9 Stochastische Modelle und Methoden

9.1 Grundbegriffe und Grundlagen
(Erinnerung)

9.1.1 Zufallsvariable und ihre Charakterisierung

- Eindimensionale (numerische) ZV Y
  mathematisches Modell einer numerischen Größe,
  - die in der Realität wiederholt beobachbar;
  - deren Wert über Beobachtungen (potentiell) variiert;
  - bei der gewisse Regelmäßigkeiten des Variierens
    zu erwarten: liegt Menge von Beobachtungen vor,
    dann liegen Beobachtungswerte
    mit bestimmten relativen Häufigkeiten
    in bestimmten Intervallen

- Gesetzmäßigkeiten des Variierens der **Zufallsvariablen** Y
  erfaßt durch deren **Verteilung**
  legt fest: für jedes Werte-Intervall I=(u,o]
  Wahrscheinlichkeit P[Y I]
  mit der **Realisierung** von Y im Intervall I
  statistische Interpretation:
  Intervall-Wahrscheinlichkeiten
  Modell der (erwarteten) relativen Häufigkeiten

- Viele Möglichkeiten, Verteilung einer ZV Y
  vollständig und sparsam zu charakterisieren:
  - Menge Wahrscheinlichkeiten / Verteilungsdichte
  - Verteilungsfunktion
  - Menge Momente / Zentralmomente
  - diverse Transformierte
    generierende Funktion
    Laplace-Transformierte …

- Axiomatische Fundierung (Kolmogorov)

- Sei
  Ω' Menge von Elementarereignissen ω' Ω'
  A' über Ω' definierte σ-Algebra von Ereignissen a' A'
  P':A' → [0,1] auf A' definiertes (auf 1 normiertes) Maß,
  genannt: **Wahrscheinlichkeitsmaß**
  dann heißt
  Maßraum (Ω',A',P') **Wahrscheinlichkeitsraum**

- Sei ferner
  Ω Bildmenge von Elementen ω Ω
  A über Ω definierte σ-Algebra von Elementen a A

  dann heißt
  jede A'-A-meßbare Abbildung
  Y: Ω' → Ω **Zufallsvariable**
  und (induziertes) Bildmaß
  P=Y(P') **Verteilung** von Y

  P (auch: PY) ist Wahrscheinlichkeitsmaß auf A,
  für a A gilt PY(a) = P[Y a]
  für Wahrscheinlichkeit, daß Y a

- Ist insbesondere
  Ω = $\mathbf{R}^1$ Menge der reellen Zahlen
  A = B^1 σ-Algebra der Borelschen Mengen des $\mathbf{R}^1$
  dann heißt eine A'-B^1-meßbare Abbildung
  Y: Ω' → $\mathbf{R}^1$ **reelle Zufallsvariable**

- Anwendungen
  arbeiten (gerne) völlig im Bildraum (Ω,A,P)
  zB Würfeln (Standardbeispiel)
  Bearbeitungsdauern (Abschn. 6.3)
  Bestellmengen (Abschn. 8.4)

- Verteilung reeller ZV Y (Bildmenge $\mathbf{R}^1$)

- charakterisierbar durch **Verteilungsfunktion**
  FY(y) := P[Y≤y] y $\mathbf{R}^1$

- **diskrete** (reelle) ZV Y nehmen Werte y_k aus
  endlicher / abzählbarer Wertemenge an:
  $\{y_k \,;\, k=0,1,\ldots\}$
  ° zugehörige Wahrscheinlichkeiten
  $p_k := P[Y=y_k]$ (= $PY(y_k)$) k=0,1,…
  charakterisieren Verteilung (PMF, Probabiliy Mass Function)
  ° Verteilungsfunktion

  $$FY(y) = \sum_{y_k \le y} p_k$$

  ist reine Treppenfunktion
  ° Intervallwahrscheinlichk'n aus beiden Charakterisier'gn

  $$P[\,Y \quad (u,o]\,] = \sum_{y_k \quad (u,o]} p_k$$

  bzw = FY(o) – FY(u)

- **kontinuierliche** (reelle) ZV Y nehmen Werte y aus
  Wertekontinuum an (idR konvexes $\mathbf{R}^1$-Intervall)
  ° **Verteilungsdichte**(funktion) (PDF, Probabiliy Density Function)
  fY(y) y $\mathbf{R}^1$
  charakterisiert Verteilung
  ° mit Verteilungsfunktion besteht Zusammenhang

  $$FY(y) = \int_{-\infty}^{y} fY(y')\,dy'$$

  bzw
  fY(y) = dFY(y)/dy
  ° Verteilungsfunktion ist stetig

° Intervallwahrscheinlichk'n aus beiden Charakterisier'gn

$$P[\,Y \quad (u,o]\,] = \int_{u}^{o} fY(y')\,dy'$$

bzw = FY(o) – FY(u)

- gemeinsame Behandlung von
  diskreten, kontinuierlichen, "gemischten" ZV
  mit Hilfe Stieltjes-Integralbegriff,
  erlaubt Darstellung

$$FY(y) = \int_{-\infty}^{y} dFY(y')$$

$$P[\,Y \quad (u,o]\,] = \int_{u}^{o} dFY(y')$$

- **Erwartungswert** einer ZV Y definiert als

$$E[Y] = \int_{-\infty}^{\infty} y'\,dFY(y')$$

diskret $= \sum_{y_k \quad (-\infty,\infty)} y_k\,p_k$

kontinuierlich $= \int_{-\infty}^{\infty} y'\,fY(y')\,dy'$

Erwartungswert einer Funktion g(Y) einer ZV Y als

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y')\,dFY(y')$$

diskret $= \sum_{y_k \quad (-\infty,\infty)} g(y_k)\,p_k$

kontinuierlich $= \int_{-\infty}^{\infty} g(y')\,fY(y')\,dy'$

- Erwartungswerte der Funktionen
  $g(Y) = Y^k \qquad k=1,2,\ldots$
  heißen (k-te) **Momente**

$$\mu_k := E\left[Y^k\right] = \int_{-\infty}^{+\infty} y'^k \, dFY(y') \qquad k=1,2,...$$

- Erwartungswerte der Funktionen
  $g(Y) = (Y-E[Y])^k \qquad k=2,3,\ldots$
  heißen (k-te) **Zentralmomente**
  $E[(Y-E[Y])^k] \qquad k=2,3,\ldots$
- existieren alle k-ten Momente k=1,2,… , einer ZV, dann charakterisieren sie Verteilung eindeutig
- oft Beschränkung auf partielle Charakterisierungen
  - ° Teilmenge Momente / Zentralmomente
  - ° aus diesen abgeleitete Kenngrößen
  - ° 1. Moment — Erwartungswert E[Y]
  - ° 2. Zentralmoment — **Varianz** V[Y]
    $V[Y] := E[(Y-E[Y])^2]$
  - ° Wurzel 2. Zentralmoment — **Streuung** σ[Y]
    $\sigma[Y] := +\sqrt{V[Y]}$
  - ° relative Streuung — **Variationskoeffizient** VK[Y]
    $VK[Y] := \sigma[Y] / E[Y]$
  - ° uam

9.1.2 Stochastische Prozesse

statt: einzelne ZV

jetzt: Menge zusammengehöriger ZV,
"Familie" von ZV

Mitglieder der Familie unterschieden durch Index t, in geordneter Indexmenge T variierend:

$(Y_t)_{t \ \ T}$ oder auch: $(Y(t);t \ \ T)$ oä

alle Y_t als Abbildungen definiert
- aus derselben Menge von Elementarereignissen
- auf denselben Wertebereich Ω
- samt ihrer Verteilungen (Wahrscheinlichkeits-Maße) P_t

T: "Parametermenge"
Ω: "Zustandsraum" des Prozesses

"Zusammengehörigkeit" der Familie (Y_t):
- Realisierungen (y_t) "simultan entstehend", aus **einem** Elementarereignis resultierend
- jedes Elementarereignis σ impliziert Abbildung $t \to y_t(\sigma), \ T \to \Omega$

Ist gar nicht so kompliziert:

- Beispiel:
- $\Omega=\mathbf{R^+}$, $T=\mathbf{R^+}$ gewählt
- Realisierungen des Prozesses betrachtet
  (Realisierungen aller ZV der Familie)
  für zwei Elementarereignisse σ_1, σ_2

- zu jedem Elementarereignis (σ_1 und σ_2)
  gehört Realisierung stochastischer Prozeß
  ($y_t(\sigma_1)$ und $y_t(\sigma_2)$),
  die gemeinsame Realisierung aller beteiligten ZV,
  die Abbildung $T \to \Omega$ ($\mathbf{R^+} \to \mathbf{R^+}$)
- häufigste Interpretation Parametermenge T ist "Zeit",
  damit Realisierung stochastischer Prozeß
  Verlauf seines Zustands (aus $\mathbf{R^+}$) über der Zeit,
  auch "Trajektorie"

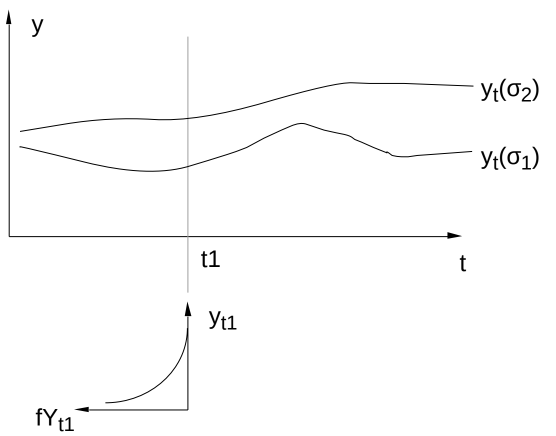


Abbildung 9.1.01 : Trajektorien stochastischer Prozeß, Zustandsverteilung zu festem Zeitpunkt

- für festen Beobachtungszeitpunkt t (t_1)
  ("quer" über alle Trajektorien / möglichen Z'Verläufe)
  Zufallsvariable Y_{t1} : Verteilung z.B. wie Dichtefunktion
- anderer Zeitpunkt $t_2 \neq t_1$: andere ZV Y_{t2}, andere Verteil'g.
- Y_{t1} und Y_{t2} aber **abhängig**: **ein** Elementarereignis !

bei praktischen Anwendungen
- Bezug zu Elementarereignissen meist unterdrückt
- Bewegung völlig im Zustandsraum Ω

Klassifizierungen stochastischer Prozesse bezüglich

(a) der Natur des Zustandsraums Ω,
häufig betrachtete Fälle
$\Omega = \mathbf{N}_0$ (zustandsdiskreter Prozeß)
$\Omega = \mathbf{R}^+$ (zustandskontinuierlicher Prozeß)

(b) der Natur der Parametermenge T,
häufig betrachtete Fälle
$T = \mathbf{N}_0$ (zeitdiskreter Prozeß)
$T = \mathbf{R}^+$ (zeitkontinuierlicher Prozeß)

(c) der Art der Abhängigkeiten der Zufallsvariablen $(Y_t)_{t \ \ T}$,
erfaßbar durch "gemeinsame Verteilung",
z.B. gemeinsame Verteilungsfunktion (τ T , r=|τ|)

$$F\mathbf{Y}(\mathbf{t}_\tau , \mathbf{y}_\tau) = P[Y(t_1) \le y_1, Y(t_2) \le y_2, ..., Y(t_r) \le y_r]$$

"kompaktere" Abhängigkeitscharakterisierungen gefragt

Einige (sehr unterschiedliche) Abhäng.Charakterisierungen:

Definition 9.1.02 : Unabhängige Prozesse

Grenzfall **keinerlei** Abhängigkeiten,
Familie **unabhängiger** ZV,
für abzählbares τ T

$$F\mathbf{Y}(\mathbf{t}_\tau , \mathbf{y}_\tau) = \prod_{t \ \ \tau} FY_t(y_t)$$

Definition 9.1.03 : Stationäre Prozesse

Zeit-Invarianz der Abhängigkeiten, formal:
Für alle τ T, und alle s>0 ("sinnvoll")
gilt: $F\mathbf{Y}(\mathbf{t}_\tau, \mathbf{y}_\tau) = F\mathbf{Y}(\mathbf{t}_\tau+\mathbf{s}, \mathbf{y}_\tau)$ wo $\mathbf{s}=(s,s,...,s)^T$
insbesondere, für |τ|=1:
alle ZV $Y(t_i)$, t_i τ, identisch verteilt (nicht: unabhäng. !)

Definition 9.1.04 : Markovsche Prozesse

Unabhängigkeit von Vergangenheit, formal:
Für alle $\tau = \{t_i ; i \;\; I\}$ T
und $t^* := \max(t_i ; i \;\; I)$
und beliebiges t T mit t>t*
gilt: $FY(t \mid \mathbf{Y}(\mathbf{t}_\tau)=\mathbf{y}_\tau) = FY(t \mid Y(t^*)=y^*)$
"gegenwärtiger Zustand sagt alles über Zukunft":
"**Gedächtnislosigkeit**"

Ab jetzt: Konzentration auf

Markov-Prozesse
/ Semi-Markov-Prozesse (Def "später")

9.1.3 Markov-Ketten mit diskreter Parametermenge
Discrete Time Markov Chains: DTMCs

- Sei
  $(Y_t)_{t \;\; T}$ stochastische **Kette**, Zustandsraum $\Omega=\mathbf{N}_0$
  T ihre Parametermenge T,
  diskret, speziell $T=\mathbf{N}_0$
  Darstellung der Kette auch als
  $(Y_n)_{n=0,1,...}$

  Kette besitze Markov-Eigenschaft,
  im vorliegenden Falle ausdrückbar als

  $$P[Y_n=i_n \mid Y_{n-1}=i_{n-1},...,Y_0=i_0] = P[Y_n=i_n \mid Y_{n-1}=i_{n-1}] \qquad n\geq 1,\ i_j \;\; \mathbf{N}_0,\ j \;\; [0..n]$$

- Größen (bedingte Wahrscheinlichkeiten)

  $$\pi_{ij}(n-1,n) := P[Y_n=j \mid Y_{n-1}=i] \qquad n\geq 1;\ i,j \;\; \mathbf{N}_0$$

  heißen (1-Schritt-) **Übergangswahrscheinlichkeiten**

- Größen (totale Wahrscheinlichkeiten)

  $$p_i(n) := P[Y_n=i] \qquad i,n \;\; \mathbf{N}_0$$

  heißen (zeitabhängige) **Zustandswahrscheinlichkeiten**

- Zusammenhang zwischen p.(.) und π.(.,.) offensichtlich

(9.1.05a) $p_j(n) = \sum_{i \;\; \mathbf{N}_0} p_i(n-1) \;\; \pi_{ij}(n-1,n) \qquad n\geq 1;\ j \;\; \mathbf{N}_0$

- sind die Übergangswahrscheinlichkeiten stationär
  (zeitunabhängig),
  also
  $\pi_{ij}(n-1,n) \equiv \pi_{ij} \qquad n\geq 1;\ i,j \;\; \mathbf{N}_0$
  heißt die Kette (zeit-)**homogen**
  und aus (9.1.05a) wird

  $$p_j(n) = \sum_{i \;\; \mathbf{N}_0} p_i(n-1) \;\; \pi_{ij} \qquad n\geq 1;\ j \;\; \mathbf{N}_0$$

  bzw. in Vektor/Matrix-Schreibweise
  (**p**: Zustandsvektor, Π: Übergangsmatrix)

(9.1.05b) $\mathbf{p}^T(n) = \mathbf{p}^T(n-1) \cdot \Pi$

- für homogene Ketten sind auch die
  m-Schritt-Übergangswahrscheinlichkeiten

  $$\pi_{ij}^{(m)} := P[Y_{n+m}=j \mid Y_n=i] \qquad m\geq 2$$

  zeitunabhängig, da aus (9.1.05b) folgt daß

  $$\begin{aligned} \mathbf{p}^T(n+m) &= \mathbf{p}^T(n+m-1) \cdot \Pi \\ &= (\mathbf{p}^T(n+m-2) \cdot \Pi) \cdot \Pi = \dots \\ &= \mathbf{p}^T(n) \cdot \Pi^m \end{aligned}$$

  und für die **m-Schritt-Übergangsmatrix** $\Pi^{(m)}$ gilt, daß
  $\Pi^{(m)} = \Pi^m$

  sowie, verallgemeinernd, die
  Chapman-Kolmogorov-Beziehungen (gerafft / detailliert)

(9.1.06a) $\Pi^{(n+m)} = \Pi^{(n)} \cdot \Pi^{(m)}$

(9.1.06b) $\pi_{ij}^{(n+m)} = \sum_k \pi_{ik}^{(n)} \;\; \pi_{kj}^{(m)}$

- der Vollständigkeit halber (konsistent mit Definitionen)
  $\Pi^{(1)} = \Pi \qquad \Pi^{(0)} = I$ (Einheitsmatrix)

- mit (9.1.05/.06) lassen sich,
  startend mit der
  Anfangsverteilung **p**(0),

  die zeitabhängigen,
  transienten Zustandswahrscheinlichkeiten
  p(n) n=1,2,…
  ermitteln (numerisch nur bei endlichem Z'Raum)

- Gleichung (9.1.05b) wirft naheliegende Frage auf,
  ob Verteilungen **p**(n) gegen Grenzverteilung

  $$\mathbf{p} := \lim_{n\to\infty} \mathbf{p}(n)$$

  streben, so daß
  Zust'ds-Wahrsch'n Prozeß für große n beschrieben
  durch (zeitunabhängige) Grenzverteilung

  bei Existenz ergäbe sich Grenzverteilung,
  gemäß (9.1.05b), als Lösung Gleichungssystem

(9.1.07) $\mathbf{p}^T = \mathbf{p}^T \;\; \Pi \qquad \sum_i p_i = 1$

  Grenzverteilung heißt (bei Existenz) auch
  Gleichgewichtsverteilung,
  da Prozeß bei Wahl der Anfangsverteilung
  p(0) = **p**
  stationär (Stationarität auch möglch
  ohne Existenz Grenzverteilung)

→ **Wann existiert** (höchst interessante) **Grenzverteilung** ??

- Markov-Kette heißt **irreduzibel**,
  wenn jeder Zustand
  sich von jedem anderen Zustand
  erreichen läßt:

  $$i,j \quad \mathbf{N}_0 : (\quad m_{ij} : \pi_{ij}^{(m_{ij})} > 0)$$

- sei Wahrscheinlichkeit, von Zustand i aus
  in genau m Schritten
  Zustand j zu erreichen
  $f_{ij}^{(m)}$:= P[Zustand j wird von Zustand i aus
  erstmals nach m Schritten erreicht]

  und Wahrscheinlichkeit, von Zustand i aus
  in genau m Schritten
  in Zustand zurückzukehren
  $f_{jj}^{(m)}$:= P[erste Rückkehr in Zustand j
  geschieht nach m Schritten]

  dann ist Wahrscheinlichkeit,
  jemals wieder zurückzukehren

  $$f_{jj} := \text{P[Rückkehr in Zustand j]} = \sum_{m=1}^{\infty} f_{jj}^{(m)}$$

→ Klassifizierung Zustände
  $f_{jj}=1$ Zustand j heißt **rekurrent**
  $f_{jj}<1$ Zustand j heißt **transient**
  bei Rückkehr nur nach r>1 Schritten
  d.h. $f_{jj}^{(m)}=0$ für m<r
  Zustand j heißt **periodisch** mit Periode r,
  sonst Zustand j heißt **aperiodisch**

- in rekurrente Zustände kehrt Prozeß
  nach (im Mittel) M_{jj} Schritten zurück, wo

  $$M_{jj} := \sum_{m=1}^{\infty} m \quad f_{jj}^{(m)}$$

→ Klassifizierung Zustände
  $M_{jj}=\infty$ Zustand j heißt **null-rekurrent**
  $M_{jj}<\infty$ Zustand j heißt **positiv rekurrent**

→ Klassifizierung Ketten
  sind sämtliche Zustände einer Kette aperiodisch
  (periodisch, positiv rekurrent, null-rekurrent)
  dann heißen Ketten aperiodisch
  (periodisch, rekurrent, null-rekurrent)
  aperiodische, rekurrente Ketten heißen **ergodisch**

- für rekurrente, irreduzible Ketten gilt,
  neben Definitionsbedingung $f_{jj}=1$,
  daß für alle Zustandspaare (i,j)

  $$f_{ij} := \sum_{m=1}^{\infty} f_{ij}^{(m)} = 1$$

Satz 9.1.08 : Zustände irreduzibler DTMC

In irreduzibler Markov-Kette sind
entweder alle Zustände transient
oder alle null-rekurrent
oder alle positiv rekurrent.
Falls Periodizität vorliegt, alle Zustände identische Perioden r

oB

→ bei Überprüfung irreduzibler Kette
  auf Eigenschaften (wird interessieren!)
  genügt Überprüfung eines Zustands
  auf Periodizität, Rekurrenz, Ergodizität

Satz 9.1.09 : Konvergenzbedingungen DTMC

Für eine irreduzible, aperiodische Markov-Kette
existieren die Grenzwahrscheinlichkeiten

$$p_j := \lim_{n \to \infty} p_j(n) \qquad j \quad \mathbf{N}_0$$

und sind unabhängig von der Anfangsverteilung.

Zusätzlich sind
entweder
- alle Zustände transient
  oder alle Zustände null-rekurrent,
  in welchen Fällen sich
  $p_j = 0 \qquad j \quad \mathbf{N}_0$
  ergibt und keine Grenzverteilung existiert

oder
- alle Zustände positiv rekurrent, in welchem Falle die
  $p_j \qquad j \quad \mathbf{N}_0$
  eine stationäre Verteilung charakterisieren.
  In diesem Falle ist auch
  $p_j = 1/M_{jj} \quad j \quad \mathbf{N}_0$
  und die p_j ergeben sich eindeutig aus Gleichungssystem

  $$p_j = \sum_i p_i \quad \pi_{ij}$$
  $$\sum_j p_j = 1$$

  \- dies ist genau Gleichung (9.1.07) - oB

ungemein hilfreich (Überprüfungen können uU entfallen):

Satz 9.1.10 : Ergodizität DTMC

Eine irreduzible, aperiodische Kette
ist genau dann positiv rekurrent,
wenn das Gleichungssystem $\mathbf{p}^T = \mathbf{p}^T \Pi$
eine Lösung besitzt mit $\Sigma\, p_j = 1$

oB

9.1.4 Markov-Ketten m. kontinuierlicher Parametermenge
Continuous Time Markov Chains: CTMCs

- Zeitkontinuierliche Markov-Ketten besitzen
  - kontinuierlichen Parameterraum
  - diskreten Zustandsraum
  - typisch für Anwendungen: $(\Omega,T) = (\mathbf{N}_0,\mathbf{R}^+)$

- Markov-Eigenschaft formal:

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

  $$P[Y(t_{n+1})=i_{n+1}|Y(t_1)=i_1,...,Y(t_n)=i_n] = P[Y(t_{n+1})=i_{n+1}|Y(t_n)=i_n] \quad t_i \; \mathbf{R}^+;\; i_j \; \mathbf{N}_0,\; j=1,2,...,n+1$$

- Prozeßzustand zu festem t
  beschrieben durch Verteilung der ZV Y(t),
  durch **Zustandswahrscheinlichkeiten**

  $$p_i(t) := P\left[X(t)=i\right] \quad i \;\; \mathbf{N}_0$$
  $$\sum_i p_i(t) = 1$$

- für unterschiedliche Zeitpunkten t_1,t_2 (sei $t_1<t_2$):
  abhängige Zustandsverteilungen

  herrsche Zustand i zum Zeitpunkt t_1
  - Wahrscheinlichkeit für Zustand zum Zeitpunkt t_2 ?
    von Vor-Zustand i abhängig
  - mit dieser Abhängigkeit aber Abhängigkeit
    von gesamter "Vorgeschichte" (bis t_1) erfaßt
    (Markov-Eigenschaft!)

- Abhängigkeitsstruktur sinnvoll durch
  (Zustands-)**Übergangswahrscheinlichkeiten** erfaßbar

  $$\pi_{ij}(t_1,t_2) := P[Y(t_2)=j|Y(t_1)=i] \qquad t_2>t_1;\; t_1,t_2 \;\; \mathbf{R}^+;\; i,j \;\; \mathbf{N}_0$$
  $$\sum_j \pi_{ij}(t_1,t_2) = 1 \qquad i \;\; \mathbf{N}_0$$

- Zusammenhang zwischen p.(.) zu Zeitpunkten t_1 und t_2
  und $\pi.(.,.)$ offensichtlich:

(9.1.11) $$p_j(t_2) = \sum_{i \;\; \mathbf{N}_0} p_i(t_1) \;\; \pi_{ij}(t_1,t_2) \qquad t_2>t_1,j \;\; \mathbf{N}_0$$

- Übergangswahrscheinlichkeiten bisher sehr allgemein:
  Abhängigkeitsstruktur darf sich über der Zeit ändern,
  ist für jedes Paar (t_1,t_2) spezifisch

  sehr viel einfacherer Fall:
  Abhängigkeitsstruktur **zeitinvariant**
  praktisch sehr relevanter Fall
  (Gesetze der Dynamik
  eines arbeitenden Systems ia konstant)

  bei Betrachtung der Zeitpunkten (t_1,t_2)
  - absolute Lage t_1 unerheblich,
  - nur noch Differenz t_2-t_1 maßgeblich

→ **(zeit-)homogener Markov-Prozeß**

Abhängigkeitsstruktur homogenen Prozesses erfaßt
(für alle t) durch Übergangswahrscheinlichkeiten
(Umbenennung!)

(9.1.12a) $\pi_{ij}(s) := P[Y(t+s)=j|Y(t)=i] \qquad s>0;\; i,j \;\; \mathbf{N}_0$

(9.1.12b) $\sum_j \pi_{ij}(s) = 1 \qquad i \;\; \mathbf{N}_0$

und aus (9.1.11)

(9.1.13a) $p_j(t+s) = \sum_{i \;\; \mathbf{N}_0} p_i(t) \;\; \pi_{ij}(s) \qquad t, s > 0;\; j \;\; \mathbf{N}_0$

- mit Schreibvereinfachung (Vektor-/Matrix-Form)
  $\Pi(s) := (\pi_{ij}(s) \;;\; i,j \;\; \mathbf{N}_0)$ Matrix Übergangsw'n
  $\mathbf{p}(t) := (p_j(t) \;;\; j \;\; \mathbf{N}_0)$ Spaltenvektor Zust'dsw'n
  kompakte Form

(9.1.13b) $\mathbf{p}^T(t+s) = \mathbf{p}^T(t) \cdot \Pi(s)$

- mit expliziter Form für Übergangswahrsch'n $\Pi(s)$
  (erhältlich!)
  sind aus beliebigen Start-Zustandsverteilungen $\mathbf{p}(0)$
  (zeitabhängige) Zustandsverteilungen für ZV Y(t) gemäß

(9.1.14) $\mathbf{p}^T(t) = \mathbf{p}^T(0) \cdot \Pi(t)$

errechenbar (mühsam!)

- statt dessen unmittelbar (vgl. DTMCs):
  - Zustandsverteilung t von
    Start-Zustandsverteilung $\mathbf{p}(0)$ abhängig
  - vernünftig, zu vermuten, daß Abhängigkeit
    "mit der Zeit" (t wachsend) abnimmt,
    "für gewisse Systeme" verschwindet

- falls zutreffend, dann
  alle Übergangsw'n $\pi_{ij}(t)$, i $\mathbf{N}_0$
  für große t zum selben Wert konvergierend !
  Matrix $\Pi(t)$ für große t
  Matrix Π mit identischen Zeilen zustrebend

  mit (9.1.14) auch

(9.1.15) $\lim_{t \to \infty} p_j(t) = p_j \quad j \;\; \mathbf{N}_0$

Prozeß mit t "stationär. Verhalten" zustrebend (s. Def.)

- falls Grenzverteilung $\mathbf{p} := (p_j;j \;\; \mathbf{N}_0)$ gemäß (9.1.15)
  existierend, dann wie folgt bestimmbar:

  aus (9.1.13) **E** Einheitsmatrix
  $$\mathbf{p}^T(t+s) - \mathbf{p}^T(t) = \mathbf{p}^T(t) \cdot (\Pi(s)-\mathbf{E})$$
  $$\frac{\mathbf{p}^T(t+s) - \mathbf{p}^T(t)}{s} = \mathbf{p}^T(t) \;\; \frac{\Pi(s) - \mathbf{E}}{s}$$
  und im Grenzübergang $s \to 0$

(9.1.16a) $$\frac{d\mathbf{p}^T(t)}{dt} = \mathbf{p}^T(t) \;\; \lim_{s \to 0} \frac{\Pi(s) - \mathbf{E}}{s} = \mathbf{p}^T(t) \;\; \mathbf{Q}$$

wo Matrix $\mathbf{Q}=(q_{ij} \;;\; i,j \;\; \mathbf{N}_0)$ mit Elementen

(9.1.16b) $$q_{ij} = \lim_{s \to 0} \left[\pi_{ij}(s)/s\right] \qquad i \neq j$$
$$q_{ii} = \lim_{s \to 0} \left[(\pi_{ii}(s) - 1)/s\right]$$

- unsere Vermutung:

  für große t Prozeßverhalten –> stationäres Verhalten

$\mathbf{p}(t)$	de facto zeitunabhängig,
$d\mathbf{p}(t)/dt$	de facto verschwindend

  aus (9.1.16) für hinreichend große t

(9.1.17a) $\mathbf{0}^T = \mathbf{p}^T \mathbf{Q}$

(9.1.17b) $\sum_{i \; \mathbf{N}_0} p_i = 1$

- lineares Gleichungssystem (9.1.17) direkt nutzbar

  - um aus Q (Matrix, deren Elemente alle Informationen über Charakteristika speziellen Prozesses, also konkreten Problems, enthalten)

  - "stationäre Verteilung" **p** des Prozesses zu bestimmen, d.h. Wahrscheinlichkeiten p_j, j $\mathbf{N}_0$, Prozeß für große t in seinen Zuständen anzutreffen (unabhängig von t und Start-Z'd)

- Überlegungen basieren auf Prämisse, daß stationäre Verteilung (9.1.15) existiert, d.h. "unter gewissen Voraussetzungen"

  dazu folgender Satz (Wißbegierige: s. Cinl75)

Satz 9.1.18 : Konvergenzbedingungen CTMC

Zeitkontinuierliche homogene Markov- Kette
heißt **irreduzibel**,
wenn jeder Zustand von jedem Zustand "erreichbar" ist
(für alle Zustandspaare (i,j) Werte t>0, so daß $\pi_{ij}(t)>0$).

Irreduzible (zeitkontinuierliche, homogene) Markov-Kette
besitzt eindeutige stationäre Grenzverteilung (9.1.17)
genau dann, wenn (9.1.17) Lösung besitzt

oB

große praktische Bedeutung:

- wenn "stabiles" Verhalten M-Kette (System) interessant (–> stationäre Grenzverteilung,
  auch "eingeschwungene", "Gleichgewichts-"Phase,
  im Unterschied zu anfänglicher "transienter" Phase)

- und wenn Kette (zeit-)homogen
  (i. allg. direkt dem Problem entnehmbar)

- und wenn Kette irreduzibel
  (i. allg. direkt dem Problem entnehmbar)

- dann lediglich Gleichungssystem (9.1.17) aufzustellen und nach **p** zu lösen

- falls gelingend, "gewisse Voraussetzungen" für Existenz stationärer Grenzverteilung erfüllt, und diese bereits eindeutig ermittelt

praktische Durchführung:

- (9.1.17) aufstellen:
  aus gegebenem Problem Matrix **Q** ableiten

  dazu erforderlich:
  Klärung der Bedeutung der
  - in (9.1.16) nur formal eingeführten -
  Elemente q_{ij} der Matrix **Q**

- (9.1.17) lösen:
  "nur" Lösung linearen Gleichungssystems,
  allerdings mit potentiell sehr großer (bis: abzählbarer) Dimension (=Mächtigkeit des Zustandsraums)

- Fortsetzung nach

  Einschub: EXPONENTIALVERTEILUNG

  wird sich
  - als inhärent mit CTMCs "verquickt"
  - als problemseitig gut "anwendbar"

  herausstellen

EXPONENTIALVERTEILUNG (+ POISSON-PROZESS)

- kontinuierliche Zufallsvariable A (Wertemenge $\mathbf{R}^+$)

  mit Verteilungsfunktion :

(9.1.19a) $FA(a) = P[A \le a] = \begin{cases} 1 - \exp(-\lambda a) & a>0 \\ 0 & a \le 0 \end{cases}$

und positivem reellem Parameter

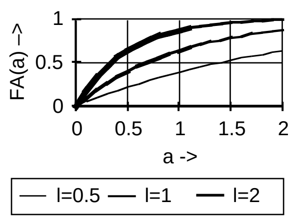
(9.1.19b) $\lambda > 0$

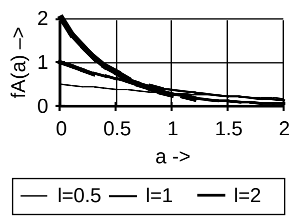
heißt (negativ) **exponentiell verteilt**

leicht ermittelbar:
Charakteristika Exponentialverteilung

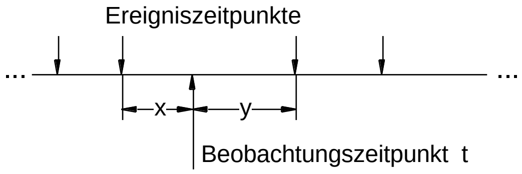
(9.1.20)		
	Dichtefunktion	$fA(a) = \lambda \cdot \exp(-\lambda a)$
	Erwartungswert	$E[A] = 1/\lambda$
	2. Moment	$E[A^2] = 2/\lambda^2$
	Varianz	$V[A] = 1/\lambda^2$
	Variationskoeffizient	$VK[A] = 1$

Erscheinungsbild





- Betrachtung Folge von Ereignissen
  - je zu **Ereigniszeitpunkt** **R**$^+$ "eintretend"
  - mit **Ereignisabständen** zwischen Folgeereignissen, welche "zufällig lang" sind, durch ZV A_k erfaßt werden, wo alle A_k u.i. (wie ZV A) verteilt, und A exponenteill verteilt mit Parameter λ
- Beobachtung des **Ereignisstroms**
  - zu zufälligem Zeitpunkt t
  - mit Interesse an Zeitspanne bis zu nächstem Ereignis



x ZE nach Ereignis:
"Rest" y ist seinerseits ZV Y
charakterisiert z.B. durch
Verteilungsfunktion FY(y)

(einfachere) Betrachtung:
komplementäre Funktion 1-FY(y) = P[Y>y]
aus Diagramm:

$$\begin{aligned} P[Y>y] &= P[A>x+y \mid A>x] \\ &= P[A>x+y] \,/\, P[A>x] \\ &= (1-FA(x+y)) \,/\, (1-FA(x)) \end{aligned}$$

$$\begin{aligned} P[Y>y] &= (1-FA(x+y)) \,/\, (1-FA(x)) \\ \text{mit (9.1.19)} \quad &= \exp(-\lambda\cdot(x+y))/\exp(-\lambda x) \\ &= \exp(-\lambda y) \end{aligned}$$

so daß

(9.1.21)
$$\begin{aligned} FY(y) &= 1-P[Y>y] \qquad y\geq 0 \\ &= 1-\exp(-\lambda y) \\ &= FA(y) \end{aligned}$$

erstaunlich !!

- kein "x" in (9.1.21):
  Länge des Restes Y unabhängig davon, wo Beobachtungszeitpunkt t im Ereignisintervall lag **??**
- FY(y) = FA(y):
  Verteilung des Restes Y identisch mit der des gesamten Ereignisintervalls A **??**

kontra-intuitiv !!

- exponentiell verteilte ZV in dem Sinne "gedächtnislos", daß sie die Dauer ihrer Vergangenheit "vergißt"

  Gedächtnislosigkeit als Bedingung:

  nur Familie der Exponentialverteilungen ist in obigem Sinn gedächtnislos (bei kontinuierlichen ZVs)

  diskretes Gegenstück (wieder exklusiv, mit analoger Interpretation): geometrische Verteilung (Wertebereich $\mathbf{N}_0$)

- weitere interessante Verteilungen:

  Annahme:
  Ereignisstrom mit u.i. exp. Ereignisabständen
  Frage:
  wie lange dauert es von Ereignis bis übernächstem, drittnächstem, k-nächstem ?
  Antwort:
  Summen von 2, 3, …, k (u.i.) exp. Ereign.Abst'n "A"
  ZV A_k k = 2, 3, ...

$$A_k = \sum_{i=1}^{k} A \qquad k=2,3,...$$

(Verteilung **nicht** gleich der von k·A)
ferner (Zeit bis zur nächsten Ankunft):
$A_1 = A$

Ermittlung A_2 (Zeit bis übernächste Ankunft)
Verteilungsfunktion

$$\begin{aligned} FA_2(a) &= P\left[A_2 \leq a\right] \\ &= \int_0^a fA_1(x)\, FA_1(a-x)\, dx \end{aligned}$$

(9.1.19, 9.1.20) eingesetzt:

(9.1.22a)
$$\begin{aligned} FA_2(a) &= \int_0^a \lambda \exp(-\lambda x) \left[1-\exp(-\lambda(a-x))\right] dx \\ &= \int_0^a \lambda \exp(-\lambda x)\, dx - \int_0^a \lambda \exp(-\lambda a)\, dx \\ &= 1-\exp(-\lambda a) - \lambda a \exp(-\lambda a) \\ FA_2(a) &= 1 - (1+\lambda a) \exp(-\lambda a) \end{aligned}$$

entsprechend für A_3:

(9.1.22b)
$$\begin{aligned} FA_3(a) &= \int_0^a fA_1(x)\, FA_2(a-x)\, dx \\ &= 1 - \left(1 + \lambda a + \frac{\lambda^2 a^2}{2}\right) \exp(-\lambda a) \end{aligned}$$

Bildungsgesetz

(9.1.22c)
$$FA_k(a) = 1 - \left[\sum_{i=0}^{k-1} \frac{(\lambda a)^i}{i!}\right] \exp(-\lambda a)$$

(9.1.22) beschreibt Familie "Erlang" -:
Erlang(k) -, E_k -**Verteilungen**
ZV entsteht als Summe von k u.i. exp. verteilten ZV,
Parameter: k, λ
weitere Charakteristika:

Dichtefunktion	$fA_k(a)$	$= \lambda \exp(-\lambda a) \dfrac{(\lambda a)^{k-1}}{(k-1)!}$
Erwartungswert	$E\left[A_k\right]$	$= k/\lambda$
zweites Moment	$E\left[A_k^2\right]$	$= k(1+k)/\lambda^2$
Varianz	$V\left[A_k\right]$	$= k/\lambda^2$
Variationskoeff.	$VK\left[A_k\right]$	$= 1/\sqrt{k}$

(9.1.22): Zeit **von Ereignis** bis zum k-tem folgenden

Zeit **von zufälligem** Beobachtungs**zeitpunkt**
bis k-tem folgenden Ereignis?
Entsprechend (9.1.22) !
denn: Rest Ereignisinterv. verteilt wie gesamtes Intervall !

inverse Betrachtung:
- von bestimmtem Beobachtungszeitpunkt an
- Zeitintervall fester Länge t
- wieviel Ankünfte in diesem Intervall ?
- "variierend": ZV N_t, Wertebereich $\mathbf{N}_0$

Verteilung erfaßbar durch Wahrscheinlichkeiten

$\{ PN_t(k);\ k\ \ \mathbf{N}_0 \}$

wo aus (9.1.22):

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

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

(9.1.23) $PN_t(k) = \dfrac{(\lambda t)^k}{k!} \exp(-\lambda t) \quad k=0,1,2,...$

Verteilungen dieser Familie:
Poisson-Verteilungen

hier: Poisson-Verteilung mit Parameterwert λt beschreibt Anzahl Ereignisse in beliebigem Intervall der Länge t für Ereignisstrom mit u.i.v. exp. (λ) Ereign.Intervallen

Unabhängig identisch exponentiell verteilte Ereignisabstände entsprechen Poisson-verteilten Ereigniszahlen und umgekehrt
"**Poisson-Strom**"

Charakteristika Poisson-Verteilung (Parameterwert λt):

- Erwartungswert:

$$\begin{aligned} E\left[N_t\right] &= \sum_{k=0}^{\infty} k\, PN_t(k) \\ &= \exp(-\lambda t) \sum_{k=1}^{\infty} k \frac{(\lambda t)^k}{k!} \\ &= \lambda t \exp(-\lambda t) \sum_{k=1}^{\infty} \frac{(\lambda t)^{k-1}}{(k-1)!} \end{aligned}$$

mit Reihendarstellung der Exponentialfunktion

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

(9.1.24) $E\left[N_t\right] = \lambda t \exp(-\lambda t) \exp(\lambda t) = \lambda t$

- spezielle Wahrscheinlichkeiten:

(9.1.25a) $PN_t(0) = \exp(-\lambda t) = 1 - \lambda t + (\lambda t)^2/2! - + ... = 1 - \lambda t + o(t)$

wo o(t), "Funktion von kleiner Ordnung gegenüber t", definiert durch Bedingung

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

für kleines t, mit hinreichender Genauigkeit:
$PN_t(0) \approx 1 - \lambda t$

(9.1.25b) $PN_t(1) = \lambda t \cdot \exp(-\lambda t) = \lambda t + o(t)$

(9.1.25c) $P[N_t > 1] = 1 - PN_t(0) - PN_t(1) = o(t)$

(9.1.25), als Bedingung gestellt,
besitzt als einzige Lösung den bekannten Poisson-Strom

folgende drei Bedingungen definieren (je für sich) einen Poisson-Strom (treffen immer gleichzeitig zu):

- Ereignisse mit u.i.v. exponentiellen Abständen, Par. λ
- Von zufälligem Zeitpunkt aus
Anzahl Ereignisse in Intervall der Länge t
entsprechend Poisson-Verteilung mit Parameter λt
- Von zufälligem Zeitpunkt aus
Wahrscheinlichkeit 0, 1, >1 Ereignisse in Interv. Länge t
entsprechend (9.1.25)

zugrunde liegt: Eigenschaft der **Gedächtnislosigkeit**

Größe λ ist zentraler Parameter Poisson-Strom,
als "Rate" des Stroms bezeichnet, aus zwei Gründen:

- "Rate": "Anzahl von Ereignissen pro Zeiteinheit"
Poisson-Strom, Intervall der Länge T:
im Mittel (9.1.24) λT Ereignisse
$\lambda T/T = \lambda$ ist Erwartungswert Ereignisrate
- (9.1.25b): Wahrscheinlichkeit,
von beliebigem Zeitpunkt aus,
innerhalb nächster t ZE Ereignis zu registrieren,
wächst bei kleinem t proportional t, Prop.Konstante λ
auch insofern: Benennung "Rate" treffend

Ende
EXPONENTIALVERTEILUNG (+ POISSON-PROZESS)

Zurück zu:
AUFSTELLUNG Q-MATRIX

- sei System (zum Zeitpunkt t) im Zustand i (Z'Raum)
- \+ zeithomogen d.h.unabhängig von t ableitbar,
wie Zustand s Zeiteinheiten später

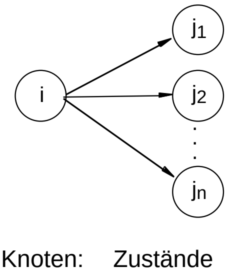
erfaßt durch	$\pi_{ij}(s)$	(9.1.12)
bzw. für		
unmittelbare Zukunft	q_{ij}	(9.1.16b)

- für unmittelbar bevorstehende Zukunft
nur unmittelbare Folgezustände relevant:

aus Zustand i –> $j_1, j_2, ..., j_n$

(ungesunde Verhältnisse "beliebig viele Z'Übergänge in beliebig kurzer Zeit"
als praxisfern ausgeschlossen)

- somit Interesse auf (Ausschnitt von)
Zustandsübergangsgraph



Knoten: Zustände
Kanten: mögliche unmittelbare Zustandsübergänge

- (bis auf weiteres) einfacher:
  genau ein unmittelbaren Folgezustand: j_1
  Zustand i zu t
  Zustand j_1 zu s+t mit (bed.)Wahrsch. $\pi_{i,j1}(s)$

- mit (gedachter) Reihenentwicklung für $\pi_{i,j1}(s)$ auch

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

$q_{i,j1}$ "Übergangsintensität", "Zustandsübergangsrate"

alle q_{ij} mit $j \neq j_1$ (bzgl. nicht unmittelbarer Folgezustände)
sollten Wert 0 haben, zugeordnete $\pi_{ij}(s)=o(s)$ sein,
Zustände $j \neq j_1$ in erster Näherung
(für kleine s de facto) nicht auftauchen

Hauptdiagonalelement q_{ii} : separat zu besprechen

- vom Problem her:
  - woher kommen Elemente $q_{ij} \neq 0$?
  - welche Werte haben diese Elemente ?

- Beispiel:

  (Modell von) Rechensystem,
  das im Moment nichts zu tun hat ("leer" ist), Zustand i
  in diesem Zustand verharrt, bis neuer Auftrag eintrifft

  einziger mögliche Folgezustand zu Zustand i
  ist Zustand nach Auftragseingang, Zustand j_1
  mit Blick auf (9.1.26) müßte gelten

  $\pi_{i,j1}(s) = q_{i,j1} \cdot s + o(s)$
  = P[bei Zustand i trifft innerhalb nächster
  s Zeiteinheiten genau ein Auftrag ein]

- unabhängig von Zeitpunkt, zu dem Zustand i festgestellt !

  … Diskussion nach (9.1.21)
  nur im Falle exponentiell verteilter
  "Zeit bis zur nächsten Ankunft" möglich !

  bei Poisson'schem Ankunftsstrom (Parameter: λ) gilt,
  P[genau eine Ankunft innerhalb s Zeiteinheiten]
  $= \lambda \cdot s + o(s)$
  fragliches $q_{i,j1}$ hat Wert λ

- (etwas) allgemeiner:
  zu Zustand i zwei unmittelbare Folgezustände (j_1 und j_2)
  Matrixelemente $q_{i,j1} \neq 0$ und $q_{i,j2} \neq 0$ gefragt (sonstige =0)

  im Problem
  - streben zwei "exponentielle Phasen" ihrem Ende zu,
    unabhängig voneinander und parallel,
  - Ende der einen Phase (Parameter: λ_1)
    resultiert (Ereignis) in Zustand j_1
    Ende der anderen Phase (Parameter λ_2) in j_2
  - wieder aus
    $\pi_{i,j1}(s) = P[\lambda_1$-Phase endet innerhalb s ZE]
    $= \lambda_1 \cdot s + o(s)$
    $\pi_{i,j1}(s) = q_{i,j1} \cdot s + o(s)$
    Matrixelement $q_{i,j1} = \lambda_1$
    analog $q_{i,j2} = \lambda_2$

- aus Beispielen:
  Elemente Q-Matrix aus (geeignetem!) Problem ablesbar,

  (zumindest Nichtdiagonal-Elemente von Q,
  Diagonalelemente –>)

- Diagonalelemente Q

  nach (9.1.12b): bedingte Überg'gs-W'n addieren sich zu 1

(9.1.27) $1-\pi_{ii}(s) = \sum_{j \neq i} \pi_{ij}(s)$ i $\mathbf{N}_0$

mit Grenzbetrachtung (9.1.16b), –> Def. Q-Elemente

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

–> Hauptdiag.Elemente = negative Summen Nichtdiag.E's
Zeilensummen Q-Matrix verschwinden

- weitere Interpretation:
  analog zu Ableitung (9.1.26), aus (9.1.27):

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

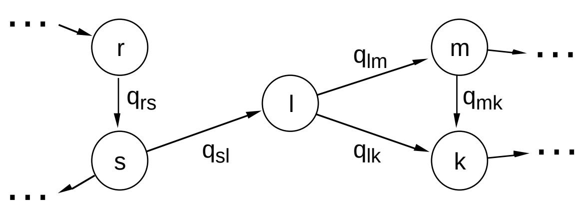
  mit (9.1.28): $= s \cdot (-q_{ii}) + o(s)$

  $\pi_{ii}(s)$ ist W., s ZE nach Antreffen i wieder i vorzufinden
  $1-\pi_{ii}(s)$ ist W., s ZE nach Antreffen i ein $j \neq i$ vorzufinden

  $-q_{ii}$ ist damit Gesamtrate (Intensität),
  mit der Zustand i seinem Ende zustrebt
  "unabhängig von t"
  "Lebenszeit" Zustand i
  kann nur exponentiell (Par.: $-q_{ii}$) sein !

- somit formulierbar:
  Technik Aufstellung Gleichungssystems (9.1.17)
  zur Ermittlung Grenzverteilung
  von homogener, irreduzibler Markoff-Kette

Startpunkt: "Graph der Zustandsübergangsraten"
Knoten: sämtliche Zustände M-Kette
Kanten: sämtl. Möglichkeiten unmittelbarer Übergänge,
bezeichnet mit zugehörigen Übergangsraten
(Intensitäten, Q-Matrix-Elementen)
z.B.



einzelne Gleichung (9.1.17) (festes i)

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

und mit (9.1.28)

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

aus Graph der (Zustands-)Übergangsraten direkt ablesbar:

- links Produkt stationäre Zustandswahrsch. Zustand i
  Summe Übergangsraten aus i heraus

- rechts Summe über Produkte aus (jeweils)
  stationäre W. Zustand j (mit: i erreichbar)
  zugehörige Übergangsrate aus j heraus

Mit Benennung "**Wahrscheinlichkeitsfluß**" (W-Fluß)
für Produkt p. · q.
dh Zustandswahrscheinlichkeit
mal Austrittsrate aus Zustand

Merkregel für Aufstellung der Gleichungen (9.1.17)

Regel 9.1.29 a: Aufstellung Gleichgewichtssystem

Für jeden Zustand ist
W-Fluß aus diesem Zustand
gleich W-Fluß in diesen Zustand

Beispiel für Zustand l:

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

aus Addition "gewisser" dieser Gleichungen
praktische Erleichterungen

mit Y Z Untermenge der Zustandsmenge:

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

–>

Regel 9.1.29 b: Aufstellung Gleichgewichtssystem

Für jede Zustands(unter-)menge ist
W-Fluß aus dieser Zustandsmenge
gleich W-Fluß in diese Zustandsmenge

(enthält keine zusätzliche Information;
führt aber gelegentlich zu einfacheren Gleichungen)

9.2 Kostenmodelle und Entscheidungsprozesse

Beschränkung auf

- zeitdiskrete Markov-Kette (DTMC)
  (X_r ; r=0,1,…) dh Parametermenge $\mathbf{N}_0$
- mit diskretem Zustandsraum
  RX = {1,2…,I} dh Zustandsram endlich, $\mathbf{N}^+$
  wo nicht abweichend gesagt
  (problemabhängig auch andere Bereiche **N**)
- Übergangsmatrix
  Π I x I - Matrix

9.2.1 Einfache Kostenmodelle

Seien mit Besuch eines Zustands i RX
Belohnungen / Bestrafungen, Erlöse / Kosten, …
verbunden in Höhe von
C(i) i {1,2…,I} unabhängige ZV
mit i-spezifischen Verteilungen

mit Erwartungswert
c(i) := E[C(i)] deterministische Kosten
trivialerweise enthalten
$\mathbf{c}$:= (c(1),…,c(I))^T Kosten-(Erwartungswert-)Vektor

Interesse liegt auf Ermittlung

- Erwartungswert Gesamt"kosten" des Prozesses
- bei endlichem / unbeschränktem "Planungshorizont"

ENDLICHER HORIZONT

- seien (naheliegenderweise)
  - Gesamtkosten Prozeß
  - bei endlichem Planunghorizont n
    := Anfangsabschnitt (X_r ; r=0,1,…,n) des Prozesses

  definiert als Summe der Besuchskosten

$$C_n := \sum_{r=0}^{n} C(X_r)$$

$$c_n := E[C_n] = \sum_{r=0}^{n} E[C(X_r)]$$

dh offensichtlich rückführbar auf Anzahl Besuche i RX

- bezeichne
  - $N_j(n)$ (ZV !)
    die Anzahl der Besuche Zustand j
    bei Planungshorizont n
  - $m_{ij}(n) := E[N_j(n) \mid X_0=i]$
    den Erwartungswert der Anzahl Besuche j
    bei Planungshorizont n
    bei Anfangszustand i
    = **Belegungszeit** j bei PH n bei AZ i
  - **M**(n) := ($m_{ij}(n)$; i,j RX) I x I - Matrix
    die **Belegungszeitmatrix**

Satz 9.2.01 : Belegungszeitmatrix + Übergangsmatrix

Es gilt

$$\mathbf{M}(n) = \sum_{r=0}^{n} \Pi^r$$

fixiere Besuchszustand j Anfangszustand i

definiere Z_r ; r=0,1,…,n ZVs
gemäß $Z_r=1$ für $X_r=j$
$Z_r=0$ für $X_r \neq j$

$$N_j(n) = \sum_{r=0}^{n} Z_r$$

$$\begin{aligned} m_{ij}(n) &= E\left[N_j(n) \mid X_0=i\right] \\ &= \sum_{r=0}^{n} E\left[Z_r \mid X_0=i\right] \\ &= \sum_{r=0}^{n} P\left[Z_r=1 \mid X_0=i\right] \\ &= \sum_{r=0}^{n} P\left[X_r=j \mid X_0=i\right] \\ &= \sum_{r=0}^{n} \pi_{ij}^{(r)} \end{aligned}$$

Behauptung

- zurück zu den Kosten

- bezeichne

$$g(i,n) := E\left[\sum_{r=0}^{n} C(X_r) \mid X_0=i\right]$$

den Erwartungswert der Gesamtkosten, bei AZ i
$\mathbf{g}(n) := (g(1,n),\dots,g(l,n))^T$
den zugehörigen Vektor

Satz 9.2.02 : Gesamtkosten + Belegungszeitmatrix
Es gilt
$\mathbf{g}(n) = \mathbf{M}(n) \cdot \mathbf{c}$

nach Definitionen ist

$$\begin{aligned} g(i,n) &= \sum_{r=0}^{n} E\left[C(X_r) \mid X_0=i\right] \\ &= \sum_{r=0}^{n} \sum_{j=1}^{l} E\left[C(X_r) \mid X_r=j\right] \; P\left[X_r=j \mid X_0=i\right] \\ &= \sum_{r=0}^{n} \sum_{j=1}^{l} c(j) \;\; \pi_{ij}^{(r)} \\ &= \sum_{j=1}^{l} \left\{ \sum_{r=0}^{n} \pi_{ij}^{(r)} \right\} \;\; c(j) \\ &= \sum_{j=1}^{l} m_{ij}(n) \;\; c(j) \end{aligned}$$

Behauptung

insgesamt:

Erwartungswert Gesamtkosten
für endl. Planungshorizont n
bei Anfangsverteilung $\mathbf{p}(0)$ deterministischer Anfangszustand trivialerw. enthalten
gegeben als

(9.2.03)
$$\begin{aligned} c_n &= \mathbf{p}^T(0) \;\; \mathbf{g}(n) \\ &= \mathbf{p}^T(0) \;\; \mathbf{M}(n) \;\; \mathbf{c} \\ &= \mathbf{p}^T(0) \;\; \sum_{r=0}^{n} \Pi^r \;\; \mathbf{c} \end{aligned}$$

ist also aus
Anfangsverteilung
Übergangsmatrix
Kostenvektor
berechenbar ("nur noch Rechnerei")

UNBESCHRÄNKTER HORIZONT

- Gesamtkosten Prozeß
bei endlichem Planunghorizont n
definiert als Summe der Besuchskosten

$$C_n := \sum_{r=0}^{n} C(X_r)$$

$$c_n := E[C_n] = \sum_{r=0}^{n} E[C(X_r)]$$

wachsen idR mit wachsendem n über alle Grenzen
wie auch zu seiner Ableitung verwendete

$$g(i,n) := E\left[\sum_{r=0}^{n} C(X_r) \mid X_0=i\right]$$

- Ausweg "Diskontierung"

hier üblicherweise aber Betrachtung
der "langfristigen Kosten pro Zeiteinheit", also der

$$g(i) := \lim_{n\to\infty} \frac{g(i,n)}{n+1}$$

und daraus abgeleiteter Größen,
falls existent

und (wieder) zu erahnen:
Zusammenhang mit Anzahl Besuchen in i RX

- bezeichne

  - $N_j(n)$ (wieder) die Anzahl der Besuche Zustand j bei Planungshorizont n

  - $m_j := \lim_{n\to\infty} \dfrac{E[\, N_j(n) \,]}{n+1}$

    den Grenzwert des relativ. Erwartungswertes von $N_j(n)$
    = langfristige **Belegung** von Zustand j
    ("falls existent", hier und idF)

  - m_j ist langfristiger relativer Anteil von Zeitpunkten (in allen Zeitpunkten)
    zu denen Zustand j vorliegt

    bei (implizit)
    "Dauer des Einnehmens eines Zustandes" = 1 ZE
    der langfristige relative Zeitanteil mit Zustand j

    $m_j = p_j$ (aus Gleichgewichtsverteilung Satz 9.1.09)

    existiert, wenn Gleichgewichtsverteilung existiert

- zurück zu den Kosten

- bezeichne (im Sinne der Vorüberlegungen)
$\mathbf{m} := (m_1,\dots,m_l)^T$
den Vektor der langfristigen Belegungen, wegen
$\mathbf{m} = \mathbf{p}$ (Gleichgewichtsverteilung Satz 9.1.09)
auch (langfristige) **Belegungsverteilung** genannt

Satz 9.2.04 : Gesamtkosten + Belegungsverteilung
Unter den Gleichgewichtsbedingungen Satz 9.1.09
(irreduzible, aperiodische, rekurrente Markov-Kette)
gilt
- für den Erwartungswert der langfristigen Kosten,
je Zeiteinheit, bei Anfangszustand i
$g(i) = \mathbf{m}^T \cdot \mathbf{c} = \mathbf{p}^T \cdot \mathbf{c}$

- sowie, wegen Unabhängigkeit dieses Ausdrucks von i,
für den Erwartungswert der langfristigen Kosten,
je Zeiteinheit, unabhängig vom Anfangszustand
$g \equiv g(i) = \mathbf{m}^T \cdot \mathbf{c} = \mathbf{p}^T \cdot \mathbf{c}$

nach Definitionen, und unter Verwendung von Satz 9.2.02:

$$\begin{aligned} g(i) &= \lim_{n\to\infty} \frac{g(i,n)}{n+1} \\ &= \lim_{n\to\infty} \frac{1}{n+1} \sum_{j=1}^{l} m_{ij}(n) \;\; c(j) \\ &= \sum_{j=1}^{l} \left[\lim_{n\to\infty} \frac{m_{ij}(n)}{n+1} \right] \;\; c(j) \\ &= \sum_{j=1}^{l} m_j \;\; c(j) \quad \left(= \sum_{j=1}^{l} p_j \;\; c(j) \right) \end{aligned}$$

Behauptung

9.2.2 Entscheidungsprozesse

Kostenmodelle erlaubten

- Berechnung von Gesamtkosten bei endlichem Planungshorizont — $\mathbf{g}(n)$ n (E[…] !)
- Berechnung von langfristigen Kosten je ZE — g
- bei stochastischer Prozeßentwicklung erfaßt durch Übergangsmatrix — Π

Möglichkeiten zur "Einwirkung"
(Entscheidungen / Politik)
bisher nicht vorgesehen

Erweiterungen der Prozesse

- zeitdiskrete Markov-Kette (DTMC)
  $(X_r\ ;\ r=0,1,\dots)$ wie gehabt
- mit diskretem endlichem Zustandsraum
  $RX = \{1,2\dots,I\}$ wie gehabt
- mit diskretem endlichem **Aktionsraum**
  $RA = \{1,2\dots,K\}$ **neu**
  (Aktionsraum statt Steuerbereich, Benenn'g a. statt u. im vorliegenden Kontext üblich)

  wo konkrete Aktion A ZV mit (bedingter) Verteilung, charakterisiert mittels (bedingter) Wahrscheinlichkeiten
  $f(i,a)$ i RX, a RA
  $\Sigma_{a\ \ RA} f(i,a) = 1$ i RX
  also ("Markovsch") allein v. Zust'd i abhängig, nicht von …

  deterministische Aktionswahl trivialerweise enthalten

- Prozeßentwicklungen charakterisiert durch (bedingte) (Ein-Schritt-) Übergangswahrscheinlichkeiten

  $\pi_{ij}(a) := P[X_{r+1}=j \mid X_r=i,\ A_r=a]$ i,j RX, a RA, r=0,1,…

  also ("Markovsch") allein v. Zust'd i abhängig, nicht von …

  erfaßt durch **Menge** von Übergangsmatrizen
  $\{ \Pi(a)\ ;\ a\ \ RA \}$
  mit
  $\Pi(a) := (\pi_{ij}(a)\ ;\ i,j\ \ RX)$ I x I - Matrizen
  Erweiterung
  (Menge $\Pi(a)$'s anstelle einer Π)

Prozeß
$\{ (X_r,A_r)\ ;\ r\geq 0 \}$
heißt **zeitdiskreter Markovscher Entscheidungsprozeß**
(Discrete-Time Markov Decision Process, DTMDP)

Funktion
$f = \{ f(i,a);\ i\ \ RX,\ a\ \ RA \}$
spielt die Rolle einer **Politik** Einwirkungsmöglichkeit auf Prozeßablauf

Politikdefinition insbesondere
- unabhängig von (Zeit-)Punkt r
- iS Kap. 8: Politik periodenunabhängig,
  Politik **stationär**

- auch Kostenmodell zu erweitern:
  mit Besuch Zustands i RX, bei Wahl Aktion a RA
  Kosten, … entstehend in Höhe von
  $C(i,a)$ i RX, a RA unabhängige ZV mit (i,a)-spez. Vert'gen
  mit Erwartungswert
  $c(i,a) := E[C(i,a)]$ deterministische Kosten trivialerweise enthalten
  zusammengefaßt zu **Kostenvektoren**
  $\mathbf{c}(a) := (c(1,a),\dots,c(I,a))^T$ a RA
  bzw **Kostenmatrix**
  $\mathbf{C} := (c(i,a)\ ;\ i\ \ RX,\ a\ \ RA)$ I x K - Matrix
  Erweiterung
  (Menge **c**(a)'s bzw Matrix **C** anstelle eines **c**)

obzwar notationsaufwendiger,
und ziemlich allgemein (Haupteinschränkung: Markovsch)

- präsentiert definierte DTMDP-Klasse
  nichts "wirklich Neues" - s.Satz 9.2.05 -
- erlaubt aber über die Auszeichnung einer "Politik" Einwirkung auf den Prozeßablauf
- und damit, zusammen mit der Kostendefinition,
  - Politik"variationen" → Gesamtkosten"veränderungen"

    Optimalitätsdefinitionen, Optimierungstechniken

Satz 9.2.05 : DTMDP ist DTMC

Sei $\{ (X_r,A_r)\ ;\ r\geq 0 \}$ zeitdiskreter Markovscher Entscheidungsprozeß
mit $RX = \{1,2\dots,I\}$ endlichem Zustandsraum (Erweiter'g → abzählbar möglich)
$RA = \{1,2\dots,K\}$ endlichem Aktionsraum (Endlichkeit wesentlich)
$\{\Pi(a)\ ;\ a\ \ RA\}$ Menge Übergangsmatrizen
und Politik
$f = \{ f(i,a);\ i\ \ RX,\ a\ \ RA \}$

dann ist der Prozeß
$\{ X_r\ ;\ r\geq 0 \}$
eine zeitdiskrete Markov-Kette
mit Übergangsmatrix
$\Pi^f := (\pi^f_{ij}\ ;\ i,j\ \ RX)$
wo $\pi^f_{ij} = P[X_{r+1}=j \mid X_r=i] = \sum_{a\ \ DA} f(i,a)\ \ \pi_{ij}(a)$

aus Definitionen:

$$P[X_{r+1}=j \mid X_r=i] = \sum_{a\ \ RA} P[X_{r+1}=j \mid X_r=i,\ A_r=a]\ \ P[A_r=a \mid X_r=i] = \sum_{a\ \ RA} f(i,a)\ \ \pi_{ij}(a)$$

(iS Kap. 8: stationäre Politik
führt zu: zeithomogenem Prozeß)

Unter Zuhilfenahme der Ergebnisse aus Abschn. 9.2.1
lassen sich auf dieser Basis Kostenmaße

- Erwartungswert Gesamtkosten, endl. Planungshorizont n, bei Anfangsverteilung **p**(0)
- Erwartungswert langfristige Kosten je Zeiteinheit

ermitteln

Konzentration auf (einfacheren) Langfristfall, auf
langfristige Kostenrate, bei Politik f, G^f

$$G^f := \lim_{n\to\infty} E\left[\frac{\sum_{r=1}^{n} c(X_r, A_r)}{n+1}\right]$$

unter der Annahme,
daß die (dem DTMDP zugeordnete)
DTMC irreduzibel, aperiodisch, rekurrent

so daß DTMC-Grenzverteilung existiert
und (folglich) unabhängig vom Anfangszustand ist

(? so daß auch langfristige Kostenrate existiert
und unabhängig vom Anfangszustand ist ?)

Satz 9.2.06 : Langfristige Kostenrate DTMDP

Sei	$\{ (X_r, A_r) ; r \geq 0 \}$	zeitdiskreter Markovscher Entscheidungsprozeß
mit	RA = {1,2…,K}	Aktionsraum
	$\{\Pi(a) ; a \;\; RA\}$	Übergangsmatrizen
und	f	Politik

sei ferner

	$\{ X_r ; r \geq 0 \}$	induzierte zeitdiskrete M-Kette irreduzibel, aperiodisch, rekurrent
mit	**p**	Belegungsverteilung

dann ist langfristige Kostenrate G^f,
unabhängig vom Anfangszustand,
gegeben durch

$$G^f = \sum_{i \;\; RX} \left\{ p_i \sum_{a \;\; RA} f(i,a) \;\; c(i,a) \right\}$$

Erwartungswert der Kosten bei Besuch Zustand i ist

$$c(i) = \sum_{a \;\; RA} f(i,a) \;\; c(i,a)$$

mit Satz 9.2.04 ergibt sich Behauptung

9.2.3 Optimierung von Entscheidungsprozessen
Ausschnitt

Politik f von DTMDPs war konstruiert
zu Zwecken der Einflußnahme auf Prozeßablauf

Optimierungsaufgabe offensichtlich

min	G^f	(bestimme bestes f)
udN	"DTMDP"	(Prozeß analysierbar ioS)

Variation f bedeutet
Variation charakterisierenden Parameter von
$f = \{ f(i,a);\ i \;\; RX,\ a \;\; RA \}$
also Variation der f(i,a)

bei	|RX| = I,	
	|RA| = K	(wie angenommen)

prinzipiell
$I \cdot K$ Entscheidungsvariablen
mit Nebenbedingungen
$\{ f(i,a);\ i \;\; RX \},\ a \;\; RA$ sind K Verteilungen

- Formalisierung unter Einführung der Größen

  $x_{ia} := p_i \cdot f(i,a) \qquad i \;\; RX,\ a \;\; RA$

  "künstlich", aber interpretierbar:
  langfristiger relativer Anteil der Zeitpunkte,
  zu denen Zustand i herrscht **und** Aktion a gewählt

- mit diesen Größen erhält man aus

  $\Sigma_{a \;\; RA} f(i,a) = 1$ (Verteilung)

  auch

  $p_i = \Sigma_{a \;\; RA}\, p_i\, f(i,a) = \Sigma_{a \;\; RA}\, x_{ia}$

  und durch Rück-Einsetzen

  $$f(i,a) = \frac{x_{ia}}{\sum_{b \;\; RA} x_{ib}}$$

- Zielfunktion G^f läßt sich in Größen x_{ia} umschreiben

  $$G^f = \sum_{i \;\; RX} \left\{ p_i \sum_{a \;\; RA} f(i,a) \;\; c(i,a) \right\}$$
  $$= \sum_{i \;\; RX} \sum_{a \;\; RA} p_i \;\; f(i,a) \;\; c(i,a)$$
  $$= \sum_{i \;\; RX} \sum_{a \;\; RA} x_{ia} \;\; c(i,a)$$

  als lineare Funktion der x_{ia}

- **p** ist (f-abhängige) Gleichgewichtsverteilung, deren Komponenten

  $p_j = \Sigma_{i \;\; RX}\, p_i\, \pi^f_{ij} \qquad j \;\; RX$

  $\Sigma_{j \;\; RX}\, p_j = 1$

  zu wahren haben (Nebenbedingungen!)
- Nebenbeding'n lassen sich in Größen x_{ia} umschreiben

  $$\sum_{a \;\; RA} x_{ja} = \sum_{i \;\; RX} \left\{ p_i \sum_{a \;\; RA} f(i,a) \;\; \pi_{ij}(a) \right\}$$
  $$= \sum_{i \;\; RX} \sum_{a \;\; RA} p_i \;\; f(i,a) \;\; \pi_{ij}(a)$$
  $$= \sum_{i \;\; RX} \sum_{a \;\; RA} x_{ia} \;\; \pi_{ij}(a) \qquad j \;\; RX$$
  $$\sum_{i \;\; RX} \sum_{a \;\; RA} x_{ia} = 1$$

  als lineare Gleichungen in den x_{ia}

- als formalisiertes Optimier'gsprobl. ergibt sich insgesamt

(9.2.07) min $\sum_{i\ RX} \sum_{a\ RA} x_{ia}\ c(i,a)$

udN $\sum_{a\ RA} x_{ja} = \sum_{i\ RX} \sum_{a\ RA} x_{ia}\ \pi_{ij}(a) \qquad j\ \ DX$

$\sum_{i\ RX} \sum_{a\ RA} x_{ia} = 1$

$x_{ia} \geq 0 \qquad i\ \ RX,\ a\ \ RA$

ein lineares Optimierungsproblem

- Lösungsmethodik bekannt,
  hier allerdings:

  - alles Nebenbedingungs**gleichungen**
    Standardumwandlung Kap.2
    1 Gleichung → Paar von ≤, ≥-Ungleichungenn
    führt zu linearen Anhängigkeiten
    degeneriertem Problem

  - Dimensionalität des Problems idR hoch
    I · K Entscheidungsvariablen
    I + 1 funktionale Nebenbedingungen
    (1 Redundanz enthalten → I Nebenbedggn)

  für Standard LP-"Pakete" dennoch idR unproblematisch

- Lösungsinterpretation
  (ggf inkl Überprüfung Existenzbedingungen)
  aber erforderlich

  (Teil-)Skizze:

  Lösung LP-Problem liefert
  G* optimaler Zielfunktionswert f. Politik f*
  x^*_{ia} zugehörige Variablenwerte

- bezeichne

  $T := \{ i\ \ RX : \Sigma_{a\ \ RA}\ x^*_{ia} > 0 \}$

- falls T = RX (mit $x_{ia} = p_i \cdot f(i,a)$:
  alle Zustände
  mit Gleichgew-Wahrsch. p_i>0
  haben Aktion(en) zugeordnet)

  ist optimale Politik f* definiert als

  $$f^*(i,a) = \frac{x^*_{ia}}{\sum_{b\ \ RA} x^*_{ib}}$$

  wo sich weitergehend zeigen läßt, daß
  für festes i genau ein x*(i,a)>0
  $x^*_{ia} > 0$ a=a*(i) ausgezeichnete Aktion
  $x^*_{ia} = 0$ a≠a*(i)

  und folglich
  f*(i,a) = 1 a=a*(i)
  f*(i,a) = 0 a≠a*(i)

  dh deterministische Politik liefert (bereits) Optimum,
  stochastische Politiken führen nicht zu Verbesserung

- falls T ≠ RX

  schwierigerer Fall, nicht von vornherein "ergebnislos",
  aber detaillierter zu untersuchen (hier nicht diskutiert)

hier: Abbruch

war: "Ausschnitt"

nicht betrachtet:

- kontinuierliche Zeit
- kontinuierliche Zustände
- andere Techniken (Politik-Iteration, …)
- "bias" der langfristigen Kosten (relative Werte, …)
- …

9.3 Wartesysteme

Erinnerung Abschn. 6.3.2. Scheduling-Probleme:

Bearbeitung
- von **Aufträge**
- durch **Maschinen**

Interesse an **Schedules**: Zuordnungen
- von Arbeits-(Zeit-)intervallen
- auf Maschinen (Stationen)
- an Jobs / Tasks (Aufträge / Teilaufträge)

mit Zielrichtung Optimierung von Schedules
hinsichtlich Ziel-(Kosten-)Funktionen wie
- Abschlußzeitpunkten
- Durchlaufzeiten
- Verspätungen
- …

getrennt diskutiert nach
- Ein-Maschinen-Problemen
- Mehr-Maschinen-Problemen
  (mehrere parallele, Flowshop, Jobshop)

immer aber
"unter im Voraus deterministisch festliegenden Informationen"

hier (Einstieg in)
- "im Voraus nur stochastisch festlegbaren Informationen"

bzw (bewußt)
- "nur stochastisch festgelegte Informationen"

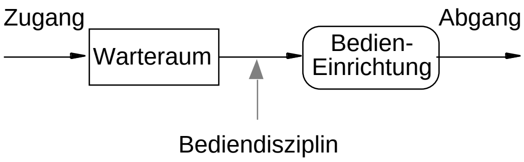
9.3.1 Einzelstationen

Ziel:
- konkrete Stationen untersuchen (Stationen "in Isolation")
- Analysetechniken des Abschn. 9.1.4 einsetzen
  –> stationäre Grenzverteilungen der Stationen
- Basis legen für "Vernetzung" von Stationen + Analyse
  (hier nur kursorische Ergebnisse, Abschn. 9.3.2)

Einzelstation:

- Strukturell:
  - Bedieneinrichtung   ein oder mehrere
    gleichartige Bediener
  - Warteraum   räumlich unbeschränkt
  - Bediendisziplin   regelt Bedienreihenfolge

    sorgt für Erledigung genau der übertragenen Aufträge
    ohne Arbeitsverweigerung / Pausen / …
    =: **arbeitserhaltend**



- Belastung:
  - Strom von "Kunden" mit **Ankunftsabständen** A_k
    unabh. identisch verteilt
    wie ZV A
  - **Bedienwünsche** Kunden S_k
    (ausgedrückt in Bedienzeiten)
    unabh. identisch verteilt
    wie ZV S

erstes / einfachstes System dieser Art:
- genau ein Bediener
- FCFS-Disziplin
- exponentiell (Parameter: λ) verteilte Ankunftsabstände
- exponentiell (Parameter: μ) verteilte Bedienwünsche

mit Kurzbezeichnung ("Kendall-Notation"):

M / M / 1 - FCFS

Bediendisziplin

1 Bediener

Charakterisierung der
Bedienwunschverteilung

Charakterisierung der
Zwischenankunftszeitverteilung
("M" von "Markoff´sch": exponentiell)

Betrachtung in kontinuierlicher Zeit
Charakterisierung Zustand:
n   Zahl anwesender Kunden

Frage: Ist $(N_t)_{t \; \mathbf{R}^+}$ zeitkontinuierliche, homogene,
irreduzible Markoff-Kette,
sodaß (mit Satz 9.1.18) stationäre Grenzverteil'g ermittelbar ?

im einzelnen:

- Zeitkontinuierlichkeit (alle Zeitpunkte t $\mathbf{R}^+$) vorausgesetzt

- stochastische Kette (diskreter Zustandsraum) gegeben,
  Zustandsraum: $\mathbf{N}_0$ (n=0, "leeres System",
  + n=1,2,... sind Systemzustände,
  ohne Grenze: unbeschr. W.raum)

- K. irreduzibel: von beliebigem Zustand n_1 aus
  jeder andere Zustand n_2 durch
  geeignete Anzahl Ankünfte/Abgänge

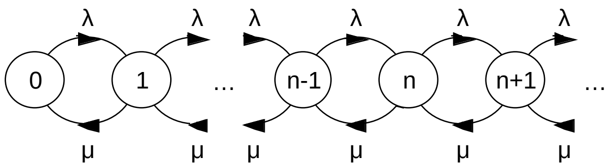
- K. Markoff'sch: bei Kenntnis Zustand n zu t
  weitere Entwicklung festgelegt
  (ohne weitere Kenntnis Vorgeschichte)

  Erläuterung:
  - zu t "läuft" (unabh. von n) **Ank.Intervall**
  - hat exponentielle (Par. λ) "Zukunft"
  - endet in (t,t+s) m. Wahrscheinl. λ s+o(s)
  - führt von Zustand n in Folgezustand n+1
  - Übergangsrate n –> n+1
    $q_{n,n+1}=\lambda$ für alle n=0,1,2,...

  - in Zuständen n≥1 "läuft" **Bed.-Intervall**
  - hat exponentielle (Par. μ) "Zukunft"
  - endet in (t,t+s) m. Wahrscheinl. μ·s+o(s)
  - führt von Zustand n in Folgezustand n-1
  - Übergangsrate n (>0) –> n-1
    $q_{n,n-1}=\mu$ für alle n=1,2,...

- K. homogen: Feststellungen unabhängig von "t"

=> Frage zufriedenstellend beantwortet !
Elemente der Q-Matrix bereits ermittelt !
(+ Kenngrößen Übergangsratendiagramm)



Zustandsübergangsratendiagramm enthält
- als Knoten alle Zustände
- als Kanten alle unmittelbar möglichen Zustandsübergänge
- als Kantenbewertung die zugehörigen Übergangsraten

Gleichungssystem (9.1.17)   (–> stationäre Verteilung)
mittels Regel 9.1.29a direkt ablesbar:

für Zustand 0: $p_0 \cdot \lambda = p_1 \cdot \mu$
für Zustand 1: $p_1 \cdot (\mu+\lambda) = p_0 \cdot \lambda + p_2 \cdot \mu$
…
für Zustand n: $p_n \cdot (\mu+\lambda) = p_{n-1} \cdot \lambda + p_{n+1} \cdot \mu$

insgesamt:

(9.3.01a) $p_0 \cdot \lambda = p_1 \cdot \mu$
$p_n \cdot (\mu+\lambda) = p_{n-1} \cdot \lambda + p_{n+1} \cdot \mu$ n=1,2,…

dazu (9.1.17b):

(9.3.01b) $\sum_{n \; \mathbf{N}_0} p_n = 1$

explizite **Lösung Gleichungssystem** (hier) leicht möglich:
- (9.3.01), leicht umgeschrieben
  $p_1 \cdot \mu - p_0 \cdot \lambda = 0$
  $p_n \cdot \mu - p_{n-1} \cdot \lambda = p_{n+1} \cdot \mu - p_n \cdot \lambda \qquad n=1,2,...$
- durch iteratives Einsetzen zunächst
  $p_n \cdot \mu - p_{n-1} \cdot \lambda = 0 \qquad n=1,2,...$
- woraus, wieder durch iteratives Einsetzen
  $p_n = (\lambda / \mu)^n \cdot p_0$

mit Abkürzung
$\rho = \lambda/\mu$
auch: $p_n = \rho^n \cdot p_0 \qquad n=0,1,2,...$

unbekanntes p_0 aus (9.3.01b)

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

unter Voraussetzung (Konvergenz der Summe!): $\rho<1$
$p_0/(1-\rho) = 1$

(9.3.02a) $p_0 = 1-\rho$

Insgesamt **Lösung** für M/M/1-FCFS-Station

(9.3.02b) $p_n = \rho^n \cdot (1-\rho) \qquad n=0,1,2,...$

- **geometrische Verteilung** für Zustandsvariable N (Anzahl in stationärer Phase anwesender Kunden)
- Erwartungswert N ("mittlere Kundenzahl"):

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

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

(9.3.02c) $E[N] = \rho/(1-\rho)$

Größe ρ
- bisher nur abkürzende Bezeichnung
- besitzt interessante Interpretationen
  zentrale Kenngröße des Systems !

$1/E[A] = \lambda$ mittlere Zahl je ZE Eintreffender
Ankunftsrate
$E[S] = 1/\mu$ im Mittel je Ankunft zu leistende Arbeit

Interpretation:
$\rho = \lambda/\mu$ im Mittel pro ZE eingebrachte
zu leistende Arbeit

"arbeitserhaltende Station" weitere Interpretation:
ρ im Mittel pro ZE von Station
geleistete Arbeit

"Bedienwünsche direkt als Bedienzeiten"
ρ mittl. Tätigkeitszeit Bediener je ZE
Auslastung der Station

- Auslastung immer: **< 1**
- war schon Voraussetzung für "Verteilung"
- physikalische Interpretation hier geliefert:
  wenn pro ZE mehr Arbeit ins System als leistbar,
  kann Warteschlange nur wachsen,
  kann stationäre Phase nicht existieren

weiter mit ρ als Auslastung
$1-\rho$ mittl. **Leerlaufzeit** Bediener pro ZE:
untätig, kein Kunde anwesend
$= p_0$ in statistischer Interpretation

$p_0=1-\rho$ war in (9.3.02a) bereits analytisch abgeleitet

Interpretationen
- für konkretes M/M/1-FCFS-System
- aber weit **allgemeinere Gültigkeit**

• arbeitserhaltende Station,
  gespeist von Kundenstrom mit
  u.i.v. Ank.abständen A, Ankunftsrate λ
  u.i.v. Bedienzeiten S, mittlere Bedienzeit $1/\mu$

• impliziert vorstehende Aussagen:

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

$\rho<1$ wesentliche Voraussetz'g für "stabilen Betrieb"

ρ im Mittel pro ZE durch Station geleistete Arbeit,
Auslastung Station (mittl. Tätigkeitszeit je ZE),

$1-\rho$ mittlere Leerlaufzeit Station pro Zeiteinheit
$\rho+p_0=1$ offensichtlich

---

für "derartige" Systeme (weitere Resultate "später")
- stationäre Verteilung der **Population** (Zahl Anwesender)
- implizierte Größen Auslastung, Leerlaufzeit

mit Techniken aus Abschn. 9.1.4 unschwer ermittelbar

andere Beurteilungsgrößen
- zB (Kunden-)**Verweilzeit** V (von **Ankunft** bis **Abgang**)

nicht unmittelbar zugänglich

---

sehr hilfreiches Resultat liefert

Satz 9.3.03 : Little's Theorem

Für die stationäre Phase eines Wartesystems mit
- mittlerem Ankunftsintervall E[A]
- mittlerer Population E[N]

gilt für die mittlere Kundenverweilzeit E[V]

$E[V] = E[A] \cdot E[N]$

betrachte eine Realisierung diverser Prozesse
- Ankunftsprozeß
- Abgangsprozeß
- Populationsprozeß

in einem Ausschnitt (Zeitintervall) der Länge T
aus der stationären Phase eines Wartesystems

Ausschnitt so gewählt, daß er
- bei leerem System (Population = 0) beginnt
- bei leerem System endet
  (Folge von) **busy periods** voll umfaßt

(solche Zeitpunkte, mit endlichem Zeitabstand, existieren
wegen $\rho<1$ $p_0>0$)

bezeichne mit

ZA(T) Anzahl der Ankünfte in diesem Intervall

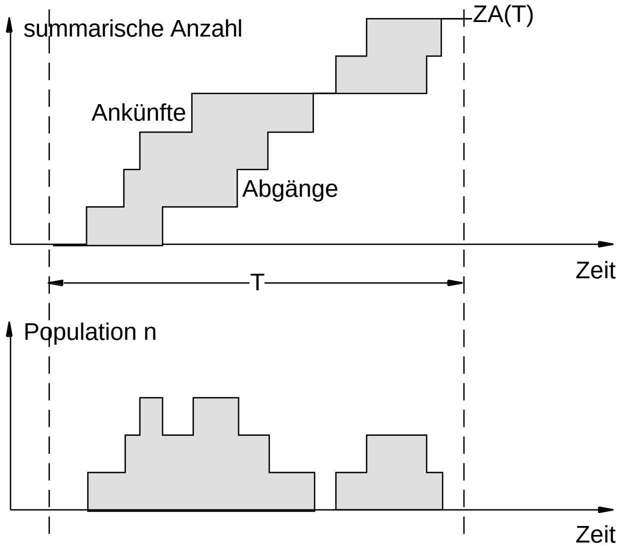
J(T) Gesamtzeit, welche Kunden "am System" sind
(s. schraffierte Flächen)

MA(T) := T / ZA(T) mittl. Ankunftsabstand im Interv.

MV(T) := J(T) / ZA(T) mittlere Verweilzeit im Intervall

MN(T) := J(T) / T mittlere Population im Intervall

Veranschaulichung:



aus Definitionen

$$MN(T) = J(T) / T = \{MV(T) \cdot ZA(T)\} / \{ZA(T) \cdot MA(T)\} = MV(T) / MA(T)$$

bei Existenz der Grenzwerte

$$\lim_{T \to \infty} MA(T) = E[A]$$

$$\lim_{T \to \infty} MV(T) = E[V]$$

existiert auch

$$\lim_{T \to \infty} MN(T) = E[N]$$

Behauptung
$E[N] = E[V] / E[A]$

---

mit Little folgt für "unsere" M/M/1-FCFS Station
als mittlere Kundenverweilzeit

$E[N] = E[V]/E[A] = E[V] \cdot \lambda$

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

ähnlich einfach: M/M/1 unter div anderen Bediendisziplinen

Satz 9.3.05 : M/M/1-Stationen + Bediendisziplinen

Stationäre Verteilung Kundenpopulation N
der M/M/1-Station unter Bediendisziplinen

(a) RANDOM: Auswahl aus Warteschlange zufällig
(b) LCFS: Last Come First Served (vollständig)
(c) LCFS-PR: LCFS-Preemptive Resume (unterbrechend)

ist identisch Populationsverteilung FCFS-Disziplin (9.3.02)

siehe Ableitung FCFS

damit ebenfalls identisch: mittlere Verweilzeit (9.3.04)
(**nicht** identisch aber **Verteilung** der Verweilzeit V **!!**)

"skizzenhaft" weiter
(Genaueres: "Leistungsbewertung von R+K-Systemen")

Mit denselben Techniken lassen sich erzielen

- allgemeine Ergebnisse für Stationen
  mit **zustandsabhängiger Arbeitskapazität**
  (bei Anwesenheit i Kunden g_i Arbeitseinh. je ZE geleistet)
  und damit die Spezialfälle
  - M/M/m-FCFS, -RANDOM, -LCFS, -LCFS-PR
  - M/M/∞ "Infinite Server" (IS), "pure delay"
  - M/M/1-PS "Processor Sharing"(PS)
  - M/M/1-PS "Processor Sharing"(PS)
- allgemeine Ergebnisse für
  nicht-exponentielle Bedienzeiten
  (Erlang-, Phasen-Verteilungen: "praktisch alle") für
  - RANDOM-, LCFS-, LCFS-PR-Disziplinen
    auch für zustandsabhängige Arbeitskapazitäten
  - **nicht** für FCFS !
- Erweiterungen obiger Ergebnisse für
  mehrere Kunden-Klassen: mehrere Ankunftsströme mit
  - spezifischen Ankunftsabständen
  - spezifischen Bedienzeiten
- gewisse Erweiterungen
  für **nicht-exponentielle Ankünfte**

9.3.2 Stationsnetze
Grobskizze

- bzgl Abschn. 6.3.2. (Scheduling-Probl.) bisher betrachtet:
  - Ein-Maschinen-Probleme
  - implizit Mehr-Maschinen-Probleme
    mit mehreren parallelen Maschinen
- Mehr-Maschinen-Probleme
  der Flowshop / Jobshop-Typen
  waren charakterisiert durch
  - mehrere Maschinen
  - Jobs, in Tasks unterteilt,
    mit vorgeschriebener Zuordnung Task / Maschine
    und bekannter Bearbeitungsdauer,
    in Varianten
    ° Flowshop: Task-Reihenfolge für alle Jobs identisch
    ° Jobshop: Task-Reihenfolge nicht notwendig identisch
- hier erfaßt durch ("Modellwelt", "Paradigma"):
  Verkehrsnetze

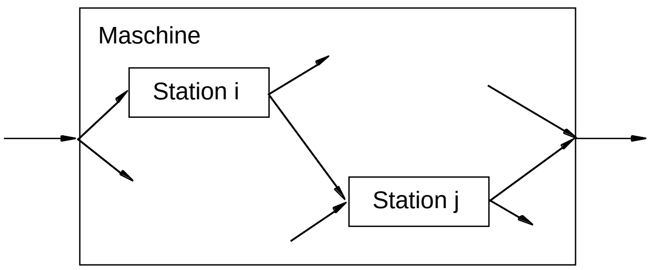
VERKEHRSNETZE

zwei Komponenten: "**Maschine**" und "**Last**"

- Maschine (zeitlich fester) gerichteter Graph
- Last (oft zeitlich variierende) Menge von Prozessen

- **Maschine** besteht aus
  - Menge von "Betriebsmitteln" ("Ressourcen", "Funktionseinheiten", "Stationen")
  - Menge von "Übergangsmöglichkeiten" ("Verbindungen", "Wegen") zwischen Paaren BMs
  - graphisch darstellbar als Netz aus Stationen

Umgebung



Betriebsmittel: "Station i" , "Station j" , …
+ Übergangsmöglichkeiten
- zwischen Stationen,
- aus Maschine (von BMs) in Umgeb'g ("Rest der Welt")
- aus Umgebung in Maschine (zu BMs)

- **Last** besteht aus
  - Menge von "Prozessen" ("Lasteinh.", "Aufträgen", "Tasks", "Jobs", "Kunden")
  - Prozeß (statisch)
    Menge von BM-Anforderungen (je für Einzel-BM)
    + Reihenfolgevorschrift (zeitl. Ordnung über BM-Anf.)
    (BM-Anforderung, Reihenfolgevorschrift zu detaillieren)

- Statische Charakterisierung Prozeß ist Verhaltensvorschrift ("**Prozeßmuster**")

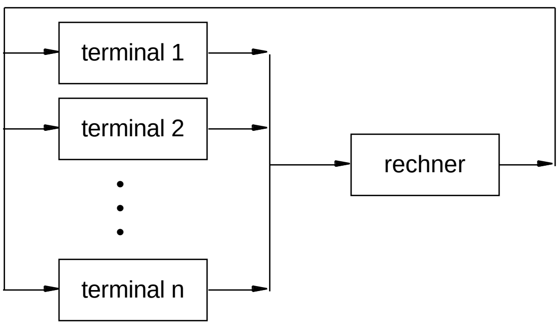
- **Prozeß** (dynamisch) entsteht, wenn
  - Maschine Prozeßmuster als Arbeitsauftrag versteht,
  - BM-Anf'g nach BM-Anf'g in geforderter Folge erfüllt

- "Entstehung" Prozeß (in dieser Modellwelt)
  - Umgebung generiert Prozeß,
    überantwortet ihn der Maschine - jetzt existent
    ("Produktionsauftrag", "Bestellung", "Benutzer")
    Prozeß wird irgendwann terminieren
    Prozeß "**temporär**"
    System/Modell "**offen**"
  - Prozeß weder generiert noch terminiert,
    existiert "dauernd" (Reihenfolge irgendwie zyklisch)
    Prozeß "**permanent**"
    System/ Modell "**geschlossen**"

Zur Veranschaulichung:

Verkehrsnetze häufig - ohne genauere Erklärung - eingesetzt in Bereichen
- Produktion, Logistik, …
- Rechnerarchitektur, Betriebssysteme, Kommunikation, …

Beispiel Teilnehmerrechensystem
- Maschine 1 Rechner
  n Terminals
- Last (an jedem Terminal: Benutzer)
  bereitet Auftrag vor,
  beauftragt Rechner → Rechner arbeitet, antwortet
  … von vorne



n+1 Stationen: Endgeräte terminal i (i=1,2,...,n) + rechner
+ Übergangsmöglichkeiten:
terminal i nach rechner,
rechner nach terminal i (i=1,2,...,n)
Modell ist geschlossen
Stationen sind formal zu charakterisieren

Geschilderte Verhaltensregelmäßigkeiten
(je Benutzer terminal - rechner - terminal - …)
sind "Prozeßmuster"
Prozeßmuster sind formal zu erfassen

Für spezifische Klassen stochastischer Verkehrsnetze
- (pseudo-)explizite analytische Ergebnisse erzielbar
- effiziente Berechnungsalgorithmen bekannt

Entwicklungshistorie in zwei (großen) Stufen
- "exponentielle" Netze (Jackson + Gordon/Newell)
- "separable" Netze (Baskett/Chandy/Muntz/Palacious)

Definition 9.3.06 : Exponentielle Netze

(a) Spezifikation Maschine
- M Stationen; Stationsmenge I = {1,2,...,M} (uU Umgebung als Pseudo-Station "0")
- Stationen unbegrenzter (räumlicher) Kapazität
- Einzelstation i charakterisiert durch
  "Geschwindigkeitsvektor" $\underline{g}_i = (g_i(1), g_i(2),...)$
  $g_i(n)$: zeitliche Gesamtarbeitskapazität Station bei Anwesenheit von n Kunden
- Stationen bedienen nach einer der Disziplinen FCFS, LCFS, RANDOM oder LCFS-PR
- Kundenzugänge / Kundenabgänge bzgl. Umgebung an jeder Station möglich
- Wechsel Kunden von jeder zu jeder Station zugelassen

(b) Spezifikation Last
- System offen oder geschlossen
- bei offenem System Kundenankünfte entsprechend Poisson-Strom, Parameter λ
- bei geschlossenem System keine Ankünfte ($\lambda = 0$), permanent N Kunden anwesend
- Kundenbewegungen entsprechend "Wechselmatrix"
  H = (h(i,j)) (routing matrix)
  h(i,j) P[Bedienung i → Bedienung j]
  H gehorcht "Zusammenhangsbedingung":
  jede Station von jeder mit endlicher Wahrsch.erreicht
  (direkt oder indirekt, + unter Einschluß Umgeb'g)
- an Station i ankommender Kunde hat exponentiellen Bedienwunsch, Parameter μ_i
- alle stochastischen Größen voneinander unabhängig

- + analoge Erweiterungen für mehrere verschiedene "Muster" ("Ketten")

Untersuchung mit Mitteln des Abschn. 9.1.4
auf Basis von

- Stationszuständen: Stationspopulation n_i
- Netzzuständen: $\mathbf{z}$: = $(n_i;i \quad I) = (n_1,n_2,...,n_M)^T$
  Netz-Zustandsraum

(9.3.07) $Z = \times_{i \quad I} \mathbf{N}_0$ bei offenem System

$Z = \{\mathbf{z}: \sum_{i \quad I} n_i = N\} \quad \times_{i \quad I} \mathbf{N}_0$ bei geschl. System

Ergebnis der Analyse exponentieller Netze
einfach und "überraschend":

als ob Stationen unabhängig voneinander arbeiteten,
unter Belastungsströmen,
welche aus simpler Überlegung
zu "Verkehrsgleichgewicht"
"soviel rein wie raus" stammen

Analoga zu (folgendem) Satz 9.3.08 existieren

- für Exponentielle Netze im Mehr-Ketten
  \+ (Mehr-Klassen, …) Fall
- für "Separable Netze":
  allgemeinere Bedienwunsch-Verteilungen
  (**nicht** für alle Stationstypen !)

Genaueres s. Warteschlangennetze,
Leistungsbewertungs von R+K-Systemen

(auch: Verkehrstheorie der Nachrichtentechnik)

Satz 9.3.08 : Exponentielle Netze (Ein-Ketten-Fall)

sei
- Netz "exponentiell" gemäß Definition 9.3.06
- mit Zustandscharakterisierung gemäß 9.3.07
- $\lambda = (\lambda_i;i=1,2,...,M)^T$ Lösungsvektor des linearen Gl'systems

$$\lambda_i = \sum_{j \quad I} \lambda_j \quad h(j,i) + \lambda \quad h(0,i) \qquad i \quad I$$

bezeichne
- $\mathbf{P}_i := (P_i(n);n \quad \mathbf{N}_0) \qquad i \quad I$

wo $P_i(n) := P[n_i=n]$

stationäre Verteilung (s.Abschn. 9.3.1) isolierter Station i,
° betrieben mit Poisson-Strom (Parameter λ_i)
° und exponentiellen Bedienwünschen (Parameter μ_i)

dann hat stationäre Verteilung Netzzustand (falls existent)
Form

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

wo $\mathbf{z}=(n_i;i \quad I)$
G "Normalisierungskonstante" $\Sigma P[\mathbf{z}] = 1$
und G = 1 im offenen Netz

oB

- Lösung insgesamt "Produktform",
  "wie unabhängig arbeitende Stationen"
  (ist sicher nicht der Fall)

Abbruch

LEER

LEER