre gemeinsame Verteilung erfaßbar. Interessieren wir uns beispielsweise für eine bestimmte abzählbare Untermenge von Zeitpunkten $\tau \subset T$ und die damit fixierte Menge von ZV

$$\{X_{ti}: t_i \in \tau\}$$

dann ist, formal ausgedrückt, die gemeinsame Verteilung der mehrdimensionalen ZV

$$\underline{X}(\underline{t}_\tau) = (X(t_1), X(t_2), \ldots, X(t_r)) \qquad r := |\tau|$$

gefragt. Diese ist beispielsweise im Falle eines angenommenen Zustandsraums $\Omega = \mathbf{R}$ eindeutig charakterisiert durch die gemeinsame Verteilungsfunktion

$$F_{\underline{X}}(\underline{t}_\tau; \underline{x}_\tau) = P[X(t_1) \leq x_1, X(t_2) \leq x_2, \ldots, X(t_r) \leq x_r]$$

Vollständige Information über den stochastischen Prozeß läge also vor, wenn diese Verteilungsfunktion für τ=T bzw. in unserer Illustration für jedes abzählbare τ T bekannt wäre. Klassifizierungen dieser an sich überwältigenden Informationsmenge müssen natürlicherweise vergleichsweise "kompakte" Abhängigkeitscharakterisierungen bemühen.

Einige (sehr unterschiedliche) Abhängigkeitscharakterisierungen:

Unabhängige Prozesse

Definition 4.5.3: Ein Grenzfall liegt offensichtlich vor, wenn keinerlei Abhängigkeiten bestehen. Wir haben dann eine Familie unabhängiger ZV vorliegen, so daß für abzählbares τ T

$$F_{\underline{X}}(\underline{t}_\tau ; \underline{x}_\tau) = \prod_{t \ \tau} F_{X_t}(x_t)$$

gilt. Solche (weitgehend uninteressanten) Prozesse nennt man **unabhängige Prozesse.**

Stationäre Prozesse

Definition 4.5.4: Eine völlig andersartige Eigenschaft von Prozessen liegt vor, wenn die Abhängigkeiten "sich über der Zeit nicht ändern". Erfaßt wird dies durch das Postulat einer Invarianz der Abhängigkeiten bzgl. Zeitverschiebungen, formal:

Für alle τ T, und alle s>0

(wir setzen einen unserer Sonderfälle entsprechend Testfrage 4.5.2 voraus, so daß mit t τ und sinnvollem s auch t+s τ gilt)

gilt: $F_{\underline{X}}(\underline{t}_\tau;\underline{x}_\tau) = F_{\underline{X}}(\underline{t}_\tau+\underline{s};\underline{x}_\tau)$ wo $\underline{s}=(s,s,...,s)^T$

Solche (äußerst interessanten) Prozesse heißen **stationäre Prozesse.** Insbesondere folgt unmittelbar (für $|\tau|=1$), daß alle ZV $X(t_i)$, t_i τ, identisch verteilt sind (aber durchaus nicht, daß sie unabhängig verteilt wären).

Markoff-Prozesse

Definition 4.5.5: Eine wiederum andere Charakterisierungsart basiert auf der Idee, die Zukunft eines Prozesses (also die ZV X(t) für alle t größer als ein bestimmtes t*) aus einer bekannten Vergangenheit (also aus den Realisierungen x(t) für alle t≤t*) "vorauszusagen". Ein Spezialfall dieser Art ist gegeben, wenn die Zukunft nur derart von der Vergangenheit beeinflußt wird, daß der gegenwärtige Zustand (also x(t*) bei obigen Bezeichnungen) bereits ausreicht, um die Zukunft zu charakterisieren. Formal drückt sich dieser Spezialfall folgendermaßen aus:

Für alle $\tau = \{t_i; i \ I\}$ T
mit $t^* = \max_{\tau}(t_i)$

und beliebiges t T mit t>t*

gilt: $F_X(t \mid \underline{X}(\underline{t}_\tau)=\underline{x}_\tau) = F_X(t \mid X(t^*)=x^*)$

In Worten: Die Verteilung von X(t) (zu "zukünftigem" Zeitpunkt t) unter der Bedingung, daß die Realisierung des Prozesses bis zum Zeitpunkt t* gegeben ist (gewisse "Vergangenheit" zu Zeitpunkten aus τ), ist bereits durch die bezüglich der Realisierung zum Zeitpunkt t* ("Gegenwart") bedingte Verteilung von X(t)

gegeben. Diese (im folgenden nahezu ausschließlich betrachteten) Prozesse heissen **gedächtnislos** bzw. **Markoff-Prozesse.**

Testfrage

Testfrage 4.5.6: Als stochastische Prozesse lassen sich auch auffassen und beschreiben die folgenden bekannten Gegebenheiten:

- die Gesetze der Stationswechsel, im Anschluß an den Besuch einer bestimmten Station i eines Verkehrsnetzes, auf seiten der Sequenz der diese Station verlassenden Kunden; vgl. (4.4.1);
- die Gesetze der Ankunftsabstände, am offenen Netz, auf seiten der Sequenz eintreffender Kunden; vgl. (4.4.2);
- die Gesetze der Bedienwunsch-Umfänge, an einer Station i, auf seiten der Sequenz der diese Station betretenden Kunden; vgl. Spezifikation Bedienwünsche in Abschn. 4.4.

Beschreiben Sie diese Gesetzmöglichkeiten als stochastische Prozesse und charakterisieren Sie diese Prozesse nach Zustandsraum, Parametermenge und Art der Abhängigkeiten innerhalb der ZV des Prozesses.

Testfrage

Testfrage 4.5.7: Treten Ereignisse mit gewissen (regelmäßigen, unregelmäßigen, zufälligen) zeitlichen Abständen auf, dann nennt man den stochastischen Prozeß $(N_t;t \in \mathbf{R}^+)$, der die Anzahl in (0,t) aufgetretener Ereignisse charakterisiert, einen "Zählprozeß". Diskutieren Sie den speziellen Zählprozeß, Poisson-Prozeß genannt, der die Ankunftszahl mit u.i.v., exponentiellen Zwischenankunftsabständen eintreffender Kunden beschreibt; vgl. (4.4.2, 4.4.3, 4.4.11).

zeitkontinuierliche Markoff-Ketten

Nachdem wir uns in den vergangenen Abschnitten die Begriffsbildung "stochastischer Prozeß" im allgemeinen in Erinnerung gerufen haben, wollen wir uns im folgenden auf zeitkontinuierliche Markoff-Ketten konzentrieren und (sehr verkürzt) Resultate und Techniken zusammenstellen, die uns ein Arbeiten mit dieser Klasse von Prozessen erlaubt. Zeitkontinuierliche stochastische Ketten besitzen einen kontinuierlichen Parameterraum sowie einen diskreten Zustandsraum; recht typisch für Anwendungen ist die Wahl $(\Omega,T)=(\mathbf{N_0},\mathbf{R}^+)$. Die geforderte Markoff-Eigenschaft drückt sich hier formal so aus, daß

(4.5.8) Für alle $n \geq 1$
und für alle Folgen $t_1<t_2<...<t_n<t_{n+1}$
gilt:

$$P[X(t_{n+1})=i_{n+1} \mid X(t_1)=i_1,...,X(t_n)=i_n] = P[X(t_{n+1})=i_{n+1} \mid X(t_n)=i_n]$$

wo $t_i \in \mathbf{R}^+$; $i_j \in \mathbf{N_0}$, $j=1,2,...,n+1$
vorausgesetzt ist.

Der Prozeßzustand zu festem Zeitpunkt t ist durch die Verteilung der ZV X(t) beschrieben, die sich im vorliegenden Falle durch die Zustandswahrscheinlichkeiten

(4.5.9)
$$p_i(t) := P[X(t)=i] \quad i \in \mathbf{N_0}$$
$$\sum_i p_i(t) = 1$$

charakterisieren läßt. Die Zustandsverteilungen zu zwei unterschiedlichen Zeitpunkten t_1,t_2 (wo $t_1<t_2$ sei) sind voneinander abhängig. Liegt z.B. zum Zeitpunkt t_1 Zustand i vor, dann wird die Wahrscheinlichkeit dafür, welcher Zustand zum

Zeitpunkt t_2 eingenommen wird, von diesem Vor-Zustand i abhängen. Mit der Charakterisierung dieser Abhängigkeit ist aber wegen der Markoff-Eigenschaft des Prozesses die Abhängigkeit von der gesamten "Vorgeschichte" (bis zu t_1) voll erfaßt. Es ist somit höchst sinnvoll, die Abhängigkeitsstruktur durch sog. (Zustands-)Übergangswahrscheinlichkeiten gemäß

Übergangswahrscheinlichkeiten

(4.5.10) $\pi_{ij}(t_1,t_2) := P[X(t_2)=j \mid X(t_1)=i] \qquad t_2>t_1;\ t_1,t_2 \in \mathbf{R}^+;\ i,j \in \mathbf{N_0}$

$$\sum_j \pi_{ij}(t_1,t_2) = 1 \quad i \in \mathbf{N}_0$$

zu beschreiben. Die Abhängigkeit der Zustandsverteilungen zu den beiden Zeitpunkten ergibt sich mit dieser Definition zu

Zustandswahrscheinlichkeiten

(4.5.11) $p_j(t_2) = \sum_{i \in \mathbf{N}_0} p_i(t_1)\ \pi_{ij}(t_1,t_2) \quad t_2>t_1, j \in \mathbf{N}_0$

In den allgemeinen Übergangswahrscheinlichkeiten nach (4.5.10) steckt noch die Vorstellung, daß sich die Abhängigkeitsstruktur über der Zeit (der Interpretation des Parameterraums T) ändern kann, also für jedes Paar (t_1,t_2) spezifisch ist. Ein sehr viel einfacherer Fall liegt vor, wenn diese Abhängigkeitsstruktur sich über der Zeit nicht verändert. Dies ist gleichzeitig ein praktisch sehr relevanter Fall, denn (wenn wir uns kurz an den eigentlichen Zweck unserer derzeitigen Überlegungen erinnern) wir sind ja an der Erfassung der Dynamik eines arbeitenden RS interessiert, die sich mit Fortschreiten der absoluten Zeit nicht ändern sollte (wenigstens ist dies eine vernünftige Annahme). Konstanz der Abhängigkeitsstruktur über der Zeit bedeutet aber, daß es bei dem betrachteten Paar von Zeitpunkten (t_1,t_2) auf die absolute Lage von t_1 nicht ankommt, sondern nur noch auf die Differenz t_2-t_1. Ein derart beschaffener Markoff-Prozeß heißt (zeit-)homogen; seine Abhängigkeitsstruktur ist erfaßt durch Übergangswahrscheinlichkeiten (Umbenennung!)

Zeithomogene Prozesse

(4.5.12a) $\pi_{ij}(s) := P[X(t+s)=j \mid X(t)=i] \qquad s>0;\ i,j \in \mathbf{N_0}$

die vom Wert von t unabhängig sind, mit anderen Worten für jedes t den gleichen Wert annehmen. Sicher muß auch

(4.5.12b) $$\sum_j \pi_{ij}(s) = 1 \quad i \in \mathbf{N}_0$$

gelten. Für homogene Prozesse wird aus (4.5.11)

(4.5.13a) $p_j(t+s) = \sum_{i \in \mathbf{N}_0} p_i(t)\ \pi_{ij}(s) \quad t,s>0; j \in \mathbf{N}_0$

Zur Schreibvereinfachung schreiben wir die letzte Gleichung nochmals in Vektor- / Matrix-Form auf und erhalten mit

$$\pi(s) = (\pi_{ij}(s); i,j \in \mathbf{N_0})$$

als Matrix der Übergangswahrscheinlichkeiten und mit

$$\underline{p}(t) = (p_j(t); j \in \mathbf{N_0})^T$$

als Spaltenvektor der Zustandswahrscheinlichkeiten aus (4.5.13a) die kompaktere Form

(4.5.13b) $\underline{p}^T(t+s) = \underline{p}^T(t) \cdot \pi(s)$

Wir könnten uns nun die Frage stellen, ob unsere Voraussetzungen ausreichen, um für die Übergangswahrscheinlichkeiten $\pi(s)$ eine explizite Form anzugeben, so daß für beliebige Start-Zustandsverteilungen $\underline{p}(0)$ des Prozesses sich die (zeitabhängigen) Zustandsverteilungen für die ZV X(t) gemäß

(4.5.14) $\underline{p}^T(t) = \underline{p}^T(0) \cdot \pi(t)$

angeben ließen. Dieser Weg läßt sich in der Tat erfolgreich beschreiten, führt allerdings zu expliziten Formen für (4.5.14), die für unsere praktisch motivierten Ziele wegen ihrer Komplexität nicht unmittelbar hilfreich sind (Neugierige seien erneut auf Cinl75 verwiesen).

Vorüberlegungen Grenzverteilung

Wir wollen statt dessen folgende, im Moment für uns tragfähigere Idee verfolgen: (4.5.14) drückt aus, daß die Zustandsverteilung zu einem Zeitpunkt t im Prinzip von der Start-Zustandsverteilung $\underline{p}(0)$ abhängt, bzw. vom Start-Zustand, wenn es einen ausgezeichneten derartigen gibt (bei festem Startzustand i wäre die Startverteilung durch $p_i(0)=1$ und $p_j(0)=0$, $j \neq i$, beschrieben). Es scheint vernünftig zu vermuten, daß diese Abhängigkeit des Zustands (zum Zeitpunkt t) vom Startzustand "mit der Zeit" (d.h. für große t) abnimmt und evtl. völlig verschwindet - zumindest unter gewissen Voraussetzungen (d.h. zumindest für gewisse zu modellierende Systeme). Gesetzt den Fall, diese Vermutung träfe zu, dann sollten alle Übergangswahrscheinlichkeiten $\pi_{ij}(t)$, $i \in \mathbf{N_0}$, für große t zum selben Wert konvergieren, d.h. es sollte

(4.5.15a) $\lim_{t \to \infty} \pi_{ij}(t) \equiv p_j \quad j \in \mathbf{N}_0$

gelten. Die Matrix $\pi(t)$ strebte, mit anderen Worten, für große t einer Matrix π mit untereinander identischen Zeilen zu, so daß mit Blick auf (4.5.14) auch

(4.5.15b) $\lim_{t \to \infty} p_j(t) = p_j \quad j \in \mathbf{N}_0$

gelten sollte. Unter zusätzlicher Beachtung der vorausgesetzten Homogenitätsbedingung (4.5.12) näherte sich der Prozeß demnach für große t einem "stationären Verhalten" (man konsultiere Def. 4.5.4).

Bestimmung Grenzverteilung

Weiterhin unter der Hypothese, es existiere eine Grenzverteilung $\underline{p}=(p_j)$ gemäß (4.5.15), dann ließe sie sich wie folgt bestimmen: (4.5.13) liefert

$$\underline{p}^T(t+s) - \underline{p}^T(t) = \underline{p}^T(t) \cdot (\pi(s) - E)$$

mit E als Einheitsmatrix passender Dimension, woraus man weiter

$$\frac{\underline{p}^T(t+s) - \underline{p}^T(t)}{s} = \underline{p}^T(t)\ \frac{\pi(s) - E}{s}$$

erhält und im Grenzübergang ($s \to 0$)

(4.5.16a) $$\frac{d\underline{p}^T(t)}{dt} = \underline{p}^T(t)\ \lim_{s \to 0} \frac{\pi(s)-E}{s}$$
$$= \underline{p}^T(t)\ Q$$

wo die Matrix $Q=(q_{ij}),i,j \in \mathbf{N_0}$ Elemente enthält gemäß

(4.5.16b) $$q_{ij} = \lim_{s \to 0} \left[\pi_{ij}(s) / s\right] \qquad i \neq j$$
$$q_{ii} = \lim_{s \to 0} \left[(\pi_{ii}(s) - 1) / s\right]$$

Im Rahmen unserer Vermutungen sollte für große t das Prozeßverhalten gegen ein stationäres Verhalten konvergieren ($\underline{p}(t)$ de facto zeitunabhängig und $d\underline{p}(t)/dt$ de facto verschwindend), also aus (4.5.16) für hinreichend große t

(4.5.17a) $$\underline{0}^T = \underline{p}^T Q$$

mit (Verteilung!):

(4.5.17b) $$\sum_{i \in \mathbf{N}_0} p_i = 1$$

folgen. Das lineare Gleichungssystem (4.5.17) wäre dann auch direkt nutzbar, um aus Q (der Matrix, deren Elemente alle notwendigen Informationen über die Charakteristika eines speziellen Prozesses, also die Charakteristika eines konkreten Problems enthalten) die "stationäre Verteilung" $\underline{p}$ des Prozesses zu bestimmen, d.h. die Wahrscheinlichkeiten $p_j, j \in \mathbf{N_0}$, den Prozeß für hinreichend große t in einem seiner Zustände j anzutreffen (unabhängig vom konkreten Zeitpunkt t und unabhängig vom Startzustand des Prozesses).

Unsere Überlegungen gingen von der Prämisse aus, daß die stationäre Verteilung eines Prozesses gemäß (4.5.15) existiert und von der Vermutung, daß diese Prämisse zumindest "unter gewissen Voraussetzungen" Gültigkeit besitzen sollte. Anstatt diesen Voraussetzungen nachzuspüren und daraus konstruktiv die Existenz einer Grenzverteilung nach (4.5.15) bzw. (4.5.17) abzuleiten, möge uns für unsere praktischen Arbeiten der im folgenden ohne Beweis angegebene Satz ausreichen (Wißbegierige: s. Cinl75).

Irreduzibilität; Existenz und Eindeutigkeit stationärer Verteilung

Satz 4.5.18 (o.B.): Eine zeitkontinuierliche homogene Markoff- Kette heißt irreduzibel, wenn jeder Zustand von jedem Zustand "erreichbar" ist, d.h. wenn es für alle Zustandspaare (i,j) Werte $t>0$ gibt, so daß $\pi_{ij}(t)>0$. Eine irreduzible (zeitkontinuierliche, homogene) Markoff-Kette besitzt eine stationäre Grenzverteilung gemäß (4.5.17) genau dann, wenn (4.5.17) eine Lösung besitzt. Diese Lösung ist eindeutig.

Interpretation

Dieser Satz ist von erheblicher praktischer Bedeutung:

- Wenn wir uns ausschließlich für das (potentiell existierende) "stabile" Verhalten einer Markoff-Kette (eines Systems!) - nach Abklingen von vorübergehenden Einflüssen des Anfangszustands - interessieren (also für die stationäre Grenzverteilung, auch "eingeschwungene" oder "Gleichgewichts-"Phase genannt im Unterschied zur anfänglichen "transienten" Phase),
- und wenn wir mit einer (zeit-)homogenen Kette arbeiten (d.h. mit einer, deren Zustandsübergangswahrscheinlichkeiten entsprechend (4.5.12) allein von den Differenzen beobachteter Zeitpunkte abhängen), wo sich die Homogenität i. allg. direkt dem Problem entnehmen läßt,
- und wenn wir eine irreduzible Kette vor uns haben (d.h. eine, deren Zustände alle "wesentlich" sind in dem Sinne, daß sie immer wieder eingenommen werden, wie lange auch das System arbeitet), wo sich die Irreduzibilität i.allg. direkt dem Problem ansehen läßt;
- dann brauchen wir lediglich das Gleichungssystem (4.5.17) aufstellen und nach $\underline{p}$ zu lösen versuchen: Gelingt dies, dann sind die "gewissen Voraussetzungen" für die Existenz einer stationären Grenzverteilung erfüllt, und wir haben sie eindeutig als Lösung von (4.5.17) ermittelt.

Probleme der Durchführung

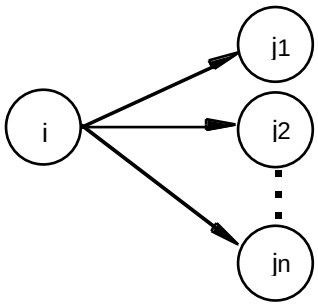
Die praktische Durchführung allerdings ist nicht ohne Dornen:

- Wir müssen erst noch lernen, (4.5.17) aufzustellen, i.w. also aus einem gegebenen Problem die Matrix Q (die sämtliche Problem-Charakteristika erfaßt) abzuleiten.
- Wir müssen weiterhin (4.5.17) lösen; dies ist zwar "nur" die Lösung eines linearen Gleichungssystems, allerdings eines mit potentiell sehr hoher (endlicher) oder sogar abzählbarer Anzahl von Variablen (=Mächtigkeit des Zustandsraums).

Aufstellung Q-Matrix

Wenden wir uns zunächst der Aufstellung der Q-Matrix zu, der Aufgabe also, "aus dem Problem" die Elemente der Q-Matrix abzulesen. Stellen wir uns dazu vor, das System sei (zum Zeitpunkt t) im Zustand i (einem bestimmten seiner diskreten Zustände). Voraussetzung für den Einsatz unserer Ergebnisse ist, daß das System "zeithomogen" ist, daß also aus der Kenntnis "Zustand i zum Zeitpunkt t", unabhängig von t, die Aussage abgeleitet werden kann, in welchem Zustand das System sich s Zeiteinheiten später befindet (bzw. die Aussage, mit welcher Wahrscheinlichkeit es sich zum Zeitpunkt t+s in welchem seiner Zustände befindet). Es war Aufgabe unserer Übergangswahrscheinlichkeiten $\pi_{ij}(s)$, vgl. (4.5.12), diese Entwicklung über der Zeit zu erfassen; ferner hatten wir in der zunächst rein formalen Definition der Elemente q_{ij} - s.(4.5.16b) - diese Entwicklung für die "unmittelbar bevorstehende" Zukunft ($s \rightarrow 0$) charakterisiert. Bei dieser unmittelbar bevorstehenden Zukunft werden nur die unmittelbaren Folgezustände zum Zustand i ins Spiel kommen, jene, die ohne Zwischenzustand von i erreichbar sind (jedenfalls, wenn wir ungesunde Verhältnisse wie "beliebig viele Zustandsübergänge in beliebig kurzer Zeit" als praxisfern ausschließen). Seien diese unmittelbaren Folgezustände zu i mit $j_1,j_2,...,j_n$ bezeichnet, dann konzentriert sich unser Interesse auf folgenden (Ausschnitt eines) "Zustandsübergangsgraphen":

Zustands-übergangs-graph



Knoten: Zustände
Kanten: mögliche unmittelbare Zustandsübergänge

einfaches Beispiel

Machen wir uns die Sache für den Moment noch etwas einfacher mit der Annahme, es gebe überhaupt nur einen unmittelbaren Folgezustand zum Zustand i, sagen wir: j_1. $\pi_{i,j1}(s)$ ist die Wahrscheinlichkeit, unter der Bedingung, zum Zeitpunkt t Zustand i vorgefunden zu haben, zum Zeitpunkt t+s Zustand j_1 einzunehmen. Der Zusammenhang zwischen $\pi_{i,j1}(s)$ und $q_{i,j1}$ läßt sich auf der Basis der Definitionsgleichung (4.5.16b) und einer (gedachten) Reihenentwicklung für $\pi_{i,j1}(s)$ auch schreiben als

(4.5.19) $\pi_{i,j1}(s) = q_{i,j1} \cdot s + o(s)$

Zustands-Übergangsrate

$q_{i,j1}$ spielt also die Rolle einer "Übergangsintensität", mit der in erster Näherung (proportional zu s) der Zustand i dem Zustand j_1 "zustrebt", die Rolle einer "Zustandsübergangsrate" im Sinne unserer Diskussion von Raten im Anschluß an (4.4.14). Zur Abrundung unseres Spezialfalls: Alle q_{ij} mit $j \neq j_1$ (also die Übergangsraten bzgl. der nicht unmittelbar möglichen Folgezustände) sollten den Wert 0 haben, daher die zugeordneten $\pi_{ij}(s)=o(s)$ sein, die Zustände $j \neq j_1$ in erster Näherung (und für kleine s de facto) zum Zeitpunkt t+s nicht auftauchen. Auf das Hauptdiagonalelement q_{ii} kommen wir separat zu sprechen.

Problem und Q-Elemente

Drehen wir den Spieß um und fragen vom Problem her: Wie kommt es zu Elementen $q_{ij} \neq 0$ und welche Werte haben diese Elemente? Wir bleiben bei unserem Spezialfall genau eines unmittelbaren Folgezustands j_1. Wir könnten diesen Spezialfall etwa vorliegen haben für ein (Modell eines) Rechensystem(s), das im Zustand i gar nichts zu tun hat ("leer" ist, keine unbearbeiteten Aufträge mehr kennt) und das in diesem Zustand solange verharrt, bis ein neuer Auftrag eintrifft. Der einzige mögliche Folgezustand zu i ist damit der Zustand j_1, der nach einem solchen Auftragseingang eingenommen wird (implizite zusätzliche Voraussetzungen: Aufträge treffen einzeln ein, es gibt nur eine Art von Aufträgen). Um die Verhältnisse von (4.5.19) in unserem Beispiel zu reproduzieren, müßte also gelten:

einfaches Beispiel

$$\begin{aligned}\pi_{i,j1}(s) &= q_{i,j1} \cdot s + o(s) \\ &= P[\text{bei eingenommenem Zustand i trifft innerhalb der nächsten s Zeiteinheiten genau ein Auftrag ein}]\end{aligned}$$

Dies unabhängig vom Zeitpunkt, zu dem Zustand i festgestellt wurde! Was, wenn Sie kurz unsere Diskussion nach (4.4.5) nachlesen, nur im Falle einer exponentiell verteilten "Zeit bis zur nächsten Ankunft" möglich ist. Hinreichend für die geforderten Verhältnisse ist beispielsweise ein Poisson'scher Ankunftsstrom gemäß (4.4.14), für den (angenommener Parameter: λ) ja unabhängig vom Systemzustand (hier: i) gilt, daß

$$P[\text{genau eine Ankunft innerhalb s Zeiteinheiten}] = \lambda \cdot s + o(s)$$

Vergleichen wir mit oben: Das fragliche $q_{i,j1}$ hat den Wert λ .

weiteres Beispiel

Dies war die Diskussion eines Spezialfalls. Werden wir etwas allgemeiner, indem wir uns vorstellen, daß zum Zustand i zwei unmittelbare Folgezustände (j_1 und j_2) existieren, also Matrixelemente $q_{i,j1},\neq 0$ und $q_{i,j2}\neq 0$ ins Spiel kommen (während alle anderen q_{ij}, mit $j \neq j_1,j_2$, die Werte 0 behalten). Ein zugehöriges Problem könnte so geartet sein, daß im Zustand i (in dem im bisherigen Spezialfall genau eine "Aktivität" exponentiell verteilter Dauer ihrem Ende zustrebte, z.B. ein "Ankunftsintervall") nun zwei "exponentielle Phasen", unabhängig voneinander und parallel, ihrem Ende zustreben, daß das Ende der einen Phase (Parameter: λ_1) als Ereignis zum Zustand j_1 führt, das der anderen Phase (Parameter λ_2) zum Zustand j_2. Wieder erhalten wir aus

$$\pi_{i,j1}(s) = P[\lambda_1\text{-Phase endet innerhalb s ZE}]$$
$$= \lambda_1 \cdot s + o(s)$$
$$\pi_{i,j1}(s) = q_{i,j1} \cdot s + o(s)$$

das Matrixelement

$$q_{i,j1} = \lambda_1$$

und aus der analogen Überlegung bzgl. der λ_2-Phase

$$q_{i,j2} = \lambda_2$$

Diagonalelemente Q-Matrix

Wir erkennen aus diesen Beispielen: Die Elemente der Q-Matrix lassen sich aus dem (geeigneten!) Problem ablesen; zumindest gilt dies für die Nichtdiagonal-Elemente von Q, die Diagonalelemente haben wir ja bisher vernachlässigt. Daher jetzt zu diesen Diagonalelementen: (4.5.12b) sagte aus, daß die bedingten Übergangswahrscheinlichkeiten sich zu 1 addieren, bzw. leicht umgeschrieben, daß

$$\textbf{(4.5.20)} \qquad 1-\pi_{ii}(s) = \sum_{j\neq i} \pi_{ij}(s) \quad i \in \mathbf{N}_0$$

Vollziehen wir die Grenzbetrachtung (4.5.16b), die zur Definition der Q-Elemente führte, im Rahmen dieser Gleichung, dann erhalten wir

$$\textbf{(4.5.21)} \qquad -q_{ii} = \sum_{j\neq i} q_{ij}$$

Damit: Die Hauptdiagonalelemente nehmen den negativ genommenen Summenwert der Nichtdiagonalelemente der betreffenden Zeile an; sie sind negativ (die Nichtdiagonalelemente sind positiv bzw. verschwinden: Verfolgen Sie die Definition); alle Zeilensummen der Q-Matrix verschwinden. Zur weiteren Interpretation: Analog zu dem Vorgehen, das zu (4.5.19) führte, erhalten wir aus (4.5.20):

$$1-\pi_{ii}(s) = \sum_{j\neq i} \pi_{ij}(s)$$
$$= s \sum_{j\neq i} q_{ij} + o(s)$$

mit (4.5.21): $= s\cdot(-q_{ii}) + o(s)$

$\pi_{ii}(s)$ ist aber die Wahrscheinlicheit, s ZE nach Antreffen von i wieder i vorzufin-

den; entsprechend $1-\pi_{ii}(s)$ die Wahrscheinlichkeit, s ZE nach Antreffen von i ein $j \neq i$ vorzufinden; $-q_{ii}$ damit die Gesamtrate (Intensität), mit der Zustand i seinem Ende zustrebt; dies unabhängig vom Beobachtungszeitpunkt t, so daß (vgl. die Diskussionen im Anschluß an (4.4.14) sowie (4.5.19)) die "Lebenszeit" des Zustands i nicht anders als exponentiell (mit Parameter: $-q_{ii}$) verteilt sein kann.

Testfrage

Testfrage 4.5.22: Gegeben seien 2 unabhängig exponentiell verteilte ZV A und B mit Parametern λ_A und λ_B. Ermitteln Sie die Verteilung der ZV C=min(A,B), die also als Realisierung jeweils den kleineren Wert der Realisierungen von A und B annimmt, etwa in Form der Verteilungsdichte

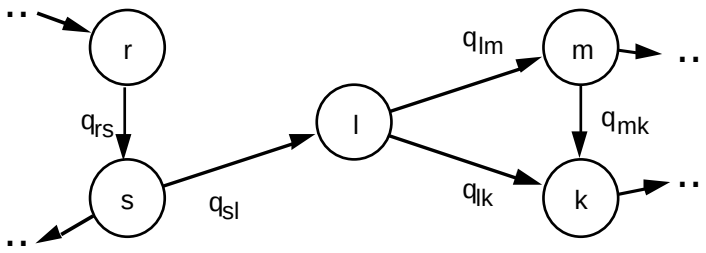
$$fC(c) = P[B>c]\cdot fA(c) + P[A>c]\cdot fB(c)$$

Interpretieren Sie das Ergebnis, zur Stützung Ihrer Vorstellungskraft, im Hinblick auf den Zustand einer Markoff-Kette, in dem zwei exponentielle Phasen laufen, sowie auf die Lebensdauer dieses Zustands im Rahmen der obigen Überlegungen.

Technik: Aufstellung Gleichungssystem für Grenzverteilung

So gestärkt, können wir "den Rahm abschöpfen" und uns eine Technik der Aufstellung des Gleichungssystems (4.5.17) zur Ermittlung der Grenzverteilung einer homogenen, irreduziblen Markoff-Kette zurechtlegen. Wir zeichnen einen sog. "Graphen der Zustandsübergangsraten", in dem als Knoten sämtliche Zustände der M-Kette auftreten sowie als gerichtete Kanten die Möglichkeiten des unmittelbaren Übergangs in Nachbarzustände eingezeichnet sind. Die Kanten sind mit den zugehörigen Übergangsraten (Intensitäten, Q-Matrix-Elementen) bezeichnet. Also etwa so (als Ausschnitt des Gesamtgraphen):

Graph der Zustandsübergangsraten



Eine einzelne Gleichung von (4.5.17) hat die Form (für festes i)

$$0 = \sum_j p_j \; q_{ji} \qquad \text{bzw.} \qquad -p_i \; q_{ii} = \sum_{j \neq i} p_j \; q_{ji}$$

und mit (4.5.21)

$$p_i \sum_{k \neq i} q_{ik} = \sum_{j \neq i} p_j \; q_{ji}$$

Ablesen der Gleichungen

Diese Gleichungen lassen sich aus dem Graphen der (Zustands-)Übergangsraten direkt ablesen: Links steht das Produkt aus der stationären Zustandswahrscheinlichkeit für Zustand i und der Summe der aus i herausführenden Übergangsraten (Pfeile aus i hinaus); rechts steht eine Summe über Produkte, deren jedes als einen Term die stationäre Wahrscheinlichkeit eines Zustands j enthält, von dem aus i direkt erreicht werden kann und als zweiten Term die zugehörige Über-

Wahrscheinlich-

keits-Fluß

gangsrate. Mit der Benennung "Wahrscheinlichkeitsfluß" (W-Fluß) für ein Produkt p.·q., also für Zustandswahrscheinlichkeit mal Austrittsrate aus diesem Zustand, können wir als Merkregel für unser Vorgehen der Aufstellung der Gleichungen von (4.5.17) festhalten:

Regeln

Regel 4.5.23a:
Für jeden Zustand ist der W-Fluß aus diesem Zustand gleich dem W-Fluß in diesen Zustand.

Als Beispiel für Zustand l des obigen Graphen:

$$p_l \cdot (q_{lm} + q_{lk}) = p_s \cdot q_{sl}$$

Regel 4.5.23a ermöglichte eine einfache Aufstellung der (aller!) Gleichungen (4.5.17). Aus (4.5.17) lassen sich weitere Beziehungen zwischen den stationären Zustandswahrscheinlichkeiten ableiten, indem man gewisse dieser Gleichungen addiert. Sei etwa $Y \subset Z$ irgendeine Untermenge der Zustandsmenge, dann ergibt sich

$$\sum_{i \in Y} \{p_i \sum_{k \in Z \setminus Y} q_{ik}\} = \sum_{j \in Z \setminus Y} \{p_j \sum_{i \in Y} q_{ji}\}$$

Auch diese Beziehungen lassen sich in eine Regel fassen, die direkt auf dem Graphen der Übergangsraten operiert:

Regel 4.5.23b:
Für jede Zustands(unter-)menge ist der W-Fluß aus dieser Zustandsmenge gleich dem W-Fluß in diese Zustandsmenge.

Um Mißverständnissen vorzubeugen: Die neue Regel liefert natürlich keine zusätzliche Information; sie führt nur gelegentlich zu einfacheren Gleichungen. Wir werden Gelegenheit haben, die in den Regeln 4.5.23 eingeführte Technik ausführlich zu üben.

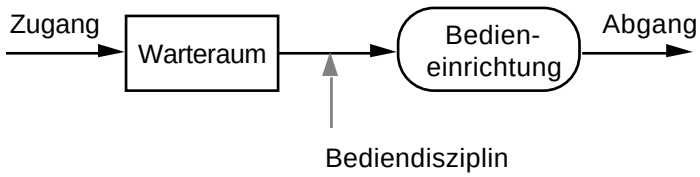
4.6 Stochastische Modelle: Analyse von Einzelstationen

Wir wollen in diesem Abschnitt
- einige konkrete Stationen untersuchen (Stationen "in Isolation", als "Kleinstsysteme"),
- dabei die in Abschnitt 4.5 vorgestellten Analysetechniken einsetzen, um die stationären Grenzverteilungen dieser Stationen zu ermitteln,
- damit die Basis für die Vernetzung von Stationen und deren Analyse legen (Netze von Stationen als "komplexere Systeme") - die wir hier nicht im einzelnen durchführen werden, deren Ergebnisse aber in Abschn. 5.3 vorgestellt sind.

Station

In Anlehnung an vorangegangene Abschnitte kennzeichnen wir eine einzelne Station und ihre Belastung wie folgt:

- Strukturell besteht eine Station aus einer Bedieneinrichtung (die einen oder mehrere gleichartige Bediener enthält), einem vorgelagerten Warteraum (unbeschränkten Fassungsvermögens) und einer Bediendisziplin, welche die Bedienungsreihenfolge regelt (wir beschränken uns auf arbeitserhaltende Disziplinen)

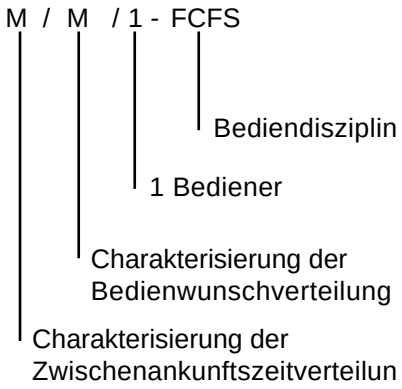


- Die Belastung einer Station ist gekennzeichnet durch einen Strom von (einzeln eintreffenden) "Kunden", deren Ankunftsabstände A unabhängig identisch verteilt sind; ebenfalls unabhängig identisch verteilt sind die Bedienwünsche S der Kunden (wir drücken die Bedienwünsche wieder direkt als Bedienzeiten aus).

Das erste und einfachste System dieser Art, dem wir uns widmen,
- besitzt genau einen Bediener,
- bedient gemäß FCFS-Disziplin,
- weist exponentiell (Parameter: λ) verteilte Ankunftsabstände
- und exponentiell (Parameter: μ) verteilte Bedienwünsche auf.

Als Kurzbeschreibung (sog. "Kendall-Notation") für dieses System dient die Bezeichnung:

M/M/1-FCFS



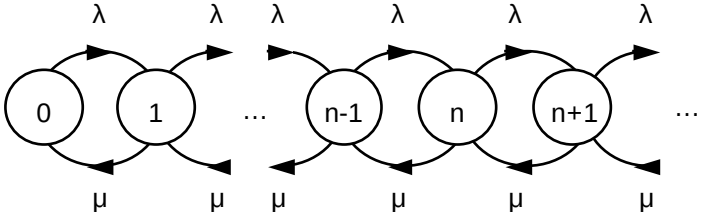
Charakterisierung (N_t;$t \in \mathbf{R}$)

Wir betrachten das System in kontinuierlicher Zeit; als Charakterisierung des Zustands sei die Zahl n insgesamt (also in Warteraum und Bediener zusammen) anwesender Kunden vorgeschlagen. Offensichtliche Frage: Ist $(N_t)_{t \in \mathbf{R}^+}$ eine zeitkontinuierliche, homogene, irreduzible Markoff-Kette, so daß wir unter Anwendung von Satz 4.5.18 nach ihrer stationären Grenzverteilung fahnden können? Nun,

- Zeitkontinuierlichkeit (Betrachtung aller Zeitpunkte $t \in \mathbf{R}^+$) haben wir vorausgesetzt.
- Eine stochastische Kette (diskreter Zustandsraum) liegt vor: der Zustandsraum ist offensichtlich $\mathbf{N_0}$ (n=0, für "leeres System", und n=1,2,... sind mögliche Systemzustände; eine obere Grenze existiert wegen des unbeschränkten Warteraums nicht).
- Die Kette ist irreduzibel, da von beliebigem Zustand n_1 aus sich jeder andere Zustand n_2 durch eine geeignete Anzahl von Ankünften bzw. Abgängen erreichen läßt.
- Die Kette ist Markoff'sch, da bei Kenntnis des Zustands n zu einem Zeitpunkt t die weitere Entwicklung festgelegt ist (und dies ohne weitere Kenntnis der Vorgeschichte zu Zeitpunkten <t, insbesondere unabhängig davon, wie lange der Zustand n vor dem Zeitpunkt t schon andauerte). Zur Erläuterung: Einmal "läuft" zum Zeitpunkt t (unabhängig von n) ein Ankunftsintervall, das wegen der Annahmen über seine Dauer (exponentiell verteilt, Parameter λ) noch eine exponentiell mit Parameter λ verteilte Dauer vor sich hat - s.(4.4.5). Es endet in der Zeit (t,t+s) mit der Wahrscheinlichkeit $\lambda \cdot s + o(s)$ und führt vom Zustand n in den unmittelbar möglichen Folgezustand n+1. Die Übergangsrate von einem Zustand n zum Zustand n+1 (vgl. (4.5.19) und darauf folgende Ausführungen) beträgt damit $q_{n,n+1}=\lambda$ für alle n=0,1,2,... Zum anderen "läuft" in allen Zuständen n≥1 eine Bedienung (der arbeitserhaltende Bediener ist ja mit der Bedienung des Erstankömmlings unter den n Anwesenden beschäftigt). Diese Bedienung hat wegen der Annahmen über ihre Dauer (exponentiell verteilt, Parameter μ) noch eine exponentiell mit Parameter μ verteilte Restdauer. Sie endet in der Zeit (t,t+s) mit der Wahrscheinlichkeit $\mu \cdot s + o(s)$ und führt zum Zustand n-1. Die Übergangsrate von einem Zustand n≥1 zum Zustand n beträgt demnach $q_{n,n-1}=\mu$ für alle n=1,2,...

Übergangsraten

Wir haben unsere Frage damit zufriedenstellend beantwortet und sogar die Elemente der Q-Matrix (bzw. die Kenngrößen des Übergangsratendiagramms) bereits ermittelt. Das Diagramm hat konkret die Form:



Zur Erinnerung: Es enthält als Knoten alle Zustände; als Kanten sind alle unmittelbar möglichen Zustandsübergänge eingezeichnet; sie sind mit den zugehörigen Übergangsraten bezeichnet.

Lösung

Das Gleichungssystem (4.5.17), das ja die (potentiell existierende) stationäre Verteilung der Kette bestimmt, läßt sich mittels Regel 4.5.23a direkt aus dem Diagramm ablesen. Wir erhalten

für Zustand 0: $p_0 \cdot \lambda = p_1 \cdot \mu$

für Zustand 1: $p_1 \cdot (\mu+\lambda) = p_0 \cdot \lambda + p_2 \cdot \mu$

…

für Zustand n: $p_n \cdot (\mu+\lambda) = p_{n-1} \cdot \lambda + p_{n+1} \cdot \mu$

Insgesamt also:

(4.6.1a) $p_0 \cdot \lambda = p_1 \cdot \mu$
$p_n \cdot (\mu+\lambda) = p_{n-1} \cdot \lambda + p_{n+1} \cdot \mu \quad n=1,2,...$

dazu natürlich (4.5.17 b):

(4.6.1b) $\sum_{n \varepsilon \mathbf{N}_0} p_n = 1$

Die explizite Lösung dieses Gleichungssystems ist leicht möglich: (4.6.1a), leicht umgeschrieben, lautet

$p_1 \cdot \mu - p_0 \cdot \lambda = 0$
$p_n \cdot \mu - p_{n-1} \cdot \lambda = p_{n+1} \cdot \mu - p_n \cdot \lambda \quad n=1,2,...$

und führt durch iteratives Einsetzen zunächst zu

$p_n \cdot \mu - p_{n-1} \cdot \lambda = 0 \qquad n=1,2,...$

woraus wir, wieder durch iteratives Einsetzen

$p_n = (\lambda / \mu)^n \cdot p_0$

erhalten, bzw. mit der abkürzenden Bezeichnung

(4.6.2a) $\rho = \lambda/\mu$

Stationäre Verteilung M/M/1

auch: $p_n = \rho^n \cdot p_0 \quad n=0,1,2,...$

Für die verbleibende Unbekannte p_0 ergibt sich mit (4.6.1b)

$p_0 \sum_{n=0}^{\infty} \rho^n = 1$

und unter der Voraussetzung (Konvergenz der Summe!)

(4.6.2b) $\rho<1$

wird $p_0 / (1- \rho) = 1$

(4.6.2c) $p_0 = 1-\rho$

Insgesamt lautet unsere Lösung also

(4.6.2d) $p_n = \rho^n \cdot (1-\rho) \qquad n=0,1,2,...$

Die M / M / 1-FCFS-Station besitzt also eine geometrische Verteilung - vgl. (4.4.6) - für die Zustandsvariable N, jene ZV, welche die Anzahl (zu zufälligem Zeitpunkt der stationären Phase) anwesender Kunden beschreibt. Der Erwartungswert von N (die "mittlere Kundenzahl") ergibt sich zu

$$E[N] = \sum_{n=0}^{\infty} n\, p_n = (1-\rho) \sum_{n=0}^{\infty} n\, \rho^n$$

Explizit erhält man mit einem kleinen "Kunstgriff":

$$\begin{aligned} E[N] &= (1-\rho) \sum_{n=1}^{\infty} \rho\, n\, \rho^{n-1} \\ &= (1-\rho)\,\rho \sum_{n=1}^{\infty} d\rho^n / d\rho \\ &= (1-\rho)\,\rho \frac{d}{d\rho}\left[\sum_{n=1}^{\infty} \rho^n \right] \\ &= (1-\rho)\,\rho \frac{d}{d\rho}\left[\rho/(1-\rho)\right] \\ &= (1-\rho)\,\rho \frac{(1-\rho)-(-\rho)}{(1-\rho)^2} \end{aligned}$$

(4.6.2e) $E[N] = \rho / (1-\rho)$

Interpretation von ρ

Die Größe ρ war in (4.6.2a) ohne weitere Begründung als abkürzende Bezeichnung eingeführt worden. Sie besitzt aber eine Reihe interessanter Interpretationen, die sie zur zentralen Kenngröße des untersuchten Systems machen. Zunächst ist ja die Ankunftsrate

(4.6.3a) $1 / E[A] = \lambda$

auch die mittlere Anzahl der je Zeiteinheit an der Station eintreffenden Kunden. Jeder von diesen bringt (unabhängig vom Ankunftsprozeß selbst) im Mittel

(4.6.3b) $E[S] = 1 / \mu$

an zu leistender Arbeit (Bedienwunsch) in die Station ein, so daß als erste Interpretation

(4.6.4a) $\rho = \lambda / \mu$ ist die im Mittel pro Zeiteinheit in die Station eingebrachte, **zu leistende Arbeit**

anfällt. Eine arbeitserhaltende Station leistet aber genau die in Auftrag gegebene Arbeit, so daß als weitere Interpretation folgt

(4.6.4b) ρ ist die im Mittel pro Zeiteinheit von der Station **geleistete Arbeit**

Wieder aufgrund der Tatsache, daß wir Bedienwünsche direkt als Bedienzeiten angegeben hatten, resultiert nun auch

(4.6.4c) ρ ist die mittlere Tätigkeitszeit des Bedieners je Zeiteinheit, d.h. die

"**Auslastung**" der Station

als Interpretation. Die Auslastung, in dieser Interpretation, kann aber maximal den Wert 1 annehmen (länger als 1 Zeiteinheit je Zeiteinheit kann der Bediener nicht tätig sein). Eine solche Voraussetzung, nämlich $\rho<1$, hatten wir in (4.6.2b) auch getroffen; für Werte $\rho\geq 1$ hatten wir nämlich keine stationäre Grenzverteilung angeben können (keine Konvergenz der Summe über ρ^n). Die physikalische Interpretation wird uns hier geliefert: Wird pro Zeiteinheit mehr Arbeit ins System gebracht, s. (4.6.4b), als das System leisten kann, s. (4.6.4c), dann kann die Warteschlange vor dem Bediener immer nur wachsen (wenigstens im Mittel) - eine Hoffnung auf Annäherung an eine eingeschwungene, stationäre Phase kann sich nicht erfüllen. Noch weiter: Mit (4.6.4c) folgt auch

(4.6.4d) $1-\rho$ ist die mittlere **Leerlaufzeit** des Bedieners pro Zeiteinheit,

also die Zeit, in welcher die Station untätig ist - was diese nur ist, wenn kein Kunde anwesend ist. Die mittlere Zeit pro Zeiteinheit, in der genau i Kunden an der Station weilen, ist aber die statistische Interpretation für die Wahrscheinlichkeit, zu beliebigem Zeitpunkt genau i Kunden vorzufinden. Im Mittel sollten also pro Zeiteinheit für die Zeitdauer p_0 keine Kunden anwesend sein, d.h. die Station nicht arbeiten. $p_0=1-\rho$ folgt aus (4.6.4d), ist andererseits in (4.6.2c) bereits analytisch abgeleitet worden.

Erweiterung der Interpretation von λ

Der vorstehende Abschnitt ist zwar als Interpretation der für das konkret betrachtete M / M / 1-FCFS-System erhaltenen Resultate (4.6.2) formuliert, besitzt aber weit allgemeinere Gültigkeit: Wird eine arbeitserhaltende Station mit einem Kundenstrom gespeist, der durch u.i.v. Ankunftsabstände A und u.i.v. Bedienzeiten S charakterisiert ist, und definieren wir Ankunftsrate λ und mittlere Bedienzeit $1/\mu$ entsprechend (4.6.3), dann gelten die Aussagen (4.6.4). Es ist demnach

$\rho = \lambda / \mu$ die im Mittel pro Zeiteinheit in die Station getragene Arbeit

und unter der (für die Existenz einer stationären Phase notwendigen) Voraussetzung $\rho<1$ ist auch

ρ die im Mittel pro Zeiteinheit durch die Station geleistete Arbeit, die **Auslastung** der Station (mittlere Tätigkeitszeit je Zeiteinheit),

$1-\rho$ die mittlere Leerlaufzeit der Station pro Zeiteinheit, d.h. $\rho+p_0=1$

Bitte machen Sie sich die Mühe, die offensichtlichen Analogien zur "arbeitserhaltenden Station konstanter Bedienkapazität 1" des Abschnitts 4.2 (Def. 4.2.8) zu ziehen. Und, wenn wir schon beim Zurückblättern sind, natürlich gilt auch weiterhin das allgemeine Resultat von Little (4.2.22), das sich (Voraussetzung: Existenz stationärer Phase) mit den hier verwendeten Notationen als

Little-Resultat

(4.6.5)
$$\begin{aligned} E[N] &= E[V] / E[A] \\ &= E[V]\cdot\lambda \end{aligned}$$

darstellt (wo die ZV V die Verweilzeit eines zufällig ausgewählten Kunden, im stationären Betrieb der Station, bezeichnet). Es gibt eine Reihe von Beweisen für das Resultat von Little. Einer davon, auf Eilon zurückgehend (vgl. z.B. Klei75), verfolgt exakt die Vorgehensweise, die uns zu (4.2.22) führte und vollzieht dann

den Grenzübergang für $T \to \infty$, was zu (4.6.5) führt.

Mit (4.6.2e,4.6.5) erhalten wir konkret für die M / M / 1-FCFS Station die mittlere Kundenverweilzeit

Verweilzeit

(4.6.6) $E[V] = 1/[\mu\cdot(1-\rho)] = 1/(\mu-\lambda)$

Ähnlich einfach wie die vorangegangene Analyse der M / M / 1-FCFS-Station verläuft auch die von M / M / 1-Stationen unter einer Reihe anderer Bediendisziplinen:

M/M/1-Resultate

Satz 4.6.7: Die stationäre Verteilung der Kundenpopulation N der M / M / 1-Station ist unter den Bediendisziplinen

(a) RANDOM: die Auswahl aus der Warteschlange erfolgt zufällig in dem Sinne, daß aus den Wartenden jeder (mit identischer Wahrscheinlichkeit) ausgewählt werden kann;

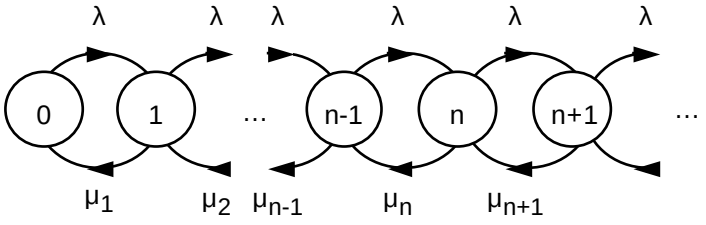
(b) LCFS: Last Come First Served: der jeweils zuletzt Angekommene wird zuerst bedient;

(c) LCFS-PR: LCFS-Preemptive Resume:
die Bedienung eines Neuankömmlings beginnt unmittelbar, der bei Ankunft Bediente wird unterbrochen, "preempted"; kann ein Kunde seine Bedienung beenden, wird jener Kunde der Warteschlange ausgewählt, der (zeitlich) als Letzter dort eingereiht wurde; er setzt seine Tätigkeit an der evtl. Unterbrechungsstelle fort, "resume"; im Unterschied zu LCFS-PR sind FCFS, RANDOM, LCFS nicht-unterbrechend, "non-preemptive".

identisch der Populationsverteilung unter der FCFS-Disziplin, also durch (4.6.2) beschrieben. Ebenfalls identisch ist bei all diesen Disziplinen die mittlere Verweilzeit gemäß (4.6.6). (Zur Warnung: Absolut **nicht** identisch sind aber die Verteilungen der Verweilzeiten V.)

Testfrage 4.6.8: Bestätigen Sie Satz 4.6.7 durch Aufstellung und Interpretation des Diagramms der Übergangsraten sowie durch die Anwendung von Little's Resultat.

Um uns eine weitere interessante Klasse von Stationen zu erschließen, betrachten wir folgendes Übergangsratendiagramm einer zeitkontinuierlichen Markoff-Kette mit Zustandsraum $\mathbf{N_0}$:



In Unterschied zum zuvor benutzten M / M / 1-Diagramm sind also hier die Rückwärtsraten nicht identisch gleich μ, sondern dürfen für jeden Zustand unterschiedlich sein. (Eine weitere Verallgemeinerung ließe auch unterschiedliche, zustandsabhängige Vorwärtsraten λ_i zu; diese weitere Verallgemeinerung trägt den Namen "allgemeiner Markoff'scher Geburts- / Todesprozeß"; wir benötigen diese allgemeinere Form aber nicht.)

Geburts-/ Todesprozeß

Aus diesem Diagramm erhalten wir (Vorgehen analog (4.6.1, 4.6.2)) das System der Gleichgewichtsgleichungen

$$p_0 \cdot \lambda = p_1 \cdot \mu_1$$
$$p_n \cdot (\lambda + \mu_n) = p_{n-1} \cdot \lambda + p_{n+1} \cdot \mu_{n+1} \qquad n=1,2,...$$

dessen vereinfachte Form

$$p_n \cdot \mu_n - p_{n-1} \cdot \lambda = 0 \qquad n=1,2,...$$

und seine Lösung

(4.6.9a) $$p_n = p_0 \ \lambda^n / \prod_{i=1}^{n} \mu_i$$ *Stationäre Verteilung*

woraus mit der normierenden Bedingung

$$\sum_{i=0}^{\infty} p_i = 1$$

sich weiterhin

(4.6.9b) $$p_0 = \left\{ 1 + \sum_{n=1}^{\infty} \left[\lambda^n / \left(\prod_{i=1}^{n} \mu_i \right) \right] \right\}^{-1}$$

ergibt (Bedingung für Existenz stationärer Lösung: Summe konvergiert zu endlichem Wert). Offensichtlich ist (4.6.2) als Sonderfall in (4.6.9) enthalten (nachprüfen!).

Wir haben ein Ergebnis, aber welche Interpretation können wir ihm unterlegen, welche praktischen Probleme sind durch dies Ergebnis erfaßt?

Interpretation

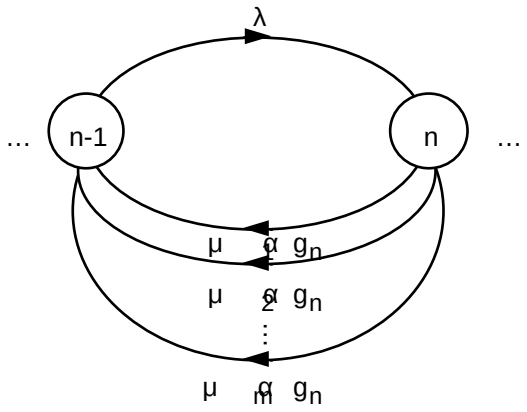
Nun, eine der Interpretationen (von vielen möglichen) ist die folgende: Nehmen wir an, wir hätten ein M/M/. System, d.h. einen Poisson-Ankunftsstrom von Kunden (Parameter λ) mit exponentiellen Bedienwünschen (Parameter μ) - wie gehabt. Bedieneinrichtung und Bediendisziplin der Station aber seien etwas abstrakter wie folgt beschrieben: Die Bedienung erfolgt mit einer zustandsabhängigen Arbeitskapazität, mit einer Arbeitskapazität also, die von der Anzahl anwesender Kunden abhängt. Will heißen: Die Station erledigt nicht (wie bisher in diesem Kapitel angenommen) immer exakt 1 Arbeitseinheit je Zeiteinheit, sondern besitzt zustandsabhängige Bediengeschwindigkeiten g_i, i=1,2,..., so daß bei Anwesenheit von i Kunden g_i Arbeitseinheiten je Zeiteinheit abgearbeitet werden. Man beachte, daß wir uns noch nicht festgelegt haben, welchem bzw. sogar welchen der Anwesenden diese Arbeit zugute kommt. Dies ist auch nicht nötig! Nehmen wir einen allgemeinen Fall an: Seien n Kunden anwesend; die Bedieneinrichtung bediene in diesem Zustand m≤n Kunden (die anderen n-m "warten"); die im Zustand n aufgebrachte Arbeitskapazität g_n komme den m bearbeiteten, beliebig numerierten Kunden in jeweiligen Bruchteilen $\alpha_1, \alpha_2, ...\alpha_m$ zugute; auf den i-ten Kunden entfällt also eine Arbeit von $\alpha_i \cdot g_n$ AE/ZE. Der Vollständigkeit halber sei angemerkt:

zustandsabhängige Bedienung

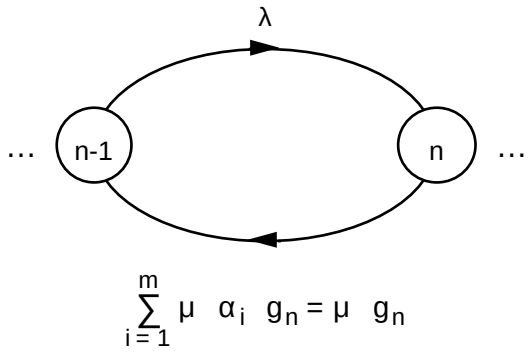
$$\sum_{i=1}^{m} \alpha_i = 1$$

$$\sum_{i=1}^{m} \alpha_i \; g_n = g_n$$

Der bzgl. einer Bediengeschwindigkeit von 1 AE/ZE ausgedrückte Bedienwunsch (exponentiell mit Parameter μ) resultiert in einer demgegenüber linear gedehnten/gestauchten Bedienzeit (exponentiell, aber mit Parameter $\mu\cdot\alpha_i\cdot g_n$). Das Ende der Bedienung des so Bedienten geschieht also innerhalb der nächsten s ZE mit Wahrscheinlichkeit $(\mu\cdot\alpha_i\cdot g_n)\cdot s+o(s)$. Zeichnen wir die so gewonnenen Übergangsraten in unser Diagramm ein, erhalten wir folgendes Bild (Ausschnitt)



welches aber zu identischen Gleichungen führt (nachvollziehen!) wie das Bild



Unser Ergebnis (4.6.9) läßt sich also auf alle M/M/. Systeme mit zustandsabhängigen (Gesamt-)Arbeitskapazitäten g_n anwenden, wenn wir nur $\mu_n = g_n\cdot\mu$ (n=1,2,...) setzen.

Ist diese Interpretation praktisch relevant? Und ob!:

M/M/m-FCFS

(4.6.10a) Die M/M/m-FCFS Station ist definiert durch eine M/M-Belastung und eine Bedieneinrichtung aus m identischen Bedienern (deren jeder eine konstante Arbeitskapazität von 1 AE/ZE besitzt). Die Bedienung erfolgt derart, daß bei n≤m anwesenden Kunden jeder genau einem Bediener zugeordnet ist, bei n>m Anwesenden die m "früher Angekommenen" je einen Bediener belegen (während die n-m "später Angekommenen" warten), bei Freiwerden eines Bedieners

(wegen Bedienungsende) aus den evtl. Wartenden der Erstankömmling diesen Bediener erhält. Die Gesamtarbeitskapazitäten dieser Station betragen also

$$g_n = n \qquad n=1,2,...,m$$
$$g_n = m \qquad n \geq m$$

Die stationäre Verteilung der Station ist durch (4.6.9), nach Substitution $\mu_n = g_n \cdot \mu$, gegeben.

(4.6.10b) Die Resultate (4.6.10a) gelten auch für M / M / m-RANDOM, M / M / m-LCFS, M / M / m-LCFS-PR-Stationen. *M/M/m-.*

Testfrage 4.6.10c: Wie lautet die explizite Form der stationären Zustandsverteilung der M / M / m-Systeme? *Testfrage*

(4.6.11) Die M / M / ∞-Station ist definiert durch eine M / M-Belastung und eine Bedieneinrichtung aus "beliebig vielen" identischen Bedienern (je 1 AE / ZE). Jeder Ankömmling bekommt unmittelbar einen eigenen Bediener zugeordnet (vgl. Abb. 2.2.10 samt Kommentar). Die Gesamtarbeitskapazitäten der Station betragen *M/M/∞*

$$g_n = n \qquad n=1,2,...$$

Die stationäre Verteilung erhalten wir aus (4.6.9) zu

$$p_n = p_0 \cdot (\lambda / \mu)^n / n!$$

mit

$$p_0 = (1 + \sum_{n=1}^{\infty} [(\lambda / \mu)^n / n!])^{-1}$$
$$= \exp(-\lambda / \mu)$$

Diese Station trägt auch die Bezeichnungen "Infinite Server" (kurz: "IS") oder "pure delay" (vgl. nochmals Abb. 4.2.10). *IS*

(4.6.12) Die M / M / 1-PS-Station ist definiert durch eine M / M-Belastung, einen einzelnen Bediener (1 AE / ZE) und eine Bediendisziplin, die sich als Grenzfall "fairer" Zeitscheibendisziplinen (z.B. von RR, vgl. Testfrage 4.2.9) ergibt: Lassen wir nämlich die Zeitscheiben, die ein einzelner Kunde an Bedienung erhält, sehr klein werden, dann teilt sich die Arbeitskapazität des Bedieners gleichmäßig auf alle Kunden auf, bei n Anwesenden wird also der einzelne mit $\alpha_n = 1 / n$ der Gesamtkapazität bedient. Die Gesamtarbeitskapazität der Station beträgt aber konstant 1, so daß wir aus (4.6.9) dieselben Ergebnisse erhalten, wie wir sie für FCFS-Bedienung erhielten, nämlich (4.6.2): *M/M/1-PS*

$$p_n = \rho^n \cdot p_0$$
$$p_0 = 1 - \rho$$

Die Disziplin trägt den Namen "Processor Sharing" (kurz: "PS)" aufgrund der geschilderten (idealerweise) gleichmäßigen Aufteilung ihrer Kapazität auf alle Anwesenden. *PS*

Fassen wir die gewonnenen Ergebnisse kurz zusammen.

Zusammenfassung Ergebnisse

Zusammenfassung 4.6.13: Sei die Belastung einer Station gemäß M / M / . festgelegt, d.h. Kunden treffen gemäß eines Poissonstroms mit Parameter λ ein und haben u.i. exponentiell verteilte Bedienwünsche des Parameters μ. Sei die Station ausgezeichnet durch zustandsabhängige Bediengeschwindigkeiten g_i, i=1,2,... . Bediene die Station gemäß Disziplin FCFS oder LCFS oder LCFS-PR oder RANDOM. In all diesen Fällen ist die stationäre Zustandsverteilung gegeben durch (4.6.9), also durch

$$p_n = p_0 \left(\frac{\lambda}{\mu}\right)^n / \prod_{i=1}^{n} g_i \quad n=1,2,...$$
$$\Sigma\, p_n = 1$$

Durch geeignete Setzung der g_i sind hierin insbesondere eingeschlossen die diversen M / M / m-Fälle sowie die Strategien IS und PS.

Literatur zu Kap. 4 und allgemein zu "operational analysis"

Bau74 Bauer,H.; Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie; de Gruyter, 1974

BrDe82 Brumfield,J.A. / Denning,P.J.; Error analysis of homogeneous mean queue and response time estimators; Performance Evaluation Review vol.11(1982 / 83) nr.4

Brya80 Bryant,R.M.; On homogeneity in M / G / 1 queueing systems; in: Proc. PERFORMANCE ´ 80: Performance Evaluation Rev. vol.9(1980) nr.2 pp.199-208

Buze71 Buzen,J.P.; Analysis of system bottlenecks using a queueing network model; in: Proc. ACM / SIGOPS workshop on system performance evaluation; ACM 1971

Buze76 Buzen,J.P.; Fundamental operational laws of computer system performance; Acta Informatica vol.7(1976) nr.2

Buze78 Buzen,J.P.; Operational analysis: An alternative to stochastic modelling; in: Ferrari,D.; Performance of Computer Installations; North Holland 1978

ChLa74 Chang,A. / Lavenberg,S.S.; Work rates in closed queueing networks with general independent servers; Operations Research vol.22(1974) nr.4

Cinl75 Cinlar,E.; Introduction to Stochastic Processes; Prentice-Hall, 1975

Cox55 Cox,D.R.; A Use of Complex Probabilities in the Theory of Stochastic Processes; Proc. Camb. Philos. Soc., vol. 51 (1955)

DeBu77 Denning,P.J. / Buzen,J.P.; Operational analysis of queueing networks; in: Proc. PERFORMANCE ´77: Beilner / Gelenbe (eds); Measuring, modelling and evaluating computer systems pp.151-172; North Holland 1977

DeBu78 Denning,P.J. / Buzen,J.P.; The operational analysis of queueing network models; Computing Surveys vol.10(1978) nr.3 pp.225-261

DeKo82 Denning,P.J. / Kowalk,W.; Error analysis of the mean busy period of a queue; in: Proc. 10th IMACS World Congress on System Simulation and Scientific Computation, 1980

Denn81 Denning,P.J. ; Performance analysis: Experimental computer science at its best; CACM vol.24(1981) nr.11 pp.725-727

Dörf78 Dörfler, W.; Mathematik für Informatiker, Band 2: Methoden aus der Analysis; C. Hanser, 1978

Gele82 Gelenbe,E.; Stationary deterministic flows in discrete systems I; Performance Evaluation Review vol.11(1982) nr.4 p.89ff

Klei75 Kleinrock,L.; Queueing Systems; Wiley, 1975 / 1976

Kowa82 Kowalk,W.; Verkehrsanalyse in endlichen Zeiträumen; Springer, IFB vol.55(1982)

Kowa83 Kowalk,W.; Erweiterte Methoden der operationalen Analyse; Angewandte Informatik vol.25(1983) nr.1

Kowa89 Kowalk W.; Operationale Wartetheorie; Springer IFB vol.196 1989

LZGS84 Lazowska,E.D. / Zahorjan,J. / Graham,G.S. / Sevcik,K.C.;
Quantitative system performance - Computer system analysis using queueing network models; Prentice Hall 1984

Pflu86 Pflug G.; Stochastische Modelle in der Informatik; Teubner 1986

Rood79 Roode,J.D.; Multiclass operational analysis of queueing networks;
in: Proc. PERFORMANCE ´79: Arato et al (eds); Performance of computer systems; North Holland 1979

ThBA85 Thareja,A.K. / Buzen,J.P. / Agrawal,S.C.; BEST / 1-SNA: A software tool for modelling and analysis of IBM SNA networks; in Potier,D. (ed): Modelling techniques and tools for performance analysis, North Holland 1985 pp.81-98

Wu82 Wu,L.T.; Operational models for the evaluation of degradable computing systems; Performance Evaluation Review vol.11(1982) nr.4 pp.179ff

Lösungshinweise zu den Testfragen

4.1.2: *4.1.2*

Temporäre Prozesse modellieren das Verhalten von Aufgaben, die von der Umgebung an die Maschine gestellt werden. Im Normalfalle werden dabei, mit gewissen zeitlichen Abständen, immer wieder Prozesse eines Prozeßmusters generiert (z.B. die Aufgaben der Stapelverarbeitung repräsentierend); solche Aufgaben sind aber irgendwann "abgearbeitet". Wären sie dies nicht, dann würde sich das System mit in Bearbeitung befindlichen Aufgaben immer weiter "anfüllen". Für den Fall einiger weniger tatsächlich "immer" vorhandener Prozesse (z.B. Verwaltungsprozesse) setzt man die permanenten Prozesse ein - und unterschlägt ggf. die explizite Generierung (die ja nur den "Systemstart" nachbilden würde).

4.2.4: *4.2.4*

Die Feststellung "Ereignis x tritt mit Rate y auf" impliziert die Vorstellung von y Ereignissen x je Zeiteinheit. Selbst wenn Sie klarstellen, daß aufgrund der Berechnungsvorschriften nur "im Mittel" y Ereignisse x pro Zeiteinheit generiert sein können, führt

- ein zu kleines T
  (Beispiel: In 1 Stunde geschehen insgesamt 10 Ereignisse, T hat die Länge 1 sec, in diese Sekunde fällt ein / kein Ereignis)

- starke statistische Unregelmäßigkeit
  (Beispiel: In 1 Stunde geschehen insgesamt 10 Ereignisse, T hat die Länge 1 h, alle 10 Ereignisse fallen in die ersten 10 sec von T)

zu (intuitiv) falschen Vorstellungen.

4.2.9: *4.2.9*

Sollten Ihnen die Bediendisziplinen nicht vertraut sein (in welchem Fall Sie unbedingt Ihre Betriebssystem-Kenntnisse auffrischen sollten!), hier eine kurze Charakterisierung:

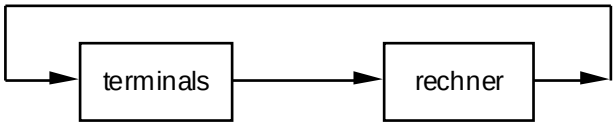
- FCFS: Kunden werden in der Reihenfolge ihrer Ankunft, jeweils vollständig und ohne Unterbrechung, bedient;
- HOL: Jeder Kunde gehört einer aus einer festen Anzahl von (Prioritäts-) Klassen an; der Warteraum der Station enthält für jede Klasse eine gesonderte Warteschlange; die Bedienung jedes Kunden geschieht vollständig und ohne Unterbrechung; höhere Prioritätsklassen werden vor niedrigeren bedient, innerhalb jeder Prioritätsklasse gilt FCFS;
- RR: Der Warteraum besteht aus einer Warteschlange; Neuankömmlinge stellen sich am Ende der Schlange an; die Bedienung erfolgt entsprechend der Ordnung der Warteschlange, ist allerdings unterbrechend in dem Sinne, daß der Bediener einem Kunden maximal für die Dauer einer "Zeitscheibe" fester Länge verfügbar ist; ist die Bedienanforderung innerhalb dieser Zeitscheibe nicht zur Gänze abgewickelt, stellt sich der Kunde (mit seiner "Rest"-Bedienanforderung) wieder am Ende der Warteschlange an und wartet auf Fortsetzung der Bedienung u.s.f.

Zur Aufgabe: Machen Sie sich (durch Zeichnungen, Rechnungen o.ä.) klar, daß

weder eine (beliebige!) Umordnung im Warteraum noch eine (beliebige!) Stückelung der Bedienung auf Gleichungen (4.2.7) einen Einfluß hat, solange nur "arbeitserhaltend" und "konstante Bedienkapazität" im Sinne von Def. 4.2.8 zutrifft, was hier insbesondere heißt, daß durch die Organisation der Disziplin keine Bedienkapazität verlorengeht: Also kein Aufwand für Sortieren, Unterbrechen, Wiederaufsetzen; kein Verlust bereits geleisteter Arbeit; keine "idle"-Zeiten trotz nichtleeren Warteraums.

4.2.11 **4.2.11:**

Alle Endgeräte werden zusammen durch eine Verzögerungsstation terminals repräsentiert, so daß sich folgende Maschine ergibt:



Wir benötigen nur noch ein Prozeßmuster:

```
LOOP
   terminals.denke;
   rechner.bearbeite
ENDLOOP
```

zu dem es n individuelle Prozesse gibt, jeder einem Benutzer (samt seinem nur noch gedachten individuellen Endgerät) zugeordnet. Die Umsetzung von "denke" in das physikalisch notierte "widme mir (z.B.) 20 sec" - die Denkzeit des Benutzers repräsentierend - bleibt unverändert (die u.U. zeitlich überlappend ablaufenden Denkvorgänge beeinflussen einander aufgrund der Verzögerungsstation-Fiktion in keiner Weise), eine Umsetzung von "bearbeite" ist, wie gehabt, nur bei weiterer Auflösung von rechner sinnvoll möglich.

4.2.28 **4.2.28:**

Die Aufgabe ist in ihrer Einfachheit fast trivial und auch ohne viel Formalismus mit gesundem Menschenverstand lösbar: Von den 31 cpu-Besuchen je Stapelauftrag endet der letzte mit einem Verlassen des Systems (nur von cpu aus ist Verlassen möglich), an die ersten 30 schließt sich je eine Anforderung an eine der E / A-Einheiten an; davon gelten 20 der disk_1, 10 der disk_2. Numerieren wir der Einfachheit halber die Stationen: disk_1 entspricht "1", disk_2 entspricht "2", cpu entspricht "3", dann erhalten wir bei einer Ankunftsrate der Stapelaufträge von c_0 (Systemankünfte / Zeiteinheit) die Durchsätze $c_3=31 \cdot c_0$, $c_2=10 \cdot c_0$, $c_1=20 \cdot c_0$.

Bei einer komplexeren Aufgabe dieser Art ist es dagegen ratsam, formaler vorzugehen wie folgt:

$h_{03} = 1$	Jede Systemankunft geht nach cpu
$h_{13} = h_{23} = 1$	Jeder E / A-Abgang geht nach cpu
$h_{31} = 20 / 31$	
$h_{32} = 10 / 31$	Entsprechend obiger Überlegungen
$h_{30} = 1 / 31$	
$h_{ij} = 0$	für alle anderen (i,j)

Gleichungssystem (4.2.27b) erhält daher das Aussehen

$$c_0 = h_{30}c_3 + 0$$
$$c_1 = h_{31}c_3 + 0$$
$$c_2 = h_{32}c_3 + 0$$
$$c_3 = h_{03}c_0 + h_{13}c_1 + h_{23}c_2 + 0$$

Wegen der Homogenität des Gleichungssystems entfernen wir eine Gleichung (z.B. die letzte) und benutzen c_0 als Parameter

$$h_{30}c_3 = c_0$$
$$c_1 - h_{31}c_3 = 0$$
$$c_2 - h_{32}c_3 = 0$$

woraus sich die bereits bekannte Lösung

$$c_3 = \frac{1}{h_{30}} c_0 = 31\, c_0$$
$$c_1 = \frac{h_{31}}{h_{30}} c_0 = 20\, c_0$$
$$c_2 = \frac{h_{32}}{h_{30}} c_0 = 10\, c_0$$

ergibt.

4.2.32: ***4.2.32***

Gegenüber 4.2.28 ändert sich nur folgendes: Eine Umgebung (also "Station 0") gibt es nicht mehr, h_{03} und h_{30} fallen also weg. Dagegen gibt es einen neuen Weg cpu-cpu und die zugehörige Übergangshäufigkeit $h_{33}=1/31$ - dem Werte nach natürlich gleich dem alten h_{30}, da ja mit diesem Weg (zuvor und jetzt) die Beendigung eines Stapelauftrags angezeigt ist. Ziehen wir die Aufgabe formal durch (sie wäre nach wie vor einfach genug, um auch ohne Formalismus durchzukommen), dann haben wir es mit dem Gleichungssystem (4.2.31) zu tun:

$$h_{31}X_3 - X_1 = 0$$
$$h_{32}X_3 - X_2 = 0$$
$$h_{13}X_1 + h_{23}X_2 + (h_{33}-1)X_3 = 0$$

mit der zusätzlichen Gleichung (j*=1 wegen Messung von c_1 an disk_1)

$$X_1 = 1$$

Lassen wir eine der oberen Gleichungen weg, z.B. die dritte, dann ergeben sich

$$X_3 = \frac{1}{h_{31}} X_1 = \frac{1}{h_{31}} = \frac{31}{20}$$
$$X_2 = h_{32} X_3 = \frac{h_{32}}{h_{31}} = \frac{1}{2}$$

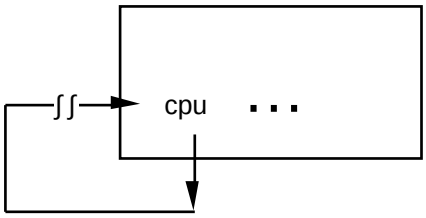
sowie nach (4.2.30) die Durchsätze (in Kunden / ZE)

$$c_2 = X_2 c_1 = \frac{40}{2} = 20$$
$$c_3 = X_3 c_1 = \frac{31\,40}{20} = 62$$

Die Abfertigungsrate für Stapelaufträge entspricht dem Verkehrsfluß cpu-cpu, hat also den Wert

$c_3 \cdot h_{33} = 62 / 31 = 2$

Stellen wir uns nun "in" den Rückkopplungsweg cpu-cpu und fassen das gesamte System als "Station" auf:



Wir kennen den Durchsatz (die Eingangsrate): $c_3 h_{33} = 2$ Kunden / ZE; die Zahl der Kunden im System: konstant gleich MP=5, d.h. auch im Mittel 5. Little's Gesetz (4.2.22b) ist anwendbar und ergibt als mittlere Bearbeitungszeit eines Stapelauftrags

$$\overline{V} = \frac{n}{a} = \frac{MP}{c_3 h_{33}} = \frac{5}{2} = 2.5 \text{ ZE}$$

4.2.34 **4.2.34:**

Der maximale Wert für c_0, nennen wir ihn $c_{0,max}$, ergab sich aus der Tatsache, daß für diesen Wert eine der Stationen, sagen wir Station k, einen Durchsatz $c_{k,max}$ erreichte derart daß, s. (4.2.30,4.2.33)

$$1 = b_{k,max} = \overline{B}_k \cdot c_{k,max}$$

diese Station also dauernd beschäftigt war. Würden wir nun c_0 über $c_{0,max}$, und damit c_k über $c_{k,max}$, hinaus steigern, dann könnte Station k die daraus resultierende Arbeit nicht mehr bewältigen: Der Warteraum von Station k würde sich fortlaufend weiter anfüllen mit Kunden, die auf Bearbeitung warten (real natürlich nur bis zu einer physikalisch gegebenen Grenze, die aber in unserem Modell nicht enthalten ist). Wie groß auch immer die Besetzung dieses Warteraums wird, an der Abgangsrate $c_{k,max}$ der Station k ändert sich nichts mehr: Die Station arbeitet ja bereits "mit voller Kraft". Da nun alle Durchsätze gemäß Verkehrsflußgleichgewicht (4.2.27) fest aneinander gebunden sind und Stationen $i \neq k$ auch die aus $c_k = c_{k,max}$ resultierenden Ankunftsraten c_i durchaus bewältigen, bleibt auch der Systemdurchsatz (jetzt am Ausgang des Systems als Abgangsrate System gemessen) an den Maximalwert $c_{k,max}$ gebunden. Insgesamt: Steigt die Ankunftsrate am System (Abgangsrate Umgebung) c_0 über den Wert $c_{0,max}$, dann nimmt die Abgangsrate vom System (Ankunftsrate Umgebung) a_0 den festen Wert $c_{0,max}$ an, das Verkehrsflußgleichgewicht ist gestört ($a_0 \neq c_0$); intern wächst die Kundenpopulation an Station k über alle Grenzen.

4.2.38 **4.2.38:**

Nein! Beschleunigen Sie nämlich eine Station i, die nicht die Flaschenhals-Station ist, dann sinkt dadurch zwar deren Auslastung

$$b_{i,neu} (= \overline{S}_{i,neu} \cdot X_i c_0) < b_{i,alt} (= \overline{S}_{i,alt} \cdot X_i c_o)$$

was aber auf den Durchsatz der Flaschenhals-Station keinerlei Einfluß hat (alle X. bleiben gleich!), somit auch nicht auf die Tatsache, daß die Flaschenhals-Station für $c_{0,max}$ voll ausgelastet ist. Wenn Sie eine Verbesserung des maximalen

Systemdurchsatzes erreichen wollen, dann müssen Sie schon die Flaschenhals-Station beschleunigen: Trage diese Station den Index k, dann ist ja durch diese Maßnahme

$$\overline{S}_{k,neu} \cdot X_k < \overline{S}_{k,alt} \cdot X_k$$

erreicht, und der Wert $c_{0,max}$ ist auf einen höheren Wert festgelegt (sowohl wenn Station k Flaschenhals bleibt, als auch, wenn eine andere Station Flaschenhals wird; vgl. (4.2.33c) mit j*=0).

4.3.2: ***4.3.2***

- Die RR-Disziplin "merkt sich" (als Zustand) alle anwesenden Kunden sowie eine lineare Ordnung über diesen Kunden. Wenn Kunden vorhanden sind, wird deren Erster (gemäß festgelegter Ordnung) mit fester Bediengeschwindigkeit r bedient - alle anderen warten, haben also Bediengeschwindigkeit 0. Wird der bediente Kunde während der zugestandenen Bedienzeitscheibe fertig, verläßt er die Station (bzgl. des lokalen Stationszustands wird er also "vergessen") - der (gemäß festgelegter Ordnung) Nächste ist jetzt Erster und wird mit Geschwindigkeit r bedient. Wird der Bediente innerhalb der Zeitscheibe nicht fertig, rutscht er "ans Ende" der linear angeordneten Kunden (hat also Geschwindigkeit 0) - der Nächste rückt vor (bekommt also Geschwindigkeit r). Ein Neuankömmling an der Station stellt sich "am Ende" an.

- Die Verzögerungsstation kennt alle Anwesenden und weist jedem von ihnen eine konstante Bediengeschwindigkeit von z.B. r (AE / ZE) zu.

4.3.3: ***4.3.3***

Die Grundidee war ja die der Messung am System. Sollte also einer der Kunden (obwohl an Station i fertig bedient und mit dem Wunsch, nach Station j zu wechseln) aufgrund einer räumlichen Überlastung der Zielstation (Station j) nicht wechseln können, dann verbleibt er wohl an der Quellstation (Station i) und wird auch nicht als Abgang bzw. Ankunft gezählt: Die Messungen beinhalten bereits alle Blockierungseffekte. Andererseits: Wenn wir, von einer Messung ausgehend, zur Ermittlung von Flaschenhälsen hypothetisch Zugangsrate (beim offenen System) oder Kundenzahl (beim geschlossenen System) erhöhen, dann verfälschen Blockierungen unsere Resultate: So könnte z.B. eine Quellstation, die einen fertigen Kunden "nicht los wird", in ihrer weiteren Arbeit anhalten müssen: Die Station arbeitet nicht, obwohl u.U. Arbeit ansteht - eine unserer Grundvoraussetzungen ("arbeitserhaltend") ist verletzt, und zwar mit steigendem Verkehr immer mehr. Es gibt weitere ähnliche Effekte; insgesamt sind die Flaschenhalsüberlegungen nur bei Ausschluß von Blockierungen anwendbar.

4.4.4: ***4.4.4***

• Dichtefunktion:

$$fA(a) = dFa(a) / da$$

$$= \begin{cases} \lambda \exp(-\lambda a) & a \geq 0 \\ 0 & a < 0 \end{cases}$$

• Erwartungswert:

$$E[A] = \int_{-\infty}^{\infty} a\, f_A(a)\, da$$

$$= \int_{0}^{\infty} a\, \lambda \exp(-\lambda a)\, da$$

partielle Integration:

$$= \lambda \left[a\,(-1/\lambda) \exp(-\lambda a) - \int (-1/\lambda) \exp(-\lambda a)\, da \right]_0^{\infty}$$

$$= \left[-a \exp(-\lambda a) + (-1/\lambda) \exp(-\lambda a) \right]_0^{\infty}$$

$$= 0 - (0 - 1/\lambda)$$

$$= 1/\lambda$$

- Zweites Moment:

$$E[A^2] = \int_{-\infty}^{+\infty} a^2 f_A(a)\, da$$

$$= \int_{0}^{\infty} a^2 \lambda \exp(-\lambda a)\, da$$

partielle Integration:

$$= \lambda \left[a^2\,(-1/\lambda) \exp(-\lambda a) - \int 2a\,(-1/\lambda) \exp(-\lambda a)\, da \right]_0^{\infty}$$

$$= 0 + (2/\lambda) \int_{0}^{\infty} a\, \lambda \exp(-\lambda a)\, da$$

s. oben:

$$= (2/\lambda)\, E[A] = 2/\lambda^2$$

- Varianz:

$$V[A] = E[(A - E[A])^2] = E[A^2] - E^2[A]$$

s. oben: $= 2/\lambda^2 - 1/\lambda$

$= 1/\lambda^2$

- Variationskoeffizient:

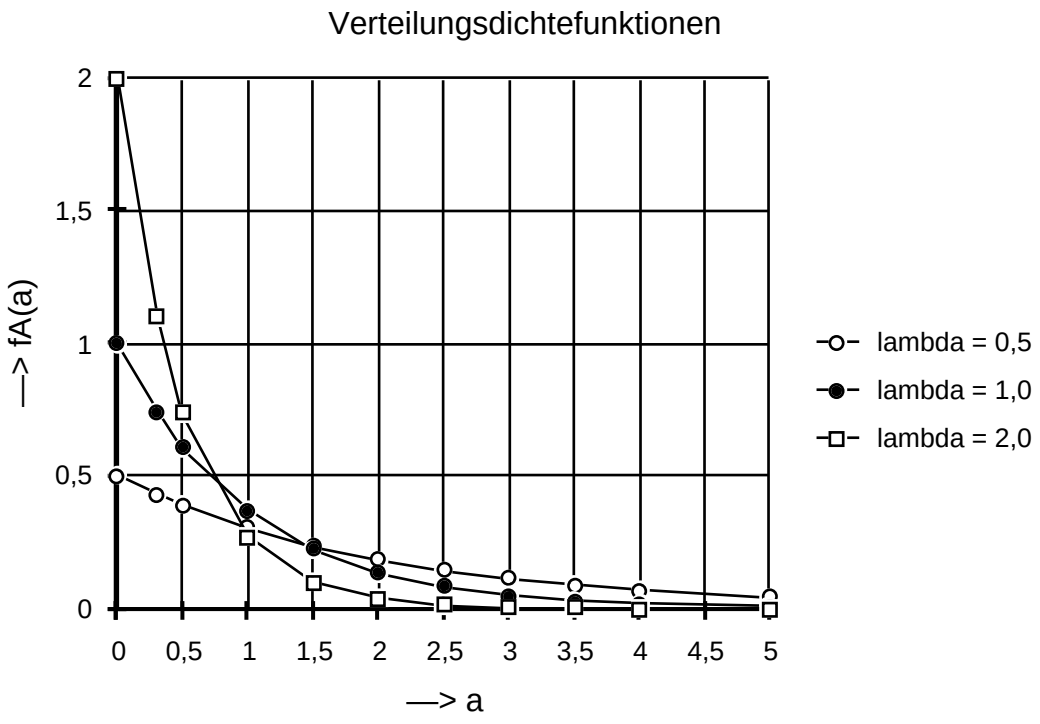
$$VK[A] = E[A]/\sqrt{V[A]}$$

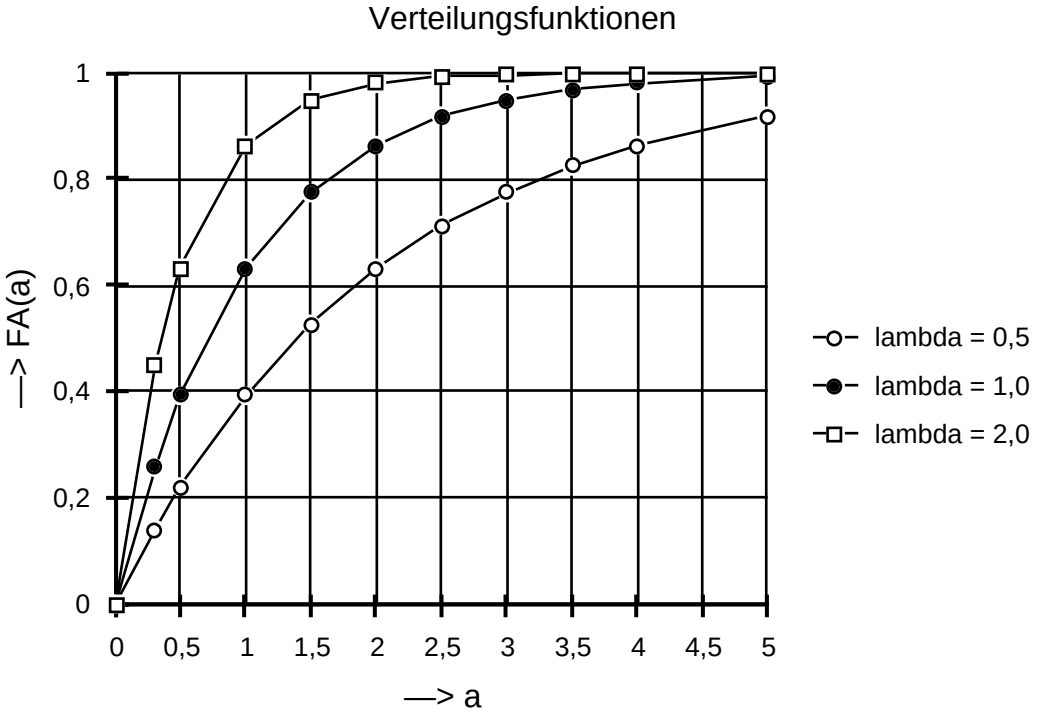
s. oben: $= 1$

- Eine kleine Wertetafel zur Erstellung der geforderten Graphen:

	λ= 0,5			λ= 1,0			λ= 2,0		
a	$e^{-\lambda a}$	$1-e^{-\lambda a}$	$\lambda e^{-\lambda a}$	$e^{-\lambda a}$	$1-e^{-\lambda a}$	$\lambda e^{-\lambda a}$	$e^{-\lambda a}$	$1-e^{-\lambda a}$	$\lambda e^{-\lambda a}$
0,0	1,000	0,000	0,500	1,000	0,000	1,000	1,000	0,000	2,000
0,3	0,861	0,139	0,430	0,741	0,259	0,741	0,549	0,451	1,098
0,5	0,779	0,221	0,389	0,607	0,393	0,607	0,368	0,632	0,736
1,0	0,607	0,393	0,303	0,368	0,632	0,368	0,135	0,865	0,271
1,5	0,472	0,528	0,236	0,223	0,777	0,223	0,050	0,950	0,100
2,0	0,368	0,632	0,184	0,135	0,865	0,135	0,018	0,982	0,037
2,5	0,287	0,713	0,143	0,082	0,918	0,082	0,007	0,993	0,013
3,0	0,223	0,777	0,112	0,050	0,950	0,050	0,002	0,998	0,005
3,5	0,174	0,826	0,087	0,030	0,970	0,030	0,001	0,999	0,002
4,0	0,135	0,865	0,068	0,018	0,982	0,018	0,000	1,000	0,001
5,0	0,082	0,918	0,041	0,007	0,993	0,007	0,000	1,000	0,000

• Graphen der Verteilungsdichten fA(a) und der Verteilungsfunktionen FA(a)





4.4.9 **4.4.9:**

- zu (4.4.8b):

$$FA_3(a) = \int_0^a fA_1(x)\ FA_2(a-x)\ dx$$

mit (4.4.3,4.4.8a)

$$= \int_0^a \lambda \exp(-\lambda x)\ \{\ 1 - [1+\lambda(a-x)]\ \exp[-\lambda(a-x)]\ \}\ dx$$

$$= \int_0^a \lambda \exp(-\lambda x)\, dx - \lambda \int_0^\infty [1+\lambda(a-x)]\ \exp[-\lambda a]\ dx$$

mit (4.4.2a):

$$= 1 - \exp(-\lambda a) - \lambda \exp(-\lambda a)\ \left[(1+\lambda a)x - \lambda x^2/2\right]_0^a$$

$$= 1 - \exp(-\lambda a) - \lambda \exp(-\lambda a)\ [a + \lambda a^2 - \lambda a^2/2]$$

$$= 1 - [1+\lambda a+\lambda^2 a^2/2]\ \exp(-\lambda a)$$

- zu (4.4.8c):

$$FA_{k+1}(a) = \int_0^a fA_1(x)\, FA_k(a-x)\ dx$$

mit (4.4.3,4.4.8c):

$$FA_{k+1}(a) = \int_0^a \lambda \exp(-\lambda x)\ \{1-\exp[-\lambda(a-x)] \sum_{i=0}^{k-1} [\lambda(a-x)]^i/i!\ \}\ dx$$

$$= 1 - \exp(-\lambda a) - \lambda \exp(-\lambda a) \int_0^a \Sigma \ldots\ dx$$

$$= 1 - \exp(-\lambda a) - \lambda \exp(-\lambda a)\ \left[\sum_{i=0}^{k-1} \frac{[\lambda(a-x)]^{i+1}}{-\lambda\,(i+1)!}\right]_0^a$$

$$= 1 - \exp(-\lambda a) - \lambda \exp(-\lambda a) \sum_{i=0}^{k-1} \frac{(\lambda a)^{i+1}}{\lambda\,(i+1)!}$$

$$= 1 - \exp(-\lambda a)\,[1+\sum_{i=1}^{k} (\lambda a)^i/i!\,]$$

$$= 1 - \exp(-\lambda a) \sum_{i=0}^{k} (\lambda a)^i/i!$$

4.4.10: ***4.4.10***

- Dichte:

$$fA_k(a) = dFA_k(a)/da$$

$$= \lambda \exp(-\lambda a) \sum_{0}^{k-1} (\lambda a)^i/i! \ - \exp(-\lambda a) \sum_{1}^{k-1} i\,\lambda\,(\lambda a)^{i-1}/i!$$

$$= \lambda \exp(-\lambda a)\ \{\sum_{0}^{k-1} (\lambda a)^i/i! - \sum_{1}^{k-1} (\lambda a)^{i-1}/(i-1)!\ \}$$

$$= \lambda \exp(-\lambda a)\ \{\sum_{0}^{k-1} (\lambda a)^i/i! - \sum_{0}^{k-2} (\lambda a)^i/i!\ \}$$

$$= \lambda \exp(-\lambda a)\ \frac{(\lambda a)^{k-1}}{(k-1)!}$$

- Erwartungswert: $E[A_k] = \int \ldots$
  Summe unabh. ZV: $= k \cdot E[A_1]$
  mit (4.4.3): $= k / \lambda$

- Varianz: $V[A_k] = E[(A_k - E[A_k])^2]$
  Summe unabh. ZV. $= k \cdot V[A_1]$
  mit (4.4.3): $= k / \lambda^2$

- zw. Moment: $E[A_k^2] = V[A_k] + E^2[A_k]$

  s. oben: $= k / \lambda^2 + k^2 / \lambda^2$
  $= k(1+k) / \lambda^2$

- Variationskoeffizient:

$$VK[A_k] = \sqrt{V[A_k]}\ / E[A_k]$$

s. oben

$$VK[A_k] = (\sqrt{k}/\lambda) / (k/\lambda)$$
$$= 1/\sqrt{k}$$

4.4.12 **4.4.12:**

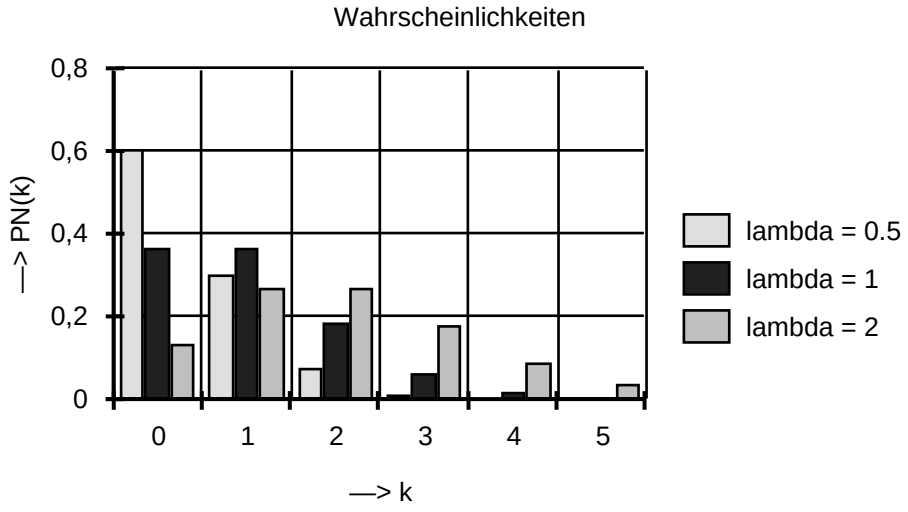
Für t=1 wird aus (4.4.11b):

$PN(k) = (\lambda^k / k!) \cdot \exp(-\lambda)$

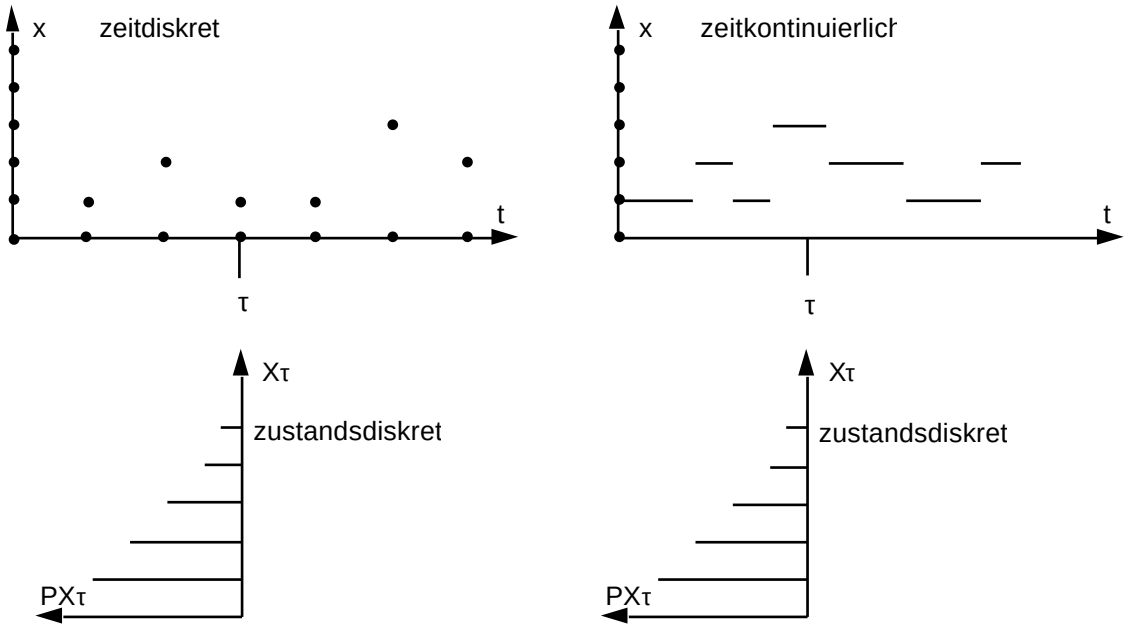
Die Tabelle der Werte für

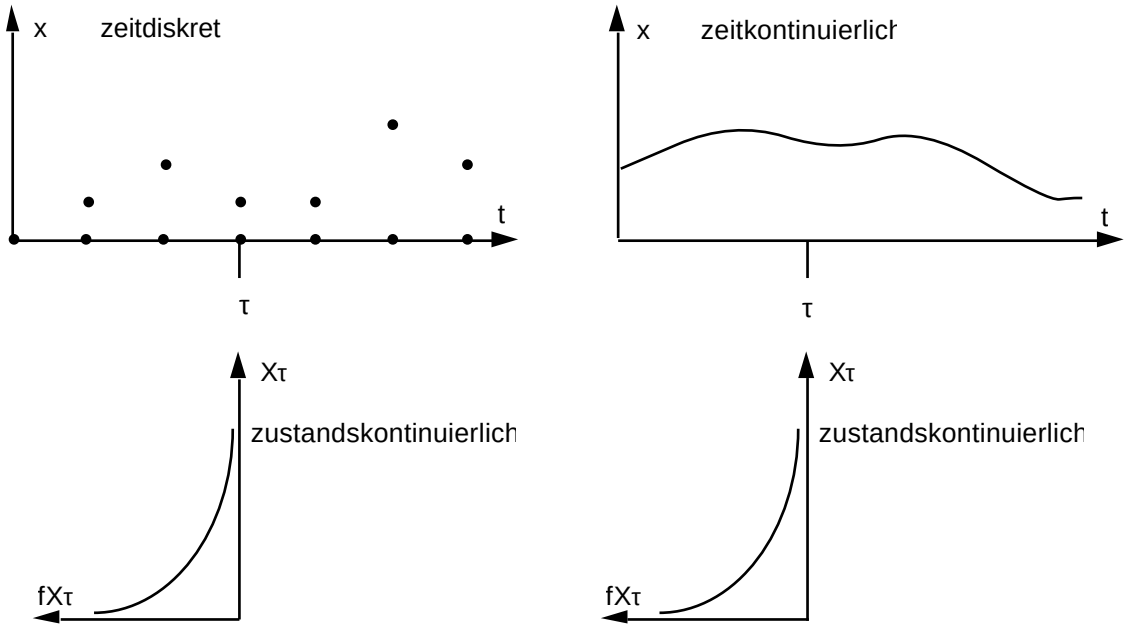
k	λ=1 PN(k)	λ=0.5 PN(k)	λ=2 PN(k)
0	0,368	0,607	0,135
1	0,368	0,303	0,271
2	0,184	0,076	0,271
3	0,061	0,013	0,180
4	0,015	0,002	0,090
5	0,003	0,000	0,036

Die graphische Darstellung:



4.5.2 ***4.5.2***





4.5.6 ***4.5.6***

- Stationswechsel:
  Seien die Kunden, die Station i verlassen, in der Reihenfolge des Verlassens und beginnend mit einem beliebigen Kunden, fortlaufend durchnumeriert, etwa startend mit Nr. 1. Sei Z_k die Zielstation des k-ten (die Station i verlassenden) Kunden; Z_k ist gemäß (4.4.1) eine ZV; die Sequenz (Z_k;k **N**) ist ein stochastischer Prozeß. Zustandsraum des Prozesses ist I, die (endliche) Menge vorhandener Stationen, bzw. I {0}: Der Prozeß ist zustandsdiskret. Parametermenge des Prozesses ist **N**: Der Prozeß ist zeitdiskret. Die ZV sind wechselsei-

tig unabhängig: Es handelt sich um einen unabhängigen Prozeß. Wegen der wechselseitigen Unabhängigkeit und der identischen Verteilung aller Z_k ist der Prozeß auch stationär.

- Ankunftsabstände:
  Sei A_1 die Zeit zwischen einer beliebigen Kundenankunft bis zur ersten darauf folgenden, A_2 die Zeit zwischen dieser Ankunft und der darauf folgenden, u.s.f.: Ankunftsintervalle seien also sequentiell durchnumeriert. Die A_k sind ZV, $(A_k;k \in \mathbf{N})$ ist stochastischer Prozeß. Zustandsraum ist $\mathbf{R}^+$ (Wertebereich von Zwischenankunftsabständen): Der Prozeß ist zustandskontinuierlich. Parametermenge des Prozesses ist **N**: Der Prozeß ist zeitdiskret. Ankunftsabstände sind wechselseitig unabhängig: Wir haben einen unabhängigen Prozeß vor uns. Darüber hinaus sind Ankunftsabstände als identisch verteilt vorausgesetzt: Der Prozeß ist stationär.
- Bedienwünsche:
  Numeriere die Bedienwünsche W bzgl. Station i sequentiell, beginnend mit beliebiger Ankunft. $(W_k;k \in \mathbf{N})$ ist zustandskontinuierlicher, zeitdiskreter, unabhängiger, stationärer Prozeß.

4.5.7 **4.5.7:**

Offensichtlich ist 0 der Startzustand des Prozesses, d.h. $P[N_0=0]=1$, und der Prozeß insgesamt der Werte aus $\mathbf{N_0}$ fähig: Es ist ein zustandsdiskreter Prozeß. Wir betrachten ihn in reeller Zeit: Der Prozeß ist zeitkontinuierlich. Wegen der Eigenschaft "hochzählend" ist sicher die Verteilung von N_{t+s}, $s>0$, abhängig von der Realisierung n_t zum Zeitpunkt t: Der Prozeß ist nicht unabhängig. Gleichermaßen (Tendenz: Werte fortlaufend steigend) kann der Prozeß nicht stationär sein. Dagegen ist der Prozeß Markoff'sch: N_{t+s} hängt von n_t ab, aber (Gedächtnislosigkeit!) nicht zusätzlich von n_τ, $\tau<t$. In Formeln haben wir $(N_t;t \in \mathbf{R}^+)$ in den Beziehungen (4.4.11) bereits beschrieben.

4.5.22 **4.5.22:**

Es ist gemäß angegebener Formel

$$\begin{aligned} fC(c) &= [1-FB(c)]\cdot fA(c) + [1-FA(c)]\cdot fB(c) \\ &= \exp(-\lambda_B\cdot c)\cdot \lambda_A\cdot\exp(-\lambda_A\cdot c) + \exp(-\lambda_A\cdot c)\cdot \lambda_B\cdot\exp(-\lambda_B\cdot c) \\ &= (\lambda_A+\lambda_B)\cdot\exp[-(\lambda_A+\lambda_B)\cdot c] \end{aligned}$$

C ist also wieder exponentiell verteilt, mit Parameter $\lambda_C=\lambda_A+\lambda_B$.

Dies ist eine Bestätigung der Überlegungen zu (4.5.21). Zwei exponentielle Phasen "laufen" im Zustand (angenommene Bezeichnung:) k; daher ist q_{kk} gleich der negativen Summe der beiden Phasenparameter, also $q_{kk} = -(\lambda_A+\lambda_B)$; $-q_{kk}$ ist Parameter der exponentiell verteilten Lebensdauer von Zustand k.

4.6.8 **4.6.8:**

- Zu RANDOM, LCFS:
  Der Populationsprozeß $(N(t);t \in \mathbf{R}^+)$ an der Station anwesender Kunden, insbesondere auch seine stationäre Phase, war ja für M / M / 1-FCFS durch das Übergangsratendiagramm vor (4.6.1) charakterisiert. Umgekehrt: Sofern sich an diesem Diagramm nichts ändert (sofern also "andere" Probleme durch dasselbe Diagramm erfaßt werden können), ändert sich natürlich auch nichts an den daraus abgeleiteten Resultaten (4.6.2,4.6.6). Das Charakteristische dieses Dia-

gramms war, daß für jeden Zustand n $\mathbf{N_0}$ die Zugangsrate λ betrug und für jeden Zustand n **N** die Abgangsrate μ herrschte. Dies gilt allerdings nicht nur für FCFS, sondern auch für RANDOM und LCFS. Ja, noch darüber hinaus: Wir hätten auch immer den Zweiten oder den Vorletzten aus der Warteschlange auswählen können oder mit identischer Wahrscheinlichkeit einen aus den ersten drei Wartenden, o.ä.: An dem Diagramm ändert sich nichts, die Resultate bleiben erhalten. Um Mißverständnissen vorzubeugen: Dies heißt nicht, daß diese Resultate immer (für alle Disziplinen) gelten; sie verändern sich, wenn in unsere Auswahl aus der Warteschlange die tatsächlichen (zukünftigen) individuellen Bedienwünsche mit eingehen (etwa bei der Disziplin SJN, Shortest Job Next, die immer den Kunden geringsten Bedienwunsches unter den Wartenden auswählt).

- Zu LCFS-PR:
  Wenn ein in Bedienung befindlicher Kunde durch einen Neuankömmling unterbrochen wird, reiht er sich wieder in die Warteschlange ein; er hat dann (später irgendwann) noch seinen "restlichen" Bedienwunsch abzuwarten. Dieser Rest ist aber (Exponentialverteilung, Gedächtnislosigkeit) genauso verteilt, wie der Bedienwunsch eines Noch-Nicht-Bedienten, vgl. (4.4.5), so daß er in der Warteschlange nicht eigens markiert werden muß: Diagramm und Ergebnisse bleiben wie gehabt.

4.6.10c: ***4.6.10c***

Aus (4.6.9) erhalten wir für die M / M / m-Stationen durch Setzung von $\mu_n = c_n \cdot \mu$ sowie $c_n = n$ ($n \le m$) und $c_n = m$ ($n \ge m$) die stationäre Verteilung

$$p_n = p_0 \lambda^n / \prod_{i=1}^{n} (c_i \mu)$$

für $n < m$:

$$= p_0 (\frac{\lambda}{\mu})^n / (n!)$$

für $n \ge m$:

$$= p_0 (\frac{\lambda}{\mu})^n / (m!\, m^{n-m})$$

Für p_0 ergibt sich:

$$p_0 = \{1 + \sum_{n=1}^{m-1} (\lambda/\mu)^n / n! + \frac{m!}{m^m} \sum_{n=m}^{\infty} (\frac{\lambda}{\mu\, m})^n\}^{-1}$$

$$= \{1 + \sum_{n=1}^{m-1} (\lambda/\mu)^n / n! + \frac{(\lambda/\mu)^m}{(m-1)!\,(m - \lambda/\mu)}\}^{-1}$$

(falls $\lambda / (\mu \cdot m) < 1$).

LEERSEITE