Lösung der Aufgaben

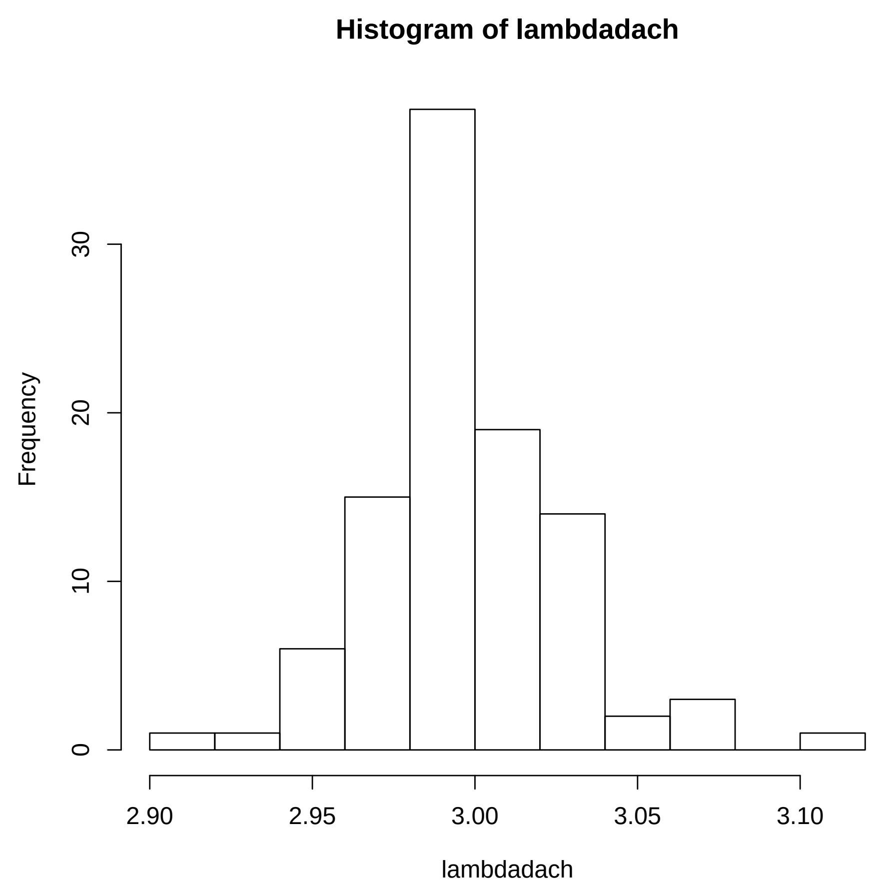
- Reproduktion des Ankunftsstroms.

- Vorüberlegung: Um die Intensität des Ausgangsstroms zu messen, muss man zu den Zwischenzeiten übergehen. Dies geht in `R` z.B. mit dem Befehl `diff()`.

```
n <- 10000 ; lambda <- 3 ; mu <- 4 ; lambdadach <- rep(NA, 100)

for (runs in 1:100){  cat(runs, "\r")
ankunftszeiten <- cumsum(rexp(n, lambda))
bearbeitungszeiten <- rexp(n, mu)
exits <- austrittsstrom(ankunftszeiten, bearbeitungszeiten)
lambdadach[runs] <- 1/mean(diff(exits))
}
hist(lambdadach)
```

Histogramm der geschätzten Intensitäten $\hat{\lambda}$



Simulation der Schlangenlänge

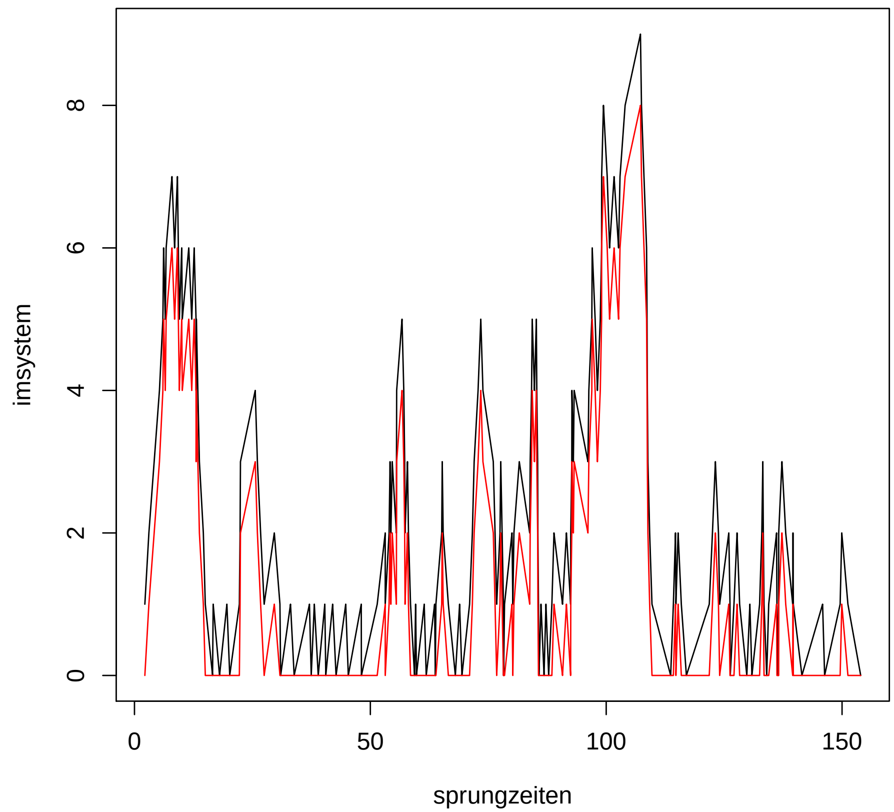
- Vorüberlegungen: Änderungen in der Schlangenlänge treten nur auf, wenn entweder ein neuer Kunde eintrifft (Ankunftszeiten) +1, oder wenn ein Kunde das System verlässt (Austrittsstrom) -1.

```
n <- 100 ; lambda <- 0.6 ; mu <- 1
ankunftszeiten <- cumsum(rexp(n, lambda))
bearbeitungszeiten <- rexp(n, mu)
exits <- austrittsstrom(ankunftszeiten, bearbeitungszeiten)

### An dieser Stelle liegen nun die Eintritts- und die
### Austrittszeiten vor. Jetzt
rein <- rep(1, n); raus <- rep(-1, n)
allezeiten <- c(ankunftszeiten, exits)
alleaenderungen <- c(rein, raus)
```

```
sortierterindex <- sort.int(allezeiten, index.return=TRUE)$ix
sprungzeiten <- allezeiten[sortierterindex]
alleaenderungen <- alleaenderungen[sortierterindex]
imsystem <- cumsum(alleaenderungen)
inschlange <- imsystem-1 ; inschlange[inschlange <0] <- 0
plot(sprungzeiten,imsystem, t="l")
lines(sprungzeiten,inschlange, t="l", col="red")
```

Kunden im System und in der Schlange



Simulation mit Inversionsmethode

- $y = F(x) = 1 - \rho e^{-(\mu-\lambda)x}$. Mittels Inversionsmethode:
- $x = F^{-1}(y) = \frac{\ln \frac{y-1}{-\rho}}{-(\mu-\lambda)}$.
- Generator:

```
xt1 <- function(lambda, mu)
              log((runif(1)-1)/-(lambda/mu))/(-(mu-lambda))
rxt <- function(n, lambda, mu) replicate(n, xt1(lambda, mu))
var(rxt(10000, 0.6, 1))
[1] 6.13298
mean(rxt(10000, 0.6,1))
[1] 1.232767
```

Ungeduldige Kunden und Warteschlangen

- Bisher sind wir davon ausgegangen, dass Kunden jeden Zustand ertragen und geduldig warten, bis sie an der Reihe sind.
- Heute werden die Modellierungsmöglichkeiten für verschieden ungeduldige Verhaltensweisen von Kunden eingeführt.
- Darüberhinaus werden einige theoretischen Ergebnisse für Spezialfälle angegeben.

Scheuende **Kunden**

- Jeder kennt das Phänomen, dass die Bereitschaft, sich hinten in eine Schlange einzureihen, mit der Länge der Schlange abnimmt.
- Auf Englisch wird das Phänomen *balking* genannt.
- Es sei G eine Funktion von $\mathcal{N} \to \mathcal{R}$, $G(n)$ bezeichne die Wahrscheinlichkeit, dass ein neuer Kunde sich hinten an eine Schlange von n Kunden anschliesst.
- $G(n)$ heißt die Scheufunktion (balking function).
- Meist nimmt man $G(0) = 1$ an.

- Achtung: G ist keine Dichte über den natürlichen Zahlen, sonder für jedes n wird ein evtl. anderes Bernoullie-Experiment mit Erfolgswahrscheinlichkeit $G(n)$ durchgeführt!

Ein einfaches Szenario mit analytischen Ergebnissen

- Gegeben sei die Scheufunktion $G(n) = \frac{1}{n+1}$ für $n \geq 0$.
- Anschaulich: Die Wahrscheinlichkeit des Anstellens ist umgekehrt proportional zur Schlangenlänge.
- Es kann dann analog zu den Ergebnissen der letzten Veranstaltung gezeigt werden, dass für die Wahrscheinlichkeiten $p_n(t)$, dass sich zum Zeitpunkt t genau n Anfragen im System befinden gilt:

$$\begin{aligned}
\frac{dp_n(t)}{dt} &= -(\lambda G(n) + \mu)p_n(t) + \mu p_{n+1}(t) + \lambda G(n-1)p_{n-1}(t), \text{ für } n \geq 1 \\
\frac{dp_0(t)}{dt} &= -\lambda p_0(t) + \mu p_1(t).
\end{aligned}$$

- Im Gleichgewichtszustand folgt:

$$\begin{aligned} p_{n+1} &= \frac{\lambda G(n) + \mu}{\mu} p_n - \frac{\lambda G(n-1)}{\mu} p_{n-1} \text{ für } n \geq 1; \\ p_1 &= \frac{\lambda}{\mu} p_0. \end{aligned}$$

- Diese sogenannte Differenzengleichung kann analytisch gelöst werden und es ergibt sich $p_n = \frac{1}{n!}(\frac{\lambda}{\mu})^n p_0$.

- Durch Einsetzen bekommt man auch noch p_0:

$$1 = p_0 + \sum_{n=1}^{\infty} p_n = p_0 \sum_{n=1}^{\infty} \left(\frac{1}{n!} (\frac{\lambda}{\mu})^n \right) = p_0 e^{\frac{\lambda}{\mu}}.$$

- Also $p_0 = e^{-\frac{\lambda}{\mu}}$.

- Wichtig ist die Beobachtung, dass die Summe $\sum_{n=1}^{\infty}\left(\frac{1}{n!}(\frac{\lambda}{\mu})^n\right)$ immer konvergiert, wenn $\mu > 0, \lambda \in \mathcal{R}$.

- Die erwartete Anzahl von Anfragen im System im Gleichgewichtszustand lautet also:

$$L = \sum_{n=0}^{\infty} p_n n = \sum_{n=1}^{\infty} \frac{1}{(n-1)!}\left(\frac{\lambda}{\mu}\right)^n e^{-\frac{\lambda}{\mu}} = e^{-\frac{\lambda}{\mu}} \sum_{m=0}^{\infty} \frac{1}{m!}\left(\frac{\lambda}{\mu}\right)^{m+1} = e^{-\frac{\lambda}{\mu}} e^{\frac{\lambda}{\mu}} \frac{\lambda}{\mu} = \frac{\lambda}{\mu}$$

Warteraumkapazität als Spezialfall des Scheuens

- Die Kapazität $K < \infty$ eines Systems lässt sich über eine spezielle Scheufunktion in des Modell aufnehmen,

- Um eine Kapazität K zu modellieren, definiere

$$G(n) = \begin{cases} 1, & \text{für } n < K, \\ 0 & \text{für } n \geq K. \end{cases}$$

- Solange die Kapazität nicht erschöpft ist, stellen sich weitere Kunden an, sobald und solange die Kapazität erschöpft ist, stellt sich niemand mehr an.

Integration der scheuenden Kunden in Simulationen

- Offensichtlich muss zu jedem Zeitpunkt die Schlangenlänge bekannt sein. (Musterlösung zu heute)

- Die Ankunftsintensität λ des Prozesses bleibt unverändert, jedoch wird bei jedem eintreffenden Kunden eine weiteres Bernoulliexperiment mit $p = G(n)$ durchgeführt und der Kunde nur dann in des System aufgenommen, wenn das Experiment erfolgreich ist.

- In `R`: `runif() < G(n)`.

Aufgabe

- Implementieren Sie die beiden Scheufunktionen aus den Beispielen.
- Lösung gleich hier.

Der *fliehende* (ungeduldige) Kunde

- Auf Englisch: *reneging*.

- Wenn in der Schlange nichts passiert, dann lässt man den Einkaufswagen irgendwann stehen.

- Die Funktion, die bestimmt, ob und wann die Schlange verlassen wird, heißt Ungeduldsfunktion.

- Ein (Der) Fall, der in der Theorie noch gut zu behandeln ist, liegt vor, wenn man auch die Zeit bis zu einem Verlassen der Warteschlange aus Ungeduld als exponentialverteilt ansieht. Die Ungeduldsfunktion ist dann die Verteilungsfunktion der Exponentialverteilung.

- Die übliche Annahme ist, dass der Kunde geht, wenn bis zu einem aus der Exponentialverteilung mit Intensität α zu bestimmenden Zeitpunkt

seine Bedienung nicht begonnen hat.
(Ankunftszeit + erträgliche Wartezeit, beide aus Expoentialverteilung)

- In der Simulation bedeutet dies, dass zu jeder Anfrage auch ein spätester Zeitpunkt festgelegt wird, bis zu dem die Anfrage gültig ist. Danach wird sie zurückgezogen.

- Unter diesen Annahmen gilt das folgende Theorem:
  Jede M/G/1 Schlange mit exponentialverteilter Ungeduldsfunktion hat einen Gleichgewichtszustand, unabhängig von der Ankunftsintensität und der Bedienverteilung. (Bacelli, Boyer, Hebturne 1984)

- Ein Speziallfall der Ungeduld ist das *Ausscheiden* (drop-off) von Kunden. Im Unterschied zur Ungeduld wird beim Ausscheiden die Entscheidung nicht vom Kunden getroffen (Beispiel: Wartelisten auf Organspenden).

Beispiel mit Scheuen und Fliehen

- Es sei folgende Scheufunktion G gegeben

$$G(n) = \begin{cases} 1 - \frac{n+1}{N}, & \text{für } n < N \\ 0, & \text{sonst, also } n \geq N \end{cases}.$$

- In diesem Modell ist die Wahrscheinlichkeit für einen neuen Kunden sich anzustellen proportional zu den Restplätzen im Wartezimmer mit Kapazität N.

- Außerdem bricht der Kunde sein Warten ab, wenn bis zu einem exponentialverteilten Zeitpunkt (Intensität α) seine Bedienung nicht begonnen hat.

- Mit Rechnung analog zu den einfacheren Fällen kann man in diesem Fall noch die p_n ausrechnen, aber es wird schon unübersichtlich:

$$\begin{aligned} p_0 &= \left[1 + \frac{NB_z(\gamma, N)}{z^{\gamma-1}(1-z)N}\right]^{-1} \\ p_n &= \frac{\rho^n \prod_{j=0}^{n-1}(N-j)}{\prod_{j=0}^{n-1} j} p_0, \end{aligned}$$

wobei $\gamma = \frac{\mu}{\alpha}$ und $z = \frac{\rho}{1+\rho}$.
$B_z(\gamma, N)$ ist die sogenannte Besselfunktion.

- Entsprechend kann man noch Erwartungswerte etc. berechnen.

FIFO und LIFO

- First-in-first-out und Last-in-last-out sind beides übliche Regimes für Warteschlangen.
- Interessant ist die Frage, wie sich diese Regimes aus die verschiedenen interessanten Größen einer Warteschlange auswirken.
- Wie üblich habe der Ankunftsprozess die Intensität λ und der Bedienprozess die Intensität $\mu > \lambda$.

FIFO

- Bisher haben wir FIFO als einziges REgime betrachtet. Zur Wiederholung:
- Für das Warteschlangenregime FIFO und die Wartezeit W im Steady-State gelten dann:

$$\begin{aligned} E(W) &= \frac{\lambda}{\mu(\mu-\lambda)}, \\ Var(W) &= \frac{\rho(2-\rho)}{\mu^2(1-\rho)^2}. \end{aligned}$$

LIFO und Vergleich

- Für LIFO ist es etwas schwieriger zu rechnen, man gelangt jedoch zu dem Ergebnis, dass Für das Warteschlangenregime LIFO und die Wartezeit W im Steady-State gilt:

$$\begin{aligned} E(W) &= \frac{\lambda}{\mu(\mu-\lambda)} \\ Var(W) &= \frac{\rho(\frac{2}{1-\rho}-\rho)}{\mu^2(1-\rho)^2}. \end{aligned}$$

- Damit gilt

$$Var(W)_{LIFO} - Var(W)_{FIFO} = \frac{2\rho^2}{\mu^2(1-\rho)^3} > 0!$$

Ordnung der Warteschlangenregimes

- Zum Abschluss noch ein sehr ästhetisches Resultat.
- Tatsächlich gilt sogar das Theorem (Kingman 1962, Tambouratzis 1968): Unter allen Warteschlangeregimen die nicht die Länge der Warteschlange beeinflussen hat das Regime FIFO die kleinste und LIFO die größte Varianz der Wartezeit!

Aufgabe zur Warteschlangensimulation

- Diese Aufgabe vereint alle Aspekte, die in dem Abschnitt über Warteschlangen behandelt wurden. Deshalb klingt sie sehr komplex, jedoch lassen sich mit Hilfe der bisherigen Aufgaben viele Teillösungen relativ unkompliziert zur Lösung dieser Aufgabe verbinden.

- Gegenben seien $\lambda = 12, \mu = 13$. Zu simulierende Länge des Stroms $n = 1000$. Dies sei die Anzahl von Anfragen, die pro Stunde eintritt.

- Das System kann höchstens $N = 10$ Anfragen aufnehmen.

- Je nach Schlangenlänge m gelangen neue Anfragen nur gemäß der Scheufunktion

$$G(m) = 1 - (m/N)^2, m \leq N, 0 \text{ sonst}$$

ins System.

- Weiterhin endet jede Anfrage gemäß eines Expoentialprozesses mit einer mittleren Wartezeit von 15 min in der Warteschlange.
- Simulieren Sie diesen Prozess! Gehen Sie dabei schrittweise vor und notieren Sie in natürlicher Sprache, wie die einzelnen Verteilungen ineinandergreifen!
- Welche Intensität (pro Stunde) besitzt der Austrittsprozess?
- Wie lang ist die Schlange im Schnitt?
- Wie groß ist die durchschnittliche Wartezeit im System bis zum Verlassen, entweder nach Bedienung oder nach Abbruch?

Ausblick

- Nur für die Poissonprozesse kann man diese schönen Ergebnisse berechnen.
- Sobald andere Randverteilungen ins Spiel kommen oder wenn die Modellannahmen kompexer werden, helfen nur noch Simulationen.
- Nicht behandelt wurden beispielsweise Mehrbedienersysteme oder ganze Netze von Warteschlangen.