2 Wahrscheinlichkeitsrechnung

Erinnerung: Die sogenannte **Potenzmenge** $\wp(\Omega)$ einer Menge Ω besteht aus allen Teilmengen von Ω, also $A \in \wp(\Omega) \iff A \subseteq \Omega$.
Beispiel $\Omega := \{x_1, x_2, x_3\}$. $\wp(\Omega) = \{\emptyset, \{x_1\}, \{x_2\}, \{x_3\}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_3\}, \Omega\}$, hat also 8 Elemente, die jeweils ihrerseits Mengen sind. Wir listen einmal $\wp(\Omega)$ systematisch auf

$x_1 \in A$	$x_2 \in A$	$x_3 \in A$	A
0	0	0	$\emptyset$
0	0	1	$\{x_3\}$
0	1	0	$\{x_2\}$
0	1	1	$\{x_2, x_3\}$
1	0	0	$\{x_1\}$
1	0	1	$\{x_1, x_3\}$
1	1	0	$\{x_1, x_2\}$
1	1	1	Ω

2.1 Kolmogorov-Axiome

Definitionen:
Ein **Zufallsexperiment** ist ein im Prinzip beliebig oft wiederholbarer Vorgang (Versuch) mit einem unbestimmten Ergebnis. Wir fassen alle möglichen Ergebnisse ω zusammen zu einer **Ergebnismenge** Ω . Die Elemente $\omega \in \Omega$ sind also die möglichen Versuchsergebnisse. Jede Teilmenge $A \subseteq \Omega$ heißt dann ein **Ereignis**. Das Ereignis A tritt genau dann ein, wenn das Ergebnis ω zur Menge A gehört, wenn also $\omega \in A$ ist. Damit sind wir auf dem sicheren Boden der Mathematik, und Ereignisse sind nun einfach Mengen. Die möglichen Ereignisse sind also damit jeweils Elemente der Menge $\wp(\Omega)$, der sogenannten Potenzmenge Ω. $\wp(\Omega)$ heißt **Ereignisraum** oder Ereignismenge über Ω. Die einelementigen Teilmengen $\{\omega\} \in \wp(\Omega)$ heißen **Elementarereignisse**. Das Elementarereignis $\{\omega\}$ tritt also genau dann ein, wenn genau das Ergebnis ω vorliegt und kein anderes.

Bemerkung: Bei unendlichen Mengen Ω ist es oft nicht sinnvoll, jede Teilmenge als Ereignis zuzulassen. Der Ereignisraum ist dann nicht die ganze Potenzmenge $\wp(\Omega)$, sondern nur eine Teilmenge $\mathcal{A} \subseteq \wp(\Omega)$, für die man aber noch folgende Eigenschaften fordert:

1) $\Omega \in \mathcal{A}$
2) $A \in \mathcal{A} \Rightarrow \bar{A} \in \mathcal{A}$
   $(\bar{A} = \Omega \backslash A)$
3) $A_1, A_2, A_3, \dots, A_i, \dots \in \mathcal{A} \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$

Ein solches Mengensystem $\mathcal{A}$ nennt man eine $\boldsymbol{\sigma}$ **− Algebra** über Ω. Der Einfachheit halber legen wir aber immer die größtmögliche $\sigma -$ Algebra zugrunde, nämlich ganz $\wp(\Omega)$, was für unsere Zwecke völlig ausreichend ist.

In Zukunft werden wir die Menge $\{1,2,3, \dots, n\}$ abgekürzt schreiben als $[1\colon n]$

$\wp_k(M)$ soll die Menge aller $k -$ elementigen Teilmengen von M sein

Beispiele für Zufallsexperimente, Ereignisse und deren Modellierung durch Mengen

a) Ereignis A: Würfeln einer geraden Augenzahl
   $\Omega = [1\colon 6]$, $A = \{2,4,6\}$. Haben wir also eine 6 gewürfelt ($\omega = 6$), dann ist $\omega \in A$, und das Ereignis ist eingetreten. Haben wir hingegen eine 5 gewürfelt ($\omega = 5$), dann ist $\omega \notin A$, und das Ereignis ist nicht eingetreten.

b) Ereignis A: Würfeln der Augensumme 7 mit zwei Würfeln
$\Omega = [1{:}6] \times [1{:}6] = [1{:}6]^2$, $A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

c) Ereignis A: Eine Ziehung der Lottozahlen („6 aus 49“), bei der alle Gewinnzahlen einstellig sind
$\Omega = \big\{\{1,2,3,4,5,6\},\{1,2,3,4,5,7\},\dots,\{2,6,9,23,31,43\},\dots,\{44,45,46,47,48,49\}\big\} = \wp_6([1{:}49])$
$\mathrm{A} = \big\{\{1,2,3,4,5,6\},\{1,2,3,4,5,7\},\dots,\{2,4,5,6,7,8\},\dots,\{4,5,6,7,8,9\}\big\} = \wp_6([1{:}9])$

d) Ereignis A: Aus einer Urne mit 9 Kugeln – 5 sind weiß, 4 sind schwarz – werden 3 Kugeln **ohne** Zurücklegen gezogen. Dabei erhält man genau eine schwarze Kugel
$\Omega = \wp_3([1{:}9])$, $\mathrm{A} = \big\{\{1,2,3\},\{1,2,5\},\{1,2,7\},\{1,3,4\},\dots,\{7,8,9\}\big\}$

e) Ereignis A: Aus einer Urne mit 9 Kugeln – 5 sind weiß, 4 sind schwarz – werden 3 Kugeln **mit** Zurücklegen gezogen. Dabei erhält man genau eine schwarze Kugel
$\Omega = [1{:}9]^3$, $\mathrm{A} = \{(1,1,2),(1,1,4),\dots,(2,1,1),(2,1,3),\dots,(2,3,3),(2,3,5),\dots,(8,9,9)\}$

An diesen Beispielen sehen wir: Die Menge Ω kann aus Zahlen (Beispiel a)), Zahlenpaaren (Beispiel b)) oder Zahlentripeln (Beispiel e)) bestehen. Die Elemente ω können aber selber auch wieder Mengen sein wie in den Beispielen c) und d).
Und man hat oft mehrere Möglichkeiten, die Ergebnisse und damit die Ergebnismenge Ω zu modellieren. So könnte man z.B. eine Ziehung der Lottozahlen auch durch ein Binärwort der Länge 49 beschreiben, welches an genau 6 Positionen eine 1 hat. Die Ziehung $\{2,6,9,23,31,43\}$ wäre dann das Binärwort 0100010010000000000000100000001000000000001000000, welches genau an den Positionen 2,6,9,23,31,43 eine Eins hat.
In den Beispielen d) und e) haben wir ausgenutzt, dass es in $[1{:}9]$ genau 4 gerade und 5 ungerade Zahlen gibt, sodass wir uns die Kugeln mit geraden Nummern als schwarz und diejenigen mit ungeraden Nummern als weiß denken konnten; denn natürlich ist es egal, welche Kugeln weiß bzw. schwarz sind. Relevant ist nur, **wie viele** weiß bzw. schwarz sind. Man hätte aber anstatt $[1{:}9]$ z.B. auch eine Menge
$K \coloneqq \{w1,w2,w3,w4,w5,s1,s2,s3,s4\}$ kreieren können. In d) wäre dann
$\Omega = \wp_3(\mathrm{K}), \mathrm{A} = \big\{\{\mathrm{w1},\mathrm{w2},\mathrm{s1}\},\{\mathrm{w1},\mathrm{w3},\mathrm{s1}\},\{\mathrm{w1},\mathrm{w4},\mathrm{s1}\},\{\mathrm{w1},\mathrm{w2},\mathrm{s2}\},\dots,\{w4,w5,s4\}\big\}$
Und in e) wäre $\Omega = \mathrm{K}^3$, $\mathrm{A} = \{(w1,w1,s1),(\mathrm{w1},\mathrm{w1},\mathrm{s2}),\dots,(\mathrm{s1},\mathrm{w2},\mathrm{w2}),(s1,w2,w3),\dots,(\mathrm{s4},\mathrm{w5},\mathrm{w5})\}$

Den üblichen Mengenverknüpfungen entsprechend, können wir nun auch Ereignisse verknüpfen. $A \cup B$ bzw. $A \cap B$ werden manchmal auch als Summenereignis $A + B$, bzw. als Produktereignis $A \cdot B$ bezeichnet. Wir werden davon keinen Gebrauch machen. A und B heißen **unvereinbar** oder **disjunkt**, wenn $A \cap B = \emptyset$ ist, wenn es also kein Ergebnis ω gibt, welches zu beiden Ereignissen gehört. Verwechseln Sie nicht Unvereinbarkeit mit Unabhängigkeit (dazu später)! Wichtig ist auch noch das sogenannte **Gegenereignis** (oder **komplementäres Ereignis**) zu A, in Zeichen $\bar{A}$. ($\bar{A} = \Omega \backslash \mathrm{A} = \mathrm{A}' = A^c$). Es tritt genau dann ein, wenn das Ergebnis ω nicht zu A, (sondern eben zu $\bar{A}$) gehört.

Wo wollen uns keine tiefen Gedanken darüber machen, was Wahrscheinlichkeit **ist**, sondern welche **Eigenschaften** eine Wahrscheinlichkeit „vernünftigerweise“ haben soll. Das ist die in der Mathematik so beliebte (und auch erfolgreiche) axiomatische Vorgehensweise. Nachdem wir Wahrscheinlichkeit – was immer das ist – meist in Prozent quantifizieren, sollte sie jedenfalls eine reelle Zahl zwischen Null und Eins sein, welche wir einem Ereignis zuordnen können:

Definition (Kolmogorov „abgespeckt“):
Ein **Wahrscheinlichkeitsmaß** über der Menge Ω ist eine Abbildung $P{:}\,\wp(\Omega) \to \mathbb{R}$ ($P =$ Probability) mit folgenden Eigenschaften: Für alle $A, B \in \wp(\Omega)$ gilt

1) $P(A) \geq 0$ (Nicht-Negativität)
2) $P(\Omega) = 1$ (Normiertheit)
3) $P(\mathrm{A} \cup \mathrm{B}) = P(A) + P(B)$, falls $A \cap B = \emptyset$ ist. (Additivität)

Axiomensysteme legt man üblicherweise so sparsam wie möglich aus, verzichtet also auf die Forderung von Eigenschaften, welche sich aus den wenigen postulierten bereits folgern lassen. Diese Ökonomie bedeutet für uns, dass wir uns Mühe sparen, wenn wir uns überzeugen wollen, ob eine konkret vorliegende Abbildung ein Wahrscheinlichkeitsmaß (mit allen Konsequenzen) ist oder nicht. Wir listen eine kleine Auswahl dieser Konsequenzen auf:

4) $P(\emptyset) = 0$
5) $A \subseteq B \quad \Longrightarrow \quad P(A) \leq P(B)$
6) $P(A) = 1 - P(\bar{A})$
7) $P(B \backslash A) = P(B) - P(B \cap A)$
8) $0 \leq P(A) \leq 1$
9) Sind $A_1, A_2, \dots, A_n$ paarweise disjunkt (d.h. $A_i \cap A_j = \emptyset$ für alle Indexpaare $i \neq j$) so gilt
   $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$
10) Allgemein gilt für alle Mengenfamilien $A_1, A_2, \dots, A_n \in \wp(\Omega)$
   $P(\bigcup_{i=1}^{n} A_i) = \sum_{\emptyset \neq T \subseteq \{1,2,\dots,n\}} (-1)^{|T|+1} P\left(\bigcap_{i \in T} A_i\right)$ mit dem Spezialfall ($n = 2$)
   $P(\mathrm{A} \cup \mathrm{B}) = P(A) + P(B) - P(A \cap B)$

Wir bringen hier keine vollständigen Beweise (, die auch nicht besonders spannend sind). Zum Beispiel sieht man 4) so: Setzen wir $A := \Omega$, $B := \emptyset$ und wenden darauf 2) und 3) an, so ergibt sich $1 = P(\Omega) = \mathrm{P}(\Omega \cup \emptyset) = P(\Omega) + \mathrm{P}(\emptyset) = 1 + \mathrm{P}(\emptyset)$, also $1 = 1 + \mathrm{P}(\emptyset)$, woraus 4) folgt.
Oder 5): Wenn $A \subseteq B$ ist, dann ist $B = A \cup (B \backslash A)$ und $A \cap (B \backslash A) = \emptyset$. Wieder mit 3) folgt
$P(B) = P\big(A \cup (B \backslash A)\big) = P(A) + P(B \backslash A) \geq P(A)$, weil $P(B \backslash A) \geq 0$ ist wegen 1).
9) folgt aus 3) durch vollständige Induktion.
10) beweist man genauso wie die Siebformel aus der Mengenalgebra. Diese folgt ihrerseits auch aus 9), wenn wir das Wahrscheinlichkeitsmaß $P(A) := |A|/|\Omega|$ zugrunde legen. Diese sogenannte Laplace-Wahrscheinlichkeit wird uns gleich noch beschäftigen.

Bemerkung: Der Vollständigkeit halber möchte ich Ihnen die Original-Kolmogorov-Axiome (1933) hier nicht vorenthalten: Dort ist ein Wahrscheinlichkeitsmaß eine Abbildung $P: \mathcal{A} \to \mathbb{R}$ auf einer $\sigma -$ Algebra $\mathcal{A}$, welche nicht-negativ und normiert ist. Und als Additivität verlangt man: Für jede paarweise disjunkte abzählbare Mengenfamilie $A_1, A_2, \dots, A_i, \dots \in \mathcal{A}$ ist $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Beispiele für Wahrscheinlichkeitsmaße

Relative Häufigkeit (statistische Wahrscheinlichkeit)
Führt man ein Zufallsexperiment n mal durch (dabei sei vorausgesetzt, dass die Einzelergebnisse keinen Einfluss aufeinander haben), und tritt dabei k mal das Ereignis A ein, so ist die relative Häufigkeit $P(A) := h_n(A) := \frac{k}{n}$ ein Wahrscheinlichkeitsmaß, d.h. es gelten die Eigenschaften 1)-3) (und damit auch alle daraus abgeleiteten Eigenschaften).
Unterbeispiel: Werfen eines Würfels, also $\Omega = [1:6]$. A = „gerade Augenzahl", also $\mathrm{A} = \{2,4,6\}$, $n = 26$ mit den Ergebnissen 1,2,1,6,5,1,3,2,4,5,3,5,1,6,6,3,5,5,6,1,2,4,3,1,5,5, so ist $h_n(A) = \frac{9}{26} \approx 0{,}346$;
$h_n(\bar{A}) = \frac{17}{26} = 1 - h_n(A)$; $h_n(\{5\}) = \frac{7}{26} \approx 0{,}269$; $h_n(\Omega) = \frac{26}{26} = 1$, usw.
Empirisches Gesetz der großen Zahlen: Für $n \to \infty$ „stabilisiert" sich $h_n(A)$ auf einen Wert $h(A)$, die sog. **empirische Wahrscheinlichkeit** (v. Mises 1931). Dies ist aber keine Konvergenz im strengen Sinne der Analysis.

Im vorigen Beispiel dürfte $h_n(A)$ bei 0,5 liegen (sonst liegt die Vermutung nahe, dass der Würfel „gezinkt" ist).

Bei **endlichen** Ergebnismengen $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ ist ein Wahrscheinlichkeitsmaß P schon eindeutig festgelegt durch seine Werte $p_i := P(\{\omega_i\})$ auf den Elementarereignissen $\{\omega_i\}$, $i = 1, \dots, n$. Denn es gilt der

Satz: Ist $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, dann gibt es zu jedem sog. **Wahrscheinlichkeitsvektor**, das ist ein Vektor $(p_1, p_2, \dots, p_n)$ mit $0 \le p_i \le 1\ \forall i$ und $\sum_{i=1}^n p_i = 1$,
genau ein Wahrscheinlichkeitsmaß P mit $P(\{\omega_i\}) = p_i\ \forall i$, nämlich

$$P(A) = \sum_{\omega_i \in A} p_i$$

Dies folgt aus 9), denn für jedes $A \subseteq \Omega$ gilt

$$A = \bigcup_{\omega_i \in A} \{\omega_i\}$$

mit paarweise disjunkten Elementarereignissen $\{\omega_i\}$.

Dies ist auch der Grundbaustein der Informationstheorie: Eine diskrete stationäre Nachrichtenquelle sendet (mit Wiederholungen) Zeichen aus einem Alphabet (einem geordneten Zeichenvorrat) $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, so dass ein stationärer stochastischer Prozess entsteht, bei dem jedes Zeichen ω_i eine Wahrscheinlichkeit p_i hat. In einem solchen Szenario lässt sich dann beispielsweise präzisieren, was man unter einer optimalen Quellencodierung versteht, und wie man eine solche realisiert.

Bei endlichen Ergebnismengen $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ betrachten wir jetzt zwei wichtige Spezialfälle, nämlich den Fall $\Omega = \{\omega_1, \omega_2\} = \{0,1\}$ und die Gleichverteilung $p_1 = p_2 = \cdots = p_n$

Ein **Bernoulli-Experiment** ist ein Zufallsversuch mit genau zwei Ausgängen, z.B. Treffer und Niete. Die Wahrscheinlichkeit für einen Treffer bezeichnet man mit p, die Wahrscheinlichkeit für eine Niete ist dann $q = 1 - p$.

Ein **Laplace-Experiment** ist ein Zufallsversuch mit endlich vielen möglichen Ergebnissen $\omega_1, \omega_2, \dots, \omega_n$, die alle gleichwahrscheinlich sind, also $p_1 = p_2 = \cdots = p_n$. Es folgt sofort, dass dann alle p_i den Wert $\frac{1}{n}$ haben müssen, und dass $\forall A \subseteq \Omega$ gelten muss $P(A) = \frac{|A|}{|\Omega|}$, was Sie vielleicht aus der Schule kennen:

$$P(A) = \frac{Anzahl\ aller\ f\ddot{u}r\ A\ g\ddot{u}nstigen\ F\ddot{a}lle}{Anzahl\ aller\ m\ddot{o}glichen\ F\ddot{a}lle}$$

⚠ Diese Abzählregel kann zu falschen Ergebnissen führen, wenn man ignoriert, dass sie nur unter der Prämisse funktioniert, dass alle Elementarereignisse $\{\omega_i\}$ (alle Ergebnisse ω_i) **gleichwahrscheinlich** sind.
Der französische Mathematiker D'Alembert (1717-1783) nahm an, dass beim zweimaligen Werfen einer („symmetrischen") Münze das Ereignis A: „Werfen von verschiedenen Seiten" die Wahrscheinlichkeit $\frac{1}{3}$ hat. Zu diesem Ergebnis kommt man, indem man beispielsweise $\Omega := \{(2,0), (1,1), (0,2)\}$ festlegt. Dabei sagt die erste Komponente eines Zahlenpaares jeweils, wie oft „Kopf", die zweite Komponente, wie oft „Zahl" geworfen wurde. Dann ist $A = \{(1,1)\}$ und $P(A) = \frac{|A|}{|\Omega|} = \frac{1}{3}$. Nimmt man aber das Modell $\Omega := \{\mathrm{KK}, \mathrm{KZ}, \mathrm{ZK}, \mathrm{ZZ}\}$, dann ist $A = \{KZ, ZK\}$, also $P(A) = \frac{|A|}{|\Omega|} = \frac{2}{4} = 0{,}5$. Es ist nicht bekannt, ob D'Alembert eine längere Versuchsserie durchgeführt hat. Er hätte dann „wahrscheinlich" eine relative Häufigkeit dieses Ereignisses von etwa 0,5 beobachten können. Im zweiten Modell wurde die Reihenfolge der Münzwürfe berücksichtigt. Nicht, weil es auf die Reihenfolge ankäme (das tut es nämlich nicht), sondern um zu gleichwahrscheinlichen Versuchsergebnissen zu kommen: Das Werfen von verschiedenen Seiten kann dann auf zwei Arten zustande kommen, was eine höhere Wahrscheinlichkeit plausibel macht.

Aber nochmal: Gleichwahrscheinlichkeit ist eine nur empirisch gestützte Annahme. Dass beim Würfeln (mit einem „idealen" Würfel) jede Zahl mit einer Wahrscheinlichkeit $\frac{1}{6}$ auftritt, ist nicht eine Frage der Mathematik. Wahrscheinlichkeits**rechnung** heißt eben nur, dass wir aus gegebenen (angenommenen) Wahrscheinlichkeiten mit Hilfe von (durch Axiome festgelegten) „Spielregeln" weitere Wahrscheinlichkeiten berechnen.
Geht man von einer Gleichverteilung der Wahrscheinlichkeiten p_i aus, also von einem Laplace-Experiment, so läuft das Bestimmen der Wahrscheinlichkeit eines Ereignisses letztlich auf das Abzählen von endlichen Mengen, also auf reine **Kombinatorik**, hinaus.

Umgekehrt können wir uns das empirische Gesetz der großen Zahlen bei rein kombinatorischen Fragestellungen zunutze machen: So können wir unsere Ergebnisse durch Computer-Simulation eines geeigneten Laplace-

Experiments (z.B. mit Hilfe eines Zufallszahlen-Generators) verifizieren (überprüfen). Das ist dann natürlich noch kein Beweis, kann aber evtl. auf einen möglichen „Denkfehler“ aufdecken.
Beispiel: Wie viele injektiven Abbildungen $f\colon [1:6] \to [1:100]$ gibt es? Antwort (siehe Kapitel „Kombinatorik) $100 \cdot 99 \cdot 98 \cdot 97 \cdot 96 \cdot 95 = 858277728000$. „Für den, der es nicht glaubt“: Man könnte das Zufallsexperiment „Wähle 6 mal hintereinander eine Zahl zwischen 1 und 100 (einschließlich)“ simulieren. Hat man dies „unendlich oft“ von einem Rechner machen lassen, dann müsste nach der Theorie die relative Häufigkeit dafür, dass 6 unterschiedliche Zahlen gewählt wurden, $\frac{858277728000}{100^6} \approx 0{,}858277728$ sein.