

Ein Warteschlangen-Problem

Robert Märki

Grosse Fastfood Restaurantketten spielen eine führende Rolle bei der Verwendung mathematischer Modelle zur Optimierung von Betriebsabläufen und die Ergebnisse der Studien führten zu organisatorischen und baulichen Veränderungen, welche zum Ziel hatten, den Kundenfluss und damit den Profit zu steigern. Das hier diskutierte Modell beruht auf einer vereinfachten modifizierten Version einer von Burger King durchgeführten Studie [1]. Es ist sehr geeignet als Abschlussprojekt in einem Gymnasium, werden doch viele wichtige Themenbereiche aus dem Unterricht der letzten zwei Jahre benötigt. Das Projekt ermöglicht es, ein sehr reizvolles Thema zu untersuchen und gleichzeitig wichtige Stoffgebiete zu repetieren und anzuwenden. Das kleine physische Simulationsspiel zur Einführung, welches mit einer Klasse vielleicht etwa 20 Minuten lang gespielt werden kann, gibt ein Gefühl für den Ablauf im System und kann durch das überraschende Ergebnis die Neugier wecken.

Warteschlangen

Das Studium von Warteschlangen ist ein wichtiges Anwendungsgebiet der Mathematik. In einem Warteschlangen-System gibt es eine oder mehrere Servicestellen, wo die ankommenden Kunden bedient werden. Die Kunden kommen zu zufälligen Zeitpunkten an und die Bedienung eines Kunden dauert eine zufällige Zeitspanne. Währendem ein Kunde bedient wird, muss ein neuer Kunde in einer Warteschlange warten. Ein bedienter Kunde wechselt zu einer anderen Servicestelle oder verlässt das System. Im vorliegenden Beispiel wird ein sehr einfaches System mit einer einzigen Warteschlange mit begrenzt vielen Plätzen untersucht.

Das Simulationsspiel

Das Spiel dient der Simulation des Warteschlangenproblems, welches nachfolgend mathematisch modelliert wird. Wenn dieses Spiel von einer Klasse etwa 20 Minuten lang gespielt wird, entwickelt sich ein gutes Gefühl für die Abläufe und für die Problematik und die mathematische Modellbildung wird dann auch besser verstanden. Die statistische Auswertung der Spielergebnisse kann zudem mit den theoretischen Modellrechnungen verglichen werden.

Spielbeschreibung

Ein Drive-In Restaurant liegt an einer Strasse. Vor dem Schalter hat es Platz für drei Autos. Wenn diese drei Plätze besetzt sind, kann ein neuer potenzieller Kunde nicht in die Warteschlange einbiegen und fährt weiter (zur Konkurrenz). Wir führen die folgenden Grössen ein:

Ankommensrate λ

Im Mittel kommt pro Minute ein Kunde an, dies wird mit einem Würfel simuliert: Jeder Wurf steht für ein Zeitintervall von 10 Sekunden, wird die Augenzahl 6 gewürfelt, heisst dies, dass ein Kunde kommt. Auf diese Weise wird gewährleistet, dass die Kunden zwar in zufälligen Zeitabständen kommen, jedoch im Mittel einer pro Minute. Mathematisch ausgedrückt ist $\lambda = 1$ Kunde pro Minute.

Service rate μ :

Die Bedienung der Kunden dauert natürlich verschieden lang, einige haben kleine Bestellungen, die sofort erledigt werden können, andere brauchen mehr Zeit. Es gibt jedoch einen mittleren Wert der Anzahl Kunden, welche pro Minute resp. pro Stunde abgefertigt werden können. Diese Grösse heisst *Service rate*. Die Service rate kann von der Firma durch organisatorische und bauliche Massnahmen beeinflusst werden. Eine Erhöhung der Service rate verlangt aber mehr oder besser ausgebildetes Personal.

Das Spiel wird in zwei Versionen gespielt:

Version I:

Die Service rate wird der Ankommensrate angepasst, d.h. es wird im Mittel pro Minute ein Kunde bedient. Dies wird wiederum mit einem Würfel simuliert, ein Wurf steht für 10 Sekunden Dauer, wird die Augenzahl 6 gewürfelt heisst dies, dass ein Service abgeschlossen wurde und der Kunde das System verlässt. Auf diese Weise dauert die Servicezeit im Mittel 1 Minute, sie kann aber auch länger oder kürzer sein.

Version II:

Die Service rate wird verdoppelt, was höhere Personalkosten bedeutet. Simuliert wird dies dadurch, dass bei Augenzahl 5 oder 6 ein Service vollendet wird.

Spielanleitung

Das Spiel wird von zwei Personen gespielt, Spieler A ist Kunde, Spieler B Servicepersonal. Zu Beginn befindet sich ein Wagen in der Warteschlange

A und B würfeln abwechselungsweise, B beginnt. Würfelt A die Augenzahl 6, dann reiht sich ein weiterer Wagen in die Warteschlange ein, falls dort ein freier Platz vorhanden ist. Zur Verfügung stehen drei Plätze Würfelt B die Augenzahl 6 (Version I) resp. die Augenzahl 5 oder 6 (Version II), dann wurde ein Service abgeschlossen und ein Wagen wird aus der Warteschlange entfernt. Natürlich kann es Situationen geben, wo kein Wagen in der Warteschlange steht.

Ein Warteschlangenproblem

Robert Märki

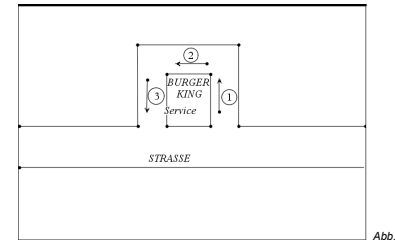


Abb. 1

Auswertung:

Wichtig ist, die genaue Anzahl der Würfe von A (oder von B) zu ermitteln, denn dies entspricht der simulierten Zeitspanne. Z.B. kann man in jeder Version 120 Mal würfeln und simuliert so eine Zeitspanne von je 20 Minuten. Lässt man eine ganze Klasse spielen, so simuliert dies eine Zeitspanne von einigen Stunden. A macht also bei jedem Wurf einen Strich in einer Liste, B jedesmal, wenn ein Kunde bedient wurde, d.h. wenn er ein Fahrzeug aus der Schlange entfernen konnte.

Berechnung des Gewinns: Jeder Kunde gibt im Mittel $c_1 = 10$ Euro aus, die Servicekosten betragen in jeder Minute c_2 Service rate mit $c_2 = 1$ Euro, d.h. in Version I kostet der Service 1 Euro pro Minute, in Version II dagegen 2 Euro pro Minute. Man bestimme den mittleren Gewinn pro Minute in beiden Versionen. Die Schülerinnen und Schüler vermuten in der Regel dass Version I besser ist, weil es vernünftig scheint, die Service rate gleich gross wie die Ankommensrate zu wählen. Das Spiel zeigt dann aber, dass bei Version I die Warteschlange recht häufig voll ist und damit Kunden an die Konkurrenz verloren gehen.

Im Folgenden werden zwei Modelle betrachtet, ein diskretes, welches direkt das Simulationsspiel widerspiegelt, eines allgemeineres kontinuierliches. Je nach fachlicher Voraussetzung kann auch nur eines der Modelle betrachtet werden. Die für die Berechnung des Gewinns entscheidende Grösse ist die *Eintrittsrate*, diese ist, weil einige ankommende Kunden nicht in die Warteschlange eintreten können, kleiner als die Ankommensrate. Um die Eintrittsrate zu berechnen, muss die Wahrscheinlichkeit, dass die Warteschlange voll ist (d.h. die maximal mögliche Länge hat), bekannt sein.

Das diskrete Modell

Sei $P_n(t)$, $n = 0, 1, 2, 3 \dots$ die Wahrscheinlichkeit, dass zur Zeit t genau n Fahrzeuge in der Schlange sind. Die Zeit t ist diskret, $t = 0, 1, 2, 3 \dots$ (Zeiteinheit 10 Sekunden). Es gilt Version I.

$$P_0(t+1) = P_0(t) \frac{5}{6} + P_1(t) \frac{1}{6} \frac{5}{6}$$

$$P_1(t+1) = P_0(t) \frac{1}{6} + P_1(t) \left(\frac{1}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \right) + P_2(t) \frac{1}{6} \frac{5}{6}$$

$$P_2(t+1) = P_1(t) \frac{5}{6} \frac{1}{6} + P_2(t) \left(\frac{1}{6} \cdot \frac{1}{6} + \frac{5}{6} \cdot \frac{5}{6} \right) + P_3(t) \frac{1}{6} \frac{5}{6}$$

$$P_3(t+1) = P_2(t) \frac{5}{6} \frac{1}{6} + P_3(t) \left(\frac{1}{6} \cdot \frac{1}{6} + \frac{5}{6} \right)$$

Das ergibt die Matrix:

$$m_1 = \frac{1}{36} \begin{pmatrix} 30 & 5 & 0 & 0 \\ 6 & 26 & 5 & 0 \\ 0 & 5 & 26 & 5 \\ 0 & 0 & 5 & 31 \end{pmatrix}$$

Version II ergibt analog die Matrix:

$$m_2 = \frac{1}{36} \begin{pmatrix} 30 & 10 & 0 & 0 \\ 6 & 22 & 10 & 0 \\ 0 & 4 & 22 & 10 \\ 0 & 0 & 4 & 26 \end{pmatrix}$$

Die Eigenvektoren zum Eigenwert 1 sind:

eig1	0,217391
	0,26087
	0,26087
	0,26087
eig2	0,516529
	0,309917
	0,123967
	0,049587
2/99	

Abb. 2

Bemerkung: TI-Nspire™ liefert den normierten Eigenvektor mit der Norm 1, der Eigenvektor muss noch angepasst werden, so dass die Summe der Elemente 1 ergibt.

Der mittlere Gewinn pro Minute beträgt demnach (Parameterwerte wie im Spielbeschreibung):

$$\text{Version I: } G_I = 10 \cdot (1 - 0,26087) - 1 = 6,39$$

$$\text{Version II: } G_{II} = 10 \cdot (1 - 0,049587) - 2 = 7,50$$

Bemerkungen: Anstelle der Eigenvektoren kann man auch (für die Schüler leichter nachvollziehbar) die Entwicklung von $(m_1)^n$ resp. $(m_2)^n$ beobachten, , daraus lässt sich die Wahrscheinlichkeit, dass die Warteschlange im Langzeitverhalten voll ist, leicht ablesen.

$n \cdot 2^0$	0,22143	0,219004	0,21997	0,219834
	0,263165	0,261957	0,260062	0,258077
	0,259164	0,260062	0,26147	0,262408
	0,256241	0,258077	0,262408	0,265219
$n \cdot 1^{100}$	0,217433	0,217413	0,217377	0,217362
	0,260893	0,260881	0,260861	0,260847
	0,260852	0,260861	0,260876	0,260886
	0,260823	0,260847	0,260886	0,260914
$n \cdot 2^{50}$	0,516583	0,516533	0,516477	0,516361
	0,30992	0,309818	0,309911	0,309904
	0,123958	0,123964	0,124035	0,124118
	0,049561	0,049588	0,049647	0,04972
$n \cdot 2^{100}$	0,516529	0,516529	0,516529	0,516529
	0,309917	0,309917	0,309917	0,309917
	0,123967	0,123967	0,123967	0,123967
	0,049587	0,049587	0,049587	0,049587
4/99				

Abb. 3

Im Spiel beginnt in jedem 10-Sekunden Intervall jeweils Spieler B.

Bemerkung: Wenn jeweils Spieler A beginnt, dann ergibt sich interessanterweise ein anderes Resultat, dies wegen der "Randbedingung": Wenn die Warteschlange voll ist, dann tritt ein ankommendes Fahrzeug nicht ein und ist als Kunde verloren. Die Wahrscheinlichkeit dafür ist bedeutend kleiner, wenn A beginnt.

Selbstverständlich kann nun das diskrete Modell angepasst und verändert werden für andere Werte der vorkommenden Parameter.

Das allgemeine kontinuierliche Modell

Ausgangslage: Ein Drive-In-Restaurant hat K Plätze in der Warteschlange. Die Autos, welche sich in die Schlange einreihen wollen, kommen in zufälligen Zeitabständen an mit einer mittleren Ankommensrate von λ Autos pro Minute. Die Servicezeit ist ebenfalls zufällig, die mittlere Servicerate sei μ Autos pro Minute. Im Mittel gibt ein Kunde c₁ Euro aus, die Servicekosten des Betriebs belaufen sich auf c₂·μ Euro pro Minute. Bei welcher Servicerate ist der Gewinn maximal?

1. Ankommens- und Service-Wahrscheinlichkeit

Wenn z.B. λ = 1 Auto/Minute beträgt, dann ist die Wahrscheinlichkeit, dass während einer beliebigen Sekunde ein Auto ankommt 1/60. Im Simulationsspiel war die Wahrscheinlichkeit, dass ein Auto in einem 10-Sekunden Intervall ankommt 1/6.

Allgemein gilt: Die Wahrscheinlichkeit, dass ein Auto in einem kleinen Zeitintervall Δt ankommt und damit in die Warteschlange eintritt, beträgt λ·Δt (mathematisch exakt gilt dies nur für Δt→0). Entsprechend gilt, dass die Wahrscheinlichkeit, dass ein Kunde in einem kleinen Zeitintervall Δt die Warteschlange verlässt (sofern es Kunden in der Schlange hat), μ·Δt beträgt.

2. Wahrscheinlichkeitsverteilung der Länge der Warteschlange

Um das "Warteschlangen-Problem" zu lösen, müssen wir die Wahrscheinlichkeitsverteilung der Schlängellänge verstehen.

Definition: P_n(t) := Wahrscheinlichkeit, dass zu Zeit t die Zahl der Autos in der Schlange n beträgt.

Für P_n(t) können wir Differenzialgleichungen aufstellen. Wir werden jedoch diese Differenzialgleichungen nicht lösen (obwohl dies möglich, jedoch aufwändig schwierig ist), sondern nur das Langzeitverhalten für t→∞ studieren.

3. Differenzialgleichungen für die Schlängellänge

Um die Differenzialgleichungen aufzustellen, ist folgendes zu beachten: Die Wahrscheinlichkeit eines Ereignisses (Ankunft eines Kunden, Wegfahrt eines Kunden) in einer kleinen Zeitspanne Δt beträgt λ·Δt resp. μ·Δt. Die Wahrscheinlichkeit, dass in einer kleinen Zeitspanne Δt zwei Ereignisse stattfinden ist dann proportional zu (Δt)². Da

wir beim Aufstellen des Differenzialquotienten dann durch Δt dividieren und anschliessend den Grenzwert für Δt→0 betrachten, werden diese Wahrscheinlichkeiten vernachlässigbar, d.h. es genügt, in jedem kleinen Zeitintervall Δt nur Einfachereignisse zu betrachten: Es findet jeweils höchstens ein einziges einzelnes Ereignis, Ankunft oder Wegfahrt eines Kunden, statt.

Stellen Sie nun die Differenzialgleichungen auf. Dazu berechnen Sie zuerst P_n(t+Δt) und gewinnen dann daraus einen Ausdruck für P_n(t+Δt) - P_n(t) rep. dann

$$\frac{P_n(t+\Delta t) - P_n(t)}{\Delta t}$$

Der letzte Ausdruck ist dann für Δt → 0 die Ableitung P'_n(t), was die Differenzialgleichung ergibt. Dabei sind drei Fälle zu unterscheiden:

(a) n = 0

Um P₀(t+Δt), also die Wahrscheinlichkeit, dass zum Zeitpunkt t+Δt kein Kunde in der Schlange ist, zu berechnen, gehen Sie wie folgt vor: Es gibt zwei Möglichkeiten, dass zum Zeitpunkt t + Δt kein Kunde in der Schlange ist:

- Zum Zeitpunkt t ist kein Kunde in der Schlange (Wahrscheinlichkeit dafür ist P₀(t)) und es tritt auch keiner ein.
- Zum Zeitpunkt t ist ein Kunde in der Schlange (Wahrscheinlichkeit dafür ist P₁(t)) und es tritt einer aus und keiner ein.

Die Differenzialgleichung für P₀(t) enthält also die Variablen λ, μ, P₀(t), P₁(t).

(b) 1 ≤ n ≤ K - 1

Jetzt gibt es drei Fälle zu unterscheiden:

- Zur Zeit t befinden sich n - 1 Kunden in der Schlange und während der Zeitspanne Δt tritt einer ein.
- Zur Zeit t befinden sich n Kunden in der Schlange und während der Zeitspanne Δt tritt keiner ein und keiner aus.
- Zur Zeit t befinden sich n + 1 Kunden in der Schlange und während der Zeitspanne Δt tritt einer aus.

(c) n = K

Jetzt gibt es wiederum nur noch zwei Fälle! Bestimmen Sie analog zu (a) die Differenzialgleichung

4. System im Gleichgewicht, Langzeitverhalten

Wir gehen davon aus, dass sich nach einiger Zeit die Wahrscheinlichkeiten P_n(t) eingespielt haben und sich nicht mehr verändern, d.h. wir studieren das Grenzverhalten für t→∞. Bezeichnen Sie nun die Langzeitwahrscheinlichkeit, dass n Kunden in der Schlange sind, mit π_n:

$$\pi_n = \lim_{t \rightarrow \infty} P_n(t)$$

Untersuchen Sie:

$$\lim_{t \rightarrow \infty} P'_n(t)$$

Aus dem System von K Differenzialgleichungen ergibt sich damit ein System von K + 1 Gleichungen in den Variablen π₀, π₁, ..., π_K. Formulieren Sie dieses System.

5. Zeigen Sie: (π_K) ist eine geometrische Folge

Zeigen Sie: Für K ≥ 3 folgt aus dem obigen Gleichungssystem

$$\pi_1 = \left(\frac{\lambda}{\mu}\right) \pi_0 \quad \pi_2 = \left(\frac{\lambda}{\mu}\right)^2 \pi_0 \quad \pi_3 = \left(\frac{\lambda}{\mu}\right)^3 \pi_0$$

Beweisen Sie mit vollständiger Induktion, dass aus dem obigen Gleichungssystem folgt:

$$\pi_n = \left(\frac{\lambda}{\mu}\right)^n \pi_0$$

Die Werte π₀, π₁, ..., π_K bilden also eine geometrische Folge mit dem Quotienten

$$q = \frac{\lambda}{\mu} (*)$$

6. Berechnung von π_n

(a) Mit Hilfe der Summenformel für die endliche geometrische Reihe und der Tatsache, dass

$$\pi_0 + \pi_1 + \dots + \pi_K = 1$$

kann π₀ durch λ und μ ausgedrückt werden. Verwenden Sie dazu den Quotienten (*). Dabei muss allerdings vorausgesetzt werden, dass q ≠ 1. Zeigen Sie, dass für q = 1 gilt, dass:

$$\pi_n = \frac{1}{K+1}$$

(b) Nachdem Sie π₀ in Abhängigkeit von q ausgedrückt haben, können Sie die Formel für π_n finden, also π_n in Abhängigkeit von q darstellen.

7. Lösung des Burger-King-Problems

π_K ist die Wahrscheinlichkeit, dass zu einem beliebigen Zeitpunkt t genau K Fahrzeuge in der Schlange sind, dass also die Warteschlange voll ist. Die Rate, mit der die Kunden in die Schlange eintreten, beträgt folglich λ·(1 - π_K) (Begründung?). Wir nennen diese Grösse "Eintrittsrate".

Wir verwenden im folgenden den Quotienten in (*) als unabhängige Variable. Der Gewinn pro Zeiteinheit ist nun: G(q) = c₁ × Eintrittsrate - c₂ × Servicerate

(a) Formulieren Sie diese Funktion. Aus dem Funktionsterm lässt sich λ ausklammern, man kann also G(q) darstellen als G(q) = λ·G*(q)

(b) Skizzieren Sie den Graphen von G*(q) für die Parameterwerte c₁=12, c₂=1, K=10 und bestimmen Sie, für welchen Wert q das Maximum von G* (und damit von G) erreicht wird. Wie gross muss also die Servicerate sein?

(c) Im Simulationsspiel waren die Werte c₁=10, c₂=1, K=3. Wo liegt hier das Maximum? Wie gross ist also die optimale Servicerate? Welcher Wert von G(q) ergibt sich für μ = 1 und für μ = 2? Vergleichen Sie mit den Ergebnissen des Spiels!

(d) Untersuchen Sie mit Hilfe von Schiebereglern den Einfluss der Parameter c₁ und K.

Lösungen

3. (a) P'₀(t) = -λ · P₀(t) + μ · P₁(t)

3. (b) P'_n(t) = λ · P_{n-1}(t) - (λ + μ) · P_n(t) + μ · P_{n+1}(t) 1 ≤ n ≤ K - 1

3. (c) P'_K(t) = λ · P_{K-1}(t) - μ · P_K(t)

4. Es ist

$$\lim_{t \rightarrow \infty} (P'_n(t)) = 0$$

und es ergibt sich das Gleichungssystem

$$0 = -\lambda \pi_0 + \mu \pi_1$$

$$0 = \lambda \pi_{n-1} - (\lambda + \mu) \pi_n + \mu \pi_{n+1} \quad 1 \leq n \leq K - 1$$

$$0 = \lambda \pi_{K-1} - \mu \pi_K$$

5. π_n = (λ/μ)ⁿ π₀

6. π_n = (1-q)/(1-q^{K+1}) · qⁿ; q = λ/μ; 0 ≤ n ≤ K

7. (a) G(q) = c₁ · λ · (1-q^K)/(1-q^{K+1}) - c₂ · λ/q = λ (c₁ · (1-q^K)/(1-q^{K+1}) - c₂/q)

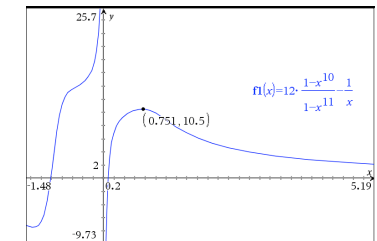


Abb. 4

7. (b) q = 0,751. Die Servicerate μ muss also etwa μ = 1,33·λ betragen.

7. (c) G(q) wird maximal für q = 0,552. Für q = 1 ist G(q) nicht definiert, jedoch beträgt der Grenzwert G(q) = 6,5 für q → 1. Für μ = 2 ist q = 0,5. G(0,5) = 7,333. Das mathematische Modell sagt also bei einer Verdoppelung der Servicerate einen um etwa 12,8% höheren Gewinn voraus. Man Vergleiche mit den Ergebnissen des Spiels!

Quellenangabe:

[1] Hastings K. J., 1997, *Queueing Systems in MAA Notes 29*, Washington, Mathematical Association of America.

Autor:

Robert Märki, Thun (Schweiz)
r.maerki@hispeed.ch