

3. Netze aus exponentiellen Stationen

Dieses Kapitel widmet sich der Spezifikation und Analyse einer bestimmten Familie stochastischer Verkehrsnetze. Diese Familie, soviel sei vorausgeschickt, ist dadurch gekennzeichnet, daß (Kunden-)Bedienwünsche bezüglich (Netz-)Stationen jeweils "exponentiell verteilt" sind.

Überblick

Unser Studienobjekt "stochastisches Verkehrsnetz" hatten wir, in allgemeiner Form, bereits in Abschn. 2.3 kennengelernt. Hier wird es nun zunächst darum gehen müssen, die spezielle Klasse "Netze exponentieller Stationen" zu konkretisieren, d.h. im Rahmen der Annahmen 2.3.1/2.3.4 genauere Festlegungen zu treffen. Insbesondere wird es um die Charakterisierung verschiedener Zufallsvariablen gehen, die in der Last-Komponente des Netzes Verwendung finden. Zu diesem Zweck betrachten wir in Abschn. 3.1 eine Reihe für uns interessanter Verteilungen, dabei auch solche, die erst in Kapitel 4 Einsatz finden werden (dort werden wir eine allgemeinere Klasse von stochastischen Netzwerken studieren).

*Charakterisierung
benötigter ZV*

Somit in den Stand gesetzt, stochastische Verkehrsnetze genau zu spezifizieren, werden wir uns um deren Analyse kümmern. Auch diese hatten wir in den Grundzügen bereits in Abschn. 2.3 skizziert: Es geht um den Verlauf des "Zustands" des zu analysierenden dynamischen Systems über der Zeit, bzw. um das Studium des mathematischen Modells dieses Zustandsverlaufs in Form eines stochastischen Prozesses $(Z(t); t \in T)$. Als Vorbereitung auf "konkrete Taten" widmen wir uns in Abschn. 3.2 einer speziellen Familie stochastischer Prozesse, den sog. "zeitkontinuierlichen Markoff-Ketten"; wir werden sehen (bzw.: Sie werden sich daran erinnern), daß diese die angenehme Eigenschaft haben, unter gewissen Vorbedingungen eine "stationäre Phase" anzustreben, d.h. für hinreichend große Zeiten t eine Charakterisierung zuzulassen, die sowohl vom Anfangszustand der Kette (z.B. für $t=0$) als auch vom spezifischen Beobachtungszeitpunkt t unabhängig ist. Wir werden ferner Techniken erarbeiten, die Charakterisierung dieser stationären Phase, die sog. "stationäre Verteilung" des Prozesses, für konkrete Fälle abzuleiten.

*Analyse zeit-
kontinuierlicher
Markoff-Ketten*

In Anwendung dieser Techniken wenden wir uns (endlich wieder!) konkreten Problemen zu: In Abschn. 3.3 dem Studium einzelner, "exponentieller" Stationen, in Abschn. 3.4 dem Studium von Netzen derartiger Stationen, "exponentiellen Netzen". Es wird uns gelingen, in diesen Fällen explizite Formeln für die stationäre Zustandsverteilung der betrachteten Modelle (Systeme) anzugeben. Die stationäre Verteilung wiederum ist eine geeignete Grundlage zur Ableitung von Leistungscharakteristika wie Durchsätzen, Verweilzeiten etc.

*Analyse
exponentieller
Stationen
und Netze*

Listen wir die Lernziele dieses Kapitels im einzelnen auf:

Lernziele

- Sie werden die Fähigkeit erwerben, (gewisse) praktische Probleme auf Markoff-Ketten abzubilden;
- Sie werden die Fertigkeit einüben, die stationäre Verteilung dieser Ketten zu ermitteln;
- Sie werden (praktisch anwendbare!) Resultate für eine relativ komplexe Problemklasse, nämlich die exponentiellen Netze, kennenlernen.

In Wiederholung nochmals der Hinweis, daß es hier nicht darum gehen kann, Ihnen die Grundlagen der Stochastik vorzutragen. Kenntnisse etwa im Umfang von Dörf78 werden vorausgesetzt. Wer sich über diese Literaturstelle hinaus stärker mit den maßtheoretischen Grundlagen der Stochastik auseinandersetzen möchte, sei auf Baue74 verwiesen. Eine sehr gute Vertiefung in die Theorie sto-

*vorausgesetzte
+ vertiefende
Literatur*

chastischer Prozesse bietet Cinl75. Und: Geradezu ein Klassiker auf dem von uns behandelten Gebiet ist Klei75, der in sehr lesbarer Form sowohl Grundlagen als auch für uns relevante Anwendungen der Stochastik beschreibt.

3.1 Ausgewählte Verteilungen

Wir hatten in Abschnitt 2.3 eine allgemeine Charakterisierung stochastischer Verkehrsnetze vorgenommen und schließlich unsere Vorstellungen von Last- und Maschinenkomponenten dieser Netze in den Annahmen 2.3.1 und 2.3.4 festgehalten. So behandelte Ann. 2.3.1 die Spezifikation der Lastkomponente stochastischer Verkehrsnetze, innerhalb derer

*benötigte
Zufallsvariable
in der
Lastkomponente*

- (i) die Spezifikation des Bedienwunschumfangs eines an einer Station i eintreffenden Kunden durch eine Zufallsvariable W_i geschah;
- (ii) die Spezifikation des nächsten Besuchsziels eines die Station i verlassenden Kunden mittels einer Zufallsvariablen (bisher nicht festgelegten Namens) erfolgte;
- (iii) die Spezifikation der (im offenen System) zwischen zwei aufeinanderfolgenden Kundenankünften verstreichenden Zeitspanne über eine Zufallsvariable A vorgenommen wurde.

Nur der Vollständigkeit halber: Für die Spezifikation der Maschinenkomponente eines stochastischen Verkehrsnetzes hatten wir Zufallsvariable nicht vorgesehen; solche sollten in diesem Fall auch nicht nötig sein, haben wir doch i. allg. eine durchaus deterministische Vorstellung von der Existenz und Vernetzung der Stationen sowie von den stationsspezifischen Bediendisziplinen und ihrer Funktionsweise (vgl. Ann. 2.3.4).

Rekapitulieren wir doch noch einmal, um festen Boden unter den Füßen zu haben, ganz kurz den Begriff "Zufallsvariable": Die Zufallsvariable ist eine Abbildung aus einer Menge von "Elementarereignissen" in eine "andere" Menge, nämlich den "Wertebereich" dieser Zufallsvariablen. Letzterer ist i. allg. numerisch, d.h. er besteht aus einer (endlichen oder abzählbaren oder überabzählbaren) Menge von Zahlen. Über der Menge von Elementarereignissen ist ein (auf 1 normiertes) Maß definiert, das sog. Wahrscheinlichkeitsmaß. Dieses Maß wird mittels erwähnter Abbildung in ein Maß über dem Wertebereich der ZV übertragen (damit dies sauber funktioniert, sind sowohl über der Menge der Elementarereignisse, als auch über dem Wertebereich der ZV gewisse "zusammenpassende" algebraische Strukturen definiert). Wir arbeiten letztlich, wenn wir mit einer ZV arbeiten, mit deren Wertebereich und einem darüber definierten Wahrscheinlichkeitsmaß, das konkret "die Wahrscheinlichkeiten" (in statistischer Interpretation: die "relativen Häufigkeiten") dafür angibt, daß die ZV einen Wert innerhalb bestimmter Untermengen ihres Wertebereichs annimmt. Dieses Maß, d.h. die Festlegung all solcher Wahrscheinlichkeiten, heißt auch die "Verteilung" der ZV. Meist machen wir uns über die zugrundeliegenden Elementarereignisse (samt zugehörigen Festlegungen: den zugrundeliegenden Wahrscheinlichkeitsraum) nicht viele Gedanken. Denken Sie nur an das übliche Beispiel mit dem Werfen eines Würfels, wo wir vom "Würfeln einer Zwei" sprechen und die Abbildung vom Elementarereignis (nämlich: Man sieht das Muster "••") auf die zugeordnete Zahl (nämlich: "2") ohne viel Aufhebens vollziehen. Die Charakterisierung des W -Maßes P (der Verteilung) einer Zufallsvariablen V kann auf verschiedenste Weise geschehen; bei (eindimensionalen) numerischen Zufallsvariablen etwa durch ihre "Verteilungsfunktion"

*Begriff
Zufallsvariable*

Verteilung

$$FV(v) := P[V \leq v]$$

*Verteilungs-
funktion*

Wahrscheinlichkeiten Daneben bei "diskreten" ZV (abzählbarer Wertebereich, z.B. N_0), durch die "Menge der Wahrscheinlichkeiten"

$$PV(v) := P[V=v]$$

Wahrscheinlichkeitsdichte oder, bei "kontinuierlichen" ZV (überabzählbarer, kontinuierlicher Wertebereich, z.B. R^+), durch die "Wahrscheinlichkeitsdichte(-funktion)", falls existent

$$fV(v) := dFV(v) / dv$$

Spezifikation Stationswechsel

Sehen wir uns die für Zwecke (i) bis (iii) benötigten ZV der Reihe nach an und beginnen mit (ii), der Spezifikation der Stationswechsel. Für jede Station i ($i \in I$; I war die Menge der vorhandenen Stationen) gibt es eine ZV H_i , die das nächste Besuchsziel eines (jedes beliebigen!) die Station i verlassenden Kunden angibt. Als Wertemenge für H_i kommt also wieder die (diskrete, endliche) Menge I der Stationen ins Spiel (bzw. eine Untermenge von I , vgl. unten). Die Verteilung jeder ZV H_i , $i \in I$, (d.h., die Gesetzmäßigkeiten, nach denen H_i einen der Werte aus der festgelegten Wertemenge I annimmt) läßt sich somit charakterisieren durch die Menge der Wahrscheinlichkeiten

$$(3.1.1a) \quad \{PH_i(j); j \in I\}$$

Damit (3.1.1a) eine Verteilung charakterisiert, muß selbstverständlich

$$(3.1.1b) \quad \begin{aligned} PH_i(j) &\geq 0 \quad j \in I \\ \sum_{j \in I} PH_i(j) &= 1 \end{aligned}$$

gelten. Ansonsten ist die Wahl der Verteilung von H_i in keiner Weise eingeschränkt und kann voll auf das darzustellende Problem ausgerichtet werden.

Einige ergänzende Anmerkungen:

Wertemenge für H_i

- Kommen als Zielstationen eines die Station i verlassenden Kunden (vom Problem her gesehen!) gewisse Stationen überhaupt nicht in Frage, dann müßten wir streng genommen die Wertemenge der ZV H_i entsprechend einschränken, also, wie oben angedeutet, mit einer Wertemenge J_i für H_i arbeiten, die nur die tatsächlich auftretenden Zielstationen enthält. Wir belassen es aber bei der vollen Wertemenge I für alle H_i , $i \in I$, und erfassen die Einschränkung durch verschwindende Wahrscheinlichkeiten, also

$$PH_i(j) = 0 \quad j \notin J_i$$

was nicht streng identisch ist, aber für unsere Zwecke den gleichen Effekt hat.

offenes System

- Bei einem offenen System taucht als Quell- und Zielstation (neben den "normalen" Stationen $i \in I$) auch die Pseudostation "Umwelt" mit dem besonderen Bezeichner "0" auf (vgl. Ann. 2.3.1). Entsprechend haben wir es in diesem Fall mit einer zusätzlichen Zufallsvariablen H_0 zu tun, welche die erste zu besuchende Station nach Systemeintritt charakterisiert; entsprechend erstreckt sich der Wertebereich aller H_i in diesem Fall auf die Menge $I \cup \{0\}$, wo die $PH_i(0)$ die Wahrscheinlichkeiten erfassen, nach Station i das System zu verlassen. Verabredungsgemäß ist ferner $PH_0(0) = 0$.

- Sie sind es wahrscheinlich nicht gewohnt, ZV mit nicht-numerischem Wertebereich zu verwenden (wir hatten ja den Wertebereich I verwendet, also eine Menge von Stationsnamen). Dies sollte Sie nicht stören, ist doch die Bedeutung der ZV H_i klar herausgearbeitet worden. Sie sollten sich aber nicht täuschen lassen von der in der Literatur häufig verwendeten Bezeichnung der Stationen mit ganzen Zahlen: Diese läßt sich einfach genug erreichen, indem wir z.B. eine willkürliche, aber feste Ordnung über die Stationen legen; bzgl. dieser Ordnung gibt es dann eine erste, zweite, ..., M-te Station (es war $|I| = M$); die Umgebung könnten wir z.B. "vorn", als "nullte" Station einreihen. Unter Verwendung dieser Ordnungszahlen als (ganzzahlige) Stationskennungen erhalten wir zwar nun Zufallsvariable $H_i, i \in \{0, 1, \dots, M\}$, mit Wertebereichen $\{0, 1, \dots, M\}$ - den "Namenscharakter" verlieren die Ordnungszahlen aber dadurch nicht. Um sich das auf intuitiver Basis klarzumachen, vergleichen Sie etwa
 - die Situation eines Spiels mit zwei möglichen Resultaten "head" und "tail", wobei Resultat head Ihnen einen Gewinn von 1 DM, Resultat tail einen von 2 DM bringt (es ist ein sehr günstiges Spiel für Sie) und die Wahrscheinlichkeiten für Resultate head, tail bekannt sind, so daß Sie sich über den Erwartungswert einer numerischen ZV G (für "Gewinn") einen Eindruck vom mittleren Gewinn pro Spiel verschaffen können,
 - mit der Situation eines Kunden, der ein System aus zwei Stationen mit Namen "head" und "tail" betritt, wobei Station head die Nummer 1 trägt, Station tail die Nummer 2 und die Wahrscheinlichkeiten für die Ziele head, tail bekannt sind, so daß Sie rein formal den Erwartungswert der ZV H_0 (für "erste Zielstation") errechnen könnten, aber mit welchem Sinn? "Mittlere" erste Zielstation ??

nicht-numerische Zufallsvariable

Von der willkürlichen, festen Anordnung der Stationen und der Verwendung zugeordneter Ordnungszahlen werden auch wir aus notationellen Gründen durchaus Gebrauch machen - und haben dies bereits getan, als wir (in Ann. 2.3.1) als Notierung für die Verteilungen aller "Wechsel"-ZV H_i die Wechselmatrix $H = (h_{ij})$ verwendeten; der Zusammenhang mit der im vorliegenden Abschnitt verwendeten Notation ist dabei durch

Wechselmatrix

$$PH_i(j) := h_{ij} \quad \begin{matrix} i, j \in I \setminus \{0\} \\ \text{bzw. } \{0, 1, \dots, M\} \end{matrix}$$

gegeben.

Soweit zur Spezifikation der Stationswechsel, d.h. zu ZV für Zweck (ii). Gehen wir einen Schritt weiter und beschäftigen uns mit (iii), der Spezifikation von Kundenankünften (im offenen System). Wir hatten uns schon darauf festgelegt (Ann. 2.3.1), die Ankunftsgesetzmäßigkeiten durch eine (einzige!) ZV A zu erfassen, die den zeitlichen Abstand zwischen (je zwei beliebigen!) unmittelbar aufeinanderfolgenden Ankunftszeiten beschreibt. Ankunftsabstände sind, vom Problem her gesehen, positiv und reellwertig, so daß wir zu ihrer Charakterisierung kontinuierliche, auf \mathbf{R}^+ beschränkte ZV ins Auge fassen sollten.

Spezifikation Kundenankünfte

Die Charakterisierung der Ankunftsabstände geschieht in der Literatur (in einfachen Fällen) nahezu ausschließlich durch Annahme einer Exponentialverteilung für A, d.h. durch die Annahme, daß die Verteilung von A durch eine Verteilungsfunktion der Art

Exponentialverteilung

$$(3.1.2a) \quad FA(a) = P[A \leq a] = \begin{cases} 1 - \exp(-a) & a \geq 0 \\ 0 & a < 0 \end{cases}$$

beschrieben ist mit positivem reellem Parameter

$$(3.1.2b) \quad > 0$$

Rechtfertigung

Die bevorzugte Verwendung der Exponentialverteilung hat zwei Gründe: Einerseits, vom Problem her gesehen, stellt sie häufig eine durchaus realistische Wahl dar; andererseits, von der analytischen Behandlung her betrachtet, ist es eine sehr vorteilhafte Wahl. Exponentielle Ankunftsabstände implizieren nämlich einen sog. "Poisson'schen Ankunftsprozeß" (auf dessen konkrete Definition wir noch zu sprechen kommen), der einerseits ein realistisches Modell "seltener Ereignisse" darstellt und andererseits die vorteilhafte Eigenschaft der "Gedächtnislosigkeit" besitzt.

Diskussion Exponential- Verteilung

Machen wir uns etwas besser mit der Exponentialverteilung vertraut; sie besitzt

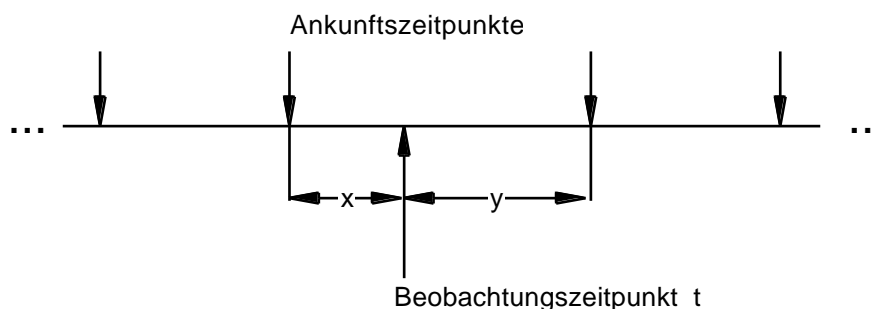
(3.1.3)	die Dichtefunktion	$f_A(a) = \lambda \exp(-\lambda a)$
	den Erwartungswert	$E[A] = 1/\lambda$
	das 2. Moment	$E[A^2] = 2/\lambda^2$
	die Varianz	$V[A] = 1/\lambda^2$
	den Variationskoeffizienten	$VK[A] = 1$

Testfrage

Testfrage 3.1.4: Bestätigen Sie auf der Basis von (3.1.2) die Ergebnisse (3.1.3); zeichnen Sie Graphen der Verteilungs- und Dichtefunktionen für die Parameterwerte $\lambda = 1, 0.5, 2$.

Gedächtnis- losigkeit

Was hat es nun mit den angekündigten Annehmlichkeiten der Exponentialverteilung (Stichwort: Gedächtnislosigkeit) auf sich? Stellen wir uns einen kontinuierlichen Ankunftsstrom mit exponentiell verteilten Ankunftsabständen vor, und versetzen wir uns in die Rolle eines Beobachters, der zu irgendeinem zufälligen Zeitpunkt t an diesen Strom herantritt und sich dafür interessiert, wie lange es bis zur nächsten Ankunft dauern wird. t liegt also im allgemeinen zwischen zwei Ankünften, sagen wir x Zeiteinheiten (ZE) nach der letzten Ankunft und y ZE vor der nächsten Ankunft:



Ebensowenig wie wir zu einem Ankunftszeitpunkt mit Sicherheit den nächsten kennen (aus diesem Grund haben wir ja Ankunftsabstände durch ZV charakterisiert), können wir x ZE nach einer Ankunft das Ende des laufenden Ankunftsintervalls mit Sicherheit voraussehen: Der "Rest" y muß seinerseits als ZV Y aufgefaßt und charakterisiert werden, also beispielsweise durch seine Verteilungsfunktion $F_Y(y)$. Sehen wir uns zu diesem Zweck (dies wird sich als einfacher herausstellen) die komplementäre Funktion $1 - F_Y(y) = P[Y > y]$ an. Aus obigem Dia-

gramm leiten wir leicht ab, daß

$$P[Y > y] = P[A > x + y \mid A > x]$$

daß Y also größer y ist, wenn A größer x+y ist unter der Bedingung, daß A die Länge x überhaupt überschreitet. Wir erhalten weiter

$$\begin{aligned} P[Y > y] &= P[A > x + y] / P[A > x] \\ &= (1 - FA(x + y)) / (1 - FA(x)) \\ \text{und mit (3.1.2)} &= \exp(-\lambda(x + y)) / \exp(-\lambda x) \\ &= \exp(-\lambda y) \end{aligned}$$

so daß

$$\begin{aligned} (3.1.5) \quad FY(y) &= 1 - P[Y > y] \quad y \geq 0 \\ &= 1 - \exp(-\lambda y) \\ &= FA(y) \end{aligned}$$

gilt. (3.1.5) enthält gleich zwei bedenkenswerte Aussagen: Zum einen ist die Länge des Restes Y offenbar unabhängig davon, wo der Beobachtungszeitpunkt t im Ankunftsintervall lag (x taucht in der Formel für FY(y) gar nicht mehr auf); zum anderen ist die Verteilung des Restes identisch mit der des gesamten Ankunftsintervalls (FY(y) = FA(y)). Dies ist erstaunlich: Sollte man nicht erwarten, daß Y um so kleiner ausfällt, je größer x war? Und: War es vorauszusehen, daß der nächste Ankunftszeitpunkt mit gleicher Wahrscheinlichkeit jenseits eines bestimmten zukünftigen Zeitpunktes liegt, gleichgültig, ob von einem Ankunftszeitpunkt oder einem Zeitpunkt im Innern eines Ankunftsintervalls aus gefragt wird?

Wir wundern uns zu Recht: Exponentiell verteilte ZV sind in dem Sinne "gedächtnislos", daß sie ihre jeweilige Vergangenheit "vergessen": x ZE nach Beginn eines exponentiell verteilten Intervalls sind alle Aussagen, die sich bezüglich des Restes (der "Zukunft") dieses Intervalls treffen lassen, unabhängig von x - und sogar identisch mit den Aussagen, die sich über das gesamte Intervall (vom "Startpunkt" aus) treffen ließen. Formulieren wir umgekehrt diese Gedächtnislosigkeit als Bedingung, suchen wir also nach einer Funktion $G(d) = P[D > d]$, die der Forderung

*Interpretation
Gedächtnis-
losigkeit*

$$(3.1.6a) \quad G(d_1 + d_2) = G(d_1) \cdot G(d_2)$$

nachkommt, dann ergibt sich als Lösung dieser Funktionalgleichung (ohne Beweis) einzig die Familie von Funktionen

$$G(d) = \exp(-\text{const} \cdot d)$$

Die Exponentialverteilungen, mit der Verteilungsfunktion

$$(3.1.6b) \quad FD(d) = 1 - G(d) = 1 - \exp(-\text{const} \cdot d)$$

sind also die einzigen kontinuierlichen Verteilungen, die in obigem Sinne gedächtnislos sind. Fürwahr eine Besonderheit!

Nur der Vollständigkeit halber: Bei den diskreten Verteilungen gibt es ein (einziges) "gedächtnisloses" Gegenstück zur Exponentialverteilung, die sog. geometrische Verteilung. Eine geometrisch verteilte ZV N (mit Wertebereich \mathbb{N}_0) ist cha-

*Geometrische
Verteilung*

Charakterisiert durch

$$P_N(n) = r^n \cdot (1-r) \quad 0 < r < 1, n \in \mathbb{N}_0$$

Wählen wir eine ähnliche Interpretation wie bei der Exponentialverteilung, indem wir die Länge eines Zeitintervalls beschreiben, das hier aus einer zufälligen Anzahl N aneinandergfügter Zeitscheiben fester (deterministischer) Länge zusammengesetzt ist, wo obige Verteilung die Anzahl dieser Zeitscheiben festlegt, dann drückt sich die Eigenschaft "gedächtnislos" wie folgt aus: Fragen wir, nach Ablauf einer beliebigen Anzahl von Zeitscheiben, nach der Wahrscheinlichkeit für eine Fortsetzung um (mindestens) eine weitere, und ist diese Wahrscheinlichkeit durch einen konstanten Wert r (unabhängig von der "Vergangenheit" des ablaufenden Zeitintervalls) gegeben, dann liegt die angegebene Verteilung vor.

*Zeit zwischen
Ankünften*

Um weitere interessante Verteilungen kennenzulernen (in einer Weise, die es uns erlauben wird, die angekündigte Definition des "Poisson-Prozesses" vorzunehmen), stellen wir uns folgende Frage: Gegeben sei ein Ankunftsstrom mit unabhängig, identisch, exponentiell gemäß (3.1.2) verteilten Ankunftsabständen; wie lange dauert es von einem Ankunftsereignis bis zum übernächsten, dritt-nächsten, k -nächsten? Diese Zeiten ergeben sich offensichtlich als Summen von zwei, drei, k unabhängig exponentiell verteilten Ankunftsabständen A ; sie sind daher wiederum ZV, die wir mit A_k , $k=2,3,\dots$ bezeichnen wollen; es ist

$$(3.1.7) \quad A_k = \sum_{i=1}^k A \quad k=2,3,\dots$$

Bitte beachten Sie: Die Verteilung der ZV A_k ist **nicht** gleich der von $k \cdot A$. Einzu-beziehen ist ferner (Zeit bis zur nächsten Ankunft):

$$A_1 = A$$

Wenden wir uns der Ermittlung von A_2 zu (der Zeit bis zur übernächsten Ankunft), charakterisiert durch die Verteilungsfunktion dieser ZV

$$\begin{aligned} F_{A_2}(a) &= P[A_2 \leq a] \\ &= \int_0^a f_{A_1}(x) F_{A_1}(a-x) dx \end{aligned}$$

[Sollte Ihnen diese Formel nicht einsichtig sein: In Anlehnung an die einleuchtende Beziehung bei diskret verteilten unabhängigen ZV X, Y

$$P[X+Y \leq z] = \sum_{x_i \leq z} P[X=x_i] P[Y \leq z-x_i]$$

ist bei den hier vorliegenden kontinuierlich verteilten ZV

$f_{A_1}(x)dx$ die inkrementelle Wahrscheinlichkeit, daß der eine (der erste) Ankunftsabstand einen Wert im Intervall $(x, x+dx)$ annimmt;

$F_{A_1}(a-x)$ die Wahrscheinlichkeit, daß der andere (der zweite) Ankunftsabstand einen Wert $a-x$ annimmt (die Summe liegt damit im Wertebereich a , wie gefordert);

\int_0^a die Summation über alle Werte a , die beim ersten Ankunftsabstand auftreten]

Setzen wir (3.1.2., 3.1.3) ein:

$$\begin{aligned} FA_2(a) &= \int_0^a \exp(-x) [1 - \exp(-(a-x))] dx \\ &= \int_0^a \exp(-x) dx - \int_0^a \exp(-a) dx \\ &= 1 - \exp(-a) - \exp(-a) a \end{aligned}$$

$$(3.1.8a) \quad FA_2(a) = 1 - (1+a) \exp(-a)$$

In entsprechender Vorgehensweise erhält man für A_3 :

$$\begin{aligned} (3.1.8b) \quad FA_3(a) &= \int_0^a fA_1(x) FA_2(a-x) dx \\ &= 1 - \left(1 + a + \frac{a^2}{2}\right) \exp(-a) \end{aligned}$$

und spätestens nach Ermittlung von $FA_4(a)$ läßt sich das Bildungsgesetz

$$(3.1.8c) \quad FA_k(a) = 1 - \left[\sum_{i=0}^{k-1} \frac{(a)^i}{i!} \right] \exp(-a)$$

erraten und bestätigen.

Testfrage 3.1.9: Leiten Sie Formel (3.1.8b) im Detail ab. Bestätigen Sie, durch Schluß von k auf $k+1$, Formel (3.1.8c). *Testfrage*

Die durch (3.1.8) beschriebenen Verteilungen stellen die Familie der sog. Erlang-Verteilungen dar. Eine Erlang(k)- (auch E_k -) verteilte Zufallsvariable entsteht als Summe von k unabhängig, identisch exponentiell verteilten ZV (neben k ist zusätzlicher Parameter). Als weitere Charakteristika von Verteilungen der Erlang-Familie halten wir fest: *Erlang-Verteilung*

(3.1.10) Dichtefunktion	$fA_k(a) = \exp(-a) \frac{(a)^{k-1}}{(k-1)!}$
Erwartungswert	$E[A_k] = k/a$
zweites Moment	$E[A_k^2] = k(1+k)/a^2$
Varianz	$V[A_k] = k/a^2$
Variationskoeffizient	$VK[A_k] = 1/k$

Testfrage 3.1.10: Bestätigen Sie (3.1.10). *Testfrage*

Gleichungen (3.1.8) beschrieben die Zeit von einem Ankunftsereignis bis zum k -ten darauf folgenden (bei unabhängig identisch verteilten - kurz: u.i.v. - exponentiellen Ankunftsabständen). Wie ist nun die Zeit von einem zufälligen Beobachtungszeitpunkt bis zum k -ten darauf folgenden Ereignis verteilt? Wieder entsprechend (3.1.8)! Denn: Der zufällige Beobachtungszeitpunkt fällt ja i. allg. in ein Intervall zwischen zwei Ankünften; der Rest dieses Intervalls ist aber nach (3.1.5) identisch verteilt wie das gesamte Intervall - die Exponentialannahme beginnt, ihre "analytische Angenehmheit" zu zeigen.

Ankunftsanzahl
in festem
Zeitintervall

Versuchen wir uns nun in einer inversen Betrachtung. Nehmen wir einen bestimmten Beobachtungszeitpunkt an und fragen uns, wieviel Ankünfte in einem Zeitintervall der festen Länge t (gemessen vom Beobachtungszeitpunkt an) erfolgen. Wegen der Stochastik des Vorgangs ist diese Ankunftsanzahl nicht konstant, sondern muß durch eine ZV, nennen wir sie N_t , beschrieben werden. N_t ist offensichtlich diskret verteilt, mit einem Wertebereich N_0 ; ihre Verteilung ist z.B. charakterisiert durch die Menge der Wahrscheinlichkeiten

$$(3.1.11a) \quad \{PN_t(k); k \in N_0\}$$

Nun ist für $k \geq 1$ - siehe (3.1.8c) -:

$$P[N_t = k] = P[A_k \leq t] - P[A_{k+1} \leq t] = FA_k(t) - FA_{k+1}(t) \quad k \geq 1$$

$$P[N_t = k+1] = FA_{k+1}(t)$$

und daher:

$$\begin{aligned} PN_t(k) &= P[N_t = k] = P[N_t = k] - P[N_t = k+1] \quad k \geq 1 \\ &= P[N_t = k] - P[N_t = k+1] \\ &= FA_k(t) - FA_{k+1}(t) \\ &= 1 - \left[\sum_{i=0}^{k-1} \frac{(\lambda t)^i}{i!} \right] \exp(-\lambda t) - \left(1 - \left[\sum_{i=0}^k \frac{(\lambda t)^i}{i!} \right] \right) \exp(-\lambda t) \\ &= \frac{(\lambda t)^k}{k!} \exp(-\lambda t) \end{aligned}$$

Für $k=0$ erhält man:

$$\begin{aligned} PN_t(0) &= P[A_1 > t] = 1 - P[A_1 \leq t] \\ &= 1 - FA_1(t) = \exp(-\lambda t) \end{aligned}$$

so daß sich insgesamt ergibt:

$$(3.1.11b) \quad PN_t(k) = \frac{(\lambda t)^k}{k!} \exp(-\lambda t) \quad k=0,1,2,\dots$$

Poisson-
Verteilung

Die Verteilungen der Familie (3.1.11b) heißen Poisson-Verteilungen. In unserem Anwendungsfall charakterisiert die Poisson-Verteilung mit Parameterwert λt die Anzahl von Ankünften eines Ankunftsstroms mit u.i.v. exponentiellen Ankunftsintervallen des Parameters λ , in einem beliebigen (Gedächtnislosigkeit!) Zeitintervall der Länge t . Unabhängig identisch exponentiell verteilte Ankunftsabstände und Poisson-verteilte Ankunftsanzahlen in beliebigen Zeitintervallen sind also zwei Seiten derselben Medaille; man nennt einen derartigen Ankunftsstrom auch einen Poisson-Strom.

Poisson-Strom

Testfrage

Testfrage 3.1.12: Machen Sie sich mit dem Aussehen der Poisson-Verteilung vertraut, indem Sie Graphen der Wahrscheinlichkeiten (3.1.11b) zeichnen für den Wert $\lambda t=1$ und die Werte $\lambda t=1, 0.5, 2$ (insgesamt also drei Graphen).

Weitere Charakteristika der Poisson-Verteilung (mit unserem Parameterwert t): *Charakteristika Poisson-Verteilung*

- Erwartungswert:

$$\begin{aligned} E[N_t] &= \sum_{k=0}^{\infty} k \cdot P_{N_t}(k) \\ &= \exp(-t) \sum_{k=1}^{\infty} k \frac{t^k}{k!} \\ &= t \exp(-t) \sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} \end{aligned}$$

und mit der bekannten Reihendarstellung der Exponentialfunktion

$$e^x = \sum_{k=0}^{\infty} x^k/k!$$

wird

$$(3.1.13) \quad \begin{aligned} E[N_t] &= t \exp(-t) \exp(t) \\ &= t \end{aligned}$$

- einige spezielle Wahrscheinlichkeiten:

$$(3.1.14a) \quad \begin{aligned} P_{N_t}(0) &= \exp(-t) \\ &= 1 - t + (t)^2/2! - + \dots \\ &= 1 - t + o(t) \end{aligned}$$

wo die Definition für $o(t)$, eine "Funktion von kleiner Ordnung gegenüber t ", durch die Bedingung

$$\lim_{t \rightarrow 0} o(t)/t = 0$$

gegeben ist (so daß im obigen Fall, für kleines t , $P_{N_t}(0) \approx 1 - t$ mit hinreichender Genauigkeit gilt).

$$(3.1.14b) \quad \begin{aligned} P_{N_t}(1) &= t \cdot \exp(-t) \\ &= t + o(t) \end{aligned}$$

$$(3.1.14c) \quad \begin{aligned} P[N_t > 1] &= 1 - P_{N_t}(0) - P_{N_t}(1) \\ &= o(t) \end{aligned}$$

Zur Abrundung sei erwähnt, daß sich aus der Voraussetzung der Gültigkeit von (3.1.14) als einzige Lösung der bekannte Poisson-Strom ergibt, so daß jede der folgenden drei Bedingungen schon für sich einen Poisson-Strom definiert (sie also auch immer gleichzeitig zutreffen): *Charakteristika Poisson-Strom*

- Ereignisse treten mit u.i.v. exponentiellen Abständen, Parameter λ , auf.
- Von beliebigem zufälligen Zeitpunkt aus wird die Anzahl von Ereignissen in einem Intervall der Länge t durch die Poisson-Verteilung mit Parameter λt beschrieben.

- Von beliebigem zufälligen Zeitpunkt aus wird die Wahrscheinlichkeit keines, eines, mehr als eines Ereignisses in einem Intervall der Länge t durch (3.1.14) beschrieben.

Allem zugrunde liegt, wie mehrfach erwähnt, die Eigenschaft der Gedächtnislosigkeit.

Interpretation Parameter

Die Größe λ , der zentrale Parameter eines Poisson-Stromes, beherrscht offenbar alle Variationsmöglichkeiten innerhalb des gegebenen Rahmens. λ wird oft als "Rate" des Stroms bezeichnet, und dies gleich aus zwei unterschiedlichen Gründen:

- Wir hatten in Abschn. 2. die Bezeichnung "Rate" reserviert für Größen der Art "Anzahl von Ereignissen pro Zeiteinheit" und berechnet als "Anzahl von Ereignissen in einem Zeitintervall der Länge T , dividiert durch diese Länge T ". In ein Intervall der Länge T fallen aber bei einem Poisson-Strom im Mittel, s. (3.1.13), λT Ereignisse; $\lambda T/T = \lambda$ als (Erwartungswert der) Ereignisrate zu bezeichnen, ist demnach konsistent mit Abschn. 2.
- Gemäß (3.1.14b) wächst die Wahrscheinlichkeit, von einem beliebigen Zeitpunkt aus innerhalb der nächsten t ZE ein Ereignis zu registrieren, bei kleinem t proportional zu t , mit Proportionalitätskonstante λ . Auch insofern ist die Benennung "Rate" für λ also treffend.

Ein Wort der Abgrenzung und Warnung: Eine Ereignisrate λ , im ersten Sinne, können die verschiedensten Ereignisströme haben (ein Strom mit festen Ereignisabständen der Länge $1/\lambda$ hat diese Rate genauso wie ein Poisson-Strom mit Parameter λ). Eine Ereignisrate λ , im zweiten Sinne, hat dagegen ausschließlich ein Poisson-Strom, wobei die Besonderheit bei dem "beliebigen Zeitpunkt" liegt, und damit indirekt wieder bei der Gedächtnislosigkeit.

Spezifikation Bedienwünsche

Suchen wir unseren roten Faden wieder: Wir waren dabei, Typen von Verteilungen kennenzulernen, die im Hinblick auf ihre Brauchbarkeit für die Spezifikation stochastischer Verkehrsnetze ausgewählt wurden. Von den drei Problembereichen (s. Anfang von Abschn. 3.1) haben wir die mit (ii) und (iii) bezeichneten inzwischen diskutiert. Verbleibt die Behandlung von (i), d.h. von ZV zur Spezifikation von Bedienwünschen.

Erinnern wir uns: An einer Station i eines Verkehrsnetzes treffen Kunden nach Maßgabe des durch die Wechselwahrscheinlichkeiten - vgl. (iii) - geregelten Verkehrs ein. Sie gelangen an die Station, um von ihr "bedient" zu werden, also um von der Station eine gewisse Arbeit geleistet zu erhalten. Um nicht jeden Kunden einzeln beschreiben zu müssen, haben wir uns entschlossen, die Größe des (jedes!) Bedienwunsches durch eine (einzige!) ZV W_i zu beschreiben. Konkreter: Auf eine Station i kommt ein Strom von u.i.v. Bedienwünschen zu. Bezüglich der Dimension von W_i verfahren wir entsprechend einer der in Abschn. 2 diskutierten Möglichkeiten: Wir beschreiben die Größe eines Bedienwunsches durch die Zeit, welche die Station zu seiner Bewältigung benötigt (wenn Zweifel bestehen: welche sie benötigen würde, wenn sie einzig und allein diesen einen Bedienwunsch zu bewältigen hätte). Bedienwünsche (jetzt: Zeiten!) sind demnach, vom Problem her gesehen, eindimensional, positiv und reellwertig, so daß zu ihrer Charakterisierung kontinuierliche, auf \mathbb{R}^+ beschränkte ZV ins Auge gefaßt werden sollten.

Kritik an Exponential- Verteilung

Wir haben uns in ähnlicher Situation, bei der Charakterisierung von Ankunftsabständen, voll auf die Exponentialverteilung gemäß (3.1.2) konzentriert. Von den beiden Gründen, die für diese Wahl sprachen, sollte die analytische Angenehm-

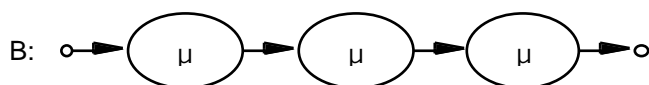
heit (Gedächtnislosigkeit!) sicher auch hier zutreffen. Der andere aber, die Realitätsadäquatheit der Exponentialverteilung, läßt sich für Bedienwunschumfänge nicht von vornherein postulieren: Warum sollten Bedienwünsche ausgerechnet exponentiell verteilt sein? Wir würden dies zumindest gerne, im Einzelfall, durch Messungen bestätigt sehen.

Aus dem Dilemma heraus, einerseits an der Exponentialverteilung festhalten zu wollen (um deren Annehmlichkeiten nicht aufgeben zu müssen), andererseits "andere" Verteilungsformen berücksichtigen zu müssen (um die Realitätstreue der Modellspezifikation nicht zu verlieren), sind die sog. "Phasenverteilungen" erdacht worden. Die Vorstellung ist die, daß eine eindimensionale, positive, reelle Größe künstlich unterstrukturiert wird in eine Folge einzelner "Phasen", daß sich also der Gesamtumfang dieser Größe als Summe der Umfänge dieser Phasen ergibt. Die Phasenumfänge werden dabei durch ZV beschrieben, so daß (wie gewünscht) der Gesamtumfang ebenfalls durch eine ZV (nämlich der Summe der Phasenumfänge) erfaßt ist; als Besonderheit werden für die Beschreibung der Phasen ausschließlich exponentielle ZV verwendet, wobei aber verschiedene Phasen verschiedene Parameter besitzen können. Unser Problem besteht aus der Charakterisierung von Bedienwünschen durch ihren zeitlichen "Umfang" - wir benutzen im folgenden die "Zeit"-Vorstellung unseres Problems.

*Abhilfe:
Phasen-
Verteilungen*

Was in der obigen allgemeinen Schilderung vielleicht etwas mysteriös wirkte, ist in Wirklichkeit ganz einfach. Stellen wir uns beispielsweise vor, wir hätten (etwa auf der Basis vorliegender Messungen von Bedienzeiten) Grund zu der Annahme, die Bedienwünsche an einer Station seien durch eine Erlang(3)-Verteilung, also gemäß (3.1.8b), zu charakterisieren. Wir hatten die Erlang(3)-verteilte ZV als Summe dreier unabhängiger exponentieller ZV konstruiert. Setzen wir unser Wissen nun in umgekehrter Richtung ein, dann ist offensichtlich der Erlang(3)-verteilte Bedienwunsch beschreibbar durch die Gesamtdauer dreier unmittelbar aufeinanderfolgender exponentieller (Teil-)Bedienwünsche: Wir haben damit den Bedienwunsch unterstrukturiert in aufeinanderfolgende Phasen, ohne an seiner Dauer etwas zu ändern. Es sei nochmals betont: Diese Strukturierung ist rein künstlich, hat mit dem Problem selbst (Arbeit gewisser Dauer) nichts zu tun. Als graphische Verdeutlichung der Charakterisierung diene folgende Skizze

*Beispiel:
E₃-Verteilung*

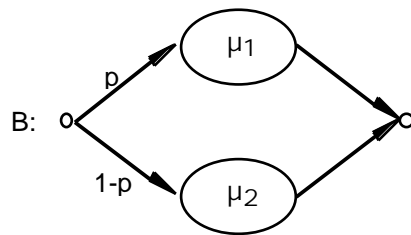


mit der oben eingeführten Interpretation, d.h.: Die Zufallsvariable B ist verteilt gemäß der Summe dreier wechselseitig unabhängiger, jeweils u.i.v. exponentieller ZV, wobei (in diesem Beispiel) alle drei Phasen denselben Parameter μ aufweisen. Oder mehr umgangssprachlich: Die Dauer von B entspricht der Dauer des Durchlaufens (in Pfeilrichtung) des obigen Graphen, wobei die Verweildauer in jedem der "Ovale" exponentiell verteilt ist mit dem betreffenden (im Oval notierten) Parameter.

Wir erweitern unsere Darstellungsmöglichkeiten noch etwas, indem wir zulassen, daß der "Weg" in unserem "Phasengraphen" sich ggf. verzweigt, und daß an einem Verzweigungspunkt jeder Zweig mit fester Wahrscheinlichkeit gewählt wird (also entsprechend einer diskreten Verteilung über den Zweigen, unabhängig von allen sonstigen Gegebenheiten). Wieder mag ein Beispiel zur Klärung

*Phasengraph
mit Verzwei-
gungen*

beitragen:



Hier ist eine ZV B charakterisiert, die mit Wahrscheinlichkeit p exponentiell mit Parameter μ_1 , mit der Alternativwahrscheinlichkeit 1-p exponentiell mit Parameter μ_2 verteilt ist. Ihre Verteilungsfunktion ist damit gegeben durch

$$FB(b) = p \cdot FB_1(b) + (1-p) \cdot FB_2(b)$$

wo B_1 bzw. B_2 die der μ_1 - bzw. μ_2 -Phase zugeordneten ZV seien. Explizit erhalten wir mit (3.1.2)

*Hyper-exponential-
verteilung*

(3.1.15a) $FB(b) = p \cdot (1 - \exp(-\mu_1 b)) + (1-p) \cdot (1 - \exp(-\mu_2 b))$

Diese Verteilung trägt den Namen "Hyperexponentialverteilung", kurz (wegen der **zwei** alternativen Phasen) die Bezeichnung H_2 -Verteilung. Leicht errechnen lassen sich weitere Charakteristika dieser Verteilung

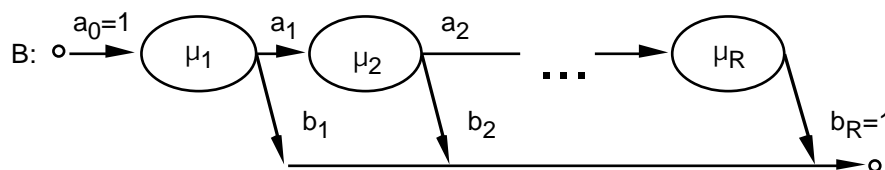
(3.1.15b) $fB(b) = p \cdot \mu_1 \cdot \exp(-\mu_1 b) + (1-p) \cdot \mu_2 \cdot \exp(-\mu_2 b)$
 $E[B] = p/\mu_1 + (1-p)/\mu_2$
 $E[B^2] = 2p/\mu_1^2 + 2(1-p)/\mu_2^2$

Testfrage

Testfrage 3.1.16: Bestätigen Sie (3.1.15). Untersuchen Sie die Verallgemeinerung der H_2 -Verteilung auf R alternative Phasen (die H_R -Verteilung). Ermitteln Sie den Variationskoeffizienten der H_R -Verteilung und vergleichen Sie ihn mit denen der Exponentialverteilung und der E_R -Verteilung.

*Cox-
Verteilung*

Eine spezielle Form von Phasenverteilungen, die wir in folgenden Kapiteln einsetzen werden, beschreibt der folgende Graph:



mit $a_i + b_i = 1 \quad i=0,1,\dots,R$

Es handelt sich um die sog. Cox-Verteilung (mit R Phasen), kurz C_R -Verteilung. Die Interpretation des Graphen: B "durchlebt" zunächst eine (erste) exponentiell verteilte Phase mit Parameter μ_1 ; dann ist B mit Wahrscheinlichkeit b_1 abgeschlossen, mit Wahrscheinlichkeit $a_1 (=1-b_1)$ schließt sich eine zweite exponentiell verteilte Phase (Parameter μ_2) an; u.s.w. Alle beteiligten ZV, d.h. die Phasendauernvariablen und die Verzweigungsvariablen, sind wechselseitig unabhängig.

*Vorzüge
Phasen-
verteilung*

Alle Phasenverteilungen besitzen den Vorzug, daß sie einerseits "fast so angenehm" sind wie die Exponentialverteilung, andererseits eine deutlich breitere

Klasse von Verteilungen überstreichen als diese. Die Angenehmheit der Exponentialverteilung lag ja in ihrer Gedächtnislosigkeit, vgl. (3.1.5), begründet. Diese ermöglichte Aussagen über die zukünftige (Rest-)Dauer eines exponentiell verteilten Intervalls unabhängig davon, zu welchem (zufälligen) Abstand vom Beginn dieses Intervalls aus nach dieser Dauer gefragt wurde. Andere Verteilungen weisen diese Eigenschaft sicher nicht auf, so daß die interessierende Rest-Dauer vom Abstand zum Intervallbeginn abhängen wird; dieser Abstand wiederum ist (wir arbeiten in kontinuierlicher Zeit) durch eine kontinuierliche Variable repräsentiert. Bei einem phasenverteilten Zeitintervall "erwischen" wir dies Intervall bei zufälligem Beobachtungszeitpunkt in irgendeiner seiner (exponentiellen) Phasen. Die Restdauer des Intervalls setzt sich zusammen aus der Restdauer dieser Phase (und diese Dauer ist wegen der Gedächtnislosigkeit vom konkreten Beobachtungszeitpunkt innerhalb der Phase unabhängig) und der Dauer der restlichen (komplett) zu durchlaufenden Phasen. Wenn wir also die Phase kennen, die "erwischt" wurde, wissen wir über die Zukunft Bescheid. Anstatt der kontinuierlichen Information (Dauer seit Intervallanfang) des allgemeinen Falls benötigen wir im phasenverteilten Fall demnach lediglich eine diskrete Information (Nummer der erwishten Phase) - wir arbeiten zwar nicht gedächtnislos, aber mit diskretem (statt kontinuierlichem) Gedächtnis; und dies wird uns enorm helfen.

Die speziellen Cox'schen Phasenverteilungen weisen darüber hinaus eine gewisse Allgemeinheit auf in folgendem Sinne: Wie Cox55 zeigt, lassen sich mit der vorgestellten (gewissermaßen kanonischen) Form alle nichtnegativen kontinuierlichen Verteilungen mit rationaler Laplace-Transformierter erfassen - für unsere Zwecke damit praktisch alle Verteilungen, die wir für Bedienwunschverteilungen in Betracht ziehen können. Wir sollten hier zumindest die Formel für den Mittelwert einer C_R -verteilten ZV B festhalten
Mit

*Anmerkungen
COX-
Verteilung*

$$e_r := P [r\text{-te Phase wird erreicht}] \\ = \sum_{i=0}^{r-1} a_i$$

erhält man ($1/\mu_r$ ist mittlere Dauer r-te Phase, $1/\mu_r^2$ deren Varianz)

$$(3.1.17) \quad E[B] = \sum_{r=1}^R e_r / \mu_r \\ V[B] = \sum_{r=1}^R e_r / \mu_r^2$$

LEERSEITE

3.2 Stochastische Prozesse (zur Auffrischung)

Hatten wir es in Abschn. 3.1 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

Familie von ZV

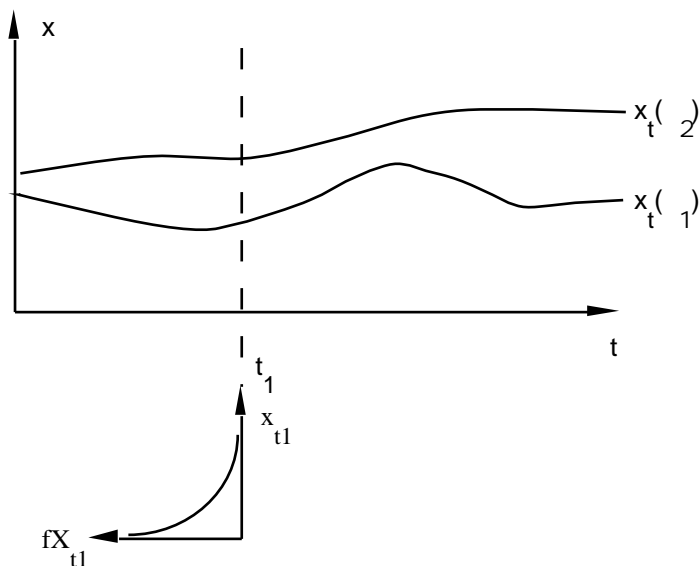
$$(X_t)_{t \in T} \quad \text{oder auch:} \quad (X(t); t \in T)$$

diese Familie von ZV, alle als Abbildungen aus derselben Menge von Elementarereignissen auf denselben Wertebereich \mathcal{X} 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 \mathcal{X} seinen "Zustandsraum".

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 ω eine Abbildung $t \mapsto x_t(\omega)$, eine Abbildung von T in \mathcal{X} , impliziert. Dies klingt wieder weit komplizierter als es tatsächlich ist. Machen wir uns die Zusammenhänge an einem kleinen Beispiel klar, in dem wir $\mathcal{X} = \mathbb{R}^+$, $T = \mathbb{R}^+$ wählen und uns mögliche Realisierungen des Prozesses (Realisierungen aller ZV der Familie) für zwei Elementarereignisse ω_1, ω_2 ansehen (Abb. 3.2.1). Zu jedem Elementarereignis (hier zu sowohl ω_1 als auch ω_2) gehört je eine Realisierung des stochastischen Prozesses (hier die Realisierungen $x_t(\omega_1)$ und $x_t(\omega_2)$). Eine Realisierung eines stochastischen Prozesses ist die gemeinsame Realisierung aller beteiligten ZV, ist eine Abbildung $T \rightarrow \mathcal{X}$ (hier: $\mathbb{R}^+ \rightarrow \mathbb{R}^+$). 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 \mathbb{R}^+) über der Zeit, auch "Trajektorie" genannt.

Veranschaulichung



Trajektorien

Abbildung 3.2.1: Trajektorien eines stochastischen Prozesses und Zustandsverteilung zu festem Zeitpunkt

Für festen Beobachtungszeitpunkt t (in Abb. 3.2.1 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_{t_1} , deren mögliche Verteilung durch eine Skizze ihrer Dichtefunktion charakterisiert ist. Für einen anderen Zeitpunkt $t_2 > t_1$ ist dementsprechend eine andere ZV, nämlich X_{t_2} maßgebend, mit einer (i. allg.) anderen Verteilung. X_{t_1} und X_{t_2} 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

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

Zustandsraum

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

$$\begin{aligned} &= \mathbf{N}_0 \text{ (zustandsdiskret)} \\ &= \mathbf{R}^+ \text{ (zustandskontinuierlich)} \end{aligned}$$

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

$$\begin{aligned} T &= \mathbf{N}_0 \text{ (zeitdiskret)} \\ T &= \mathbf{R}^+ \text{ (zeitkontinuierlich)} \end{aligned}$$

Stochastische Folgen

Zeitdiskrete stochastische Prozesse nennt man auch "stochastische Folgen".

Testfrage

Testfrage 3.2.2: Skizzieren Sie (analog Abb. 3.2.1) das prinzipielle Aussehen von stochastischen Prozessen in den vier interessanten Sonderfällen

$$(T, \Omega) \in \{(\mathbf{N}_0, \mathbf{N}_0), (\mathbf{N}_0, \mathbf{R}^+), (\mathbf{R}^+, \mathbf{N}_0), (\mathbf{R}^+, \mathbf{R}^+)\}$$

Abhängigkeiten

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

$$\{X_{t_i}; t_i \in T\}$$

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

$$\underline{X}(t) = (X(t_1), X(t_2), \dots, X(t_r)) \quad r := |T|, \text{ endlich}$$

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

$$F\underline{X}(t; \underline{x}) = P[X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_r) \leq x_r]$$

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

Einige (sehr unterschiedliche) Abhängigkeitscharakterisierungen:

Definition 3.2.3: Ein Grenzfall liegt offensichtlich vor, wenn keinerlei Abhängigkeiten bestehen. Wir haben dann eine Familie unabhängiger ZV vorliegen, so daß für abzählbares $t \in T$

$$F_{\underline{X}}(t; \underline{x}) = \prod_t F_{X_t}(x_t)$$

Unabhängige Prozesse

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

Definition 3.2.4: 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:

Stationäre Prozesse

Für alle $t \in T$, und alle $s > 0$

(wir setzen einen unserer Sonderfälle entsprechend Testfrage 3.2.2 voraus, so daß mit $t \in T$ und sinnvollem s auch $t+s \in T$ gilt)

$$\text{gilt: } F_{\underline{X}}(t; \underline{x}) = F_{\underline{X}}(t+s; \underline{x}) \quad \text{wo } \underline{s} = (s, s, \dots, s)$$

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

Definition 3.2.5: 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 $t \leq t^*$) "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:

Markoff-Prozesse

Für alle $i = \{t_j; j \in I\} \in T$
mit $t^* = \max(t_j)$

und beliebiges $t \in T$ mit $t > t^*$

$$\text{gilt: } F_{\underline{X}}(t | \underline{X}(t^*) = \underline{x}^*) = F_{\underline{X}}(t | X(t^*) = x^*)$$

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 I), 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 3.2.6: 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. (3.1.1);
- die Gesetze der Ankunftsabstände, am offenen Netz, auf seiten der Sequenz eintreffender Kunden; vgl. (3.1.2);
- die Gesetze der Bedienwunsch-Umfänge, an einer Station i , auf seiten der Sequenz der diese Station betretenden Kunden; vgl. Spezifikation Bedienwünsche in Abschn. 3.1.

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 3.2.7: 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 Ankunftsanzahl mit u.i.v., exponentiellen Zwischenankunftsabständen eintreffender Kunden beschreibt; vgl. (3.1.2, 3.1.3, 3.1.11).

*zeit-
kontinuierliche
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 $(\mathcal{S}, T) = (\mathbf{N}_0, \mathbf{R}^+)$. Die geforderte Markoff-Eigenschaft drückt sich hier formal so aus, daß

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

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

wo $t_j \in \mathbf{R}^+$; $i_j \in \mathbf{N}_0$, $j=1, 2, \dots, n+1$
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

$$(3.2.9) \quad p_i(t) := P[X(t)=i] \quad i \in \mathbf{N}_0$$

$$\sum_i p_i(t) = 1$$

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

$$(3.2.10) \quad \begin{aligned} p_{ij}(t_1, t_2) &:= P[X(t_2)=j | X(t_1)=i] && t_2 > t_1; t_1, t_2 \in \mathbf{R}^+; i, j \in \mathbf{N}_0 \\ \sum_j p_{ij}(t_1, t_2) &= 1 && i \in \mathbf{N}_0 \end{aligned}$$

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

Zustandswahrscheinlichkeiten

$$(3.2.11) \quad p_j(t_2) = \sum_{i \in \mathbf{N}_0} p_i(t_1) p_{ij}(t_1, t_2) \quad t_2 > t_1; j \in \mathbf{N}_0$$

(Diese Beziehung stellt einen Spezialfall der sog. Kolmogoroff-Gleichungen dar.)

In den allgemeinen Übergangswahrscheinlichkeiten nach (3.2.10) 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

$$(3.2.12a) \quad p_{ij}(s) := P[X(t+s)=j | X(t)=i] \quad s > 0; i, j \in \mathbf{N}_0$$

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

$$(3.2.12b) \quad \sum_j p_{ij}(s) = 1 \quad i \in \mathbf{N}_0$$

gelten. Für homogene Prozesse wird aus (3.2.11)

$$(3.2.13a) \quad p_j(t+s) = \sum_{i \in \mathbf{N}_0} p_i(t) p_{ij}(s) \quad t, s > 0; j \in \mathbf{N}_0$$

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

$$P(s) = (p_{ij}(s); i, j \in \mathbf{N}_0)$$

als Matrix der Übergangswahrscheinlichkeiten und mit

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

als Spaltenvektor der Zustandswahrscheinlichkeiten aus (3.2.13a) die kompaktere Form

$$(3.2.13b) \quad \underline{p}^T(t+s) = \underline{p}^T(t) \cdot \underline{P}(s)$$

Wir könnten uns nun die Frage stellen, ob unsere Voraussetzungen ausreichen, um für die Übergangswahrscheinlichkeiten $\underline{P}(s)$ 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

$$(3.2.14) \quad \underline{p}^T(t) = \underline{p}^T(0) \cdot \underline{P}(t)$$

angeben ließen. Dieser Weg läßt sich in der Tat erfolgreich beschreiten, führt allerdings zu expliziten Formen für $\underline{P}(t)$, 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: (3.2.14) 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 $p_{ij}(t), i \in N_0$, für große t zum selben Wert konvergieren, d.h. es sollte

$$(3.2.15a) \quad \lim_t p_{ij}(t) = p_j \quad j \in N_0$$

gelten. Die Matrix $\underline{P}(t)$ strebe, mit anderen Worten, für große t einer Matrix mit untereinander identischen Zeilen zu, so daß mit Blick auf (3.2.14) auch

$$(3.2.15b) \quad \lim_t p_j(t) = p_j \quad j \in N_0$$

gelten sollte. Unter zusätzlicher Beachtung der vorausgesetzten Homogenitätsbedingung (3.2.12) näherte sich der Prozeß demnach für große t einem "stationären Verhalten" (man konsultiere Def. 3.2.4).

Bestimmung
Grenzverteilung

Weiterhin unter der Hypothese, es existiere eine Grenzverteilung $\underline{p}=(p_j)$ gemäß (3.2.15), dann ließe sie sich wie folgt bestimmen: (3.2.13) liefert

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

mit E als Einheitsmatrix passender Dimension, woraus man weiter

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

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

$$(3.2.16a) \quad \frac{d\underline{p}^T(t)}{dt} = \underline{p}^T(t) \lim_{s \rightarrow 0} \frac{(s) - E}{s} = \underline{p}^T(t) Q$$

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

$$(3.2.16b) \quad q_{ij} = \lim_{s \rightarrow 0} \left[\frac{ij(s) - ij}{s} \right] \quad i \neq j$$

$$q_{ii} = \lim_{s \rightarrow 0} \left[\frac{(ii(s) - 1) - (ii - 1)}{s} \right]$$

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 (3.2.16) für hinreichend große t

$$(3.2.17a) \quad \underline{0}^T = \underline{p}^T Q$$

mit (Verteilung!):

$$(3.2.17b) \quad \sum_{i \in N_0} p_i = 1$$

folgen. Das lineare Gleichungssystem (3.2.17) 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 $p_j, j \in N_0$, 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 (3.2.15) 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 (3.2.15) bzw. (3.2.17) abzuleiten, möge uns für unsere praktischen Arbeiten der im folgenden ohne Beweis angegebene Satz ausreichen (Wißbegierige: s. Cinl75).

Satz 3.2.18 (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 $p_{ij}(t)>0$. Eine irreduzible (zeitkontinuierliche, homogene) Markoff-Kette besitzt eine stationäre Grenzverteilung gemäß (3.2.17) genau dann, wenn (3.2.17) eine Lösung besitzt. Diese Lösung ist ein-

*Irreduzibilität;
Existenz und
Eindeutigkeit
stationärer
Verteilung*

deutig.

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 (3.2.12) 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 (3.2.17) aufstellen und nach 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 (3.2.17) ermittelt.

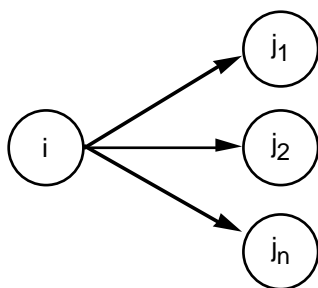
Probleme der Durchführung

Die praktische Durchführung allerdings ist nicht ohne Dornen:

- Wir müssen erst noch lernen, (3.2.17) aufzustellen, i.w. also aus einem gegebenen Problem die Matrix Q (die sämtliche Problem-Charakteristika erfaßt) abzuleiten.
- Wir müssen weiterhin (3.2.17) 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 $q_{ij}(s)$, vgl. (3.2.12), diese Entwicklung über der Zeit zu erfassen; ferner hatten wir in der zunächst rein formalen Definition der Elemente $q_{ij} - s$.(3.2.16b) - diese Entwicklung für die "unmittelbar bevorstehende" Zukunft ($s = 0$) charakterisiert. Bei dieser unmittelbar bevorstehenden Zukunft werden nur die unmittelbaren Folgezustände zum Zustand i ins Spiel kommen, also jene, welche ohne Zwischenzustand von i aus 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, \dots, j_n bezeichnet, dann konzentriert sich unser Interesse auf folgenden (Ausschnitt eines) "Zustandsübergangsgraphen":



Zustands-
übergangs-
graph

Knoten: Zustände

Kanten: mögliche unmittelbare Zustandsübergänge

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 . $i_{j_1}(s)$ 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 $i_{j_1}(s)$ und q_{i,j_1} läßt sich auf der Basis der Definitionsgleichung (3.2.16b) und einer (gedachten) Reihenentwicklung für $i_{j_1}(s)$ auch schreiben als

einfaches
Beispiel

$$(3.2.19) \quad i_{j_1}(s) = q_{i,j_1} \cdot s + o(s)$$

q_{i,j_1} 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 (3.1.14). 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 $i_j(s) = o(s)$ 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.

Zustands-
Übergangsrate

Drehen wir den Spieß um und fragen vom Problem her: Wie kommt es zu Elementen $q_{ij} > 0$ 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 (3.2.19) in unserem Beispiel zu reproduzieren, müßte also gelten:

Problem und
Q-Elemente

$$\begin{aligned}
 i_{j_1}(s) &= q_{i,j_1} \cdot s + o(s) \\
 &= P[\text{bei eingenommenem Zustand } i \text{ trifft innerhalb der} \\
 &\quad \text{nächsten } s \text{ Zeiteinheiten genau ein Auftrag ein}]
 \end{aligned}$$

einfaches
Beispiel

Dies unabhängig vom Zeitpunkt, zu dem Zustand i festgestellt wurde! Was, wenn Sie kurz unsere Diskussion um (3.1.6) 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 (3.1.14), für den (angenommener Parameter: λ) ja unabhängig vom Systemzustand (hier: i) gilt, daß

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

Vergleichen wir mit oben: Das fragliche q_{i,j_1} 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 $q_{i,j_1} > 0$ und $q_{i,j_2} > 0$ 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

$$q_{i,j_1}(s) = P[\lambda_1\text{-Phase endet innerhalb } s \text{ ZE}] \\ = \lambda_1 \cdot s + o(s)$$

$$q_{i,j_1}(s) = q_{i,j_1} \cdot s + o(s)$$

das Matrixelement

$$q_{i,j_1} = -\lambda_1$$

und aus der analogen Überlegung bzgl. der λ_2 -Phase

$$q_{i,j_2} = -\lambda_2$$

Wir erkennen aus diesen Beispielen: Die Elemente der Q -Matrix lassen sich aus dem (geeigneten!) Problem ablesen - wir werden es darin noch zu einiger Übung bringen; zumindest gilt dies für die Nichtdiagonal-Elemente von Q , die Diagonalelemente haben wir ja bisher vernachlässigt. Daher jetzt zu diesen Diagonalelementen: (3.2.12b) sagte aus, daß die bedingten Übergangswahrscheinlichkeiten sich zu 1 addieren, bzw. leicht umgeschrieben, daß

Diagonal-
elemente
Q-Matrix

$$(3.2.20) \quad 1 - q_{ii}(s) = \sum_{j \neq i} q_{ij}(s) \quad i \in N_0$$

Vollziehen wir die Grenzbetrachtung (3.2.16b), die zur Definition der Q -Elemente führte, im Rahmen dieser Gleichung, dann erhalten wir

$$(3.2.21) \quad -q_{ii} = \sum_{j \neq i} q_{ij}$$

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 (3.2.19) führte, erhalten wir aus (3.2.20):

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

$$\text{mit (3.2.21):} \quad = s \cdot (-q_{ii}) + o(s)$$

$q_{ii}(s)$ ist aber die Wahrscheinlichkeit, s ZE nach Antreffen von i wieder i vorzufinden; entsprechend $1 - q_{ii}(s)$ 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 (3.1.14) sowie (3.2.19)) die "Lebenszeit" des Zustands i nicht anders als exponentiell (mit Parameter: $-q_{ii}$) verteilt sein kann.

Testfrage 3.2.22: 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

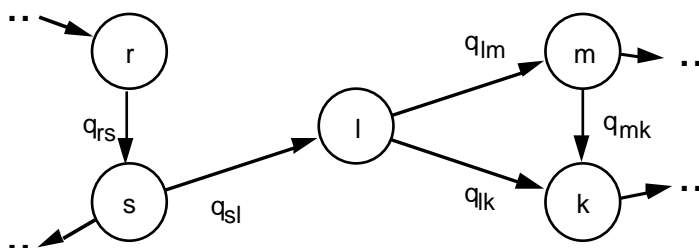
Testfrage

$$f_C(c) = P[B > c] \cdot f_A(c) + P[A > c] \cdot f_B(c)$$

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.

So gestärkt, können wir "den Rahm abschöpfen" und uns eine Technik der Aufstellung des Gleichungssystems (3.2.17) 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):

*Technik:
Aufstellung
Gleichungs-
system für
Grenz-
verteilung*



*Graph der
Zustands-
übergangsraten*

Eine einzelne Gleichung von (3.2.17) hat die Form (für festes i)

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

und mit (3.2.21)

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

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 Übergangsrate. Mit der Benennung "Wahrscheinlichkeitsfluß" (W -Fluß) für ein Produkt $p \cdot q$, also für Zustandswahrscheinlichkeit mal Austrittsrate aus diesem Zustand, können wir als Merkregel für unser Vorgehen der Aufstellung der Gleichungen von (3.2.17) festhalten:

*Ablesen der
Gleichungen*

*Wahrscheinlich-
keits-Fluß*

*Regeln***Regel 3.2.23a:**

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 3.2.23a ermöglichte eine einfache Aufstellung der (aller!) Gleichungen (3.2.17). Aus (3.2.17) lassen sich weitere Beziehungen zwischen den stationären Zustandswahrscheinlichkeiten ableiten, indem man gewisse dieser Gleichungen addiert. Sei etwa $Y \subseteq Z$ irgendeine Untermenge der Zustandsmenge, dann ergibt sich

$$\sum_{i \in Y} \left\{ \sum_{k \in Z \setminus Y} p_{ik} q_{ik} \right\} = \sum_{j \in Z \setminus Y} \left\{ \sum_{i \in Y} p_{ji} q_{ji} \right\}$$

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

Regel 3.2.23b:

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 3.2.23 eingeführte Technik ausführlich zu üben.

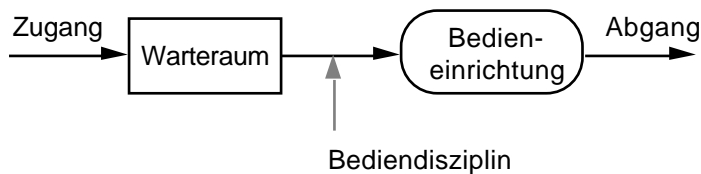
3.3 Stationsverteilungen

Wir wollen in diesem Abschnitt

- einige konkrete Stationen untersuchen (Stationen "in Isolation", als "Kleinstsysteme")
- dabei die in Abschnitt 3.2 vorgestellten Analysetechniken einsetzen, um die stationären Grenzverteilungen dieser Stationen zu ermitteln,
- damit die Basis für die im Abschnitt 3.4 folgende Vernetzung von Stationen und deren Analyse legen (Netze von Stationen als "komplexere Systeme").

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

- 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)

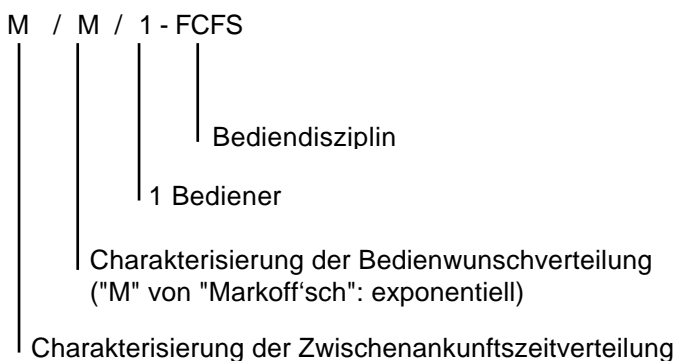


- Die Belastung einer Station ist gekennzeichnet durch einen Strom von (einzeln eintreffenden) "Kunden", deren Ankunftsabstände A unabhängig identisch verteilt sind; ebenfalls unabhängig identisch verteilt sind die Bedienwünsche S der Kunden (wir drücken die Bedienwünsche wieder direkt als Bedienzeiten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

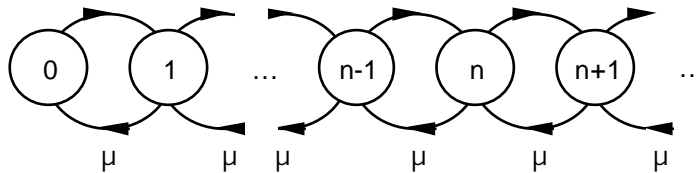
Charakterisierung $(N_t; t \in \mathbb{R})$

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 $(N_t)_{t \in \mathbb{R}^+}$ eine zeitkontinuierliche, homogene, irreduzible Markoff-Kette, so daß wir unter Anwendung von Satz 3.2.18 nach ihrer stationären Grenzverteilung fahnden können? Nun,

- Zeitkontinuirlichkeit (Betrachtung aller Zeitpunkte $t \in \mathbb{R}^+$) haben wir vorausgesetzt.
- Eine stochastische Kette (diskreter Zustandsraum) liegt vor: der Zustandsraum ist offensichtlich \mathbb{N}_0 ($n=0$, für "leeres System", und $n=1,2,\dots$ 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.(3.1.5). Es endet in der Zeit $(t,t+s)$ mit der Wahrscheinlichkeit $\lambda \cdot e^{-\lambda s}$ und führt vom Zustand n in den unmittelbar möglichen Folgezustand $n+1$. Die Übergangsrate von einem Zustand n zum Zustand $n+1$ (vgl. (3.2.19) und darauf folgende Ausführungen) beträgt damit $q_{n,n+1} = \lambda$ für alle $n=0,1,2,\dots$ Zum anderen "läuft" in allen Zuständen $n \geq 1$ 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 $\mu \cdot e^{-\mu s}$ und führt zum Zustand $n-1$. Die Übergangsrate von einem Zustand $n \geq 1$ zum Zustand n beträgt demnach $q_{n,n-1} = \mu$ für alle $n=1,2,\dots$

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 (3.2.17), das ja die (potentiell existierende) stationäre Verteilung der Kette bestimmt, läßt sich mittels Regel 3.2.23a direkt aus dem Diagramm ablesen. Wir erhalten

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

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

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

Insgesamt also:

$$(3.3.1a) \quad \begin{aligned} p_0 \cdot \mu &= p_1 \cdot \mu \\ p_n \cdot (\mu + \lambda) &= p_{n-1} \cdot \mu + p_{n+1} \cdot \mu \quad n=1,2,\dots \end{aligned}$$

dazu natürlich (3.2.17 b):

$$(3.3.1b) \quad \sum_{n=0}^{\infty} p_n = 1$$

Die explizite Lösung dieses Gleichungssystems ist leicht möglich: (3.3.1), leicht umgeschrieben, lautet

$$\begin{aligned} p_1 \cdot \mu - p_0 \cdot \mu &= 0 \\ p_n \cdot \mu - p_{n-1} \cdot \mu &= p_{n+1} \cdot \mu - p_n \cdot \mu \quad n=1,2,\dots \end{aligned}$$

und führt durch iteratives Einsetzen zunächst zu

$$p_n \cdot \mu - p_{n-1} \cdot \mu = 0 \quad n=1,2,\dots$$

woraus wir, wieder durch iteratives Einsetzen

$$p_n = \left(\frac{\lambda}{\mu} \right)^n \cdot p_0$$

erhalten, bzw. mit der abkürzenden Bezeichnung

$$(3.3.2a) \quad p_n = \rho^n \cdot p_0$$

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

Für die verbleibende Unbekannte p_0 ergibt sich mit (3.3.1b)

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

und unter der Voraussetzung (Konvergenz der Summe!)

$$(3.3.2b) \quad \rho < 1$$

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

$$(3.3.2c) \quad p_0 = 1 - \rho$$

Insgesamt lautet unsere Lösung also

$$(3.3.2d) \quad p_n = (1 - \rho) \cdot \rho^n \quad n=0,1,2,\dots$$

*Stationäre
Verteilung
M/M/1*

Die M/M/1-FCFS-Station besitzt also eine geometrische Verteilung 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":

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

$$(3.3.2e) \quad E[N] = \rho / (1-\rho)$$

*Interpretation
von*

Die Größe ρ war in (3.3.2a) 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

$$(3.3.3a) \quad \lambda / E[A] =$$

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

$$(3.3.3b) \quad E[S] = 1/\mu$$

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

$$(3.3.4a) \quad \rho = \lambda / \mu \text{ ist die im Mittel pro Zeiteinheit in die Station eingebrachte, zu leistende Arbeit}$$

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

$$(3.3.4b) \quad \rho \text{ ist die im Mittel pro Zeiteinheit von der Station geleistete Arbeit}$$

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

$$(3.3.4c) \quad \rho \text{ ist die mittlere Tätigkeitszeit des Bedieners je Zeiteinheit, d.h. die "Auslastung" der Station}$$

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 (3.3.2b) auch getroffen; für Werte $\rho = 1$ hatten wir keine stationäre Grenzverteilung angeben können (keine Konvergenz der Summe über n). Die physikalische Interpretation wird uns hier geliefert: Wird pro Zeiteinheit mehr Arbeit ins System gebracht, s. (3.3.4b), als das System leisten kann, s. (3.3.4c), 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 (3.3.4c) folgt auch

(3.3.4d) p_0 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 (3.3.4d), ist andererseits in (3.3.2c) bereits analytisch abgeleitet worden.

Der vorstehende Abschnitt ist zwar als Interpretation der für das konkret betrachtete M/M/1-FCFS-System erhaltenen Resultate (3.3.2) 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 (3.3.3), dann gelten die Aussagen (3.3.4). Es ist demnach

*Erweiterung
der
Interpretation
von*

$\rho = \lambda / \mu$ 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),

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

Bitte machen Sie sich die Mühe, die offensichtlichen Analogien zur "arbeitserhaltenden Station konstanter Bedienkapazität 1" des Abschnitts 2.2 zu ziehen. Und, wenn wir schon beim Zurückblättern sind, natürlich gilt auch weiterhin das allgemeine Resultat von Little (2.2.22), das sich (Voraussetzung: Existenz stationärer Phase) mit den hier verwendeten Notationen als

*Little-
Resultat*

$$(3.3.5) \quad E[N] = E[V] / E[A] \\ = E[V] \cdot \lambda$$

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 (2.2.22) führte und vollzieht dann den Grenzübergang für $T \rightarrow \infty$, was zu (3.3.5) führt.

Mit (3.3.2e,3.3.5) erhalten wir konkret für die M/M/1-FCFS Station die mittlere Kundenverweilzeit

Verweilzeit **(3.3.6)** $E[V] = 1/[\mu \cdot (1 - \rho)] = 1/(\mu - \lambda)$

Lassen Sie uns zur Übung die Aussage (3.3.6) explizit, ohne Bezugnahme auf Little, herleiten:

*Verweilzeit
explizit*

Ein neu ankommender Kunde findet die Station in einem ihrer möglichen Zustände $n \in \mathbf{N}_0$ vor, wo n die Anzahl unmittelbar vor der Ankunft anwesender Kunden bezeichnet; einer der Kunden befindet sich in Bedienung, die anderen warten. Die Verweilzeit des Ankömmlings setzt sich aufgrund der angenommenen FCFS-Disziplin zusammen als Summe

- (i) der Restbedienzeit des Bedienten (falls $n > 0$),
- (ii) der Bedienzeiten aller Wartenden (falls $n > 1$),
- (iii) der Bedienzeit des Ankömmlings selbst (immer).

Erinnern wir uns, daß wir die Bedienzeiten als exponentiell verteilt angenommen hatten, also mit Verteilungsdichte

$$f_S(s) = \mu \cdot \exp(-\mu \cdot s)$$

Die Restbedienzeit des Bedienten (i) ist (Ankunft zu zufälligem Zeitpunkt) exponentiell verteilt mit Parameter μ - vgl. (3.1.5). Damit ergibt sich die Verweilzeit V_n (ZV!) eines im Zustand n Eintreffenden als Summe von

$$\begin{aligned} & \text{(i)} + \text{(ii)} + \text{(iii)} \\ & 1 + (n-1) + 1 = n+1 \end{aligned}$$

unabhängig identisch exponentiell verteilten ZV gleichen Parameters, also mit Verteilungsdichte - vgl. (3.1.10)

(3.3.7a) $fV_n(v) = \mu \cdot \exp(-\mu v) \cdot (\mu v)^n / n!$

In welchem Zustand nun trifft der Ankömmling ein? Er findet, bei Ankunft in der stationären Phase, zu zufälligem Zeitpunkt, einen Zustand $n \in \mathbf{N}_0$ entsprechend den Wahrscheinlichkeiten p_n - vgl. (3.3.2d). Die Verweilzeit V eines beliebigen Kunden gehorcht demnach der Verteilungsfunktion

(3.3.7b)
$$\begin{aligned} fV(v) &= \sum_{n \in \mathbf{N}_0} p_n fV_n(v) \\ &= (1 - \rho) \mu \exp(-\mu v) \sum_{n \in \mathbf{N}_0} \left[(\mu v)^n / n! \right] \\ &= (1 - \rho) \mu \exp(-\mu v) \exp(\mu v) \\ &= (\mu - \lambda) \exp[-(\mu - \lambda)v] \end{aligned}$$

Die Kundenverweilzeiten im M/M/1-FCFS-System sind also exponentiell verteilt mit Parameter $(\mu - \lambda)$. Der Erwartungswert der Verweilzeiten beträgt demnach - vgl.(3.1.3)

(3.3.7c) $E[V] = 1/(\mu - \lambda)$

in Bestätigung von (3.3.6).

Die große Hilfe, die Little's Resultat zur Ermittlung der mittleren Kundenverweilzeit bietet, ist darin begründet, daß wir für (im Vergleich zur betrachteten M/M/1-FCFS-Station) kompliziertere Stationen oder gar für komplexe Systeme die mittlere Kundenzahl $E[N]$ wesentlich leichter ermitteln können als $E[V]$ über die Verteilung von V (also etwa analog der Ableitung von (3.3.7)).

Wert des Little-Resultats

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:

Satz 3.3.8: Die stationäre Verteilung der Kundenpopulation N der M/M/1-Station ist unter den Bediendisziplinen

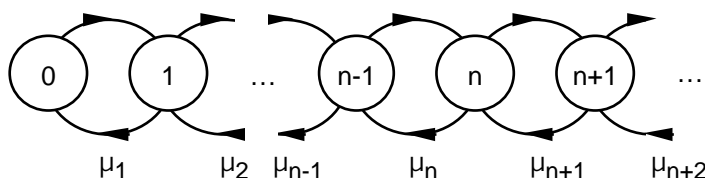
M/M/1-Resultate

- (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 eingereicht 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 (3.3.2) beschrieben. Ebenfalls identisch ist bei all diesen Disziplinen die mittlere Verweilzeit gemäß (3.3.6). (Zur Warnung: Absolut **nicht** identisch sind aber die Verteilungen der Verweilzeiten V .)

Testfrage 3.3.9: Bestätigen Sie Satz 3.3.8 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 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 (3.3.1, 3.3.2))

das System der Gleichgewichtsgleichungen

$$\begin{aligned} p_0 \cdot \mu_0 &= p_1 \cdot \mu_1 \\ p_n \cdot (\mu_n + \mu_{n+1}) &= p_{n-1} \cdot \mu_{n-1} + p_{n+1} \cdot \mu_{n+1} \quad n=1,2,\dots \end{aligned}$$

dessen vereinfachte Form

$$p_n \cdot \mu_n - p_{n-1} \cdot \mu_{n-1} = 0 \quad n=1,2,\dots$$

und seine Lösung

Stationäre
Verteilung

$$(3.3.10a) \quad p_n = p_0 \prod_{i=1}^n \mu_i$$

woraus mit der normierenden Bedingung

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

sich weiterhin

$$(3.3.10b) \quad p_0 = \left(1 + \sum_{n=1}^{\infty} \prod_{i=1}^n \mu_i \right)^{-1}$$

ergibt (Bedingung für Existenz stationärer Lösung: Summe konvergiert zu endlichem Wert). Offensichtlich ist (3.3.2) als Sonderfall in (3.3.10) enthalten (nachprüfen!).

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

Interpretation

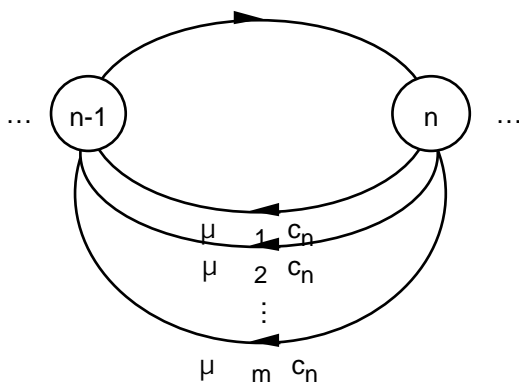
zustands-
abhängige
Bedienung

Nun, eine (von vielen möglichen) Interpretationen 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 c_i , $i=1,2,\dots$, so daß bei Anwesenheit von i Kunden c_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 aufgebrauchte Arbeitskapazität c_n komme den m bearbeiteten, beliebig nummerierten Kunden in jeweiligen Bruchteilen $\alpha_1, \alpha_2, \dots, \alpha_m$ zugute; auf den i -ten Kunden entfällt also eine Arbeit von $\alpha_i \cdot c_n$ AE/ZE. Der Vollständigkeit halber sei angemerkt:

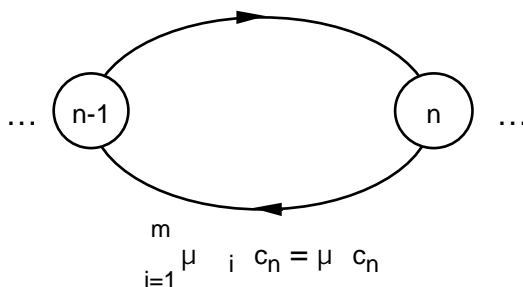
$$\prod_{i=1}^m \mu_i = \mu$$

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

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 $\mu \cdot i \cdot c_n$). Das Ende der Bedienung des so Bedienten geschieht also innerhalb der nächsten s ZE mit Wahrscheinlichkeit $(\mu \cdot i \cdot c_n) \cdot s + o(s)$. 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 (3.3.10) lässt sich also auf alle M/M/. Systeme mit zustandsabhängigen (Gesamt-)Arbeitskapazitäten c_n anwenden, wenn wir nur $\mu_n = c_n \cdot \mu$ ($n=1,2,\dots$) setzen.

Ist diese Interpretation nun praktisch relevant? Und ob!:

(3.3.11a) 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

M/M/m-FCFS

Bediener erhält. Die Gesamtarbeitskapazitäten dieser Station betragen also

$$\begin{aligned} c_n &= n & n=1,2,\dots,m \\ c_n &= m & n > m \end{aligned}$$

Die stationäre Verteilung der Station ist durch (3.3.10), nach Substitution $\mu_n = c_n \cdot \mu$, gegeben.

M/M/m-

(3.3.11b) Die Resultate (3.3.11a) gelten auch für *M/M/m-RANDOM*, *M/M/m-LCFS*, *M/M/m-LCFS-PR*-Stationen.

Testfrage 3.3.11c: Wie lautet die explizite Form der stationären Zustandsverteilung der *M/M/m*-Systeme?

M/M/

(3.3.12) 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

$$c_n = n \quad n=1,2,\dots$$

Die stationäre Verteilung erhalten wir aus (3.3.10) zu

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

mit

$$\begin{aligned} p_0 &= \left(1 + \sum_{n=1}^{\infty} \left[(\rho / \mu)^n / n! \right] \right)^{-1} \\ &= \exp(-\rho / \mu) \end{aligned}$$

IS

Diese Station trägt auch die Bezeichnungen "Infinite Server" (kurz: "IS") oder "pure delay" (vgl. nochmals Abb. 2.2.10).

M/M/1-PS

(3.3.13) 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 2.2.9) 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 $\mu_n = 1/n$ der Gesamtkapazität bedient. Die Gesamtarbeitskapazität der Station beträgt aber konstant 1, so daß wir aus (3.3.10) dieselben Ergebnisse erhalten, wie wir sie für FCFS-Bedienung erhielten, nämlich (3.3.2):

$$\begin{aligned} p_n &= \rho^n \cdot p_0 \\ p_0 &= 1 - \rho \end{aligned}$$

PS

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

Fassen wir die gewonnenen Ergebnisse kurz zusammen.

*Zusammenfassung
Ergebnisse*

Zusammenfassung 3.3.14: Sei die Belastung einer Station gemäß $M/M/1$ 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 $c_i, i=1,2,\dots$. Bediene die Station gemäß Disziplin FCFS oder LCFS oder LCFS-PR oder RANDOM. In all diesen Fällen ist die stationäre Zustandsverteilung gegeben durch (3.3.10), also durch

$$p_n = p_0 \left(\frac{\lambda}{\mu} \right)^n \prod_{i=1}^n c_i \quad n=1,2,\dots$$

$$p_0 = 1$$

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

Wie Sie sich leicht überzeugen können, gilt für die gefundene stationäre Verteilung, daß

*Weitere
Eigenschaften*

(3.3.15) $p_n \cdot \mu_n = p_{n+1} \cdot \mu_{n+1} \quad n=0,1,\dots$

Dieses Gleichungssystem hatten wir als "vereinfachte Form" des Systems gemäß (3.2.17) auf dem Wege zur Lösung (3.3.10) bereits einmal zu Papier gebracht, dort in der Notation

$$p_n \cdot \mu_n - p_{n-1} \cdot \mu_{n-1} = 0 \quad n=1,2,\dots$$

Das nach Regel 3.2.23a aufgestellte (kompliziertere) Gleichungssystem gemäß (3.2.17) trägt den Namen "System der globalen Gleichgewichtsgleichungen", das offensichtlich äquivalente (einfachere) System (3.3.15) wird als "System partieller Gleichgewichtsgleichungen" bezeichnet. Die Namensgebung "partiell" rührt daher, daß jede der Gleichungen (3.2.17) in unabhängigen Teilen erfüllt wird; daß, im vorliegenden Falle (vgl. erneut die Ableitung von (3.3.10a)), in den Gleichungen

*globales/
partielles
Gleichgewicht*

$$p_n \cdot \mu_n - p_{n-1} \cdot \mu_{n-1} = p_{n+1} \cdot \mu_{n+1} - p_n \cdot \mu_n$$

sowohl rechte als auch linke Seite je für sich verschwinden. Daß dies (im vorliegenden Falle) so sein muß, erkennen wir leicht, wenn wir unsere Regel 3.2.23b auf die Zustandsuntermengen $Y_n = \{0,1,\dots,n\}; n=0,1,2,\dots$ unseres Übergangsraten-diagramms (vor (3.3.10)) anwenden: System (3.3.15) steht unmittelbar da (wo $\mu_n = \mu \cdot c_n$).

Es ist verlockend, Regeln zur direkten Aufstellung des (einfacheren!) partiellen Systems (3.3.15) zu formulieren. Hier "funktionieren" gleich zwei davon:

Regel 3.3.16a: "lokales Gleichgewicht"

*lokales
Gleichgewicht*

Für jeden Zustand ist der W-Fluß in diesen Zustand aufgrund der "Geburt" einer exponentiellen Phase (eines bestimmten Typs) gleich dem W-Fluß aus diesem Zustand aufgrund des "Todes" einer Phase (dieses Typs).

Erklärung bzgl. (3.3.15): Betrachten Sie Zustand "n+1"; in diesen Zustand gelangt man durch Geburt einer (bevorstehenden) Bedienphase anlässlich einer Ankunft, Fluß: $p_n \cdot \mu \cdot c_{n+1}$; aus diesem Zustand gelangt man durch Tod einer Bedienphase, Fluß: $p_{n+1} \cdot \mu \cdot c_{n+1}$.

*Stations-
gleichgewicht*

Regel 3.3.16b: "Stationsgleichgewicht"

Für jeden Zustand ist der W-Fluß aus diesem Zustand aufgrund einer Stationsankunft gleich dem W-Fluß in diesen Zustand aufgrund eines Stationsabgangs.

Erklärung bzgl. (3.3.15): Betrachten Sie Zustand "n"; der W-Fluß aus n aufgrund einer Ankunft ist $p_n \cdot \mu \cdot c_{n+1}$; der W-Fluß nach n aufgrund eines Abgangs ist $p_{n+1} \cdot \mu \cdot c_{n+1}$.

Obwohl diese Regeln hier so gut "passen": Niemand gibt uns die Gewähr, daß sie allgemeingültig sind! (Und sie sind es auch nicht!) Immerhin: Man kann sie bei komplizierten Fällen einmal "ausprobieren" - und dies werden wir im weiteren Verlauf auch tun, teils mit, teils ohne Erfolg.

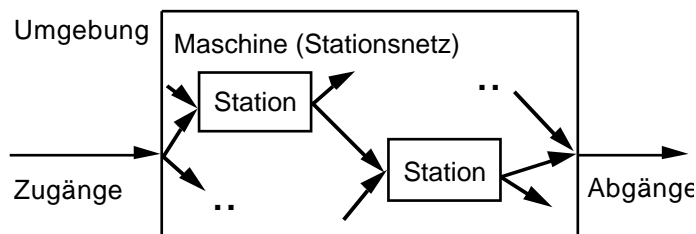
3.4 Exponentielle Netze

Wir wenden uns der Definition und Analyse einer ersten Klasse von Netzen zu, die wir exponentielle Netze nennen wollen, da wir an allen Stationen dieser Netze nur exponentiell verteilte Bedienwünsche der dort zu bearbeitenden Kunden zulassen werden.

In teilweiser Wiederholung von vorangegangenen Abschnitten definieren wir exponentielle Netze wie folgt:

Spezifikation exponentielle Netze

(3.4.1a) Graphische Repräsentation der Struktur:



(3.4.1b) Spezifikation der Maschine:

Maschine

Die Maschine bestehe aus M Stationen, die wir zur Notationsvereinfachung willkürlich von 1 bis M durchnummerieren; wir haben es mit einer Stationsmenge $I = \{1, 2, \dots, M\}$ zu tun. Die Stationen weisen eine unbegrenzte (räumliche) Kapazität auf. Jede Einzelstation i ist durch einen "Geschwindigkeitsvektor" $\underline{c}_i = (c_i(1), c_i(2), \dots)$ charakterisiert, wo jede einzelne Komponente $c_i(n)$ dieses Vektors die (zeitliche) Gesamtarbeitskapazität der Station bei Anwesenheit von n Kunden festlegt. Jede Station bedient nach einer der Disziplinen FCFS, LCFS, RANDOM oder LCFS-PR (vgl. 3.3.14). Kundenzugänge zur und Kundenabgänge von der Maschine (aus der und zur Umgebung) sind an jeder Station möglich. Ebenso ist, innerhalb der Maschine, ein Wechsel von Kunden von jeder Station zu jeder Station zugelassen.

(3.4.1c) Spezifikation der Last:

Last

Das System wird als offenes oder geschlossenes System betrieben. Im Falle eines offenen Systems erfolgen Kundenankünfte aus der Umgebung entsprechend eines Poisson-Stroms mit Parameter λ ; bei geschlossenem System gibt es keine derartigen Kundenankünfte, was wir gelegentlich auch durch die Setzung $\lambda = 0$ erfassen wollen. Bei geschlossenem System sind (permanent) N Kunden anwesend. Die Kunden bewegen sich im Netz entsprechend der Einträge einer "Wechselmatrix" (routing matrix) $H = (h(i,j))$; jedes Element $h(i,j)$ dieser Matrix sagt aus, mit welcher Wahrscheinlichkeit ein Kunde, dessen Bedienung an Station i abgeschlossen ist, einen unmittelbar folgenden Bedienwunsch an Station j stellt, d.h. nach Station j wechselt. Zeilen- und Spaltenindizes von H laufen über $I^* = I \cup \{0\}$, wo "0" die Umgebung kennzeichnet (so daß Elemente $h(0,i)$ die Aufteilung von Kunden bei Maschineneintritt, Elemente $h(i,0)$ den Austritt aus der Maschine regeln). H gehorcht insgesamt einer "Zusammenhangsbedingung" derart, daß jede Station von jeder Station aus mit endlicher Wahrscheinlichkeit erreicht wird, direkt oder indirekt auf dem Weg über andere Zwischenstationen unter evtl. Einfluß der Umgebungsstation 0. (Als Matrix der Übergangswahrscheinlichkeiten einer zeitdiskreten Markoff-Kette interpretiert, ist diese Kette irreduzibel.)

Sinnvollerweise gilt:

$$\begin{aligned} h(i,j) &= 1 && i \in I^* \\ &= 0 && h(i,0) = 0 \quad i \in I \\ h(0,0) &= 0 \end{aligned}$$

Jeder an einer Station i ankommende Kunde stellt einen exponentiell mit Parameter μ_i verteilten Bedienwunsch an diese Station. Es herrscht völlige Unabhängigkeit aller involvierten stochastischen Größen.

*Analyse-
schritte*

Die Analyse des so spezifizierten Systems "exponentielles Netz" nehmen wir in folgenden Schritten vor:

- (i) Definition von Zustandsbeschreibung und Zustandsraum einer zugeordneten zeitkontinuierlichen Markoff-Kette;
- (ii) Ermittlung der Ereignisse, die zu Zustandsübergängen führen und Ableitung der zugehörigen Übergangsraten;
- (iii) Skizzierung des Diagramms der Übergangsraten;
- (iv) Aufstellung und Lösung des Systems von Gleichgewichtsgleichungen zur Ermittlung der stationären Verteilung.

Zustand

(i) Zustandsbeschreibung / Zustandsraum

Wir wählen als Zustandsbeschreibung Vektoren

$$(3.4.2a) \quad z = (n_i; i \in I) = (n_1, n_2, \dots, n_M)$$

wo n_i die Zahl an Station i weilender Kunden bezeichnet; diese Beschreibung ist hinreichend, um dem System eine zeitkontinuierliche, irreduzible Markoff-Kette zuordnen zu können (nachdenken!). Der (diskrete) Zustandsraum ergibt sich zu

$$(3.4.2b) \quad Z = \prod_{i \in I} N_0 \quad \text{bei offenem System}$$

$$(3.4.2c) \quad Z = \{z: \prod_{i \in I} n_i = N\} \quad \prod_{i \in I} N_0 \quad \text{bei geschlossenem System}$$

(NB: Ein Zustand ist also nicht mehr, wie in Abschn. 3.3, durch einen Skalar gekennzeichnet!) Um notationelle Schwierigkeiten in den folgenden Betrachtungen zu vermeiden, vereinbaren wir generell, daß die stationären Wahrscheinlichkeiten des Einnehmens unzulässiger (non-feasible) Zustände verschwinden, d.h. wir setzen

$$(3.4.2d) \quad P[z] = 0 \quad z \notin Z$$

*Zustands-
übergänge*

(ii) Ereignisse, Zustandsübergänge, Übergangsraten.

Offensichtlich führen ausschließlich die Ereignisse

- Ankunft eines Kunden aus der Umgebung
 - Beendigung einer Kundenbedienung an einer Station
- zu Zustandsänderungen im Rahmen unserer Zustandsbeschreibung, wobei wir letzteren Ereignistyp in die beiden Unterfälle
- Bedienungsende mit nachfolgendem Verlassen des Netzes,
 - Bedienungsende mit nachfolgendem Wechsel zu einer anderen Station aufspalten wollen.

(a) Netzankünfte

Bei Ankunft eines Kunden im Systemzustand z erfolgt ein Übergang in einen Zustand z' , der einer Zustandsmenge $A(z)$ angehört, in deren Zuständen an einer der Stationen ein Kunde mehr vorhanden ist, als im Zustand z vorhanden waren. Wir benutzen hierzu folgende Notationen

$$z \in Z \quad z' \in A(z)$$

wo $A(z) = \{z(i+1); i \in I\}$

(Also: $z(i+1)$ ist der Zustand, der im Vergleich zum Zustand z an Station i einen Kunden mehr aufweist, $A(z)$ ist die Menge dieser Zustände, wo i über die Stationen variiert.)

Eine Ankunft ist in jedem Systemzustand möglich. Die Übergangsrate für den Übergang von z nach einem $z'=z(i+1)$ wollen wir als $q(z,z')$ notieren (im Prinzip als Element einer Q-"Matrix", wobei wir irgendeine beliebige lineare Anordnung aller Zustände $z \in Z$ voraussetzen, über die wir uns aber nicht konkret äußern müssen). Diese Rate ergibt sich als Produkt der Ankunftsrate und der Wahrscheinlichkeit $h(0,i)$, daß die Ankunft an Station i erfolgt, daß also der erste Bedienwunsch des Netzankömmlings an Station i gerichtet ist. Wir erhalten

$$\text{für } z \in Z \quad z'=z(i+1)$$

(3.4.3a) $q(z,z') = \lambda \cdot h(0,i)$

(b) Netzabgänge

Nun etwas kompakter: Die Übergänge sind von der Art

$$z \in Z \quad z' \in D(z)$$

wo $D(z) = \{z(i-1); i \in I, n_i > 0\}$

Die Übergangsrate ist

$$\text{für } z \in Z \quad z'=z(i-1)$$

(3.4.3b) $q(z,z') = \mu_i \cdot c_i(n_i) \cdot h(i,0)$

da mit der Rate $\mu_i \cdot c_i(n_i)$ eine Bedienung an Station i endet, die mit der Wahrscheinlichkeit $h(i,0)$ zu einem Abgang führt. Um für die weiteren Überlegungen nicht immer die Bedingung "Übergang aus einer Station nur falls dort ein Kunde vorhanden ist" mitschleppen zu müssen, verlängern wir alle Geschwindigkeitsvektoren c_i um eine Komponente $c_i(0)=0$, so daß (3.4.3b) generell verwendet werden kann - in gewisse unzulässige Folgezustände wird mit Rate "0", de facto also nie, übergegangen.

(c) Stationswechsel

Noch etwas kompakter: Übergänge der Art

$$z \in Z \quad z' \in C(z)$$

wo $C(z) = \{z(i-1,j+1); i,j \in I, n_i > 0\}$

(Wechsel von i nach j : Bei i ein Kunde weniger, bei j einer mehr)

Übergangsrate

$$\text{für } z \in Z \quad z'=z(i-1,j+1)$$

(3.4.3c) $q(z,z') = \mu_i \cdot c_i(n_i) \cdot h(i,j)$

(iii) Diagramm der Übergangsraten

Im Prinzip könnten wir natürlich versuchen, für konkretes M (= Zahl von Stationen) und, bei geschlossenem System, konkretes N (=Zahl von Kun-

Übergangs-
raten

den) das volle Übergangsdiagramm zu zeichnen; dies ist allerdings ein umfangreiches Unterfangen (versuchen Sie es ruhig einmal) und auch eine unnötige Arbeit. Wir sehen uns statt dessen einen einzelnen Zustand z (als Prototyp aller Zustände $z \in Z$) und seine "Diagrammumgebung", d.h. seine direkten Vorgänger- und Nachfolgerzustände an:

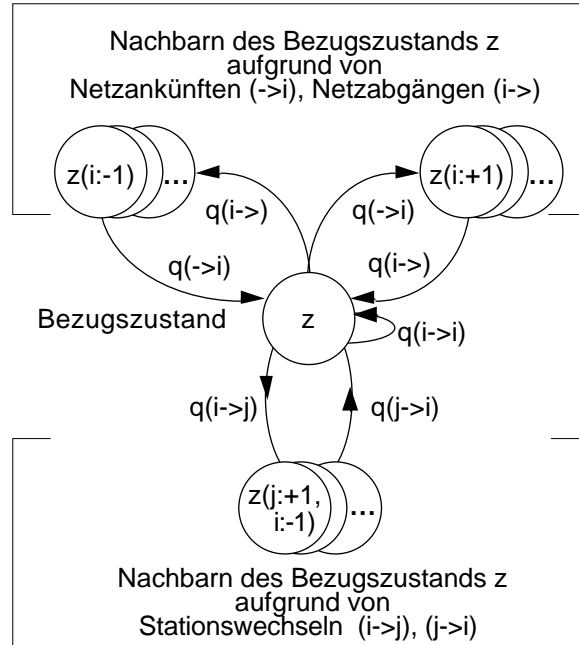


Abb. 3.4.4: Ausschnitt Übergangsdiagramm exponentieller Netze

Gleichungssystem

(iv) **Aufstellung und Lösung des Gleichgewichtssystems**

Entsprechend Regel 3.2.23a erhalten wir das Gleichungssystem zur Bestimmung der stationären Verteilung als

$$W\text{-Fluß nach } z = W\text{-Fluß aus } z \quad z \in Z$$

wobei unsere Vorsichtsmaßnahmen ($\mu_i = 0$ für geschlossenes System, $P[z]=0$ für unzulässige Zustände, $\mu_i(0)=0$) dafür Sorge tragen, daß der in vorangegangener Skizze verwendete Zustand z wirklich prototypisch für alle Zustände ist. Das Gleichungssystem wird natürlich einigermaßen komplex werden (kein Vergleich etwa mit 3.3.1!); schlimmer noch: Eine explizite Lösung für derartige Gleichungssysteme läßt sich sicher nur in Sonderfällen finden. Andererseits: Wir hätten uns nicht soviel Mühe mit der Definition exponentieller Netze gemacht, wenn sie nicht einen derartigen Sonderfall darstellten. Die Lösung wird im folgenden Satz angegeben. Der Satz vereint die Ergebnisse von Jackson (Jack63), und Gordon-Newell (GoNe67); die offenen exponentiellen Netze tragen auch den Namen "Jackson-Netze", die geschlossenen den Namen "Gordon-Newell-Netze".

Satz 3.4.5: Gegeben sei ein exponentielles Netz gemäß Spezifikation (3.4.1), dessen Zustände gemäß (3.4.2) charakterisiert seien. Sei ferner $\underline{\mu} = (\mu_i; i=1,2,\dots,M)$ ein Vektor, der das lineare Gleichungssystem

$$\mu_i = \sum_{j \in I} \mu_j h(j,i) + h(0,i) \quad i \in I$$

löst, und stehe jeweils

$$P_i := (P_i(n); n \in \mathbb{N}_0) \quad i \in I$$

(wo $P_i(n) = P[n_i=n]$ gesetzt ist) für die stationäre Verteilung der isolierten Station i , wenn sie mit einem Poisson-Strom (Parameter λ_i) und exponentiellen Bedienungswünschen (Parameter μ_i) betrieben wird; solche Verteilungen hatten wir in Abschn. 3.3 abgeleitet. Unter diesen Voraussetzungen hat die stationäre Verteilung des Netzzustands (falls existent) die Form

$$P[z] = \prod_{i \in I} [P_i(n_i)] / G \quad z \in Z$$

(wo $z=(n_i; i \in I)$ gesetzt ist). G ist eine "Normalisierungskonstante", die

$$\sum_{z \in Z} P[z] = 1$$

sicherstellt. Die Lösung hat eine sog. "Produktform", eine Lösung also, die man auch erhielte, wenn alle Stationen unabhängig voneinander arbeiteten, was aber im vorliegenden Falle sicher nicht zutrifft.

Zum Beweis stellen wir nun doch das Gleichungssystem nach (iv) auf und zeigen, daß es durch die Aussagen des Satzes gelöst wird. Zunächst strukturieren wir das Gleichungssystem etwas; wir haben es mit folgenden Termen zu tun (konsultieren Sie Abb. 3.4.4): Bezüglich festen Zustands $z \in Z$ sei *Beweis*

(3.4.6a):

- A_{1i} = W-Fluß nach z aufgrund eines Netzzugangs zu Station i
- A_{2i} = W-Fluß nach z aufgrund eines Netzabgangs von Station i
- A_{3i} = W-Fluß nach z aufgrund eines Kundenwechsels zu Station i
- B_{1i} = W-Fluß aus z aufgrund eines Netzabgangs von Station i
- B_{2i} = W-Fluß aus z aufgrund eines Netzzugangs zu Station i
- B_{3i} = W-Fluß aus z aufgrund eines Kundenwechsels von Station i aus

Mit diesen Abkürzungen hat das zu lösende Gleichungssystem die Gestalt

$$(3.4.6b) \quad \sum_{i \in I} (A_{1i} + A_{2i} + A_{3i}) = \sum_{i \in I} (B_{1i} + B_{2i} + B_{3i}) \quad z \in Z$$

Konkret erhält man für die einzelnen Terme (bitte im Detail nachvollziehen!)

$$(3.4.6c) \quad \begin{aligned} A_{1i} &= P[z(i:-1)] \quad h(0,i) \\ A_{2i} &= P[z(i:+1)] \quad \mu_i \cdot c_i(n_i+1) \cdot h(i,0) \\ A_{3i} &= \sum_{j \in I \setminus i} \{P[z(j:+1; i:-1)] \quad \mu_j \cdot c_j(n_j+1) \cdot h(j,i)\} \\ B_{1i} &= P[z] \quad \mu_i \cdot c_i(n_i) \cdot h(i,0) \\ B_{2i} &= P[z] \quad h(0,i) \\ B_{3i} &= \sum_{j \in I \setminus i} \left\{ P[z] \quad \mu_i \cdot c_i(n_i) \cdot h(i,j) \right\} \end{aligned}$$

(NB: Der Wechsel $i \leftrightarrow i$ wäre sowohl in A_{3i} als auch in B_{3i} zu berücksichtigen und fällt daher heraus.)

Wir dividieren (3.4.6b) durch $P[z]$ (per def 0) und erhalten Gleichungssystem

$$(3.4.7a) \quad \sum_{i \in I} (A'_{1i} + A'_{2i} + A'_{3i}) = \sum_{i \in I} (B'_{1i} + B'_{2i} + B'_{3i}) \quad z \in Z$$

mit

$$A'_{1i} = \lambda_i h(0,i) \cdot P[z(i:-1)] / P[z]$$

woraus wegen der Behauptung des Satzes 3.4.5

$$\begin{aligned} \frac{P[z(i:-1)]}{P[z]} &= \frac{\left\{ \prod_{j \in I \setminus i} P_j(n_j) \right\} P_i(n_i-1) / G}{\left\{ \prod_{j \in I \setminus i} P_j(n_j) \right\} P_i(n_i) / G} \\ &= P_i(n_i-1) / P_i(n_i) \end{aligned}$$

und damit

$$A'_{1i} = \lambda_i h(0,i) \cdot P_i(n_i-1) / P_i(n_i)$$

wird; entsprechend auch

$$\begin{aligned} A'_{2i} &= \mu_i \left\{ c_i(n_i+1) h(i,0) \frac{P_i(n_i+1)}{P_i(n_i)} \right\} \\ A'_{3i} &= \sum_{j \in I \setminus i} \left\{ \mu_j c_j(n_j+1) h(j,i) \frac{P_j(n_j+1) P_i(n_i-1)}{P_j(n_j) P_i(n_i)} \right\} \\ B'_{1i} &= \mu_i \left\{ c_i(n_i) h(i,0) \right\} \\ B'_{2i} &= \lambda_i h(0,i) \\ B'_{3i} &= \sum_{j \in I \setminus i} \left\{ \mu_j c_j(n_j) h(j,i) \right\} \end{aligned}$$

Nun haben die stationären Verteilungen isolierter exponentieller Stationen (und nur solche sind im Netz vorhanden) gemäß (3.3.15) alle die Eigenschaften, daß

$$P_i(n_i) \cdot \lambda_i = P_i(n_i+1) \cdot \mu_i \cdot c_i(n_i+1)$$

so daß wir erhalten:

$$\begin{aligned} (3.4.7b) \quad A'_{1i} &= \left\{ \lambda_i h(0,i) \mu_i c_i(n_i) \right\} / \left\{ \mu_i c_i(n_i+1) \right\} \\ A'_{2i} &= \left\{ \mu_i c_i(n_i+1) h(i,0) \right\} / \left\{ \mu_i c_i(n_i+1) \right\} \\ &= h(i,0) \\ A'_{3i} &= \sum_{j \in I \setminus i} \left\{ \mu_j c_j(n_j+1) h(j,i) \right\} / \left\{ \mu_j c_j(n_j+1) \right\} \\ &= \sum_{j \in I \setminus i} \left\{ h(j,i) \right\} \\ B'_{1i}, B'_{2i}, B'_{3i} & \quad \text{wie oben} \end{aligned}$$

Addition von A'_{1i} und A'_{3i} ergibt wegen der Voraussetzungen des Satzes 3.4.5 bezüglich der i

$$A'_{1i} + A'_{3i} = \mu_i \cdot c_i(n_i) \cdot (1 - h(i,i))$$

Andererseits ist wegen der Voraussetzungen bezüglich der $h(i,j)$ auch

$$B'_{1i} + B'_{3i} = \mu_i \cdot c_i(n_i) \cdot (1 - h(i,i))$$

so daß, für alle Gleichungen des Systems (3.4.7a), und Term für Term der Summen jeder Einzelgleichung

$$A'_{1i} + A'_{3i} = B'_{1i} + B'_{3i}$$

gilt. Bleibt bezüglich (3.4.7a) zu bestätigen, daß

$$\sum_{i \in I} A'_{2i} = \sum_{i \in I} B'_{2i} \quad z \in Z$$

mit (3.4.7b) also

$$\sum_{i \in I} h(i,0) = \sum_{i \in I} h(0,i) = 1$$

sein sollte. Nun ist, wieder wegen der i -Voraussetzungen des Satzes,

$$\begin{aligned} \sum_{i \in I} h(i,0) &= \sum_{i \in I} \left\{ \sum_{j \in I} h(j,i) + h(0,i) \right\} \\ \sum_{j \in I} h(j,0) &= \sum_{j \in I} \left\{ \sum_{i \in I} h(j,i) \right\} + \sum_{i \in I} h(0,i) \\ \sum_{j \in I} \left\{ \sum_{i \in I} [1 - h(j,i)] \right\} &= 1 \end{aligned}$$

d.h. tatsächlich

$$\sum_{j \in I} h(j,0) = 1$$

womit Satz 3.4.5 insgesamt bestätigt ist.

Wenden wir uns nun den Möglichkeiten der praktischen Anwendung von Satz 3.4.5 zu. Bei Betrachtung des angegebenen Ergebnisses für die stationäre Zustandsverteilung exponentieller Netze fällt zunächst auf, daß über die Normalisierungskonstante G nichts Explizites ausgeführt wurde. Tatsächlich wird uns diese Konstante bei geschlossenen Netzen einigen Kummer bereiten (wir kommen in Kapitel 4 in breiterem Kontext darauf zurück). Bei offenen Netzen dagegen hat sie schlicht den Wert "1" - wie wir leicht einsehen, wenn wir uns zunächst die Frage nach der stationären Verteilung der Population an einer beliebigen Einzelstation des Netzes stellen. Sei diese Verteilung an einer Station i gekennzeichnet durch die Wahrscheinlichkeiten

*Anwendung
offenes
Netz*

$$PN_i(n) = P[\text{an Station } i \text{ sind } n \text{ Kunden}]$$

$\{PN_i(n); n \in \mathbb{N}_0\}$ ist eine "Randverteilung" der gemeinsamen Verteilung $\{P[z]; z \in Z\}$ nach Satz 3.4.5. Sie läßt sich gewinnen gemäß

$$PN_i(n) = \sum_{z: n_i=n} P[z]$$

mit Z gemäß (3.4.2b), da wir uns hier ausschließlich mit offenen Systemen beschäftigen. Unter Einsetzen der Ergebnisse des Satzes 3.4.5 erhalten wir weiter

$$PN_i(n) = \frac{1}{G} P_i(n) \prod_{j \in I \setminus i} \left\{ \sum_{n_j=0}^{N_0} P_j(n_j) \right\}$$

woraus wegen der Tatsache, daß die $\{P_j(n_j)\}$ echte Stationsverteilungen darstellen und daher jede der Summen den Wert 1 hat,

$$PN_i(n) = P_i(n)/G$$

sich ergibt; auch $\{P_i(n)\}$ ist eine Verteilung: G muß den Wert 1 haben. Fassen wir diese Resultate zusammen:

Korollar 3.4.8: Bezüglich der stationären Verteilung von exponentiellen Netzen, wie in Satz 3.4.5 angegeben, ergibt sich im offenen Netz für die Normalisierungskonstante:

$$G=1$$

und für die Stationsverteilungen

$$PN_i(n) = P_i(n) \quad n \in N_0, i \in I$$

Beispiel

Bemühen wir uns auf dieser Basis um das Beispiel unserer Testfrage 2.2.28 als Anwendungsfall. Die Struktur des Netzes wollen Sie bitte dort nachschlagen. Die weitere Spezifikation, entsprechend unseren jetzigen Bedürfnissen (aber in Anlehnung an die genannte Testfrage):

cpu (Station Nr. 3) bedient gemäß PS (Grenzfall Zeitscheibenverfahren) mit konstanter Bedienkapazität von 1 AE/ZE
 disk_1, disk_2 (Stationen Nr. 1 und 2) bedienen gemäß FCFS (also sequentiell), ebenfalls mit konstanter Bedienkapazität 1.

Kunden treffen an der Maschine gemäß eines Poisson-Stromes mit Parameter ein. Sie bewegen sich im Netz entsprechend der Wechselwahrscheinlichkeiten:

$$\begin{aligned} h(0,3) &= 1 \\ h(3,1) &= 20/31, \quad h(3,2) = 10/31, \quad h(3,0) = 1/31 \\ h(1,3) &= 1, \quad h(2,3) = 1 \\ h(i,j) &= 0, \text{ sonst} \end{aligned}$$

Die Bedienwünsche an allen Stationen sind exponentiell verteilt mit den Mittelwerten

$$1/\mu_3=2, \quad 1/\mu_1=1/\mu_2=20 \quad \text{Dimension: "ZE"}$$

Welche Ergebnisse lassen sich für dies Anwendungsbeispiel erreichen? Nun, eigentlich "fast alles", woran man denken mag; im folgenden nur eine Auswahl:

Durchsätze

- Satz 3.4.5 forderte zunächst die Lösung eines Gleichungssystems in (bisher uninterpretierten) Größen $\lambda_i, i \in I$. Wenn Sie zu Gleichungen (2.2.27) zurückblättern, finden Sie dort ein identisches Gleichungssystem in Größen $c_i, i \in I$. Die Interpretation dort war, bei offenem System: Die c_i sind die Stationsdurchsätze (mittlere Zahl von Ankünften/Abgängen je ZE). Interpretieren wir hier den Parameter des Poisson-Zugangstroms (vgl. diesbezügliche Diskussion in Abschn. 3.1) entsprechend als Systemdurchsatz (mittlere Zahl

von Systemankünften/-abgängen je ZE), dann entsprechen die ρ_i offenbar den Stationsdurchsätzen. Als Lösung des ρ_i -Gleichungssystems ergeben sich die Durchsätze

$$\rho_1=20 \cdot \lambda, \quad \rho_2=10 \cdot \lambda, \quad \rho_3=31 \cdot \lambda.$$

- Nach Sätzen 3.4.5/3.4.8 ist die stationäre Populationsverteilung einer Station i identisch mit der stationären Verteilung, welche die nämliche Station isoliert unter einem Poisson-Zugangstrom des Parameters ρ_i aufwiese. Die Verteilungen unserer isolierten Stationen (Bedienung PS, FCFS) kennen wir aus (3.3.14), welches Ergebnis bei der getroffenen Annahme konstanter Bediengeschwindigkeiten mit (3.3.2) zusammenfällt, in unseren jetzigen Bezeichnungen also lautet:

Populations-Verteilungen

$$PN_i(n) = P_i(n) = \binom{n}{i} (\rho_i/\mu_i)^n (1 - \rho_i/\mu_i) \quad i=1,2,3$$

Die Größen

Auslastungen

$$\text{d.h. } \rho_i := \lambda/\mu_i \quad i=1,2,3$$

$$\rho_1 = 400, \quad \rho_2 = 200, \quad \rho_3 = 62$$

sind als Belastung/Auslastung der betreffenden Station zu interpretieren, vgl. dazu (3.3.4). Als mittlere Population je Station liefert uns (3.3.2e):

$$E[N_i] = \rho_i / (1 - \rho_i) \quad i=1,2,3$$

Die Konvergenzbedingungen, vgl. (3.3.2b), nämlich

$$\rho_i / \mu_i < 1 \quad i=1,2,3$$

sind identisch mit den Flaschenhalsüberlegungen des Abschn. 2.2 und begrenzen die Systemzugangsrate (den Systemdurchsatz) auf

$$\lambda > \max(400, 200, 62)$$

$$\text{d.h.: } \lambda < 1/400$$

- Die mittleren Verweilzeiten $E[V_i]$ (je Kundenbesuch an Station i) liefert uns schließlich Little's Resultat zu

Stations-Populationen, Verweilzeiten

$$E[V_i] = E[N_i] / \mu_i$$

Nehmen wir etwa eine Zugangsrate von $\lambda = 0.001$ an, erhalten wir die konkreten Werte:

$$\rho_1 = 0.02, \quad \rho_2 = 0.01, \quad \rho_3 = 0.031$$

$$\rho_1 = 0.4, \quad \rho_2 = 0.2, \quad \rho_3 = 0.062$$

$$E[N_1] = 0.667, \quad E[N_2] = 0.25, \quad E[N_3] = 0.0661$$

$$E[V_1] = 33.3, \quad E[V_2] = 25.0, \quad E[V_3] = 2.13$$

- Bezüglich des Gesamtsystems erhalten wir eine mittlere Population von

Gesamt-Population, Verweilzeit

$$E[N] = E[N_1] + E[N_2] + E[N_3]$$

$$\text{hier: } \lambda = 0.983$$

und, wieder mit Little, die mittlere Gesamtverweilzeit (Bearbeitungszeit je

job, "turnaroundtime")

$$\begin{aligned} E[V] &= E[N] / \\ &= 983 \end{aligned}$$

"Fast alles", sagten wir, läßt sich ermitteln; obiger Ausschnitt sollte diese Behauptung belegt haben in dem Sinne, daß **Mittelwerte** interessanter Charakteristika und Leistungsgrößen sich de facto leicht gewinnen lassen.

Literatur zu Kap.3

- Bau74 Bauer,H.; Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie; de Gruyter, 1974
- 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)
- Dörf78 Dörfler,W.; Mathematik für Informatiker, Band 2; Hanser, 1978
- GoNe67 Gordon,W.J./Newell,G.J.; Closed Queueing Systems with Exponential Servers;
Operations Research, vol. 15 (1967)
- Jack63 Jackson,J.R.; Jobshop-like Queueing Systems; Management Science, vol. 10 (1963)
- Klei75 Kleinrock,L.; Queueing Systems; Wiley, 1975/1976

LEERSEITE

Lösungshinweise zu den Testfragen

3.1.4:

- Dichtefunktion:

$$f_A(a) = dF_A(a) / da$$

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

- Erwartungswert:

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

$$= \int_0^{\infty} a \exp(-a) da$$

partielle Integration:

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

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

$$= 0 - (0 - 1/a)$$

$$= 1/a$$

- Zweites Moment:

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

$$= \int_0^{\infty} a^2 \exp(-a) da$$

partielle Integration:

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

$$= 0 + (2/a) \int_0^{\infty} a \exp(-a) da$$

s. oben:

$$= (2/a) E[A] = 2/a^2$$

- Varianz:

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

s. oben:

$$= 2/a^2 - 1/a^2$$

$$= 1/a^2$$

- Variationskoeffizient:

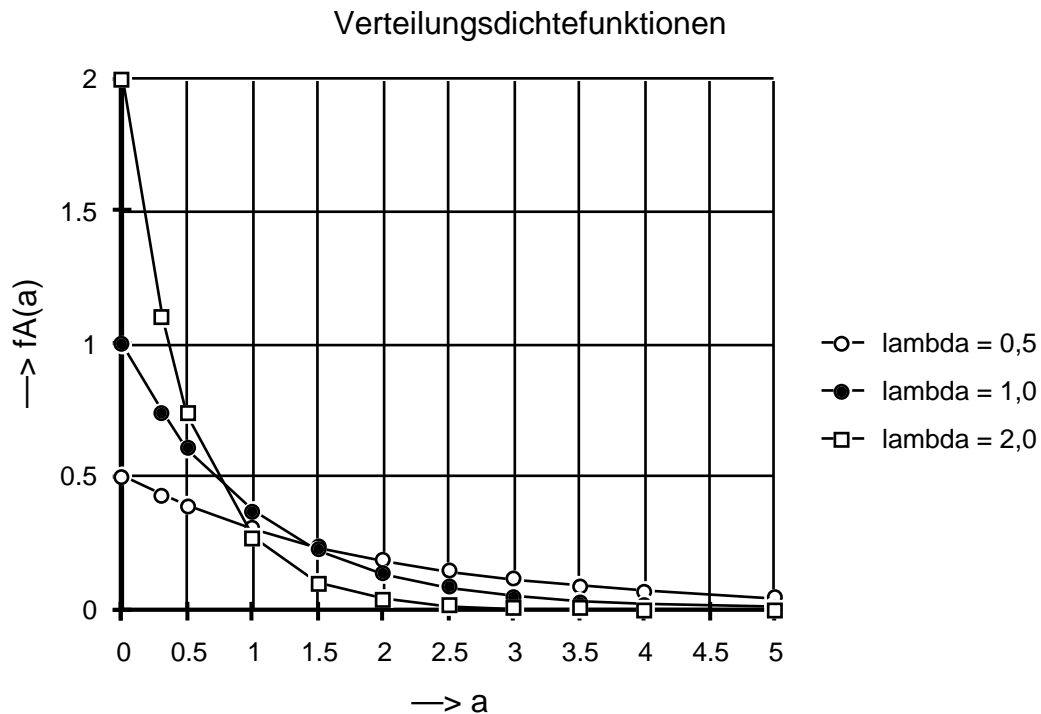
$$VK[A] = E[A] / \sqrt{V[A]}$$

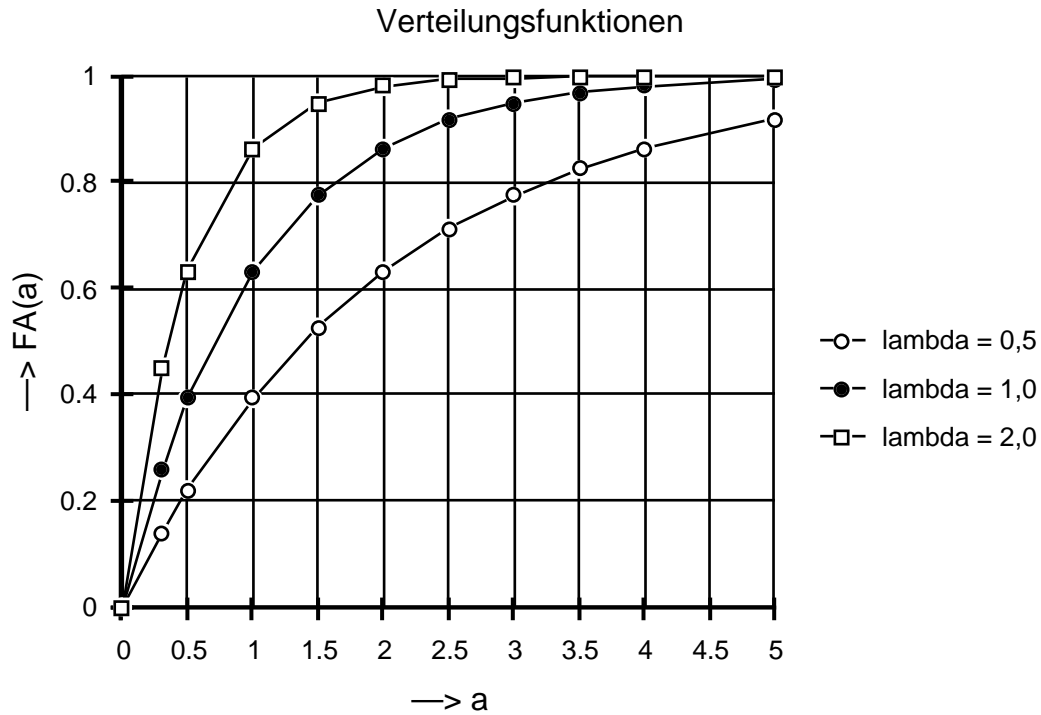
s. oben: = 1

- Eine kleine Wertetafel zur Erstellung der geforderten Graphen:

a	= 0.5			= 1.0			= 2.0		
	e^{-a}	$1-e^{-a}$	e^{-a}	e^{-a}	$1-e^{-a}$	e^{-a}	e^{-a}	$1-e^{-a}$	e^{-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 $f_A(a)$ und der Verteilungsfunktionen $F_A(a)$





3.1.9:

- zu (3.1.8b):

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

mit (3.1.3, 3.1.8a)

$$\begin{aligned}
 &= \int_0^a \exp(-x) \{ 1 - [1 + (a-x)] \exp[-(a-x)] \} dx \\
 &= \int_0^a \exp(-x) dx - \int_0^a [1 + (a-x)] \exp[-a] dx
 \end{aligned}$$

mit (3.1.2a):

$$\begin{aligned}
 &= 1 - \exp(-a) - \exp(-a) \left[(1+a)x - \frac{x^2}{2} \right]_0^a \\
 &= 1 - \exp(-a) - \exp(-a) \left[a + \frac{a^2}{2} - \frac{a^2}{2} \right] \\
 &= 1 - [1 + \frac{a^2}{2}] \exp(-a)
 \end{aligned}$$

- zu (3.1.8c):

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

mit (3.1.3,3.1.8c):

$$\begin{aligned}
 FA_{k+1}(a) &= \int_0^a \exp(-x) \left\{ 1 - \exp[-(a-x)] \sum_{i=0}^{k-1} \frac{(a-x)^i}{i!} \right\} dx \\
 &= 1 - \exp(-a) - \exp(-a) \int_0^a \dots dx \\
 &= 1 - \exp(-a) - \exp(-a) \left[\sum_{i=0}^{k-1} \frac{(a-x)^{i+1}}{(i+1)!} \right]_0^a \\
 &= 1 - \exp(-a) - \exp(-a) \sum_{i=0}^{k-1} \frac{(a)^{i+1}}{(i+1)!} \\
 &= 1 - \exp(-a) \left[1 + \sum_{i=1}^k \frac{(a)^i}{i!} \right] \\
 &= 1 - \exp(-a) \sum_{i=0}^k \frac{(a)^i}{i!}
 \end{aligned}$$

3.1.10:

- Dichte:

$$\begin{aligned}
 fA_k(a) &= dFA_k(a)/da \\
 &= \exp(-a) \sum_{i=0}^{k-1} \frac{(a)^i}{i!} - \exp(-a) \sum_{i=0}^{k-1} i \frac{(a)^{i-1}}{i!} \\
 &= \exp(-a) \left\{ \sum_{i=0}^{k-1} \frac{(a)^i}{i!} - \sum_{i=1}^{k-1} \frac{(a)^{i-1}}{(i-1)!} \right\} \\
 &= \exp(-a) \left\{ \sum_{i=0}^{k-1} \frac{(a)^i}{i!} - \sum_{i=0}^{k-2} \frac{(a)^i}{i!} \right\} \\
 &= \exp(-a) \frac{(a)^{k-1}}{(k-1)!}
 \end{aligned}$$

- Erwartungswert: $E[A_k] = \dots$
 Summe unabh. ZV: $= k \cdot E[A_1]$
 mit (3.1.3): $= k/$
- Varianz: $V[A_k] = E[(A_k - E[A_k])^2]$
 Summe unabh. ZV: $= k \cdot V[A_1]$
 mit (3.1.3): $= k/2$
- zw. Moment: $E[A_k^2] = V[A_k] + E^2[A_k]$
 s. oben: $= k/2 + k^2/2$
 $= k(1+k)/2$
- Variationskoeffizient:

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

s. oben

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

$$= 1 / \sqrt{k}$$

3.1.12:

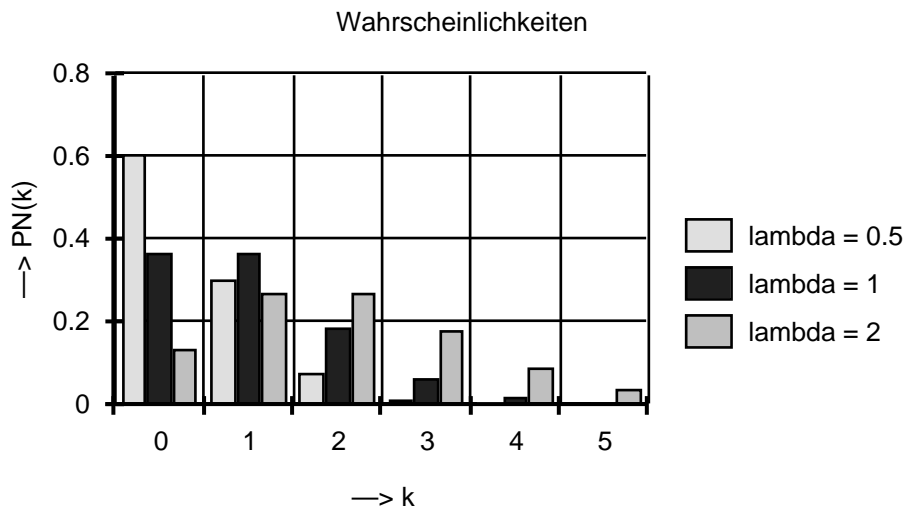
Für $t=1$ wird aus (3.1.11b):

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

Die Tabelle der Werte für

k	$\lambda = 1$ PN(k)	$\lambda = 0.5$ PN(k)	$\lambda = 2$ 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:



3.1.16:

(3.1.15a,b) direkt aus der Tatsache, daß genau zwei Möglichkeiten existieren (deren eine mit p , deren andere mit $1-p$ eintritt) und den Charakteristika der beiden (rein exponentiellen) Phasen B_1 und B_2 .

Bei H_R haben wir R exponentielle Phasen (Parameter: $\mu_1, \mu_2, \dots, \mu_R$), von denen genau eine mit Wahrscheinlichkeit p_r ($r=1, 2, \dots, R$; $\sum p_r=1$) ausgewählt wird. Wir erhalten für H_R -verteiltes B

$$FB(b) = \sum p_r \cdot (1 - \exp(-\mu_r b))$$

$$fB(b) = \sum p_r \cdot \mu_r \cdot \exp(-\mu_r b)$$

$$\begin{aligned} E[B] &= p_r / \mu_r \\ E[B^2] &= 2p_r / \mu_r \end{aligned}$$

$$\begin{aligned} \text{VK}[B] &= \sqrt{\text{V}[B]} / E[B] \\ &= \sqrt{E[B^2] - E^2[B]} / E[B] \end{aligned}$$

$$\begin{aligned} \text{VK}^2[B] &= (E[B^2] - E^2[B]) / E^2[B] \\ &= E[B^2] / E^2[B] - 1 \\ &= \{ 2p_r / \mu_r^2 \} / \{ p_r / \mu_r \}^2 - 1 \end{aligned}$$

Die sog. Cauchy-Ungleichung sagt aus, daß für reelle a_r, b_r

$$\left(\sum_r a_r b_r \right)^2 \leq \left(\sum_r a_r^2 \right) \left(\sum_r b_r^2 \right)$$

Setzen wir $a_r = p_r$ und $b_r = p_r / \mu_r$, dann ergibt sich

$$\left(\sum_r p_r / \mu_r \right)^2 \leq \left(\sum_r p_r \right) \left(\sum_r p_r / \mu_r^2 \right)$$

und da sich die p_r zu 1 addieren

$$\left(\sum_r p_r / \mu_r \right)^2 \leq \sum_r p_r / \mu_r^2$$

Somit erhalten wir für den Variationskoeffizienten der Hyperexponentialverteilung

$$\begin{aligned} \text{VK}^2[B] &= 2 - 1 \\ \text{VK}[B] &= 1 \end{aligned}$$

Im Unterschied zu den Erlang-Verteilungen, die ja durch $\text{VK} < 1$ gekennzeichnet waren, haben wir bei den Hyperexponentialverteilungen das Charakteristikum $\text{VK} = 1$ vorliegen; die Exponentialverteilung, mit $\text{VK} = 1$, liegt "in der Mitte" dieser beiden Verteilungsfamilien (und ist Grenzfall beider).

Lösung Testfrage 3.2.2 s. S. 3-59

3.2.6

- Stationswechsel:

Seien die Kunden, die Station i verlassen, in der Reihenfolge des Verlassens und beginnend mit einem beliebigen Kunden, fortlaufend durchnummeriert, etwa startend mit Nr. 1. Sei Z_k die Zielstation des k -ten (die Station i verlassenden) Kunden; Z_k ist gemäß (3.1.1) eine ZV; die Sequenz $(Z_k; k \in \mathbb{N})$ 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 \mathbb{N} : Der Prozeß ist Zeitdiskret. Die ZV sind wechselseitig unabhängig: Es handelt sich um einen unabhängigen Prozeß. Wegen der wechselseitigen Unabhängigkeit und der identischen Verteilung aller Z_k ist der Prozeß auch stationär.

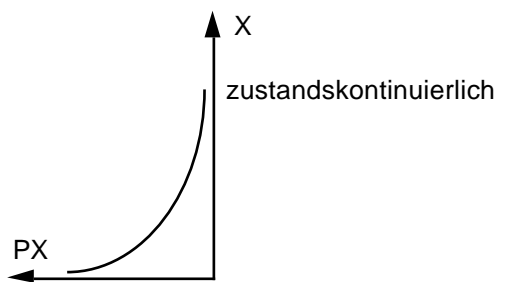
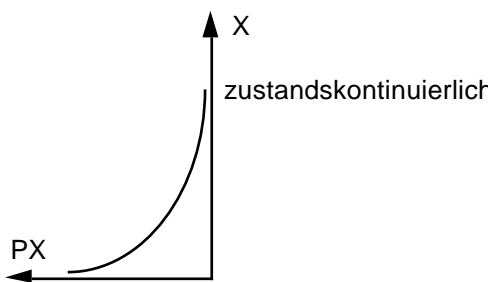
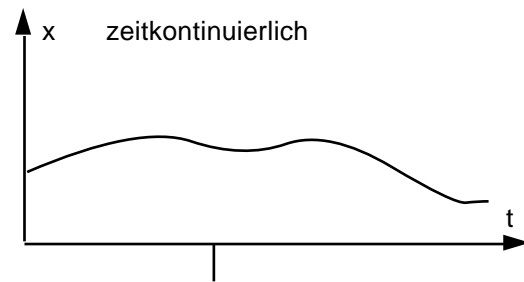
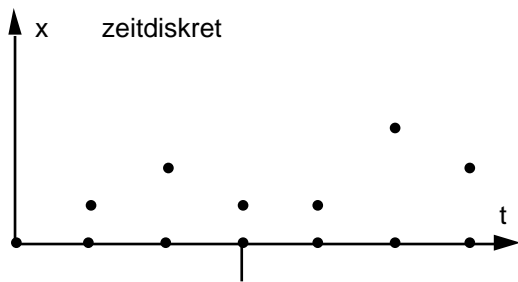
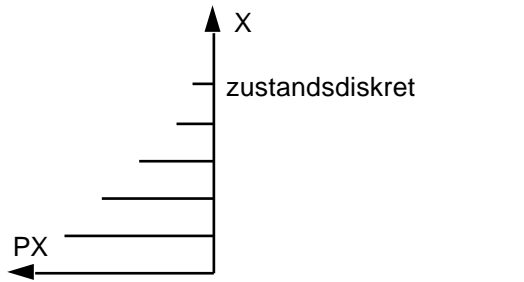
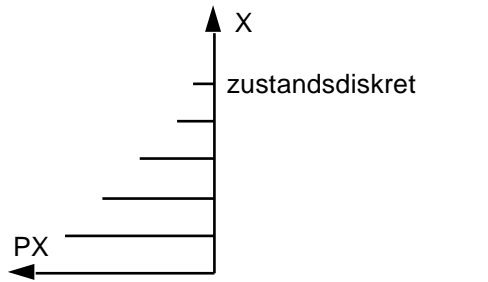
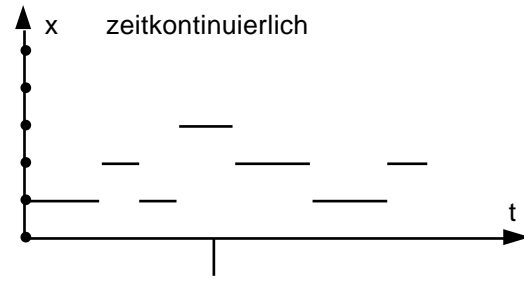
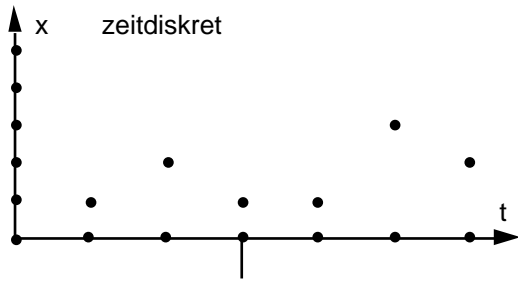
- Ankunftsabstände:

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 durchnummeriert. Die A_k sind ZV, $(A_k; k \in \mathbb{N})$ ist stochastischer Prozeß. Zustandsraum ist \mathbb{R}^+ (Wertebereich von Zwischenankunftsabständen): Der Prozeß ist Zustandskontinuierlich. Parametermenge des Prozesses ist \mathbb{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. $(W_k; k \in \mathbb{N})$ ist Zustandskontinuierlicher, Zeitdiskreter, unabhängiger, stationärer Prozeß.

3.2.2



3.2.7:

Offensichtlich ist 0 der Startzustand des Prozesses, d.h. $P[N_0=0]=1$, und der Prozeß insgesamt der Werte aus 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. Gleichermäßen (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 $n, <t$. In Formeln haben wir $(N_t; t \in \mathbf{R}^+)$ in den Beziehungen (3.1.11) bereits beschrieben.

3.2.22:

Es ist gemäß angegebener Formel

$$\begin{aligned} f_C(c) &= [1-F_B(c)] \cdot f_A(c) + [1-F_A(c)] \cdot f_B(c) \\ &= \exp(-B \cdot c) \cdot A \cdot \exp(-A \cdot c) + \exp(-A \cdot c) \cdot B \cdot \exp(-B \cdot c) \\ &= (A + B) \cdot \exp[-(A + B) \cdot c] \end{aligned}$$

C ist also wieder exponentiell verteilt, mit Parameter $C = A + B$.

Dies ist eine Bestätigung der Überlegungen zu (3.2.21). Zwei exponentielle Phasen "laufen" im Zustand (angenommene Bezeichnung:) k; daher ist q_{kk} gleich der negativen Summe der beiden Phasenparameter, also $-(A + B)$; $-q_{kk}$ ist Parameter der exponentiell verteilten Lebensdauer von Zustand k.

3.3.9:

- Zu RANDOM, LCFS:

Der Populationsprozeß $(N(t); t \in \mathbb{R}^+)$ an der Station anwesender Kunden, insbesondere auch seine stationäre Phase, war ja für M/M/1-FCFS durch das Übergangsratendiagramm vor (3.3.1) 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 (3.3.2, 3.3.6). Das Charakteristische dieses Diagramms war, daß für jeden Zustand $n \in \mathbb{N}_0$ die Zugangsrate λ betrug und für jeden Zustand $n \in \mathbb{N}$ 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. (3.1.5), so daß er in der Warteschlange nicht eigens markiert werden muß: Diagramm und Ergebnisse bleiben wie gehabt.

3.3.11c:

Aus (3.3.10) erhalten wir für die M/M/m-Stationen durch Setzung von $\mu_n = c_n \cdot \mu$ sowie $c_n = n$ ($n < m$) und $c_n = m$ ($n \geq m$) die stationäre Verteilung

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

für $n < m$:

$$= p_0 \left(\frac{\rho}{m}\right)^n / (n!)$$

für $n \geq m$:

$$= p_0 \left(\frac{\rho}{m}\right)^n / (m! m^{n-m})$$

Für p_0 ergibt sich:

$$\begin{aligned} p_0 &= \left\{ 1 + \sum_{n=1}^{m-1} \left(\frac{\rho}{m}\right)^n / n! + \frac{m!}{m^m} \sum_{n=m}^{\infty} \left(\frac{\rho}{m}\right)^n \right\}^{-1} \\ &= \left\{ 1 + \sum_{n=1}^{m-1} \left(\frac{\rho}{m}\right)^n / n! + \frac{(\rho/m)^m}{(m-1)! (m - \rho/m)} \right\}^{-1} \end{aligned}$$

(falls $\rho/m < m$).

LEERSEITE