Seminar über Bayes-Statistik

SUBJEKTIVE WAHRSCHEINLICHKEIT

Alain Helfenstein & Thomas Krabichler

1 Einleitung

In der Statistik und Wahrscheinlichkeitsrechnung werden Entscheidungen anhand einer Annahme an die zugrundeliegende Wahrscheinlichkeitsverteilung getroffen. Jedes Problem beinhaltet jedoch die Handhabung der spezifischen Wahrscheinlichkeit von gewissen fundamentalen Ereignissen. Man trifft stets eine Modellannahme und arbeitet auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Betrachtet man darauf eine Zufallsvariable Y mit Werten in einem messbaren Raum $(S, \mathcal{A})$, so liefert das Wahrscheinlichkeitsmass $P \circ Y^{-1}$ eine Verteilung auf dem Raum $(S, \mathcal{A})$. $P \circ Y^{-1}$ weist jedem Ereignis $A \in \mathcal{A}$ eine Wahrscheinlichkeit bezüglich dessen Auftreten zu. In unserem Seminar-Vortrag steht nun folgende Frage zu Grunde: 'Wie weist der Statistiker anfänglich den Ereignissen A die Wahrscheinlichkeiten zu, auf welchen die weiteren Berechnungen fundieren?' Ein Beispiel:

Beispiel 1.1 Falls wir einen 'normalen' Würfel (also einen Würfel mit den Augenzahlen 1 - 6 auf jeweils gleich grossen Flächen) betrachten und an dem Ereignis $A =$ 'Gerade Zahl' $= \{2, 4, 6\}$ interessiert sind, folgt unmittelbar, dass $P[A] = \frac{3}{6} = \frac{1}{2}$ gelten muss. Dies gilt jedoch nur, weil wir durch die Annahme an den Würfel bereits eine Gleichverteilung gewählt haben. Falls wir nun aber mit einem Würfel spielen, bei dem nicht alle Flächen gleich gross sind, sieht das ganze Problem wieder anders aus, und wir müssen ein anderes Modell suchen.

Dieses Beispiel zeigt, dass in den meisten Fällen die Zuweisung eines Wahrscheinlichkeitsmasses reine Routine geworden ist. In den verschiedenen Bereichen haben sich viele Standards entwickelt. Tradition und Erfahrung spielen in diesem Bereich eine grosse Rolle. Oftmals ist es durch die Problemstellung klar, welche Verteilung gewählt werden kann. Es gibt aber auch Situationen, in welchen die Wahl schwierig ist, und nur wenige Leute die gleiche Meinung teilen. In diesen Problemen braucht es viel Erfahrung, und die gewählte Verteilung ist oft eine sehr subjektive Entscheidung. In unserem Vortrag werden wir zeigen, wie der Statistiker dieses Problem angehen kann, und unter welchen Umständen überhaupt eine gewünschte Verteilung existieren kann. Dazu führen wir eine Relation und Axiome ein, die dann zu dem gewünschten Resultat führen.

2 Relative Wahrscheinlichkeit

Wir bewegen uns nun in einem Stichprobenraum S zusammen mit einer σ-Algebra $\mathcal{A}$. Der Kerngedanke der Konstruktion einer Verteilung auf $\mathcal{A}$ liegt darin, dass man entscheiden kann, ob das Eintreten eines Ereignisses A als wahrscheinlicher, gleich wahrscheinlich oder weniger wahrscheinlich betrachtet wird als das Eintreten eines Ereignisses B. Dies bringt uns zur ersten Definition:

Definition 2.1 Wenn wir zwei Ereignisse A und B vergleichen, schreiben wir $A \prec B$ um auszudrücken, dass B mit grösserer Wahrscheinlichkeit auftritt als A und $A \sim B$ um auszudrücken, dass A und B mit der gleichen Wahrscheinlichkeit auftreten. Ferner schreiben wir $A \precsim B$, falls wir denken, dass B zumindest mit der gleichen Wahrscheinlichkeit auftritt wie A oder äquivalent, dass A mit höchstens derselben Wahrscheinlichkeit auftritt wie B.

Mit dieser Definition ist auch klar, welche Struktur ein mögliches Wahrscheinlichkeitsmass auf $\mathcal{A}$ erfüllen muss; für alle Ereignisse A und B aus $\mathcal{A}$ soll folgende Äquivalenz gelten:

$$P[A] \leq P[B] \qquad \Longleftrightarrow \qquad A \precsim B. \tag{1}$$

Ein Mass P, welches die Äquivalenz (1) erfüllt, heisst der Relation $\precsim$ *entsprechend.* Wir geben im Folgenden Bedingungen an, welche die Existenz und die Eindeutigkeit eines der Relation $\precsim$ entsprechenden Wahrscheinlichkeitsmasses garantieren. Dazu formulieren wir Axiome, welche wir (als Referenz zu *Subjective Probability*) SP1 bis SP4 nennen.

SP 1 *Für je zwei beliebige Ereignisse $A, B \in \mathcal{A}$ muss genau eine der drei Relationen $A \prec B$, $A \succ B$ oder $A \sim B$ erfüllt sein.*

SP 2 *Sind A_1, A_2, B_1, B_2 vier Ereignisse mit $A_1 \cap A_2 = B_1 \cap B_2 = \varnothing$ und $A_i \precsim B_i$ für $i = 1, 2$, so gilt $A_1 \cup A_2 \precsim B_1 \cup B_2$. Gilt darüber hinaus $A_1 \prec B_1$ oder $A_2 \prec B_2$, so haben wir $A_1 \cup A_2 \prec B_1 \cup B_2$.*

SP 3 *Für ein beliebiges Ereignis $A \in \mathcal{A}$ gilt $\varnothing \precsim A$. Ferner gilt $\varnothing \prec S$.*

SP 4 *Ist $A_1 \supset A_2 \supset \ldots$ eine fallende Folge von Ereignissen und B ein Ereignis mit $A_i \succsim B$ für $i = 1, 2, \ldots$, so gilt*

$$\bigcap_{i=1}^{\infty} A_i \succsim B.$$

SP2 kann man zum Beispiel folgendermassen interpretieren: Falls wir annehmen, dass die Ereignisse A und B auf genau zwei sich gegenseitig ausschliessende Arten auftreten können, und falls in beiden Fällen das Auftreten von A nicht wahrscheinlicher ist als das Auftreten von B, dann gilt $A \precsim B$.
Wir werden nun einige kurze und interessante Konsequenzen der Axiome präsentieren:

Lemma 2.2 *Seien A, B und C Ereignisse mit $A \cap C = B \cap C = \varnothing$. Dann gilt $A \precsim B$ genau dann, wenn $A \cup C \precsim B \cup C$ gilt.*

Beweis.
$\Rightarrow$: Wir haben $A \precsim B$. Wenn wir in SP2 $A_2 = B_2 = C$ setzen, folgt die Behauptung unmittelbar.
$\Leftarrow$: Wir haben $A \cup C \precsim B \cup C$. Nehmen wir an, es gelte $A \succ B$. Wenn wir wiederum SP2 anwenden, erhalten wir, dass $A \cup C \succ B \cup C$ gilt. Dies steht aber im Widerspruch zur Voraussetzung. □

Dieses Lemma liefert uns das wichtige Resultat der *Transitivität* unserer Relation:

Satz 2.3 *Sind A, B und C Ereignisse mit $A \precsim B$ sowie mit $B \precsim C$, so gilt $A \precsim C$.*

Beweis. Wir teilen die Vereinigung $A \cup B \cup C$ in 7 disjunkte Ereignisse auf, wie in Abbildung 1 skizziert. Nun wissen wir, dass $A \precsim B$ gilt. Diesen Sachverhalt können wir wie folgt schreiben

$$\begin{aligned} &(A \cap B^c \cap C^c) \cup (A \cap B^c \cap C) \cup (A \cap B \cap C^c) \cup (A \cap B \cap C) \\ &\precsim (A^c \cap B \cap C^c) \cup (A^c \cap B \cap C) \cup (A \cap B \cap C^c) \cup (A \cap B \cap C). \end{aligned}$$

Wenn wir nun Lemma 2.2 anwenden, um auf beiden Seiten die jeweils gleichen Ereignisse (die nach Konstruktion ja alle disjunkt zu einander sind) 'zu kürzen', erhalten wir

$$(A \cap B^c \cap C^c) \cup (A \cap B^c \cap C) \precsim (A^c \cap B \cap C^c) \cup (A^c \cap B \cap C). \tag{2}$$

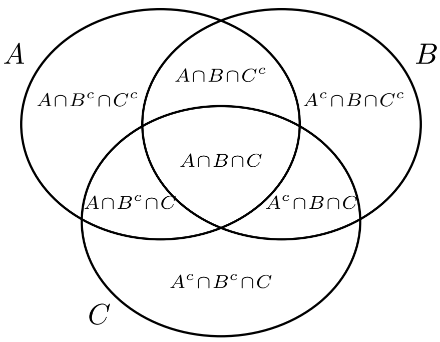

Abbildung 1: Schnittmengen

Genauso erhält man wegen $B \precsim C$

$$(A \cap B \cap C^c) \cup (A^c \cap B \cap C^c) \precsim (A \cap B^c \cap C) \cup (A^c \cap B^c \cap C). \tag{3}$$

Da nun jeweils die linken und rechten Seiten von (2) und (3) disjunkt sind, erhält man mit SP2

$$\begin{aligned}&(A \cap B^c \cap C^c) \cup (A \cap B^c \cap C) \cup (A \cap B \cap C^c) \cup (A^c \cap B \cap C^c)\\ &\precsim (A^c \cap B \cap C^c) \cup (A^c \cap B \cap C) \cup (A \cap B^c \cap C) \cup (A^c \cap B^c \cap C).\end{aligned}$$

Wenn wir nun wiederum nach Lemma 2.2 auf beiden Seiten $(A \cap B^c \cap C) \cup (A^c \cap B \cap C^c)$ 'kürzen', erhalten wir $(A \cap B^c \cap C^c) \cup (A \cap B \cap C^c) \precsim (A^c \cap B \cap C) \cup (A^c \cap B^c \cap C)$. Nach Lemma 2.2 können wir nun auf beiden Seiten das Ereignis $(A \cap B^c \cap C) \cup (A \cap B \cap C)$ hinzufügen und erhalten die Behauptung. □

Dieser Satz zusammen mit SP1 zeigt also, dass die Relation $\precsim$ eine totale Quasiordnung der Ereignisse in $\mathcal{A}$ liefert. Ausgehend von SP2 erhält man durch vollständige Induktion das folgende Resultat:

Satz 2.4 *Für je n disjunkte Ereignisse $A_1, \ldots, A_n$ und $B_1, \ldots, B_n$ mit $A_i \precsim B_i$ für $i = 1, \ldots, n$ gilt*

$$\bigcup_{i=1}^{n} A_i \precsim \bigcup_{i=1}^{n} B_i.$$

Gilt ferner $A_i \prec B_i$ für mindestens ein $1 \leq i \leq n$, so haben wir

$$\bigcup_{i=1}^{n} A_i \prec \bigcup_{i=1}^{n} B_i.$$

Auch eine weitere wichtige Eigenschaft für Wahrscheinlichkeiten ist erfüllt:

Satz 2.5 *Für zwei beliebige Ereignisse A und B gilt genau dann $A \precsim B$, wenn $A^c \succsim B^c$ gilt.*

Beweis.
$\Rightarrow$: Wir haben $A \precsim B$. Nehmen wir nun an, es gelte $A^c \prec B^c$. Dann folgt mit SP2 für $A_1 = A, A_2 = A^c$ und $B_1 = B, B_2 = B^c$, dass $A \cup A^c \prec B \cup B^c$ sprich $S \prec S$ gilt. Dies ist offensichtlich ein Widerspruch zu SP1.
$\Leftarrow$: Diese Richtung zeigt man komplett analog. □

Bis jetzt haben wir jeweils nur die ersten beiden Axiome benützt. Wenn wir nun SP3 hinzunehmen, erhalten wir ein für das entsprechende Wahrscheinlichkeitsmass zentrales Resultat:

Satz 2.6 *Für zwei Ereignisse A und B mit $A \subset B$ gilt $A \precsim B$. Für jedes Ereignis $A \in \mathcal{A}$ gilt insbesondere $\varnothing \precsim A \precsim S$.*

Beweis. Sei $A \subset B$. Wegen SP3 gilt $\varnothing \precsim A^c \cap B$. Da $A^c \cap B$ und A disjunkt sind, folgt mit SP2 $A \precsim A \cup (A^c \cap B) = B$. $\square$

Zur Veranschaulichung von Axiom SP4 dient das folgende Beispiel:

Beispiel 2.7 Wir betrachten die Teilmengen der reellen Zahlen als Ereignisse. Wir nehmen an, dass das Auftreten jedes unendlichen Intervalls der Form (n, ∞) für $n = 1, 2 \ldots$ wahrscheinlicher ist, als das Auftreten einer fixen und kleinen Teilmenge B. Der Schnitt über alle unendlichen Intervalle wie oben ist klarerweise die leere Menge $\varnothing$. Aus SP4 folgt daher, dass für die Teilmenge B gelten muss: $B \sim \varnothing$. Oder anders ausgedrückt: Falls B irgendeine Teilmenge ist mit $B \succsim \emptyset$, dann folgt, dass nicht jedes unendliche Intervall (n, ∞) zumindest mit der gleichen Wahrscheinlichkeit wie B auftreten kann, egal wie klein B auch sein mag.

Satz 2.5 liefert eine zu SP4 äquivalente Aussage:

Satz 2.8 *Ist $A_1 \subset A_2 \subset \ldots$ eine aufsteigende Folge von Ereignissen und B ein Ereignis mit $A_i \precsim B$ für $i = 1, 2, \ldots$, so gilt*

$$\bigcup_{i=1}^{\infty} A_i \precsim B.$$

Beweis. Nach Annahme und Satz 2.5 folgt, dass $A_1^c \supset A_2^c \supset \ldots$ eine fallende Folge von Ereignissen ist mit $A_i^c \succsim B^c$ für $i = 1, 2, \ldots$. Also folgt mit SP4 und den Regeln von de Morgan, dass gilt

$$\bigcap_{i=1}^{\infty} A_i^c = \left(\bigcup_{i=1}^{\infty} A_i \right)^c \succsim B^c.$$

Satz 2.5 liefert nun die Behauptung. $\square$

Wir wollen nun Satz 2.4 auf Vereinigungen von unendlich vielen disjunkten Ereignisse ausweiten.

Satz 2.9 *Für zwei Folgen disjunkter Ereignisse $A_1, A_2, \ldots$ und $B_1, B_2, \ldots$ mit $A_i \precsim B_i$ für $i = 1, 2, \ldots$ gilt*

$$\bigcup_{i=1}^{\infty} A_i \precsim \bigcup_{i=1}^{\infty} B_i.$$

Gilt ferner $A_i \prec B_i$ für mindestens ein $1 \leq i < \infty$, so haben wir

$$\bigcup_{i=1}^{\infty} A_i \prec \bigcup_{i=1}^{\infty} B_i.$$

Beweis. Aus Satz 2.4 wissen wir, dass die Behauptung für endliche Vereinigungen gilt, d.h. es gilt

$$\bigcup_{i=1}^{n} A_i \precsim \bigcup_{i=1}^{n} B_i.$$

Da $\bigcup_{i=1}^{n} B_i \subset \bigcup_{i=1}^{\infty} B_i$ gilt, folgt mit Satz 2.6 für $n = 1, 2, \ldots$

$$\bigcup_{i=1}^{n} A_i \precsim \bigcup_{i=1}^{n} B_i \precsim \bigcup_{i=1}^{\infty} B_i. \tag{4}$$

Auf der linken Seite von (4) handelt es sich für $n = 1, 2, \ldots$ um eine aufsteigende Folge von Ereignissen. Folglich können wir mit Satz 2.8 schreiben

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{n\in\mathbb{N}} \bigcup_{i=1}^{n} A_i \precsim \bigcup_{i=1}^{\infty} B_i,$$

und die erste Behauptung ist bewiesen.
Gilt nun $A_j \prec B_j$ für mindestens ein $j \in \mathbb{N}$, so folgt nach Satz 2.4 für jedes $n \geq j$

$$\bigcup_{i=1}^{n} A_i \prec \bigcup_{i=1}^{n} B_i. \tag{5}$$

Wenn wir den bereits bewiesenen Teil von Satz 2.9 für die Indexmenge $i = n+1, n+2, \ldots$ anwenden, können wir schreiben

$$\bigcup_{i=n+1}^{\infty} A_i \precsim \bigcup_{i=n+1}^{\infty} B_i. \tag{6}$$

Nach Voraussetzung sind die involvierten Mengen disjunkt, und mit SP2 angewandt auf (5) und (6) folgt

$$\bigcup_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{n} A_i\right) \cup \left(\bigcup_{i=n+1}^{\infty} A_i\right) \prec \left(\bigcup_{i=1}^{n} B_i\right) \cup \left(\bigcup_{i=n+1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} B_i.$$

□

Es ist klar, dass die Axiome SP1 bis SP4 notwendig sind für die Existenz eines der Relation $\precsim$ entsprechenden Wahrscheinlichkeitsmasses. Hinreichend sind diese vier Annahmen nicht, wie ein Gegenbeispiel von Kraft, Pratt und Seidenberg (siehe [2]) zeigte. Im nächsten Abschnitt werden wir deshalb eine weitere Annahme hinzunehmen.
In die Entscheidung, welche Ereignisse der Statistiker nun als wahrscheinlicher einschätzt, können auch zusätzliche Überlegungen einfliessen. Nehmen wir an, dass man entscheiden muss, welches der Ereignisse A und B mit grösserer Wahrscheinlichkeit auftritt. Der Statistiker kann sich zum Beispiel fragen, ob er lieber an einem Wettbewerb teilnehmen will, bei dem er einen Gewinn einfährt, falls das Ereignis A eintrifft, aber nichts gewinnt, falls A ausbleibt. Oder möchte er lieber an einem Wettbewerb teilnehmen, wo man mit derselben Gewinnaussicht auf das Ereignis B setzen kann. Um seine Chancen auf den Gewinn zu maximieren, wird sich der Statistiker für das Ereignis entscheiden, welches er als wahrscheinlicher betrachtet. Diese Entscheidungsgrundlage hat jedoch ihre Grenzen und bringt Gefahren mit sich. Das folgende Beispiel soll illustrieren, wie diese Vorgehensweise zu widersprüchlichen Ergebnissen führen kann.
Ein normaler Bürger der USA nimmt an, dass die Wahrscheinlichkeit, innerhalb der nächsten zehn Jahre in einem Atomkrieg zu sterben, grösser ist, als innerhalb derselben Zeit zum Präsidenten der USA gewählt zu werden. Wenn man aber jemanden fragt, auf welches Ereignis er im Stile eines Wettbewerbes wie oben beschrieben, setzen würde, ist es klar, dass jeder lieber einen Preis bei seiner Amtseinführung entgegennehmen will, als bei seiner Beerdigung.
Dieses Beispiel ist sicherlich sehr extrem, aber es zeigt klar die Schwierigkeiten, die entstehen können, wenn man zwei Ereignisse vergleichen muss, die (nach Ramsey, 1926) nicht 'ethisch neutral' sind. Man betrachtet nicht nur die Ereignisse per se, sondern bringt automatisch auch deren Konsequenzen ins Spiel, und diese können zu falschen Annahmen führen. Man sollte daher, wenn möglich, bloss die Wahrscheinlichkeiten betrachten, mit denen die Ereignisse auftreten können, und nicht deren Konsequenzen.

3 Das Hilfsexperiment

Im folgenden Abschnitt präsentieren wir eine letzte Annahme, die es schlussendlich ermöglicht, jedem Ereignis in eindeutiger Weise eine Wahrscheinlichkeit zuordnen zu können. Mit einem Wahrscheinlichkeitsmass kann man nicht nur angeben, welches von zwei Ereignissen wahrscheinlicher ist, sondern auch in welchem Mass das eine Ereignis wahrscheinlicher ist als das andere. Letzterem wird im Allgemeinen nicht Rechnung getragen, wenn man ausschliesslich die Annahmen SP1 bis SP4 betrachtet. Dies soll das folgende Beispiel illustrieren.

Beispiel 3.1 Betrachte ein Experiment mit zwei möglichen Ausgängen A und A^c. Nach Ansicht eines Statistikers mag $A \prec A^c$ gelten, aber augenscheinlich ist es unmöglich, ohne weitere Betrachtungen eine sinnvolle Zuordnung der Wahrscheinlichkeiten $P[A]$ und $P[A^c]$ zu treffen.

Die grundlegende Idee ist nun ein System $\mathcal{B} \subset \mathcal{A}$ zu finden, welches folgenden Anforderungen genügt:

1. Jedem $B \in \mathcal{B}$ kann in eindeutiger Weise eine Wahrscheinlichkeit zugeordnet werden.
2. Zu jedem $A \in \mathcal{A}$ existiert ein $B \in \mathcal{B}$ mit $A \sim B$.

Diese Idee macht insofern den Anschein, eine sehr grosse Einschränkung zu sein, als dass unklar ist, wie der Statistiker zu einem solchen System gelangt. Wie in diesem und im nächsten Abschnitt gezeigt wird, findet diese Idee implizit ihre Realisierung, indem wir die Existenz einer gewissen „Zufallsvariable“ fordern.
Für eine messbare Funktion X auf $(S, \mathcal{A})$ und für Intervalle I_1, I_2 sind $\{X \in I_1\}$ und $\{X \in I_2\}$ Elemente von $\mathcal{A}$. Gemäss Annahme SP1 muss deshalb entweder $\{X \in I_1\} \precsim \{X \in I_2\}$ oder $\{X \in I_1\} \succsim \{X \in I_2\}$ erfüllt sein. Für ein Intervall I mit Endpunkten a und b bezeichnen wir im Folgenden mit $\lambda(I) := b-a$ die *Länge* von I. Dabei bleibt unbeachtet, ob keiner, einer oder beide Endpunkte im Intervall enthalten sind. Wir werden jetzt die Definition einer Zufallsvariablen mit einer Gleichverteilung auf dem Intervall $[0,1]$ geben, ohne dass dabei ein Mass involviert ist.

Definition 3.2 Wir sagen, eine $[0,1]$-wertige Zufallsvariable X auf $(S, \mathcal{A})$ mit der Relation $\precsim$ besitze eine *Gleichverteilung* auf dem Intervall $[0,1]$, wenn für alle Intervalle $I_1, I_2 \subset [0,1]$ gilt

$$\{X \in I_1\} \precsim \{X \in I_2\} \qquad \Longleftrightarrow \qquad \lambda(I_1) \leq \lambda(I_2).$$

Nach diesen Vorbereitungen kann die letzte Annahme einfach formuliert werden.

SP 5 *Es existiert eine Zufallsvariable auf* $(S, \mathcal{A})$ *mit einer Gleichverteilung auf* $[0,1]$.

Wie SP5 den obigen Überlegungen Rechnung trägt, wird im nächsten Abschnitt ausgearbeitet.

4 Konstruktion des Wahrscheinlichkeitsmasses

In diesem Abschnitt nehmen wir an, dass der Stichprobenraum $(S, \mathcal{A})$ mit der Relation $\precsim$ die Annahmen SP1 bis SP5 erfüllt. Unter diesen Voraussetzungen werden wir zeigen, dass auf $(S, \mathcal{A})$ ein eindeutiges Wahrscheinlichkeitsmass P existiert, welches der Relation $\precsim$ entspricht.
Gemäss Annahme SP5 existiert eine Zufallsvariable X auf $(S, \mathcal{A})$, welche eine Gleichverteilung auf $[0,1]$ besitzt. Für ein offenes Intervall $(a,b) \subset [0,1]$ bezeichne $G(a,b)$ das Ereignis, dass X im Intervall (a,b) liegt, d.h. $G(a,b) := \big\{X \in (a,b)\big\}$. Da X gleichverteilt ist, gilt für je zwei Intervalle $(a_1,b_1), (a_2,b_2) \subset [0,1]$ die Äquivalenz

$$G(a_1,b_1) \precsim G(a_2,b_2) \qquad \Longleftrightarrow \qquad b_1 - a_1 \leq b_2 - a_2.$$

Ferner gilt mit den entsprechenden Definitionen $G(a_1,b_1) \sim G[a_1,b_1) \sim G(a_1,b_1] \sim G[a_1,b_1]$. Für die Definition des Wahrscheinlichkeitsmasses P ist der folgende Satz zentral.

Satz 4.1 *Sei $A \in \mathcal{A}$ ein Ereignis. Dann existiert ein eindeutiges $a^* \in [0,1]$, so dass gilt*

$$A \sim G[0, a^*].$$

Beweis. In einem ersten Schritt zeigen wir die *Existenz* eines solchen a^*. Definiere

$$U(A) := \big\{a \in [0,1] \,\big|\, A \precsim G[0,a]\big\}.$$

Es gilt $A \precsim S = G[0,1]$ und folglich $1 \in U(A)$, womit die Menge $U(A)$ nicht leer ist. Setze

$$a^* := \inf U(A).$$

Ist $(a_k)_{k\in\mathbb{N}}$ eine monoton fallende Folge in $U(A)$, die gegen a^* konvergiert, dann gilt

$$G[0, a^*] = \bigcap_{k=1}^{\infty} G[0, a_k].$$

Nach Annahme SP4 gilt also

$$A \precsim G[0, a^*]. \tag{7}$$

Umgekehrt gilt es zwei Fälle zu betrachten:
Fall 1: $a^* = 0$. Es gilt $G[0,0] \sim \varnothing \precsim A$, und es folgt mit der Ungleichung (7) $A \sim G[0,0]$.
Fall 2: $a^* > 0$. Wegen der Definition von a^* gilt $G[0,a] \prec A$ für jedes $a \in [0, a^*)$. Ist $(a_k)_{k\in\mathbb{N}}$ eine monoton wachsende Folge in $[0, a^*)$, welche gegen a^* konvergiert, so gilt

$$G[0, a^*) = \bigcup_{k=1}^{\infty} G[0, a_k].$$

Mit der von Satz 2.8 dualen Formulierung von SP4 folgt also $G[0,a^*] \sim G[0,a^*) \precsim A$. Wiederum wegen Ungleichung (7) haben wir demzufolge $A \sim G[0,a^*]$.
Es bleibt die *Eindeutigkeit* von a^* zu zeigen. Nach Definition der Gleichverteilung gilt für zwei Zahlen $a_1, a_2 \in [0,1]$ mit $a_1 < a^* < a_2$ stets $G[0,a_1] \prec G[0,a^*] \prec G[0,a_2]$, wonach nur eines dieser drei Ereignisse äquivalent zu A sein kann. Demzufolge ist a^* eindeutig. □

Mit Hilfe von Satz 4.1 sind wir nun in der Lage, das der Relation $\precsim$ entsprechende Wahrscheinlichkeitsmass P zu definieren. Dies geschieht mit der eindeutig bestimmten Zahl a^*. Und zwar charakterisieren wir für ein Ereignis $A \in \mathcal{A}$ die Wahrscheinlichkeit $P[A]$ durch

$$A \sim G\big[0, P[A]\big]. \tag{8}$$

Definition 4.2 Wir definieren auf $\mathcal{A}$ die Gewichtsfunktion

$$P : \mathcal{A} \longrightarrow [0,1]\,,\; A \longmapsto P[A] := \inf \big\{a \in [0,1] \,\big|\, A \precsim G[0,a]\big\}.$$

Bevor wir zeigen, dass es sich bei P tatsächlich um ein Wahrscheinlichkeitsmass handelt, beweisen wir, dass P der Relation $\precsim$ entspricht.

Satz 4.3 *Seien $A, B \in \mathcal{A}$. Dann gilt die Äquivalenz*

$$A \precsim B \qquad \Longleftrightarrow \qquad P[A] \leq P[B].$$

Beweis. Gemäss der charakterisierenden Gleichung (8) haben wir $A \precsim B$ genau dann, wenn $G\big[0, P[A]\big] \precsim G\big[0, P[B]\big]$ gilt. Letzteres ist aber nach Definition der Gleichverteilung äquivalent zu $P[A] \leq P[B]$. □

Satz 4.4 *P ist ein Wahrscheinlichkeitsmass.*

Der Beweis von Satz 4.4 bedarf einiger Vorbereitungen.

Lemma 4.5 *Seien $A, B \in \mathcal{A}$ disjunkte Ereignisse. Dann gilt $P[A \cup B] = P[A] + P[B]$.*

Beweis. Gemäss (8) gilt $A \sim G\big[0, P[A]\big]$ und $A \cup B \sim G\big[0, P[A \cup B]\big]$. Wegen $A \subset A \cup B$ gilt ferner $A \precsim A \cup B$ und somit nach Satz 4.3 $P[A] \leq P[A \cup B]$. Wir zeigen nun, dass gilt

$$B \sim G\big(P[A], P[A \cup B]\big]. \tag{9}$$

Wäre $B \prec G\big(P[A], P[A \cup B]\big]$, so hätten wir wegen der Annahme SP2

$$A \cup B \prec G\big[0, P[A]\big] \cup G\big(P[A], P[A \cup B]\big] = G\big[0, P[A \cup B]\big],$$

was offensichtlich ein Widerspruch zu den einleitenden Bemerkungen wäre. Genauso kann man zeigen, dass $B \succ G\big(P[A], P[A \cup B]\big]$ *nicht* gelten kann. Demzufolge ist die Korrektheit von (9) verifiziert.

Nach der Definition der Gleichverteilung gilt $G\big(P[A], P[A \cup B]\big] \sim G\big[0, P[A \cup B] - P[A]\big]$. Ferner haben wir $B \sim G\big[0, P[B]\big]$. Mit der Transitivität von $\precsim$ folgt schliesslich

$$G\big[0, P[B]\big] \sim G\big[0, P[A \cup B] - P[A]\big].$$

Eine weitere Anwendung der Definition der Gleichverteilung ergibt $P[B] = P[A \cup B] - P[A]$, und Lemma 4.5 ist bewiesen. □

Mit Induktion folgt die Additivität von P:

Korollar 4.6 *Seien $A_1, \ldots, A_n \in \mathcal{A}$ disjunkte Ereignisse. Dann gilt*

$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} P[A_k].$$

Es ist ein wohlbekanntes Resultat aus der Masstheorie, dass die folgende Aussage zusammen mit der endlichen Additivität von P äquivalent zur σ-Additivität von P ist.

Lemma 4.7 *Sei $A_1 \supset A_2 \supset \ldots$ eine fallende Folge von Ereignissen mit $\bigcap_{k=1}^{\infty} A_k = \varnothing$. Dann gilt*

$$\lim_{k \to \infty} P[A_k] = 0.$$

Beweis. Weil P der Relation $\precsim$ entspricht, ist $\big(P[A_k]\big)_{k \in \mathbb{N}}$ eine monoton fallende Folge in $[0, 1]$, welche gegen ein $b \in [0, 1]$ konvergiert. Es gilt $P[A_k] \geq b$ für alle $k \in \mathbb{N}$, womit wiederum mit der Entsprechung von P und $\precsim$ folgt, dass für alle $k \in \mathbb{N}$ gilt $A_k \succsim G[0, b]$. Die Annahme SP4 impliziert nun

$$G[0, b] \precsim \bigcap_{k=1}^{\infty} A_k = \varnothing. \tag{10}$$

Wäre b eine strikt positive Zahl, so hätten wir $G[0, b] \succ G[0, b/2] \succsim \varnothing$, was der Ungleichung (10) widersprechen würde. Folglich gilt $b = 0$, und Lemma 4.7 ist bewiesen. □

Beweis von Satz 4.4 Gemäss (8) gilt $G[0, 1] = S \sim G\big[0, P[S]\big]$ und somit $P[S] = 1$. Lemma 4.7 garantiert die σ-Additivität von P, womit gezeigt ist, dass es sich bei P um ein Wahrscheinlichkeitsmass handelt. □

Der folgende Satz fasst die Resultate aus diesem Abschnitt zusammen und geht schliesslich noch auf die Eindeutigkeit von P ein.

Satz 4.8 *Wenn der Stichprobenraum $(S, \mathcal{A})$ und die Relation $\precsim$ die Annahmen SP1 bis SP5 erfüllen, dann ist P aus Definition 4.2 das eindeutige Wahrscheinlichkeitsmass, welches der Relation $\precsim$ entspricht.*

Beweis. Es bleibt lediglich die Eindeutigkeit von P zu zeigen. Sei P' ein weiteres Wahrscheinlichkeitsmass, das der Relation $\precsim$ entspricht. Wir zeigen im Folgenden, dass dann gelten muss

$$P'\big[G[0,a]\big] = a. \tag{11}$$

Nehmen wir vorerst an, es gäbe ein $a = \frac{p}{q} \in \mathbb{Q} \cap [0,1]$ mit $P'\big[G[0,a]\big] < a$. Wir können $[0,a)$ als disjunkte Vereinigung

$$[0,a) = \bigcup_{k=1}^{p} \big[(k-1)/q, k/q\big).$$

schreiben. Wegen der Definition der Gleichverteilung gilt dann

$$\frac{p}{q} = a > P'\big[G[0,a)\big] = P'\left[\bigcup_{k=1}^{p} G\big[(k-1)/q, k/q\big)\right] = \sum_{k=1}^{p} P'\Big[G\big[(k-1)/q, k/q\big)\Big] = p \cdot P'\big[G[0,1/q)\big],$$

also

$$P'\big[G[0,1/q)\big] < \frac{1}{q}.$$

Mit derselben Überlegung erhalten wir

$$P'[S] = P'\left[\bigcup_{k=1}^{q} G\big[(k-1)/q, k/q\big)\right] = \sum_{k=1}^{q} P'\Big[G\big[(k-1)/q, k/q\big)\Big] = q \cdot P'\big[G[0,1/q)\big) < q \cdot \frac{1}{q} = 1,$$

was im Widerspruch zu $P'[S] = 1$ steht. Analog kann man zeigen, dass $P'\big[G[0,a]\big] > a$ *nicht* gelten kann für alle $a \in \mathbb{Q} \cap [0,1]$. Folglich haben wir bewiesen, dass die Gleichung (11) für alle $a \in \mathbb{Q} \cap [0,1]$ erfüllt ist. Für ein beliebiges $a \in [0,1]$ können wir eine monoton fallende Folge $(a_k)_{k\in\mathbb{N}}$ in $\mathbb{Q} \cap [0,1]$ finden, welche gegen a konvergiert. Dann gilt $G[0,a] = \bigcap_{k=1}^{\infty} G[0,a_k]$, und wir können wegen $G[0,a_1] \supset G[0,a_2] \supset \ldots$ und der σ-Additivität von P' schreiben

$$P'\big[G[0,a]\big] = \lim_{k\to\infty} P'\big[G[0,a_k]\big] = \lim_{k\to\infty} a_k = a.$$

Demnach ist die Gültigkeit der Gleichung (11) gezeigt.
Sei nun $A \in \mathcal{A}$ ein Ereignis. Satz 4.1 garantiert uns die Existenz eines $a^* \in [0,1]$, so dass $A \sim G[0,a^*]$ gilt. Weil P' der Relation $\precsim$ entspricht, gilt also nach Gleichung (11)

$$P'[A] = P'\big[G[0,a^*]\big] = a^* = P[A].$$

Folglich stimmen P und P' überein, und die Eindeutigkeit von P ist bewiesen. □

5 Bedingte Wahrscheinlichkeit

In diesem Abschnitt erweitern wir die Betrachtung der Relation $\precsim$ in gewisser Weise. Wir betrachten nämlich nicht nur die Relation $A \precsim B$, sondern ferner auch $(A|C) \precsim (B|C)$ für Ereignisse $A, B, C \in \mathcal{A}$. Diese letzte Relation betrifft bedingte Wahrscheinlichkeiten, und zwar kann ihr die folgende Bedeutung beigemessen werden: Das Ereignis B ist mindestens so wahrscheinlich wie das Ereignis A, gegeben, dass das Ereignis C eintritt.

Im Folgenden untersuchen wir die Frage, unter welchen zusätzlichen Annahmen die bedingten Wahrscheinlichkeiten von P der erweiterten Relation entsprechen, sprich wann ist die folgende Äquivalenz garantiert:

$$(A|C) \precsim (B|C) \qquad \Longleftrightarrow \qquad P[A|C] \leq P[B|C].$$

Beachte, dass diese Relation lediglich für Ereignisse C mit $P[C] > 0$ Sinn macht. Für ein Ereignis C mit $P[C] > 0$ ist $P[A|C] \leq P[B|C]$ äquivalent zu $P[A \cap C] \leq P[B \cap C]$. Diese Ungleichung ihrerseits ist äquivalent zu $A \cap C \precsim B \cap C$. Diese Ausführungen führen direkt zu der folgenden Annahme.

SP 6 *Für je drei Ereignisse $A, B, C \in \mathcal{A}$ gilt die Äquivalenz*

$$(A|C) \precsim (B|C) \qquad \Longleftrightarrow \qquad A \cap C \precsim B \cap C.$$

Wie oben angetönt, würde es unter den Annahmen SP1 bis SP5 genügen, wenn man SP6 lediglich für Ereignisse C mit $P[C] > 0$ formulieren würde. Um aber die Annahmen unabhängig voneinander zu halten, wurde die etwas allgemeinere Formulierung gewählt. Gemäss Annahmen SP1 und SP6 ist nun mindestens eine der beiden Relationen $(A|C) \precsim (B|C)$ oder $(A|C) \succsim (B|C)$ erfüllt. Indem wir die Überlegungen dieses Abschnitts mit Satz 4.8 kombinieren, erhalten wir direkt die folgende Aussage:

Satz 5.1 *Wenn der Stichprobenraum $(S, \mathcal{A})$ und die Relation $\precsim$ die Annahmen SP1 bis SP6 erfüllen, dann ist P aus Definition 4.2 das eindeutige Wahrscheinlichkeitsmass, welches die folgende Eigenschaft erfüllt: Für je drei Ereignisse $A, B, C \in \mathcal{A}$ mit $P[C] > 0$ gilt die Äquivalenz*

$$(A|C) \precsim (B|C) \qquad \Longleftrightarrow \qquad P[A|C] \leq P[B|C].$$

Literatur

[1] Morris H. DeGroot (2004). Optimal Statistical Decisions, Ch. 6. *John Wiley & Sons Inc.*

[2] Charles H. Kraft, John W. Pratt, A. Seidenberg (1959). Intuitive Probability on Finite Sets. *Annals of Mathematical Statistics.* **30** 408-419.