1.2 Das Axiomensystem von Kolmogoroff

Beschreibung eines Zufallsvorganges:

a) was kann alles passieren?

b) mit welcher Wahrscheinlichkeit?

zu a) Festlegen eines *Ergebnisraums* (Grundraum, Stichprobenraum) Ω, der die Menge aller möglichen *Ergebnisse* ω enthält.

Teilmengen A von Ω bezeichnet man als *Ereignisse*,

, einelementige Teilmengen als *Elementarereignisse*,

.

Ereignisse sind also bestimmte *Mengen von Ergebnissen*.

zu b) Eine Wahrscheinlichkeitsbewertung ordnet jedem Ereignis seine Wahrscheinlichkeit(skomponente) zu. Dabei sollen gewisse „fundamentale Rechenregeln“ gelten:

Def 1.1 *Axiome von Kolmogorov* (1933)
Eine Funktion P, die Ereignissen aus Ω (reelle) Zahlen zuordnet, heißt *Wahrscheinlichkeit*, wenn gilt

$(K1)$	$P(A)$	für alle Ereignisse $A \subset \Omega$	(1.2.1)
$(K2)$	$P(\Omega)$		(1.2.2)
$(K3)$		dann ist $P(A \cup B) =$.	(1.2.3)

(P steht für Probability.)

Anmerkungen

a) Dies ist eine reine Definition, die sich zunächst im „luftleeren“ Raum bewegt. Es wird rein formal festgelegt, was eine Wahrscheinlichkeit sein soll.
Es gab/gibt (wie in Kap. 1.1 ausgeführt) verschiedene Versuche, Wahrscheinlichkeiten operational zu definieren (also durch eine Messvorschrift) und verschiedene *Interpretationen*, die die Axiomatik mit Leben füllen (sollen).
Die Axiomatik ist *verträglich* sowohl mit der *Häufigkeits-* als auch mit der *Wettinterpretation*.
Die Axiome von Kolmogoroff geben an, wie man mit Wahrscheinlichkeiten rechnet. Welche Phänomene man durch Wahrscheinlichkeiten beschreiben darf und wie die Ergebnisse zu interpretieren sind, ist eine Frage des Wahrscheinlichkeitsbegriffs.

b) In der Tat gibt es auch Kritik an dieser Axiomatik: „zu streng und überpräzise“ $\longrightarrow$ aktueller Forschungsgegenstand (*Imprecise Probabilities, Intervallwahrscheinlichkeit*); hier nicht näher thematisiert: Kolmogorov als absolute „Wahrheit“.

c) Aus hier nicht zu erörternden mathematischen Gründen

* darf man bei überabzählbar unendlichen Ergebnisräumen, z.B. also im Fall $\Omega = \mathbb{R}$, nicht alle Teilmengen von Ω als Ereignisse zulassen. Alle Mengen, „an die man normalerweise denkt“ sind aber zugelassen.
* muss man bei unendlichen Ergebnisräumen in $(K3)$ eigentlich unendliche Summen zulassen.

* Wir werden uns darum aber nicht kümmern. (Nur daran denken, wenn man in etwas formalere Bücher schaut.)

Bem. 1.2 **Zentrale Rechenregeln**

i) $P(\overline{A}) =$, insb. $P(\emptyset)$ (1.2.4)

ii) $P(A \cup B) =$ (1.2.5)
für nicht notwendig disjunkte A, B.

iii) Falls $A_1, A_2, \ldots, A_n$ paarweise disjunkt sind, also $A_i \cap A_j = \emptyset$ für $i \neq j$, dann gilt:

$$\begin{aligned} P(\bigcup_{i=1}^{n} A_i) &= P(A_1 \cup A_2 \cup \ldots \cup A_n) \\ &= \end{aligned} \qquad (1.2.6)$$

Insbesondere folgt hieraus, dass, sofern Ω endlich ist, die Wahrscheinlichkeitsbewertung durch die Bewertung auf den Elementarereignissen vollständig bestimmt ist:

$$P(A) = \sum_{\omega \in A} p(\{\omega\})$$

Beweisidee zu (1.2.6):

iv) Für die Modellbildung wird später noch folgende Folgerung aus (iii) wichtig:

Ist $A_1, A_2, \ldots, A_n$ eine *vollständige Zerlegung* von Ω, d.h. gilt $\bigcup_{i=1}^{n} A_i = \Omega$ und $A_i \cap A_j = \emptyset$ für $i \neq j$, so gilt für jedes Ereignis B:

$$P(B) = \sum_{i=1}^{n} P(B \cap A_i) \tag{1.2.7}$$

Beispiel 1.3

Würfelwurf, fairer Würfel: $\Omega = \{1, 2, 3, 4, 5, 6\}$
Alle Elementarereignisse gleichwahrscheinlich

Sei $A = \{1, 3\}$:

$$P(A) = P(\{1, 3\})$$

Alternativ: direkt über Laplace

$$P(\{A\}) =$$

Aber, wie gesagt, Axiomatik nötig, um

Beispiel 1.4

Betrachtet werde ein Land, in dem die Wahlberechtigten die Wahl zwischen den Parteien Nr 1, Nr 2,..., Nr 6 (Nichtwähler) haben. Dabei entfallen auf die Parteien $2, 4, 6$ jeweils 25% der Stimmen; die restlichen Stimmen verteilen sich gleichmäßig auf die Parteien $1, 3, 5$. Seien $f_1, \dots f_6$ die entsprechenden relativen Häufigkeiten:

$$f_2 = f_4 = f_6 = \frac{1}{4} \quad , \quad f_1 = f_3 = f_5 = \frac{1}{12}$$

Es wird zufällig (im Sinne einer reinen Zufallsauswahl) eine Person ausgewählt und ihre Parteipräferenz ermittelt.
Geben Sie die sich ergebende Wahrscheinlichkeitsbewertung an.
Betrachten Sie ferner die Ereignisse

A	die Person wählt Partei	1 oder 3
B	"	4 oder 6
C	"	3 oder 4

und berechnen Sie $P(A), P(B), P(B \cap C)$ und $P(A \cup C)$!

- *Ergebnisraum*

- Durch die reine Zufallsauswahl hat insbesondere jede Person dieselbe Wahrscheinlichkeit gezogen zu werden: $\frac{1}{N}$ mit $N =$ Umfang der Grundgesamtheit.

  Wahrscheinlichkeit, Partei j zu wählen:

d.h. allgemein:

$$\boxed{P(\{j\}) = f_j \, , \, j = 1, \ldots 6.}$$

Was bedeutet das?

<u>Hier:</u> $P(\{1\}) = P(\{3\}) = P(\{5\}) =$

$P(\{2\}) = P(\{4\}) = P(\{6\}) =$

Modell entspricht einem verfälschtem Würfel.

- Überprüfe:

  $P(\Omega) \stackrel{!}{=} 1$

$P(A) =$

$P(B) =$

$P(C) =$

$P(B \cap C) =?$

$P(A \cup C) =$

alternative Berechnung von $P(A \cup C)$:

1.3 Stochastische Unabhängigkeit, bedingte Wahrscheinlichkeiten und Koppelung

(Fahrmeir et al. (2003), Kap 4.4 bis 4.7)

1.3.1 Stochastische Unabhängigkeit

Def 1.5 Zwei Ereignisse A und B heißen *(stochastisch) unabhängig beim Wahrscheinlichkeitsmaß* P, wenn gilt

$$P(A \cap B) = P(A) \cdot P(B) \qquad (1.3.1),$$

widrigenfalls heißen sie *stochastisch abhängig.*

Bem. 1.6:

a) Vorsicht: Nicht Unabhängigkeit und Disjunktheit verwechseln! Disjunktheit ist eine Eigenschaft der engen (Ereignisse), Unabhängigkeit ist eine Eigenschaft er Mengen (Ereignisse) *und* des Wahrscheinlichkeitsmaßes.

b) Stochastische Abhängigkeit bedeutet nicht kausale Abhängigkeit.

c) Die (stochastische) Unabhängigkeit ist eine symmetrische Beziehung in dem Sinne, dass A und B genau dann unabhängig sind, wenn B und A unabhängig sind.

   Genauer gilt die Äquivalenz folgender Aussagen:

   A und B sind stoch. unabhängig
   A und $\overline{B}$ "
   $\overline{A}$ und B "
   $\overline{A}$ und $\overline{B}$ "

d) Die Verwandtschaft zur empirischen Unabhängigkeit (Stat. I, (3.7): $f_{ij} = f_{i\bullet} f_{j\bullet}$ für alle i, j) ergibt sich, enn man sich durch A und B eine $(2 \times 2)-$Tafel erzeugt denkt, in der anstatt der Häufigkeiten die Wahrscheinlichkeiten stehen:

	B	$\overline{B}$	
A	$P(A \cap B)$	$P(A \cap \overline{B})$	$P(A)$
$\overline{A}$	$P(\overline{A} \cap B)$	$P(\overline{A} \cap \overline{B})$	$P(\overline{A})$
	$P(B)$	$P(\overline{B})$	

Unabhängigkeit:

$$\begin{aligned} \text{gemeinsame Verteilung} &= \text{Produkt der Randverteilungen} \\ P(A \cap B) &= P(A) \cdot P(B) \end{aligned}$$

Beispiel 1.7:
Fortsetzung von Beispiel 1.4
Sind die Ereignisse $A = \{1, 3\}$ und $C = \{3, 4\}$ stochastisch unabhängig?

Unabhängigkeit ist eine Eigenschaft, die vom Wahrscheinlichkeitsmaß abhängt. Betrachtet man wieder $\Omega = \{1, 2, 3, 4, 5, 6\}$, jetzt aber mit $P(\{1\}) = \frac{1}{9}$, $P(\{2\}) = \frac{1}{9}$, $P(\{3\}) = \frac{1}{18}$, $P(\{4\}) = \frac{5}{18}$, $P(\{5\}) = \frac{2}{9}$, $P(\{6\}) = \frac{2}{9}$,

also sind hier A und C stochastisch **un**abhängig.

1.3.2 Bedingte Wahrscheinlichkeiten

Def 1.8 Gegeben seien zwei Ereignisse A und B, wobei $P(B) > 0$ sei:
Dann heißt:

$$P(A|B) := \frac{P(A \cap B)}{P(B)} \qquad (1.3.2)$$

bedingte Wahrscheinlichkeit von A gegeben B oder *bedingte Wahrscheinlichkeit von A unter der Bedingung B.*

Interpretation: Wahrscheinlichkeit von A unter der Bedingung, dass B.

z.B.:
A: zufällig herausgegriffene Person wählt Partei X
B: zufällig herausgegriffene Person gehört der Oberschicht an

$P(A)$:
$P(A|B)$:

Nochmals deutlich:
$P(A \cap B)$:

Bem 1.9

a) Die Beziehung zu Statistik I und den bedingten relativen Häufigkeiten

b) An Stichprobenmodell denken
(Grundgesamtheit Ω, $P(B) \widehat{=} f_{\bullet 1}$, $P(A \cap B) \widehat{=} f_{11}$) :

	1	2	
1	f_{11}	f_{12}	$f_{1\bullet}$
2	f_{21}	f_{22}	$f_{2\bullet}$
	$f_{\bullet 1}$	$f_{\bullet 2}$	

c) Es ergibt sich auch wieder eine analoge Charakterisierung der stochastischen Unabhängigkeit über bedingte Wahrscheinlichkeiten:
Sind $P(A)$ und $P(B) > 0$, so sind äquivalent:

i) A und B sind stochastisch unabhängig

Merken!

Inhaltlich:

Nachweis zu ii)

d) Die Def. der Unabhängigkeit über (1.3.1) besitzt den Vorteil, dass man nicht $P(A) = 0$, $P(B) = 0$ ausschließen muss; dies kommt bei Wahrscheinlichkeiten durchaus vor.

e) „unter der Bedingung B“:
Anstatt aller Ergebnisse in Ω sind nur mehr die Ergebnisse in B möglich. Das Betrachten bedingter Wahrscheinlichkeiten entspricht also einer Änderung des Grundraumes: B statt Ω.
$P(A|B) = P(A \cap B|B)$ ist als Funktion in A bei festem B wieder eine Wahrscheinlichkeitsbewertung, erfüllt also wieder die Axiome von Kolmogorov.

1.3.3 Koppelung von unabhängigen Experimenten, unabhängige Wiederholungen

Formaler als hier: Rohwer und Pötter (2002): *Wahrscheinlichkeit. Begriff und Rhetorik in der Sozialforschung*

(teilweise implizit) Begriff: Behnen und Neuhaus (1984) *Grundkurs Stochastik*, Teubner

in Fahrmeir et al.: später bei Zufallsvariablen Kap 8.4

Mit dem Begriff der Unabhängigkeit (und der bedingten Wahrscheinlichkeit, siehe später) kann man komplexere Situationen aus „Einzelbausteinen“ zusammensetzen:

- bisher:

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) \quad \Longrightarrow \text{ unabhängig.}$$

- jetzt Unabhängigkeit zum Ausdruck bringen: gegeben $P(A_1)$, $P(A_2)$, setzt man $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$, so ist Unabhängigkeit gegeben.

- übliches didaktisches Beispiel: Werfen eines Würfels ($\Omega_1 = \{1, \ldots, 6\}$) und eines Oktaeders ($\Omega_2 = \{1, \ldots, 8\}$) unabhängig voneinander:

$$A_1 \subset \Omega_1 : \ A_1 = \{5, 6\} \,, \quad A_2 \subset \Omega_2 : \ A_2 = \{7, 8\}$$

$A_1 \cap A_2$: „eine 5 oder 6 mit dem Würfel und eine 7 oder 8 mit dem Oktaeder zu würfeln"
Also bei fairem Würfel $P_1(\{j\}) = \frac{1}{6}$, $i = 1, \ldots, 6$, und bei fairem Oktaeder $P_2(\{j\}) = \frac{1}{8}$, $i = 1, \ldots, 8$ setzt man fest:

$$P(A_1 \cap A_2) \; [:= P(A_1 \times A_2) =] \; = P_1(A_1) \cdot P_2(A_2) = \frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}.$$

Bem 1.10: (Unabhängige Koppelung)
Gegeben sei eine Reihe von Zufallsexperimenten, beschrieben durch die Ergebnisräume Ω_i, $i = 1, \ldots, n$ und die Wahrscheinlichkeitsbewertungen P_i, $i = 1, \ldots, n$. Faßt man die Experimente zusammen, so ergibt sich der Ergebnisraum

$$\Omega := \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n \,,$$

und die Elemente $\omega := (\omega_1, \omega_2, \ldots, \omega_n)$. Sind die Experimente unabhängig (Dies ist inhaltlich zu entscheiden!), so setzt man für beliebige $A_i \subset \Omega_i, i = 1, \ldots, n$

$$P\left(A_1 \cap A_2 \cap \ldots \cap A_n\right) = P_1(A_1) \cdot P_2(A_2) \cdot \ldots \cdot P_n(A_n) \,.$$

Dies beschreibt ein Wahrscheinlichkeitsmaß auf Ω, bei dem – per Konstruktion – beliebige Ereignisse aus den einzelnen Ω_i voneinander unabhängig sind. Es ist in der Literatur üblich, die Indizes $_{1,2,\ldots,n}$ bei den P´s wegzulassen, wie wir es auch oben im Beispiel getan haben.

Von besonderer Bedeutung ist der Fall *unabhängiger und identischer Wiederholungen*, bei dem dasselbe Experiment wiederholt durchgeführt wird.

Bsp und Bem 1.11: (Zufallsstichprobe vom Umfang n)
Stichprobe vom Umfang n:
Das Experiment „Ziehen einer Person und Ermittelung ihrer Parteipräferenz" wird n-mal unabhängig (Befragte dürfen sich nicht gegenseitig beeinflussen!) durchgeführt.
Betrachtet werde die Situation von Bem 1.4 (endliche Grundgesamtheit $\tilde{\Omega}$, Merkmal $\tilde{X}$, Ausprägungen $a_1, \ldots a_k$, Häufigkeiten $f_1, \ldots f_k$). Es werde eine reine *(geordnete) Zufallsstichprobe vom Umfang n bezüglich des Merkmals* $\tilde{X}$ entnommen, d.h. eine (geordnete) Zufallsauswahl (mit Zurücklegen) von n Elementen $\omega_1, \ldots \omega_n$ mit $\omega_i \in \tilde{\Omega},\ i = 1 \ldots n$, entnommen und die diesbezüglichen Ausprägungen $\tilde{X}(\omega_i)$ von X erhoben.
Sei für $j = 1, \ldots k,\ i = 1, \ldots n$ mit A_{ij} das Ereignis „$\tilde{X}(\omega_i) = a_j$" („Die i-te gezogene Person hat Ausprägung a_j, z.B. wählt Partei B") bezeichnet, so gilt für beliebige $j_1, j_2, \ldots j_n$

$$\begin{aligned} &P(A_{1j_1} \cap A_{2j_2} \cap \ldots A_{nj_n}) = \\ &= P(A_{1j_1}) \cdot P(A_{2j_2}) \cdot \ldots \cdot P(A_{nj_n}) \overset{(1.2)}{=} f_{j_1} \cdot f_{j_2} \cdot \ldots \cdot f_{j_n} \end{aligned} \qquad (1.3.3)$$

Aufgabe: Veranschaulichen Sie sich Beispiel und Bemerkung 1.11 anhand von Beispiel 1.12

Bsp 1.12 (Zur reinen Zufallsauswahl vom Umfang n: Wahlumfrage)
Unmittelbar nach Schließung der Wahllokale 2002 habe man eine reine Zufallsauswahl vom Umfang 10 (damit von Hand rechenbar) unter den Wählern vorgenommen und ihre Wahlentscheidung erfragt.
Wie groß/gering ist die Wahrscheinlichkeit, mindestens 9 PDS-Anhänger in der Stichprobe zu haben?

Daten, amtliches Endergebnis:

SPD:	$38,5\%$
CDU/CSU:	$38,5\%$
B90/Grüne:	8.6%
FDP:	$7,4\%$
PDS:	$4,0\%$
Sonstige:	$3,0\%$

„Mindestens 9 mal PDS“ heißt: