Kapitel 3

§3.2 Markov-Modelle mit unendlicher Population

Das M/M/m Modell

Dabei handelt es sich um ein System mit *m* gleichartigen Servern.
Bei $m>1$ Bedienern ist die Zeit bis zur nächsten Ankunft bzw. dem nächsten Abgang wieder exponentiell verteilt, damit liegt auch hier ein GTP vor.
$\lambda_n = \lambda_{n\,n+1} = \lambda$ ergibt sich wie oben.
Die *m* Bediener seien identisch und unabhängig, jeweils mit Bedienrate μ. Im Zustand *i* (d.h. *i* Jobs im System) ergibt sich die Zahl *n* der aktiven Benutzer als $n=\min\{i,m\}$. Ein Übergang in den Zustand i-1 geschieht also hier, wenn keine Ankunft stattfindet und genau einer der Bediener einen Job abschließt:

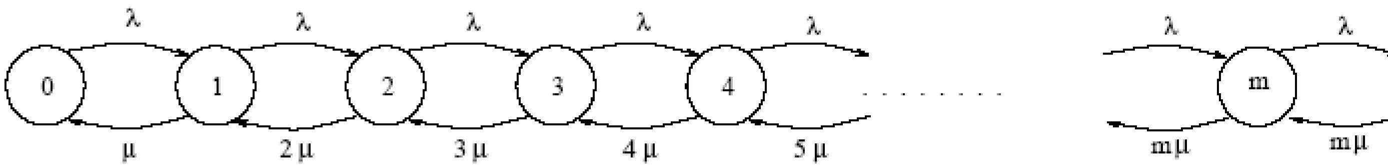
$$\lambda_{n\,n-1} = \mu_n = \lim_{\Delta t \to 0} \frac{\binom{n}{1}\left(1-e^{-\mu\Delta t}\right)\left(e^{-\mu\Delta t}\right)^{n-1} e^{-\lambda\Delta t}}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{\binom{n}{1}\mu\, e^{-\mu\Delta t}}{1} = n\cdot\mu$$

$$\lambda_{n\,n-1} = \mu_n = \begin{cases} n\cdot\mu \text{ falls } 1 \le n \le m \\ m\cdot\mu \text{ falls } n \ge m \end{cases}$$

Als Generatormatrix ergibt sich daher:

$$\Lambda = \begin{pmatrix} -\lambda & \lambda & 0 & \dots & \square & \square & \square & \square & \square \\ \mu & -\lambda-\mu & \lambda & 0 & \dots & \square & \square & \square & \square \\ 0 & 2\mu & -\lambda-2\mu & \lambda & 0 & \dots & \square & \square & \square \\ \dots & 0 & 3\mu & -\lambda-3\mu & \lambda & 0 & \dots & \square & \square \\ \vdots & \vdots & \square & \square & \square & \square & \square & \square & \square \\ \square & \dots & 0 & m\mu & -\lambda-m\mu & \lambda & 0 & \dots & \square \\ \square & \square & \dots & 0 & m\mu & -\lambda-m\mu & \lambda & 0 & \dots \\ \vdots & \vdots & \square & \square & \square & \square & \square & \square & \square \end{pmatrix}$$



Wir haben schon auf M/M/1 eingegangen. Andere Systeme wie M/M/m, M/M/∞, M/M/1/k oder M/M/k/k können ähnlich behandelt werden und werden als Übungsaufgaben gestellt.
Die Gleichgewichtsverteilung

$$p_n = \begin{cases} \frac{\lambda^n}{\mu^n\, n!}\, p_0 & \text{falls } 1 \le n \le m \\ \frac{\lambda^n}{\mu^n\, m!\, m^{n-m}}\, p_0 & \text{falls } n \ge m \end{cases}$$

ergibt sich aus der Formel $p_n = p_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i}$ für den GTP, was sich für $\lambda<\mu$m normieren läßt mit

$$p_0 = \left(\sum_{n=0}^{m-1} \frac{\lambda^n}{\mu^n\, n!} + \frac{\lambda^m}{\mu^m\, m! \left(1-\frac{\lambda}{\mu \mathrm{m}}\right)} \right)^{-1}$$

Für die erwartete Länge der Warteschlange gilt

$$L_q = \frac{\lambda^{m+1}}{\mathrm{m}\mu^{m+1}\, m! \left(1-\frac{\lambda}{\mu \mathrm{m}}\right)^2}\; p_0$$

Daraus können nun mittels der allgemeinen Resultate L, W_q, W berechnet werden.

Auch für das M/M/m Modell im Gleichgewicht kann die Verteilung der (Zufallsvariable) Wartezeit $Q_q$eines (virtuellen) Kunden, der zum Zeitpunkt 0 eintrifft, durch folgende Überlegungen gewonnen werden.

Satz: Die Verteilung der Wartezeit eines Kunden besitzt die folgende Verteilungsfunktion

$$\mathbb{F}_{Q_q}(t) = \mathbb{P}\left[Q_q \le t\right] = 1 - \frac{\left(\frac{\lambda}{\mu}\right)^m e^{-(\mu \mathrm{m}-\lambda)\, t}}{m! \left(1-\frac{\lambda}{\mu \mathrm{m}}\right)}\; p_0$$

Beweis: Wegen der Gedächtnislosigkeit der Exponentialverteilung gilt für $n \in \mathbb{N}_0$bei einer FCFS Strategie

$$\mathbb{P}\left[Q_q \le t \,\middle|\, N = n\right] = \begin{cases} 1 & \text{für } n < m \\ 1 - \sum_{k=0}^{n-m} \frac{e^{-\mathrm{m}\mu t}(\mathrm{m}\mu\, t)^k}{k!} & \text{für } n \ge m \end{cases}$$

Daraus folgt mit dem Satz von der totalen Wahrscheinlichkeit

$$\begin{aligned}
\mathbb{P}\left[Q_q \le t\right] &= \sum_{n=0}^{\infty} \mathbb{P}\left[Q_q \le t \,\middle|\, N = n\right] p_n \\
&= 1 - \sum_{n=m}^{\infty} \left(\sum_{k=0}^{n-m} \frac{e^{-\mathrm{m}\mu t}(\mathrm{m}\mu\, t)^k}{k!} \right) p_n \\
&= 1 - \sum_{n=m}^{\infty} \left(\sum_{k=0}^{n-m} \frac{e^{-\mathrm{m}\mu t}(\mathrm{m}\mu\, t)^k}{k!} \right) \frac{\lambda^n}{\mu^n\, m!\, m^{n-m}}\; p_0 \\
&= 1 - \frac{e^{-\mathrm{m}\mu\, t}\, m^m}{m!}\; p_0 \sum_{k=0}^{\infty} \frac{(\mathrm{m}\mu\, t)^k}{k!} \left(\sum_{n=m+k}^{\infty} \frac{\lambda^n}{\mu^n\, m^n} \right) \\
&= 1 - \frac{e^{-\mathrm{m}\mu\, t}\, m^m}{m!}\; p_0 \sum_{k=0}^{\infty} \frac{(\mathrm{m}\mu\, t)^k}{k!} \frac{\left(\frac{\lambda}{\mathrm{m}\mu}\right)^{m+k}}{1-\left(\frac{\lambda}{\mathrm{m}\mu}\right)} \\
&= 1 - \frac{e^{-\mathrm{m}\mu\, t}}{m! \left(1-\frac{\lambda}{\mathrm{m}\mu}\right)} \left(\frac{\lambda}{\mu}\right)^m p_0 \sum_{k=0}^{\infty} \frac{(\lambda\, t)^k}{k!} \\
&= 1 - \frac{\left(\frac{\lambda}{\mu}\right)^m e^{-(\mathrm{m}\mu-\lambda)\, t}}{m! \left(1-\frac{\lambda}{\mathrm{m}\mu}\right)}\; p_0
\end{aligned}$$

```
λ = 0.8; μ = 0.2; t = 5.5;
p0[m_] := (Sum[(λ/(μ n!))^n, {n, 0, m - 1}] + λ^m/(μ^m m! (1 - λ/(m μ))))^-1;
pQ[m_] := 1 - ((λ/μ)^m p0[m])/(m! (1 - λ/(m μ))) Exp[-(m μ - λ) t];
Table[pQ[m], {m, 5, 10}]
{0.726684, 0.928271, 0.983428, 0.996819, 0.999499, 0.999933}
```

Interessant ist auch die bedingte Wahrscheinlichkeit

$$\mathbb{P}\left[Q_q > t \,\middle|\, Q_q > 0\right] = \frac{\mathbb{P}[Q_q > t, Q_q > 0]}{\mathbb{P}[Q_q > 0]} = \frac{\mathbb{P}[Q_q > t]}{\mathbb{P}[Q_q > 0]} = e^{-(m\mu-\lambda)\,t}$$

Ein Kunde muss also mit Wahrscheinlichkeit

$$1 - \frac{\left(\frac{\lambda}{\mu}\right)^m}{m!\left(1-\frac{\lambda}{m\mu}\right)}\ p_0$$

gar nicht warten und, wenn er warten muss, so muss er eine exponentialverteilte Zeit mit Mittelwert

$$\frac{1}{m\mu-\lambda}$$

Die Wahrscheinlichkeit dafür, daß ein Kunde warten muss, kann man wie folgt bestimmen

$$\mathbb{P}[\text{Kunde muss warten in der Warteschlange}]=\sum_{k=m}^{\infty} p_k=\sum_{k=m}^{\infty} \frac{\lambda^k}{\mu^k\, m!\, m^{k-m}}\ p_0$$

$$=\frac{\left(\frac{(m\,\rho)^m}{m!}\right)\left(\frac{1}{1-\rho}\right)}{\left[\sum_{k=0}^{m-1}\frac{(m\,\rho)^k}{k!}+\left(\frac{(m\,\rho)^m}{m!}\right)\left(\frac{1}{1-\rho}\right)\right]},\ \text{wobei } \rho=\frac{\lambda}{m\mu}$$

Diese Formula heißt ***Erlang C Formula***, kurz $C(m,\frac{\lambda}{\mu})$

λ = 0.8; μ = 0.2; m = 5;

$$\text{p0} = \left(\text{Sum}\left[\left(\frac{\lambda}{\mu\, \text{n!}}\right)^{\text{n}}, \{\text{n}, 0, \text{m}-1\}\right] + \frac{\lambda^{\text{m}}}{\mu^{\text{m}}\, \text{m!}\left(1-\frac{\lambda}{\text{m}\,\mu}\right)}\right)^{-1};$$

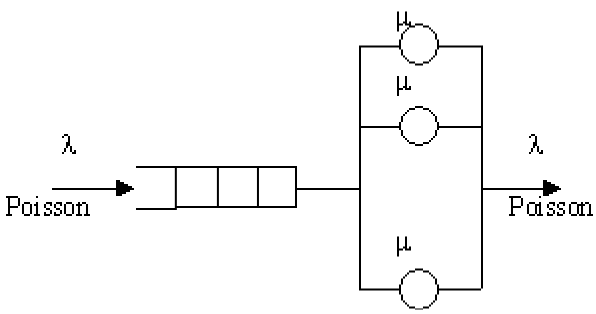
$$\text{pN} = \frac{\left(\frac{\lambda}{\mu}\right)^{\text{m}}}{\text{m!}\left(1-\frac{\lambda}{\text{m}\,\mu}\right)}\ \text{p0}$$

$$\text{pW} = \frac{1}{\text{m!}}\left(\frac{\lambda}{\mu}\right)^{\text{m}}\frac{1}{1-\frac{\lambda}{\text{m}\,\mu}}\ \text{p0}$$

```
0.821085
0.821085
```

Bemerkung: Man beachte, dass alle Formeln für das M/M/m Modell für m=1 in die entsprechenden Formeln für das M/M/1 Modell übergehen.

Satz (Satz von Burke für M/M/m-Systeme): Für jedes M/M/m-System im Gleichgewicht stimmen Ankunfts- und Abgangsprozeß stochastisch übereit.



Zusammenfassung (Leistungsmerkmale bei M/M/m): Betrachte ein M/M/m-System mit Ankunftsrate λ, m gleichartigen Servern, Bedienrate μ,

$$p_0 = \left(\sum_{n=0}^{m-1} \frac{\lambda^n}{\mu^n\, n!} + \frac{\lambda^m}{\mu^m\, m! \left(1 - \frac{\lambda}{\mu \mathrm{m}}\right)} \right)^{-1}$$

Dann gilt

- mittlere Anzahl von Kunden in der Warteschlange:

$$L_q = \mathbb{E}\left[N_q\right] = \frac{\lambda^{m+1}}{\mathrm{m}\mu^{m+1}\, m! \left(1 - \frac{\lambda}{\mu \mathrm{m}}\right)^2}\; p_0$$

- mittlere Wartezeit

$$W_q = \mathbb{E}\left[Q_q\right] = \frac{L_q}{\lambda} = \frac{\lambda^m}{\mathrm{m}\mu^{m+1}\, m! \left(1 - \frac{\lambda}{\mu \mathrm{m}}\right)^2}\; p_0$$

- mittlere Systemzeit

$$W = \mathbb{E}[Q] = W_q + \mathbb{E}[B] = \frac{\lambda^m}{\mathrm{m}\mu^{m+1}\, m! \left(1 - \frac{\lambda}{\mu \mathrm{m}}\right)^2}\; p_0 + \frac{1}{\mu}$$

- mittlere Anzahl von Kunden im System:

$$L = \mathbb{E}[N] = \mathrm{W}\lambda = \lambda\left(W_q + \mathbb{E}[B]\right) = L_q + \frac{\lambda}{\mu} = \frac{\lambda^{m+1}}{\mathrm{m}\mu^{m+1}\, m! \left(1 - \frac{\lambda}{\mu \mathrm{m}}\right)^2}\; p_0 + \frac{\lambda}{\mu}$$

- die Verteilungsfunktion der Wartezeit eines Kunden

$$\mathbb{F}_{Q_q}(t) = \mathbb{P}\left[Q_q \le t\right] = 1 - \frac{\left(\frac{\lambda}{\mu}\right)^m e^{-(\mathrm{m}\mu - \lambda)\, t}}{m! \left(1 - \frac{\lambda}{\mathrm{m}\mu}\right)}\; p_0$$

```
λ = 0.3; μ = 0.5; m = 2;
```

$$\mathbf{p0} = \left(\mathbf{Sum}\left[\left(\frac{\lambda}{\mu\, \mathbf{n!}}\right)^{\mathbf{n}}, \{\mathbf{n}, \mathbf{0}, \mathbf{m-1}\}\right] + \frac{\lambda^{\mathbf{m}}}{\mu^{\mathbf{m}}\, \mathbf{m!}\, \left(\mathbf{1} - \frac{\lambda}{\mathbf{m}\,\mu}\right)} \right)^{-1};$$

$$\frac{\lambda^{\mathbf{m}}}{\mathbf{m}\, \mu^{\mathbf{m+1}}\, \mathbf{m!}\, \left(\mathbf{1} - \frac{\lambda}{\mathbf{m}\,\mu}\right)^{2}}\; \mathbf{p0} + \frac{\mathbf{1}}{\mu}$$

```
2.1978
```

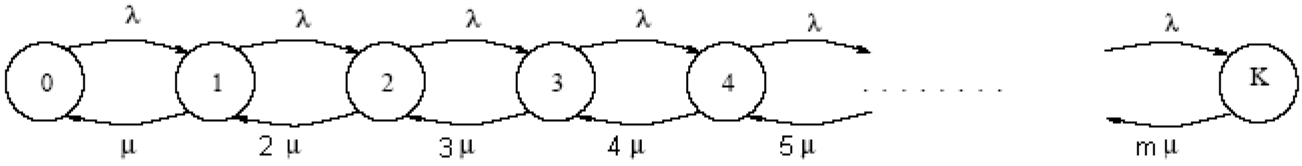
Das M/M/m/K Modell

Dabei handelt es sich um ein System mit m gleichartigen Servern mit einer Systemkapazität K, d.h. Kunden, die bereits K Kunden im System vorfinden, werden abgewiesen. Ein Mathematische Model ist ein Geburts- und Todesprozess mit den Raten

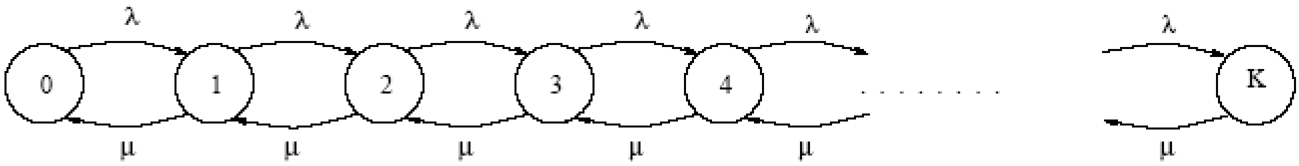
$$\lambda_{n,n+1} = \lambda_n = \begin{cases} \lambda & \text{falls } \; 0 \le n < K \\ 0 & \text{falls } \; n \ge K \end{cases}$$

$$\lambda_{n\,n-1} = \mu_n = \begin{cases} n\,\mu & \text{falls } 1 \le n \le m \\ m\,\mu & \text{falls } n \ge m \end{cases}$$

Dieser Prozess hat nur die Zustände I={0,1,...,K}.



Wenn m=1,



dann ergibt sich als Generatormatrix für K=6

$$\Lambda=\begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & 0 & 0 \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & 0 & 0 \\ 0 & \mu & -(\lambda+\mu) & \lambda & 0 & 0 & 0 \\ 0 & 0 & \mu & -(\lambda+\mu) & \lambda & 0 & 0 \\ 0 & 0 & 0 & \mu & -(\lambda+\mu) & \lambda & 0 \\ 0 & 0 & 0 & 0 & \mu & -(\lambda+\mu) & \lambda \\ 0 & 0 & 0 & 0 & 0 & \mu & -\mu \end{pmatrix}$$

Die Gleichgewichtsverteilung

$$p_n = \begin{cases} \frac{\lambda^n}{\mu^n\, n!}\; p_0 \;\text{ falls }\; 1 \le n \le m \\ \frac{\lambda^n}{\mu^n\, m!\, m^{n-m}}\; p_0 \;\text{ falls } m \le n \le K \end{cases}$$

ergibt sich wieder aus der Formel $p_n = p_0 \prod_{i=1}^{n} \frac{\lambda_{i-1}}{\mu_i}$ für den GTP. Sie existiert in jedem Fall, wobei

$$p_0 = \left(\sum_{n=0}^{m-1} \frac{\lambda^n}{\mu^n\, n!} + \frac{\lambda^m \left(1-\left(\frac{\lambda}{\mathrm{m}\mu}\right)^{K-m+1}\right)}{\mu^m\, m! \left(1-\frac{\lambda}{\mathrm{m}\mu}\right)}\right)^{-1} \quad \text{für } \lambda \neq \mathrm{m}\,\mu \text{ bzw.}$$

$$p_0 = \left(\sum_{n=0}^{m-1} \frac{\lambda^n}{\mu^n\, n!} + \frac{\lambda^m}{\mu^m\, m!}\,(K-m+1)\right)^{-1} \quad \text{für } \lambda = \mathrm{m}\,\mu.$$

Man beachte, dass die Formel für K→∞ in die betreffende Formel für das M/M/m Modell übergeht.
Die Formel für die erwartete Länge der Warteschlange lautet für $\lambda \neq \mathrm{m}\,\mu$

$$L_q = \frac{\lambda^{m+1}\, p_0 \left(1-\left(\frac{\lambda}{\mathrm{m}\mu}\right)^{K-m+1} - \left(1-\frac{\lambda}{\mathrm{m}\mu}\right)(K-m+1)\left(\frac{\lambda}{\mathrm{m}\mu}\right)^{K-m}\right)}{\mathrm{m}\mu^{m+1}\, m! \left(1-\frac{\lambda}{\mathrm{m}\mu}\right)^2}$$

Die entsprechende Formel für $\lambda = \mathrm{m}\,\mu$ wird durch die Regel von de L'Hospital gewonnen. Die erwarteten Wartezeiten werden wieder über die Formel von Little berechnet, allerdings muss man beachtet, dass die mittlere Anzahl von eintreffenden Kunden je Zeiteinheit hier gleich $\lambda(1-p_K)$ ist, weil ja nicht jeder eintreffende Kunde tatsächlich in das System eintritt.
Wir erhalten daher

$W_q = \frac{L_q}{\lambda(1-p_K)}$ und $W = W_q + \mathbb{E}[B] = W_q + \frac{1}{\mu}$,
sowie
$L = \mathbb{E}[N] = W\lambda(1 - p_K) = L_q + \frac{\lambda(1-p_K)}{\mu}$.
Auch in diesem Fall kann die Verteilung der Wartezeit berechnet werden.

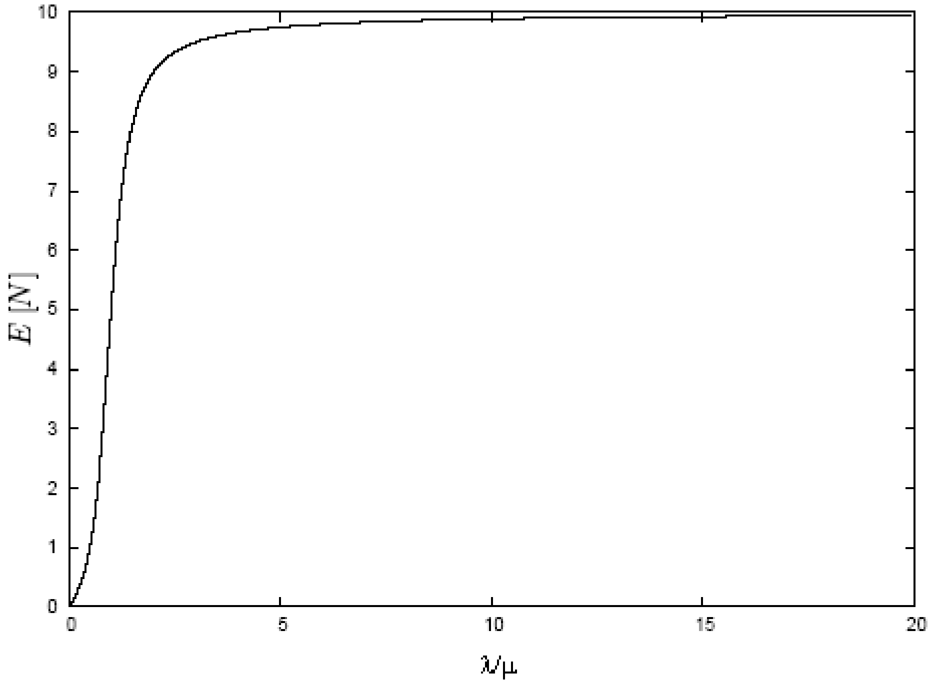


Figure Mean number of Customers in the system for M/M/1/10-queue

Die Wahrscheinliechkeit dafür, daß ein Kunde abgewiesen wird (Blokierwahrscheinlichkeit), kann für System M/M/1/K wie folgt bestimmen werden

$$\mathbb{P}[\text{ein Kunde wird verloren}] = p_K = \begin{cases} \frac{\left(\frac{\lambda}{\mu}\right)^K - \left(\frac{\lambda}{\mu}\right)^{K+1}}{1-\left(\frac{\lambda}{\mu}\right)^{K+1}} & \text{falls } \frac{\lambda}{\mu} \neq 1 \\ \frac{1}{K+1} & \text{falls } \frac{\lambda}{\mu} = 1 \end{cases}$$

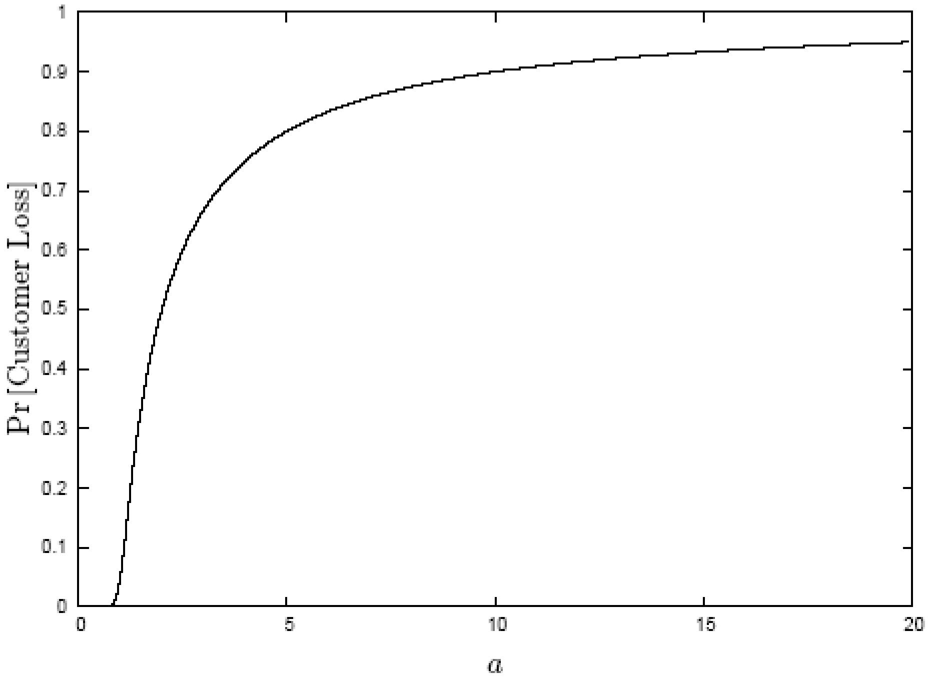


Figure Loss Probability for M/M/1/20

Beispiel 16. Um die Auswirkung eines begrenzten Speichers abzuschätzen, berechnen Sie ab welcher Anzahl von Pakete n im System die Wahrscheinlichkeit das noch eines hinzukommt 1% oder weniger beträgt.

Lösung:

Die Wahrscheinlichkeit, dass sich mehr als n Pakete im System befinden, lässt sich als totale Wahrscheinlichkeit ausdrücken:

$$\mathbb{P}[N > n] \geq p_0 \sum_{k=n+1}^{\infty} \left(\frac{\lambda}{\mu}\right)^k$$

Günstiger ist es jedoch, dass durch die etsprechende Gegenwahrscheinlichkeit auszudrücken:

$$\mathbb{P}[N > n] \geq 1 - p_0 \sum_{k=0}^{n} \left(\frac{\lambda}{\mu}\right)^k$$

mit $\rho=\frac{\lambda}{\mu}$, $p_0 = 1 - \rho$, $\mathbb{P}[N > n] = 0.01$

und der endlichen geometrischen Reihe

$$\sum_{k=0}^{n} q^k = \frac{q^{n+1}-1}{q-1}$$

folgt

$$\begin{aligned} 0.01 &\geq 1 - (1-\rho) \sum_{k=0}^{n} \rho^k \\ &\geq 1 - (1-\rho) \frac{\rho^{n+1}-1}{\rho-1} \\ &\geq 1 - (1-\rho)(-1) \frac{(1-\rho^{n+1})}{(-1)(1-\rho)} \\ &\geq 1 - \left(1-\rho^{n+1}\right) \\ &\geq \rho^{n+1} \end{aligned}$$

umstellen nach n

$$\begin{aligned} \log 0.01 &\geq (n+1) \log \rho \\ n &\geq \frac{\log 0.01}{\log \rho} - 1 \\ &\geq 2.3219 \approx 3 \end{aligned}$$

D.h. mit einer Wahrscheinlichkeit von unter 1% befinden sich mehr als drei Pakete im System.

Beispiel 17. Berechnen Sie für eine M/M/1/5 - Warteschlange die Blokierwahrscheinlichkeit für
(a) $\rho = \frac{\lambda}{\mu} = 0.25$ bzw.
(b) $\rho = \frac{\lambda}{\mu} = 1$.

Lösung:

Für eine M/M/1/K Warteschlange (Systemkapazität K) ist die Blockierwahrscheinlichkeit
$p_B = \mathbb{P}[N > K] =$ (falls $N > K$ die new hinzukommenden Paket werden verworfen,
daher können sich nie mehr als K Pakete im System aufhalten) $= \mathbb{P}[N = K]$

$$= p_K = \begin{cases} \frac{\rho^K - \rho^{K+1}}{1-\rho^{K+1}} & \text{falls } \rho \neq 1 \\ \frac{1}{K+1} & \text{sonst} \end{cases}$$

Damit ist
(a) $p_B = \frac{0.25^5 - 0.25^6}{1-0.25^6} \approx 0.73 \cdot 10^{-3}$ für ρ=0.25, und
(b) $p_B = 1/6$ für ρ=1.

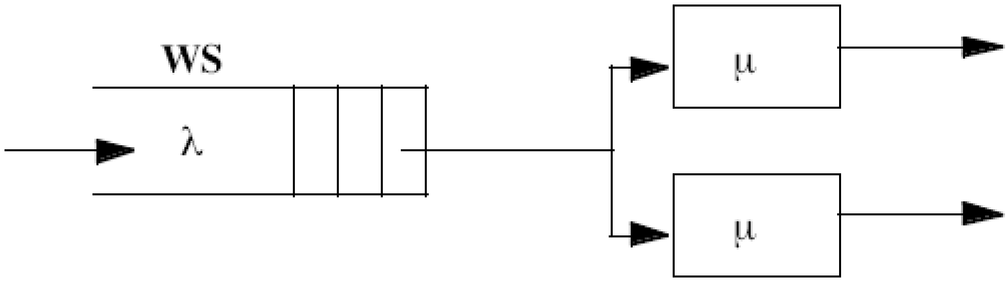
Beispiel 18. Vergleichen Sie drei Alternativen:
1. M/M/1 Modell mit Bedienungsrate 2μ;
2. M/M/2 Modell mit gemeinsamer Warteschlange und Bedienungsraten μ für jedes Server.
3. Zwei getrennte Warteschlange M/M/1 mit Bedienungsraten μ.
Hinweis: Benutzen Sie als Maß für die Leistungsfähigkeit den Erwartungswert für die Anzahl der Pakete, die versendet werden bzw. auf ihren Versand warten.

Lösung:
Um die drei Alternativen vergleichen zu können, sei ρ in diesem Fall definiert als $\rho = \frac{\lambda}{2\mu}$

1. M/M/1 mit 2μ

$$\mathbb{E}[N] = \frac{\rho}{1-\rho}$$

2. M/M/2

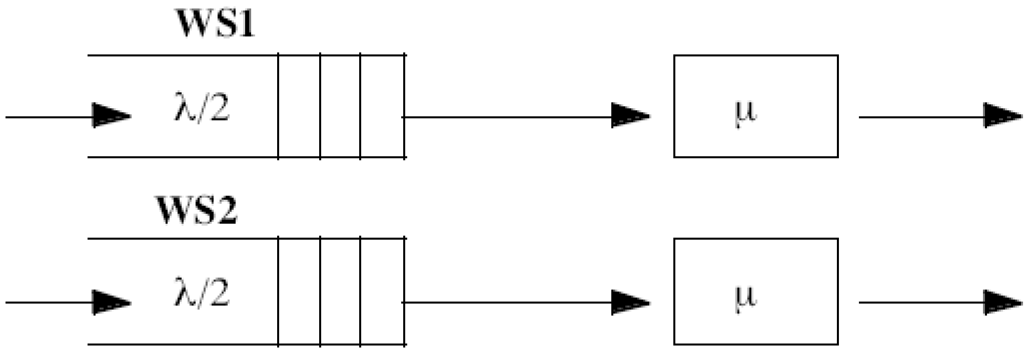


In diesem Fall m=2,

$$p_0 = \left(\sum_{n=0}^{m-1} \frac{\lambda^n}{\mu^n\, n!} + \frac{\lambda^m}{\mu^m\, m! \left(1-\frac{\lambda}{\mu \mathrm{m}}\right)}\right)^{-1} = \left(1 + 2\,\rho + 2\,\rho^2\, \frac{1}{(1-\rho)}\right)^{-1} = \frac{1-\rho}{1+\rho}$$

$$\mathbb{E}[N] = \frac{\lambda^{m+1}}{\mathrm{m}\mu^{m+1}\, m! \left(1-\frac{\lambda}{\mu \mathrm{m}}\right)^2}\, p_0 + \frac{\lambda}{\mu} = \frac{2\,\rho^3}{(1-\rho)^2} \left(\frac{1-\rho}{1+\rho}\right) + 2\,\rho = \frac{2\,\rho}{1-\rho^2}$$

3. Zwei M/M/1 Systeme,



für jede der beiden Bedienstationen gilt für die mittlere Anzahl von Aufträgen:

$$\frac{\frac{(1/2)\,\lambda}{\mu}}{1-\frac{(1/2)\,\lambda}{\mu}} = \frac{\rho}{1-\rho}$$

Für das Gesamtsystem folgt deshalb:

$$\mathbb{E}[N] = 2\frac{\rho}{1-\rho}$$

In der folgenden Abbildung ist für die drei Alternativen die mittlere Anzahl an Aufträgen im System abhängig von der Auslastung ρ des Systems dargestellt (setze ρ=rho)

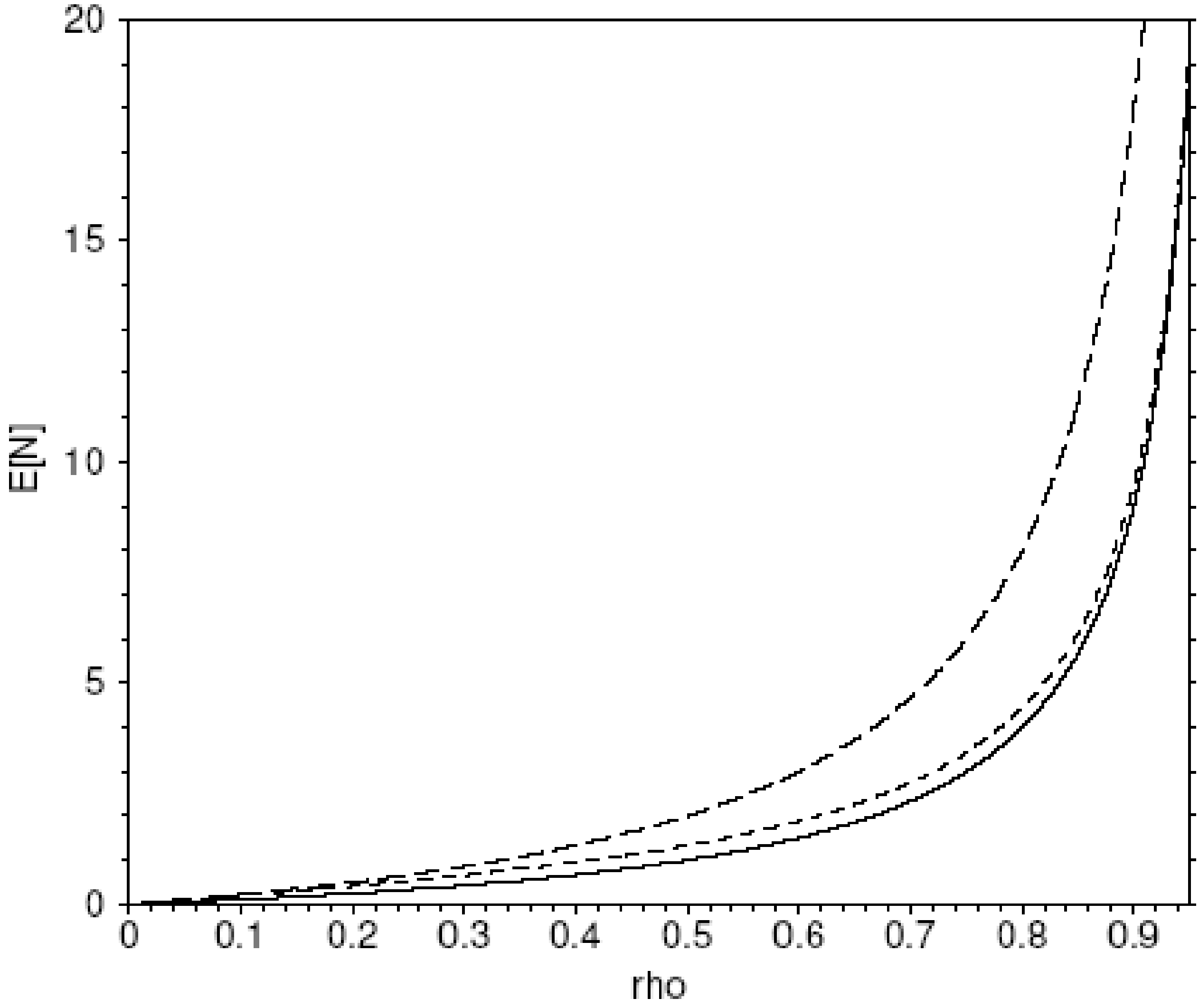


Abbildung Vergleich der drei Alternativen

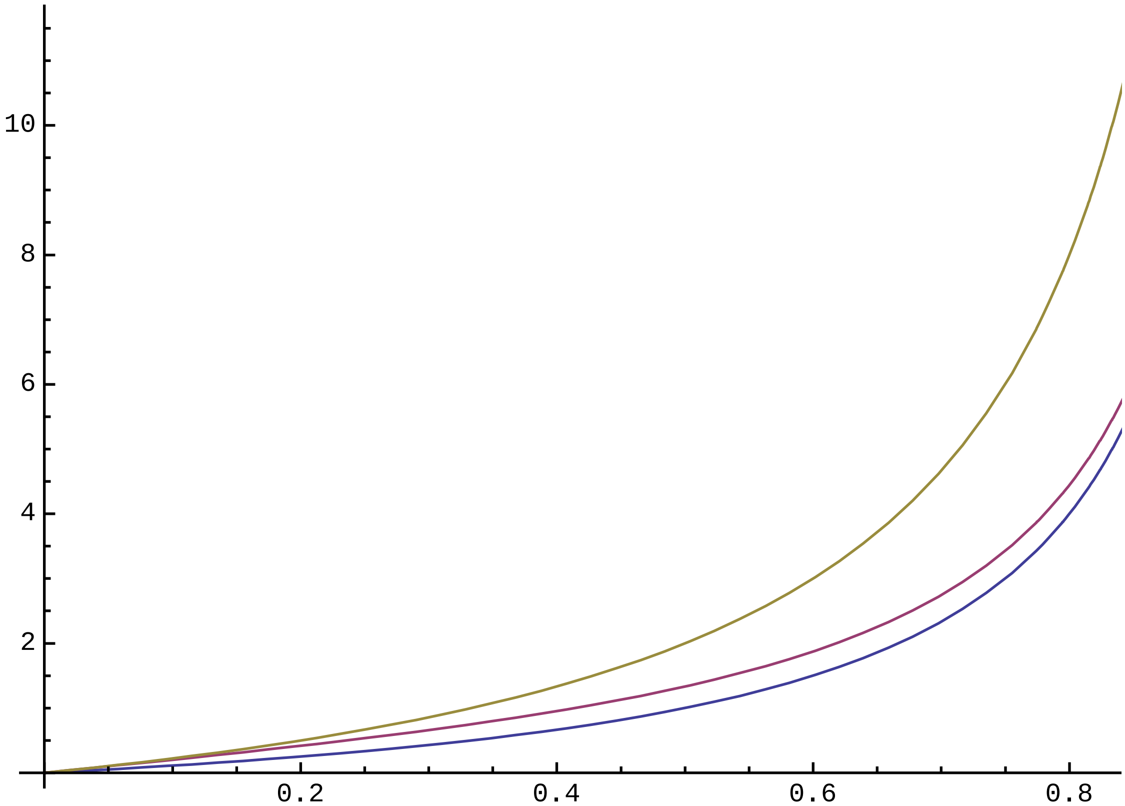
Die durchgezogene Linie steht für die Alternative 1, die kurz gestrichelte Linie steht für die Alternative 2, und die lang gestrichelte Linie steht für die Alternative 3.

$$EN1[\rho\_] := \frac{\rho}{1-\rho};$$

$$EN2[\rho\_] := \frac{2\,\rho}{1-\rho^2};$$

$$EN3[\rho\_] := 2\,\frac{\rho}{1-\rho};$$

```
Plot[{EN1[ρ], EN2[ρ], EN3[ρ]}, {ρ, 0, 0.95},
 PlotLegend → {"Model 1", "Model 2", "Model 3"},
 LegendPosition → {0.8, -0.8}]
```



Das M/M/m/m Modell

Dieses Model wurde bereits 1917 von Erlang (Agner Krarup Erlang, 1878-1929, dänischer Ingenieur und Mathematiker) für eine Telefonvermittlung verwendet. Dabei werden die ankommenden Telefongespräche durch einen Poisson-Prozess modeliert und die Sprechzeiten als exponentialverteilt angenommen.
Das mathematische Modell ist ein Geburts- und Todesprozess mit den Raten

$$\lambda_{n,n+1} = \lambda_n = \begin{cases} \lambda & \text{falls} \;\; 0 \le n < m \\ 0 & \text{falls} \;\; n \ge m \end{cases}$$

$$\lambda_{n\,n-1} = \mu_n = n\,\mu, \;\; 1 \le n \le m$$

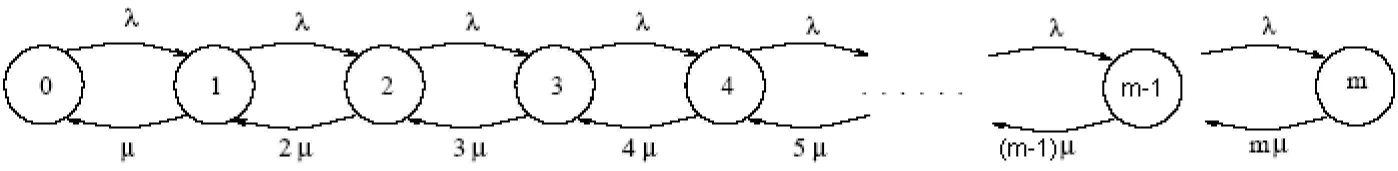


Figure the M/M/m/m queue

Die Gleichgewichtsverteilung

$$p_n = \begin{cases} \frac{\lambda^n}{\mu^n\, n!}\; p_0 & \text{falls } 1 \le n \le m \\ 0 & \text{falls } n > m \end{cases},$$

wobei

$$p_0 = \left(\sum_{n=0}^{m} \frac{\lambda^n}{\mu^n\, n!}\right)^{-1}$$

Die von Erlang hergeleiteten Formeln dienten dazu, die Telefonzentrale richtig zu dimensionieren. Seine sogenannten Erlangschen Verlustformeln lauten

$$p_m = \frac{\lambda^m}{\mu^m\, m!} \left(\sum_{n=0}^{m} \frac{\lambda^n}{\mu^n\, n!}\right)^{-1}$$

und heißen ***Erlang B formula***, kurz $B(m, \frac{\lambda}{\mu})$.

Mit Wahrscheinlichkeit p_m geht der Kunder verloren, weil er keine freie Telefonleitung vorfindet. Die große Bedeutung dieser Formeln liegt darin, daß sie sogar für ein M/G/m/m Modell gelten, in dem die mittlere Bedienungszeit $\frac{1}{\mu}$ ist. Erlang nahm bei seinem Untersuchungen die Bedienungszeit als konstant an, sein Beweis für die Formeln war aber nicht ganz korrekt.

Das M/M/∞ Modell (unendlich viel Servern)

In diesem Modell wird jeder Kunde sofort bedient, oft ist es gut geeignet, um eine Selbstbedienungsanlage zu beschreiben. Mathematisch handelt es sich um einen Geburts- und Todesprozess mit

$$\lambda_{n\,n+1} = \lambda_n = \lambda, \quad n = 0, 1, 2, \ \ldots$$
$$\lambda_{n\,n-1} = \mu_n = n\,\mu, \quad n = 1, 2, 3, \ \ldots$$

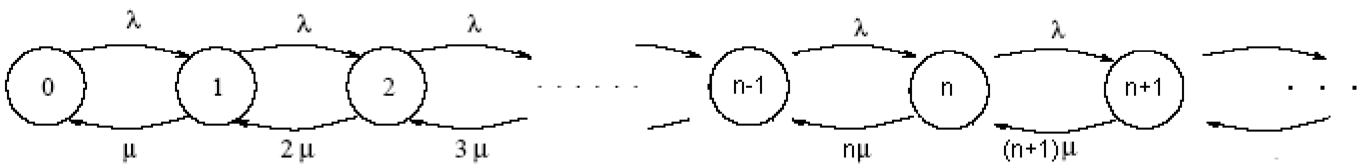


Figure the M/M/∞ queue

Damit gilt

$$p_n = \frac{\lambda^n}{\mu^n\, n!}\, e^{\left(-\frac{\lambda}{\mu}\right)}.$$

Die Anzahl der Kunden im System ist Poisson-verteilt mit Parameter $\frac{\lambda}{\mu}$. Auch diese Formel gild sogar für ein M/G/1 Modell.

$$L = \mathbb{E}[N] = \frac{\lambda}{\mu} \text{ und}$$
$$W = \mathbb{E}[Q] = \text{(Formel von Little) } \frac{L}{\lambda} = \frac{1}{\mu}$$

Zustandsabhängiges Service

Kunden treffen nach einem Poisson-Prozess mit Rate λ in ein System mit einem Server ein, der negative exponential verteilt mit unterschiedlicher Intensität arbeitet. Sind weniger als k Kunden im System, so erbeitet der Server langsam mit Rate μ_l. Bei mindestens k Kunden arbeitet er schnell mit Rate μ_s. Die Kapazität des System sei unbeschränkt. Wir erhalten die Raten

$$\lambda_{n\,n+1} = \lambda_n = \lambda,$$
$$\lambda_{n\,n-1} = \mu_n = \begin{cases} \mu_l & \text{falls } 1 \le n < k \\ \mu_s & \text{falls } n \ge k \end{cases}.$$

Im Fall $\frac{\lambda}{\mu_s} < 1$ existiert die Gleichgewichtsverteilung und für sie gilt

$$p_n = \begin{cases} \frac{\lambda^n}{\mu_l^n}\, p_0 & \text{falls } 1 \le n < k \\ \frac{\lambda^n}{\mu_l^{k-1}\,\mu_s^{n-k+1}}\, p_0 & \text{falls } n \ge k \end{cases}$$

mit

$$p_0 = \begin{cases} \left(\frac{1-(\lambda/\mu_l)^k}{1-\lambda/\mu_l} + \frac{(\lambda/\mu_l)^{k-1}\,\lambda/\mu_s}{1-\lambda/\mu_s}\right)^{-1} & \text{falls } \frac{\lambda}{\mu_l} \ne 1 \\ \left(k + \frac{\lambda/\mu_s}{1-\lambda/\mu_s}\right)^{-1} & \text{falls } \frac{\lambda}{\mu_l} = 1 \end{cases}.$$

Verallgemeinerungen dieses Modells liegen nahe und sind genauso zu behandeln: Man kann statt einem m Server einsetzen, diese Können je nach Anzahl der vorhandenen Kunden (Aufträge) mehrere verschiedene Arbeitsintensitäten aufweisen. Man kann auch bei Vorhandensein von k_1, k_2, ... Kunden einen 2.,3., ... Server einsetzen.

Ungedultige Kunden

Zumindest drei Typen von ungedultigen Kunden sind denkbar.

● Solche, die sich bei einer gewissen Länge der Warteschlange nur ungern anstellen (*balking*), d.h. für jeden Systemzustand n gibt es eine *Warte* wahrscheinlichkeit w_n, mit der ein eintrefender Kunde in das System eintritt.

● Ein nächster Typ von ungeduldigen Kunden stellt sich zuerst an, verläßt aber, wenn zu lange warten muss, das System ohne Bedienung. Diese Art von Ungeduld (*reneging*) kann man dadurch modellieren, dass in jedem Systemzustand n jeder Kunde mi Rate r_n das System verläßt.

● Wenn es für jeden Server eine eigene Warteschlange gibt, tritt eine dritte Art von Ungedult auf, nämlich dasWechseln den Warteschlangen (*jockeying*).