Technische Universität Wien
Vienna University of Technology

Inst. f. Statistik u.Wahrscheinlichkeitstheorie

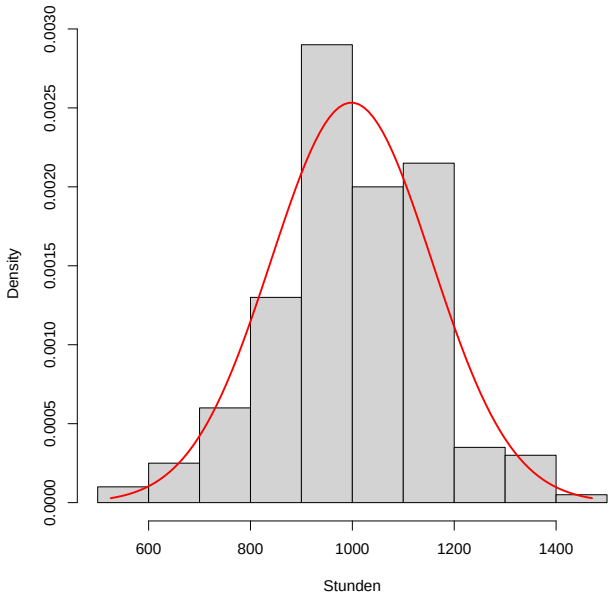
# Statistik und Wahrscheinlichkeitstheorie

Unterlagen zur UEBUNG

W. Gurker

LVA–Nr.: 107.369 (2h)

W 2006|7




A–1040 Wien
Wiedner Hauptstrasse 8-10|107

Kontakt:

Ass.Prof. Dipl.-Ing. Dr.techn. Werner GURKER
Inst. f. Statistik u. Wahrscheinlichkeitstheorie
Technische Universität Wien
Wiedner Hauptstr. 8-10 | 107
A-1040 Wien

Tel.: 58801 - 10724
E-Mail: W.Gurker@tuwien.ac.at
Spr.: Di u. Do von 11-12

# Vorwort

Die folgende Aufgabensammlung dient als Grundlage für die UE *Statistik u. Wahrscheinlichkeitstheorie* [107.369] im WS 2006|7. Sie orientiert sich in Kapitelstruktur und Inhalt an der gleichnamigen *Vorlesung* [107.254].

Auch wenn prinzipiell alle Beispiele mit „Bleistift und Lineal“ (und einem Taschenrechner) lösbar sind, so empfiehlt sich gelegentlich doch die Verwendung eines Programmsystems. Aus der Vielzahl der dafür in Frage kommenden Systeme sei insbesondere auf das unter der GNU–Lizenz frei verfügbare R hingewiesen:

`http://www.r-project.org`

Die einigen Beispielen beigefügten Programmscripts wenden sich an R–User und sind in der Regel nur als Anregung zu verstehen; die Studierenden sind aufgefordert, diese Programmfragmente zu verbessern und auszubauen. Beispiel(teil)e mit einem (*) sind etwas schwieriger, umfangreicher oder behandeln Konzepte, die in der Vorlesung nur gestreift werden; sie sind als Ergänzung gedacht und können auch übersprungen werden.

Im *Anhang* sind einige diskrete und stetige Verteilungen mit ihren wichtigsten Eigenschaften zusammenfassend dargestellt. Auf Tabellen wurde verzichtet; häufig benötigte Quantile finden sich in den meisten Lehrbüchern bzw. sind Bestandteil einschlägiger Programmsysteme.

Wien, Oktober 2006 W.G.

# 1 EINLEITUNG

## 1.1 Historisches und Grundsätzliches

1. Suchen Sie Beispiele für „Statistiken“ (Tageszeitungen, Zeitschriften, Lehrbücher, ...), erläutern Sie die tabellarische und/oder graphische Darstellung und recherchieren Sie den Hintergrund bzw. den Zweck der Untersuchung. Interessieren Sie sich auch dafür, woher die Daten stammen bzw. auf welche Weise sie erhoben wurden.

## 1.2 Beschreibende Statistik

### 1.2.1 Diskrete Merkmale

1. Im WS 2005 verteilten sich die Inskriptionszahlen der Studienrichtungen, für die die VO/UE *Statistik u. Wahrscheinlichkeitstheorie* anrechenbar bzw. verpflichtend ist, wie unten angegeben. Erstellen Sie ein Balkendiagramm und ein Kreisdiagramm für die Gesamtzahlen.

| KNR | Studienrichtung | Inländer Forts. | Inländer Anf. | Ausländer Forts. | Ausländer Anf. | Summe |
|---|---|---|---|---|---|---|
| 526 | Wirtschaftsinformatik | 497 | 106 | 142 | 24 | 769 |
| 531 | Data Egineering & Statistics | 55 | 14 | 22 | 3 | 94 |
| 532 | Medieninformatik | 645 | 210 | 180 | 33 | 1068 |
| 533 | Medizinische Informatik | 285 | 104 | 101 | 24 | 514 |
| 534 | Software & Information Engineering | 702 | 167 | 336 | 61 | 1266 |
| 535 | Technische Informatik | 427 | 117 | 205 | 37 | 786 |

Quelle: TUWIEN

[inf05.r]

2. Betrachten Sie die Zahl der Inversionen einer zufälligen Permutation von $1, 2, \ldots, n$. Das Paar $(i, j)$ ist eine *Inversion*, wenn $i < j$ aber $j$ in der Permutation vor $i$ liegt. Ist $n = 5$ und lautet die Permutation beispielsweise:

$$2 \quad 4 \quad 1 \quad 3 \quad 5$$

so gibt es drei Inversionen: $(1, 2)$, $(1, 4)$ und $(3, 4)$. Die folgende Tabelle enthält die Anzahl der Inversionen von 200 zufälligen (simulierten) Permutationen der Zahlen $1, 2, \ldots, 10$ (UE–Homepage: `inversions.dat`). Erstellen Sie eine Strichliste (Häufigkeitstabelle) und bereiten Sie den Datensatz mittels geeigneter Diagramme graphisch auf (Stabdiagramm, Summenkurve, etc.).

Inversionen

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 18 | 15 | 17 | 27 | 22 | 21 | 23 | 14 | 22 | 11 | 34 | 23 | 19 | 20 | 14 | 31 | 23 | 15 | 19 |
| 30 | 27 | 22 | 35 | 30 | 27 | 17 | 23 | 34 | 19 | 27 | 22 | 25 | 19 | 19 | 19 | 19 | 19 | 28 | 28 |
| 24 | 33 | 18 | 28 | 15 | 17 | 23 | 26 | 32 | 23 | 16 | 14 | 22 | 22 | 26 | 27 | 20 | 26 | 21 | 23 |
| 28 | 29 | 21 | 16 | 21 | 22 | 20 | 31 | 16 | 22 | 17 | 31 | 23 | 21 | 32 | 21 | 24 | 23 | 26 | 23 |
| 20 | 24 | 24 | 30 | 21 | 21 | 20 | 29 | 26 | 22 | 9 | 21 | 19 | 18 | 25 | 26 | 14 | 32 | 18 | 28 |
| 32 | 26 | 18 | 21 | 16 | 26 | 23 | 21 | 21 | 22 | 23 | 28 | 15 | 16 | 25 | 17 | 21 | 24 | 20 | 27 |
| 28 | 26 | 22 | 15 | 17 | 22 | 22 | 15 | 26 | 16 | 30 | 21 | 14 | 24 | 33 | 26 | 29 | 20 | 24 | 28 |
| 25 | 22 | 25 | 16 | 34 | 16 | 23 | 22 | 31 | 27 | 14 | 32 | 31 | 28 | 23 | 22 | 11 | 26 | 25 | 16 |
| 32 | 23 | 26 | 26 | 19 | 22 | 12 | 18 | 22 | 19 | 23 | 17 | 28 | 26 | 20 | 24 | 23 | 25 | 20 | 20 |
| 19 | 31 | 19 | 19 | 20 | 22 | 20 | 27 | 11 | 30 | 12 | 22 | 20 | 27 | 27 | 23 | 24 | 23 | 22 | 14 |

[inversions.r]

3. Ist $x_1, \ldots, x_n$ eine Binärfolge, so heißt die Teilfolge $(x_{i+1}, \ldots, x_{i+k})$ ein *Lauf* der Länge $k$, wenn:
   (1) $i = 0$ oder $x_i \neq x_{i+1}$
   (2) $x_{i+1} = \cdots = x_{i+k}$
   (3) $i + k = n$ oder $x_{i+k} \neq x_{i+k+1}$
   Ist beispielsweise $n = 10$ und lautet die Binärfolge:

   $$0\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 0\ 0$$

   so gibt es 5 Läufe mit den Längen 2, 1, 3, 2, 2. Die folgende Tabelle enthält für 250 (simulierte) Binärfolgen der Länge 60 (mit gleicher Wahrscheinlichkeit für 0 und 1) die Zahl der Läufe (UE–Homepage: `numberruns.dat`). Bilden Sie eine Häufigkeitstabelle und erstellen Sie ein Stabdiagramm und die Summenkurve.

   **Zahl der Läufe**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 28 | 34 | 42 | 23 | 27 | 37 | 31 | 31 | 32 | 29 | 33 | 31 | 34 | 35 | 37 | 28 | 30 | 24 | 38 | 30 |
| 32 | 34 | 35 | 22 | 37 | 37 | 25 | 31 | 30 | 28 | 38 | 31 | 28 | 31 | 31 | 26 | 21 | 31 | 35 | 30 |
| 29 | 26 | 28 | 24 | 30 | 35 | 26 | 32 | 27 | 34 | 25 | 30 | 37 | 32 | 34 | 30 | 36 | 28 | 32 | 33 |
| 29 | 29 | 30 | 29 | 32 | 31 | 30 | 26 | 25 | 35 | 33 | 34 | 29 | 30 | 29 | 29 | 35 | 32 | 30 | 30 |
| 33 | 25 | 32 | 27 | 29 | 34 | 32 | 35 | 32 | 29 | 30 | 29 | 31 | 41 | 32 | 28 | 29 | 35 | 31 | 40 |
| 25 | 31 | 30 | 31 | 26 | 32 | 28 | 31 | 33 | 30 | 28 | 34 | 32 | 25 | 36 | 24 | 29 | 33 | 32 | 29 |
| 33 | 29 | 29 | 33 | 25 | 28 | 29 | 27 | 27 | 40 | 30 | 30 | 30 | 37 | 34 | 35 | 29 | 30 | 29 | 30 |
| 35 | 29 | 35 | 31 | 29 | 31 | 27 | 28 | 35 | 32 | 35 | 27 | 32 | 25 | 33 | 25 | 28 | 26 | 31 | 35 |
| 29 | 32 | 36 | 27 | 36 | 31 | 24 | 33 | 37 | 26 | 25 | 28 | 30 | 33 | 36 | 30 | 30 | 30 | 31 | 34 |
| 30 | 36 | 33 | 27 | 35 | 31 | 34 | 29 | 39 | 31 | 23 | 25 | 39 | 28 | 31 | 27 | 26 | 34 | 26 | 29 |
| 30 | 27 | 32 | 34 | 34 | 31 | 36 | 35 | 31 | 30 | 33 | 34 | 32 | 30 | 34 | 28 | 35 | 29 | 28 | 32 |
| 27 | 30 | 35 | 28 | 29 | 28 | 31 | 27 | 23 | 32 | 30 | 30 | 36 | 27 | 25 | 25 | 25 | 24 | 30 | 35 |
| 35 | 35 | 29 | 29 | 28 | 21 | 26 | 26 | 30 | 34 | | | | | | | | | | |

   [`runs.r`]

4. [Fortsetzung des vorhergehenden Beispiels] Neben der Anzahl der Läufe wurde auch die Länge des *längsten* Laufs ermittelt (UE–Homepage: `maxruns.dat`). Bilden Sie eine Häufigkeitstabelle und erstellen Sie ein Stabdiagramm und die Summenkurve.

   **Länge des längsten Laufs**

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 6 | 10 | 5 | 6 | 7 | 5 | 6 | 6 | 5 | 6 | 6 | 7 | 6 | 4 | 5 | 5 | 5 | 9 | 4 | 6 |
| 6 | 5 | 7 | 10 | 5 | 8 | 10 | 5 | 7 | 6 | 5 | 5 | 5 | 5 | 11 | 6 | 8 | 4 | 4 | 6 |
| 7 | 8 | 6 | 10 | 6 | 4 | 7 | 5 | 6 | 7 | 7 | 6 | 5 | 6 | 6 | 5 | 4 | 8 | 5 | 6 |
| 7 | 7 | 8 | 6 | 4 | 4 | 6 | 9 | 9 | 3 | 4 | 5 | 6 | 8 | 6 | 6 | 6 | 6 | 7 | 8 |
| 7 | 6 | 5 | 8 | 7 | 4 | 6 | 6 | 5 | 10 | 8 | 10 | 5 | 3 | 6 | 7 | 7 | 5 | 7 | 5 |
| 7 | 7 | 5 | 5 | 8 | 5 | 4 | 5 | 5 | 5 | 6 | 4 | 5 | 7 | 5 | 7 | 8 | 4 | 5 | 7 |
| 7 | 15 | 5 | 6 | 8 | 6 | 10 | 6 | 6 | 4 | 5 | 6 | 6 | 5 | 7 | 4 | 7 | 4 | 6 | 5 |
| 5 | 6 | 4 | 9 | 7 | 5 | 9 | 4 | 5 | 7 | 5 | 11 | 5 | 6 | 5 | 6 | 11 | 5 | 11 | 5 |
| 6 | 5 | 5 | 7 | 5 | 7 | 7 | 8 | 3 | 5 | 8 | 5 | 7 | 4 | 4 | 7 | 6 | 5 | 4 | 5 |
| 10 | 7 | 5 | 5 | 6 | 5 | 4 | 7 | 4 | 4 | 7 | 10 | 4 | 5 | 7 | 7 | 8 | 5 | 5 | 7 |
| 6 | 4 | 8 | 5 | 7 | 9 | 4 | 5 | 6 | 7 | 4 | 4 | 6 | 9 | 6 | 7 | 7 | 5 | 6 | 6 |
| 9 | 7 | 4 | 8 | 12 | 6 | 8 | 11 | 10 | 5 | 4 | 5 | 5 | 6 | 8 | 5 | 7 | 7 | 6 | 7 |
| 4 | 4 | 6 | 5 | 7 | 10 | 7 | 5 | 5 | 5 | | | | | | | | | | |

   [`runs.r`]

*5. Eine Möglichkeit zur automatischen Erkennung, in welcher Sprache ein Text abgefaßt ist, ist der Vergleich der (relativen) Buchstabenhäufigkeiten des Textes mit den bekannten (relativen) Häufigkeiten der jeweiligen Sprache. Lesen Sie als Beispiel einen längeren deutschen Text ein, verwandeln Sie Groß– in Kleinbuchstaben, separieren Sie den Text in die einzelnen Zeichen (Buchstaben, Satzzeichen, Sonderzeichen, etc.) und ermitteln Sie die relativen Häufigkeiten der Buchstaben des Alphabets (inklusive Umlaute und Zwischenraum; ß=ss). Stellen Sie die ermittelten Häufigkeiten den für die deutsche Sprache charakteristischen Häufigkeiten gegenüber. Die Angaben bezüglich letzterer schwanken etwas; *eine* Liste finden Sie unter `lettersd.dat` auf der UE–Homepage.

   [`lettersd.r`]

### 1.2.2 Kontinuierliche Merkmale

1. Der Datensatz `resistor.dat` (UE–Homepage) enthält Messungen (in Ohm) von 80 Widerständen:

```
74.4 77.3 77.4 74.3 69.6 71.3 72.5 77.4 73.5 76.1
75.6 75.5 73.3 74.3 77.2 75.1 75.1 79.3 75.8 77.1
76.9 75.8 73.3 74.8 74.7 76.4 78.7 78.5 77.4 74.0
73.2 72.3 76.9 76.2 76.3 74.1 77.0 74.1 72.5 72.7
73.6 78.4 77.7 71.6 73.2 76.4 71.8 78.1 75.6 74.0
75.0 77.5 76.4 72.5 72.8 71.8 73.1 75.0 77.7 77.9
76.2 75.0 76.4 76.3 73.1 74.7 76.0 75.6 75.3 79.6
74.6 77.0 72.1 75.2 75.7 74.7 73.6 75.2 76.6 74.6
```

   (a) Ermitteln und zeichnen Sie die empirische Verteilungsfunktion.

   (b) Stellen Sie die Verteilung der relativen Klassenhäufigkeiten in Form eines Histogramms dar; nehmen Sie dazu beispielsweise die folgende äquidistante Klasseneinteilung:

$$[69.55,\ 70.65], (70.65,\ 71.75], \ldots, (79.45,\ 80.55]$$

   Wählen Sie die Darstellung so, daß die Fläche unter dem Histogramm gleich 1 ist („Dichtehistogramm“ oder „flächentreues“ Histogramm genannt).

   (c) Zeichnen Sie auf Basis der obigen Klasseneinteilung das Summenpolygon.

   [`resistor.r`]

2. Die folgende Tabelle enthält Angaben zu Geschlecht (GE), Körpergröße (GR) und Körpergewicht (GW) für Studierende einer medizinischen Universität (erweiterter Datensatz auf der UE–Homepage: `meddat.dat`):

| | GE | GR | GW | | GE | GR | GW | | GE | GR | GW |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | M | 178 | 70 | 24 | M | 170 | 66 | 47 | W | 170 | 54 |
| 2 | M | 178 | 75 | 25 | M | 173 | 68 | 48 | W | 170 | 60 |
| 3 | M | 179 | 60 | 26 | M | 174 | 74 | 49 | W | 175 | 65 |
| 4 | M | 179 | 71 | 27 | M | 177 | 81 | 50 | W | 176 | 65 |
| 5 | M | 187 | 75 | 28 | M | 178 | 72 | 51 | W | 161 | 56 |
| 6 | M | 178 | 67 | 29 | M | 183 | 72 | 52 | W | 168 | 54 |
| 7 | M | 179 | 76 | 30 | M | 190 | 83 | 53 | W | 170 | 57 |
| 8 | M | 181 | 73 | 31 | M | 178 | 66 | 54 | W | 172 | 56 |
| 9 | M | 186 | 70 | 32 | M | 180 | 72 | 55 | W | 174 | 64 |
| 10 | M | 186 | 75 | 33 | M | 183 | 69 | 56 | W | 176 | 60 |
| 11 | M | 188 | 77 | 34 | M | 198 | 90 | 57 | W | 178 | 65 |
| 12 | M | 189 | 80 | 35 | M | 179 | 70 | 58 | W | 164 | 60 |
| 13 | M | 189 | 89 | 36 | M | 180 | 79 | 59 | W | 165 | 52 |
| 14 | M | 190 | 75 | 37 | M | 187 | 82 | 60 | W | 173 | 62 |
| 15 | M | 190 | 80 | 38 | M | 192 | 80 | 61 | W | 165 | 58 |
| 16 | M | 192 | 96 | 39 | M | 182 | 98 | 62 | W | 177 | 60 |
| 17 | M | 170 | 68 | 40 | W | 158 | 46 | 63 | W | 160 | 48 |
| 18 | M | 175 | 75 | 41 | W | 166 | 63 | 64 | W | 163 | 65 |
| 19 | M | 178 | 64 | 42 | W | 168 | 63 | 65 | W | 167 | 72 |
| 20 | M | 180 | 67 | 43 | W | 170 | 57 | 66 | W | 168 | 56 |
| 21 | M | 184 | 88 | 44 | W | 165 | 50 | 67 | W | 169 | 59 |
| 22 | M | 186 | 73 | 45 | W | 166 | 49 | 68 | W | 162 | 57 |
| 23 | M | 192 | 94 | 46 | W | 168 | 56 | 69 | W | 164 | 65 |

   Betrachten Sie das Merkmal „Körpergröße“:

   (a) Stellen Sie die Verteilung in Form eines (Dichte–) Histogramms dar. (Die Klasseneinteilung sollte dabei nicht zu fein aber auch nicht zu grob sein.)

   (b) Zeichnen Sie auf Basis der obigen Klasseneinteilung das Summenpolygon.

   Zeichnen Sie Histogramm und Summenpolygon sowohl für alle Beobachtungseinheiten zusammen als auch getrennt nach Geschlecht.

   [`meddat.r`]

3. [Fortsetzung des vorhergehenden Beispiels] Wiederholen Sie die Aufgabe für das Merkmal „Körpergewicht“. Zeichnen Sie Histogramm und Summenpolygon sowohl für alle Beobachtungseinheiten zusammen als auch getrennt nach Geschlecht.

   [`meddat.r`]

4. Ein wichtiges Qualitätskriterium von Wasser ist die Konzentration [ppm] von schwebenden festen Teilchen. Die folgende Tabelle enthält 40 Messungen dieser Größe für einen Badesee (UE–Homepage: `solid.dat`):

```
3.41 0.45 0.29 1.15 4.40 1.98 1.08 0.99 1.23 1.52
0.98 2.47 0.30 1.79 0.44 0.37 1.66 3.49 0.15 0.82
1.93 0.12 0.95 0.44 0.16 6.94 0.61 3.47 0.36 2.12
1.31 1.40 0.29 0.51 0.86 0.55 0.85 1.32 0.57 0.27
```

   (a) Ermitteln und zeichnen Sie ein Histogramm sowie die empirische Verteilungsfunktion.

   *(b) Wie sich zeigt, ist die Verteilung stark (rechts–) schief. Mittels einer geeigneten Transformation kann man versuchen, die Verteilung zu symmetrisieren. Welche der folgenden Transformationen eignet sich dafür am besten?

$$x^* = \sqrt{x}, \quad x^* = \ln x, \quad x^* = 1/x$$

   [`solid.r`]

5. Der Datensatz `bulb.dat` (UE–Homepage) umfaßt die Lebensdauern [Stunden] von 200 Glühbirnen.

   (a) Ermitteln Sie eine Häufigkeitstabelle; nehmen Sie als Klassengrenzen: 500(100)1500.

   (b) Zeichnen Sie auf Basis der obigen Klasseneinteilung das Histogramm und das Summenpolygon.

   (c) Zeichnen Sie die empirische Verteilungsfunktion.

   [`bulb.r`]

### 1.2.3 Lageparameter

1. Bestimmen Sie für den Datensatz `inversions.dat` die folgenden Lageparameter: Mittelwert, Median, Modalwert (Modus).

   [`invers.r`]

2. Bestimmen Sie für die Datensätze (a) `NumberRuns.dat` und (b) `MaxRuns.dat` die folgenden Lageparameter: Mittelwert, Median, Modalwert.

   [`runs.r`]

3. Bestimmen Sie für den Datensatz `resistor.dat` die folgenden Lageparameter: Mittelwert (unklassierte und klassierte Daten), Modalwert (klassierte Daten).

   [`resistor.dat`]

4. Bestimmen Sie für den Datensatz `bulb.dat` die Quartile, d.h. das 25%–, das 50%– (= Median) und das 75%–Quantil (Fraktil); stützen Sie sich dabei auf die klassierten Daten. *Wie lautet ein allgemeiner Ausdruck für das $p$–Quantil (Fraktil) auf Basis eines klassierten Datensatzes?

   [`bulb.r`]

5. Zeigen Sie für einen beliebigen Datensatz (diskret oder stetig) $x_1, \ldots, x_n$, daß:

   (a) der Mittelwert $\overline{x}_n$ die folgende Minimumeigenschaft hat:

$$\sum_{i=1}^{n} (x_i - \overline{x}_n)^2 \leq \sum_{i=1}^{n} (x_i - c)^2, \quad c \in \mathbb{R}$$

   *(b) der Median $\widetilde{x}$ die folgende Minimumeigenschaft hat:

$$\sum_{i=1}^{n} |x_i - \widetilde{x}| \leq \sum_{i=1}^{n} |x_i - c|, \quad c \in \mathbb{R}$$

### 1.2.4 Streuungsparameter

1. Bestimmen Sie für den Datensatz `inversions.dat` die mittlere quadratische Abweichung (empirische Varianz) und die empirische Streuung.

   [`inversions.r`]

2. Bestimmen und vergleichen Sie für das Merkmal „Größe“ (Datensatz: `meddat.dat`) die Varianz, die Streuung und den Variationskoeffizienten für Männer und Frauen.

   [`meddat.r`]

3. [Fortsetzung des vorhergehenden Beispiels] Wiederholen Sie die Aufgabe für das Merkmal „Gewicht“.

   [`meddat.r`]

4. Ermitteln Sie für den Datensatz `bulb.dat` (UE–Homepage) die folgenden Streuungsparameter: Spannweite, Quartilabstand, MAD, Varianz, Streuung.

   **Zusatz:* Wie in der VO kurz diskutiert, versucht man in der schließenden Statistik den empirisch gegebenen Verteilungen (Histogrammen) theoretische Verteilungen (Dichten) anzupassen. Man versuche dies hier mit der Anpassung einer *Normaldichte*; für die beiden Parameter dieser Verteilung ($\mu$, $\sigma^2$) nehme man die entsprechenden empirischen Größen ($\overline{x}$, $s^2$).

   [`bulb.r`]

5. Zeigen Sie den *Verschiebungssatz* für die Varianz (Daten: $x_1, \ldots, x_n$):

$$\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = \sum_{i=1}^{n}(x_i - c)^2 - n(c - \overline{x}_n)^2, \quad c \in \mathbb{R}$$

   Speziell für $c = 0$ gilt:

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_n)^2 = \frac{1}{n}\left[\sum_{i=1}^{n} x_i^2 - n\overline{x}_n^2\right] = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \overline{x}_n^2$$

6. Angenommen, es gibt Daten (Meßwerte) $x_1, \ldots, x_j$ und Sie benötigen den (empirischen) Mittelwert $\overline{x}_j$ und die (empirische) Varianz $s_j^2$. Nun kommt eine weitere Beobachtung $x_{j+1}$ dazu. Im Sinne einer *Realtime*–Berechnung ist es nicht notwendig, den Mittelwert und die Varianz für alle $j+1$ Daten neu zu berechnen, sondern man kann auf die bereits vorhandenen Werte zurückgreifen. Zeigen Sie:

   (a) Rekursion für den Mittelwert:

$$\overline{x}_{j+1} = \overline{x}_j + \frac{1}{j+1}(x_{j+1} - \overline{x}_j), \quad j = 1, 2, \ldots; \ \overline{x}_1 = x_1$$

   *(b) Rekursion für die Varianz:

$$s_{j+1}^2 = \left(1 - \frac{1}{j+1}\right) s_j^2 + j(\overline{x}_{j+1} - \overline{x}_j)^2, \quad j = 1, 2, \ldots; \ s_1^2 = 0$$

   Überprüfen Sie die Gültigkeit der beiden Rekursionen konkret an den Werten: 3, 4, 7, 2, 9, 6.

## 1.3 Wahrscheinlichkeitsbegriffe

### 1.3.1 Klassische Wahrscheinlichkeitsdefinition

1. (a) Bei einem Picknick von 50 Personen hatten 30 Hamburger, 25 hatten Hotdogs und 15 hatten beides. Wieviele hatten weder noch?

   (b) Auf wieviele Arten kann ein 20 Fragen wahr/falsch–Test beantwortet werden?

   (c) In einer Lade sind 12 schwarze und 12 weiße Socken. Welche minimale Anzahl von Socken muß man (zufällig) entnehmen, um zumindest ein passendes Paar zu bekommen?

   (d) In einem Behälter befinden sich 3 rote, 4 weiße und 5 blaue Kugeln. Auf wieviele Arten kann man 4 Kugeln entnehmen, sodaß von jeder Farbe eine Kugel darunter ist?

2. In einer Schuhablage gibt es 5 verschiedene Paare und 5 Schuhe werden willkürlich herausgegriffen. Mit welcher Wahrscheinlichkeit befindet sich darunter kein Paar, 1 Paar, 2 Paare?
3. Angenommen, alle 365 Tage eines Jahres kommen mit gleicher Wahrscheinlichkeit als Geburtstage in Frage. Wie groß ist die Wahrscheinlichkeit, daß in einer Gruppe von 60 Personen genau 5 Personen am selben Tag geboren sind?
4. Sieben Personen betreten einen Lift im Erdgeschoß eines 11–stöckigen Gebäudes. Angenommen, die Personen steigen unabhängig voneinander und zufällig auf einem der Stockwerke 1 bis 11 aus. Mit welcher Wahrscheinlichkeit steigen alle auf verschiedenen Stockwerken aus?
5. In einem Feld der Länge $n$ werden zufällig $k$ ($\leq n$) Daten abgelegt. Mit welcher Wahrscheinlichkeit kommt es dabei zu Kollisionen (d.h. Mehrfachbelegungen)? Wie groß muß $k$ konkret für $n = 100$ mindestens sein, damit diese Wahrscheinlichkeit größer als 0.5 (0.9) ist?
6. (a) Wieviele Möglichkeiten gibt es, $r$ unterscheidbare Kugeln auf $n$ Fächer so zu verteilen, daß das $i$–te Fach $r_i$ Kugeln enthält ($i = 1, \ldots, n$, $\sum_{i=1}^{n} r_i = r$)? Bestimmen Sie die Wahrscheinlichkeit für dieses Ereignis.

   *(b) Wieviele Möglichkeiten gibt es, $r$ *ununterscheidbare* Kugeln auf $n$ Fächer zu verteilen? Sind diese Möglichkeiten alle gleichwahrscheinlich?

### 1.3.2 Geometrische Wahrscheinlichkeiten

1. Zwei Personen haben die gleiche Wahrscheinlichkeit an einem bestimmten Ort zu einem beliebigen Zeitpunkt im Intervall $[0, T]$ einzutreffen. Die zuerst eintreffende Person wartet auf die andere. Ermitteln Sie die Wahrscheinlichkeit, daß keine Person länger als $t$ auf die andere warten muß.
2. Zwei Wanderer erreichen aus unterschiedlichen Richtungen einen Aussichtspunkt und halten sich dort 10 Minuten (Wanderer 1) bzw. 20 Minuten (Wanderer 2) auf. Ihre Ankunftszeitpunkte liegen – unabhängig voneinander – zufällig zwischen 11 und 12 Uhr.

   (a) Mit welcher Wahrscheinlichkeit begegnen sie einander am Aussichtspunkt?

   (b) Wie groß ist die Wahrscheinlichkeit, daß sich um 11:30 (1) keiner, (2) genau einer, (3) beide am Aussichtspunkt befinden?
3. M verläßt zufällig zwischen 15 und 17 Uhr seinen/ihren Arbeitsplatz und begibt sich zur U–Bahn. Seine/Ihre Mutter lebt am einen Ende der Stadt, seine/ihr Freundin/Freund am anderen. Er/Sie will fair sein und nimmt jeweils diejenige U–Bahn, welche als erste eintrifft. Nach einiger Zeit beklagt sich die Mutter darüber, daß er/sie nur ganz selten zum Abendessen kommt; an den letzten 20 Arbeitstagen nur zweimal. Kommt dieses Ungleichgewicht zufällig zustande oder gibt es eine andere Erklärung dafür?
4. Zwei Punkte werden willkürlich auf der Oberfläche einer Kugel mit Radius $R$ markiert. Wie groß ist die Wahrscheinlichkeit, daß der (Groß–) Kreisbogen, der die beiden Punkte miteinander verbindet, einen Winkel kleiner als $\alpha$ ($< \pi$) einschließt?
5. Bestimmen Sie die Wahrscheinlichkeit, daß die Wurzeln der quadratischen Gleichung $x^2 + 2ax + b = 0$ reell sind, wenn bekannt ist, daß die Koeffizienten mit gleicher Wahrscheinlichkeit aus dem Rechteck $|a| \leq A$, $|b| \leq B$ stammen. Bestimmen Sie unter diesen Bedingungen auch die Wahrscheinlichkeit dafür, daß die Wurzeln beide positiv sind.

*6. Drei Geradensegmente, deren Länge nicht größer als $L$ ist, werden zufällig gewählt. Wie groß ist die Wahrscheinlichkeit, daß man mit ihnen ein Dreieck bilden kann?

### 1.3.3 Grenzwert von Häufigkeiten

1. Der französische Offizier und Schriftsteller Chevalier de Méré (1607 – 1684) wandte sich im Jahre 1654 mit der folgenden Frage an Blaise Pascal (1623 – 1662): Was ist vorteilhafter, beim Spiel mit einem Würfel auf das Eintreten mindestes eines Sechsers in vier Würfen oder beim Spiel mit zwei Würfeln auf das Eintreten eines Doppelsechsers in 24 Würfen zu setzen? Er wußte aus Erfahrung, daß die erste Wette für ihn vorteilhaft ist; bei der zweiten Wette, von der er annahm, daß sie nur eine Variante der ersten sei, gestalteten sich die Einnahmen aber nicht nach seinen Vorstellungen.

   [`demere.r`]

2. Ein ähnliches (Wett–) Problem hatte der englische Abgeordnete und berühmte Tagebuchschreiber („Die geheimen Tagebücher“) Samuel Pepys (1633 – 1703). Welches der folgenden Ereignisse ist am wahrscheinlichsten?

   (1) Mindestens 1 Sechser beim Werfen von 6 Würfeln.

   (2) Mindestens 2 Sechser beim Werfen von 12 Würfeln.

   (3) Mindestens 3 Sechser beim Werfen von 18 Würfeln.

   Als Präsident der *Royal Society* gehörte u.a. Isaac Newton (1643 – 1727) zu seinem Bekanntenkreis. Letzterem legte er sein Problem in einem kompliziert formulierten Brief dar.

   [`pepys.r`]

### 1.3.4 Axiomatische Wahrscheinlichkeiten

1. Bestimmen Sie das komplementäre Ereignis $A^\perp$ zu $A = \{(x,y) : x^2 + y^2 < 2\}$ in Bezug auf das sichere Ereignis $e = \{(x,y) : |x| + |y| \le 2\}$.
2. Bestimmen Sie $\bigvee_{k=1}^{\infty} A_k$ für die folgenden Ereignisse:

   (a) $A_k = \{x : 1/k \le x \le 3 - 1/k\}$, $k = 1, 2, 3, \ldots$

   (b) $A_k = \{(x,y) : 1/k \le x^2 + y^2 \le 4 - 1/k\}$, $k = 1, 2, 3, \ldots$
3. Bestimmen Sie $\bigwedge_{k=1}^{\infty} A_k$ für die folgenden Ereignisse:

   (a) $A_k = \{x : 2 - 1/k < x \le 2\}$, $k = 1, 2, 3, \ldots$

   (b) $A_k = \{x : 2 < x \le 2 + 1/k\}$, $k = 1, 2, 3, \ldots$

   (c) $A_k = \{(x,y) : 0 \le x^2 + y^2 \le 1/k\}$, $k = 1, 2, 3, \ldots$
4. Für jedes (eindimensionale) Ereignis $A$ sei eine Wahrscheinlichkeit wie folgt definiert:

$$W(A) = \sum_A f(x) \quad \text{mit} \quad f(x) = \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^x, \; x = 0, 1, 2, \ldots \quad (f(x) = 0 \text{ sonst})$$

   Wenn $A_1 = \{x : x = 0, 1, 2, 3\}$ und $A_2 = \{x : x = 0, 1, 2, \ldots\}$, bestimmen Sie $W(A_1)$ und $W(A_2)$.
5. Für jedes (eindimensionale) Ereignis $A$ sei eine Wahrscheinlichkeit wie folgt definiert:

$$W(A) = \int_A f(x)\,dx \quad \text{mit} \quad f(x) = 6x(1-x)\,I_{(0,1)}(x)$$

   (Existiert das Integral nicht, ist $W(A)$ nicht definiert.) Wenn $A_1 = \{x : \frac{1}{4} < x < \frac{3}{4}\}$, $A_2 = \{\frac{1}{2}\}$ und $A_3 = \{x : 0 < x < 10\}$, bestimmen Sie $W(A_1)$, $W(A_2)$ und $W(A_3)$.

### 1.3.5 Subjektive Wahrscheinlichkeiten

1. Wahrscheinlichkeiten – insbesondere subjektive – werden häufig in Form von Verhältnissen angegeben. So sagt man beispielsweise, die *Chancen* (engl. *odds*) für den Eintritt von $A$ stehen 1 zu 10. Heißt das nun, daß die Wahrscheinlichkeit $W(A)$ von $A$ gleich 1/9, 1/10 oder 1/11 ist? Umgekehrt, wenn beispielsweise $W(A) = 2/3$, stehen dann die Chancen (für den Eintritt von $A$) $2:3$, $3:2$, oder $2:1$ ?
2. Wenn die Chancen $2:3$ für $A$, $3:7$ für $B$ und $1:4$ für $A \cap B$ stehen, wie stehen dann die Chancen für $A \cup B$ ?
3. Ein Buchmacher offeriert in einem bestimmten Rennen für *Lucky Star* eine Quote von $99:1$ auf Sieg. Bedeutet das, daß die Siegeschance des Pferdes in diesem Rennen (nach Einschätzung des Buchmachers) größer, kleiner oder gleich 1/100 ist ?
4. In einem Rennen mit $n$ Pferden offeriert ein Buchmacher Quoten von $r_i : 1$, $i = 1, \ldots, n$.

(a) Betrachten Sie die Summe:

$$S = \sum_{i=1}^{n} \frac{1}{r_i + 1}$$

Ist $S$ größer, kleiner oder gleich 1 ?

*(b) Diskutieren Sie den Fall $S < 1$. Könnte man (und wenn ja, wie) einen Vorteil daraus ziehen?

5. Im Wetterbericht hören Sie, daß für morgen in Wien die Wahrscheinlichkeit für Regen 30% beträgt. Was bedeutet das?

(a) In 30% des Stadtgebiets wird es morgen regnen.

(b) Zu 30% der Zeit wird es morgen regnen.

(c) An 30% der Tage wie morgen regnet es.

### 1.3.6 Unscharfe Wahrscheinlichkeiten

*1. Ein Experiment bestehe in der Wahl eines Punktes aus dem Intervall $[0, 1]$. Für eine Menge $C \subseteq [0, 1]$ sei eine unscharfe Wahrscheinlichkeit wie folgt definiert: Die Zugehörigkeitsfunktion von $P^*(C)$ ist dreiecksförmig mit den Eckpunkten ($|C|$ = Länge von $C$) $(0, 0)$, $(|C|, 1)$, $(2|C|, 0)$, falls $|C| \leq 0.5$, und den Eckpunkten $(2|C| - 1, 0)$, $(|C|, 1)$, $(1, 0)$, falls $|C| > 0.5$. Zeigen Sie, daß es sich um eine unscharfe Wahrscheinlichkeitsverteilung handelt.