Von englischen und deutschen Postämtern

Teilnehmer:

Ricky Burzlaff	Wilhelm-Ostwald-Gymnasium, Leipzig
Robert Butz	Herder-Oberschule, Berlin
Jan Putzig	Heinrich-Hertz-Oberschule, Berlin
Antoni Schilling	Herder-Oberschule, Berlin
Jakob Steinbrück	Käthe-Kollwitz-Oberschule, Berlin
Patrick Zielonka	Herder-Oberschule, Berlin

Gruppenleiter:

André Henning	Humboldt-Universität zu Berlin
Elke Warmuth	Humboldt-Universität zu Berlin Mitglied im DFG-Forschungszentrum MATHEON „Mathematik für Schlüsseltechnologien“

Dass die Engländer fast immer in den technischen und wirtschaftlichen Bereichen die Nase vor den Deutschen hatten, ist schon längst aus der Geschichte bekannt. Nur in der Mathematik schien es bisher so zu sein, dass das Land von Gauß und Ries kaum zu schlagen war. Doch auch da können – wie das folgende Beispiel beweist – die Engländer eine Vorreiterrolle spielen. Wenn man sich anschaut, wie beispielsweise die Warteschlangen auf Postämtern gehandhabt werden, so kann man zwei unterschiedliche Varianten der beiden Staaten finden. Um zu untersuchen, welcher dieser Fälle der günstigere ist, haben wir zunächst an die Binomialverteilung erinnert, bevor wir uns mit dem Poissonprozess und den Markov-Ketten beschäftigt haben. Das Ziel dieser Gruppenarbeit war es also, das Problem der Warteschlangen in Postämtern mit Hilfe von mathematischen Modellen zu untersuchen.

1 Stochastischer Prozess

Zunächst stellt sich die Frage nach dem Begriff des „Stochastischen Prozesses“.

Definition: Ein stochastischer Prozess ist eine Familie von Zufallsgrößen (X_t), die auf dem gleichen Wahrscheinlichkeitsraum definiert sind und Werte in einem Zustandsraum E annehmen. Die Variable t stellt dabei häufig die Zeit dar.

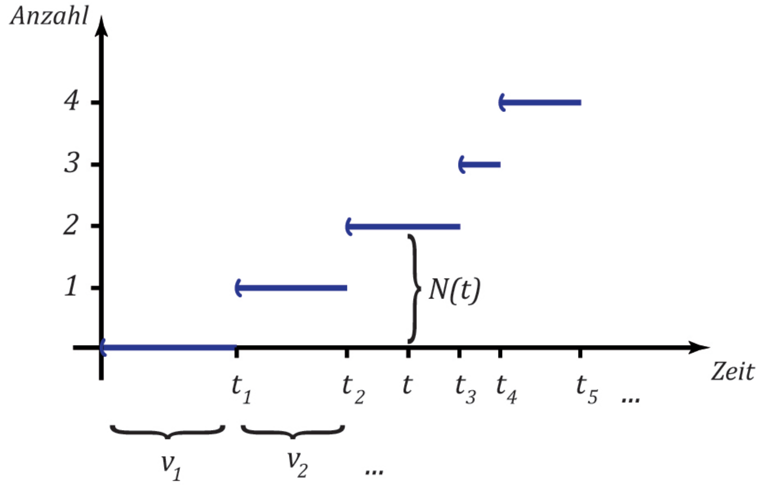
Das einfachste Beispiel eines stochastischen Prozesses ist die Bernoulli-Kette. Diese ist ein Zufallsversuch, bei dem es zu jedem Zeitpunkt $t = 1, 2, \ldots, n$ nur relevant ist, ob ein gewisses Ereignis (Treffer) eintritt oder nicht (z.B. Treffer $= 1$; Niete $= 0$). Dabei gilt für die Wahrscheinlichkeit, k Treffer bei n Zufallsversuchen mit der Trefferwahrscheinlichkeit p zu erzielen:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Die Teilversuche einer Bernoulli-Kette sind voneinander unabhängig. Damit ist sie das einfachste Beispiel einer Markov-Kette.

2 Poissonprozess

Der Poissonprozess $(N(t))$ ist ein diskreter stochastischer Prozess in stetiger Zeit, d.h. $N(t) \in E \subset \mathbb{N}, t \geq 0$ und $t \in \mathbb{R}$. Ein typischer Verlauf sieht so aus:



Hierbei bezeichnet v_i die i-te Wartezeit und $N(t)$ die Anzahl der Ankünfte bis zum Zeitpunkt t.

- **Annahme 1**: Der Prozess ist stochastisch zeitinvariant, d.h.

$$P(N(t_1+c)-N(t_1)=k)=P(N(t_2+c)-N(t_2)=k)$$

  für alle $t_1, t_2, c \geq 0$.

- **Annahme 2**: Die Wartezeiten zwischen zwei Ereignissen sind unabhängige Zufallsvariablen.

- **Annahme 3**: Die Anzahlen der Ereignisse in disjunkten Intervallen sind unabhängige Zufallsvariablen. Der Erwartungswert $E(N(1))$ der Anzahl der Ereignisse pro Zeiteinheit heißt Intensität und wird mit λ bezeichnet.

Der Prozess startet mit $N(0)=0$. Uns interessiert die Wahrscheinlichkeit

$$P(N(t)=k) \text{ für } k=1,2,3,\ldots.$$

Um diese Wahrscheinlichkeit zu berechnen, wählen wir ein beliebiges aber festes t. Das Intervall $I=[0;t]$ wird in n disjunkte Teilintervalle der Länge $\frac{t}{n}$ zerlegt. Für die Anzahl N_i der Ereignisse in jedem Teilintervall gilt dann:

$$N_i = N\left(\frac{t}{n}\right) = \begin{cases} 0 \\ 1 \\ >1. \end{cases}$$

Wir definieren das Ereignis A als $\bigcap\limits_{i=1}^{n}\{N_i \leq 1\}$. Das Gegenereignis $\overline{A}$ tritt also ein, sobald ein i mit $N_i > 1$ existiert. Ziel ist es sicherzustellen, dass in einem sehr kurzen Teilintervall maximal ein Ereignis stattfindet, um das Problem mathematisch beherrschen zu können. Deshalb treffen wir folgende

- (Regularitäts-)**Annahme 4**:
  Für $h \to 0$ gilt: $P(N(h)>1)=o(h)$[1]

Für $n \to \infty$ liefert Annahme 4: $P(N\left(\frac{t}{n}\right)>1)=o(\frac{t}{n})$.

Aus den Annahmen 1 bis 3 folgt

$$\begin{aligned} P(N(t)=k) &= P(N_1+N_2+\ldots+N_n=k) \\ &= P(\{N_1+N_2+\ldots+N_n=k\}\cap A) \\ &\quad +P(\{N_1+N_2+\ldots+N_n=k\}\cap \overline{A}). \end{aligned}$$

[1]Ein Term $t(h)$ heißt von der Größenordnung $o(h)$ für $h \to 0$, falls $\lim\limits_{n\to\infty} \frac{t(h)}{h}=0$.

Der zweite Summand ist wegen Annahme 4 von der Größenordnung $o(h)$. Deshalb betrachten wir nur den ersten Summanden. In diesem Summanden sind die N_i Bernoulli-Variablen und damit ist ihre Summe binomialverteilt mit n und $p = \frac{\lambda t}{n}$.

Daraus ergibt sich dann für diese Verteilung

$$P(N(t) = k) = \binom{n}{k} \left(\lambda \cdot \tfrac{t}{n}\right)^k \cdot \left(1 - \lambda \cdot \tfrac{t}{n}\right)^{n-k} + o(1)$$

und für den Grenzwert bei $n \to \infty$

$$P(N(t) = k) \quad = \quad \tfrac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

Anschließend werden die Wahrscheinlichkeiten für die Länge der Wartezeiten in einem Poissonprozess gesucht. Zunächst gilt, dass $N(t) - N(s)$ und $N(t-s)$ für $0 < s < t$ identisch verteilt sind und für $0 < s_1 < s_2 < s_3 < \ldots$ sind $N(s_1), N(s_2 - s_1), \ldots$ unabhängig.

Die Wartezeiten $v_1, v_2, ...$ sind unabhängig und identisch verteilt, mit

$$P(v_1 \leq x) = 1 - e^{-\lambda \cdot x},$$

da

$$P(t_n \leq t) = 1 - P(t_N > t) = 1 - P(N(t) < n) = 1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}.$$

3 Markov-Ketten

Eine Markov-Kette ist ein stochastischer Prozess, bei dem das zukünftige Verhalten nur vom aktuellen Zustand und nicht von vorherigen Zuständen abhängt.

Definition: $X(t)$ sei der Zustand des Systems zum Zeitpunkt t, $t \in \mathbb{R}$, $X(t) \in E \subset \mathbb{N}$.

$(X(t))$ heißt Markov-Kette mit dem Zustandsraum E genau dann, wenn:

$$\begin{aligned} P(X(t_n) = i_n \mid X(t_1) = i_1, X(t_2) = i_2, ..., X(t_{n-1}) = i_{n-1}) \\ = P(X(t_n) = i_n \mid X(t_{n-1}) = i_{n-1}) \end{aligned} \tag{1}$$

für alle endlichen Auswahlen von Zeitpunkten $t_1 < t_2 < \ldots < t_n$ und Zuständen gilt.

Eine Markov-Kette ist durch ihre **Anfangsverteilung**: $\{p_i = P(X(0) = i)\}$ und die **Übergangswahrscheinlichkeiten** $\{p_{ij}(s,t) = P(X(t) = j \mid X(s) = i)\}$ eindeutig bestimmt. Aus diesen Bestimmungsstücken kann insbesondere für jedes t und jedes k die Wahrscheinlichkeit $P(X(t) = k)$ berechnet werden.

Der Poissonprozess als Markov-Kette

Sei $(N(t))$ ein Poissonprozess. Aus der Annahme 3 und der Poissonwahrscheinlichkeit folgt:

$$\begin{aligned}
&P(N(t_n) = i_n \mid N(t_1) = i_1, N(t_2) = i_2, \ldots, N(t_{n-1}) = i_{n-1}) \\
&= P(N(t_n) - N(t_{n-1}) = i_n - i_{n-1} \mid N(t_{n-1}) = i_{n-1}) \\
&= P(N(t_n) = i_n \mid N(t_{n-1}) = i_{n-1}) \\
&= \frac{(\lambda \cdot (t_n - t_{n-1}))^{i_n - i_{n-1}}}{(i_n - i_{n-1})!} \cdot e^{-\lambda(t_n - t_{n-1})}.
\end{aligned}$$

Somit ist der Poissonprozess eine Markov-Kette und außerdem hängt die Übergangswahrscheinlichkeit $p_{ij}(s,t)$ nur von der Zeitdifferenz $t-s$ ab und nicht von der Lage der Zeitpunkte s und t auf der Achse. Diese Eigenschaft heißt **Zeithomogenität**.

Für eine homogene Markov-Kette liefert die Formel für die totale Wahrscheinlichkeit

$$P(X(t) = k) = \sum_{i \in E} P(X(t) = k | X(0) = i) \cdot P(X(0) = i)$$

$$p_k(t) = \sum_{i \in E} p_{ik}(t) \cdot p_i.$$

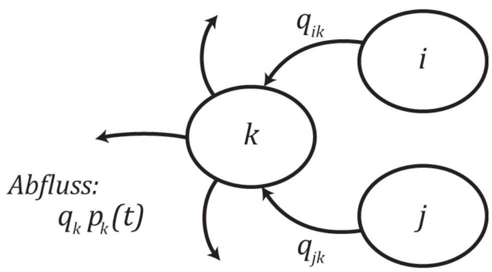
Um die zeitliche Entwicklung durch ein System von Differentialgleichungen beschreiben zu können, müssen wir das Verhalten in kurzen Zeitintervallen erfassen. Dies geschieht durch die **Intensitäten**:

$$q_{ij} = \lim_{h \to 0} \frac{p_{ij}(h)}{h} = p'_{ij}(0),$$

$$q_i = \lim_{h \to 0} \frac{1 - p_{ii}(h)}{h} = -p'_{ii}(0),$$

wobei q_{ij} die Intensität ist, mit der das System vom Zustand i in den Zustand j wechselt, und q_i die Intensität ist , mit der das System den Zustand i verlässt.

Durch sogenannte Übergangsgraphen werden die möglichen Übergänge mit ihren Intensitäten veranschaulicht:



Eine homogene Markov-Kette mit $\sum\limits_{i \neq j} q_{ij} = q_i$ heißt **konservativ**.

Der Poissonprozess ist konservativ, weil $q_i = q_{ii+1} = \lambda$ und für alle anderen Übergangsintensitäten $q_{ik} = 0$ gilt.

Wir wollen nun eine Differentialgleichung für die Wahrscheinlichkeiten $p_k(t)$ herleiten. Dazu teilen wir das Intervall $[0; t+h]$ in die Teilintervalle $[0; t]$ und $[t; t+h]$. Dann ergibt sich für die Wahrscheinlichkeit zum Zeitpunkt $t + h$ im Zustand k zu sein:

$$p_k(t+h) = \sum_{i \in E} p_{ik}(h) \cdot p_i(t).$$

Nach einigen Umformungen ergibt sich die Differentialgleichung:

$$p_k'(t) = -p_k(t) \cdot q_k + \sum_{i \in E \setminus \{k\}} p_i(t) \cdot q_{ik}$$

mit der Anfangsbedingung

$$p_k(0) = p_k.$$

Stationäre Prozesse

Ein stochastischer Prozess $(X(t))$ heißt stationär, wenn sich alle endlich-dimensionalen Verteilungen bei einer Verschiebung auf der Zeitachse um einen beliebigen, aber festen Betrag h nicht ändern.

Wenn wir nur einen festen Zeitpunkt t betrachten, gilt insbesondere

$$P(X_{t+h} = x) = P(X_t = x).$$

Ist die stationäre Verteilung von der Anfangsverteilung unabhängig, so nennt man den Prozess **ergodisch**. Wenn Ergodizität vorliegt, folgt aus der obigen

Differentialgleichung wegen $p'_k(t) = 0$ für die stationäre Verteilung (x_k) das lineare Gleichungssystem

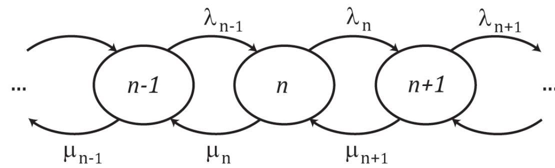
$$0 = -x_k \cdot q_k + \sum_{i \in E \setminus \{k\}} x_i \cdot q_{ik} \qquad \text{mit} \qquad \sum_{k=0}^{\infty} x_k = 1.$$

4 Bedienungsmodelle

Mathematisch gesehen kann man die Bedienungstheorie als die stochastische Charakterisierung von bestimmten Warteschlangen deuten. Dabei besteht ein solches Bedienungsmodell stets aus einem oder mehreren Bedienungskanälen, welche wir im Folgenden Schalter nennen werden. Als Beispiele können die Schalter auf dem Postamt, die Kassen im Supermarkt oder die Telefonleitungen eines Kundenservice genannt werden. Aus diesen Bedienungssystemen geht hervor, dass hinzukommende Kunden jeweils einen Schalter in Anspruch nehmen möchten, weshalb sich schließlich durch die Bedienungszeit eine sogenannte Wartezeit für die Kunden ergibt. Dies kennt man auch vom Check-In auf Flughäfen oder von Bestellungen in einem Restaurant.

Um die Frage zu klären, wann das Bedienungssystem am kundenfreundlichsten ist bzw. wann die Gesamtbedienungszeit bei endlich vielen Kunden einen möglichst kleinen Wert annimmt, muss man die Faktoren nennen, die ein solches Modell beeinflussen und charakterisieren. Auf der einen Seite hängt das System von dem Kundenstrom ab, auf der anderen Seite aber auch von der Bedienungsanlage. Das heißt also, die Anzahl der Kunden k, die Bedienungszeit t, die Anzahl der Schalter s und die Art der Warteschlangen (je Schalter eine Schlange oder für alle Schalter genau eine Warteschlange) bestimmen, wie letztendlich das Bedienungsmodell aussieht.

Für die von uns betrachteten Systeme ist der Forderungenstrom ein Poissonprozess mit der Intensität λ. Die Bedienungszeit eines Kunden nehmen wir als exponentialverteilt mit dem Parameter μ an. μ beschreibt die Intensität, mit der ein Kunde das System verlässt. Wir modellieren das ganze System als sogenannte **Geburts- und Todesprozess**. Solche Prozesse sind konservative Markov-Ketten, bei denen nur die Übergänge von k nach $k+1$ (Geburt) und von k nach $k-1$ (Tod) möglich sind. Die zugehörigen Intensitäten bezeichnen wir mit $q_{k,k+1} = \lambda_k$ und $q_{k,k-1} = \mu_k$



Das lineare Gleichungssystem für die stationäre Verteilung nimmt dann folgende Gestalt an:

$$(\mu_k + \lambda_k)x_k = \mu_{k+1}x_{k+1} + \lambda_{k-1}x_{k-1}$$

Wie in der obigen Abbildung zu erkennen ist, stellt der Term $(\mu_k + \lambda_k)x_k$ den Abfluss aus dem Zustand k und $\mu_{k+1}x_{k+1} + \lambda_{k-1}x_{k-1}$ den Zufluss in den Zustand k dar.

Wenn man nun alle Gleichungen bis zu Nummer k addiert erhält man

$$\sum_{i=0}^{k}(\mu_i + \lambda_i)x_k = \sum_{i=0}^{k}\mu_{i+1}x_{i+1} + \sum_{i=0}^{k}\lambda_{i-1}x_{i-1}.$$

und

$$\lambda_0 x_0 = \mu_1 x_1.$$

Umgeformt ergibt sich dann

$$\mu_1 x_1 + \lambda_k x_k = \lambda_0 x_0 + \mu_{k+1}x_{k+1}$$

bzw.

$$\lambda_k x_k = \mu_{k+1}x_{k+1}.$$

Wenn man diese Gleichung nach x_{k+1} umstellt, erhält man

$$x_{k+1} = \frac{\lambda_k}{\mu_{k+1}}x_k.$$

Die Auflösung dieser rekursiven Gleichung liefert

$$x_k = (\prod_{i=0}^{k}\frac{\lambda_i}{\mu_{i+1}}) \cdot x_0$$

mit der Anfangsbedingung

$$x_0 = (\sum_{k=1}^{\infty}\prod_{i=0}^{k}\frac{\lambda_i}{\mu_{i+1}} + 1)^{-1}.$$

Beispiele für Bedienungssysteme

Die von uns betrachteten Bedienungssysteme werden mit der Schreibweise $M|M|s|w$ beschrieben, wobei M den Poissonschen Forderungsstrom, M die exponentialverteilte Bedienzeit mit dem Parameter μ, s die Anzahl der Schalter und w die Anzahl der Plätze in der Warteschlange beschreiben. Untersuchen wir die Bedienungssysteme $M|M|1$ und $M|M|2|2$, wobei im ersten Fall ein unendlich großer Warteraum zur Verfügung steht. Für das Verhältnis der Intensitäten λ und μ schreiben wir $\frac{\lambda}{\mu} = \rho$. Für $M|M|1$ erhalten wir aus den obigen allgemeinen Gleichungen die stationäre Verteilung

$$\begin{aligned} x_0 &= 1 - \rho \\ x_k &= (1-\rho)\rho^k, \qquad k = 1, 2, \ldots . \end{aligned}$$

Im System $M|M|2|2$ gibt es 2 Schalter zur Bearbeitung eintreffender Kunden, außerdem bestehen nur für 2 Kunden Warteplätze. Das heißt also, sobald mehr Kunden ankommen, verlassen sie das System.

$$\begin{aligned} x_0 &= \frac{1-\frac{\rho}{2}}{1+\frac{\rho}{2}-\rho(\frac{\rho}{2})^4} \\ x_1 &= \rho x_0 \\ x_2 &= \frac{1}{2}\rho^2 x_0 \\ x_3 &= \frac{1}{4}\rho^3 x_0 \\ x_4 &= \frac{1}{8}\rho^4 x_0 \end{aligned}$$

Aus dieser stationären Verteilung kann man Systemkenngrößen wie die mittlere Schlangenlänge und die mittlere Wartezeit berechnen.

5 Die Simulation

Bei wachsender Schalter- und Warteplatzanzahl wachsen auch die zugehörigen Gleichungssysteme quadratisch bezüglich der Platzanzahl, was herkömmliche Software bald überfordert.
Um die ursprüngliche Frage zu klären, müssen wir aber auch Fälle mit mehr als 2 Schaltern und 2 Warteplätzen betrachten. Damit gewinnt das **Werkzeug Simulation** schnell an Attraktivität. In Ermangelung einer Simulationssoftware wurde eine eigene Variante implementiert, deren grundlegende Überlegungen im Folgenden vorgestellt werden.

Wahrscheinlichkeiten

Im Bedienungssystem liegt ein poissonverteilter Zustrom an Kunden und eine exponentialverteilte Reihe von Bedienungszeiten vor. Bei der Implementierung bestand aber nur die Möglichkeit, auf gleichverteilte Zufallszahlen aus $[0;1]$ zurückzugreifen.
Um den Kundenstrom zu simulieren, werden die Wahrscheinlichkeiten für genau 0, maximal 1, maximal 2, usw. bis maximal 5 Kundenankünfte innerhalb einer Zeiteinheit bestimmt:

$$P(X \leq k) = \sum_{i=0}^{k} \frac{\lambda^i \cdot e^{-\lambda}}{i!}.$$

Da die Wahrscheinlichkeit für die Ankunft von mehr als 5 Kunden bei kleinem λ verschwindend gering ist, werden diese Ereignisse vernachlässigt. Wir haben nun das Intervall [0;1] in 7 Teilintervalle zerlegt:

$$[0; P(X=0)], (P(X=0); P(X=1)], ..., (P(X=5), 1].$$

Je nachdem, in welchem Intervall die gleichverteilte Zufallszahl nun liegt, werden 0 bis 6 neue Kunden simuliert.
Eine exponentialverteilte Bedienungszeit X lässt sich durch die Umkehrung der Formel $Y = 1 - e^{-\mu X}$, also $X = -\frac{\ln(1-Y)}{\mu}$ erzeugen. Wenn Y gleichverteilt auf $[0;1]$ ist, dann ist

$$X = -\frac{\ln Y}{\mu}$$

exponentialverteilt mit dem Parameter μ.

Ergebnisse

Zur Validierung der Software konnten wir auf das gerechnete Beispiel $M|M|2|2$ bei $\lambda = 2$ und $\mu = 3$ bauen. Der dort erhaltene Erwartungswert der Wartezeit von 0,36 Zeiteinheiten stimmte annähernd mit der simulierten mittleren Wartezeit überein. Auch beim deutschen System stellte sich in der Simulation dieses kleinen Systems etwa dieser Wert ein. Als größere Variante wurde nun $M|M|2|40$ simuliert. Nach 300 bearbeiteten Kunden stellte sich eine stabile mittlere Wartezeit von etwa 5,8 Zeiteinheiten beim englischen und etwa 6,3 beim deutschen System ein.

Literatur

[1] F. Topsøe: Spontane Phänomene. Stochastische Modelle und ihre Anwendungen. Vieweg, Braunscheig, 1990.

[2] E. Warmuth: Mathematische Modelle diskreter stochastischer Systeme. Ministerium für Bildung, Wissenschaft und Weiterbildung, Land Rheinland-Pfalz, 1997.