

Ein Kurz-Skript

H. P. Wolf

July 23, 2003, file: stat3.rev

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einleitung | 3 |
| 1.1 | Warum? | 3 |
| 1.2 | Literatur | 5 |
| 2 | Einige Bemerkungen zum Birthday-Spacings-Test | 6 |
| 2.1 | Algorithmus | 6 |
| 2.2 | Theoretischen Häufigkeiten – per Simulation | 7 |
| 2.3 | Testeinsatz | 8 |
| 2.4 | Weitere Implementationsvariationen | 9 |
| 2.5 | Implementierung des BS-Tests | 10 |
| 2.6 | Fazit | 12 |
| 3 | Warten auf die nächste Party | 13 |
| 3.1 | Problemstellung | 13 |
| 3.1.1 | Fehlversuch: Geometrische Verteilung | 13 |
| 3.1.2 | Versuch einer Simulation der Wartezeit | 13 |
| 3.1.3 | Vorgehensstrategie | 16 |
| 3.2 | Umskalierungen | 18 |
| 3.3 | Klassen der Modellierung | 21 |
| 3.4 | Gedankenexperiment: das Jahr als Kreis | 21 |
| 3.5 | Übergänge zwischen Intervall- und Kreiswelt | 22 |
| 3.6 | Verteilung von Wartezeiten in der linearen Welt | 24 |

| | |
|----------|--|
| 0 | 2 |
| 3.7 | Abweichung diskreter von kontinuierlicher Modellierung 26 |
| 3.8 | Experimente zur Frage der verlängerten Wartezeit über den Jahreswechsel 29 |
| 3.9 | Die letzten warten länger: wieso und wie viel? 34 |
| 3.10 | Die Verteilung von $U_{(i)} - U_{(i-1)}$ in der linearen Welt 42 |
| 3.11 | Fazit 47 |
| 4 | Warten auf den nächsten Bus 49 |
| 4.1 | Bushaltestellenlatein? 49 |
| 4.2 | Vergeudete Zeiten an Haltestellen und vor Ampeln 50 |
| 4.3 | Warten beim Autostopp mit Bernoulli und Co. 53 |
| 4.3.1 | Per Autostopp zur momenterzeugenden Funktion 54 |
| 4.3.2 | Varianz der Wartezeit auf das nächste Auto 56 |
| 4.3.3 | Experimentieren bis zum Erfolg 58 |
| 4.4 | Warten beim Autostopp mit Stoppuhr 59 |
| 4.5 | Eins, zwei, drei, ..., viele Wartezeiten zusammenaddiert 61 |
| 4.6 | Hand aufs Herz: Technik und Kompromisse 65 |
| 4.7 | Gestaltversteinerung nach dem x-ten Auto 66 |
| 4.8 | Autoanzahl bei vorgegebener Zeit aus Poissonscher Sicht 70 |
| 4.9 | Verteilung von Autos in einem vorgegebenen Zeitintervall 73 |
| 4.9.1 | Das kommt uns bekannt vor! 74 |
| 4.10 | Was, wenn der Bus auch zufällig kommt? 75 |
| 4.11 | Straßenbahnen – dann nehme ich eben die nächste 77 |
| 4.12 | Unpünktliche Linienbusse 79 |
| 4.12.1 | Spezialfall: Gleichverteilte Intervalllängen 80 |
| 4.12.2 | Experimentelle Überprüfung 81 |
| 5 | Axiome 84 |
| 6 | Anhang: Geburtstage von 500 Studierenden 88 |

1 Einleitung

1.1 Warum?

Warum ist eine Beschäftigung mit elementaren statistischen Konzepten wichtig?

- Antwort auf logischer Ebene: ohne Fundament läßt sich schlecht bauen
- Antwort auf inhaltlicher Ebene: mit Kenntnissen über Wahrscheinlichkeiten, Zufallsvariablen, Verteilungen, Momenten, Transformationen usw. lassen viele Fragen, in denen Zufälligkeiten eine Rolle spielen, angehen.
- Antwort auf exemplarischer Ebene:
 - Ein Roulette-Spieler will seine Strategie nicht aufgeben. Er will solange ROT spielen, bis er ein Spiel gewinnt. Wie viele Versuche muß er (erwartungsgemäß) unternehmen? Was ist von einer beständigen Verdoppelung des Einsatzes zu halten? Ohne zu rechnen: Wie viele SCHWARZ-Serien bestehend aus 5 bzw. k aufeinander folgenden *schwarzen* Ergebnissen wird es wohl bei 1000 Roulette-Durchgängen geben?
 - Ein Hausbesitzer denkt über die Versicherung seines Hauses gegen Hochwasser nach, da er direkt an einem Fluß wohnt. Er überlegt, wie wahrscheinlich es wohl ist, daß die Wassermenge eine kritische Marke übersteigt und versucht, den einzutretenden Schaden abzuschätzen. Hierdurch will er herauskriegen, ob die übliche Versicherungsprämie in seinem Fall akzeptabel ist.
 - Zufallszahlen sollen Eigenschaften von zufällig generierten Zahlen besitzen. Um sicher zu gehen gibt es verschiedene Tests auf Zufälligkeit. Der *Birthday-Spacings-Test* ist ein solcher. Dazu werden zum Beispiel $m = 512$ ZZ ermittelt – oder aus dem Zufallsstrom m Bitfolgen entnommen –, sortiert, die Differenzen der sortierten Folge gebildet und gezählt, wie viele dieser Differenzen gleich sind. Durch Wiederholung dieser Prozedur erhalten wir eine empirische Verteilung der Anzahl gleichlanger Wartezeiten bis zum nächsten Geburtstagsfest. Diese lassen sich mit Hilfe des χ^2 -Tests auf Stimmigkeit mit der theoretischen Verteilung testen. Gibt es $n = 2^{25}$ verschiedene Bitfolgen, dann lautet die theoretische Verteilung nach Knuth (3.3.2), p.71

| | | | | |
|------------------|------------|------------|------------|------------|
| $R =$ | 0 | 1 | 2 | > 2 |
| with probability | .368801577 | .369035243 | .183471182 | .078691997 |

Was halten wir davon?

- Wie groß ist die Wahrscheinlichkeit, daß morgen die Sonne scheint?
- Ansicht des Autors über den Zusammenhang *Problemschwierigkeitsgrad* \longleftrightarrow *Erfordernisse*:

- einfache Probleme → Intuition ist ausreichend
- mittlere Probleme → Standardverfahren ausreichend
- schwere Probleme → Ergebnisinterpretation erfordert tiefere Einsichten besonders auch von statistischen Konzepten
- ganz schwere Probleme → Verfahren müssen (neu) zusammengebaut werden, was ohne fundamentales Wissen nicht klappt

Sowohl zur Modellierung realer Problemlagen wie auch zum Verständnis verschiedener statistischer Prozeduren ist Basis-Wissen notwendig sowie eine aktive Beschäftigung mit den Grundlagen. Soviel zur Warum-Frage.

In diesem noch in der Aufbauphase befindlichem Skript wird der Leser an verschiedene elementare Konzepte der Statistik herangeführt. Zusätzlich stehen eine Reihe von R-Experimente zur Untersuchung der diskutierten Zusammenhänge bereit. Thematisch ist die Diskussion aufgehängt an einem Teil eines der ganz zentralen Begriffe aus der Statistik:

Wahrscheinlichkeiten → WAhRscheinlichkeiTEN → WARTEN

Wie das? Warten Sie's ab.

1.2 Literatur

- Venables, W.N., Ripley, B.D. (2002): Modern applied statistics with S, 4rd Ed., Springer
- Knuth, D.E (verschiedene Bände / verschiedene Jahre / Auflagen): The Art of Computer Programming, Addison-Wesley
- Mood, A.M., Graybill, F.A., Boes, D.C. (1963): Introduction to the Theory of Statistics, 3rd Ed., McGraw-Hill
- Lindgren, B.W. (1976): Statistical Theory, 3rd Ed., Collier Macmillan
- Fisz, M. (1989): Wahrscheinlichkeitsrechnung und mathematische Statistik, 11rd Ed., Deutscher Verlag der Wissenschaften
- Feller, W. (1968): An Introduction to Probability Theory and its Applications, 3rd Ed., Wiley
- Dudewicz, E.J., Mishra, S.N. (1988): Modern Mathematical Theory, Wiley
- Mardia, K.V. (1972): Directional Statistics, Academic Press
- Bickel, P.J., Doksum, K.A. (1977): Mathematical Statistics: Basic Ideas and selected Topics, Holden-Day
- Bhat, U.N. (1984): Elements of Applied Stochastic Processes, 2rd Ed., Wiley. Mardia, K.V. (1972): Directional Statistics, Academic Press.

2 Einige Bemerkungen zum Birthday-Spacings-Test

2.1 Algorithmus

Zur Lösung verschiedenster Probleme kann uns der Rechner helfen: Berechnungen, Graphik-Erstellung und symbolische Umformungen gehören mit zu den elementaren Einsatzzwecken. Darüber hinaus lassen sich erst mit dem Rechner spezielle computerintensive Verfahren umsetzen und Simulationen und numerische Experimente durchführen. Hierzu wollen wir den Birthday-Spacings-Test als Beispiel aufgreifen.

In diesem Abschnitt wird demonstriert, wie es im täglichen Alltag erforderlich sein kann, verschiedenste Elemente zusammenzufügen, um damit seine Frage zu beantworten. Stellen wir uns vor, daß wir einen Zahlenstrom von Zufallszahlen vorliegen haben und wir seine Qualität überprüfen sollen. Nehmen wir weiter an, daß uns jemand zugeflüstert hat, daß wir den *Birthday-Spacings-Test* einsetzen sollen. Fein. Klingt nicht schwer. *Frisch Gesellen* ...

Für den Einsatz benötigen wir eine entsprechende Testprozedur. Dafür müssen wir die verbalen Formulierungen zum Test und dessen Einsatzes in eine algorithmische Form überführen. Dieses kann zum Beispiel zu folgender Hinschreibung führen.

Algorithmus:

1. *stelle Anzahl der Tage und Personanzahl / Anzahl der Geburtstage fest*
2. *ermittle Geburtstage für viele Gruppen von Personen*
3. *sortiere Geburtstage in jeder Gruppe*
4. *stelle Differenzen zwischen den Geburtstagen jeder Gruppe fest*
5. *sortiere Differenzen jeder Gruppe*
6. *stelle in jeder Gruppe die Anzahl gleicher Differenzen fest*
7. *stelle Häufigkeitstabelle für die Anzahl gleicher Differenzen auf*
8. *ermittle theoretische Häufigkeiten*
9. *vergleiche die beobachtete Anzahl mit der theoretischen mittels χ^2 -Test*

Diese durchnummerierte Liste suggeriert eine gewisse Wissenschaftlichkeit – aber haben wir damit nach dem *Teile-und-herrsche-Prinzip* schon alle Probleme gelöst? Nicht ganz. Denn wenn die Parameter etwas anders sind als die aus dem Knuth-Beispiel, kennen wir die theoretische Verteilung nicht.

Für eine korrekte Interpretation der Ergebnisse müssen natürlich auch die Zusammenhänge rund um den χ^2 -Test verstanden worden sein, also muß diesbezüglich statistisches Grundwissen vorliegen. Für die Durchführung des Test kann dann auf eine statistische Software zurückgegriffen werden, die einen χ^2 -Anpassungstest anbietet oder zumindest die Bestimmung von Quantilen oder Verteilungsfunktionswerten der χ^2 -Verteilung anbietet. Wenn das nicht der Fall ist, muß auch die χ^2 -Verteilung implementiert werden.

Verfolgen wir zunächst die Ermittlung der Verteilung von mehrfachen gleichlangen Geburtstagswartezeiten.

2.2 Theoretischen Häufigkeiten – per Simulation

Die theoretischen Häufigkeiten wollen wir auf simulationstechnischem Wege mit Hilfe einer R-Funktion ermitteln. Dieser sind folgende Inputs zu übergeben:

| | |
|----------|-------------------------------------|
| n.pers | Personenanzahl |
| n.days | Tage im Jahr |
| n.groups | Anzahl der Gruppen / Wiederholungen |

```
1  (* 1)≡
  bsvert1<-function(n.pers, n.days, n.groups){
    equal.spacings<-rep(0,n.groups)
    for(j in 1:n.groups){
      days          <- sample(1:n.days, n.pers, replace=T)
      days          <- sort(days)
      spacings      <- diff(days)
      spacings      <- sort(spacings)
      equal.spacings[j] <- sum(spacings[-1]==spacings[-(n.pers-1)])
    }
    table(equal.spacings)/n.groups
  }
```

Für längere Jahre dauert dieses Vorgehen sehr lange, weil der Ziehungsprozeß nur zeitaufwendig zu erledigen ist. Deshalb wollen wir diesen aus der Schleife herausnehmen. Das Ergebnis der Stichprobenziehung wird in einer Matrix abgelegt, deren j -te Spalte die Geburtstage der j -ten Gruppe (Wiederholung) zeigt.

```
2  (* 1)+≡
    bsvert2<-function(n.pers, n.days, n.groups){
      days <- sample(1:n.days, n.pers*n.groups, replace=T)
      days <- matrix(days,n.pers,n.groups)
      equal.spacing<-rep(0,n.groups)
      for(j in 1:n.groups){
        days.sort      <- sort(days[,j])
        spacings      <- diff(days.sort)
        spacings      <- sort(spacings)
        equal.spacing[j] <- sum(spacings[-1]==spacings[-(n.pers-1)])
      }
      table(equal.spacing)/n.groups
    }
```

2.3 Testeinsatz

Wir wollen sehen, ob die Funktionen vernünftig funktionieren.

```
3  (* 1)+≡
    n.groups <- 5000
    set.seed(13)
    bsvert2(512 , 2^25 , n.groups)
```

Für $n.groups==50000$ ergibt sich ein Ergebnis, das der von Knuth vorgeschlagenen Verteilung sehr nahe kommt. Die Berechnung dauert natürlich eine gewisse Zeit.

```
equal.spacing
      0      1      2      3      4      5      6      7
0.37210 0.36618 0.18392 0.05978 0.01458 0.00292 0.00050 0.00002
```


2.4 Weitere Implementationsvariationen

In der Statistik gibt es für Fragen oft viele Wege, um zu einer Antwort zu gelangen. Dieses gilt auch für algorithmische Fragen, wie in diesem Abschnitt demonstriert wird. Weniger ist die optimale Umsetzung von Interesse als viel mehr das Aufzeigen des Möglichkeitsraumes, der sich schon für ein kleines Detail einstellt.

Es können lohnenswert sein, die Funktion `bsvert2` noch weiter zu verbessern. Deshalb werden weitere Implementationsmöglichkeiten vorgeschlagen: Die Berechnungen könnten schneller gehen, wenn die Schleife von `bsvert2` ganz vermieden wird:

```
4  (* 1)+≡
    bsvert3<-function(n.pers, n.days, n.groups){
      days      <- sample(1:n.days, n.pers*n.groups, replace=T)
      days      <- matrix(days,n.pers,n.groups)
      days      <- apply(days,2,sort)
      spacings  <- apply(days,2,diff)
      spacings  <- apply(spacings,2,sort)
      equal.spacings<- apply(spacings,2,function(x) sum(x[-1]==x[-(n.pers-1)]))
      table(equal.spacings)/n.groups
    }
```

... oder das ganze noch etwas kompakter gestaltet wird:

```
5  (* 1)+≡
    bsvert4<-function(n.pers, n.days, n.groups){
      days      <- matrix(sample(1:n.days, n.pers*n.groups,T), n.pers,n.groups)
      spacings  <- apply(days,2,function(x) sort(diff(sort(x))))
      equal.spacings<- apply(spacings,2,function(x) sum(x[-1]==x[-(n.pers-1)]))
      table(equal.spacings)/n.groups
    }
```

Auch ließe sich der doppelte `apply`-Aufruf vermeiden.

```
6  (* 1)+≡
    bsvert5<-function(n.pers, n.days, n.groups){
      days      <- matrix(sample(1:n.days, n.pers*n.groups,T), n.pers,n.groups)
      equal.spacings<- apply(days,2,function(x){
                                                x<-sort(diff(sort(x)))
                                                sum(x[-1]==x[-(n.pers-1)])
                                              })
      table(equal.spacings)/n.groups
    }
```

Weitere Versuche zeigten, daß folgender Vorschlag trotz Schleife den größten Geschwindigkeitsvorteil besitzt. Jedoch war diese Lösung nur unwesentlich schneller als `bsvert2`.

```
7 < * 1)+≡
  bsvert6<-function(n.pers, n.days, n.groups){
    days <- matrix(sample(1:n.days, n.pers*n.groups, replace=T),n.pers,n.groups)
    equal.spacings<-rep(0,n.groups)
    for(j in 1:n.groups){
      spacings      <- sort(diff(sort(days[,j])))
      equal.spacings[j] <- sum(spacings[-1]==spacings[-(n.pers-1)])
    }
    table(equal.spacings)/n.groups
  }
```

Man sollte schon an dieser Stelle darauf hinweisen, daß die Abstände vom jeweils letzten Geburtstag eines Jahres bis zum ersten des nächsten Jahres nicht berücksichtigt worden sind.

2.5 Implementierung des BS-Tests

Nach diesen Vorarbeiten können wir die Testprozedur zum Birthday-Spacings-Test definieren. Im Prinzip wird der oben skizzierte Algorithmus mit Hilfe von R umgesetzt. Wir stellen uns vor, daß die Daten in einer Matrix abgelegt sind. Jede Spalte der Matrix zeigt die Geburtstage für eine Gruppe an. Die Zeilenzahl liefert die Personenanzahl.

Inputs:

| | |
|---------------------|---------------|
| <code>data</code> | Beobachtungen |
| <code>n.days</code> | Tage im Jahr |

```
8 < * 1)+≡
  bst <- function(data, n.days=365){
    <stelle Personen- und Gruppenanzahl fest 9>
    <sortiere Geburtstage jeder Gruppe 10>
    <stelle Differenzen zwischen den Geburtstagen jeder Gruppe fest 11>
    <sortiere Differenzen jeder Gruppe 12>
    <stelle in jeder Gruppe die Anzahl gleicher Differenzen fest 13>
    <erstelle empirische Häufigkeitstabelle der Anzahl gleicher Differenzen 14>
    <ermittle theoretische Häufigkeiten 15>
    <vergleiche mittels  $\chi^2$ -Test 16>
    <liefere Ergebnis ab 20>
  }
```

- 9 *<stelle Personen- und Gruppenanzahl fest 9>*≡
n.groups<-ncol(data)
n.pers <-nrow(data)
- 10 *<sortiere Geburtstage jeder Gruppe 10>*≡
data<-apply(data,2,sort)
- 11 *<stelle Differenzen zwischen den Geburtstagen jeder Gruppe fest 11>*≡
spacings<-apply(data,2,diff)
- 12 *<sortiere Differenzen jeder Gruppe 12>*≡
spacings<-apply(spacings,2,sort)
- 13 *<stelle in jeder Gruppe die Anzahl gleicher Differenzen fest 13>*≡
equal.spacings<-apply(spacings,2,function(x) sum(x[-1]==x[-(n.pers-1)]))
- 14 *<erstelle empirische Häufigkeitstabelle der Anzahl gleicher Differenzen 14>*≡
f.equal.spacings.emp<-table(equal.spacings)
- 15 *<ermittle theoretische Häufigkeiten 15>*≡
f.equal.spacings.theo<-bsvert2(n.pers,n.days,n.groups=1000)
- Der Vergleich mittels χ^2 -Test ist jetzt nur noch eine Fingerübung.
- 16 *<vergleiche mittels χ^2 -Test 16>*≡
<extrahiere gemeinsame Klassen 17>
<kombiniere zu dünn besetzte Klassen 18>
<berechne Prüfgröße 19>
- 17 *<extrahiere gemeinsame Klassen 17>*≡
values<-intersect(names(f.equal.spacings.emp),names(f.equal.spacings.theo))
n.emp <-f.equal.spacings.emp[values]
n.theo<-f.equal.spacings.theo[values]

Vorsicht: Falls Klassen zu dünn besetzt sind, müssen sie zusammengelegt werden.

- ```
18 <kombiniere zu dünn besetzte Klassen 18>≡
 repeat{
 to.combine <- which(n.emp < 5)[1]
 if(is.na(to.combine)) break
 delta <- if(to.combine==length(n.emp)) -1 else 1
 n.emp [to.combine+delta] <- n.emp [to.combine+delta] + n.emp [to.combine]
 n.theo[to.combine+delta] <- n.theo[to.combine+delta] + n.theo[to.combine]
 n.emp<-n.emp[-to.combine]; n.theo<-n.theo[-to.combine]
 }

19 <berechne Prüfgröße 19>≡
 pg <- sum(n.emp-n.theo)^2/n.theo
 n.klassen <- length(n.emp)
 prob <- pchisq(pg, n.klassen-1)

20 <liefere Ergebnis ab 20>≡
 paste("Wahrscheinlichkeit einer extremeren Beobachtung unter H_0:",1-prob)
```

## 2.6 Fazit

Fertig. Wir haben oben die Frage gestellt, wie wir Zufallszahlen mit Hilfe eines BS-Tests überprüfen können. Dazu haben wir skizzenhaft beschrieben, was für diesen Test zu tun ist und wie er implementiert werden könnte. Das läßt sich mit geringen Grundwissen verstehen. Für einen Einsatz sind natürlich noch Einsatzüberlegungen anzustellen, die zum Beispiel mit der Parameterwahl zusammenhängen. Zur Interpretation der Ergebnisse muß der Anwender das Phänomen eines statistischen Tests verstanden haben. Auch sollte er wissen, warum der  $\chi^2$ -Test gerade  $\chi^2$ -Test heißt, was er mit der Normalverteilung zu tun hat und welche Annahmen dem Test zugrundeliegen. Antworten hierauf erfordern ein Tiefenverständnis und Methodenwissen. Eine genaue Betrachtung solcher spannenden Fragen zeigt oft, daß sie sich auf Zusammenhänge von Zufallsvariablen zurückführen lassen und damit auf die Frage:

- Welche Verteilung besitzt die Zufallsvariable  $W$  und warum?

Deshalb stehen in diesem Skript *Zufallsvariablen* und der Umgang mit ihnen im Zentrum. Als tragendes Medium wurden fast mehr zufällig verschiedene *Warte-Fragen* ausgewählt.

## 3 Warten auf die nächste Party

In diesem Abschnitt wollen wir uns der Verteilung der Wartezeit auf die nächste Party in Form einiger unsortierter Gedanken widmen. Für die Herleitung der theoretischen Wartezeitverteilung stolpern wir über eine Reihe von technischen Problemen. Das ist sehr schön, da wir so im Zuge der Diskussion verschiedene Techniken der Statistik kennenlernen können.

### 3.1 Problemstellung

Im Zentrum dieses Abschnittes steht die Frage: Wie lange muß man auf den nächsten Geburtstag warten, wenn gerade ein Geburtstag gefeiert worden ist?

Erste Überlegungen ergeben, daß für einen einzelnen Geburtstages als Modell eine Gleichverteilung unterstellt werden kann. In Nicht-Schalt-Jahren ergibt sich damit eine Wahrscheinlichkeit von  $1/365$ , daß eine zufällig ausgewählte Person an einem bestimmten Tag, zum Beispiel heute, Geburtstag hat. Wenn sich heute ein Geburtstag eingestellt hat, welche Verteilung besitzt dann die Wartezeit auf den nächsten?

#### 3.1.1 Fehlversuch: Geometrische Verteilung

Die gesuchte Verteilung muß offensichtlich von diskreter Natur sein. Aus der Erinnerung könnte eine geometrische Verteilung (Warten auf eine Sechs beim Mensch-ärgere-Dich-nicht-Spiel) als Kandidat vorgeschlagen werden. Dieser Vorschlag wird jedoch schnell per Widerspruch abgelehnt werden. Denn spätestens in genau einem Jahr ist wieder eine Geburtstagsfeier fällig und – sofern nicht alle Personen der betrachteten Gruppe an einem Tag ihren Geburtstag haben –, muß sogar vor Ablauf eines Jahres eine weitere Party steigen. Bei einer geometrischen Verteilung können sich dagegen beliebig große Wartezeiten – wenn auch mit zunehmend geringer werdenden Wahrscheinlichkeiten – einstellen. Die geometrische Verteilung ist deshalb kein geeigneter Kandidat für die Wartezeitverteilung. Eventuell kann die geometrische Verteilung zur Approximation herangezogen werden.

#### 3.1.2 Versuch einer Simulation der Wartezeit

In diesem Abschnitt wollen wir experimentell Wartezeiten ermitteln. Dabei werden wir feststellen, daß eine einfache Simulation zu seltsamen Ergebnissen führt. Deshalb ist es angebracht, sich die Zusammenhänge etwas näher anzuschauen.

Geburtstage lassen sich leicht für Gruppen von Personen ermitteln und in einer Matrix ablegen.

```
21 <ziehe n.groups vom Umfang n.pers aus 1 bis n.days: stpr 21>≡
 stpr<-matrix(sample(1:n.days, size=n.groups*n.pers, replace=T),n.pers,n.groups)
```

Realisationen generieren wir durch:

```
22 <* 1>+≡
 n.pers<-5
 n.groups <- 7
 n.days<-365
 set.seed(13)
 <ziehe n.groups vom Umfang n.pers aus 1 bis n.days: stpr 21>
```

```
Thu Jun 5 18:27:46 2003
 [,1] [,2] [,3] [,4] [,5] [,6] [,7]
[1,] 81 115 168 179 250 165 281
[2,] 233 328 30 103 317 225 277
[3,] 165 200 154 272 153 137 47
[4,] 79 282 126 175 141 43 239
[5,] 244 251 31 155 301 357 193
```

**Experiment zur *verwunderlichen* Wartezeitverteilung** Betrachten wir einmal folgende Idee: Wir simulieren die Geburtstage verschiedener Gruppen gleicher Personenanzahl. Dann sortieren wir in jeder Gruppe die Geburtstage. Zum Schluß stellen wir die empirische Verteilung dar, die zeigt, wie lange nach dem  $j$ -ten Geburtstag auf die nächste Party gewartet werden muß. Hierbei darf nicht vergessen werden, daß die Person, die im Jahr zuletzt Geburtstag hat, sich auch auf die nächste Party freut.

Nach der Intuition müssen die Abstände zwischen zwei Geburtstagen ( Warten auf die nächste Fete ) immer identisch verteilt sein.

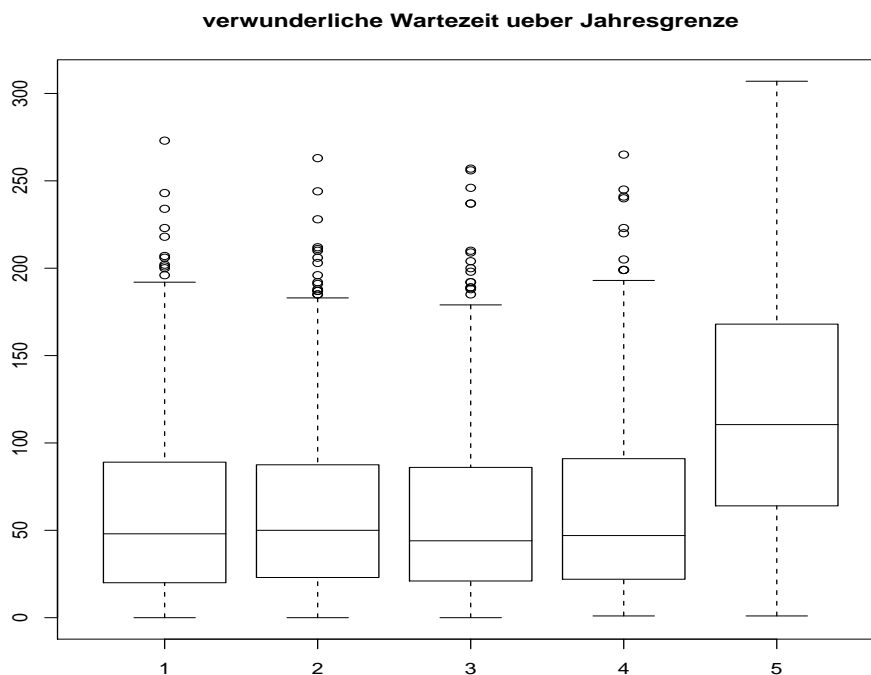
**Algorithmus:**

1. ziehe wiederholt (für viele Personengruppen) Stichproben aus  $\{1, \dots, 365\}$
2. sortiere Geburtstage innerhalb jeder Gruppe
3. berechne Wartezeiten nach dem jeweils  $j$ -ten Geburtstag innerhalb einer Personengruppe
4. stelle für jedes  $j$  die empirische Verteilung der Wartezeiten dar.

o

Hier die Umsetzung:

```
23 <* 1>+≡
Vorbereitung
n.pers<-5
n.groups <- 100
n.days<-365
set.seed(13)
Stichprobenziehung
<ziehe n.groups vom Umfang n.pers aus 1 bis n.days: stpr 21>
Berechnung der Wartezeiten
tage<-apply(stpr,2,sort)
dtage<-apply(tage,2,diff)
Wartezeit ueber die Jahresgrenze
dtage<-rbind(dtage, (n.days+tage[1,])-tage[n.pers,])
Darstellung
boxplot(split(dtage,row(dtage)))
title("verwunderliche Wartezeit ueber Jahresgrenze")
```



Was ist schiefgegangen? Oder läßt sich der seltsame Boxplot der Personen, die zuletzt im Jahr Geburtstag feiern, erklären?

### 3.1.3 Vorgehensstrategie

Wir wollen dieses Geburtstagproblem Schritt für Schritt verstehen. Dazu wollen wir anhand folgender Fragen vorgehen:

1. Was ändert sich, wenn wir die Länge des Jahres umskalieren? Es hat sich nämlich oft als vereinfachend erwiesen, ein Problem in einen standardisierten Bereich zu überführen. Weiterhin sind mittels Restandardisierungsoperationen allgemeine Aussagen über Bereiche von  $a$  bis  $b$  zu erwarten.
2. Welche Modell-Klassen lassen sich zur Modellierung der Party-Wartezeiten unterscheiden? In der Regel gibt es immer verschiedene Ansätze der Modellierung. Manche sind besser zu handhaben, andere realitätsnäher. In diesem Abschnitt soll dargestellt werden, daß wir zum einen mit einer diskreten oder kontinuierlichen Weltvorstellung, zum anderen das Jahr mittels eines Kreises oder eines Zahlenintervalls modellieren können.



3. Welche Erkenntnisse liefert ein Gedankenexperiment mit einer Modellierung des Jahres als Kreis? Genau wie ein Kreis sich durch seine Symmetrie auszeichnet, folgt, daß die Wartezeiten nach einem zufällig gewählten Geburtstag identisch sein müssen.
4. Wie lassen sich Erkenntnisse der Kreiswelt auf die lineare Welt übertragen und umgekehrt? Im Umgang mit Intervallen liegt eine große Erfahrungssammlung vor. Wenn es gelingt, Erkenntnisse zu transferieren, ist viel gewonnen. Zum Beispiel kann dann die Wartefrage der Kreiswelt mit Mitteln der linearen Welt beantwortet werden.
5. Welche eindimensionale Verteilung besitzen  $U_{(1)}$  und die Abstände zwischen zwei Geburtstagen? Wenn wir zwischen Intervall- und Kreiswelt hin- und herspringen können und (alle) Verteilungen aus der linearen Welt kennen, sind auch die entsprechenden Probleme der Kreiswelt gelöst.
6. Welche Abweichungen bestehen zwischen diskreter und kontinuierlicher Welt? Wenn die Abweichungen klein sind, kann ohne Probleme ein kontinuierliches Modell für eine diskrete Fragestellung verwendet werden.
7. Wie steht es denn nun mit der verlängerten Wartezeit zum Jahreswechsel? Hier wird nach einigen Experimenten entlarvt, daß die Verwendung des Wortes Jahreswechsel in der Frage zu einer Bedingung führt, die einen vergrößernden Einfluß auf die Wartezeit hat.
8. Läßt sich die Verteilung der Geburtstagsabstände in einer linearen Welt auch direkt ermitteln? Eine direkte Ermittlung ist, wie gezeigt wird, möglich, jedoch komplizierter. Als Entschädigung bekommt der Leser eine weitere Demonstration für den Umgang mit Transformationen im Multivariaten sowie einige Verteilungszusammenhänge geboten.

Wir werden sehen, daß sich Modelle zu unserer Problematik durch drei Eigenschaften unterscheiden können. Diese werden auch zur Gliederung der weiteren Ausführungen beitragen.

- Definitionsbereich: diskret  $\leftrightarrow$  kontinuierlich
- Bereich möglicher Realisationen: von 1 bis 365  $\leftrightarrow$  von 0 bis 1
- Modellierung auf Basis: Zahlen-Intervall  $\leftrightarrow$  Kreis

Zunächst wenden wir uns der Bereichsfrage zu.

## 3.2 Umskalierungen

Wenden wir uns zunächst der Bereichsfrage zu: Welche Auswirkungen haben Umskalierungen auf Modelle?

Diese wollen wir allgemein angehen, um dann die Übertragung auf das konkrete Problem hinzukriegen. Die konkrete Frage lautet: Wenn wir ein Jahr mit einer diskreten Zufallsvariablen modellieren, die in einem Bereich von  $\{1, \dots, 365\}$  Wahrscheinlichkeiten größer 0 besitzt, wie verändern sich die Wahrscheinlichkeiten beim Übergang auf einen anderen Bereich, zum Beispiel auf  $\{1/365, \dots, 1\}$  oder bei einer Reduktion der Problemstellung auf Wochen? Entsprechend: Wenn wir ein Jahr mit einer kontinuierlichen Zufallsvariablen mit einem Bereich positiver Dichte von  $[0, 365]$  modellieren, wie können wir auf einen Bereich von  $[a, b]$  bzw.  $[0, 1]$  schließen.

Allgemein läßt sich dieses als Transformationsproblem formulieren: Gegeben ist eine diskrete oder kontinuierliche Zufallsvariable mit Wahrscheinlichkeitsfunktion oder Dichtefunktion  $f(x)$ . Wie erhalten wir dann die Wahrscheinlichkeitsfunktion oder Dichtefunktion von  $g(f(x))$ ?

**Diskreter Fall** Betrachten wir als Beispiel einen Würfel, der einmal geworfen wird. Als Modell verwenden wir eine Zufallsvariable  $X$  mit:

$$f_X(x) = 1/6 \quad \text{mit} \quad x = 1, \dots, 6$$

Eine einfache Verschiebung ändert an den Wahrscheinlichkeiten nichts. Für  $Y = X + a$  folgt deshalb:

$$f_Y(y) = P_Y(Y = y) = P_X(X + a = y) = P_X(X = y - a) = 1/6 \quad \text{mit} \quad y = 1 + a, \dots, 6 + a$$

Auch eine Umskalierung der Form  $Z = bX$  ist intuitiv beherrschbar.

$$f_Z(z) = P_Z(Z = z) = P_X(bX = z) = P_X(X = z/b) = 1/6 \quad \text{mit} \quad y = 1 \cdot b, \dots, 6 \cdot b$$

Das Transformationsproblem läßt sich häufig mittels der sogenannten Verteilungsfunktionsmethode lösen: Problemstellung

$$X \sim F_X(x) \quad \wedge \quad Y = g(X) \quad \Rightarrow \quad F_Y(y) = ?$$

Lösung, falls  $g(\cdot)$  umkehrbar und streng monoton steigend ist, folgt

$$F_Y(y) = P_Y(Y \leq y) = P_X(g(X) \leq y) = P_X(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

**Reflexion** Warum ist das so? Um eine Wahrscheinlichkeits-Frage bezüglich  $Y$  oder  $Z$  zu beantworten, schauen wir, wie wir die Frage in äquivalenter Form mit Hilfe der Zufallsvariablen  $X$  formulieren können. Nach Beantwortung der Frage mittels  $X$  wird die Antwort zurückübertragen auf die Welt von  $Y$  bzw.  $Z$ . Die Regel lautet:

Zur Beantwortung einer Frage über  $Y = g(X)$  übertrage diese in die Welt von  $X$ , finde die Antwort und übertrage das Ergebnis zurück auf  $Y$ .

Der in der Regel formulierte Rückgriff ist nicht nur für eine konkrete Frage hilfreich. Es lassen sich ebenfalls ganz allgemeine Zusammenhänge zum Beispiel für Wahrscheinlichkeitsfunktionen lösen. Hierbei kann die Anwendung elementarer Gesetze der Wahrscheinlichkeitsrechnung auf mit dem ursächlichen Zufallsexperiment verbundenen Wahrscheinlichkeiten helfen. Formal ausgedrückt erhalten wir:

$$P_X(g(X) \leq y) = P_\Omega(\{\omega | g(X(\omega)) \leq y\})$$

Wie läßt sich die Korrektheit der Regel absichern? Die kürzeste Antwort ist mit dem Stichwort Wahrscheinlichkeitsrechnung gegeben. Deshalb ist an dieser Stelle ein Ausflug in die Grundlagen zur Erinnerung an einige elementare Begriffe angeraten:

- Stichprobenraum
- Ereignisse als Teilmenge des Stichprobenraums
- Algebra von Ereignissen
- Axiome der Wahrscheinlichkeitsrechnung
- Gesetz von der Gegenwahrscheinlichkeit
- Wahrscheinlichkeiten für nicht schnittlere Ereignisse
- stochastische Unabhängigkeit
- bedingte Wahrscheinlichkeit
- Zufallsvariable
- Verteilungsfunktion, Wahrscheinlichkeitsfunktion und Dichtefunktion

**Kontinuierlicher Fall** Die Problemstellung lautet wieder:

$$X \sim F_X(x) \quad \wedge \quad Y = g(X) \quad \Rightarrow \quad F_Y(y) = ?$$

Unter der Annahme, daß die Dichte  $f(x)$  gutmütig und  $g(x)$  differenzierbar und streng monoton steigend ist, ergibt sich:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dP_Y(Y \leq y)}{dy} = \frac{dP_X(g(X) \leq y)}{dy} = \frac{dP_X(X \leq g^{-1}(y))}{dy} = \frac{dF_X(g^{-1}(y))}{dy}$$

Nach den Ableitungsregeln erhalten wir:

$$\frac{dF_X(g^{-1}(y))}{dy} = f_X(g^{-1}(y)) \cdot \frac{dg^{-1}(y)}{dy}$$

kurz:

$$f_Y(y) = f_X(g^{-1}(y)) \cdot g^{-1}(y)'$$

Dieser Zusammenhang wird in vielen Büchern in dem *Transformationssatz für Dichten* zusammengefaßt. Technisch entspricht die Regel der Substitutionsmethode zur Berechnung von Integralen.

Oft werden noch einige Verallgemeinerungen abgehandelt. Falls zum Beispiel  $g(\cdot)$  streng monoton fallend ist, ist die Ableitung der Umkehrfunktion negativ, so daß eine Betragsbildung notwendig wird. Am Prinzip ändert sich jedoch nichts. Weiterhin lassen sich Transformationen betrachtet, die nur abschnittsweise umkehrbar sind. In diesen Fällen kann es zu einem  $y$ -Wert mehrere Urbilder geben, so daß sich eine Formulierung mit einer Summe über die Abschnitte ergibt. Jeder Summand ist wieder von der gerade betrachteten Struktur.

**Aufgabe 1:** *Suche Darstellungen und Anwendungen des Transformationssatz für Dichten in dem Buch von Mood, Graybill, Boes.* ⊥

Mit diesem Wissen sind im Besonderen lineare Transformationen kein Problem, da die Transformationsfunktion – bis auf  $g(x) = 0$  – die Voraussetzungen problemlos erfüllt. Für

$$Y = g(X) = b \cdot X + a \Rightarrow g^{-1}(y) = \frac{y - a}{b} \Rightarrow f_Y(y) = |1/b| f_X((y - a)/b)$$

Die Transformationsfunktion kommt uns von Standardisierungsprozessen bekannt vor.

**Aufgabe 2:** *Zur Übung können wir uns folgende Aufgaben stellen:*

- $U \sim U(0, 1)$ , wie ist dann  $bU + a$  verteilt und warum?
- $U \sim U(0, 365)$ , wie läßt sich die Beziehung zu  $U(0, 1)$  formulieren und begründen?
- zeige, daß  $Y = X^2$  die folgende Dichte hat, wenn  $X \sim N(0, 1)$ :

$$f_Y(y) = \frac{0.5^{(k/2)} y^{k/2-1} e^{-(y/2)}}{\Gamma(k/2)} I_{[0, \infty)}(x), \quad k = 1$$

- lese – sofern noch nicht geschehen – Mood, Graybill, Boes, Kapitel V, Abschnitt 6.

⊥

### 3.3 Klassen der Modellierung

Für unsere Ausgangsfrage erscheint es angemessen, die Tage des Jahres auf einem Kreis anzuordnen, Werte von 1 bis 365 abzutragen und ein diskretes Modell zu untersuchen. Ein kontinuierliches Modell, das sich auf ein Intervall von 0 bis 1 bezieht, bietet uns jedoch einen besseren Zugang. Wenn wir nun das Bereichsproblem als beherrschbar einstufen, verbleibt eine Diskussion folgender vier Fälle. Die Erkenntnisse sind dann in einem zweiten Schritt auf die Party Frage zu übertragen.

|                | Intervall | Kreis |
|----------------|-----------|-------|
| diskret        | A         | B     |
| kontinuierlich | C         | D     |

Genauer wollen wir uns mit folgenden Fragen beschäftigen:

- Welche Erkenntnisse lassen sich für ein 0-1-Intervallmodell formulieren?
- Wie lassen sich Erkenntnisse eines Intervallmodells auf ein Kreismodell übertragen?
- Läßt sich die diskrete Welt gut durch ein kontinuierliches Modell annähern?
- Wie lassen sich diese Erkenntnisse auf die Party-Frage übertragen?

### 3.4 Gedankenexperiment: das Jahr als Kreis

Wir wollen mit einem Modell mit den Eigenschaften *kontinuierlich, von 0 bis 1, kreisbasiert* (Fall D) beginnen. Denn dieses Modell liefert einen leichten Ansatzpunkt per Gedankenexperiment.

Stellen wir uns vor, das Jahr wird als Kreislinie dargestellt und die Geburtstage werden durch zufällige unabhängige Einschlüge auf der Kreislinie festgelegt. Dann ist aus Symmetriegründen einleuchtend, daß die Verteilungen der Wartezeiten auf die jeweils nächste Party identisch sein müssen. Und zwar unabhängig davon, an welcher Stelle der Jahresbeginn festgelegt wird bzw. wo wir – wie bei einer Uhr – den Nullpunkt plazieren. Wir wissen jedoch nicht, um welche Verteilung es sich dabei handelt.

Nach der Diskussion über Bereiche ist es weiter einleuchtend, daß sich durch Umskalierung zum Beispiel durch einen Wechsel der Einheit zu Grad, also zu einer Skala von 0 bis 360, prinzipiell nichts ändert. Weiter müssen die Wartezeitverteilungen identisch bleiben, wenn wir alle Zahlen auf ganze Zahlen runden und so zu einem Glücksrad mit 365 Fächern übergehen.

Als Zwischenergebnis können wir festhalten, daß in allen Kreismodellen die Wartezeiten auf die nächste Party durch eine einzige Verteilung beschreibbar ist. Unklar ist, welche Gestalt diese Verteilung besitzt.

### 3.5 Übergänge zwischen Intervall- und Kreiswelt

Dieser Abschnitt handelt von Beziehungen zwischen Fall D und Fall C. Die folgenden Ausführungen basieren auf Mardia (1972): *Directional Statistics*.

Betrachten wir ein Zufallsexperiment, bei dem sich gleichverteilt Positionen auf einem Kreis realisieren können. Diese Positionen lassen sich nach Festlegung eines Nullpunktes und eines Umlaufsinnns durch ihren Winkel (aus dem Intervall  $[0, 2\pi]$ ) charakterisieren. Kurz wollen wir für eine Stichprobe vom Umfang  $n$  schreiben:

$$\theta_1, \dots, \theta_n \stackrel{iid}{\sim} CU(0, 2\pi)$$

Hierbei soll die Abkürzung *CU* stehen für *circular uniform*. Diese Winkel wollen wir durch Rotation so transformieren, daß der neue Nullpunkt mit  $\theta_n$  zusammenfällt:

$$\theta_i^* \leftarrow \theta_i - \theta_n \pmod{2\pi}$$

Mit dieser Transformation geht  $(\theta_1, \dots, \theta_n)$  über in  $(\theta_1^*, \dots, \theta_{n-1}^*, 0)$ . Nun suchen wir die Verteilung von  $\theta^* = (\theta_1^*, \dots, \theta_{n-1}^*)$ . Diese Verteilung erhalten wir, indem wir die Dichte von  $(\theta_1, \dots, \theta_n)$  transformieren und  $\theta_n$  herausintegrieren:

$$f_{\theta}(\theta_1, \dots, \theta_n) = \prod_{i=1}^n f_{\theta_i}(\theta_i) = f_{\theta_n}(\theta_n) \prod_{i=1}^{n-1} f_{\theta_i}(\theta_i)$$

Durch Einsetzen der Transformation für  $(\theta_1, \dots, \theta_{n-1})$  unter Beibehaltung von  $\theta_n$  erhalten wir:

$$f_{\theta_n}(\theta_n) \prod_{i=1}^{n-1} f_{\theta_i^*}(\theta_i^* + \theta_n)$$

$\theta_n$  können wir nun herausintegrieren:

$$f_{\theta^*}(\theta_1^*, \dots, \theta_{n-1}^*) = \int_0^{2\pi} f_{\theta_n}(\theta_n) \prod_{i=1}^{n-1} f_{\theta_i^*}(\theta_i^* + \theta_n) d\theta_n$$

Wegen Unabhängigkeit und Gleichverteilung erhalten wir:

$$f_{\theta^*}(\theta_1^*, \dots, \theta_{n-1}^*) = 1/(2\pi)^{n-1}$$

Weiter bekommen wir die gemeinsame Dichte der Ordnungsstatistik  $\theta_0^*$  leicht durch Multiplikation der Anzahl der Urbilder (mit gleicher Dichte) mit dem Dichtewert eines speziellen

Urbildes. Als spezielles Urbild betrachten wir den Stichprobenvektor, bei dem die  $i$ -te Stichprobenvariable gerade die  $i$ -t kleinste Variable ist.

$$f_{\theta_0^*}(\theta_{(1)}^*, \dots, \theta_{(n-1)}^*) = (n-1)! \cdot 1/(2\pi)^{n-1} \cdot I(\theta_{(1)}^*, \dots, \theta_{(n-1)}^*)$$

– hierbei sind für die Variablen die in mit  $I(\cdot)$  abgekürzten Bedingungen

$$0 < \theta_{(1)}^* < \theta_{(2)}^* < \dots < \theta_{(n-2)}^* < \theta_{(n-1)}^* < 2\pi$$

zu beachten.

**Aufgabe 3:** *Wie als Übung nachgeprüft werden kann, ist die Verteilung dieser Ordnungsstatistik identisch mit der Verteilung der Ordnungsstatistik  $(U_{(1)}, U_{(2)}, \dots, U_{(n-1)})$  einer iid Stichprobe  $(U_1, U_2, \dots, U_{n-1})$  vom Umfang  $(n-1)$  aus  $U(0, 2\pi)$ . Außerdem ist Verteilung der ungeordneten Stichprobe identisch mit der Verteilung der von  $(\theta_1^*, \dots, \theta_{n-1}^*)$ .  $\perp$*

Wir können mit dieser Erkenntnis Fragestellungen zu einem Stichprobenumfang von  $n$ , welche in der Kreiswelt angesiedelt sind, mit Hilfe der bekannten linearen Welt unter Verwendung einer Stichprobe vom Umfang  $n-1$  beantworten.

Mit Hilfe des Zusammenhangs linearer / Kreiswelt können wir nun zufällige (gleichverteilte) Punkte auf einen Kreis durch Übertragung aus der linearen Welt generieren, zum Beispiel, um  $n$  *Geburtstage* in einem Jahr mit dem Bereich  $[0, 2\pi)$  zu simulieren:

**Algorithmus:**

1. ziehe  $n-1$  Zufallszahlen  $(U_1, U_2, \dots, U_{n-1})$  aus  $U(0, 2\pi)$ ,

2. ergänze den Wert 0:  $\rightarrow (U_1, U_2, \dots, U_{n-1}, 0)$ ,

3. ziehe eine weitere Zufallszahl  $U_n$  aus  $U(0, 2\pi)$  und

4. addiere diese zu dem Vektor, so daß entsteht:

$$\rightarrow (U_1 + U_n, U_2 + U_n, \dots, U_{n-1} + U_n, U_n).$$

5. vermindere zum Schluß Werte, die größer als  $2\pi$  sind:

$$\rightarrow (U_1 + U_n, U_2 + U_n, \dots, U_{n-1} + U_n, U_n) \pmod{2\pi}.$$

o

**Lohnenswerte Erinnerungen** Im Rahmen der gerade durchgeführten Überlegungen lassen sich wieder einige zentrale Grundkonzepte diskutieren, die sich wesentlich um den Umgang mit mehrdimensionalen Verteilungen drehen.

- Betrachtung von 2-dimensionalen Kontingenztabelle mit relativen Häufigkeiten mit Diskussion der Begriffe: gemeinsame Verteilung, Randverteilungen und Randverteilungen.
- Gegenüberstellung von Fragen zu Kontingenztabelle zu elementaren Zusammenhängen der Wahrscheinlichkeitsrechnung sowie Sätzen über Wahrscheinlichkeitsfunktionen und Sätzen über Dichten unter den Aspekten:
  - Randverteilungen
    - \* Bildung von Randverteilungen einer Kontingenztabelle
    - \*  $P(A) = \sum_j P(A \cap B_j)$ , falls  $B_j$  Zerlegung von  $\Omega$  bilden
    - \* Summation über eine Variable:  $f(x_i) = \sum_j f(x_i, y_j)$
    - \* Herausintegration einer Variable  $f(x) = \int_{-\infty}^{\infty} f(x, y) dy$
  - bedingte und Rand-Verteilungen
    - \* Falls  $B_j$  Zerlegung bilden:  $P(A) = \sum_j P(A|B_j)P(B_j)$
    - \* Summation über eine Variable:  $f(x_i) = \sum_j f(x_i|y_j)f_Y(y_j)$
    - \* Herausintegration einer Variable:  $f(x) = \int_{-\infty}^{\infty} f_Y(y)f(x|y) dy$
  - Wiederholung der Begriffe der stochastischen Unabhängigkeit:
    - \* Zellenhäufigkeiten über Ränder bestimmen bei Unabhängigkeit
    - \*  $P(A) = P(A|B) \Rightarrow P(A) \cdot P(B) = P(A \cap B)$  bei Unabhängigkeit
    - \*  $f(x_i, y_j) = f_X(x_i) \cdot f_Y(y_j)$  bei Unabhängigkeit
    - \*  $f(x, y) = f_X(x) \cdot f_Y(y)$  bei Unabhängigkeit
- Zusammenhang Dichte / Verteilungsfunktion
- Diskussion, was bei einer Statistik für Winkel anders sein muß.

### 3.6 Verteilung von Wartezeiten in der linearen Welt

In diesem Abschnitt steht der Fall C im Mittelpunkt. Stellen wir uns ein Intervall von 0 bis 1 vor, in das zufällig und unabhängig  $k$  Markierungen eingeschlagen werden. Hierdurch erhalten wir eine Stichprobe

$$U_1, \dots, U_k \sim U(0, 1)$$

Durch Sortierung gehen wir über zu:

$$U_{(1)}, \dots, U_{(k)}$$



$U_{(1)}$  ist das Minimum der Stichprobe,  $U_{(k)}$  das Maximum. Es erscheint nach dem Gedankenexperiment mit dem Kreis auch hier plausibel, daß die Differenzen  $U_{(i)} - U_{(i-1)}$  für unterschiedliche Werte von  $i$  identische Verteilungen besitzen, nämlich gerade die Verteilung von  $U_{(1)}$ . Um nicht einer falschen Intuition aufzusitzen, wollen wir uns die Zusammenhänge näher ansehen.

**Die Verteilung von  $U_{(1)}$ .** Welche Verteilung hat  $U_{(1)}$ ? Für die Verteilungsfunktion von  $U_{(1)}$  gilt:

$$\begin{aligned}
 P(U_{(1)} \leq u) &= 1 - P(U_{(1)} > u) \\
 &= 1 - P(U_1 > u, \dots, U_k > u) \\
 &= 1 - P(U_1 > u) \cdots P(U_k > u) \\
 &= 1 - P(U > u)^k \\
 &= 1 - (1 - P(U \leq u))^k \\
 &= 1 - (1 - F_U(u))^k \\
 &= 1 - (1 - u)^k
 \end{aligned}$$

Für die Dichte ergibt sich:

$$f_{U_{(1)}}(u) = F'_{U_{(1)}}(u) = k \cdot (1 - u)^{k-1}$$

Dies ist die Dichte einer Beta-Verteilung mit den Parametern 1 und  $k$ . Aus Symmetriegründen ist damit übrigens auch die Verteilung von  $1 - U_{(k)}$  gefunden.

Dieses Ergebnis können wir auf die Kreiswelt zurückübertragen.  $k$  Markierungen in  $[0,1]$  lassen sich mit  $n = k + 1$  Geburtstagen in der Kreiswelt identifizieren. Für die Differenz von aufeinander folgenden Geburtstagen stellt übrigens die Nullpunktdebatte kein Problem dar. Aufgrund der Symmetrieeigenschaften in der Kreiswelt kann nun gefolgert werden, daß die Abstände zwischen aufeinanderfolgenden Geburtstagen einer Beta-Verteilung mit den Parametern 1 und  $n - 1$  gehorchen.

Für diejenigen, die bei der saloppen Argumentation ein gewisses Unbehagen empfinden, sei darauf hingewiesen, daß dieses Ergebnis noch einmal etwas später unter der Überschrift *Die Verteilung von  $U_{(i)} - U_{(i-1)}$  in der linearen Welt* direkt abgeleitet wird. Zuvor wollen wir ein paar Bemerkungen zu Ungenauigkeiten durch eine kontinuierliche Modellierung anstellen und das Paradoxon der verlängerten Wartezeit über den Jahreswechsel klären.

### 3.7 Abweichung diskreter von kontinuierlicher Modellierung

Wir wollen nun der Frage nachgehen, ob sich bei einer diskreten Modellierung etwas ändert.

Versuchen wir wieder die Umformung mit  $U^d \sim U(1, 2, 3, \dots, n)$ .

$$\begin{aligned}
 P(U_{(1)}^d \leq u) &= 1 - P(U_{(1)}^d > u) \\
 &= 1 - P(U_1^d > u, \dots, U_k^d > u) \\
 &= 1 - P(U_1^d > u) \cdots P(U_k^d > u) \\
 &= 1 - P(U_1^d > u)^k \\
 &= 1 - (1 - P(U_1^d \leq u))^k \\
 &= 1 - (1 - F_{U_1}^d(u))^k \\
 &= 1 - (1 - u/n)^k
 \end{aligned}$$

Für die Wahrscheinlichkeitsfunktion folgt:

$$f_{U_{(1)}}(u) = 1 - (1 - u/n)^k - (1 - (1 - (u - 1)/n)^k) = (1 - (u - 1)/n)^k - (1 - u/n)^k$$

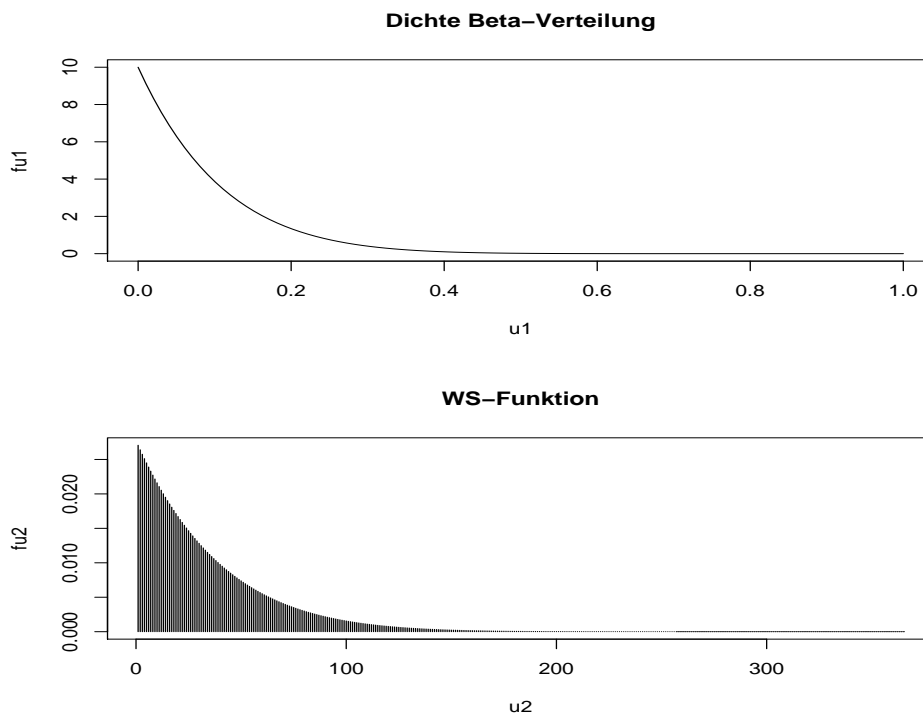
Da die Verteilungsfunktionen bis auf die Streckung gleich aussehen, entspricht die obige Dichtefunktion der Wahrscheinlichkeitsfunktion.

**Experiment** Zur Überprüfung der letzten Behauptung zeichnen wir die Dichte der Beta-Verteilung und die ermittelte Wahrscheinlichkeitsfunktion untereinander.

```

24 <* 1>+≡
 k<-10
 par(mfrow=2:1)
 # Dichte Beta-Verteilung
 u1 <-seq(0,1,length=100)
 fu1<-dbeta(u1,1,k)
 plot(u1,fu1,type="l",main="Dichte Beta-Verteilung")
 # Wahrscheinlichkeitsfunktion
 n.days<-365
 u2<-1:n.days
 fu2<-((1-(u2-1)/n.days)^k-(1-u2/n.days)^k)
 plot(u2,fu2,type="h",main="WS-Funktion")

```



**Interpretation** Die Bilder gleichen sich. Besser ist es, die Darstellungen in einem Bild zusammenzuführen. Um sie dazu skalenmäßig zur Deckung zu bringen, können wir zum Beispiel die Beta-Verteilung strecken. Wieder wird eine (lineare) Transformation erforderlich. Zur Erinnerung allgemein:

$$X \sim f_X \quad \wedge \quad Y = g(X) = a + bX \quad \Rightarrow \quad Y \sim f_Y(y) = \frac{1}{|b|} f_X((y - a)/b)?$$

Einsatz bei der Normalverteilung:

$$Z = \frac{(X - \mu)}{\sigma} \quad \Rightarrow \quad X = \mu + \sigma Z$$

mit

$$F_X(x) = \Phi((x - \mu)/\sigma)$$

sowie

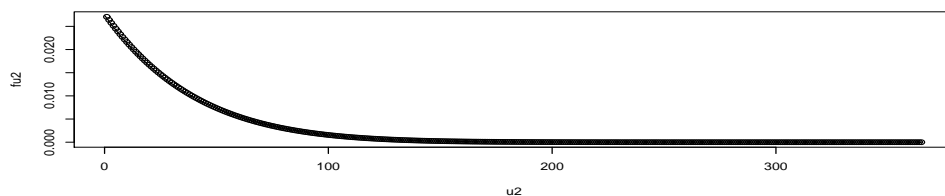
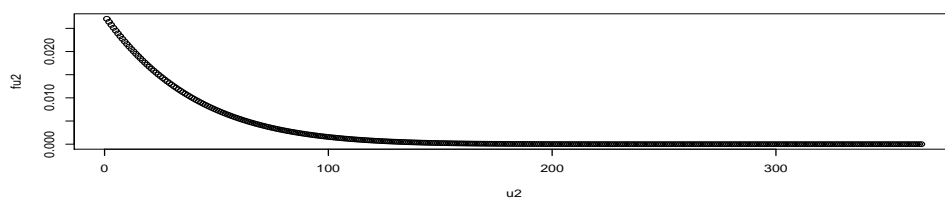
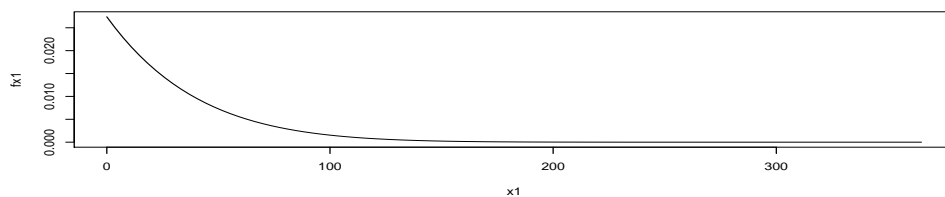
$$f_X(x) = \phi((x - \mu)/\sigma) \cdot \frac{1}{\sigma}$$

Damit korrespondierende Flächen unter der Dichte gleich bleiben, müssen natürlich bei einer Streckung der  $x$ -Achse gleichzeitig die Werte der neuen Dichte verkleinert, genauer mit  $1/\sigma$  multipliziert werden.

**Experiment** Hiermit läßt sich unser Bild verbessern:

25

```
< * 1)+≡
k<-10
n.days<-365
par(mfrow=c(3,1))
kontinuierliche Welt
u1<-seq(0,1,length=100)
fu1<-dbeta(u1,1,k)
x1<-u1*n.days
fx1<-fu1/n.days
plot(x1,fx1,type="l")
diskrete Welt
u2<-1:n.days
fu2<-(((1-(u2-1)/n.days)^k-(1-u2/n.days)^k)
plot(u2,fu2)
beide Welten
plot(u2,fu2); lines(x1,fx1)
par(mfrow=c(1,1))
```



Wie wir sehen, hat sich unser Transformationswissen ausgezahlt.

**Aufgabe 4:** *Was ist eigentlich ein QQ-Plot? Wie können wir die Idee eines QQ-Plots hier zum Einsatz bringen?* ⊥

### 3.8 Experimente zur Frage der verlängerten Wartezeit über den Jahreswechsel

Nach dem Experiment zur Wartezeit sind alle Abstände zwischen zwei Geburtstagen ( Warten auf die nächste Party ) identisch verteilt, bis auf die Wartezeit, die sich an den letzten Geburtstag im Jahr anschließt.

In diesem Abschnitt wollen wir zunächst experimentell diesem Phänomen nachgehen, um Ideen zu bekommen, warum die Wartezeiten über den Jahreswechsel länger ausfallen. Oder liegt in den bisherigen Überlegungen ein Denkfehler vor?

**Vorschlag 1: ziehe Geburtstag auf einem alternativen Weg** Wir haben herausgearbeitet, daß sich Kreis- und lineare Welt aufeinander abbilden lassen. Hierbei wurde oben schon ein alternativer Algorithmus formuliert, Geburtstage festzusetzen. Deshalb wollen wir ein Experiment ersinnen, in dem Geburtstag auf dem alternativen Weg ermittelt werden.

Für die Ziehung von `n.groups` Gruppen a `n.pers` Personen erhalten wir folgende Umsetzung:

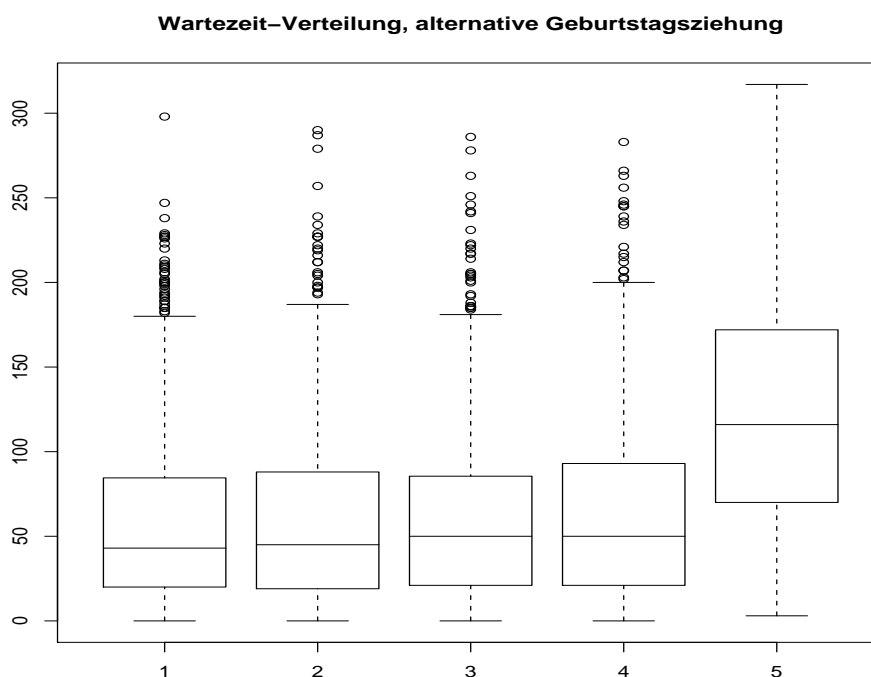
```
26 <ziehe n.groups Umfang n.pers aus Kreis, n.days Tage: stpr 26>≡
ziehe n.groups Stichproben a n.pers aus 1:n.days
stpr <- sample(1:n.days, (n.pers-1)*n.groups, T)
stpr <- matrix(stpr, n.pers-1, n.groups)
definiere 0-Punkt als weiteren Geburtstag
tage.neu <- rbind(stpr, 0)
bestimme zufaellig Jahresanfang
shift <- matrix(sample(1:n.days, n.groups, replace=T), n.pers, n.groups, T)
stpr <- -1 + ((tage.neu + shift - 1) %% n.days)
```

Wenn die Stichproben vorliegen, können wir empirischen Wartezeitverteilungen darstellen.

```
27 <stelle empirische Wartezeitverteilungen aufgrund von stpr dar 27>≡
Berechnung der Wartezeiten
tage<-apply(stpr,2,sort)
dtage<-apply(tage,2,diff)
Wartezeit ueber die Jahresgrenze
dtage<-rbind(dtage, (n.days+tage[1,])-tage[n.pers,])
Darstellung
boxplot(split(dtage,row(dtage)))
title("Wartezeit-Verteilung, alternative Geburtstagsziehung")
```

Hiermit können wir nun das oben angeregte Experiment durchführen:

```
28 <* 1>+≡
n.groups<-1000; n.pers<-5; n.days<-365; set.seed(13)
<ziehe n.groups Umfang n.pers aus Kreis, n.days Tage: stpr 26>
<stelle empirische Wartezeitverteilungen aufgrund von stpr dar 27>
```



Diese Abbildung sieht der obigen leider (?) ziemlich ähnlich.

**Vorschlag 2: setze Jahresbeginn zufällig** Vielleicht haben wir uns irgendwie ungeschickt bei der Auswahl des Jahresbeginns verhalten. Deshalb wollen wir versuchen, nachträglich einen anderen Jahresbeginn umzusetzen.

```
29 <* 1>+≡
 n.groups<-1000; n.pers<-5; n.days<-365; set.seed(13)
 <ziehe n.groups Umfang n.pers aus Kreis, n.days Tage: stpr 26>
 # Festlegung der Verschiebung des Jahresbeginns
 cat("Verschiebung des Jahresbeginns eingeben:")
 shift <- floor(abs(as.numeric(readline()))))
 stpr<-1+((stpr+shift-1) %% n.days)
 # stelle Ergebnis dar
 <stelle empirische Wartevertelungen aufgrund von stpr dar 27>
 print(summary(t(dtage)))
```

Verschiebung des Jahresbeginns eingeben:

readline-Eingabe:

111

Fri Jun 6 14:19:58 2003

|          | X1     | X2     | X3     | X4     |
|----------|--------|--------|--------|--------|
| Min. :   | 0.00   | 0.00   | 0.00   | 0.00   |
| 1st Qu.: | 21.00  | 19.00  | 20.00  | 21.75  |
| Median : | 46.00  | 45.00  | 44.00  | 49.00  |
| Mean :   | 61.62  | 58.46  | 57.39  | 63.72  |
| 3rd Qu.: | 87.00  | 82.00  | 84.00  | 94.00  |
| Max. :   | 273.00 | 250.00 | 262.00 | 277.00 |
|          | X5     |        |        |        |
| Min. :   | 4.0    |        |        |        |
| 1st Qu.: | 73.0   |        |        |        |
| Median : | 119.0  |        |        |        |
| Mean :   | 124.8  |        |        |        |
| 3rd Qu.: | 170.0  |        |        |        |
| Max. :   | 317.0  |        |        |        |

Erstaunlicherweise verändert sich die Grundstruktur der Bilder kaum. Die zusammenfassenden Statistiken zeigen, daß der Faktor 2 im Spiel sein könnte.

**Vorschlag 3: ein empirischer Test** Bisher haben wir die Experimente in einer reinen Modellwelt durchgeführt. Wir könnten dabei einen so fundamentalen strukturellen Fehler begangen haben, der bei allen bisherigen Experimententwürfen durchschlug. Deshalb wollen wir als letztes Experiment reale Daten nehmen und mit diesen empirische Wartezeiten ermitteln. Zunächst beschaffen wir uns ein paar Geburtstage – von Studierenden.

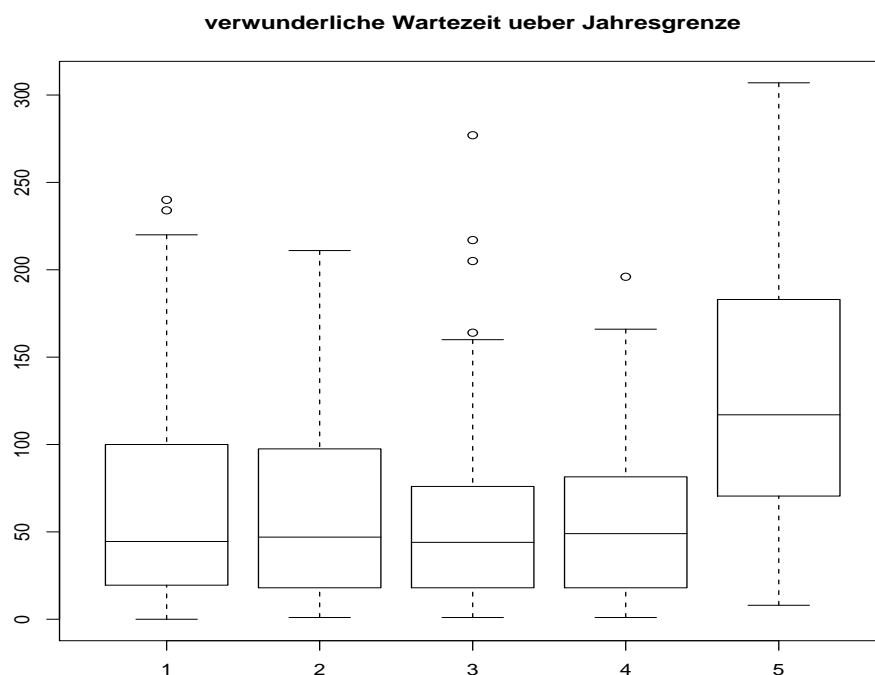
```
30 <stelle Geburtstagsstichprobe bereit 30>≡
 if(any("geburt.stud.R"==list.files())){
 geb.stud.stpr<-source("geburt.stud.R")$value
 }else{
 <lege auf geb.stud.stpr 500 Geburtstage von Studierenden ab 39>
 }
```

Jetzt können wir – wie bei Vorschlag 1 – experimentieren:

```
31 <* 1>+≡
 # Vorbereitung
 n.pers<-5; n.groups <- 100; n.days<-365; set.seed(13)
 # Stichprobenziehung
 <stelle Geburtstagsstichprobe bereit 30>
 stpr<-matrix(sample(geb.stud.stpr, size=n.groups*n.pers, replace=F),n.pers,n.groups)
 # Berechnung der Wartezeiten
 cat("Geburtstage in den Gruppen\n"); print(stpr[,1:10])
 tage<-apply(stpr,2,sort)
 cat("sortierte Geburtstage in den Gruppen\n"); print(tage[,1:10])
 dtage<-apply(tage,2,diff)
 # Wartezeit ueber die Jahresgrenze
 dtage<-rbind(dtage, (n.days+tage[1,])-tage[n.pers,])
 cat("Differenzen zwischen Geburtstagen in den Gruppen\n"); print(dtage[,1:10])
 # Darstellung
 boxplot(split(dtage,row(dtage)))
 title("verwunderliche Wartezeit ueber Jahresgrenze")
```



Wir erhalten:



und zum Nachprüfen die Zwischenprodukte für die ersten 10 Gruppen:

Geburtstage in den Gruppen

|      | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] | [,8] | [,9] | [,10] |
|------|------|------|------|------|------|------|------|------|------|-------|
| [1,] | 78   | 131  | 352  | 123  | 12   | 310  | 266  | 332  | 18   | 148   |
| [2,] | 342  | 86   | 248  | 152  | 126  | 287  | 106  | 46   | 70   | 178   |
| [3,] | 294  | 222  | 262  | 182  | 325  | 355  | 296  | 166  | 266  | 4     |
| [4,] | 196  | 318  | 58   | 134  | 266  | 112  | 320  | 78   | 185  | 249   |
| [5,] | 46   | 50   | 63   | 57   | 209  | 279  | 248  | 42   | 306  | 39    |

sortierte Geburtstage in den Gruppen

|      | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] | [,8] | [,9] | [,10] |
|------|------|------|------|------|------|------|------|------|------|-------|
| [1,] | 46   | 50   | 58   | 57   | 12   | 112  | 106  | 42   | 18   | 4     |
| [2,] | 78   | 86   | 63   | 123  | 126  | 279  | 248  | 46   | 70   | 39    |
| [3,] | 196  | 131  | 248  | 134  | 209  | 287  | 266  | 78   | 185  | 148   |
| [4,] | 294  | 222  | 262  | 152  | 266  | 310  | 296  | 166  | 266  | 178   |
| [5,] | 342  | 318  | 352  | 182  | 325  | 355  | 320  | 332  | 306  | 249   |

Differenzen zwischen Geburtstagen in den Gruppen

|      | [,1] | [,2] | [,3] | [,4] | [,5] | [,6] | [,7] | [,8] | [,9] | [,10] |
|------|------|------|------|------|------|------|------|------|------|-------|
| [1,] | 32   | 36   | 5    | 66   | 114  | 167  | 142  | 4    | 52   | 35    |
| [2,] | 118  | 45   | 185  | 11   | 83   | 8    | 18   | 32   | 115  | 109   |
| [3,] | 98   | 91   | 14   | 18   | 57   | 23   | 30   | 88   | 81   | 30    |
| [4,] | 48   | 96   | 90   | 30   | 59   | 45   | 24   | 166  | 40   | 71    |

[5,] 69 97 71 240 52 122 151 75 77 120

**Aufgabe 5:** *Der Leser möge sich mit den Daten weitere Experimente einfallen lassen, umsetzen und interpretieren.*  $\perp$

### 3.9 Die letzten warten länger: wieso und wie viel?

Jetzt wird es Zeit, die verlängerte Wartezeit desjenigen, der als letzter im Jahr Geburtstag feiert, zu klären. Ausgehend von einem Gedankenexperiment werden wir die Frage in der linearen Welt reformulieren und dann mit Hilfe eines Satzes über Ordnungsstatistiken einer Antwort zuführen. Nebenbei werden wir uns im Transformieren von Zufallsvariablen üben.

**Gedankenexperiment** Als Modell für ein Jahr stellen wir uns eine völlig leere Kreisscheibe vor. Auf diese werfen wir einen Dart-Pfeil ab. Der im Mittelpunkt beginnende und durch die getroffene Stelle verlaufende Strahl soll den Jahresanfang kennzeichnen. Entsprechend der heutigen Kultur wählen wir den Umlaufsinn analog analoger Uhren. Jetzt können wir  $n$  Geburtstag durch  $n$  weitere Dart-Pfeil-Würfe zufällig festlegen. Jedem Pfeil oder Geburtstag läßt sich ein Winkel oder auch ein Tag im Jahr zuordnen. Durch Runden können wir das Experiment mit diskreten Ausgängen ausstatten. Was meint der Leser: Ist durch den Experimentaufbau ein geeignetes Modell für das zugrundeliegende Problem formuliert worden?

**Kreismodell** Nach positiver Beantwortung der letzten Frage können wir das Experiment in der Kreiswelt (mit einer Jahreslänge von 1) formalisieren. Wir betrachten als Jahresanfang:

$$\tilde{U}_0 \sim U(0, 1)$$

und modellieren die Geburtstage durch:

$$\tilde{U}_1, \tilde{U}_2, \dots, \tilde{U}_n \stackrel{iid}{\sim} U(0, 1)$$

Natürlich sollen auch die Geburtstage vom Ausgang des ersten Zufallsexperiments unabhängig sein. Wird diese Formalisierung dem Problem gerecht?

Ja? Dann dürfte sich hieran auch nichts ändern, wenn wir die Zufallsvariablen umbenennen:

$$\tilde{U}_0 \rightarrow U_{n+1}$$

$$\tilde{U}_1 \rightarrow U_1$$

$$\tilde{U}_2 \rightarrow U_2$$

$$\dots \rightarrow \dots$$

$$\tilde{U}_n \rightarrow U_n$$

**Lineare Welt** Der Gewinn dieser Umbenennung liegt darin, daß wir jetzt die Erkenntnisse aus dem Übergang von der Kreis- in die lineare Welt direkt einsetzen können. So können wir die von Mardia (1972) übernommene Transformation anwenden und erhalten:

$$U_1^* \leftarrow U_1 - U_{n+1}$$

$$\dots \leftarrow \dots$$

$$U_n^* \leftarrow U_n - U_{n+1}$$

Diese neuen Zufallsvariablen besitzen dieselbe Verteilung wie  $n$  unabhängig aus einer 0-1-Gleichverteilung gezogene Zufallsvariablen. Entsprechendes gilt für die zugehörigen Ordnungsstatistiken.

Nach diesen Vorüberlegungen läßt sich die Wartezeit  $W$  desjenigen, der als letzter im Jahr Geburtstag hat, identifizieren mit der Summe aus der noch verbleibenden Restzeit bis zum Jahresende und der Zeit von Neujahr bis zum ersten Geburtstag im Jahr:

$$W = (1 - U_{(n)}^*) + U_{(1)}^*$$

**Satz über Ordnungsstatistiken** Für die Wartezeiten, die nicht das Jahresende überschreiten, hatten wir eine Beta-Verteilung mit den Parametern 1 und  $n$  abgeleitet. Wenn  $(1 - U_{(n)}^*)$  und  $U_{(1)}^*$  unabhängig wären, hätten wir die Verteilung der Summe von zwei identisch verteilten unabhängigen beta-verteilten Zufallsvariablen zu ermitteln. Deshalb erscheint es naheliegend, daß sich für große  $n$ , ganz grob eine verdoppelte Wartezeit einstellt. Auch läßt sich überlegen, daß eine sehr große Realisation von  $U_{(1)}^*$  einen verringenden Effekt auf  $(1 - U_{(n)}^*)$  hat – diese beiden Zufallsterme sind also negativ korreliert.

Zur genauen Ermittlung der Verteilung der Summe  $W$  können wir auf Mood, Graybill, Boes, Seite 254, zurückgreifen. Dort lesen wir:

**Satz:**

Sei  $X_1, X_2, \dots, X_n$  eine einfache Zufallsstichprobe aus einer Grundgesamtheit mit Dichte  $f(\cdot)$  und Verteilungsfunktion  $F(\cdot)$  und sei  $Y_1 \leq Y_2 \leq \dots \leq Y_n$  die zugehörige Ordnungsstatistik, dann gilt:

$$f_{Y_\alpha, Y_\beta}(x, y) = \frac{n!}{(\alpha - 1)!(\beta - \alpha - 1)!(n - \beta)!} \times [F(x)]^{\alpha-1} [F(y) - F(x)]^{\beta-\alpha-1} \times [1 - F(y)]^{n-\beta} f(x) f(y) I_{(x, \infty)}(y)$$

▽

**Exakte Verteilung der Wartezeiten** Diesen Satz gilt es für unseren Fall mit Leben zu füllen. Wir identifizieren:

| Satz       | Problem                | im Folgenden abgekürzt durch |
|------------|------------------------|------------------------------|
| $\alpha$   | 1                      |                              |
| $\beta$    | $n$                    |                              |
| $Y_\alpha$ | $U_{(1)}^*$            | $X$                          |
| $Y_\beta$  | $U_{(n)}^*$            | $Y$                          |
| $f(x)$     | $I_{(0,1)}(x)$         |                              |
| $F(x)$     | $x \cdot I_{(0,1)}(x)$ |                              |

So erhalten wir die gemeinsame Dichte von Minimum ( $X$ ) und Maximum ( $Y$ ):

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{n! \times [x]^{1-1} [y - x]^{n-1-1} \times [1 - y]^{n-n} I_{(0,1)}(x) I_{(0,1)}(y) I_{(x, \infty)}(y)}{(1 - 1)!(n - 1 - 1)!(n - n)!} \\ &= \frac{n! x^0 (y - x)^{n-2} (1 - y)^0}{(n - 2)!} \times I_{(0,1)}(x) I_{(0,1)}(y) I_{(x, \infty)}(y) \\ &= n(n - 1)(y - x)^{n-2} \times I_{(0,1)}(x) I_{(0,1)}(y) I_{(x,1)}(y) \end{aligned}$$

Für unsere Frage können wir jetzt diese Dichte so im zweidimensionalen Raum transformieren, daß eine Variable mit dem gesuchten  $W$  übereinstimmt:

$$\left. \begin{array}{l} W \leftarrow X + (1 - Y) \\ Z \leftarrow X - (1 - Y) \end{array} \right\} \Rightarrow \begin{array}{l} X = (W + Z)/2 \\ Y = 1 - (W - Z)/2 \end{array}$$

Für die Determinante der partiellen Ableitungen folgt:

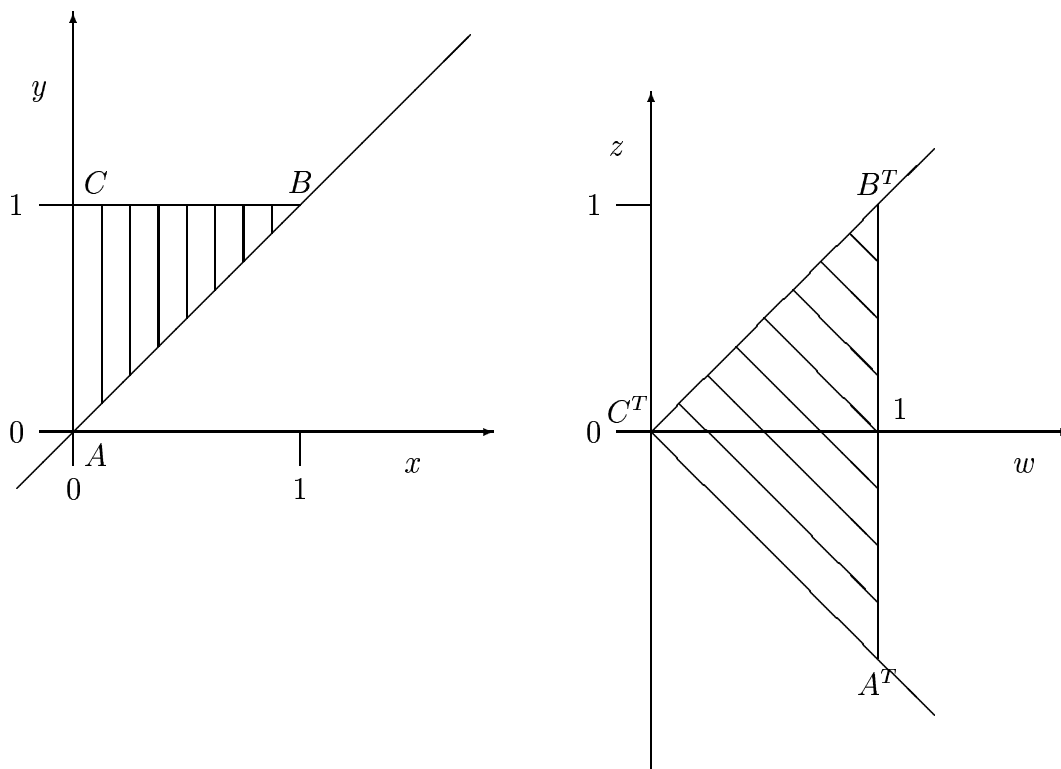
$$|J| = \begin{vmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{vmatrix} = |1/2|$$

Nun liefert der Transformationssatz für Dichten:

$$\begin{aligned} f_{W,Z}(w, z) &= \left| \frac{1}{2} \right| n(n-1) \left( 1 - \frac{w-z}{2} - \frac{w+z}{2} \right)^{n-2} \times I(w, z) \\ &= n(n-1)/2 (1-w)^{n-2} \times I(w, z) \end{aligned}$$

Hierbei faßt  $I(w, z)$  abkürzend die Bedingungen zusammen, die  $w$  und  $z$  einhalten müssen. Diese Bedingungen definieren den Bereich positiver Dichte.

Zunächst wollen wir diesen relevanten Bereich für  $x$  und  $y$  sowie  $w$  und  $z$  graphisch darstellen:



Durch die Transformation wird das schraffierte Dreieck  $ABC$  in der linken Graphik auf das Dreieck  $A^T B^T C^T$  abgebildet.

Die Bedingungen lassen sich auch formal ermitteln. Setzen wir in  $I_{(0,1)}(x)I_{(0,1)}(y)I_{(x,1)}(y)$  die Rücktransformationsbeziehung ein, erhalten wir:

$$I(w, z) = I_{(0,1)}((w + z)/2) \cdot I_{(0,1)}(1 - (w - z)/2) \cdot I_{((w+z)/2,1)}(1 - (w - z)/2)$$

Für Indikatorfunktionen gelten einfache Umformungsregeln:

---

|                      |                    |   |                                              |
|----------------------|--------------------|---|----------------------------------------------|
| Verschiebung:        | $I_{(a,b)}(x + c)$ | = | $I_{(a-c,b-c)}(x)$                           |
| Skalierung:          | $I_{(a,b)}(cx)$    | = | $I_{(a/c,b/c)}(x) \quad \wedge \quad c > 0$  |
| Vorzeichenumkehrung: | $I_{(a,b)}(-x)$    | = | $I_{(-b,-a)}(x)$                             |
| Argumentaustausch:   | $I_{(a,b)}(c)$     | = | $I_{(-\infty,c)}(a) \cdot I_{(c,\infty)}(b)$ |

---

Durch Anwendung solcher Regeln läßt sich  $I(w, z)$  vereinfachen:

$$\begin{aligned} I(w, z) &= I_{(0-w/2, 1-w/2)}(z/2) \cdot I_{(0-1+w/2, 1-1+w/2)}(z/2) \cdot I_{((w+z)/2-1+w/2, 1-1+w/2)}(z/2) \\ &= I_{(-w, 2-w)}(z) \cdot I_{(-2+w, w)}(z) \cdot I_{(2w+z-2, w)}(z) \end{aligned}$$

Die ersten beiden Indikatorfunktionen besagen, daß  $z$  zwischen  $-w$  und  $w$  liegen muß:  $-w < z < w$ . Damit darf  $w$  nicht negativ ausfallen, was sich mit problembereichs-spezifischen Überlegungen deckt. Betrachten wir die dritte Indikatorfunktion. Die aus der Obergrenze resultierende Bedingung  $z < w$  folgt bereits aus der mittleren Indikatorfunktion. Deshalb können wir die dritte ersetzen durch  $I_{(-\infty, 1)}(w)$ , denn es gilt:

$$\begin{aligned} I_{(2w+z-2, w)}(z) &= I_{(-\infty, w)}(z) \cdot I_{(2w+z-2, \infty)}(z) \\ &= I_{(-\infty, w)}(z) \cdot I_{(-\infty, z)}(2w+z-2) \\ &= I_{(-\infty, w)}(z) \cdot I_{(-\infty, 2)}(2w) \\ &= I_{(-\infty, w)}(z) \cdot I_{(-\infty, 1)}(w) \end{aligned}$$

Dieses besagt, daß  $w$  kleiner als 1 bleiben muß, und wir können die Bedingungen äquivalent ausdrücken durch:

$$I(w, z) = I_{(-w, 2-w)}(z) \cdot I_{(-2+w, w)}(z) \cdot I_{(0, 1)}(w)$$

Glücklicherweise decken sich die formal abgeleiteten und die graphisch gewonnenen Ergebnisse.

Mit etwas Mühe lassen sich also die Bedingungen für  $X$  und  $Y$  auf  $W$  und  $Z$  übertragen und auch auf formalem Weg vereinfachen.

**Aufgabe 6:** Die Umformungen sollte der Leser als Übung noch einmal überprüfen und jeweils die verwendete Regel angeben. ⊥

Die Dichte von  $W$  ergibt sich durch Integration über  $z$ .

$$\begin{aligned} f_W(w) &= \int_{\{z|1=I(w, z)\}} \frac{n(n-1)}{2} (1-w)^{n-2} dz \\ &= \frac{n(n-1)}{2} (1-w)^{n-2} \int_{\{z|1=I(w, z)\}} dz \\ &= \frac{n(n-1)}{2} (1-w)^{n-2} z \Big|_{\min\{z|1=I(w, z)\}}^{\max\{z|1=I(w, z)\}} \end{aligned}$$

Bei der letzten Umformung wurde schon mit Blick auf die Graphik zurecht unterstellt, daß der zu integrierende Bereich zusammenhängend ist. Es ist für die Integration also notwendig, die Bedingungen genau zu benennen, die für vorgegebenes  $w$  greifen. Auch an der obigen Zeichnung läßt sich feststellen, daß gelten müssen:  $0 < w < 1$  und  $-w < z < w$ .

Damit folgt:

$$\begin{aligned} f_W(w) &= \frac{n(n-1)}{2} (1-w)^{n-2} z \Big|_{-w}^{+w} \cdot I_{(0,1)}(w) \\ &= \frac{n(n-1)}{2} (1-w)^{n-2} (w - (-w)) \cdot I_{(0,1)}(w) \\ &= n(n-1) (1-w)^{n-2} w \cdot I_{(0,1)}(w) \end{aligned}$$

Die Forderung nach der Verteilung der Wartezeit, die über das Jahresende hinüberreicht, ist damit abgeschlossen: Die gesuchte Dichte ist gefunden. Es handelt sich offenbar um eine Beta-Verteilung mit den Parametern 2 und  $(n-1)$ , denn diese ist gegeben durch

$$f(x) = \frac{1}{B(a,b)} x^{a-1} (1-x)^{b-1} I_{(0,1)}(x)$$

mit der notwendigen Konstante  $\beta(a,b)$ :

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1} (1-x)^{b-1} dx$$

wobei

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dt \quad t > 0$$

Für ganzzahlige Argumente erhalten wir

$$\Gamma(n) = (n-1)!$$

so daß folgt:

$$B(2, n-1) = \frac{\Gamma(2)\Gamma(n-1)}{\Gamma(n+1)} = \frac{1 \cdot (n-2)!}{n!} = \frac{1}{n(n-1)}$$



**Probe** Übrigens können wir jetzt als kleine Probe berechnen, was sich als Erwartung der Summe aller Wartezeiten einstellt. Bezeichnen wir die gerade berechnete Wartezeit mit  $W_n$  und mit  $W_i$  die Wartezeit nach Geburtstag  $i$  in der sortierten Folge der Geburtstage, dann können wir zunächst allgemein berechnen:

$$\begin{aligned}
 E(\text{Beta}(a, b)) &= \int_{-\infty}^{\infty} x \cdot f(x) dx \\
 &= \int_{-\infty}^{\infty} x \cdot \frac{1}{B(a, b)} x^{a-1} (1-x)^{b-1} I_{(0,1)}(x) dx \\
 &= \frac{1}{B(a, b)} \int_0^1 x^{a+1-1} (1-x)^{b-1} dx \\
 &= \frac{1}{B(a, b)} \cdot B(a+1, b) \int_0^1 \frac{1}{B(a+1, b)} x^{a+1-1} (1-x)^{b-1} dx \\
 &= \frac{1}{B(a, b)} \cdot B(a+1, b) \\
 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \cdot \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+1+b)} \\
 &= \frac{\Gamma(a+b)\Gamma(a+1)}{\Gamma(a)\Gamma(a+1+b)} \\
 &= \frac{a}{a+b}
 \end{aligned}$$

Hiermit fällt es nicht mehr schwer, die Summe der Wartezeiten zu berechnen, denn bereits oben hatten wir als Ergebnis

$$W_1, \dots, W_{n-1} \sim \text{Beta}(1, n)$$

ermittelt. Somit folgt:

$$E\left(W_n + \sum_1^{n-1} W_i\right) = \frac{2}{2+n-1} + (n-1) \cdot \frac{1}{1+n} = 1$$

Die Summe der Wartezeiten entspricht also auch bezüglich der Erwartungswerte der Jahreslänge von 1.

**Ergebnis** Hier noch einmal die Zusammenfassung des Ergebnisses:

$$W \sim \text{Beta}(2, n-1)$$

### 3.10 Die Verteilung von $U_{(i)} - U_{(i-1)}$ in der linearen Welt

Die folgende Ableitung dieser Verteilung läuft in mehreren Schritten ab:

S1: Berechnung der  $k$ -dimensionalen Dichte von  $U_{()} = (U_{(1)}, \dots, U_{(k)})$

S2: Bezug dieser Dichte zur Dichte von  $(X_1/S_{k+1}, \dots, X_k/S_{k+1})$  mit  $X_i \stackrel{iid}{\sim} \text{Exp}(1)$  und  $S_{k+1} = \sum_{j=1}^{k+1} X_j$

S3: Darstellung von  $U_{(i)} - U_{(i-1)}$  mit  $S_i - S_{i-1}$

S4: Identifikation der Verteilung von  $U_{(j)} - U_{(j-1)}$  als Beta-Verteilung mit den Parametern 1 und  $k$

**Aufgabe 7:** Leses zur Vorbereitung das Kapitel Multivariate Distributions aus dem Buch von B.W.Lindgren: *Statistical Theory*. ⊥

**S1:** Berechnung der  $k$ -dimensionalen Dichte von  $U_{()} = (U_{(1)}, \dots, U_{(k)})$ .

Für die Dichte der  $k$ -dimensionalen ZV

$$U = (U_1, \dots, U_k)$$

gilt:

$$f_U(u_1, \dots, u_k) = 1 \quad u_i \in [0, 1] \quad \text{mit} \quad i = 1, \dots, k$$

Nach Sortierung ergibt sich

$$U_{()} = (U_{(1)}, \dots, U_{(k)})$$

mit

$$f_{U_{()}}(u_{(1)}, \dots, u_{(k)}) = k! \quad 0 < u_{(1)} < \dots < u_{(i-1)} < u_{(i)} < \dots < u_{(k)}, i = 2, \dots, k$$

Denn jede Realisation von  $u_{(1)}, \dots, u_{(k)}$  kann durch eines von  $k!$  Urbildern hervorgerufen worden sein.

Betrachten wir zur Vorbereitung noch folgende Umformung:

$$k! = \Gamma(k+1) = \int_0^\infty w_{k+1}^k e^{-w_{k+1}} dw_{k+1}$$

**S2:** Bezug dieser Dichte zur Dichte von  $(X_1/S_{k+1}, \dots, X_k/S_{k+1})$  mit  $X_i \stackrel{iid}{\sim} \text{Exp}(1)$  und  $S_{k+1} = \sum_{j=1}^{k+1} X_j$

Gegeben seien die  $(k+1)$ -dimensionale ZV  $X = (X_1, \dots, X_{k+1})$  mit  $X_i \stackrel{iid}{\sim} \text{Exp}(1)$ ,  $i = 1, \dots, k+1$ . Betrachten wir nun die Transformation

$$T = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & & & \ddots & & \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix} X = A \cdot X$$

Nach  $X$  aufgelöst erhalten wir:

$$X = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & & & \ddots & & \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ -1 & -1 & -1 & \cdots & -1 & 1 \end{pmatrix} T = A^{-1} \cdot T$$

Die ersten  $k$  Variablen  $T_i$  sind also mit den entsprechenden  $X_i$  identisch,  $T_{k+1}$  ist als Summe der  $(k+1)$  exponentialverteilten ZV festgelegt und oben als  $S_{k+1}$  bezeichnet worden. Dann ergibt sich die transformierte Dichte nach dem Transformationssatz für Dichten:

$$f_T(t_1, \dots, t_{k+1}) = |\mathbf{J}_T| \cdot f_X(g_1^{-1}(t_1, \dots, t_{k+1}), \dots, g_{k+1}^{-1}(t_1, \dots, t_{k+1}))$$

mit

$$|\mathbf{J}_T| = \begin{vmatrix} \partial x_1 / \partial t_1 & \cdots & \partial x_1 / \partial t_{k+1} \\ \cdots & \ddots & \cdots \\ \partial x_{k+1} / \partial t_1 & \cdots & \partial x_{k+1} / \partial t_{k+1} \end{vmatrix}$$

hierbei liefert  $g_j^{-1}(t_1, \dots, t_{k+1})$  zu  $(t_1, \dots, t_{k+1})$  von seinem Urbild  $(x_1, \dots, x_{k+1})$  dessen  $j$ -te Dimension. Im vorliegenden Fall ist  $\mathbf{J}_T = 1$ , so daß sich die gesuchte Dichte schnell hinschreiben läßt. Da für  $X$  gilt:

$$f_X(x_1, \dots, x_{k+1}) = e^{-\sum_j^{k+1} x_j}$$

folgt für  $T$ :

$$f_T(t_1, \dots, t_{k+1}) = 1 \cdot e^{-\left[ \sum_j^k t_j + \left( \sum_j^k -t_j + t_{k+1} \right) \right]} = e^{-t_{k+1}} \quad \wedge \quad \sum_j^k t_j < t_{k+1}$$

Das ist ein erstaunlich kurzer Ausdruck. Jedoch darf bei der Interpretation die angegebene Bedingung nicht vergessen werden. Von dieser neuen Variablen gehen wir über zu einer normierten  $(k+1)$ -dimensionalen ZV  $W = (W_1, \dots, W_{k+1})$  mittels

$$W_j = T_j / T_{k+1}, \quad j = 1, \dots, k \quad \text{und} \quad W_{k+1} = T_{k+1}$$

Die Umkehrung ergibt:

$$T_j = W_j/W_{k+1}, \quad j = 1, \dots, k \quad \text{und} \quad T_{k+1} = W_{k+1}$$

Die Determinante der Ableitungen besitzt damit folgende Gestalt:

$$|\mathbf{J}_{\mathbf{T}}| = \begin{vmatrix} w_{k+1} & 0 & 0 & \cdots & 0 & w_1 \\ 0 & w_{k+1} & 0 & \cdots & 0 & w_2 \\ 0 & 0 & w_{k+1} & \cdots & 0 & w_3 \\ & \cdots & & \ddots & & \cdots \\ 0 & 0 & 0 & \cdots & w_{k+1} & w_k \\ 0 & 0 & 0 & \cdots & 0 & 1 \end{vmatrix} = w_{k+1}^k$$

und als neue Dichte erhalten wir:

$$f_W(w_1, \dots, w_{k+1}) = w_{k+1}^k \cdot e^{-w_{k+1}}$$

Die gemeinsame Dichte der  $W(w_1, \dots, w_k)$  erhalten wir durch Integration:

$$f_W(w_1, \dots, w_k) = \int_0^\infty f_W(w_1, \dots, w_{k+1}) dw_{k+1} = \int_0^\infty w_{k+1}^k \cdot e^{-w_{k+1}} dw_{k+1}$$

Wir erkennen in dem Integral die  $\Gamma$ -Funktion und erhalten:

$$\int_0^\infty w_{k+1}^k \cdot e^{-w_{k+1}} dw_{k+1} = \Gamma(k + 1) = k!$$

Als Bedingung ist hierbei zu beachten, daß die Summe der  $\sum_{j=1}^k W_j$  den Wert 1 nicht überschreiten darf.

Wie dürfen wir das Ergebnis dieses Schrittes interpretieren? Interessanterweise besitzt die Dichte an Stellen positiver Dichte überall den Wert  $k!$ , das kommt uns bekannt vor. Wir müssen jetzt nur noch die Beziehung herausarbeiten. Die  $W_j$  sind aus exponentialverteilten ZV  $X_j$  durch Division mit  $S_{k+1} = \sum_{j=1}^{k+1} X_j$  entstanden. Plazieren wir die einzelnen  $W_j$ 's hintereinander, so ergeben sich die Summen

$$Y_i = \sum_{j=1}^i W_j = \sum_{j=1}^i \frac{X_j}{S_{k+1}} = \frac{S_i}{S_{k+1}}$$

Da beim Übergang von den  $X_i$  zu den  $Y_i$  nur die Bereiche positiver Dichte maßstabsgetreu verschoben werden, bleiben die Funktionswerte selbst erhalten. Für die ZV  $Y_i$  gelten die Bedingungen:

$$0 < Y_1 < Y_2 < \dots < Y_k < Y_{k+1} = 1$$

Dieses sind dieselben Bedingungen, die für die Ordnungsstatistik gelten. Da auch die Dichten gleich sind, hat der Vektor  $U_{()} = U_{(1)}, \dots, U_{(k)}$  und der Vektor  $(Y_1, \dots, Y_k)$  die gleiche Verteilung, sie sind also identisch verteilt.

Wir können also nun die Ordnungsstatistik  $U_{()} = U_{(1)}, \dots, U_{(k)}$  als aufgereichte normierte unabhängig exponentialverteilte ZV interpretieren und ggf. Eigenschaften übertragen.

**S3:** Darstellung von  $U_{(i)} - U_{(i-1)}$  mit  $S_i - S_{i-1}$

Wenn wir  $U_{(i)}$  mittels  $S_i/S_{k+1}$  ausdrücken können, können wir auch schreiben:

$$U_{(i)} - U_{(i-1)} = \frac{S_i}{S_{k+1}} - \frac{S_{i-1}}{S_{k+1}} = \frac{S_i - S_{i-1}}{S_{k+1}} = \frac{S_i - S_{i-1}}{(S_i - S_{i-1}) + [S_{k+1} - (S_i - S_{i-1})]}$$

Damit können wir die Frage nach der Verteilung der Differenzen von  $U_{(i)} - U_{(i-1)}$  mit Hilfe von Erkenntnissen über  $S_i$  beantworten.

**S4:** Identifikation der Verteilung von  $U_{(j)} - U_{(j-1)}$  als Beta-Verteilung mit den Parametern 1 und  $k$

Erster Teilschritt:

Es gilt:

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Exp}(1) \quad \Rightarrow \quad \sum_{j=1}^n X_j \sim \text{Gamma}(n, 1)$$

und folgt:

$$(S_i - S_{i-1}) = X_i \sim \text{Gamma}(1, 1)$$

sowie

$$S_{k+1} - (S_i - S_{i-1}) = S_{k+1} - X_i \sim \text{Gamma}(k, 1)$$

Zweiter Teilschritt:

Für den zweiten Teilschritt betrachten wir folgendes Theorem:

**Satz:**

*Falls  $X_1$  und  $X_2$  unabhängig gammadverteilt sind mit  $\Gamma(p, \lambda)$  bzw.  $\Gamma(q, \lambda)$ , dann gilt:*

$$Y = \frac{X_1}{X_1 + X_2} \sim \text{Beta}(p, q)$$

▽

**Beweis:** Für den Beweis wird nur der Fall  $\lambda = 1$  betrachtet. Dann ergibt sich als gemeinsame Dichte von  $X_1$  und  $X_2$ :

$$f_X(x_1, x_2) = \frac{e^{-(x_1+x_2)} x_1^{p-1} x_2^{q-1}}{\Gamma(p)\Gamma(q)}$$

Nun definiere  $(Z, Y) = (X_1 + X_2, X_1/(X_1 + X_2))$ . Als Umkehrabbildung<sup>1</sup> berechnen wir:

$$g^{-1}(z, y) = (zy, z - zy)$$

---

<sup>1</sup> $1 - Y = 1 - X_1/(X_1 + X_2) = X_2/(X_1 + X_2) \rightarrow Z(1 - Y) = X_2$

Folglich erhalten wir als Betrag der Determinante der Ableitungen:

$$|J| = \left| \begin{vmatrix} y & 1-y \\ z & -z \end{vmatrix} \right| = |-zy - z + zy| = z$$

Nach dem Transformationssatz folgt für die Dichte von  $(Z, Y)$ :

$$f_{(Z,Y)}(z, y) = z \cdot \frac{e^{-z}(zy)^{p-1}(z-zy)^{q-1}}{\Gamma(p)\Gamma(q)} = \frac{z^{1+p-1+q-1}y^{p-1}(1-y)^{q-1}e^{-z}}{\Gamma(p)\Gamma(q)}$$

Durch Umformung erhalten wir:

$$f_{(Z,Y)}(z, y) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}y^{p-1}(1-y)^{q-1} \cdot \frac{1}{\Gamma(p+q)}z^{p+q-1}e^{-z}$$

Da sich  $f_{(Z,Y)}$  in die Komponenten

$$\frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}y^{p-1}(1-y)^{q-1}$$

und

$$\frac{1}{\Gamma(p+q)}z^{p+q-1}e^{-z}$$

faktorisieren läßt, sind  $Z$  und  $Y$  unabhängig und  $Z$  besitzt, schon wegen Teilschritt 1, eine Gammaverteilung mit den Parametern  $p+q$  und 1. Uns interessiert jedoch hier die ZV  $Y$ , die sich als beta-verteilt mit den Parametern  $p$  und  $q$  entpuppt.  $\Delta$

Jetzt müssen wir nur noch den Satz für unsere Situation anwenden. Oben wurde abgeleitet:

$$U_{(i)} - U_{(i-1)} = \frac{S_i - S_{i-1}}{(S_i - S_{i-1}) + [S_{k+1} - (S_i - S_{i-1})]}$$

In den Kategorien des Satzes können wir auch schreiben:

$$U_{(i)} - U_{(i-1)} \sim \frac{\text{Gamma}(1, 1)}{\text{Gamma}(1, 1) + \text{Gamma}(k, 1)} \sim \text{Beta}(1, k)$$

Damit ist – bis auf ein paar kleine Ungenauigkeiten, wie die genaue Betrachtung von Definitionsbereichen, die Verteilung von  $U_{(i)} - U_{(i-1)}$  abgeleitet und gezeigt, daß diese mit der Verteilung des Minimums übereinstimmt.

Hinweis: Dieser Satz inclusive Beweis wie auch der rote Faden zur Ableitung der Verteilung von Differenzen der Ordnungsstatistik von  $U$  wurden in Anlehnung an Bickel/Doksum formuliert, vgl. insbesondere Theorem 1.2.3 und Aufgabe 1.2.9.

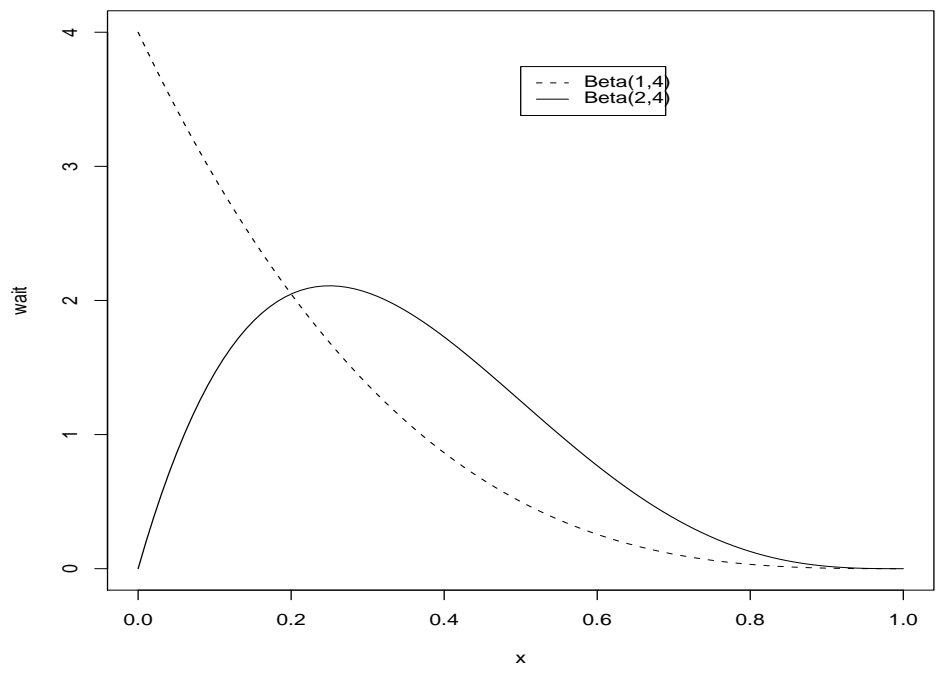
**Reflexion** Für dieses Ergebnis war es offensichtlich wieder notwendig, Gesetze der elementaren Wahrscheinlichkeitsrechnung zu verwenden. Daneben kam Wissen über Zufallsvariablen, Verteilungsfunktionen, Dichten, Gleich-, Gamma- und Beta-Verteilung zum Einsatz. Es ist also nützlich, solche elementaren Dinge parat zu haben. Weiterhin können wir noch einmal darauf hinweisen, daß Übergänge von einer Zufallsvariable zu einer anderen eine wesentliche Rolle spielten. Im vorliegenden Fall haben wir die Verteilung der Ordnungsstatistik  $U_{(1)}$  mit Hilfe der Verteilungsfunktionstechnik ermittelt. Weitere Transformationen wurden mit Hilfe des Transformationssatzes für Dichten abgehandelt.

### 3.11 Fazit

Was haben wir gelernt? Uns interessierte die Verteilung der Wartezeit, die bis zur nächsten Party vergeht. Dazu können wir jetzt zusammenfassen: Da Geburtstage nach einem Jahr wiederkehren, ist eine Modellierung in einer Kreiswelt gerechtfertigt. Aus Symmetriegründen erwarten wir, daß die Verteilungen nicht von der Person abhängen, sprich: alle Wartezeiten identisch verteilt sind. Es ist uns gelungen, das Kreismodell mit  $n$  Geburtstagen auf ein Intervallmodell mit  $n - 1$  Markierungen abzubilden, wobei der verbleibende  $n$ -te Geburtstag auf den Nullpunkt abgebildet wird. In der linearen Welt wird die erste Wartezeit durch  $U_{(1)}$ , dem Minimum, repräsentiert. Die Verteilung des Minimums läßt sich schnell finden: Beta-Verteilung mit den Parametern 1 und  $n - 1$ . Durch Ausspielung der Kreiswelt-Symmetriespielkarte sind auch die anderen Wartezeitverteilungen erkannt. Für Skeptiker wurde jedoch in einer längeren Abhandlung diese Erkenntnis zusätzlich noch einmal bewiesen, so daß jetzt mit Sicherheit ausgesprochen werden kann: Die Wartezeiten in einer kontinuierlichen Kreiswelt sind beta-verteilt mit den Parametern 1 und  $n - 1$ . Nur die arme Person, die zuletzt im Jahr Geburtstag feiert, muß in der Regel mehr Geduld aufbringen, da die Wartezeit Beta( $(2, n - 1)$ -verteilt ist. Zum Abschluß sollte dem PC-losen Leser ein Darstellung von Dichten der Beta-Verteilung nicht vorenthalten werden:

```
32 <* 1)+≡
 n<-5
 x<-seq(0,1,length=100)
 wait.last<-dbeta(x,2,n-1)
 wait<-dbeta(x,1,n-1)
 plot(x,wait,type="l",lty=2)
 lines(x,wait.last,lty=1)
 legend(.5,.9*par()$usr[4],c(paste("Beta(1,",n-1,")",sep=""),
 paste("Beta(2,",n-1,")",sep="")),lty=2:1)
 title(paste("Dichten der Wartezeitverteilungen n=",n,sep=""))
```

Dichten der Wartezeitverteilungen n=5





## 4 Warten auf den nächsten Bus

### 4.1 Bushaltestellenlatein?

Wir alle kennen die Situation, daß wir zu einer Bushaltestelle kommen und fast ewig auf den nächsten Bus warten müssen. Das Gefühl, daß wir mehr als den halben zeitlichen Abstand zwischen zwei fahrplanmäßigen Ankunftszeiten warten müssen, hat sich bei erfahrenden Busfahrern tief im Unterbewußtsein festgesetzt. Ist an diesem Gefühl etwas dran oder ist es nur Ausdruck eines allgemeinen Pessimismus? Und wenn das Gefühl stimmen sollte, wie läßt es sich quantifizieren?

In diesem Abschnitt werden wir uns solchen Wartezeiten zuwenden und können sogar Beziehungen zu den Erkenntnissen des letzten Kapitels aufdecken. Damit gesellt sich zu den rein theoretischen Elementen der Diskussion der Party-Wartezeit eine neue, auch für die Praxis relevante Bedeutungsdimension.

**Aufgabe 8:** *Messen Sie bei Ihren nächsten Fahrten mit Bus und Bahn ihre Wartezeit. Notieren Sie die Verspätungen sowie die Zeitintervalle zwischen aufeinander folgenden Linienverkehrsmitteln.* ⊥

**Aufgabe 9:** *Beschreiben Sie ein Experiment, mit dem Sie Busse simulieren und Mittelwerte für Wartezeiten ermitteln können.* ⊥

Zur Motivation sei diesem Kapitel eine Sammlung von Fragen vorangestellt, die den Inhalt widerspiegeln:

- Wie lange muß ein Busfahrgast auf den nächsten Bus warten, wenn er – der Fahrgast zufällig – kommt?
- Was ändert sich, wenn sich der Bus nicht exakt an den Fahrplan hält?
- Was läßt sich im Besonderen über die zu erwartende Wartezeit sagen?
- Wie käßt sich das Warten eines per Autostopp-Reisenden feststellen?
- Welche Wartezeitverteilung beschreibt das Warten auf das 1000-ste Auto, die 1000-ste Mail, die 1000-ste Sternschnuppe, ...?
- Was haben Poisson-, Exponential-, Gamma-, Beta- Gleich- und Geometrische Verteilung in diesem Problemkomplex zu suchen?
- Was sind erzeugende Funktionen und was läßt sich mit diesen anstellen?
- Welche Experimente helfen die diskutierten Phänomene zu begreifen?

## 4.2 Vergeudete Zeiten an Haltestellen und vor Ampeln

Stellen wir uns eine Bushaltestelle vor, an der die Busse sich genau an den Fahrplan halten und genau alle 10 Minuten eintreffen. Was meinen Sie, wie lange muß man im Durchschnitt auf den nächsten Bus warten, wenn Sie ohne Plan, also zufällig an der Bushaltestelle eintreffen? In der heutigen Zeit ist der Trend zum Selbstfahren ungebrochen. Deshalb wollen wir für Autofahrer die Busfrage in gleichwertiger Form formulieren und diskutieren. Und mit Sicherheit: Wartezeitfragen beschäftigen den autofahrenden Teil der Leser permanent. Besonders wird die Frage interessieren, wie viel Zeit wir vor roten Ampelanlagen vergeuden. Hierzu folgen nun verschiedene kleine Fragen und Antworten zum Themenkreis Warten und Verkehr.

Betrachten wir ein Auto, das zufällig an einer Ampelkreuzung eintrifft.

Frage 1: Wie groß ist die Wahrscheinlichkeit, daß die Ampel grün zeigt? Nach klassischen Vorstellungen erhalten wir:

$$P(\text{Ampel zeigt grün}) = \frac{\text{Länge der Grünphase}}{\text{Länge der Grün- plus Länge der Rotphase}}$$

Frage 2: Welche Verteilung besitzt die Wartezeit auf das nächste grüne Licht, wenn Sie bei ROT eintreffen?

$$\text{Wartezeit} |_{\text{bei ROT eingetroffen}} \sim U(0, \text{Länge der Rotphase}) =: U(0, l)$$

Frage 3: Was ist als durchschnittliche Wartezeit zu erwarten?

$$E(\text{Wartezeit}) = \int_{-\infty}^{\infty} x \cdot f(x) dx = \int_0^l x \cdot 1/l dx = l/2$$

Frage 4: Wie steht es mit der durchschnittlichen Gesamtwartezeit  $S_n$ , falls wir auf unserem Weg an  $n$  Ampeln rotes Licht antreffen?

$$E(S_n) = E\left(\sum_{i=1}^n U_i\right) = n \cdot l/2$$

– sofern für alle  $U_i$  gilt:  $U_i \sim U(0, l) \wedge i = 1, \dots, n$ .

Frage 5: Und wie groß ist die Varianz von  $\text{Var}(S_n)$ ? Bei Unabhängigkeit (\*) gilt:

$$\text{Var}(S_n) = \sum_i \sum_j \text{Cov}(U_i, U_j) \stackrel{(*)}{=} \sum_i^n \text{Var}(U_i) = n \text{Var}(U)$$

mit  $\text{Var}(U_i) = \text{Var}(U)$ . Weiter erinnern wir uns:

$$\text{Var}(U) = \text{E}[g(U)] = \text{E}[(U - \text{E}(U))^2] = \text{E}(U^2) - 2\text{E}[U\text{E}(U)] + \text{E}(U)^2 = \text{E}(U^2) - \text{E}(U)^2$$

Damit folgt für  $U \sim U(0, l)$ :

$$\text{Var}(U) = \text{E}(U^2) - \text{E}(U)^2 = \int_0^l x^2 \cdot 1/l \, dx - \frac{l^2}{4} = \frac{x^3}{3l} \Big|_0^l - \frac{l^2}{4} = \frac{l^3}{3l} - \frac{l^2}{4} = \frac{4l^3 - 3l^3}{12l} = l^2/12$$

und wir erhalten als Varianz von  $S_n$ :

$$\text{Var}(S_n) = \frac{nl^2}{12}$$

sowie die Standardabweichung:

$$\sigma(S_n) = \sqrt{\frac{nl^2}{12}}$$

Frage 6: Welche Momente ergeben sich für die durchschnittliche Wartezeit für  $n$  rote Ampeln?

$$\text{E}(S_n/n) = l/2$$

sowie

$$\text{Var}(S_n/n) = \text{Var}(S_n)/n^2 = \frac{n}{n^2} \frac{l^2}{12} = \frac{l^2}{12n}$$

**Eine kleine Experimentierumgebung** Für Experimente soll ein kleiner Experimentiertisch angeboten werden. Dieser kann im Rahmen eigener Übungen genutzt oder als Ausgangspunkt dienen:

```

33 (* 1)+≡
 # Ampeln definieren
 # 10 Ampeln mit 20 Sekunden GRUEN und 30 Sekunden ROT
 n.ampeln<-10
 rot.gruen.phasen <- cbind(red=rep(30,n.ampeln),
 green=rep(20,n.ampeln))

 # Anzahl Wiederholungen
 n.autos <- 100
 # Zufallsstart
 set.seed(13)
 # Umsetzung
 zyklus.dauer<-apply(rot.gruen.phasen,1,sum)
 result<-rep(0,100)
 for(i in 1:n.autos){
 ankunft<-runif(n.ampeln)*zyklus.dauer
 result[i] <- sum((ankunft<rot.gruen.phasen[, "red"])*ankunft)
 }
 # Auswertung
 plot(result,main="vergeudete Wartezeit")
 summary(result)

```

Mit den angegebenen Parametern erhalten wir zum Beispiel als Ergebnis:

```

Fri Jul 4 13:56:26 2003
 Min. 1st Qu. Median Mean 3rd Qu. Max.
 24.48 64.87 86.87 90.79 119.30 179.20

```

Die *vergeudete* Zeit beträgt im Mittel 90 Sekunden, wobei Minimum und Maximum erheblich differieren.

**Aufgabe 10:** *Listen Sie alle Zusammenhänge auf, die Sie über  $E(g(X))$  kennen? Welche von diesen sind bei den letzten Umformungen zum Einsatz gekommen?* ⊥

**Aufgabe 11:** *Entwerfen Sie weitere Experimente und setzen Sie diese um, um auf experimentellen Wege die oben angesprochenen Fragen anzugehen.* ⊥

**Aufgabe 12:** *Formulieren Sie für Busbenutzer zu den angegebenen Zusammenhängen entsprechende Fragen und Antworten.* ⊥

**Aufgabe 13:** *Listen Sie Argumente auf, warum die gegebenen Antworten der Realität nicht ganz gerecht werden.* ⊥

**Aufgabe 14:** *Und wie ist  $S_n/n$  verteilt – approximativ?* ⊥

### 4.3 Warten beim Autostopp mit Bernoulli und Co.

Sind Sie schon einmal per Autostopp unterwegs gewesen? Das ist eine spannende Sache. Denn Sie stehen an einer Straße und warten mit erhobenen Daumen darauf, daß ein freundlicher Autofahrer anhält und Sie mitnimmt. Wem das zu sehr nach Aussteigerkultur klingt, mag sich vorstellen, daß er sich mitten in New York befindet und für seine Weiterreise ein Taxi anhalten möchte. ...

Natürlich haben Autos bzw. Autofahrer kein Gedächtnis, nach dem Motto: neues Auto neues Glück. Aber: das wievielte Auto wird anhalten? Im Durchschnitt? Und: wie unterschiedlich wird sich diese Zahl realisieren? Diesen Fragen wollen wir hier auf den Grund gehen und dabei ganz nebenbei etwas über den Umgang mit Erwartungswerten erfahren.

Wie können wir die Situation von Anhalten modellieren? Betrachten wir jedes Auto (bzw. Taxi für die salonfähige Einkleidung), das die Straße entlang kommt, als Bernoulli-Experiment. Hält das Auto, verbuchen wir *Erfolg eingetreten*, andernfalls Pech gehabt, *Niete gezogen*. Wie ist die Anzahl der Versuche  $N$  bis zum ersten Erfolg verteilt? Falls  $p$  die Erfolgswahrscheinlichkeit ist, berechnen wir:

$$P(N = 1) = p, \quad P(N = 2) = (1 - p)p, \quad \dots, \quad P(N = k) = (1 - p)^{k-1}p$$

Damit ist die Verteilung gefunden und wir können die erwartete Autoanzahl ermitteln:

$$E(N) = \sum_{j=1}^{\infty} j \cdot (1 - p)^{j-1}p = p \sum_{j=1}^{\infty} j \cdot q^{j-1} \quad \wedge \quad q = 1 - p$$

Einfache Formel, schweres Problem – oder? Nicht für den, der den Trick kennt. Die Struktur erinnert uns an  $f'(x) = k \cdot x^{k-1}$  mit  $f(x) = x^k$ . Weiter wissen wir: Die Ableitung einer Summe ist gleich der Summe der Ableitungen. Umgekehrt können wir die Summe von Ableitungen finden, indem wir die Summe der Stammfunktionen ableiten:

$$E(N) = p \sum_{j=1}^{\infty} j \cdot q^{j-1} = p \frac{d}{dq} \sum_{j=1}^{\infty} q^j = p \frac{d}{dq} \left( \sum_{j=0}^{\infty} q^j - 1 \right) = p \frac{d}{dq} \frac{1}{1 - q} = pp^{-2} = 1/p$$

Dieser Rauf-Runter-Lösungs-Trick ist interessant: Um eine schwierige Operation zu erledigen, steige eine (Abstraktions-) Stufe nach oben, löse das Problem kehre dann auf die Anfangs-Stufe wieder zurück.

### 4.3.1 Per Autostopp zur momenterzeugenden Funktion

Inspiziert durch diesen Trick wollen nun eine weitere Technik erarbeiten. Die zentrale Schwierigkeit bestand darin, die Struktur

$$\sum j q^{j-1}$$

aufzuknacken. Dabei stört, daß  $j$  sowohl als Faktor als auch im Exponent auftaucht. Durch den Trick wurde  $j$  als Faktor entfernt bzw. erzeugt. Nun die Anschlußfrage: Kennen Sie einen anderen Trick, bei dem  $j$  als Faktor generiert wird? Richtig! Falls wir

$$f_1(t) = e^{jt}$$

ableiten, erhalten wir:

$$f_1'(t) = j e^{jt}$$

Experimentell übertragen auf obige Summe können wir schreiben:

$$f_2(t) = \sum e^{jt} q^{j-1} \Rightarrow f_2'(t) = \sum j e^{jt} q^{j-1}$$

Diese Funktion besitzt an der Stelle 0 die Eigenschaften:

$$f_2(0) = \sum e^{j0} q^{j-1} = \sum q^{j-1} \quad \wedge \quad f_2'(0) = \sum j e^{j0} q^{j-1} = \sum j q^{j-1}$$

Der Kern unserer gesuchten Summe ist an der Stelle 0 der Ableitung zu finden.

Zusammengefaßt folgt:

$$E(N) = \sum_{j=1}^{\infty} j \cdot q^{j-1} p = \left. \frac{d}{dt} \sum_{j=1}^{\infty} e^{tj} \cdot q^{j-1} p \right|_{t=0}$$

Ob dieses wirklich funktioniert? Wir benennen die Funktion, die uns zum Ziel bringen soll, mit  $m(t)$  und rechnen:

$$\begin{aligned}
m(t) &= \sum_{j=1}^{\infty} e^{tj} \cdot q^{j-1} p \\
&= p \sum_{j=1}^{\infty} (e^t)^j q^j / q \\
&= p/q \sum_{j=1}^{\infty} (qe^t)^j \\
&= p/q \left[ \sum_{j=0}^{\infty} (qe^t)^j - 1 \right] \\
&= p/q \left[ \frac{1}{1 - (qe^t)} - 1 \right]
\end{aligned}$$

Als Ableitung erhalten wir:

$$m'(t) = p/q(-1) \cdot (1 - (qe^t))^{-2} \cdot (-qe^t) = p(1 - (qe^t))^{-2} e^t$$

sowie die Betrachtung der Stelle  $t = 0$ :

$$m'(0) = p(1 - (qe^0))^{-2} e^0 = p(1 - q)^{-2} = pp^{-2} = 1/p$$

In der Tat, wir haben noch einmal gefunden, was wir schon wußten:

$$E(N) = 1/p$$

Fraglich ist, ob dieser Trick auch an anderen Stellen voranhilft. Dazu sei darauf hingewiesen, daß bei verschiedenen Verteilungen die freie Variable im Exponenten zu finden ist.

**Aufgabe 15:** Suchen Sie in einer Liste über Verteilungen entsprechende Kandidaten heraus, bei denen die freie Variable im Exponenten steht.  $\perp$

Diese Technik hat sich in der Tat in verschiedenen Situationen bewährt. Deshalb erhielt das zentrale Element, die Funktion  $m(t)$ , einen prägnanten Namen bekommen:

**Definition: Momenterzeugende Funktion**

Sei  $X$  eine Zufallsvariable mit Dichte  $f_X(\cdot)$ . Der Erwartungswert von  $e^{tX}$  heißt momenterzeugende Funktion von  $X$ , sofern dieser in einer Umgebung um 0 existiert:

$$m_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$$

Für diskrete Zufallsvariable ist das Integral durch eine Summe zu ersetzen, so daß sich ergibt:

$$m_X(t) = E[e^{tX}] = \sum_{x_j} e^{tx_j} P(X = x_j)$$

▽

Der Autostopp hat uns damit erfolgreich zur momenterzeugende Funktion mitgenommen.

### 4.3.2 Varianz der Wartezeit auf das nächste Auto

Der Name *momenterzeugende Funktion* wäre etwas übertrieben, wenn man nur in besonderen Fällen den Erwartungswert berechnen könnte. Jedoch können wir auch andere Momente, wie zum Beispiel die Varianz berechnen. Denn es gilt neben  $\text{Var}(X) = E(X^2) - E(X)^2$  folgender Satz:

**Satz:**

Falls die momenterzeugende Funktion  $m_X(\cdot)$  von  $X$  existiert und in 0 genügend häufig differenzierbar ist, gilt:

$$\mu_k = E(X^k) = m_X^{(k)}(0) = \frac{d^k m_X(0)}{dt^k}$$

▽

Das Moment  $\mu'_k$  heißt übrigens  $k$ -tes Moment um 0. Ein Moment der Form

$$\mu_k = E[(X - E(X))^k]$$

bezeichnet man dagegen als  $k$ -tes zentrales Moment und

$$E[X(X-1)\dots(X-k+1)]$$

als  $k$ -tes faktorielles Moment. Offensichtlich gilt:  $\text{Var}(X) = \mu_2$ .



**Aufgabe 16:** *Wie läßt sich die Varianz mit den faktoriellen Momenten berechnen?*  $\perp$

Mit dem obigen Satz erhalten wir für unsere Autostopp-Problematik:

$$\begin{aligned} E(N^2) &= m_N^2(0) = \frac{d}{dt} m_N'(0) = \frac{d}{dt} p (1 - qe^t)^{-2} e^t \Big|_{t=0} \\ &= p (1 - qe^t)^{-2} e^t + p(-2) (1 - qe^t)^{-3} (-qe^t) e^t \Big|_{t=0} \\ &= p (1 - q)^{-2} + p(-2) (1 - q)^{-3} (-q) \\ &= pp^{-2} + 2pp^{-3}q \\ &= \frac{1 + q}{p^2} \end{aligned}$$

Hiermit berechnen wir:

$$\text{Var}(N) = \frac{1 + q}{p^2} - \left(\frac{1}{p}\right)^2 = \frac{q}{p^2}$$

**Probe.** Als Probe können wir überlegen, daß die *Anzahl der Fehlversuche* beim Trampen nach unserer Vorstellung geometrisch verteilt ist mit dem Parameter  $p$ . Wie an anderer Stelle nachzulesen, ist der Erwartungswert dieser Fehlversuche  $q/p$  und die Varianz  $q/p^2$ . Unsere Wartezeit  $N$  gemessen in Autos ist um 1 größer als die Anzahl der Fehlversuche, so daß sich der Erwartungswert von  $N$  auch mittels  $1 + q/p$  berechnen läßt, die Varianz beider Zufallsvariablen muß dagegen identisch sein, was der Fall ist.

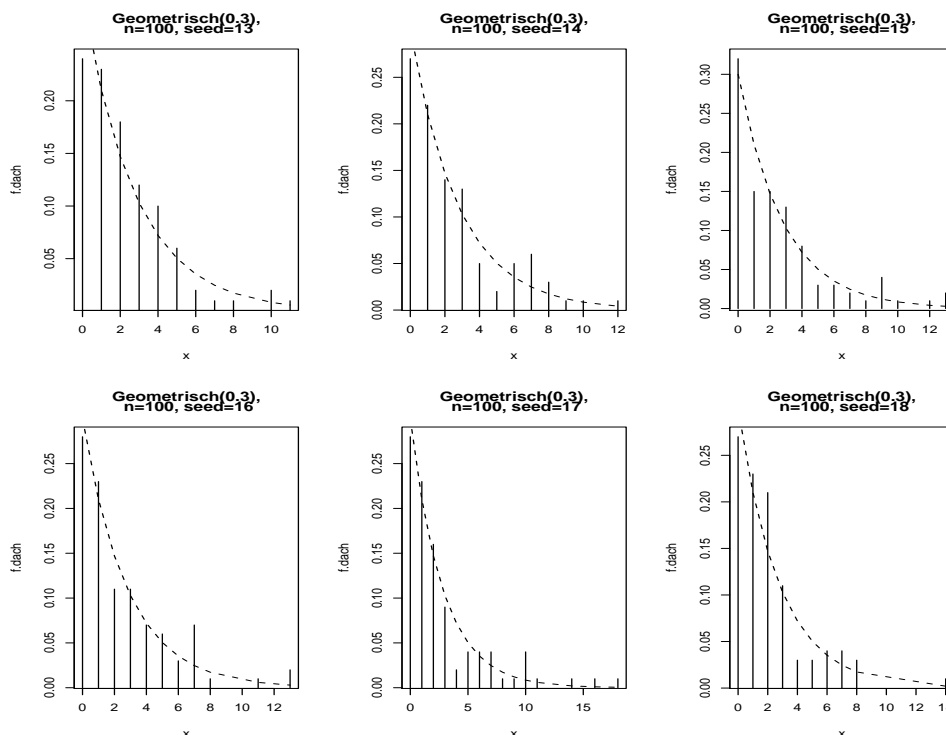
Damit hat sich die momenterzeugende Funktion auch für die Berechnung der Varianz von  $N$  als hilfreich erwiesen.

### 4.3.3 Experimentieren bis zum Erfolg

Die Struktur der geometrischen Verteilung ist nicht sehr spektakulär. Vielleicht ist es interessant zu sehen, wie unterschiedlich Stichproben aus geometrisch verteilten Grundgesamtheiten aussehen. Deshalb folgt hierzu ein einfaches Experiment:

```
34 (* 1)+≡
 # Experimentparameter
 p<-.3; q<-1-p
 n<-100
 seed<-12
 # Umsetzung
 par(mfrow=2:3)
 for(i in 1:6){
 set.seed(seed<-seed+1)
 stpr<-rgeom(n,p)
 # Auswertung
 f.tab <- table(stpr)/n
 x<-as.numeric(names(f.tab))
 f.dach<-f.tab[]
 plot(x,f.dach,type="h",
 main=paste("Geometrisch(",p,")\n",
 " n=",n,", seed=",seed,sep=""))
 lines(x,q^x*p,lty=2)
 print(summary(stpr))
 }
 par(mfrow=c(1,1))
```

|      |         |        |      |         |       |
|------|---------|--------|------|---------|-------|
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max.  |
| 0.00 | 1.00    | 2.00   | 2.23 | 3.00    | 11.00 |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max.  |
| 0.00 | 0.00    | 2.00   | 2.46 | 3.00    | 12.00 |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max.  |
| 0.00 | 0.00    | 2.00   | 2.55 | 3.25    | 13.00 |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max.  |
| 0.00 | 0.00    | 1.00   | 2.48 | 4.00    | 13.00 |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max.  |
| 0.00 | 0.00    | 1.00   | 2.78 | 3.00    | 18.00 |
| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max.  |
| 0.00 | 0.00    | 1.50   | 2.15 | 3.00    | 14.00 |



Selbst bei einem Stichprobenumfang von 100 drängt sich aufgrund eines Stabdiagramms der Verteilungstyp *geometrische Verteilung* nicht gerade auf!

#### 4.4 Warten beim Autostopp mit Stoppuhr

Was interessiert schon die Anzahl der Autos, die vorüberauschen? Wichtiger ist letztlich die Zeit, die wir am Straßenrand stehen. Die richtige Wartezeit gilt es zu modellieren! Doch wie lautet das zutreffende Modell? In diesem Abschnitt wollen wir es durch scharfes Nachdenken erarbeiten.

Die einzelnen Autos wie auch das gesamten System kann kein Gedächtnis besitzen. Falls wir mit  $W$  die Wartezeit bezeichnen, muß diese Wartezeit also folgende Bedingung erfüllen:

$$P(W \leq w + w_0 \mid W > w_0) \stackrel{!}{=} P(W \leq w)$$

Selbst wenn wir bis  $w_0$  erfolglos gewartet haben, wird sich nichts für die Zukunft ändern. Die Aussichten in 0 und in  $t_0$  sind also gleich. Hieraus ergeben sich die äquivalenten Forderungen:

$$\begin{aligned} P(W > w + w_0 \mid W > w_0) &\stackrel{!}{=} P(W > w) \\ P(W > w + w_0 \wedge W > w_0) &\stackrel{!}{=} P(W > w) \cdot P(W > w_0) \\ P(W > w + w_0) &\stackrel{!}{=} P(W > w) \cdot P(W > w_0) \end{aligned}$$

Gesucht ist also eine Funktion mit den Eigenschaften

$$g(w + w_0) = g(w)g(w_0)$$

Weiterhin wissen wir, daß für die monoton fallende Funktion  $g(\cdot)$  gelten muß:  $g(0) = 1$  und  $g(\infty) = 0$ .

Setzen wir  $w_0 = \Delta w$ , können wir die Beziehung oben überführen in einen Differenzenquotienten und erhalten:

$$\begin{aligned} \frac{g(w + \Delta w) - g(w)}{\Delta w} &= \frac{g(w)g(\Delta w) - g(w)}{\Delta w} \\ &= \frac{g(w)(g(\Delta w) - 1)}{\Delta w} \\ &= \frac{g(w)(g(0 + \Delta w) - g(0))}{\Delta w} \end{aligned}$$

Der Grenzübergang  $\Delta w \rightarrow 0$  beschert uns hieraus die Gleichung:

$$g'(w) = g(w) \cdot g'(0)$$

Welche Funktion ist uns bekannt, die mit ihrer Ableitung bis auf eine Konstante identisch ist? Richtig, die Exponentialfunktion! Probieren wir

$$f(x) = e^{ax} \Rightarrow f'(x) = ae^{ax} = a \cdot f(x)$$

Mit  $f'(0) = a$  folgt:

$$f'(x) = f'(0) \cdot f(x)$$

Die Funktion  $g(\cdot)$  ist also gerade die Exponentialfunktion mit einem Parameter  $a$  und die Verteilung ist identifiziert: Es handelt sich um die Exponentialverteilung. Damit  $g(\cdot)$  eine fallende Gestalt besitzt, muß  $a < 0$  sein. In der Regel wird als Parameter für  $-a$  der griechische Buchstabe  $\lambda$  verwendet und wir erhalten:

$$P(W > w) = g(w) = e^{-\lambda w} \Rightarrow F_W(w) = P(W \leq w) = 1 - e^{-\lambda w}, \quad w \geq 0$$

Durch Differenziation finden wir die Dichte

$$f_W(w) = \lambda e^{-\lambda w} I_{(0,\infty)}(w)$$

Mit Hilfe der momenterzeugenden Funktion

$$m_W(t) = \int_0^\infty e^{wt} \lambda e^{-\lambda w} dw = \lambda \int_0^\infty e^{-w(\lambda-t)} dw = \lambda \left. \frac{-e^{-w(\lambda-t)}}{\lambda-t} \right|_0^\infty = \frac{\lambda}{\lambda-t}$$

erhalten wir den Erwartungswert:

$$E(W) = m'_W(0) = \lambda(-1)(\lambda-t)^{-2}(-1) \Big|_{t=0} = \frac{1}{\lambda}$$

Zwischen der Exponentialverteilung und der geometrischen Verteilung muß es aus inhaltlichen Gründen eine Verwandtschaft geben. Diese zeigt sich auf formaler Ebene:  $q^x = e^{x \ln q}$  mit  $\ln q < 0$ . Für den kleinsten zulässigen Wert erhalten wir in der diskreten Welt eine Wahrscheinlichkeit von  $p$ , im kontinuierlichen einen Dichtewert von  $\lambda$ . Beides sind gerade die Kehrwerte der Erwartungswerte. Für beide läßt sich auch eine sehr ähnliche Rekursionsbeziehung ermitteln, um aus  $P(N = n)$  den nächsten Term  $P(N = n + 1)$  bzw. um aus  $f_W(w)$  den Term  $f_W(w + 1)$  zu berechnen.

**Aufgabe 17:** Formulieren Sie die erwähnte Rekursionsbeziehung ⊥

**Aufgabe 18:** Zeichnen Sie die Wahrscheinlichkeitsfunktion von  $N$  und die Dichte von  $W$  und vergleichen Sie. ⊥

**Aufgabe 19:** Konstruieren Sie entsprechend zur geometrischen Verteilung ein Experiment, mit dem Sie die Unterschiedlichkeit von Stichproben aus der Exponentialverteilung demonstrieren können. ⊥

**Aufgabe 20:** Zeigen Sie, daß auch die geometrische Verteilung an Gedächtnislosigkeit leidet. ⊥

## 4.5 Eins, zwei, drei, ..., viele Wartezeiten zusammenaddiert

Jetzt wissen wir, daß wir die Zeit zwischen ankommenden Autos unter bestimmten Bedingungen mit Hilfe der Exponentialverteilung modellieren können. Dieses gilt insbesondere auch für die Autos, die uns als Anhalter nicht mitnehmen. Für die Summe der unnützen Wartezeiten, die auf eine feste Anzahl vorbeirauschender Autos zurückgeht, müßte sich auch eine zusammenfassende Verteilung ermitteln lassen. Welche ist es und warum?

Sei  $X$  die Wartezeit, bis das erste Auto passiert, und  $Y$  die Wartezeit bis zum Eintreffen des nachfolgenden Autos. Der bisherigen Modellierung folgend wollen wir für die Wartezeiten eine Exponentialverteilung unterstellen. Doch wie ist die Summe beider Zufallsvariablen verteilt? Wenn wir diese Frage gelöst haben, kann die Summe weiterer unabhängiger exponentialverteilter Zufallsvariablen kein großes Problem mehr sein.

Die Verteilung der Summe der Wartezeiten können wir für die Zufallsvariablen  $X$  und  $Y$  mit dem bisherigem Wissen nach folgender Strategie berechnen:

1. transformiere:  $(X, Y) \rightarrow (S, T) = (X + Y, Y)$
2. integriere  $T$  aus der zweidimensionalen Dichte  $f_{(S,T)}(\cdot)$  heraus
3. identifiziere ggf. Verteilung aufgrund von  $f_S$

Mit der momenterzeugenden Funktion eröffnet sich ein alternativer Weg, denn es gilt der:

**Satz:**

*Sind  $X$  und  $Y$  zwei unabhängige Zufallsvariablen mit den momenterzeugenden Funktionen  $m_X(t)$  und  $m_Y(t)$ , dann ist die momenterzeugende Funktion von  $S = X + Y$  gegeben durch*

$$m_{X+Y}(t) = m_X(t) \cdot m_Y(t)$$

▽

**Beweis:** – für den kontinuierlichen Fall –

$$\begin{aligned} m_{X+Y}(t) &= \mathbb{E} \left( e^{t(X+Y)} \right) \\ &= \mathbb{E} \left( e^{tX} \cdot e^{tY} \right) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{tx} \cdot e^{ty} f_{X+Y}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{tx} \cdot e^{ty} f_X(x) f_Y(x) dx dy \\ &= \int_{-\infty}^{\infty} e^{tx} f_X(x) dx \cdot \int_{-\infty}^{\infty} e^{ty} f_Y(x) dy \\ &= \mathbb{E} \left( e^{tX} \right) \mathbb{E} \left( e^{tY} \right) \\ &= m_X(t) \cdot m_Y(t) \end{aligned}$$

△

Wir können also mit wenig Aufwand die momenterzeugende Funktion einer Summe unabhängiger Zufallsvariablen berechnen. Als Spezialfall kann diese Erkenntnis auf einfache Zufallsstichproben angewendet werden:

$$X_1, \dots, X_n \stackrel{iid}{\sim} F_X \quad \Rightarrow \quad m_{\sum X_i}(t) = (m_X(t))^n$$

Falls also – wie oben angenommen – gilt:

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Exp}(\lambda)$$

folgt:

$$m_{\sum X_i}(t) = (m_X(t))^n = \left( \frac{\lambda}{\lambda - t} \right)^n$$

Und was machen wir mit dieser Aussage? Es drängt sich der Verdacht auf, daß mit der momenterzeugenden Funktion uns eine Charaktereigenschaft vorliegt, die zur Identifizierung der gesuchten Verteilung beiträgt. Das noch fehlende Glied der Argumentationskette schließt folgender Satz:

**Satz:**

*Es gibt genau eine Verteilung, deren momenterzeugende Funktion mit einer vorgegebenen momenterzeugenden Funktion übereinstimmt.*  $\nabla$

Wir können mit diesem Satz die Übersichtliste, die die Verteilungen mit ihren momenterzeugenden Funktionen zeigt, durchgehen und so die gesuchte Verteilung ermitteln. Zum Beispiel in Mood, Graybill, Boes werden wir im Anhang fündig und können als Ergebnis festhalten:

**Satz:**

*Die Summe  $n$  unabhängig exponentialverteilter Zufallsvariablen mit dem Parameter  $\lambda$  ist gammaverteilt mit den Parametern  $n$  und  $\lambda$ :*

$$X_1, \dots, X_n \stackrel{iid}{\sim} \text{Exp}(\lambda) \quad \Rightarrow \quad \sum_1^n X_j \sim \text{Gamma}(n, \lambda)$$

$\nabla$

**Beweis:** Die momenterzeugende Funktion der Gamma-Verteilung ist gegeben durch

$$\begin{aligned}
 m(t) &= E(e^{tX}) \\
 &= \int_0^\infty e^{tx} \frac{\lambda^n x^{n-1} e^{-\lambda x}}{\Gamma(n)} dx \\
 &= \lambda^n \int_0^\infty \left(\frac{\lambda-t}{\lambda-t}\right)^n \frac{x^{n-1} e^{-(\lambda-t)x}}{\Gamma(n)} dx \\
 &= \frac{\lambda^n}{(\lambda-t)^n} \int_0^\infty (\lambda-t)^n \frac{x^{n-1} e^{-(\lambda-t)x}}{\Gamma(n)} dx \\
 &= \left(\frac{\lambda}{\lambda-t}\right)^n
 \end{aligned}$$

Genau diese Funktion wurde oben als die der Summe von  $n$  unabhängig exponentialverteilter Zufallsvariablen mit Parameter  $\lambda$  ermittelt.  $\Delta$

Für den Eindruck mag eine Darstellung der momenterzeugenden Funktion der Gammaverteilung helfen. Hier ein Entwurf:

```

35 <* 1>+≡
 # Experimentparameter
 lambda<-.3; n<-3
 # Umsetzung
 t<-seq(-.1,.1,length=100)
 f<-lambda/(lambda-t)^n
 # Ergebnis
 plot(t,f, main=paste("momenterzeugende Funktion\n von Gamma(",n,",",lambda,")"))

```

Das Ergebnis dieses Abschnitts lautet: Wenn 20 Autos in Abständen an uns vorbeirauschen, deren Zwischenankunftszeiten unabhängig exponentialverteilt sind mit dem Parameter  $\lambda$ , dann ist die Zeit bis zum 20-ten Auto gammaverteilt mit den Parametern  $n = 20$  und  $\lambda$ .

**Aufgabe 21:** Sei  $X$  gammaverteilt mit den Parametern  $m$  und  $\lambda$  und  $Y$  gammaverteilt mit  $m$  und  $\lambda$ . Ermitteln Sie die Verteilung von  $X+Y$  unter Verwendung des soeben bewiesenen Satzes und auch mit Hilfe des Transformationsatzes für Dichten.  $\perp$

**Aufgabe 22:** Zeichnen Sie verschiedene Gammaverteilungen. Wie hängt die Gestalt von den Parametern ab?  $\perp$



## 4.6 Hand aufs Herz: Technik und Kompromisse

Wer hat bei einer spannenden Geschichte die Zeit, sich bis ins letzte Details um genau selbige zu kümmern. Wahrscheinlich niemand. Es ist also immer ein zweckmäßiger Kompromiß zwischen Zielverfolgung und letzter Genauigkeit erforderlich. So ging es auch hier bei Einsatz und Diskussion der momenterzeugenden Funktion. Deshalb müssen bezüglich Exaktheit einige Randbemerkungen ergänzt werden.

1. Es gibt momenterzeugende Funktionen, die in 0 nicht existieren oder nicht differenzierbar sind. Betrachten wir zum Beispiel unsere geliebte Gleichverteilung. Ist  $U \sim U(0, 1)$ , dann folgt für die momenterzeugende Funktion:

$$m_U(t) = E(e^{tU}) = \int_0^1 e^{tu} du = \frac{e^{tu}}{t} \Big|_0^1 = (e^t - 1)/t$$

Offensichtlich ist diese Funktion an der Stelle 0 nicht definiert und nicht ableitbar. Probleme erwarten wir auch bei Verteilungen, die keinen Erwartungswert besitzen.

2. Für die tiefergehende Diskussionen ist die sogenannte *charakteristische Funktion* vorzuziehen, die durch

$$\phi(t) = E\left(e^{itX}\right) \quad \wedge \quad i = \sqrt{-1}$$

definiert ist. Die charakteristische Funktion existiert immer und besitzt denselben Informationsgehalt wie die Verteilungsfunktion, so daß im Prinzip aus  $\phi(t)$  die Verteilungsfunktion rekonstruierbar ist.

3. Mittels der charakteristischen Funktion lassen sich ebenfalls die Momente ermitteln, so sie existieren. Es gilt:

$$E(X^k) = i^{-k} \phi^{(k)}(0)$$

4. Weiter erweist sich manchmal die *factorial moment generating function*

$$\eta_X(t) = E(t^X) = E\left(e^{X \ln(t)}\right) = m_X(\ln(t))$$

als nützlich. Für  $\eta_X(t)$  gilt:

$$\eta_X^{(k)}(1) = E[X(X-1)\cdots(X-k+1)]$$

5. Da wir  $\eta_X(t)$  für Verteilungen, bei denen die Zufallsvariable die Werte  $0, 1, \dots$  annehmen kann, auch schreiben können als:

$$\eta_X(t) = E(t^X) = \sum_k t^k P(X = k),$$

lassen sich aus den Ableitungen an der Stelle 0 die einzelnen Wahrscheinlichkeiten ablesen. Deshalb ist auch der Name *wahrscheinlichkeitserzeugende Funktion* für  $\eta_X(t)$  plausibel.

6. Zum Abschluß sei noch die momenterzeugende Funktion einer linearen Transformation angegeben:

**Satz:**

*Hat  $X$  die momenterzeugende Funktion  $m_X(t)$ , dann besitzt  $Y = a + bX$  die momenterzeugende Funktion*

$$m_Y(t) = E(e^{t(a+bX)}) = e^{ta} E(e^{bt(X)}) = e^{ta} m_X(bt)$$

▽

Wir sehen, daß es rund um die erzeugenden Funktionen viel zu entdecken gibt, was wir im Rahmen dieses Papiers aber nur zart andeuten können.

**Aufgabe 23:** *Berechnen Sie die charakteristische Funktion und die wahrscheinlichkeitserzeugende Funktion der Poisson-Verteilung* ⊥

## 4.7 Gestaltversteinerung nach dem x-ten Auto

Wir haben gesehen, daß die Summe von *iid* exponentialverteilten Zufallsvariablen gammaverteilt ist. Dabei schlägt sich die Anzahl der Summanden in dem einen Parameter der Gammaverteilung nieder. Wenn wir die Anzahl der Autos sehr groß wählen, dann müßte sich ein ausgleichender Effekt einstellen und Abweichungen um das Summenmittel ungefähr symmetrisch verteilt sein – oder? Wir wollen in diesem Abschnitt der Frage nachgehen, was mit der Verteilung der Summenzeiten passiert, wenn die Anzahl der Autos wächst.

Wenn wir die Gestalten von Gammaverteilungen mit wachsendem Gestalt-Parameter verfolgen, scheinen sich die Bilder immer mehr dem einer Normalverteilung anzunähern. Stimmt das? Wir können dazu jetzt untersuchen, wie eigentlich die momenterzeugende Funktion der Normalverteilung aussieht. Sei  $X \sim N(\mu, \sigma^2)$ . Dann folgt:

$$E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Den Exponenten können wir zusammenfassen zu:

$$\begin{aligned} tx - \frac{(x-\mu)^2}{2\sigma^2} &= -\frac{-tx2\sigma^2 + (x-\mu)^2}{2\sigma^2} \\ &= -\frac{-tx2\sigma^2 + (x^2 - 2x\mu + \mu^2)}{2\sigma^2} \\ &= -\frac{x^2 - 2x(t\sigma^2 + \mu) + \mu^2}{2\sigma^2} \\ &= -\frac{x^2 - 2x(t\sigma^2 + \mu) + (t\sigma^2 + \mu)^2 - (t\sigma^2 + \mu)^2 + \mu^2}{2\sigma^2} \\ &= -\frac{(x - (t\sigma^2 + \mu))^2 - (t\sigma^2 + \mu)^2 + \mu^2}{2\sigma^2} \\ &= -\frac{(x - (t\sigma^2 + \mu))^2}{2\sigma^2} + \frac{(t\sigma^2 + \mu)^2 - \mu^2}{2\sigma^2} \\ &= -\frac{(x - (t\sigma^2 + \mu))^2}{2\sigma^2} + \frac{(t^2\sigma^4 + 2t\sigma^2\mu + \mu^2) - \mu^2}{2\sigma^2} \\ &= -\frac{(x - (t\sigma^2 + \mu))^2}{2\sigma^2} + \frac{t^2\sigma^4 + 2t\sigma^2\mu + 0\mu^2}{2\sigma^2} \\ &= -\frac{(x - (t\sigma^2 + \mu))^2}{2\sigma^2} + \frac{t^2\sigma^2}{2} + t\mu \end{aligned}$$

und schreiben:

$$\begin{aligned} E(e^{tX}) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-(t\sigma^2+\mu))^2}{2\sigma^2} + \frac{t^2\sigma^2}{2} + t\mu} dx \\ &= e^{\frac{t^2\sigma^2}{2} + t\mu} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-(t\sigma^2+\mu))^2}{2\sigma^2}} dx \\ &= e^{t\mu + \frac{t^2\sigma^2}{2}} \end{aligned}$$

Die Standardnormalverteilung hat demnach die momenterzeugende Funktion  $\exp(t^2/2)$ .

Mit diesen Überlegungen müßte mit  $X \sim \text{Gamma}(n, \lambda)$  die momenterzeugende Funktion von  $Y = (X - E(X))/\sqrt{\text{Var}(X)}$  gegen die momenterzeugende Funktion der Standardnormalverteilung gehen. Da gilt:

$$m_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^n$$

folgt mit  $E(X) = n/\lambda$  und  $\text{Var}(X) = n/\lambda^2$  für  $Y$ :

$$Y = \frac{X - n/\lambda}{\sqrt{n/\lambda^2}} = \frac{(X - n/\lambda)\lambda}{\sqrt{n}} = \frac{X\lambda}{\sqrt{n}} - \frac{n}{\sqrt{n}} = \frac{X\lambda}{\sqrt{n}} - \sqrt{n}$$

Hiermit können wir die momenterzeugende Funktion von  $Y$  berechnen:

$$m_Y(t) = e^{-\sqrt{n}t} m_X\left(\frac{\lambda}{\sqrt{n}}t\right) = e^{-\sqrt{n}t} \left(\frac{\lambda}{\lambda - \frac{\lambda}{\sqrt{n}}t}\right)^n = e^{-\sqrt{n}t} \left(\frac{1}{1 - \frac{t}{\sqrt{n}}}\right)^n$$

Versuchen wir die große Klammer in den Exponenten zu verschieben:

$$\begin{aligned} e^{-\sqrt{n}t} \left(\frac{1}{1 - \frac{t}{\sqrt{n}}}\right)^n &= e^{-\sqrt{n}t} e^{(n \cdot (\ln(1) - \ln(1 - t/\sqrt{n})))} \\ &= e^{-\sqrt{n}t - n \cdot \ln(1 - t/\sqrt{n})} \end{aligned}$$

Störend ist die Logarithmusfunktion im Exponenten. Diese wollen wir mit Hilfe von Taylor approximieren. Zur Erinnerung:

$$f(x) = \sum_{k=0}^{\infty} f^{(k)}(x_0) \frac{(x - x_0)^k}{k!}$$

Für  $f(x) = \ln(x)$  und  $x_0 = 1$  erhalten wir als Annäherung:

$$\begin{aligned} \ln(x) &= \ln(1) \frac{(x-1)^0}{0!} + 1/1 \frac{(x-1)^1}{1!} - 1/1^2 \frac{(x-1)^2}{2!} + \sum_{k=3}^{\infty} \ln^{(k)}(x-1) \frac{(x-1)^k}{k!} \\ &= (x-1) - (x-1)^2/2 + \sum_{k=3}^{\infty} \ln^{(k)}(x-1) \frac{(x-1)^k}{k!} \\ &\approx (x-1) - (x-1)^2/2 \end{aligned}$$

Setzen wir für  $x$  den Term  $(1 - t/\sqrt{n})$  ein, so erhalten wir:

$$\begin{aligned} \ln(1 - t/\sqrt{n}) &\approx ((1 - t/\sqrt{n}) - 1) - ((1 - t/\sqrt{n}) - 1)^2/2 \\ &= -t/\sqrt{n} - (-t/\sqrt{n})^2/2 \\ &= -t/\sqrt{n} - t^2/(2n) \end{aligned}$$

Diese Erkenntnis können wir jetzt einsetzen:

$$\begin{aligned}
 m_Y(t) &= e^{-\sqrt{nt} - n \cdot \ln(1 - t/\sqrt{n})} \\
 &\approx e^{-\sqrt{nt} - n \cdot (-t/\sqrt{n} - t^2/(2n))} \\
 &= e^{-\sqrt{nt} + n \cdot (t/\sqrt{n} + t^2/(2n))} \\
 &= e^{-\sqrt{nt} + \sqrt{nt} + t^2/2} \\
 &= e^{t^2/2}
 \end{aligned}$$

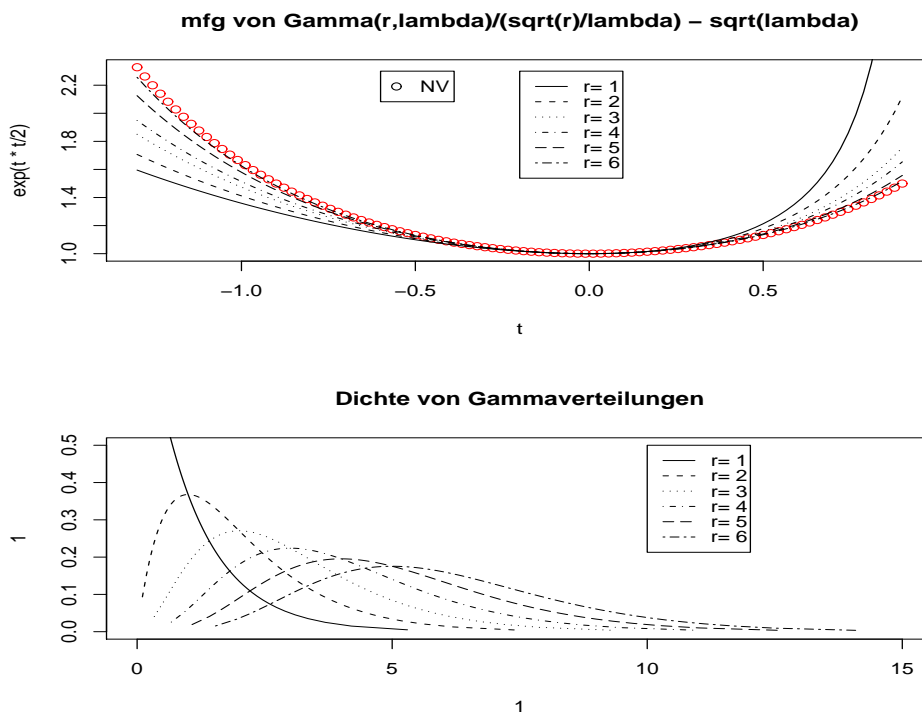
Damit haben wir mit etwas Mühe die momenterzeugende Funktion der Standardnormalverteilung als Grenzverteilung für eine standardisierte Gamma-Verteilung gefunden. Dieses Ergebnis sollte nicht verwundern, da die Fragestellung offensichtlich ein Spezialfall des *zentralen Grenzwertsatzes* darstellt.

**Demonstration** Wir wollen die Entwicklung der momenterzeugenden Funktion von  $(X_r - E(X_r))/\sqrt{X_r}$  mit  $X_r \sim \text{Gamma}(r, \lambda)$  für verschiedene Werte von  $r$  zusammen mit der momenterzeugenden Funktion der Standardnormalverteilung darstellen. Zur Orientierung werden weiterhin Dichten einiger Gammaverteilungen (nichtnormiert) gezeigt.

```

36 (* 1)+≡
Experimentparameter
lambda<-1;r<-1;t<-seq(-1.3,0.9,length=100)
Umsetzung und Darstellung
par(mfrow=2:1)
plot(t,exp(t*t/2),col="red",type="b",
 main="mfg von Gamma(r,lambda)/(sqrt(r)/lambda) - sqrt(lambda)")
r<-c(1,2,5,10,50,500)
legend(-.2,2.3,paste("r=",c(1:length(r))),lty=1:length(r))
legend(-.6,2.3,"NV",pch=1)
for(i in 1:length(r))
lines(t,exp(-r[i]^0.5*t)*(lambda/(lambda-(t*lambda/r[i]^0.5)))^r[i],lty=i)
prob<-(1:100)/100-1/200
plot(1,1,xlim=c(0,15),ylim=c(0,.5),main="Dichte von Gammaverteilungen")
for(i in 1:length(r)){
 qs<-qgamma(prob,i,lambda)
 lines(qs,dgamma(qs,i,lambda),lty=i)
}
legend(10,.5,paste("r=",c(1:length(r))),lty=1:length(r))
par(mfrow=c(1,1));invisible()

```



**Aufgabe 24:** Lesen Sie Darstellungen rund um den zentralen Grenzwertsatz nach!  $\perp$

**Aufgabe 25:** Erstellen Sie für die geometrische Verteilung ein Experiment wie es für die Gammaverteilung beschrieben wurde.  $\perp$

### 4.8 Autoanzahl bei vorgegebener Zeit aus Poissonscher Sicht

Zurück zur Wartesituation am Straßenrand. Zuletzt haben wir uns an festen Anzahlen von Autos, die nicht angehalten, orientiert. Wahrscheinlich sind dagegen Überlegungen, die sich auf *feste Zeitintervallen* – zum Beispiel mit einer Länge von 10 Minuten – beziehen, wesentlich interessanter. Wenn jedes Auto als Bernoulli-Experiment anzusehen ist und mit einer bestimmten Wahrscheinlichkeit anhält, hat deshalb die Frage einen hohen Stellenwert: Wie viele Autos treffen denn nun in 10 Minuten ein?

Die Situation läßt sich so beschreiben: Wir warten  $t$  Zeiteinheiten, zum Beispiel 10, und wollen wissen, wie groß die Wahrscheinlichkeit ist, genau  $n$  – zum Beispiel 20 – Autos zu beobachten, die uns mit Staub einnebeln.

Sei  $N$  die Anzahl der Autos, die in  $[0, t]$  eintreffen, und  $X_j$  die Wartezeit, bis Auto  $j$  eintrifft. Dann ist  $P(N = n)$  offensichtlich gesucht. Wir formulieren um:

$$P(n \text{ Autos in } [0, t]) = P(N = n) = P\left(\sum_{j=1}^n X_j \leq t \wedge \sum_{j=1}^{n+1} X_j > t\right)$$

Wegen unserer Annahmen sind die  $X_j$  unabhängig exponentialverteilt mit  $\lambda$ , so daß gilt

$$Y = \sum_{j=1}^n X_j \sim \text{Gamma}(n, \lambda)$$

Schreiben wir für  $X_{n+1}$  kurz  $X$ , so erhalten wir

$$\begin{aligned} P(N = n) &= P(Y \leq t \wedge Y + X > t) \\ &= 1 - P(Y > t \text{ oder } X + Y \leq t) \\ &= 1 - (1 - P(Y \leq t)) - P(X + Y \leq t) \\ &= P(Y \leq t) - P(X + Y \leq t) \end{aligned}$$

Manche werden jetzt sagen: *Hätte ich doch gleich so hingeschrieben!* – doch auch der Umweg führte voran. Für weitere Umformungen wird die Dichte der Gamma-Verteilung aus dem Hut gezaubert und das Wissen angewendet, daß  $X + Y \sim \text{Gamma}(n + 1, \lambda)$  gilt:

$$\begin{aligned} P(N = n) &= \int_0^t \frac{x^{n-1} \lambda^n e^{-\lambda x}}{\Gamma(n)} dx - \int_0^t \frac{x^n \lambda^{n+1} e^{-\lambda x}}{\Gamma(n+1)} dx \\ &= \frac{\lambda^n}{\Gamma(n+1)} \left( \int_0^t n x^{n-1} e^{-\lambda x} dx - \int_0^t x^n \lambda e^{-\lambda x} dx \right) \end{aligned}$$

Jetzt heißt es: *scharf hinschauen und wo haben wir solch eine Struktur schon gesehen?*  
Probieren:

$$(x^n e^{-\lambda x})' = n x^{n-1} e^{-\lambda x} - \lambda x^n e^{-\lambda x}$$

Aha! Es folgt:

$$\begin{aligned} P(N = n) &= \frac{\lambda^n}{\Gamma(n+1)} x^n e^{-\lambda x} \Big|_0^t \\ &= \frac{(\lambda t)^n e^{-\lambda t}}{n!} \end{aligned}$$

Fertig. Die Anzahl der Autos im Intervall  $[0, t]$  ist Poisson-verteilt mit dem Parameter  $\lambda t$  – wer hätte das gedacht?

**Ein Experiment** Als Experiment können wir Stichproben aus einer exponentialverteilten Grundgesamtheit ziehen, die Realisationen kumulieren und zählen, wie viele unter einer Grenze  $t_0$  bleiben. Solche Anzahlen müssen dann Eigenschaften wie Realisationen Poisson-verteilter Zufallsvariable besitzen. Zur Überprüfung werden wir einen QQ-Plot erstellen.

```
37 (* 1)+≡
 # Experimentparameter
 lambda <- 1
 n <- 100
 t.0 <- 10
 set.seed(13)
 # Umsetzung
 m <- 3*t.0*lambda # grob notwendige Anzahl zu summierender Exp-ZV
 stpr <- matrix(rexp(n*m,lambda),m,n)
 stpr <- split(stpr,col(stpr))
 anz <- unlist(lapply(stpr,function(x)sum(cumsum(x) <= t.0)))
 q.poisson <- qpois((1:50)/50-.01,lambda*t.0)
 # Ergebnisdarstellung
 qqplot(anz,q.poisson)
 print(table(anz))
```

Was meinen Sie zu dem QQ-Plot?

**Rückblick** Ausgehend von der einleuchtenden Annahme der Gedächtnislosigkeit einzelner Vorkommnisse (Autos) sind wir auf exponentialverteilte Zwischenankunftszeiten gestoßen. Die Summe von unabhängigen (exponentialverteilten) Zwischenankunftszeiten gehorcht der Gamma-Verteilung. Wird nun ein Zeitintervall vorgegeben, dann ist die Anzahl der eintreffenden Autos Poisson-verteilt.

Oft werden diese Zusammenhänge in einer anderen Reihenfolge eingeführt. Ausgehend von einigen plausiblen Annahmen über den erzeugenden Prozeß, den Poisson-Prozeß, wird die Poisson-Verteilung abgeleitet und danach die Verteilung der Zwischenankunftszeiten ermittelt. Zum Schluß wird bisweilen ein Übergang zur Gamma-Verteilung vorgenommen.

**Aufgabe 26:** *Schau nach bei Peter Naeve: Stochastik für Informatik und U. Narayan Bhat: Elements of applied stochastic Processes und studierre die dort zu findenden Bemerkungen zum Poisson-Prozeß.* ⊥



## 4.9 Verteilung von Autos in einem vorgegebenen Zeitintervall

Wenn wir ein Zeitintervall vorgeben und die Frage stellen, mit welcher Wahrscheinlichkeit genau  $n$  Autos in diesem Intervall eintreffen werden, dann können wir jetzt diese Wahrscheinlichkeiten mit der Poisson-Verteilung ermitteln. Wir wollen jetzt noch einen Schritt weiter- und der Frage nachgehen, was sich über die Verteilung der Zeitpunkte sagen läßt, wenn bekannt ist, daß in dem Beobachtungszeitraum genau  $n$  Autos beobachtet wurden.

Genauer können wir fragen, welche Verteilung die Ankunftszeit  $X$  des Autos Nummer  $k$  besitzt, wenn am Ende der Beobachtung im Zeitpunkt  $X + Y$  gerade das  $n$ -te Auto vorbeifährt. Wir sind damit auf der Suche nach der Verteilung von

$$X|X + Y.$$

Was wissen wir? Da die Summe  $n$  unabhängiger, exponentialverteilter Zufallsvariablen mit gleichem  $\lambda$  gammaverteilt ist mit dem Gestaltparameter  $n$ , ist die Summe von  $S = X + Y$  gammaverteilt mit den Parametern  $k + (n - k)$  und  $\lambda$ . Auch ist einleuchtend, daß durch die Bedingung der Bereich positiver Dichte auf das Intervall  $[0, s]$  begrenzt ist. Die gesuchte Verteilung liefert folgender Satz:

**Satz:**

Sei  $X \sim \text{Gamma}(\alpha, \lambda)$  und  $Y \sim \text{Gamma}(\beta, \lambda)$ . Dann gilt:

$$X|(X + Y = s) \sim s \cdot \text{Beta}(\alpha, \beta)$$

▽

**Beweis:** Die gesuchte Dichte  $f_{X|X+Y}(\cdot)$  läßt sich zurückführen auf den Quotienten:

$$f_{X|X+Y=s}(x) = \frac{f_{X,X+Y}(x, s)}{f_{X+Y}(x, s)}$$

Wir benötigen also die gemeinsame Verteilung von  $X$  und  $X + Y$  sowie deren Randdichte  $f_{X+Y}$ . Die gemeinsame Verteilung von  $X$  und  $S = X + Y$  ergibt sich durch die Transformation:

$$\left. \begin{array}{l} S \leftarrow X + Y \\ W \leftarrow X \end{array} \right\} \Rightarrow \begin{array}{l} X = W \\ Y = S - W \end{array} \Rightarrow |J| = \left| \begin{array}{cc} 0 & 1 \\ 1 & -1 \end{array} \right| = 1$$

Damit folgt:

$$f_{S,W}(s, w) = \frac{w^{\alpha-1}(s-w)^{\beta-1} \cdot e^{-w-(s-w)}}{\Gamma(\alpha)\Gamma(\beta)} \quad \wedge \quad w \in [0, s]$$

und wir können schreiben:

$$\begin{aligned} f_{X|X+Y=s}(x) &= \frac{f_{S,W}(x, s)}{f_S(x, s)} \\ &= \frac{x^{\alpha-1}(s-x)^{\beta-1} \cdot e^{-s}}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha + \beta)}{s^{\alpha+\beta-1}e^{-s}} \\ &= \frac{1}{B(\alpha, \beta)} \left(\frac{x}{s}\right)^{\alpha-1} \left(1 - \frac{x}{s}\right)^{\beta-1} \cdot \frac{1}{s} \end{aligned}$$

Dies ist eine um den Faktor  $s = x + y$  gestreckte Beta-Verteilung mit den Parametern  $\alpha$  und  $\beta$ . △

Wenn wir also die Zufallsvariable von  $X|X + Y = s$  durch  $s$  teilen, erhalten wir eine Beta-Verteilung mit den Parametern  $\alpha$  und  $\beta$ . Gedanklich können wir für  $X$  nun die Zeiten bis zum ersten, zum zweiten usw. Auto einsetzen. Sei  $X_j$  die Zeit bis zum Auto  $j$ , dann gilt für die *normierte* Wartezeitverteilung:

$$\frac{X_j|X_j + Y = s}{s} \sim \text{Beta}(j, k - j).$$

Damit gilt zum Beispiel für das erste Auto bei einer Gesamtzahl von  $k$  Autos:

$$\frac{X_1|Wartezeit \text{ bis Auto Nummer } k = s}{s} \sim \text{Beta}(1, k - 1).$$

#### 4.9.1 Das kommt uns bekannt vor!

- Zunächst zum formalen Déjà-vu-Erlebnis:

$$\frac{X_j|X_j + Y = s}{s} \sim \text{Beta}(j, k - j) \sim \frac{X_j}{X_j + X_{k-j}}$$

Auf Seite 45 haben wir einen Satz formuliert, der einen Zusammenhang von Gamma- und Beta-Verteilung beschreibt. Dieser ist hier in einem neuen Gewand wieder aufgetaucht.

- Inhaltliches Wiedererkennen. Wir können folgern, daß die durch das  $k$ -te Auto normierten Wartezeiten genauso verteilt sind wie eine Zufallsstichprobe vom Umfang  $k - 1$  aus einer 0-1-Gleichverteilung. Nach den bisherigen Erkenntnissen sind damit die normierten Zwischenankunftszeiten betaverteilt mit den Parametern 1 und  $k - 1$ .

**Zurück zu den Geburtstagen:**  $k - 1$  Ziehungen aus  $U(0, 1)$  konnten wir mit der Frage nach den Abständen zwischen den Geburtstagen von  $k$  Personen identifizieren. Zusammenfassend lassen sich also unsere Erkenntnisse aus dem Problemkomplex *birthday spacings* übertragen auf Wartezeitfragen, die einem Anhalter während seiner Wartezeit einfallen könnten. Durch die Verbindung zum Poisson-Prozeß können diese Überlegungen auf alle möglichen Gegenstandsbereiche ausgedehnt werden, bei denen dieser spezielle Prozeß eine geeignete Modellierung darstellt.

Wenn zum Beispiel Maschinenausfälle oder Nachrichtensendungen durch einen Poisson-Prozeß beschrieben werden können, können wir jetzt folgern, daß die Zeit zwischen den Ereignissen exponentialverteilt ist. Wenn als Bedingung die Anzahl der Ereignisse pro Tag feststeht, läßt sich überlegen, daß dann die Intervalllänge des Intervalls, das einen bestimmten Zeitpunkt beinhaltet, in der Regel größer ausfallen wird als die übrigen Zwischenintervalllängen.

**Aufgabe 27:** Wiederholen Sie das Kapitel zu der Geburtstagsproblematik. ⊥

**Aufgabe 28:** Wie ließe sich diese Parallelität zwischen Autos und Geburtstagen durch ein Experiment unterstützen? ⊥

#### 4.10 Was, wenn der Bus auch zufällig kommt?

Damit selbstfahrende Leser auch Verständnis für die Probleme anderer Verkehrsteilnehmer entwickeln, wollen wir mit unserer Auseinandersetzung zu öffentlichen Verkehrsmitteln zurückkehren. Es ist nämlich noch die Frage beantworten, was sich denn nun über die Wartezeit zum Beispiel auf einen Bus bei variablem Buseintreffen aussagen läßt. Konkret: Wie viele Zeiteinheiten  $W$  müssen wir warten?

Es lassen sich verschiedene Situationen unterscheiden:

1. Wenn gerade Mittagszeit (12.00) ist und der nächste Bus um 12.10 kommen soll, müssen wir genau 10 Minuten warten.
2. Wenn der nächste Bus um 12.10 kommt, wir jedoch zufällig nach dem Modell  $U(12.00, 12.10)$  eintreffen, müssen wir im Schnitt 5 Minuten warten.

3. Wenn der Bus gemäß  $U(12.00, 12.20)$  kommt, also gleichverteilt mit einer festen Intervalllänge von  $A = 20$ , wir aber um 12.00 eintreffen, müssen wir im Schnitt 10 Minuten warten.
4. Die komplizierteste Form ergibt sich, wenn sowohl Fahrgäste als auch Busse zufällig eintreffen. Wenn wir zufällig in dem Intervall ankommen, das sich der Bus durch Realisation seines Ankommens ausgesucht hat, weiter  $B$  der Zeitpunkt des Bus-Eintreffens und durch die Dichte  $f_B(t)$  beschreibbar ist, läßt sich die zu erwartende Wartezeit wie folgt berechnen:

Sei  $W$  die Zufallsvariable, die unsere Wartezeit beschreibt. Dann müssen wir im Falle der Realisation  $B = b$  im Durchschnitt  $b/2$  Zeiteinheiten warten:

$$E(W|B = b) = \int_0^b dx = b/2$$

Hat sich  $B$  noch nicht realisiert, ist der bedingte Erwartungswert  $E(W|B)$  eine Zufallsvariable.  $E(W)$  ergibt sich deshalb durch eine weitere Erwartungswertberechnung:

$$E(W) = E_B(E(W|B)) = \int_0^\infty b/2 f_B(b) db$$

Unterstellen wir zum Beispiel unrealistischerweise, daß gilt  $B \sim \text{Exp}(\lambda)$ , folgt:

$$E(W) = \int_0^\infty b/2 \lambda e^{-\lambda b} db = \frac{1}{2\lambda}$$

Wird das Ankommen des Busses andererseits durch  $U(0, A)$  beschrieben, erhalten wir:

$$E(W) = \int_0^A b/2 \cdot 1/A db = \frac{1}{2A} \cdot \frac{b^2}{2} \Big|_0^A = \frac{A}{4}$$

Für  $A = 20$  erhalten wir eine zu erwartende Wartezeit von 5 Minuten. Das klingt gut, doch ist dieses Modell der Realität nicht besonders überzeugend. Denn es wird unterstellt, daß wir im Prinzip das Ankommen des Busses erraten, daß die Busankunft also das erste Zufallsexperiment ist. Im zweiten abhängigen Experiment trifft dann der Fahrgast so geschickt ein, daß er den Bus auf keinen Fall verpassen kann.

**Aufgabe 29:** Beschreiben Sie Situationen, in denen die Modelle aus der Aufzählung brauchbar sind. ⊥

## 4.11 Straßenbahnen – dann nehme ich eben die nächste

Betrachten wir in diesem Abschnitt den Linienverkehr, der sich dadurch auszeichnet, daß sich in mehr oder weniger regelmäßig Abständen Busse oder auch Straßenbahnen auf festgelegten Strecken durch die Straßen wälzen. Wir kommen zufällig an der Haltestelle an und wissen aus der Erfahrung, daß die letzte Bahn gerade weg ist und wir überdurchschnittlich lange auf die nächste warten müssen. Stimmt das? Dieser Frage werden wir hier nachgehen und halten die geschilderte Ausgangssituation für realistisch und damit für relevant.

Wie können wir diese Situation angemessen modellieren? Beginnen wir mit der Vorstellung, daß ein Arbeits-Tag genau  $600 = 10 \times 60$  Minuten umfaßt. Wenn im Schnitt alle 10 Minuten eine Bahn eintreffen soll, passieren im Laufe des Tages 60 Bahnen unsere Haltestelle.

In Erinnerung an unsere Geburtstagsüberlegungen kann eine Modellierung dadurch geschehen, daß wir in dem Intervall  $[0, 600]$  zufällig, d.h. gemäß  $U(0, 600)$  60 Zufallsrealisationen bestimmen. Die sortierte Stichprobe  $B_1, \dots, B_{60}$  repräsentiert jetzt die Ankunftszeitpunkte der 60 Busse:

$$0 \leq B_1 < B_2 < \dots < B_{60} \leq 600$$

Die Differenzen zwischen eintreffenden Bussen gehorchen dann

$$B_j - B_{j-1} \sim 600 \cdot \text{Beta}(1, 60)$$

Diese Situation läßt sich wie geübt in die Kreiswelt überführen und durch 61 zufällige Punkte auf einem Kreis erfassen.

Unser Eintreffen an der Haltestelle wollen wir in der linearen Welt durch ein weiteres Zufallsexperiment darstellen mit  $K \sim U(0, 600)$ . In der Kreiswelt entspricht dieses, einen weiteren Punkt auf dem Kreis zufällig auszuwählen, wie wir das mit den Jahresanfängen fabriziert haben. Mit hoher Wahrscheinlichkeit werden wir dabei ein großes Intervall treffen. Die genaue Verteilung hatten wir oben durch eine Formulierung in der linearen Welt als Suche nach der Verteilung des Minimums gefunden. Hiernach ist die Intervalllänge von unserem Eintreffen bis zum nächsten Bus beschrieben durch

$$600 \cdot \text{Beta}(1, 61)$$

mit einem Erwartungswert von

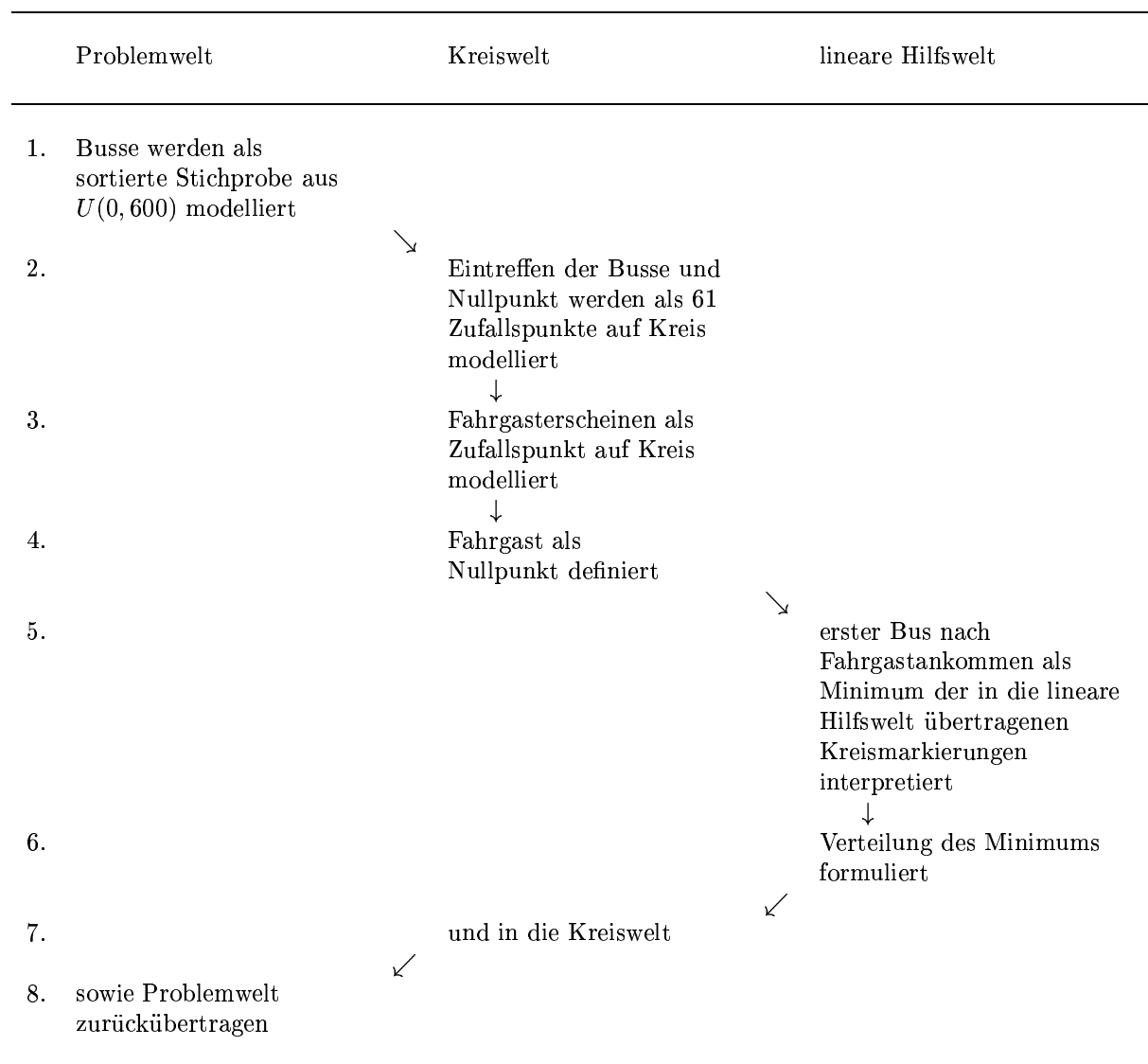
$$600 \cdot \frac{1}{1 + 61} = 600/62$$

Dieser bleibt auch bei der Rückübertragung in unsere lineare Welt erhalten.

Somit kommen wir unter den getroffenen Annahmen zu folgendem Ergebnis: Die zu erwartende Wartezeit beträgt  $600/62 \approx 9.7$  Minuten, obwohl die Bahnen im Schnitt alle 10 Minuten erscheinen, also fast doppelt soviel wie die oben berechneten 5 Minuten.

Wir können hoffen, daß sich die Realität irgendwo in der Mitte zwischen diesen knapp 10 Minuten und den zu erwartenden 5 Minuten bei einem fahrplanteuren Linienverkehr aufhält.

Damit haben sich die Studien der Kreisstatistik für die Diskussion der Bus- / Bahn-Wartefrage bezahlt gemacht. Hier noch einmal semigraphisch die Argumentationsstruktur:



## 4.12 Unpünktliche Linienbusse

Im vorherigen Abschnitt wurde die unrealistische Annahme getätigt, daß die Straßenbahnen über den Arbeitstag an der Haltestelle genauso eintreffen, wie gleichverteilte Zufallspunkte über ihr Intervall positiver Dichte verteilt sind. Wir wollen jetzt noch einen letzten Ansatz verfolgen, bei dem wir von einer gegebenen Verteilung der Zwischenankunftszeiten  $f_L(l)$  ausgehen. Es ist einsichtig, daß – wenn wir zufällig an der Haltestelle erscheinen – große Intervalle mit einer höheren Wahrscheinlichkeit ausgewählt werden als kleinere Intervall. Genauer ist einsichtig, daß die Chance, ein Intervall der Länge  $l$  zu treffen, proportional zu  $l$  ist.

Wie können wir *proportional zu  $l$*  operationalisieren? Die Dichte der Trefferwahrscheinlichkeit  $T$  für eine spezielle Länge  $l$  muß nach dieser Idee proportional zu  $l$  und natürlich auch proportional zur Auftrittswahrscheinlichkeit der Länge  $l$  sein, salopp also:

$$\text{Intervalltrefferwahrscheinlichkeit} \sim \text{Länge} \times \text{Häufigkeit(Länge)}$$

Falls  $f_L(\cdot)$  die Dichte der Intervalllängen und  $f_{\text{Treffer}}(\cdot)$  die Treffer-Dichte ist, folgt:

$$f_{\text{Treffer}}(l) = c \cdot l \cdot f_L(l)$$

Ist  $D_L$  der Bereich, in dem sich Intervalllängen realisieren können, muß  $c$  die Bedingung

$$1 = \int_{D_L} f_{\text{Treffer}}(l) dl = \int_{D_L} c \cdot l \cdot f_L(l) dl$$

erfüllen. Die zu erwartende Wartezeit können wir hiermit berechnen:

$$E(W) = \int_{D_L} \frac{l}{2} \cdot f_{\text{Treffer}}(l) dl = \int_{D_L} \frac{l}{2} \cdot c \cdot l \cdot f_L(l) dl$$

### 4.12.1 Spezialfall: Gleichverteilte Intervalllängen

Für eine  $u$ - $o$ -Gleichverteilung mit  $f_l(l) = I_{[u,o]}(l)/(o-u)$  erhalten wir zum Beispiel:

$$1 = \int_u^o c \cdot l \cdot \frac{1}{o-u} dl$$

aus diesem folgt:

$$c = \frac{o-u}{\frac{l^2}{2} \Big|_u^o} = \frac{2(o-u)}{o^2-u^2} = \frac{2}{o+u},$$

so daß sich  $f_{\text{Treffer}}(\cdot)$  schreiben läßt in der Form:

$$f_{\text{Treffer}}(l) = \frac{2}{o+u} \cdot l \cdot \frac{I_{[u,o]}(l)}{(o-u)} = \frac{2l \cdot I_{[u,o]}(l)}{o^2-u^2}$$

Wenn wir im Mittel die Hälfte des getroffenen Intervalls warten müssen, ergibt sich die Gesamterwartung  $W$  durch

$$E(W) = \int_{D_L} \frac{l}{2} \cdot f_{\text{Treffer}}(l) dl = \int_{D_L} \frac{l}{2} \cdot \frac{2l \cdot I_{[u,o]}(l)}{o^2-u^2} dl = \frac{1}{o^2-u^2} \cdot \int_u^o l^2 dl$$

**Spezialfall:  $U(10-a, 10+a)$ -verteilte Intervalllängen** Wir wollen zur Demonstration  $f_L(l)$  als Dichte einer um 10 symmetrischen Gleichverteilung festlegen:

$$L \sim U(10-a, 10+a)$$

Dann folgt für die zu erwartende Wartezeit:

$$\begin{aligned} E(W) &= \frac{1}{o^2-u^2} \int_u^o l^2 dl = \frac{1}{o^2-u^2} \int_{10-a}^{10+a} l^2 dl \\ &= \frac{1}{(10+a)^2 - (10-a)^2} \cdot \frac{l^3}{3} \Big|_{10-a}^{10+a} \\ &= \frac{(10+a)^3 - (10-a)^3}{3 \cdot ((10+a)^2 - (10-a)^2)} \\ &= \frac{(1000 + 300a + 30a^2 + a^3) - (1000 - 300a + 30a^2 - a^3)}{3 \cdot [(100 + 20a + a^2) - (100 - 20a + a^2)]} \\ &= \frac{600a + 2a^3}{3 \cdot 40a} = \frac{600 + 2a^2}{120} = \frac{600}{120} + \frac{a^2}{60} = 5 + \frac{a^2}{60} \end{aligned}$$

Für  $a = 0$  erhalten erhalten wir glücklicherweise die erwarteten 5 Minuten als Antwort. Wächst  $a$  auf 10, wächst die zu erwartende Wartezeit auf  $5 + 100/60 = 6 \frac{2}{3}$  Minuten an. Zu tun bleibt, auf empirischen Wege die Verteilung von  $f_L(\cdot)$  für reale Situationen zu schätzen.



### 4.12.2 Experimentelle Überprüfung

Was können wir jetzt noch tun? Natürlich ein Experiment zu ersinnen, das das Ergebnis untermauert. Wir ziehen zuerst eine Grundgesamtheit von Intervallen. Dann berechnen wir Gewichtungen für die Realisationen gemäß den Längen. Das Experiment besteht nun darin, wiederholt Intervalle zu ziehen und unter der Annahme, daß man im Schnitt die halbe Intervalllänge warten muß, Wartezeiten auszurechnen. Zur Kontrolle werden außerdem Wartezeiten ohne die Gewichtung ermitteln.

```
38 <* 1)+≡
Experimentierparameter
a<-10 # Schwankung des Buseintreffens
n<-1000 # Anzahl der Fahrgaeste
anz.intervalle <-100
set.seed(11)
Grundgesamtheit der Intervalle
L<-runif(anz.intervalle,10-a,10+a)
Gewichtung nach Laenge
p.l <- L/sum(L)
Umsetzung
result.mg<-result.og<-rep(0,n)
for(i in 1:n){
 l.real<-sample(L,size=1)
 result.og[i]<-l.real/2
 l.real<-sample(L,size=1,prob=p.l)
 result.mg[i]<-l.real/2
}
Experimentierbericht
boxplot(list(result.og,result.mg))
cat("Ergebnis ohne Gewichtung:\n")
print(summary(result.og))
cat("Ergebnis mit Gewichtung:\n")
print(summary(result.mg))
```

Hier ist das numerische

Ergebnis ohne Gewichtung:

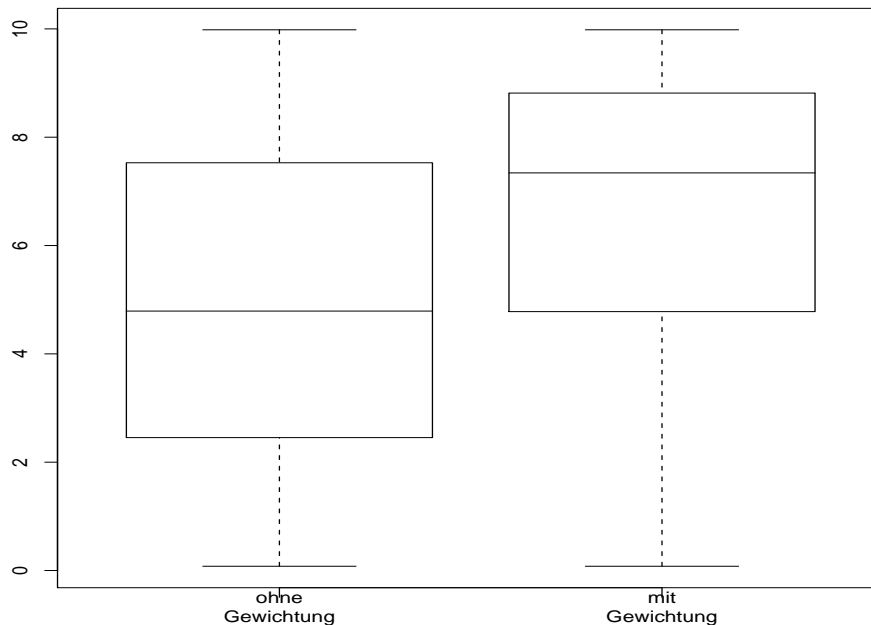
| Min.    | 1st Qu. | Median  | Mean    | 3rd Qu. | Max.    |
|---------|---------|---------|---------|---------|---------|
| 0.07852 | 2.45400 | 4.79000 | 5.03300 | 7.52800 | 9.98300 |

Ergebnis mit Gewichtung:

| Min.    | 1st Qu. | Median  | Mean    | 3rd Qu. | Max.    |
|---------|---------|---------|---------|---------|---------|
| 0.07852 | 4.78000 | 7.34100 | 6.77800 | 8.81500 | 9.98300 |

Thu Jul 3 10:46:46 2003

... und hier die zugehörigen Boxplots.



Wieder Glück. Die Simulationsergebnisse unterstützen die theoretischen Überlegungen.

**Aufgabe 30:** Welche Warteerwartung ergibt sich, falls die Intervalllängen gemäß  $U(a, b)$  verteilt sind? Wie müssten die Parameter des Experiments zur Überprüfung bestückt werden?

⊥

**Aufgabe 31:** Stellen Sie sich einen Linienbus vor, der pünktlich an der Anfangsstation losfährt. Zwischen zwei Haltestellen wird er mehr oder weniger durch den Verkehr und natürlich auch durch Lichtzeichenanlagen aufgehalten. Weiterhin dauert das Ein- und Aussteigen unterschiedlich lange. Formulieren Sie Ideen zu einer Modellierung. Bauen Sie weiterhin ein Experiment auf, daß diese Situation modelliert. Um wie viel pünktlicher wird der Bus an den Haltestellen ankommen, falls er alle Ampeln für sich auf GRÜN stellen darf?  $\perp$

**Aufgabe 32:**  $n$ . pos Personen feiern an verschiedenen Samstagen zu Hause Geburtstag. Sie wissen nicht, wer wann Geburtstag feiert, nehmen sich aber vor, an jedem Samstag einer anderen dieser Personen einen Besuch abzustatten. Wie groß ist die Wahrscheinlichkeit, daß Sie mindestens eine Feier erwischen?  $\perp$

**Aufgabe 33:** Verfassen Sie eine Zusammenfassung dieses Kapitels.  $\perp$

## 5 Axiome

*In diesem Kapitel ist zu sehen, wie es mit diesem Skript anfing. Nach kurzer Zeit wurde deutlich, daß ein solches Skript vielleicht zu einem brauchbaren Nachschlagewerk wachsen kann, doch sowohl das Lesen wie auch das Schreiben kann keine besonders große Spannung erzeugen. Deshalb wäre der Autor wahrscheinlich auf halben Wege verdurstet. Glücklicherweise entwickelte sich jedoch aus dem Wunsch, zusätzliche Motivation zu schaffen, eine immer größere Ausweitung der Einleitung, in der verschiedene Wartefragen aufgegriffen wurden. Das kleine Bruchstück des ursprünglichen ersten Kapitels mit dem Namen Axiome, hat jetzt mehr Erinnerungswert und wartet darauf, passend aufgepeppt zu werden. Näheres zu diesbezüglicher Wartezeit kann noch nicht angegeben werden. In gewisser Weise ist also das Kurzschrift aus dem Kapitel Einleitung nicht herausgekommen, jedoch doch ganz nett.*

Zum Birthday-Spacings-Problem: Was benötigen wir? Erinnerung: Was ist bekannt aus Statistik I?

Der Wahrscheinlichkeitsbegriff.

Roulette-Spieler setzt einmal auf eine Zahl. Wie groß ist die Wahrscheinlichkeit für einen Gewinn?

Antwort:  $1/37$

Trivial? Doch wie berechnet? Unter welchen Annahmen?

Annahme: Es gibt 37 verschiedene Ergebnisse. Diese werden in der Ergebnismenge / im Ergebnisraum / im Stichprobenraum zusammengefaßt:

### **Definition: Ergebnismenge**

*Die Menge der möglichen Ergebnisse eines Zufallsvorgangs heißt Ergebnismenge. Oft wird für diese das Zeichen  $\Omega$  verwendet.  $\nabla$*

Für das Experiment ein Roulette-Spiel können wir damit vorschlagen:

$$\Omega = \{1, \dots, 36, 0\}$$

Wie kommen wir nun zu den Wahrscheinlichkeiten?

- Wir könnten jedem Element dieser Menge eine Wahrscheinlichkeit zuordnen. Doch wie erhält man dann Wahrscheinlichkeiten für Kombinationen? Siehe zum Beispiel *Münzwurf*: jede Augenzahl bekommt  $1/6$  als Wahrscheinlichkeit, wie erhalten wir dann diejenige für eine gerade Augenzahl? Müßten dann nicht solche Kombinationen mitüberlegt werden? Offensichtlich wird damit Elementen wie auch Kombinationen von Elementen, also Mengen, etwas zugeordnet. Damit ist aus mathematischer

Sicht die Definitionsmenge heterogen. Wenn Kombinationen zusätzliche Wahrscheinlichkeiten zugewiesen bekommen, wie kann die Stimmigkeit gewährleistet werden? Klingt verwickelt. Deshalb:

- Wir betrachten alle denkbaren Teilmengen von  $\Omega$  und ordnen diesen Wahrscheinlichkeiten nach festen Regeln zu.
  1. bilde Potenzmenge  $\rightarrow$  für diskrete Ergebnismengen ok, für kontinuierliche noch unklar
  2. ordne einelementigen Teilmengen eine Wahrscheinlichkeit zu und berechne die Wahrscheinlichkeit anderer Teilmengen durch Addition der Wahrscheinlichkeiten derjenigen einelementigen Teilmengen, die durch Vereinigung die jeweils betrachtete Menge ergeben

Wir sehen, der zweite Vorschlag besitzt konstruktive Aspekte, löst nicht alle Probleme und zeigt, daß rein verbale Beschreibungen schnell unübersichtlich sind und Ungenauigkeiten aufweisen können. Wenden wir uns zunächst einer Vereinfachung des Potenzmengen gedanken zu: Mit diesem Ansatz kann keiner mehr eine Wahrscheinlichkeitsfrage stellen, die nicht im Rahmen des Teilmengensystems abgedeckt ist. Denn mit jeweils zwei Mengen sind deren Schnitt, ihre Vereinigung und ihre Komplemente mit in der Menge vorhanden. Wenn man nur an bestimmten Teilmengen interessiert ist, kann man sich ausgehend von den Konstruktionsideen ein Teilmengensystem vorstellen, daß einerseits für die direkt interessierenden zusammengesetzten Ereignisse eine entsprechende Teilmenge enthält, andererseits nach den Prinzipien: Schnitt, Vereinigung und Komplement abgeschlossen ist. Diese Vorstellungen werden gerade durch das Wort *Algebra* umgesetzt:

**Definition: Algebra**

$\mathcal{F}$  ist eine Algebra über  $\Omega$ , falls gilt:

1.  $(A \in \mathcal{F}) \Rightarrow (\bar{A} \in \mathcal{F})$
2.  $(A, B \in \mathcal{F}) \Rightarrow (A \cup B \in \mathcal{F})$
3.  $(A, B \in \mathcal{F}) \Rightarrow (A \cap B \in \mathcal{F})$
4.  $(\Omega \in \mathcal{F}), (\{\} \in \mathcal{F})$

▽

**Aufgabe 34:** Zeige, daß die dritte Bedingung überflüssig ist. ⊥

Sind die ersten beiden Bedingungen erfüllt, wird  $\mathcal{F}$  als *Feld* bezeichnet.

Gilt zusätzlich die Bedingung:

$$(A_i \in \mathcal{F}, i \in \mathbb{N}) \Rightarrow \left( \left( \bigcup_i^\infty A_i \right) \in \mathcal{F} \right),$$

heißt  $\mathcal{F}$  auch  $\sigma$ -Algebra.

**Aufgabe 35:** Zeige, daß für eine  $\sigma$ -Algebra gilt:  $((\bigcap_i^\infty A_i) \in \mathcal{F})$

⊥

Ereignisraum und Algebra lassen sich zusammenfassen:

**Definition: Meßraum**

Das Paar  $(\Omega, \mathcal{F})$  heißt Meßraum.

∇

Jetzt können wir uns der Wahrscheinlichkeitsfestlegung zuwenden.

**Definition: Wahrscheinlichkeitsmaß**

Sei  $(\Omega, \mathcal{F})$  ein Meßraum. Dann heißt eine Abbildung der  $\sigma$ -Algebra  $\mathcal{F}$  in das Zahlenintervall von 0 bis 1

$$P : \mathcal{F} \rightarrow [0, 1]$$

Wahrscheinlichkeitsmaß, sofern die folgenden Bedingungen erfüllt sind:

i) Für jedes  $A \in \mathcal{F}$  gilt:  $0 \leq P(A) \leq 1$ . (Axiom 1)

ii)  $P(\Omega) = 1$ . (Axiom 2)

iii)  $(A_1 \cap A_2) = \{\}$   $\rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2)$ . (Axiom 3)

Die zugeordnete Zahl  $P(A)$  heißt dann Wahrscheinlichkeit von  $A$ .

∇

**Definition: Wahrscheinlichkeitsraum**

Das Tripel  $(\Omega, \mathcal{F}, P)$  heißt Wahrscheinlichkeitsraum.

∇

Bemerkungen:

- Die axiomatische Begründung der Wahrscheinlichkeit geht auf Kolmogoroff zurück.
- Sollen auch Vereinigungen unendlich vieler Teilmengen betrachtet werden, muß das dritte Axiom verallgemeinert werden zu:
  - iii)\*  $(A_i \text{ disjunkt}) \rightarrow P(\bigcup_i^\infty A_i) = \sum_i^\infty P(A_i)$ . (Axiom 3a)
- Man findet äquivalente andere Formulierungen der Reihenfolgen, zum Beispiel zu Axiom 1:  $P(A) \geq 0$  für alle  $A \in \mathcal{F}$ .
- Die Eigenschaft iii)\* heißt  $\sigma$ -Additivität.
- Es lassen sich auch Maße oder Systeme konstruieren, die Zufälligkeiten etwas anders modellieren.:

- Fuzzy-Set-Ansätze: die Funktionswerte der sogenannte Membershipfunction addieren sich normalerweise zu einer Summe größer 1:  $M(\Omega) > 1$ .
- Belief-Funktions der Dempster-Shafer-Theorie:

$$B\left(\bigcup_i^\infty A_i\right) \geq \sum_i B(A_i)$$

## 6 Anhang: Geburtstage von 500 Studierenden

```
39 <lege auf geb.stud.stpr 500 Geburtstage von Studierenden ab 39>≡
 "geb.stud.stpr" <-
 c(238, 217, 64, 43, 212, 94, 175, 101, 318, 340, 60, 200, 322,
 210, 108, 183, 162, 146, 107, 221, 251, 225, 14, 216, 178, 314,
 12, 85, 14, 78, 313, 111, 362, 162, 302, 305, 12, 136, 329, 271,
 215, 334, 4, 263, 81, 273, 246, 67, 34, 197, 213, 100, 185, 281,
 155, 145, 123, 224, 38, 303, 328, 193, 220, 174, 296, 92, 332,
 43, 252, 149, 134, 123, 261, 265, 326, 237, 294, 321, 125, 45,
 122, 326, 136, 165, 232, 52, 125, 107, 94, 22, 284, 10, 220,
 357, 305, 135, 264, 97, 293, 328, 115, 87, 223, 341, 27, 84,
 55, 254, 139, 57, 251, 90, 360, 312, 324, 134, 66, 75, 100, 117,
 37, 288, 141, 149, 32, 211, 327, 272, 244, 35, 148, 287, 162,
 214, 288, 19, 118, 91, 135, 10, 164, 229, 93, 78, 253, 73, 292,
 324, 261, 244, 317, 316, 160, 321, 360, 205, 138, 37, 161, 202,
 290, 362, 178, 165, 173, 65, 208, 335, 215, 18, 73, 133, 228,
 105, 114, 7, 100, 285, 172, 365, 64, 43, 315, 84, 260, 118, 190,
 222, 239, 76, 311, 136, 124, 45, 91, 317, 277, 178, 77, 252,
 291, 55, 184, 260, 124, 194, 189, 169, 290, 314, 175, 194, 286,
 117, 281, 236, 267, 68, 315, 46, 334, 307, 235, 210, 274, 160,
 117, 232, 22, 218, 266, 41, 142, 264, 75, 343, 55, 319, 266,
 65, 59, 26, 364, 3, 84, 290, 189, 365, 97, 16, 341, 307, 19,
 318, 112, 161, 258, 248, 238, 297, 313, 127, 33, 181, 18, 155,
 5, 101, 159, 216, 96, 317, 4, 52, 118, 286, 46, 253, 284, 71,
 184, 263, 37, 36, 228, 41, 294, 151, 215, 20, 301, 320, 268,
 318, 250, 285, 362, 263, 294, 308, 88, 267, 317, 252, 291, 9,
 83, 62, 190, 237, 314, 288, 190, 338, 151, 284, 277, 143, 217,
 290, 239, 347, 94, 342, 146, 3, 340, 133, 47, 90, 64, 30, 223,
 136, 359, 301, 305, 160, 277, 333, 58, 28, 187, 118, 162, 339,
 98, 127, 213, 262, 273, 113, 237, 67, 249, 76, 350, 13, 64, 126,
 87, 269, 184, 198, 45, 356, 310, 260, 178, 75, 70, 168, 16, 153,
 241, 175, 129, 49, 161, 175, 281, 1, 103, 106, 140, 329, 97,
 166, 90, 90, 154, 131, 70, 307, 167, 307, 310, 97, 99, 101, 303,
 59, 134, 187, 175, 222, 162, 47, 196, 350, 214, 10, 136, 227,
 143, 196, 20, 24, 327, 5, 38, 321, 95, 344, 242, 86, 221, 342,
 308, 274, 298, 336, 130, 78, 216, 96, 113, 65, 165, 156, 282,
 199, 69, 283, 232, 26, 163, 289, 168, 310, 316, 47, 188, 148,
 334, 112, 265, 106, 214, 302, 199, 324, 272, 196, 330, 170, 77,
 181, 4, 166, 98, 288, 5, 170, 154, 311, 240, 134, 156, 64, 343,
 26, 55, 34, 208, 88, 268, 309, 154, 268, 237, 217, 148, 6, 88,
 17, 99, 280, 125, 29)
```