3.3 Birth-and-Death Prozesse

M/M/1-Warteschlangen stellen einen Spezialfall so genannter Birth-and-Death Prozesse dar. Darunter versteht man kontinuierliche Markov-Ketten mit einem Übergangsdiagramm der in Abbildung 6 angegebenen Form.

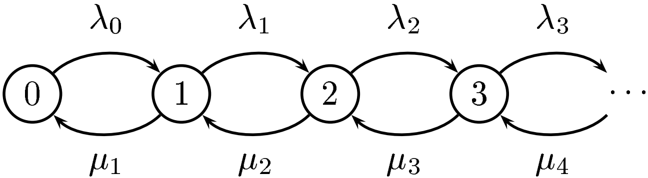

Abbildung: Ein Birth-and-Death Prozess

Bei solchen Prozessen erhalten wir das folgende Gleichungssystem für den Gleichgewichtszustand:

$$\begin{aligned} 0 &= \lambda_{k-1}\pi_{k-1} + \mu_{k+1}\pi_{k+1} - (\lambda_k + \mu_k)\pi_k \quad \text{für alle } k \geq 1, \\ 0 &= \mu_1\pi_1 - \lambda_0\pi_0. \end{aligned}$$

Dieses System können wir mit derselben Technik wie bei den M/M/1-Warteschlangen auflösen und erhalten

$$\pi_k = \pi_0 \cdot \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}} \quad \text{für alle } k \geq 1. \tag{18}$$

Die Normierungsbedingung $\sum_{k\geq 0}\pi_k = 1$ liefert

$$\pi_0 = \frac{1}{1+\sum_{k\geq 1}\prod_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}}}, \tag{19}$$

sofern $\sum_{k\geq 1}\prod_{i=0}^{k-1}\frac{\lambda_i}{\mu_{i+1}}$ nicht divergiert. Ansonsten hat das Gleichungssystem wiederum nur die triviale Lösung $\pi_0 = \pi_1 = \ldots = 0$.

Viele interessante Probleme lassen sich einfach als Birth-and-Death Prozess modellieren. Wir betrachten zwei Beispiele.

Beispiel 157

Abbildung 7 zeigt eine M/M/1-Warteschlange mit beschränktem Warteraum. Dieser liegt das Modell zu Grunde, dass ankommende Jobs nur dann ins System aufgenommen werden, wenn im aktuellen Zustand weniger als N Jobs auf ihre Bearbeitung warten. Neben den klassischen Beispielen einer Arztpraxis oder Ähnlichem ist dieses Modell auch für viele Probleme in der Informatik zutreffend, da hier für die Verwaltung der auf Bearbeitung wartenden Jobs oft fest dimensionierte Arrays vorgesehen werden.

Beispiel

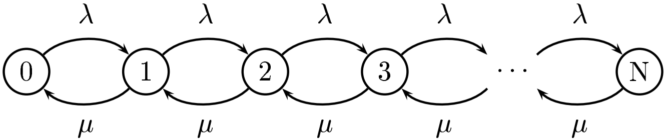

Abbildung: M/M/1-Warteschlange mit beschränktem Warteraum

Beispiel

Die Verteilung im Gleichgewichtszustand erhalten wir sofort, in dem wir in (18) *und* (19) *die entsprechenden Werte für* λ_i *und* μ_i *einsetzen:*

$$\pi_k = \rho^k \cdot \pi_0 \quad \textit{für alle } 1 \leq k \leq N \textit{ mit } \rho = \lambda/\mu$$

und

$$\pi_0 = \frac{1}{1 + \sum_{i=1}^{N} \rho^i} = \frac{1}{\sum_{i=0}^{N} \rho^i} = \begin{cases} \frac{1}{N+1} & \textit{für } \rho = 1, \\ \frac{1-\rho}{1-\rho^{N+1}} & \textit{sonst.} \end{cases}$$

In diesem Fall konvergiert das System für alle Werte von ρ *in einen stationären Zustand. Auch für* $\rho \geq 1$ *kann die Warteschlange nicht beliebig lang werden, da im Zustand* N *keine weiteren Jobs mehr entgegengenommen werden. Für* $\rho < 1$ *und* $N \to \infty$ *konvergiert das System gegen eine „normale" M/M/1-Warteschlange.*

Beispiel 158

Wir modellieren ein Anfragesystem mit einem einzelnen Server, an den M Terminals angeschlossen sind. An den Terminals treffen Anfragen mit der Rate λ ein und werden an den Server weitergeleitet. Wenn ein Terminal eine Anfrage abgeschickt hat, die noch nicht bearbeitet wurde, so bleibt es blockiert, bis es eine Antwort vom Server erhalten hat.

Wir stellen dieses System durch eine kontinuierliche Markov-Kette dar, deren Zustände $S = \{0, \ldots, M\}$ der Anzahl von Anfragen entsprechen, die gerade beim Server in Bearbeitung sind (die Bearbeitungsrate bezeichnen wir wieder wie gewohnt mit μ).

Im Zustand 0 treffen beim Server Anfragen mit der Rate $M\lambda$ ein, da sich die Anfragen aller M Terminals addieren. Im Zustand i warten i Terminals auf Antwort vom Server und sind deshalb blockiert. Somit muss der Server nur noch eine Anfragerate von $(M-i)\lambda$ entgegennehmen. Abbildung 8 zeigt das resultierende System.

Beispiel

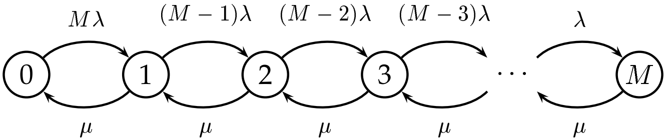

Abbildung: Markov-Kette zu einem Server mit M Terminals

Beispiel

Auch hier finden wir die stationäre Verteilung durch Einsetzen der entsprechenden Werte für λ_i und μ_i in (18) und (19):

$$\pi_k = \pi_0 \cdot \prod_{i=0}^{k-1} \frac{\lambda(M-i)}{\mu} = \pi_0 \cdot \left(\frac{\lambda}{\mu}\right)^k \cdot M^{\underline{k}} \quad \text{für alle } k \geq 1$$

und

$$\pi_0 = \frac{1}{\sum_{k=0}^{M} \left(\frac{\lambda}{\mu}\right)^k \cdot M^{\underline{k}}}.$$

Hierbei bezeichnet $M^{\underline{k}} := M(M-1)\ldots(M-k+1)$ die k-te fallende Faktorielle von M (siehe Vorlesung DS).