§3 Relative Häufigkeit

▼

```
RelativeTally[liste_] := Module[{h, n, m},
   h = Tally[liste];
   n = Length[liste];
   m = Length[h];
   Table[{h[[i, 1]], N[h[[i, 2]]/n]}, {i, 1, m}]]
```

Wird ein Zufallsexperiment oft wiederholt, so ist das dabei erzeugte Datenmaterial nicht mehr überschaubar. Meistens ist man an diesem konkreten Datenmaterial selbst aber gar nicht interessiert. Vielmehr möchte man oft nur wissen, mit welchen relativen Häufigkeiten die einzelnen Realisierung dieses Zufallsexperiments auftreten. Außerdem möchte man diese relativen Häufigkeiten sowohl tabellarisch als auch graphisch veranschaulichen.

An Hand von Beispielen werden wir experimentell zeigen, dass die (durch Simulation gewonnene) relative Häufigkeit $\mathbb{H}[A]$ eines Ereignisses $A \subseteq \Omega$ mit zunehmender Anzahl von Wiederholungen gegen die tatsächliche Wahrscheinlichkeit $\mathbb{P}[A]$ dieses Ereignisses "konvergiert". Da es sich bei der Abbildung $\mathbb{H}$, welche jedem Ereignis $A \subseteq \Omega$ die relative Häufigkeit $\mathbb{H}[A]$ zuordnet, offenbar um ein W-Maß auf Ω handelt, kann man das W-Maß $\mathbb{H}$ als Modell für den das Zufallsexperiment steuernden Zufall verwenden. Dieses Modell $\mathbb{H}$ entspricht dabei umso besser dem tatsächlichen (oft nicht explizit bekannten) Modell $\mathbb{P}$, je größer die Anzahl der Wiederholungen ist.

3.1 Statistische Auswertung von Daten

Wir beginnen mit einer wichtigen Begriffsbildung:

3.1.1 Definition: Wird ein Zufallsexperiment mit Ereignisraum Ω (etwa durch Simulation) oft wiederholt, so nennt man für jedes Ereignis $A \subseteq \Omega$ die reelle Zahl

$$\mathbb{H}[A] := \frac{\text{Anzahl der Wiederholungen, in denen das Ereignis } A \text{ eintritt}}{\text{Anzahl der Wiederholungen}}$$

die relative Häufigkeit von A. Dabei beachte man, dass die relative Häufigkeit eines Ereignisses $A \subseteq \Omega$ von der vorliegenden Versuchsreihe abhängt und sich damit von Versuchsreihe zu Versuchsreihe ändert.

Wird ein Zufallsexperiment (beispielsweise durch Simulation) sehr oft wiederholt, so ist das dabei erzeugte Datenmaterial nicht mehr überschaubar. Meistens ist man aber ohnehin nur daran interessiert, mit welchen absoluten bzw relativen Häufigkeiten die einzelnen Werte auftreten. Zur Berechnung dieser Häufigkeiten dienen die Befehle Tally bzw RelativeTally:

■ Tally[*liste*]

erzeugt eine Liste der absoluten Häufigkeiten der Liste *liste*. Darunter versteht man eine Matrix, bei der die in der Liste *liste* vorkommenden Werte zusammen mit ihren absoluten Häufigkeiten aufgelistet sind.

■ RelativeTally[*liste*]

erzeugt eine Liste der relativen Häufigkeiten der Liste *liste*. Darunter versteht man eine Matrix, bei der die in der Liste *liste* vorkommenden Werte zusammen mit ihren relativen Häufigkeiten aufgelistet sind.

▼

Beispielsweise gilt (mit Hilfe von Sort bringen wir die zu behandelnde Liste in ihre natürliche Reihenfolge)

```
n = 500;
liste = Sort[RandomInteger[{1, 6}, n]];
```

```
Tally[liste]
RelativeTally[liste]
Clear[n, liste]
```

```
{{1, 80}, {2, 81}, {3, 81}, {4, 101}, {5, 76}, {6, 81}}

{{1, 0.16}, {2, 0.162}, {3, 0.162}, {4, 0.202}, {5, 0.152}, {6, 0.162}}
```

Eine statistische Auswertung von in einer Liste zusammengefassten Daten besteht in der tabellarischen Angabe der in dieser Liste vorkommenden Werte zusammen mit ihren relativen Häufigkeiten sowie der graphischer Veranschaulichung dieser relativen Häufigkeiten durch ein geeignetes Diagramm.

Wir machen dazu einige Beispiele:

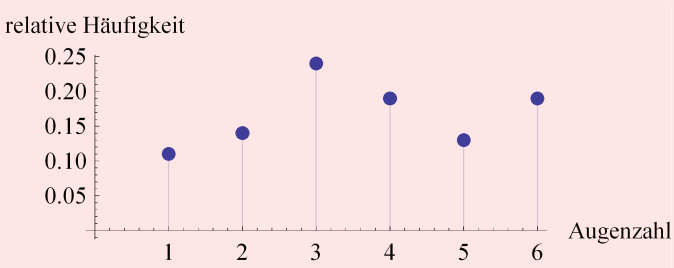
3.1.2 Beispiel: Unser Zufallsexperiment besteht im Werfen eines Würfels.
- Man simuliere dieses Zufallsexperiment n mal und werte dieses Datenmaterial statistisch aus.
- Man beobachte, dass sich die relativen Häufigkeiten der einzelnen Augenzahlen zwar von Versuch zu Versuch ändern, sich aber mit zunehmendem n der theoretischen Wahrscheinlichkeit $1/6 = 0.166\ldots$ nähern.

▼

Lösung: Wir simulieren das n-malige Werfen eines Würfels in der üblichen Weise, ordnen dieses Datenmaterial mit Hilfe von Sort und ermitteln anschließend mit Hilfe von RelativeTally eine Liste mit den relativen Häufigkeiten der dabei geworfenen Augenzahlen. Diese relativen Häufigkeiten geben wir einerseits mit Hilfe von TableForm als Tabelle aus und veranschaulichen sie andererseis graphisch unter Verwendung von ListPlot:

```
n = 100;
data = RelativeTally[Sort[RandomInteger[{1, 6}, n]]];
TableForm[data, TableHeadings → {None, {"Augenzahl", "relative Häufigkeit"}}, TableSpacing → {1, 10}]
ListPlot[data, PlotStyle → PointSize[0.03], Filling → Axis, AspectRatio → 0.4, AxesOrigin → {0, 0},
  PlotRange → All, AxesLabel → {"Augenzahl", "relative Häufigkeit"}, ImageSize → {250, 100}]
Clear[n, data]
```

Augenzahl	relative Häufigkeit
1	0.11
2	0.14
3	0.24
4	0.19
5	0.13
6	0.19



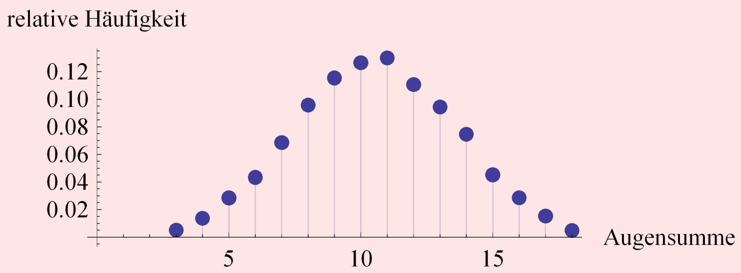
3.1.3 Beispiel: Unser Zufallsexperiment besteht im gleichzeitigen Werfen von k Würfeln und der anschließenden Ermittlung der geworfenen Augensumme.
- Man simuliere dieses Zufallsexperiment n mal und werte dieses Datenmaterial statistisch aus.
- Man beobachte wieder, wie sich die relativen Häufigkeiten zwar von Versuch zu Versuch ändern, sich mit zunehmendem n aber gewissen Werten annähern. Diese Tatsache zeigt sich auch dadurch, dass das Diagramm der relativen Häufigkeiten mit zunehmendem n immer "stabiler" wird.

▼

Lösung: Wir simulieren das gleichzeitige Werfen von *k* Würfeln und die anschließende Ermittlung der dabei geworfenen Augensumme in der üblichen Weise *n* mal, ordnen dieses Datenmaterial mit Hilfe von Sort und ermitteln anschließend mit Hilfe von RelativeTally eine Liste mit den relativen Häufigkeiten der dabei geworfenen Augensumme. Diese Liste geben wir mit Hilfe von TableForm als Tabelle aus und veranschaulichen sie graphisch unter Verwendung von ListPlot:

```
k = 3; n = 10000;
data = RelativeTally[Sort[Table[Apply[Plus, RandomInteger[{1, 6}, k]], {n}]]];
TableForm[data, TableHeadings → {None, {"Augensumme", "relative Häufigkeit"}}, TableSpacing → {1, 10}]
ListPlot[data, PlotStyle → PointSize[0.03], Filling → Axis, AspectRatio → 0.4, AxesOrigin → {0, 0},
  PlotRange → All, AxesLabel → {"Augensumme", "relative Häufigkeit"}, ImageSize → {300, 100}]
Clear[k, n, data]
```

Augensumme	relative Häufigkeit
3	0.005
4	0.0136
5	0.0284
6	0.0433
7	0.0685
8	0.0958
9	0.1154
10	0.1265
11	0.1299
12	0.1106
13	0.0944
14	0.0746
15	0.0452
16	0.0286
17	0.0153
18	0.0049



Der Glücksritter und Spieler Chevalier de MERÉ diskutierte einst mit dem Mathematiker B. PASCAL die folgende Frage: Mit drei Würfeln wird gleichzeitig gewürfelt. Besitzen die beiden Ereignisse "es wird dabei die Augensumme 11 gewürfelt" und "es wird dabei die Augensumme 12 gewürfelt" die gleiche Wahrscheinlichkeit?

De MERÉ machte nämlich im Laufe seiner Spielerkarriere die Erfahrung, dass die Augensumme 11 etwas häufiger auftritt als die Augensumme 12, war aber auf Grund seiner (fehlerhaften) Berechnungen davon überzeugt, dass diese beiden Ereignisse mit der gleich Wahrscheinlichkeit eintreten sollten. Aus der oben durchgeführten Simulation geht hervor, dass die Augensumme 11 tatsächlich mit etwas höherer Wahrscheinlichkeit eintreten dürfte, als die Augensumme 12.

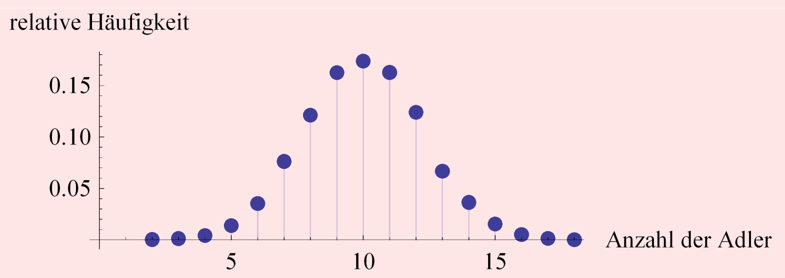
3.1.4 Beispiel: Unser Zufallsexperiment besteht im *k*-maligen Werfen einer Münze und der anschließenden Feststellung, wie oft dabei das Ereignis "Adler" eingetreten ist.
- Man simuliere dieses Zufallsexperiment *n* mal und werte dieses Datenmaterial statistisch aus.
- Man beobachte wieder, wie sich die relativen Häufigkeiten zwar von Versuch zu Versuch ändern, sich mit zunehmendem *n* aber gewissen Werten annähern. Diese Tatsache zeigt sich wieder dadurch, dass das Diagramm der relativen Häufigkeiten mit zunehmendem *n* immer "stabiler" wird.

▼

Lösung: Wir simulieren das k-malige Werfen einer Münze und die anschließende Feststellung, wie oft dabei das Ereignis "Adler" eingetreten ist, in der üblichen Weise n mal, ordnen dieses Datenmaterial mit Hilfe von Sort und ermitteln mit Hilfe von RelativeTally die Liste der relativen Häufigkeiten dafür, wie oft das Ereignis "Adler" eingetreten ist. Diese Liste geben wir mit Hilfe von TableForm als Tabelle aus und veranschaulichen sie graphisch unter Verwendung von ListPlot:

```
k = 20; n = 10 000;
data = RelativeTally[Sort[Table[Apply[Plus, RandomInteger[{0, 1}, k]], {n}]]];
TableForm[data, TableHeadings → {None, {"Anzahl der Adler", "relative Häufigkeit"}}, TableSpacing → {1, 10}]
ListPlot[data, PlotStyle → PointSize[0.03], Filling → Axis, AspectRatio → 0.4, AxesOrigin → {0, 0},
  PlotRange → All, AxesLabel → {"Anzahl der Adler", "relative Häufigkeit"}, ImageSize → {300, 100}]
Clear[k, n, data]
```

Anzahl der Adler	relative Häufigkeit
2	0.0003
3	0.0012
4	0.0041
5	0.0138
6	0.0352
7	0.0762
8	0.1211
9	0.1625
10	0.1738
11	0.1627
12	0.124
13	0.0667
14	0.0364
15	0.0154
16	0.0051
17	0.0013
18	0.0002



3.2 Näherungsweise Ermittlung der Wahrscheinlichkeit eines Ereignisses durch dessen relative Häufigkeit

Die Beispiele aus dem vorhergehenden Abschnitt zeigen insgesamt, dass sich die relative Häufigkeit $\mathbb{H}[A]$ eines Ereignisses $A \subseteq \Omega$ zwar von Versuchsreihe zu Versuchsreihe ändert, sich mit zunehmender Anzahl von Wiederholungen aber einem Wert nähert, den man als theoretische Wahrscheinlichkeit $\mathbb{P}[A]$ dieses Ereignisses interpretieren kann.

Wie viele Wiederholungen dabei notwendig sind, damit die (etwa durch Simulation) experimentell bestimmbare relative Häufigkeit $\mathbb{H}[A]$ von der (oft unbekannten) theoretischen Wahrscheinlichkeit $\mathbb{P}[A]$ um weniger als eine vorgegebene Schranke abweicht, wird durch die folgende Faustregel der Simulation beschrieben:

3.2.1 Bemerkung: Soll die theoretische Wahrscheinlichkeit $\mathbb{P}[A]$ eines Ereignisses A durch die relative Häufigkeit $\mathbb{H}[A]$ näherungsweise bestimmt werden, wobei der Fehler kleiner als 10^{-k} sein soll, so sind dazu etwa $n = 10^{-2k}$ Wiederholungen erforderlich.

Wir veranschaulichen diese Faustregel an einem einfachen Beispiel (hinsichtlich einer theoretischen Bestätigung dieser Faustregel vergleiche man Beispiel 23.1.5):

3.2.2 Beispiel: Unser Zufallsexperiment besteht im Werfen von zwei homogenen Würfeln und der anschließenden Beobachtung der dabei auftretenden Augensumme.
- Man berechne die theoretischen Wahrscheinlichkeiten der möglichen Augensummen 2, 3, …, 12.
- Man ermittle mit Hilfe der Simulation Näherungswerte für diese theoretischen Wahrscheinlichkeiten und ermittle deren Abweichung von den theoretischen Wahrscheinlichkeiten, wenn zur Bestimmung der relativen Häufigkeiten n Wiederholungen herangezogen werden.

▼

Lösung:

a) Als Ereignisraum für unser Zufallsexperiment bietet sich die Menge (vgl dazu Beispiel 2.2.7)

$$\Omega = \{\{i, k\} \mid i, k \in \{1, 2, 3, 4, 5, 6\}\}$$

an, wobei alle Realisierungen gleichwahrscheinlich sind. Die Augensummen 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 treten daher mit den theoretischen Wahrscheinlichkeiten

$$\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36}$$

auf. Wir fassen dieses Resultat in der Liste *wahr* zusammen, und verwenden dazu die Befehle Thread und Range:

```
wahr = Thread[List[Range[2, 12], {1, 2, 3, 4, 5, 6, 5, 4, 3, 2, 1}/36 // N]]
```

```
{{2, 0.0277778}, {3, 0.0555556}, {4, 0.0833333},
 {5, 0.111111}, {6, 0.138889}, {7, 0.166667}, {8, 0.138889},
 {9, 0.111111}, {10, 0.0833333}, {11, 0.0555556}, {12, 0.0277778}}
```

b) Wie in Beispiel 3.1.3 simulieren wir das gleichzeitige Werfen von $k = 2$ Würfeln und die anschließende Ermittlung der jeweils gewürfelten Augensumme n mal, berechnen von diesem Datenmaterial die relativen Häufigkeiten der einzelnen Werte und vergleichen diese relativen Häufigkeiten mit den oben ermittelten theoretischen Wahrscheinlichkeiten:

```
k = 2; n = 10 000;
relh = RelativeTally[Sort[Table[Apply[Plus, RandomInteger[{1, 6}, k]], {n}]]];
Table[Abs[wahr[[i, 2]] - relh[[i, 2]]], {i, 1, 11}]
Clear[k, n, wahr, relh]
```

```
{0.00142222, 0.00305556, 0.00383333, 0.00368889, 0.00208889,
 0.00406667, 0.00681111, 0.00728889, 0.00373333, 0.00195556, 0.000477778}
```

Im Fall $n = 10\,000$ ist die Abweichung der relativen Häufigkeiten von den theoretischen Wahrscheinlichkeiten offenbar stets deutlich kleiner als 0.01.

Schließlich beachte man:

3.2.3 Bemerkung: Wir betrachten ein Zufallsexperiment mit Ereignisraum Ω. Die Abbildung $\mathbb{H}$, welche jedem Ereignis $A \subseteq \Omega$ die (etwa durch Simulation experimentell gewonnene) relative Häufigkeit $\mathbb{H}[A]$ zuordnet, ist offenbar ein W-Maß auf Ω. Ist die Anzahl der Wiederholungen genügend groß, so ist dieses W-Maß $\mathbb{H}$ ein gutes Modell für den dieses Zufallsexperiment steuernden Zufall.

Wir haben damit eine erste Möglichkeit gefunden, mit der sich zu einem vorgegebenen Zufallsexperiment ein passendes mathematisches Modell konstruieren lässt.