2 Prozessanalyse I: Zeiten und Bestände

2.1 Prozessorientierte Betrachtung von Warte- und Bediensystemen

Bislang haben wir uns Wertschöpfungsprozesse nur sehr abstrakt als Kombinationsprozesse für Produktionsfaktoren vorgestellt, die zur Erzeugung neuer und wertgesteigerter Güter dienen.[1] Nun wollen wir uns einige bedeutende Modellierungstechniken und Gesetzmäßigkeiten von Prozessabläufen ansehen. Dazu verwenden wir elementare Modelle und Begriffe der Analyse von Warte- und Bediensystemen. In der Abbildung 2.1 wird dazu ein Grundmodell dargestellt. 2.1

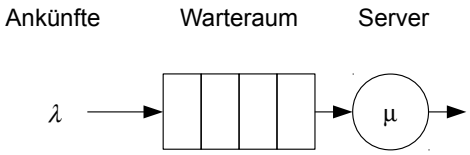

Abbildung 2.1: Grundmodell eines Warte- und Bediensystems

Die grundlegende Vorstellung besteht darin, dass im Zeitablauf *Jobs* eintreffen, die von einem als *Server* bezeichneten abstrakten (Bedien-)System bedient werden. Typischerweise gibt man zu dem Ankunftsprozess an, wie viele Jobs pro Zeiteinheit eintreffen. Das Symbol für diese sogenannte Ankunftsrate ist λ. Sie ist der Kehrwert der mittleren Zwischenankunftszeit. Die Bedienrate μ eines Servers ist analog der Kehrwert der mittleren Bearbeitungs- oder Bedienzeit.

Sind der oder die Server belegt, so muss der Job in einem Warteraum oder Eingangspuffer warten. Diese Betrachtungsperspektive ist offenbar recht allgemein. Die Jobs können E-Mails von Kunden sein, die von Kundenbetreuern zu beantworten sind. Es kann sich auch um Patienten handeln, die in der Notaufnahme eines Krankenhauses von einem Arzt zu untersuchen und behandeln sind. Es könnten auch metallische Werkstücke sein, deren Form durch einen automatischen Fräsprozess verändert werden soll.

Wir betrachten nun drei Varianten eines einfachen Beispiels, an dem man schon einiges über das Verhalten eines derartigen System lernen kann. Dazu unterstellen wir, dass über einen Zeitraum von 16 Zeiteinheiten insgesamt vier Jobs bearbeitet werden. Die drei Fälle

[1] Vgl. Abbildung 1.1 auf Seite 3.

unterscheiden sich jedoch hinsichtlich der Zwischenankunftszeiten sowie der Bedienzeiten der Jobs:

Fall 1 - Identische Zwischenankunfts- und Bedienzeiten: Nach einer Zeiteinheit kommt der erste Job an. Alle vier Zeiteinheiten trifft ein weiterer Job ein. Jeder Job hat eine Bearbeitungszeit von drei Zeiteinheiten.

Fall 2 - Verschiedene Zwischenankunftszeiten, identische Bedienzeiten: Nach zwei Zeiteinheiten kommt der erste Job an, der zweite und der dritte Job treffen gleichzeitig eine Zeiteinheit nach dem ersten Job ein, nach einer weiteren Zeiteinheit folgt dann der vierte Job. Die Bearbeitungszeiten betragen weiterhin jeweils drei Zeiteinheiten.

Fall 3 - Verschiedene Zwischenankunfts- und Bedienzeiten: Nun treffen die Jobs erneut wie im zweiten Fall ein, allerdings betragen ihre Bearbeitungszeiten jetzt fünf, vier, eine und zwei Zeiteinheiten.

Diese Daten werden auch in den Tabellen 2.1 bis 2.3 dargestellt. Wir nehmen an, dass die Jobs in der Reihenfolge ihres Eintreffens und so rasch wie möglich von dem einen Server des Systems bearbeitet werden. Die Abbildungen 2.2 bis 2.4 zeigen, wie sich die Anzahl der Jobs im Zeitablauf entwickelt.

Tabelle 2.1: Daten zu den Jobs von Fall 1

	Job 1	Job 2	Job 3	Job 4
Ankunftszeitpunkte [ZE]	1	5	9	13
Zwischenankunftszeiten [ZE]	-	4	4	4
Bedienzeiten [ZE]	3	3	3	3
Wartezeiten [ZE]	0	0	0	0

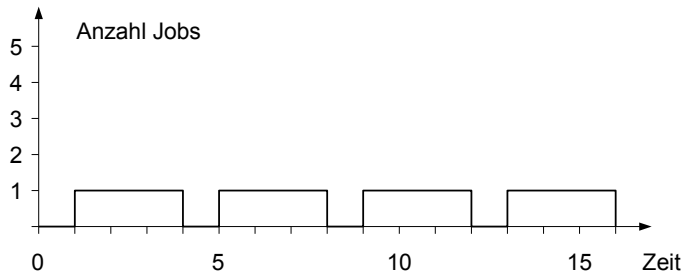

Abbildung 2.2: Anzahl der Jobs im System im Zeitablauf im Fall 1

Man erkennt deutlich, dass sich im Fall 2 und insbesondere im Fall 3 im zeitlichen Mittel mehr Jobs im System befinden. Aus der Sicht der Jobs bedeutet dies nun, dass diese im Fall 1 nie warten müssen und im Fall 3 die größte mittlere Wartezeit auftritt.

Tabelle 2.2: Daten zu den Jobs von Fall 2

	Job 1	Job 2	Job 3	Job 4
Ankunftszeitpunkte [ZE]	2	3	3	4
Zwischenankunftszeiten [ZE]	-	1	0	1
Bedienzeiten [ZE]	3	3	3	3
Wartezeiten [ZE]	0	2	5	7

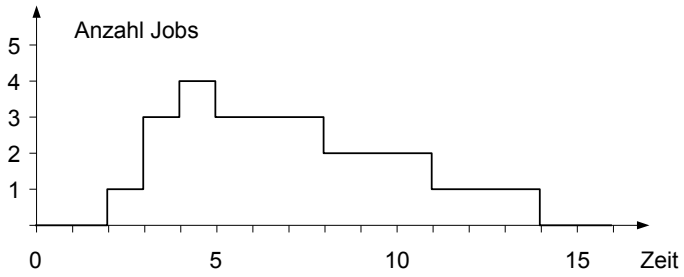

Abbildung 2.3: Anzahl der Jobs im System im Zeitablauf im Fall 2

Tabelle 2.3: Daten zu den Jobs von Fall 3

	Job 1	Job 2	Job 3	Job 4
Ankunftszeitpunkte [ZE]	2	3	3	4
Zwischenankunftszeiten [ZE]	-	1	0	1
Bedienzeiten [ZE]	5	4	1	2
Wartezeiten [ZE]	0	4	8	8

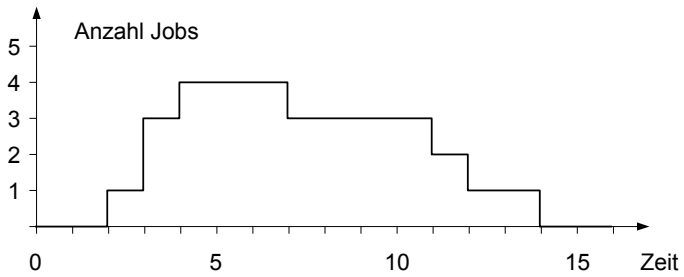

Abbildung 2.4: Anzahl der Jobs im System im Zeitablauf im Fall 3

Dies liegt aber nicht daran, dass der Server mehr zu tun hätte: In allen drei Fällen sind über den Betrachtungszeitraum von 16 Zeiteinheiten (ZE) vier Jobs angekommen. Die Summe der Bedienzeiten beträgt ebenfalls in allen drei Fällen 12 Zeiteinheiten. Damit ist der Server in allen drei Fällen offenbar zu $\frac{12}{16} = 0{,}75\%$ ausgelastet. Die allgemeine Formel für die **Auslastung** eines solchen Systems mit einem einzelnen Server lautet

$$\rho = \frac{\lambda}{\mu}. \tag{2.1}$$

Im konkreten Fall beträgt die **Ankunftsrate** $\lambda = \frac{4}{16}\ \mathrm{ZE}^{-1}$, weil im Betrachtungszeitraum von 16 Zeiteinheiten ja vier Aufträge eintreffen. Die mittlere Zeit zwischen zwei Ankünften ist also $\frac{1}{\lambda} = 4$ ZE. Die **Bedienrate** $\mu = \frac{1}{3}\ \mathrm{ZE}^{-1}$ ist der Kehrwert der mittleren Bedienzeit von 3 Zeiteinheiten. Auch auf diesem Weg erhält man den identischen Wert für die Auslastung

$$\rho = \frac{\frac{4}{16}\ \mathrm{ZE}^{-1}}{\frac{1}{3}\ \mathrm{ZE}^{-1}} = 0{,}75\%. \tag{2.2}$$

Wie kann es nun sein, dass in allen drei betrachteten Fällen der eine Server identisch stark ausgelastet ist, sich aber die Wartezeiten der Aufträge (und auch die Bestände von Jobs im System) so dramatisch unterscheiden? Die Antwort auf diese Frage liegt in der **Variabilität** des Ankunftsprozesses und des Bedienprozesses, genauer gesagt, in den (oft unvorhersehbaren) Schwankungen der Zwischenankunftszeiten und der Bedienzeiten.

2.2 Bestimmungsgrößen und Berechnungsverfahren von Wartezeiten

2.2a

Wenn Größen unvorhersehbaren Schwankungen unterliegen, so modellieren wir sie typischerweise durch **Zufallsvariablen**.[2] Stellen wir uns vor, dass T_s eine nicht-negative Zufallsvariable sei, welche die zufällige Bearbeitungsdauer (oder Servicezeit) eines Jobs mit Bedienrate μ beschreibe. Dann sei $\mathrm{E}[T_s] = \frac{1}{\mu}$ der Erwartungswert, also gewissermaßen der auf lange Sicht im Mittel zu erwartende Wert der Bearbeitungsdauer. Ferner sei σ_{T_s} die Standardabweichung der zufälligen Bearbeitungsdauer T_s. Je größer diese ist, desto stärker schwankt die Zufallsvariable.

Diese Schwankung σ_{T_s} der zufälligen Servicezeit kann man auf deren Erwartungswert $\mathrm{E}[T_s]$ beziehen, wenn man ein **relatives Maß der Variabilität** erhalten will, den sogenannten **Variationskoeffizienten**:

$$c_s = \frac{\sigma_{T_s}}{\mathrm{E}[T_s]} = \sigma_{T_s} \cdot \mu \tag{2.3}$$

Wenn man also weiß, dass eine Größe einen Variationskoeffizienten von 0 aufweist, so heißt das, dass sie *nicht* schwankt. Ist dagegen der Variationskoeffizient gleich 1, so sind

[2] Schauen Sie noch einmal in Ihre Unterlagen aus dem Kurs in Statistik und Wahrscheinlichkeitsrechnung hinein, wenn Ihnen nicht klar sein sollte, was eine Zufallsvariable ist. Eine kompakte Darstellung finden Sie z. B. in Bleymüller (2012, Kap. 7).

Erwartungswert und Standardabweichung gleich groß, was auf erhebliche Schwankungen hindeutet.

Analog kann man mit einer anderen Zufallsvariablen T_a die (zufälligen) Zwischenankunftszeiten eines Prozesses mit Ankunftsrate λ bezeichnen, mit $\mathrm{E}[T_a] = \frac{1}{\lambda}$ als Erwartungswert und entsprechend σ_{T_a} als der Standardabweichung sowie

$$c_a = \frac{\sigma_{T_a}}{\mathrm{E}[T_a]} = \sigma_{T_a} \cdot \lambda \tag{2.4}$$

als dem Variationskoeffizienten der Zwischenankunftszeiten.

Wir bezeichnen im Folgenden mit der Zufallsvariablen W_q die zufällige Wartezeit in der Warteschlange („Queue“). Im Zuge der Prozessanalyse möchten wir u. a. herausfinden, wie groß diese Zeit im Mittel ist. Dieser mittlere Wert von W_q wird durch den Erwartungswert $\mathrm{E}[W_q]$ beschrieben.

Nun betrachten wir eine kleine Serie von Simulationsexperimenten, durch die wir herausfinden wollen, wie dieser Erwartungswert der Wartezeit vor dem Server von den Mittelwerten und die Variationskoeffizienten der Zwischenankunfts- sowie Bedienzeiten der Jobs abhängt.

Wir untersuchen dazu ein einstufiges System wie in der Abbildung 2.1 auf S. 25 mit den Parameterkonstellationen in Tabelle 2.4. In dem Experiment halten wir den mittleren Wert $\mathrm{E}[T_s]$ der Servicezeit konstant und variieren zum einen die Auslastungen

$$\rho = \frac{\lambda}{\mu} = \frac{\frac{1}{\mathrm{E}[T_a]}}{\frac{1}{\mathrm{E}[T_s]}} = \frac{\mathrm{E}[T_s]}{\mathrm{E}[T_a]}, \tag{2.5}$$

so dass wir über

$$\mathrm{E}[T_a] = \frac{\mathrm{E}[T_s]}{\rho} \tag{2.6}$$

auf unterschiedliche mittlere Werte $\mathrm{E}[T_a]$ der Zwischenankunftszeiten kommen. So beträgt im Fall der Auslastung von 98% die erwartete Zwischenankunftszeit $\mathrm{E}[T_a] = \frac{10}{0{,}98}$ ZE $=$ 10,204 ZE. Zum anderen variieren wir die Variationskoeffizienten c_a und c_s der zufälligen Zwischenankunfts- und Bedienzeiten, und zwar jeweils im gleichen Maße zwischen 0,25 und 2,0.

Tabelle 2.4: Parameter der Simulationsstudie

Größe	Ausprägung(en)
$\mathrm{E}[T_s] = \frac{1}{\mu}$ [ZE]	10
$c_a = c_s$	0,25 / 0,5 / 1,0 / 2,0
$\rho = \frac{\mathrm{E}[T_s]}{\mathrm{E}[T_a]}$	50% / 65% / 80% / 95% / 98%

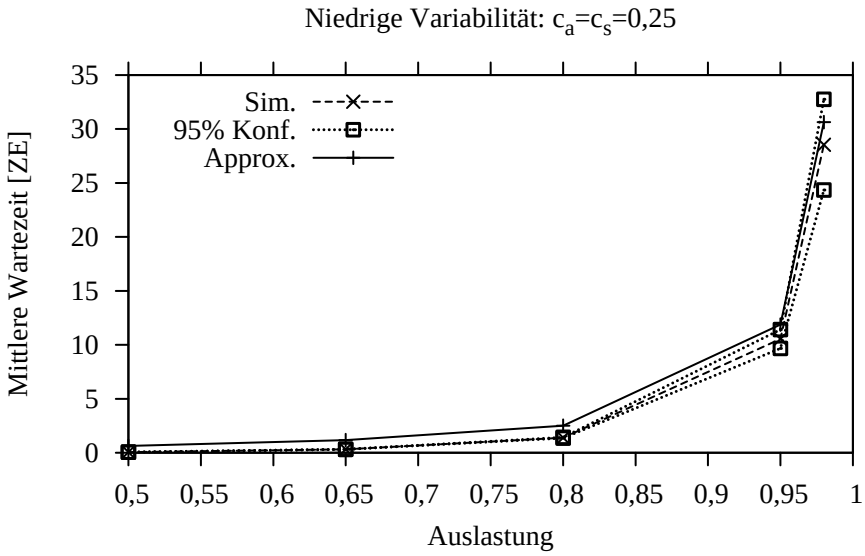

Abbildung 2.5: Mittlere Wartezeit bei niedriger Variabilität

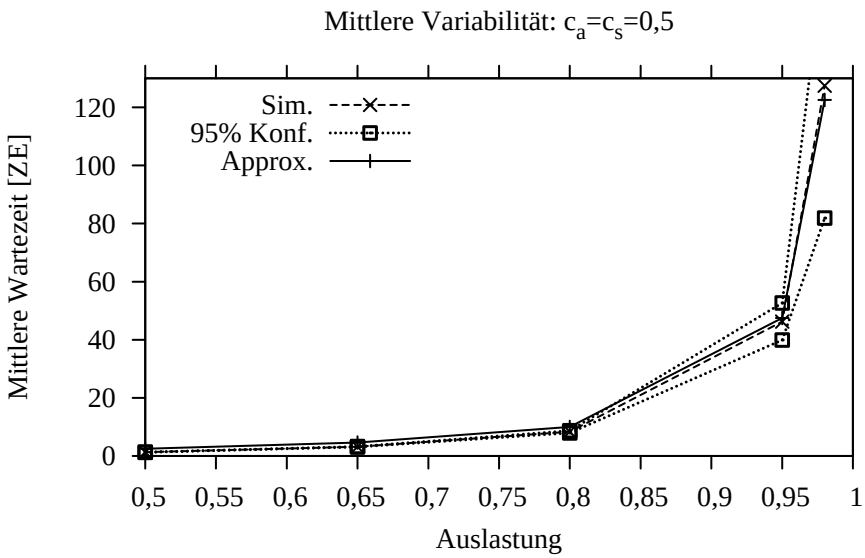

Abbildung 2.6: Mittlere Wartezeit bei mittlerer Variabilität

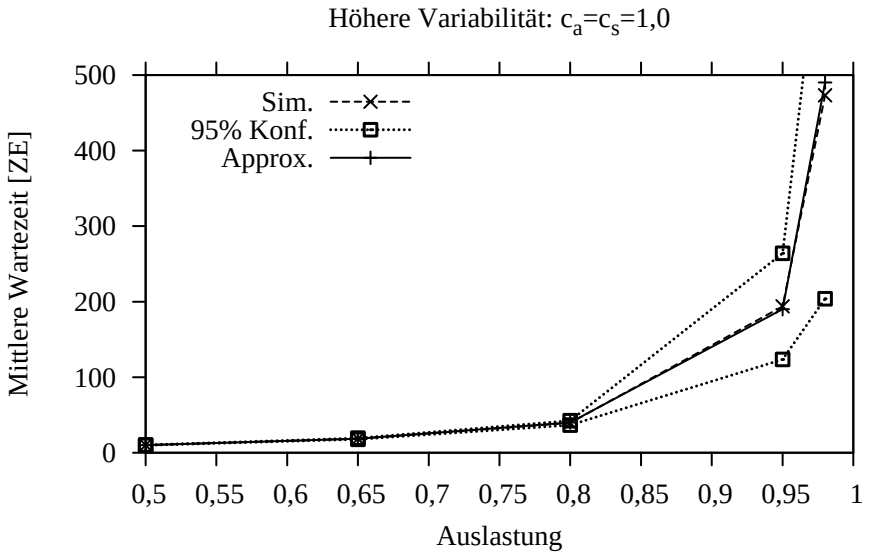

Abbildung 2.7: Mittlere Wartezeit bei höherer Variabilität

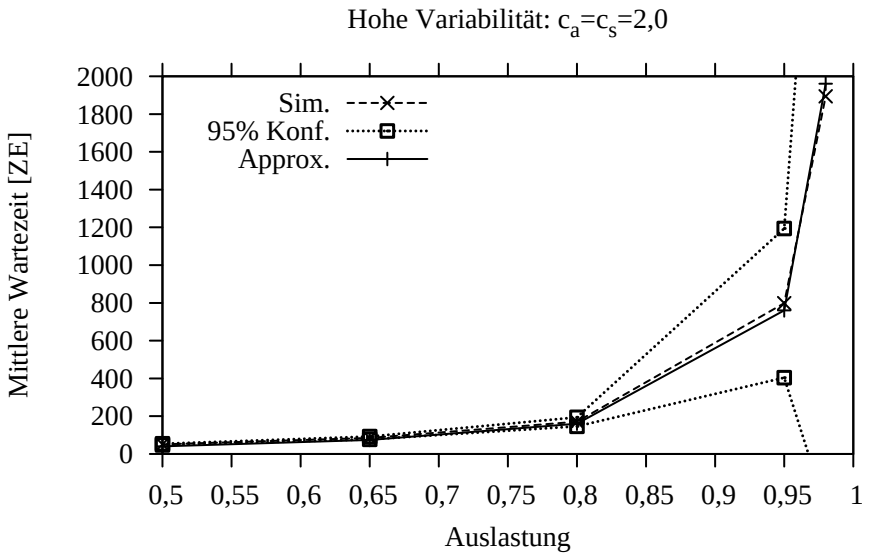

Abbildung 2.8: Mittlere Wartezeit bei hoher Variabilität

Unter Verwendung einer geeigneten Simulationssoftware simulieren wir nun dieses System. Für jede Parameterkombination aus der Tabelle 2.4 führen wir dazu mehrere voneinander unabhängige Simulationsläufe durch. In jedem dieser Läufe werden für eine größere Anzahl von Jobs zufällige Realisationen der Zwischenankunftszeiten und der Bedienzeiten ermittelt. Daraus ergibt sich für jeden Job seine Wartezeit und letztlich über die Betrachtung vieler Jobs ein Schätzwert der mittleren Wartezeit.

Die Abbildungen 2.5 bis 2.8 zeigen die Ergebnisse dieser Simulationen. Man erkennt, dass die Wartezeit der Jobs mit zunehmender Auslastung stark ansteigt. Darüber hinaus zeigt sich, dass offenbar die Variabilität der Zwischenankunfts- und Bearbeitungszeiten auch einen extrem starken Einfluss auf die mittleren Wartezeiten hat.

Zusätzlich zu den Mittelwerten der Simulation (dicke gebrochene Linien) sind in den Abbildungen auch die Ober- und Untergrenzen der 95%-Konfidenzintervalle angegeben (dünne gepunktete Linien). Diese zeigen, dass die Simulationen offenbar im Bereich niedriger Auslastungen recht präzise sind, während sie im Bereich hoher Auslastungen jedoch recht ungenau scheinen.

Die Abbildungen enthalten darüber hinaus noch jeweils eine weitere Kurve mit der Bezeichnung „Approx.“. Diese Kurve ist *nicht* das Ergebnis einer Simulation, sondern das Ergebnis einer approximativen analytischen Berechnung, die so einfach ist, dass man sie auch noch ohne Taschenrechner hinbekommen sollte. Bevor wir uns dieser Berechnung zuwenden, wollen wir mit einem letzten Blick auf die Abbildungen 2.5 bis 2.8 festhalten, dass offenbar in allen Fällen die mittleren Wartezeiten der Jobs aus der Simulation („Sim.“) recht gut mit den berechneten Werten („Approx.“) übereinstimmen. Mit dem Wissen, wie man diese Berechnung durchführt, hätten wir auf die Simulation offenbar auch verzichten können.

2.2b Für diese approximative Bestimmung des Erwartungswertes der Wartezeit verwenden wir die folgende **Approximationsformel von Kingman**:[3]

$$\mathrm{E}[W_q] \approx \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1-\rho} \cdot \frac{1}{\mu} \tag{2.7}$$

Sie erlaubt es, die im Mittel bei einem Bediensystem zu erwartende Wartezeit $\mathrm{E}[W_q]$ abzuschätzen. Die Herleitung dieser Formel liegt weit jenseits des Anspruchs dieses Tutorials, die Anwendung ist aber gleichermaßen einfach wie erhellend. Dazu ist es offenbar ausreichend, eine Vorstellung vom Erwartungswert $\mathrm{E}[T_a] = \frac{1}{\lambda}$ der Zwischenankunftszeit sowie vom Erwartungswert $\mathrm{E}[T_s] = \frac{1}{\mu}$ der Bedienzeit und den jeweiligen Standardabweichungen σ_{T_a} und σ_{T_s} zu haben. Daraus kann man die Variationskoeffizienten c_a und c_s sowie die Auslastung $\rho = \frac{\lambda}{\mu}$ berechnen und alles in die Approximationsformel (2.7) einsetzen.

[3] Vgl. z. B. Curry und Feldman (2011, S. 94).

Die *gesamte* vom Job im System verbrachte Zeit W ist die Wartezeit plus der Bearbeitungszeit

$$W = W_q + T_s \tag{2.8}$$

und für deren Erwartungswert gilt wegen $\mathrm{E}[T_s] = \frac{1}{\mu}$ offenbar

$$\mathrm{E}[W] = \mathrm{E}[W_q] + \mathrm{E}[T_s] \approx \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1-\rho} \cdot \frac{1}{\mu} + \frac{1}{\mu}. \tag{2.9}$$

Beispiel zur telefonischen Kundenbetreuung:

Die Kunden der *Möllix GmbH* beschweren sich beim Geschäftsführer, Franz Meier, dass die Erreichbarkeit der telefonischen Kundenbetreuung eine absolute Zumutung sei. Meier kann das nicht verstehen. Häufig sieht er, wie der eine Kundenbetreuer untätig herumsitzt und auf Anrufe wartet. Früher gab es so etwas nicht. Daher hatte er schon überlegt, dem Kundenbetreuer zusätzliche Aufgaben zu übertragen. Kürzlich hat der Geschäftsführer mit Thorben Schneider erstmals einen Assistenten eingestellt. Diese hat gerade an der Universität Hannover sein Studium der Wirtschaftswissenschaften abgeschlossen. Meier beauftragt ihn, der Sache nachzugehen.

Das Computerprotokoll der Telefonanlage zeigt, dass der Kundenbetreuer der Möllix GmbH pro Stunde im Mittel drei telefonische Anfragen erhält. Die Standardabweichung der Zeit zwischen zwei Anrufen beträgt 20 Minuten. Er benötigt im Mittel 15 Minuten für eine Anfrage, die Standardabweichung der Gesprächsdauer beträgt ebenfalls 15 Minuten. Die Analyse des Assistenten ergibt Folgendes:

1. Für die Zwischenanrufzeiten gilt:

$$\begin{aligned} \mathrm{E}[T_a] &= 20 \text{ min} \\ \sigma_{T_a} &= 20 \text{ min} \\ c_a &= \frac{\sigma_{T_a}}{\mathrm{E}[T_a]} = 1 \\ \lambda &= \frac{1}{20} \text{ min}^{-1} \end{aligned}$$

2. Für die Bearbeitungszeiten gilt:

$$\begin{aligned} \mathrm{E}[T_s] &= 15 \text{ min} \\ \sigma_{T_s} &= 15 \text{ min} \\ c_s &= \frac{\sigma_{T_s}}{\mathrm{E}[T_s]} = 1 \\ \mu &= \frac{1}{15} \text{ min}^{-1} \end{aligned}$$

3. Für die Auslastung folgt:

$$\rho = \frac{\lambda}{\mu} = 75\%$$

4. Für die Wartezeit folgt:

$$\begin{aligned} \mathrm{E}[W_q] &\approx \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1-\rho} \cdot \frac{1}{\mu} \\ &= \frac{1+1}{2} \frac{0{,}75}{1-0{,}75} \cdot 15 \text{ min} \\ &= 45 \text{ min} \end{aligned}$$

5. Für die Summe aus Wartezeit und Bedienzeit folgt:

$$\mathrm{E}[W] = 45 \text{ min} + 15 \text{ min} = 60 \text{ min}$$

Aus dieser Sicht erscheinen die Klagen der Kunden nachvollziehbar. Aber was ist zu tun?

Die Kingman'sche Approximationsformel für die Wartezeit

$$\mathrm{E}[W_q] \approx \frac{c_a^2 + c_s^2}{2} \cdot \frac{\rho}{1-\rho} \cdot \frac{1}{\mu} \tag{2.10}$$

zeigt, dass es zur Reduzierung der Wartezeit eines Bediensystems offenbar drei Anknüpfungspunkte gibt:

- Verringerung der **Variabilität** $c_a^2 + c_s^2$
- Verringerung der **Auslastung** ρ
- Verringerung der **mittleren Servicezeit** $\mathrm{E}[T_s] = \frac{1}{\mu}$

Wenn es gelingt, die Variabilität zu senken, also bei gleichen Erwartungswerten und somit auch gleicher Auslastung die *Schwankungen* der Zwischenankunfts- und Bearbeitungszeiten der Jobs zu reduzieren, so reduzieren sich die Wartezeiten. Schwanken diese Zeiten nicht, so treten offenbar keine Wartezeiten auf.

Beispiel zur telefonischen Kundenbetreuung (Fortsetzung):

Thorben Schneider bespricht die Angelegenheit mit dem Kundenbetreuer. Dabei wird deutlich, dass der Kundenbetreuer bei komplizierteren Problemen häufig in der Entwicklungsabteilung nachfragen muss, dort aber nicht immer jemanden erreicht, was zu den langen und stark schwankenden Bedienzeiten führt. Thorben kommt zu dem Schluss, dass der Kundenbetreuer eine bessere Schulung benötigt, um fast alle Anfragen eigenständig beantworten zu können. Er schätzt, dass so die mittlere Bedienzeit auf 10 Minuten reduziert werden kann bei einer reduzierten Standardabweichung von 5 Minuten, und macht die folgende Rechnung auf:

1. An den Zwischenanrufzeiten kann nichts geändert werden, für sie gilt weiterhin:

$$\begin{aligned} \mathrm{E}[T_a] &= 20 \text{ min} \\ \sigma_{T_a} &= 20 \text{ min} \\ c_a &= \frac{\sigma_{T_a}}{\mathrm{E}[T_a]} = 1 \\ \lambda &= \frac{1}{20} \text{ min}^{-1} \end{aligned}$$

2. Für die Bearbeitungszeiten würde gelten:

$$\begin{aligned} \mathrm{E}[T_s] &= 10 \text{ min} \\ \sigma_{T_s} &= 5 \text{ min} \\ c_s &= \frac{\sigma_{T_s}}{\mathrm{E}[T_s]} = 0{,}5 \\ \mu &= \frac{1}{10} \text{ min}^{-1} \end{aligned}$$

3. Für die Auslastung würde folgen:

$$\rho = \frac{\lambda}{\mu} = \frac{10}{20} = 50\%$$

4. Für die Wartezeit ergäbe sich:

$$\begin{aligned} \mathrm{E}[W_q] &\approx \frac{c_a^2 + c_s^2}{2} \frac{\rho}{1-\rho} \frac{1}{\mu} \\ &= \frac{1 + 0{,}5^2}{2} \frac{0{,}5}{1-0{,}5} \cdot 10 \text{ min} \\ &= 6{,}25 \text{ min} \end{aligned}$$

5. Für die Summe aus Wartezeit und Bedienzeit würde folgen:

$$\mathrm{E}[W] = 6{,}25 \text{ min} + 10 \text{ min} = 16{,}25 \text{ min}$$

Thorben Schneider berichtet dem Geschäftsführer Franz Meier, dass es sich um ein selbstgemachtes Problem handelt, und dass eine verbesserte Schulung des Kundenbetreuers erforderlich ist. Dem Geschäftsführer ist es zwar ein Dorn im Auge, dass der Kundenbetreuer nach der Schulung nur noch zu 50% ausgelastet sein soll, aber er willigt ein, weil er die Kundenbeschwerden noch schlimmer findet.

In dem gerade vorgestellten Beispiel wurden durch die vorgeschlagenen Maßnahmen die Variabilität, die Auslastung und die mittlere Servicezeit gleichzeitig reduziert, mit ganz erheblichen Auswirkungen für die Wartezeiten der Kunden.

2.3 Bestände und Zeiten: Das Gesetz von Little

2.3 Die Alltagserfahrung lehrt uns, dass es lange dauert, wenn viele andere auch da sind (und vor uns stehen, z. B. im Supermarkt oder an der Tankstelle). Diese bewusst unscharfe Formulierung ist Ausdruck einer allgemeinen Gesetzmäßigkeit, auf die wir uns in der Prozessanalyse regelmäßig stützen, das sogenannte **Gesetz von Little**.[4] In einer natürlichsprachlichen Fassung sieht dieses Gesetz folgendermaßen aus:

$$\boxed{\text{Mittlerer Bestand} = \text{Ankunftsrate} \cdot \text{Mittlere Warte- bzw. Durchlaufzeit}}$$

Wir sehen uns dieses extrem wichtige Gesetz nun etwas genauer an. Sei $\mathrm{E}[W]$ die erwartete Zeit für das Passieren eines Systems, z. B. einer Warteschlange und eines Bediensystems, und sei $\mathrm{E}[L]$ die erwartete Anzahl von Jobs oder Kunden etc. in diesem System sowie λ die Rate, mit der Jobs in dem System ankommen. Sofern das System sich in einem stabilen („eingeschwungenen“) Zustand befindet, gilt stets die folgende Beziehung:

$$\boxed{\mathrm{E}[L] = \lambda \cdot \mathrm{E}[W]} \tag{2.11}$$

Der Bestand $\mathrm{E}[L]$ ist also proportional zur (Warte- bzw. Durchlauf-)Zeit $\mathrm{E}[W]$, der Proportionalitätsfaktor ist dabei die Ankunftsrate λ. Kann man zwei dieser drei Größen bestimmen, so lässt sich die dritte offenbar leicht berechnen.

Beachten Sie bitte, dass das Gesetz von Little für *beliebig gezogene Systemgrenzen* gilt. Entscheidend ist nur, dass sich die Zufallsvariablen L für den Bestand im System

[4] Siehe zum Beweis und für die folgende Erläuterung Little, John D. C. (1961) und Little, John D. C. (2011) sowie die Darstellung in Curry und Feldman (2011, Kap. 2.1.2).