Kapitel 1

Statistische und klassische Definition von „Wahrscheinlichkeit"

1.1 Klassische Definition

Die erste quantitative Definition des Begriffes „Wahrscheinlichkeit" geht auf Blaise Pascal (1623 - 1662) und Pierre de Fermat (1601 - 1665) zurück. Sie wurden von Antoine Gombauld Chevalier de Méré, Sieur de Baussay (1607-1685) ermahnt, dass „die Mathematik nicht im Einklang mit dem praktischen Leben sei". Das „praktische Leben" war für den Edelmann das Glücksspiel. Ihn interessierte besonders ein damals übliches Würfelspiel, bei dem die Bank gewinnt, wenn bei viermaligem Würfeln mindestens einmal die Augenzahl 6 erscheint. Pascal und Fermat haben sich mit dieser und ähnlichen Fragen auseinandergesetzt und damit die klassische Wahrscheinlichkeitstheorie begründet.

Def. 1.1 (Klassische Definition von Wahrscheinlichkeit) *Ein Ereignis trete zufällig auf. Die Wahrscheinlichkeit für das Auftreten des Ereignisses E ist definiert durch den Quotienten aus der Anzahl g der für das Ereignis günstigen und der Zahl m der möglichen Fälle*

$$P = \frac{g}{m} \qquad .$$

Beispiel: Wahrscheinlichkeit für Herz in einem Skat-Spiel ist $P = \frac{8}{32} = \frac{1}{4}$.

Für die klassische Wahrscheinlichkeit lassen sich für zwei beliebige Ereignisse A und B folgende Regeln ableiten

$$P(A \vee B) = \frac{n_A + n_B - n_{A\wedge B}}{N} = P(A) + P(B) - P(A \wedge B) \tag{1.1a}$$

$$P(N) = \frac{0}{N} = 0 \qquad \text{unmögliches Ereignis, N: Null-Element} \tag{1.1b}$$

$$P(E) = \frac{N}{N} = 1 \qquad \text{sicheres Ereignis, E: Eins-Element} \tag{1.1c}$$

$$0 \leq P(A) \leq 1 \qquad \text{folgt aus der Definition} \tag{1.1d}$$

$$P(A|B) = \frac{n_{A\wedge B}}{n_B} = \frac{P(A \wedge B)}{P(B)} \qquad . \tag{1.1e}$$

Das Symbol $P(A|B)$ steht für die BEDINGTE WAHRSCHEINLICHKEIT für A, vorausgesetzt das Ereignis B liegt vor. In der klassischen Wahrscheinlichkeitstheorie heißt das, dass wir die möglichen Ereignisse vorsortieren. Wir berücksichtigen nur noch solche, die die Bedingung B erfüllen, das seien n_B Ereignisse. Von diesen interessieren uns nun gerade solche, die auch gleichzeitig A erfüllen. Damit ist die Zahl der günstigen Fälle gleich der Zahl der Fälle, die sowohl A als auch B erfüllen, also $n_{A\wedge B}$.

Def. 1.2 (Exklusive Ereignisse) *Sich gegenseitig ausschließende Ereignisse, d.h. $A \wedge B = N$, wollen wir in Anlehnung an die englische Nomenklatur kurz* EXKLUSIV *nennen.*

Def. 1.3 (Komplementäre Ereignisse) *Ein Ereignis $\overline{A}$ ist komplementär zu A, wenn gilt*

$$\overline{A} \vee A = E; \qquad \textit{und} \qquad \overline{A} \wedge A = N \qquad .$$

Aus der Summenregel folgt für exklusive Ereignisse eine vereinfachte Summenregel und daraus wiederum der Zusammenhang zwischen den Wahrscheinlichkeitenkomplementärer Ereignisse

$$P(A \vee B) = P(A) + P(B) \tag{1.2}$$

$$P(\overline{A}) = 1 - P(A) \tag{1.3}$$

Diese klassische Definition der Wahrscheinlichkeit wurde später von Jacob Bernoulli (1654-1705) in seinem Buch Ars Conjectandi (1713 posthum publiziert) weiterentwickelt. Dieses Buch enthält wegweisende Beiträge zur Wahrscheinlichkeitstheorie, unter anderem eine ausführliche Diskussion der wahren Bedeutung des Wahrscheinlichkeitsbegriffes: *Wahrscheinlichkeit ist ein messbares Maß der Gewissheit. Bernoulli unterschied bereits klar zwischen Prior- und Posterior-Wahrscheinlichkeit*. Neben Pascal, Fermat und Bernoulli war Pierre-Simon Laplace (1749-1827) einer der führenden Begründer der modernen Wahrscheinlichkeitstheorie. Laplace benutzte die Wahrscheinlichkeitstheorie u.a. für inverse Schlüsse (z.B. die Straße ist nass, wie groß ist die Wahrscheinlichkeit, dass es geregnet hat?). Die Formel, die er hierzu verwendet hat, geht auf Reverend Thomas Bayes (1702-1761) zurück. Die nach ihm benannte Formel (Bayes Theorem, sogenannt nach Pointcaré viele Jahre später) wurde erst posthum (1764) publiziert. Eingereicht wurde sie von einem seiner Freunde (Richard Price). Die Formel war allerdings in einer wenig transparenten Form und wurde erst durch Laplace in die heute bekannte Gestalt gebracht. Im Zusammenhang mit der Himmelsmechanik beruhen wesentliche seiner Ergebnisse auf Wahrscheinlichkeitsüberlegungen. Weitere Anwendungen seiner Wahrscheinlichkeitstheorie sind: Buffonsches Nadel-Problem, Glücksspiele, naturwissenschaftliche Probleme, juristische Probleme, Kindersterblichkeit, etc. Für die Ableitung seiner Ergebnisse führte er das Prinzip der Ununterscheidbarkeit (principle of indifference) ein. Auf Laplace gehen auch die Anfänge der Fehlerrechnung zurück.

Es war bereits Bernoulli bekannt, dass die Zahlen m aller Ereignisse und die der Zahl g der günstigen Ereignisse nicht immer eindeutig festgelegt werden können. Zum Beispiel: Es werden zwei Würfel geworfen und wir suchen die Wahrscheinlichkeit, dass die Summe der Augenzahlen 7 ist.

a) Wir betrachten als mögliche Ergebnisse die 11 möglichen Summen der Augenzahlen $(2, 3, \ldots, 12)$. Von diesen Summen ist nur ein Ergebnis (die Summe 7) das positive Ergebnis $\Rightarrow P = 1/11$.

b) Wir betrachten als mögliche Ergebnisse alle <u>der Größe nach sortierten</u> Zahlen-Paare, ohne zwischen dem ersten und zweiten Würfel zu unterscheiden, $(1, 1), (1, 2), \ldots (1, 6),\ (2, 2), (2, 3), \ldots (6, 6)$. Davon gibt es $6+5+4+3+2+1 = 21$. Es gibt 3 positive Ereignisse $(1, 6), (2, 5), (3, 4)$. Daraus folgt $P = 1/7$.

c) Wir unterscheiden zusätzlich, welcher Würfel welche Augenzahl liefert. Der Unterschied besteht darin, dass es nun 2 Möglichkeiten für ungleiche Augenzahlen und weiterhin nur eine für gleiche Augenzahlen gibt. Nun ist $N = 36$ und $g = 6$, mit $(1, 6), (6, 1), (2, 5), (5, 2), (3, 4), (4, 3)$. Das heißt, $P = 1/6$.

In diesem Beispiel ist offensichtlich die letzte Variante die korrekte. Man ist überzeugt, dass das a-priori so sein muss. Die Quanten-Physik lehrt uns aber etwas anderes. Es gibt hier Situationen, in denen die zweite Variante korrekt ist (Bosonen). Hierauf gehen wir später näher ein. Das obige Beispiel zeigt, dass die Definition der Wahrscheinlichkeit präzisiert werden muss:

Def. 1.4 (Präzisierte Definition der klassischen Wahrscheinlichkeit) *Die Wahrscheinlichkeit eines Ereignisses ist gegeben durch das Verhältnis der Zahl der günstigen Ergebnisse zu der aller möglichen; vorausgesetzt, alle Ergebnisse sind gleich-wahrscheinlich.*

Diese Definition eliminiert aber nicht wirklich die Probleme:

- Konzeptionell unbefriedigend, Wahrscheinlichkeit darüber definiert, dass Ereignisse dieselbe Wahrscheinlichkeit haben (Ringschluss).

- Das obige Beispiel hat gezeigt, dass die Gleich-Wahrscheinlichkeit nicht eindeutig ist.

- Kann nur in den Fällen angewandt werden, in denen alle Elementar-Ereignisse dieselbe Wahrscheinlichkeit haben. Kann in vielen Fällen leicht geklärt werden. (Glücksspiel, auch in der statistischen Physik)

Es wurden zur Zuweisung von Wahrscheinlichkeiten zwei Prinzipien vorgeschlagen: das „Principle of Insufficient Reasoning" von Jacob Bernoulli und das „Principle of Indifference" von Pierre-Simon Laplace. Sie besagen: wenn es kein Vorwissen gibt, das die einzelnen Elementar-Ereignisse unterscheidet, kann man sie neu durchnummerieren, ohne die Situation zu verändern. Daraus muss man schließen, dass alle Ereignisse, zwischen denen a-priori nicht unterschieden werden kann, „gleichwahrscheinlich" sind. Es gibt aber Gegenbeispiele für diskrete Probleme aus der Physik. Besonders augenfällig sind die Schwierigkeiten bei kontinuierlichen Problemen, da hier das Maß in's Spiel kommt. Ein prominentes Beispiel ist das sogenannte BERTRAND-PARADOXON (1888), das erst 1968 von E.T.Jaynes (1922-1998) zufriedenstellend gelöst wurde.

1.2 Bertrand Paradoxon

Über einen Kreis hinweg werden zufällig Geraden gezeichnet. Wie groß ist die Wahrscheinlichkeit, dass der Abstand vom Zentrum kleiner ist als die Hälfte des Radius r? Ohne Beschränkung der Allgemeinheit setzen wir $r = 1$ in passend gewählten Einheiten. Es bieten sich mehrere Lösungen an. Wir wollen hier zunächst nur drei diskutieren.

a) Wir nehmen an, dass – wie in Abbildung (1.1) dargestellt – der Abstand vom Zentrum mit gleicher Wahrscheinlichkeit alle Werte zwischen 0 und 1 annehmen kann. Unter dieser Voraussetzung ist das Intervall der günstigen Ereignisse $1/2$ und somit die gesuchte Wahrscheinlichkeit $P = 1/2$.

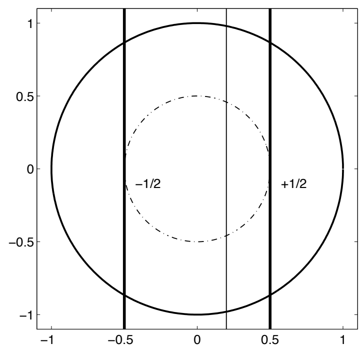

Abbildung 1.1: *Skizze 1 zum Bertrand-Paradoxon.*

b) Alternativ können wir, wie in Abbildung (1.2) skizziert, den Winkel zwischen der Geraden und der Tangente an den Kreis als Größe betrachten, die gleich-wahrscheinlich alle Werte zwischen 0 und π annehmen kann. Der „günstige" Winkelbereich ist $\pi/3$ und die gesuchte Wahrscheinlichkeit ist demnach in diesem Fall $P = 1/3$.

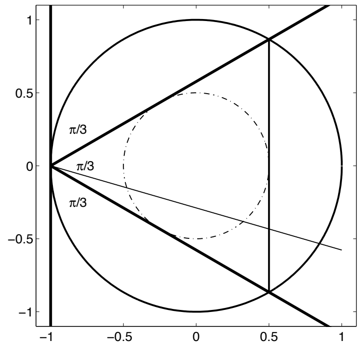

Abbildung 1.2: *Skizze 2 zum Bertrand-Paradoxon.*

c) Eine weitere Möglichkeit ist, die Fläche A des konzentrischen Kreises, der die Gerade – wie in Abbildung (1.3) – berührt, als gleich-verteilt zwischen 0 und π anzunehmen. Der günstige Bereich ist $A \in [0, \pi/4]$, woraus sich eine Wahrscheinlichkeit $P = 1/4$ ergibt.

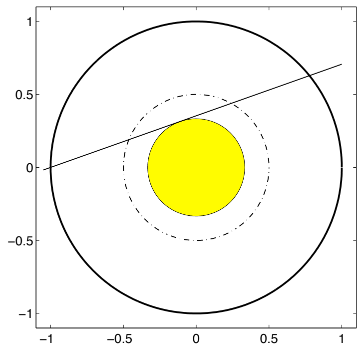

Abbildung 1.3: *Skizze 3 zum Bertrand-Paradoxon.*

Es gibt noch andere plausibel erscheinende Darstellungen, die zu wieder anderen Ergebnissen führen. Wie kann es sein, dass ein eindeutig definiertes Problem keine eindeutige Antwort liefert? Hinter dem Bertrand-Paradoxon verbirgt sich allgemein das Problem, wie man Unwissenheit bei kontinuierlichen Freiheitsgraden zu beschreiben hat. Nehmen wir an, wir wollen etwas über eine Größe x aussagen, die die reellen Zahlen zwischen 0 und 1 annehmen kann. Es ist naheliegend das klassische Konzept zu verwenden, dass, wenn wir nichts weiter wissen, die Wahrscheinlichkeit für

$x \in (x, x + dx)$ als

$$P(x \in (x, x + dx)) = \frac{dx}{1} = p_x(x)dx$$

anzusetzen ist. Hierbei ist $p_x(x) = 1$ die Wahrscheinlichkeitsdichte, die später genauer behandelt wird. Unwissenheit wird hier also durch eine konstante Wahrscheinlichkeitsdichte beschrieben. Die Wahrscheinlichkeitsdichte für x^n ist dann aber nicht mehr konstant sondern, wie wir später zeigen werden,

$$p_z(z = x^n) = p_x(x) \left| \frac{dx}{dz} \right| = \frac{1}{n} z^{\frac{1}{n}-1} \qquad .$$

Die Wahrscheinlichkeitsdichte für die Variable z ist demnach scharf bei $z = 0$ gepeakt. Das heißt unsere flache Verteilung in x impliziert, dass es eine sehr große Wahrscheinlichkeit für kleine Werte von z gibt.
Ganz allgemein erkennen wir, dass nichtlineare Transformationen offensichtlich Probleme machen. In welcher Darstellung sind die Ereignisse gleicher Größe (Länge, Fläche, ...) gleich-wahrscheinlich, und wie beschreibt man völliges Unwissen? Auf das Problem PRIOR-WAHRSCHEINLICHKEITEN zuzuweisen kommen wir im Teil III zu sprechen. Hierzu gibt es einen Zugang über Transformationsgruppen[1] und einen differentialgeometrischen, welche beide äquivalent sind. In diesem Zusammenhang werden wir auch das wichtige Prinzip der maximalen Entropie behandeln.
Die klassische Definition der Wahrscheinlichkeit war trotz dieser Probleme mehre Jahrhunderte in Gebrauch und wird auch heute noch auf viele Probleme angewandt; insbesondere solche kombinatorischer Natur.
Bereits Bernoulli hatte sich die Frage gestellt, wie die Wahrscheinlichkeit mit der relativen Häufigkeit zusammenhängt. Beispiel: Die Wahrscheinlichkeit für eine gerade Augenzahl beim Würfeln ist 1/2. Wenn man, das Würfel-Experiment N-mal wiederholt, wie häufig wird man dann eine gerade Augenzahl vorfinden. Wir werden zeigen, dass diese relative Häufigkeit im Limes $N \to \infty$ gegen die „intrinsische" Wahrscheinlichkeit konvergiert. Bernoulli hat aber eigentlich das umgekehrte Problem beschäftigt. Gegeben eine „Stichprobe" vom endlichen Umfang N, was kann man dann über die zugrundeliegende (intrinsische) Wahrscheinlichkeit aussagen?

1.3 Statistische Definition

Um die Prior-Probleme der klassischen Wahrscheinlichkeitstheorie zu vermeiden, verfolgten Ellis (1842), Boole (1854), Venn (1866) und von Mises (1928) eine andere Definition. Da, wie Bernoulli gezeigt hatte, die relative Häufigkeit im Limes $N \to \infty$ gegen die „intrinsische" Wahrscheinlichkeit konvergiert, führten sie folgende Definition der Wahrscheinlichkeit ein

[1] Dieser Zugang wurde von E.T. Jaynes verwendet, um das Bertand-Paradoxon zu klären.

Def. 1.5 (Statistische Definition von Wahrscheinlichkeit) *Ein „Ereignis" A trete zufällig auf. Die Wahrscheinlichkeit ist definiert durch die relative Häufigkeit*

$$P(A) = \lim_{N\to\infty} \frac{n}{N} \qquad , \tag{1.4}$$

dass das Ereignis bei N Versuchen n-mal auftritt, im Limes $N \to \infty$.

Diese Definition soll vermeiden, dass man je in die Verlegenheit kommt, „Prior-Wahrscheinlichkeit" angeben zu müssen, sondern die Wahrscheinlichkeit eine dem zu untersuchenden Objekt anhaftende intrinsische Eigenschaft ist, die nur experimentell bestimmt werden kann, nämlich indem man ein Stichprobe von unendlichem Umfang untersucht. Da eine unendlich Stichprobe nie untersucht werden kann, ist diese Definition von rein hypothetischem Charakter. Diese Definition des Wahrscheinlichkeitsbegriffes ist allerdings weit verbreitet und bildet die Grundlage der „orthodoxen Statistik".

Nachteile des statistischen Wahrscheinlichkeitsbegriffes

- Für viele Probleme gibt es keine Häufigkeitsverteilung
  - War Herr X der Täter?
  - Liegt die Temperatur des Fusionsplasmas $T \in (1.0, 2.0)10^8 K$?
  - War Julius Cäsar Linkshänder?
- Nur in den wenigsten Fällen ist $N \gg 1$
  - Großexperimente, z.B. Fusionsforschung (Kostenfrage)
  - Extraterrestische Physik, Stichprobenumfang von der Natur vorgegeben.
- der Limes $N \to \infty$ ist in der Praxis nicht möglich, und Gl. (1.4) ist keine Definition für $P(A)$ sondern muss als Hypothese gesehen werden. Es handelt sich nicht um eine experimentell bestimmbare Größe.
- Interpretationsprobleme
  - Was bedeutet die Aussage, die Wahrscheinlichkeit ist 0.1, dass Herr X einen Virus hat?
    (100 mal klonen, 10 Klone haben dann den Virus?!)
  - Die Wahrscheinlichkeit im Lotto zu gewinnen ist $7 \cdot 10^{-8}$
    (10^8 mal einzahlen, um einmal zu gewinnen?!)

Man erkennt an den beiden Beispielen auch, wie unterschiedlich Wahrscheinlichkeiten, je nach Kontext, bewertet werden. Im ersten Fall wird man geneigt sein, die Möglichkeit des Virus nahezu auszuschließen, während der Zuspruch des Lottospiels zeigt, dass man ein Wahrscheinlichkeit von $7 \cdot 10^{-8}$ als hinreichend groß bewertet. Wie Wahrscheinlichkeiten tatsächlich zu bewerten sind, besagt die ENTSCHEIDUNGSTHEORIE. Hierbei wird zusätzlich eine Bewertungsfunktion (Kostenfunktion) einfließen, die den unterschiedlichen Ausgängen eines „Experiments" Werte (Kosten) zuweist.

Mit der statistischen Definition kann nur eine stark eingeschränkte Klasse von Problemen behandelt werden, nämlich solche, bei denen ein „Versuch" zumindest theoretisch beliebig oft wiederholt werden kann.
Die statistische Definition der Wahrscheinlichkeit führt allerdings zu denselben Rechenregeln für Wahrscheinlichkeiten, wie die klassische Definition.

Wir haben bei der obigen Diskussion gesehen, dass einige Begriffe und Ideen im Zusammenhang mit der Wahrscheinlichkeitstheorie auftreten, die genauer definiert bzw. analysiert und genauer hinterfragt werden müssen. Hierauf werden wir in einem späteren Kapitel eingehen. Zunächst benötigen wir ein paar wichtige Begriffe zur Charakterisierung von Wahrscheinlichkeitsverteilungen, die im nächsten Kapitel dargestellt werden.