

Einführung

1.1 Ein Beispiel

Wirft man einen Würfel bis zur ersten Sechs, so kann man die Wahrscheinlichkeit, dass dies gerade beim n -ten Wurf passiert, berechnen, indem man die Menge $\Omega_n := \{1, \dots, 6\}^n$ aller n -Tupel betrachtet, die man mit den Augenzahlen $1, \dots, 6$ bilden kann. Ω_n besteht aus $|\Omega_n| = 6^n$ Elementen und bei einem fairen Würfel sollte jedes n -Tupel gleich wahrscheinlich sein. Die erste Sechs erscheint gerade dann beim n -ten Wurf, wenn das n -Tupel der Wurfresultate in $A_n := \{(x_1, \dots, x_{n-1}, 6) : x_i \in \{1, \dots, 5\} \ \forall i = 1, \dots, n-1\}$ liegt. Wegen $|A_n| = 5^{n-1}$ folgt dann aus der klassischen Wahrscheinlichkeitsdefinition

$$P(A_n) = \left(\frac{\text{günstige Fälle}}{\text{mögliche Fälle}} \right) = \frac{5^{n-1}}{6^n}.$$

Um die Wahrscheinlichkeiten der einzelnen Ausgänge zu bestimmen, haben wir für jedes n einen anderen Wahrscheinlichkeitsraum Ω_n verwendet.

Man kann dies nur umgehen, wenn man als Raum der Versuchsausgänge die Menge $\Omega := \{(x_1, x_2, \dots) : x_i \in \{1, \dots, 6\} \ \forall i \in \mathbb{N}\}$ aller Folgen, die mit den Zahlen $1, \dots, 6$ gebildet werden können, betrachtet.

Ersetzt man in diesen Folgen jede Sechs durch eine Null, so kann man die entsprechende Folge (x_1, x_2, \dots) interpretieren als Zahl $x := \sum_{i=1}^{\infty} x_i 6^{-i}$, angeschrieben im 6-adischen Zahlensystem. Bei Zahlen der Form $\sum_{i=1}^n x_i 6^{-i}$ mit $x_n \neq 0$, die auch periodisch als $\sum_{i=1}^{n-1} x_i 6^{-i} + (x_n - 1) 6^{-n} + 5 \sum_{i=n+1}^{\infty} 6^{-i}$ angeschrieben werden können, wollen wir immer die endliche Form verwenden. Dadurch entspricht jeder Zahl aus $[0, 1)$ eine eindeutige Folge.

Wir werden etwas später sehen, dass es praktisch keine Rolle spielt, wenn wir damit den Folgen $(x_1, \dots, x_n, 5, 5, \dots)$, $x_n < 5$ keine Zahl zuordnen kön-

nen. Aber auf Grund der obigen Ausführungen ist klar, dass unser Raum Ω überabzählbar sein muss.

Wir haben angenommen, dass jedes konkrete n -Tupel $(x_1, \dots, x_n) \in \Omega_n$ mit der gleichen Wahrscheinlichkeit $P((x_1, \dots, x_n)) := 6^{-n}$ auftreten kann. Die Menge aller Folgen, deren erste n Würfe durch das n -Tupel (x_1, \dots, x_n) festgelegt sind, bezeichnen wir mit $A(x_1, \dots, x_n)$, d.h.

$$A(x_1, \dots, x_n) := \{(x_1, \dots, x_n, x_{n+1}, \dots) : x_{n+i} \in \{0, \dots, 5\} \quad \forall i \in \mathbb{N}\}.$$

Der Folge $(x_1, \dots, x_n, 0, \dots)$ entspricht die Zahl $x := \sum_{i=1}^n x_i \cdot 6^{-i}$ und der Folge $(x_1, \dots, x_n, 5, \dots)$ ist die Zahl $x + 6^{-n}$ zugeordnet. Da wir keine periodischen Darstellungen der Form $(x_1, \dots, x_n, 5, \dots)$ zulassen, entsprechen den Folgen aus $A(x_1, \dots, x_n)$ die Zahlen aus dem Intervall $[x, x + 6^{-n})$, und die Länge dieses Intervalls ist gerade die Wahrscheinlichkeit von $A(x_1, \dots, x_n)$, d.h. $P(A(x_1, \dots, x_n)) = \frac{1}{6^n}$.

Von einem sinnvollen Wahrscheinlichkeitsbegriff wird man verlangen, dass keine Untermenge wahrscheinlicher als eine sie enthaltende Obermenge sein sollte. Man nennt das die Monotonie der Wahrscheinlichkeit.

Da für jede Folge (x_1, x_2, \dots) gilt $(x_1, x_2, \dots) \in A(x_1, \dots, x_n) \quad \forall n \in \mathbb{N}$, muss daraus folgen $P(\{(x_1, x_2, \dots)\}) \leq \frac{1}{6^n} \quad \forall n \in \mathbb{N}$, d.h. jede Folge hat Wahrscheinlichkeit $P((x_1, x_2, \dots)) = 0$. Damit ist klar, dass die Wahrscheinlichkeitsverteilung P nicht durch die Wahrscheinlichkeiten der einzelnen Punkte von Ω festgelegt werden kann. Außerdem kann man überabzählbar viele Terme nicht aufsummieren, d.h. eine Summe der Gestalt

$$\sum_{(x_1, x_2, \dots) \in A(x_1, x_2, \dots, x_n)} P((x_1, x_2, \dots))$$

ergibt keinen Sinn.

Die Menge der Folgen $(x_1, \dots, x_n, 5, 5, \dots)$, $x_n < 5$, $n \in \mathbb{N}$ ist abzählbar. Daher kann man die Summe der Wahrscheinlichkeiten der einzelnen Punkte dieser Menge bilden und erhält Wahrscheinlichkeit 0, was durchaus unserer Intuition entspricht, denn man wird es für ausgeschlossen halten, dass bei einem fairen Würfel ab einem bestimmten Zeitpunkt nur mehr Fünfen geworfen werden. Somit ist es praktisch irrelevant sich mit dieser Menge zu beschäftigen.

Ist nun $[a, b)$ ein beliebiges Teilintervall von $[0, 1)$ mit $a = \sum_{i=1}^{\infty} a_i 6^{-i}$ und $b = \sum_{i=1}^{\infty} b_i 6^{-i}$, und bezeichnet man die auf n Stellen abgerundeten Werte von

a und b mit \hat{a}_n bzw. \hat{b}_n (d.h. $\hat{a}_n = \sum_{i=1}^n a_i 6^{-i}$ bzw. $\hat{b}_n = \sum_{i=1}^n b_i 6^{-i}$), so bilden

die Intervalle $[\hat{a}_n, \hat{a}_n + 6^{-n})$, $[\hat{a}_n + 6^{-n}, \hat{a}_n + 2 \cdot 6^{-n})$, \dots , $[\hat{b}_n, \hat{b}_n + 6^{-n})$ eine disjunkte Überdeckung des Intervalls $[a, b)$, deren Wahrscheinlichkeit der Summe $\hat{b}_n + 6^{-n} - \hat{a}_n$ der Längen der Teilintervalle entspricht. Ohne die beiden Randintervalle $[\hat{a}_n, \hat{a}_n + 6^{-n})$, $[\hat{b}_n, \hat{b}_n + 6^{-n})$ reduziert sich die Gesamtlänge der Vereinigung der verbleibenden Intervalle auf $\hat{b}_n - \hat{a}_n - 6^{-n}$ und diese Vereinigung liegt nun zur Gänze in $[a, b)$. Wegen der Monotonie der Wahrscheinlich-

keitsverteilung sollte daher gelten $\hat{b}_n - \hat{a}_n - 6^{-n} \leq P([a, b)) \leq \hat{b}_n - \hat{a}_n + 6^{-n}$.
Daraus folgt wegen $\lim_{n \rightarrow \infty} \hat{a}_n = a$, $\lim_{n \rightarrow \infty} \hat{b}_n = b$ und $\lim_{n \rightarrow \infty} 6^{-n} = 0$

$$P([a, b)) = b - a.$$

Diese Verteilung, die jedem Teilintervall $[a, b) \subseteq [0, 1)$, $a \leq b$ seine Länge $b - a$ als Wahrscheinlichkeit zuordnet, wird stetige Gleichverteilung auf $[0, 1)$ genannt. Der Name rührt daher, dass jedes Teilintervall mit einer gegebenen Länge dieselbe Wahrscheinlichkeit besitzt, unabhängig von seiner Lage in $[0, 1)$. Man sagt auch, die stetige Gleichverteilung ist translationsinvariant.

Wir zeigen nun, dass es unmöglich ist, durch P jeder Teilmenge von $[0, 1)$ eine Wahrscheinlichkeit zuzuordnen, wenn man fordert, dass man die Wahrscheinlichkeiten abzählbar vieler disjunkter Mengen aufsummieren darf, und, wenn man die Forderung der Translationsinvarianz aufrecht erhalten möchte.

Mit den Bezeichnungen $[x] := \max\{z \in \mathbb{Z} : z \leq x\}$, $x \bmod 1 := x - [x]$ ist $x \sim y \Leftrightarrow (x - y) \bmod 1 \in \mathbb{Q} \cap [0, 1)$ eine Äquivalenzrelation. und bestimmt daher eine Klassenzerlegung von $[0, 1)$.

Man nimmt nun aus jeder Klasse genau ein Element und bildet damit eine Menge A (das Auswahlaxiom A.2 besagt, dass dies möglich ist). Somit gilt $x \neq y$, $x, y \in A \Rightarrow (x - y) \bmod 1 \notin \mathbb{Q}$.

Ist $A + x := \{y = (a + x) \bmod 1 : a \in A\}$, dann bilden die $\{A + q : q \in \mathbb{Q}\}$ eine disjunkte Zerlegung von $[0, 1)$, denn für $q_1 \neq q_2$, $q_i \in \mathbb{Q}$ gilt klarerweise $A + q_1 \cap A + q_2 = \emptyset$, und für jedes $x \in [0, 1)$ gibt es ein $y \in A$, sodass $x \sim y \Rightarrow \exists q : x - y \bmod 1 = q \in \mathbb{Q} \Rightarrow x \in A + q$. Also $[0, 1) = \bigcup_{q \in \mathbb{Q} \cap [0, 1)} A + q$.

Die Translationsinvarianz bedeutet $P(A + q) = P(A) \quad \forall q \in \mathbb{Q}$.

Darf man nun die Wahrscheinlichkeiten der $A + q$ aufsummieren, so gilt

$$P([0, 1)) = \begin{cases} 0 & , \text{ wenn } P(A) = 0 \\ \infty & , \text{ wenn } P(A) > 0. \end{cases}$$

Das widerspricht $P([0, 1)) = 1$, womit unsere Behauptung bewiesen ist.

Wir müssen also für die stetige Gleichverteilung einen kleineren Definitionsbereich als die Potenzmenge von $[0, 1)$ suchen.



<http://www.springer.com/978-3-642-45386-1>

Maß- und Wahrscheinlichkeitstheorie

Eine Einführung

Kusolitsch, N.

2014, XI, 353 S., Softcover

ISBN: 978-3-642-45386-1