



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

**„Wie man die besten MitarbeiterInnen wählt - ein
Problem aus stochastischer Sicht“**

verfasst von / submitted by

Nerma Hasanovic, BEd.

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Education (MEd)

Wien, 2024 / Vienna 2024

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 199 520 525 02

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Lehramt Sek (AB)
UF Mathematik UF Psychologie und Philosophie

Betreut von / Supervisor:

ao. Univ.-Prof. Mag. Dr. Peter Raith

Kurzfassung

Die vorliegende Arbeit beschäftigt sich mit der Frage, wie man als Chef/in einer Firma die/den besten/n Mitarbeiter/in einstellt. Dabei sind einige Voraussetzungen gegeben, wie zum Beispiel, dass der Chef/in sich direkt entscheiden muss, ob die/der Bewerber genommen wird, oder nicht.

Hinter diesem Szenario, welches besser unter dem Namen *SekretärInnen Problem* oder *Heiratsparadoxon* bekannt ist, steckt viel Stochastik. Aus diesem Grund beziehen sich die ersten Kapitel auf stochastische Grundlagen. Es werden Begriffe der elementaren Stochastik und der Maßtheorie geklärt. Der bedingte Erwartungswert und stochastische Prozesse, wie die Filtration oder die Stoppzeit, sind wichtige Aspekte, die für die Lösung der optimale Stoppzeit relevant sind.

Um das anfänglich genannte Problem zu lösen, gibt es mehrere mögliche Strategien, die auch von bestimmten Rahmenbedingungen abhängen. Im Zuge der Arbeit werden diese erklärt und anhand von Beispielen verdeutlicht.

Die wohl bekannteste Strategie bezieht sich auf ein Szenario mit einer bekannten Anzahl an BewerberInnen, die 37-Prozent-Regel von Geoffrey Miller. Die Strategie besagt, dass der Chef etwa die ersten 37% der BewerberInnen nur beobachten soll und sich den Besten davon merkt. Die/der nächste BewerberIn, die/der besser ist als die/der beste unter den ersten 37%, soll vom Chef genommen werden. Mit dieser Strategie wählt der Chef mit einer Wahrscheinlichkeit von 37% die/den beste/n Bewerber/in.

Abstract

This paper deals with the question of how the boss of a company can hire the best employee. There are a number of prerequisites, such as the boss having to decide directly whether or not to accept the applicant.

There is a lot of stochastics behind this scenario, which is better known as the *secretary problem* or *marriage paradox*. For this reason, the first chapters refer to stochastic basics. The concepts of elementary stochastics and measure theory are clarified. The conditional expected value and stochastic processes, such as filtration or stopping time, are important aspects that are relevant for solving the optimal stopping time.

There are several possible strategies for solving the problem mentioned at the beginning, which also depend on certain framework conditions. These are explained and illustrated using examples in the course of the paper.

Probably the most common strategy, the Geoffrey Miller's 37% rule, relates to a scenario with a known number of applicants. The strategy states that the boss should only observe the first 37% of applicants and remember the best one. The next applicant who is better than the best of the first 37% should be hired by the boss. With this strategy, the boss chooses the best applicant from the group with a probability of 37%.

Inhaltsverzeichnis

1	Wichtige mathematische Grundlagen	4
1.1	Folgen und Reihen	4
1.2	Elementare Stochastik	13
1.3	Maßtheorie	19
1.4	Bedingte Erwartung	27
1.4.1	Bedingte Wahrscheinlichkeit	27
1.4.2	Erwartungswert	29
1.4.3	Bedingter Erwartungswert	30
1.5	Stochastische Prozesse	32
1.5.1	Stochastische Matrix - Übergangsmatrizen	33
1.5.2	Filtration	34
1.5.3	Stoppzeit	35
2	Markovs Entscheidungstheorie zum optimalen Stoppen	40
2.1	Genauere Problemstellung	42
2.2	Die Lösung der optimalen Stoppzeit	44
3	Das Stopp-Problem	48
3.1	Trefferwahrscheinlichkeit	48
3.2	Strategie von Lindley (1961)	51
3.3	Odds-Strategie / Bruss-Algorithmus	57
3.4	x-Strategie	59
3.5	1/e-Strategie	60
3.6	Die 37-Prozent-Regel - Geoffrey Miller	60

1 Wichtige mathematische Grundlagen

Um auf die folgenden Kapitel vorzubereiten, werden in diesem Kapitel die wichtigsten stochastischen Ergebnisse und Definitionen aufgezeigt. Es folgt nach den allgemeinen mathematischen Grundlagen, die für die Arbeit von Bedeutung sind, auch ein Überblick über die elementare Stochastik, sowie ein Einblick in die Maßtheorie und die bedingte Erwartung.

Zu Beginn werden nun einige mathematische Grundlagen angeführt, die vor allem in späterer Folge der Arbeit von Nutzen sein werden.

1.1 Folgen und Reihen

Definition 1 Folgen

Eine Folge $(a_n)_{n \in \mathbb{N}}$ ist eine geordnete Folge von Zahlen a_1, a_2, a_3, \dots . Es gibt abzählbar viele Folgenglieder. Jeder natürlichen Zahl n wird ein reelles Folgenglied zugeordnet (Funktion von $\mathbb{N} \rightarrow \mathbb{R}$).

Man kann eine Folge auf verschiedene Arten darstellen.

Die rekursive Darstellung berechnet aus einem bereits bekanntem Folgenglied das nächste. Dabei ist das erste Folgenglied angegeben.

Die explizite Darstellung ermöglicht die direkte Berechnung des n -ten Folgenglieds, ohne Voraussetzung vorherige Folgenglieder zu kennen.

Bemerkung

Es gibt Spezialfälle von Folgen. Im Folgenden werden zwei Beispiele, die arithmetische und geometrische Folge, näher beschrieben.

Arithmetische Folgen

Arithmetische Folgenglieder haben eine konstante Differenz (d). Man stellt arithmetische Folgen durch lineare Funktionen dar. Folgendes Bildungsgesetz wird bei arithmetischen Folgen verwendet:

rekursive Darstellung: $a_{n+1} = a_n + d$

explizite Darstellung: $a_n = a_0 + n \cdot d$

Beispiel:

$$a_1 = 5, a_{n+1} = a_n + 5 \text{ (rekursiv)}$$

$$a_n = 5 \cdot n \text{ (explizit)}$$

$$a_n = 5, 10, 15, 20, \dots$$

Geometrische Folgen

Bei geometrischen Folgen berechnet man das nächste Folgenglied durch eine Multiplikation einer Konstanten q mit dem vorherigen Folgenglied. Geometrische Folgen stellt man mittels Exponentialfunktion dar. Man verwendet dafür folgendes Bildungsgesetz:

rekursive Darstellung: $a_{n+1} = a_n \cdot q$

explizite Darstellung: $a_n = a_0 \cdot q^{n-1}$

Beispiel:

$$a_1 = -1, a_{n+1} = a_n \cdot (-1) \text{ (rekursiv)}$$

$$a_n = (-1)^n \text{ (explizit)}$$

$$a_n = -1, 1, -1, 1, \dots$$

Notation

(i) Der Ausdruck $(a_i)_{i=1, \dots, n} = (a_1, \dots, a_n)$ steht für eine endliche Folge. Für unendliche Folgen schreibt man $(a_i)_{i \in \mathbb{N}} = (a_1, a_2, \dots)$. Die runde Klammer fasst die Folgenglieder zusammen. Hier kommt es auf die Reihenfolge an und es können mehrere Folgenglieder den gleichen Wert haben.

(ii) Der Ausdruck $\{a_i | i = 1, \dots, n\}$ steht für die Menge der Folgenglieder einer endlichen Folge. Der Ausdruck $\{a_i\}_{i \in \mathbb{N}}$ bzw. $\{a_i | i \in \mathbb{N}\}$ steht für die Menge der Folgenglieder einer unendlichen Folge. Dabei wird nicht auf die Reihe der Folgenglieder geachtet.

Definition 2 Reihen [1]

Sei $\{a_i\}_{i=n_0}^{\infty}$ eine Folge, dann heißt $s_n := \sum_{i=n_0}^n a_i$ die n -te Partialsumme dieser Folge. Die

Folge der Partialsummen $\{s_i\}_{i=n_0}^{\infty}$ heißt Reihe und wird geschrieben als $\sum_{i=n_0}^{\infty} a_i$.

Bemerkung

Summiert die Reihe die Folgenglieder einer arithmetischen Folge, so nennt man diese eine *arithmetische Reihe*.

Summiert die Reihe die Folgenglieder einer geometrischen Folge, so nennt man diese eine *geometrische Reihe*.

Eine Reihe ist endlich, wenn gilt:

$$\sum_{i=1}^n a_i = a_1 + \dots + a_n$$

Eine Reihe ist unendlich, falls:

$$\sum_{i=1}^{\infty} a_i = a_1 + \dots + a_n + a_{n+1} + \dots$$

Es existierten Folgen, deren Folgenglieder sich einer Zahl (Grenzwert) beliebig nähern, sodass fast alle Folgenglieder in jeder beliebig großen Umgebung des Grenzwerts liegen. Falls eine Folge einen solchen Grenzwert (auch Limes genannt) besitzt, nennt man die Folge *konvergent*.

Falls die Folge keinen Grenzwert hat, nennt man die Folge *divergent*.

Notation

In der vorliegenden Arbeit werden für Reihen folgende gleichwertige Notationen verwendet.

$$(i) \sum_{n=1}^{\infty} a_n \quad (ii) \sum_{n \geq 1} a_n \quad (iii) \sum_{n \in \mathbb{N}} a_n$$

Definition 3 Konvergenz von Folgen und Reihen

(i) Man nennt eine Folge $(a_n)_{n \in \mathbb{N}}$ konvergent gegen $a \in \mathbb{R}$, wenn

$$\forall \varepsilon > 0 \exists n_0 \in \mathbb{N}, \text{ sodass } |a_n - a| < \varepsilon, \forall n \geq n_0.$$

Man schreibt $\lim_{n \rightarrow \infty} (a_n)_{n \in \mathbb{N}} = a$.

Man sagt a ist der Grenzwert der Folge $(a_n)_{n \in \mathbb{N}}$.

(ii) Man nennt eine Reihe $\sum_{n=1}^{\infty} a_n$ konvergent, falls $\left(\sum_{k=1}^n a_k \right)_{n \in \mathbb{N}}$ konvergent ist. Diesen

Grenzwert nennt man auch Wert der Reihe und verwendet häufig $\sum_{n=1}^{\infty} a_n := \lim_{n \rightarrow \infty} \sum_{k=1}^n a_k$.

Bemerkung

Der folgende Satz und der dazugehörige Beweis werden zeigen, dass eine konvergente Folge nur einen eindeutigen Grenzwert (Limes) besitzen kann.

Satz 1 Eindeutigkeit des Grenzwerts

Falls der Grenzwert einer Folge existiert, so ist dieser eindeutig.

Beweis (indirekt)

Angenommen der Grenzwert einer konvergenten Folge $(a_n)_{n \in \mathbb{N}}$ ist nicht eindeutig, wodurch zwei Grenzwerte a und a' existieren und es gilt $a \neq a'$. Dadurch ergibt sich $\lim_{n \rightarrow \infty} a_n = a$ und

$$\lim_{n \rightarrow \infty} a_n = a'.$$

Aus der Definition folgt, dass ein ε mit der Bedingung

$$0 < \varepsilon < \frac{|a - a'|}{2} \Rightarrow 2\varepsilon < |a - a'|$$

existiert. Weiters muss laut Definition ein n_0 existieren, sodass

$$\forall n \geq n_0 : |a_n - a| < \varepsilon$$

und

$$\forall n \geq n_0 : |a_n - a'| < \varepsilon.$$

Daraus folgt jedoch

$$|a - a'| = |a - a_n + a_n - a'| \leq |a_n - a| + |a_n - a'| < 2\varepsilon.$$

Dabei ergibt sich ein Widerspruch zur Voraussetzung $a \neq a'$, da hier $a = a'$ gelten muss. \square

Im Laufe der Arbeit sind Kenntnisse über die Rechenregeln bei Folgen und Reihen von Vorteil. Die wichtigsten Regeln werden nun angeführt.

Satz 2 Rechenregeln für Grenzwerte von Folgen

Gegeben seien zwei konvergente Folgen $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}}$ und $c \in \mathbb{R}$ eine Konstante, dann gelten folgende Rechenregeln:

$$(i) \lim_{n \rightarrow \infty} (a_n + b_n) = \lim_{n \rightarrow \infty} (a_n) + \lim_{n \rightarrow \infty} (b_n)$$

$$(ii) \lim_{n \rightarrow \infty} c \cdot a_n = c \cdot \lim_{n \rightarrow \infty} (a_n)$$

$$(iii) \lim_{n \rightarrow \infty} a_n \cdot b_n = \lim_{n \rightarrow \infty} (a_n) \cdot \lim_{n \rightarrow \infty} (b_n)$$

$$(iv) \lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{\lim_{n \rightarrow \infty} (a_n)}{\lim_{n \rightarrow \infty} (b_n)}, \text{ wenn } b_n \neq 0, \forall n \in \mathbb{N} \text{ und } \lim_{n \rightarrow \infty} b_n \neq 0.$$

Beweis

Es gilt $a = \lim_{n \rightarrow \infty} a_n$ und $b = \lim_{n \rightarrow \infty} b_n$.

(i) Sei $\varepsilon > 0$ vorausgesetzt, dann folgt daraus auch $\frac{\varepsilon}{2} > 0$. Die Folgen $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}}$ sind konvergent, demnach gibt es laut Definition $n_a, n_b \in \mathbb{N}$, sodass gilt

$$(1) |a_n - a| < \frac{\varepsilon}{2} \text{ für } n \geq n_a \text{ und}$$

$$(2) |b_n - b| < \frac{\varepsilon}{2} \text{ für } n \geq n_b.$$

Es folgt somit für $n \geq n_a$ bzw. $n \geq n_b$, dass

$$|(a_n + b_n) - (a + b)| \leq |a_n - a| + |b_n - b| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

\square

(ii) Fasst man $(a_n)_{n \in \mathbb{N}}$ als konstante Folge auf, so ergibt sich $\lim_{n \rightarrow \infty} c \cdot a_n = c \cdot \lim_{n \rightarrow \infty} (a_n)$ aus (iii).

□

(iii) Die Folgen $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}}$ sind beschränkt. Es sei $T_a \geq a_n$ und $T_b \geq b_n$, wobei $T = \max(T_a, T_b)$ und $n \in \mathbb{N}$. Es sei $\varepsilon > 0$, dann ist auch $\frac{\varepsilon}{2T} > 0$.

Da die Folgen $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}}$ konvergent sind, existiert laut Definition $n_a, n_b \in \mathbb{N}$, sodass

$$|a_n - a| < \frac{\varepsilon}{2T}, \text{ mit } n \geq n_a \text{ und}$$

$$|b_n - b| < \frac{\varepsilon}{2T}, \text{ mit } n \geq n_b.$$

Somit gilt für $n \geq n_a$ bzw. $n \geq n_b$, dass

$$|a_n b_n - ab| = |a_n b_n - a_n b + a_n b - ab|$$

Durch herausheben ergibt sich

$$|a_n b_n - ab| = |a_n(b_n - b) + (a_n - a) \cdot b|$$

Weil $|a \cdot b| \leq |a| \cdot |b|$ gilt, folgt

$$|a_n b_n - ab| \leq |a_n| \cdot |b_n - b| + |a_n - a| \cdot |b|$$

Wegen der Voraussetzung, dass $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}}$ konvergent sind folgt schlussendlich

$$|a_n b_n - ab| < T \cdot \frac{\varepsilon}{2T} + \frac{\varepsilon}{2T} \cdot T = \varepsilon.$$

Damit wäre (iii) $\lim_{n \rightarrow \infty} a_n \cdot b_n = \lim_{n \rightarrow \infty} (a_n) \cdot \lim_{n \rightarrow \infty} (b_n)$ bewiesen.

□

(iv) Wegen $\frac{a_n}{b_n} = a_n \cdot \frac{1}{b_n}$ hilft uns (iii) beim Beweis, vorausgesetzt $\frac{1}{b_n} \rightarrow \frac{1}{b}$, mit $n \rightarrow \infty, b \neq 0$.

Es folgt aus $b \neq 0$, dass $\frac{|b|}{2} > 0$. Somit gilt für $n_b \in \mathbb{N}$, dass

$$\forall n \geq n_b : |b_n - b| < \frac{|b|}{2}.$$

Weiters gibt es zu einem beliebigen ε ein $n_0 \in \mathbb{N} : |b_n - b| < \frac{\varepsilon |b|^2}{2}, \forall n \geq n_0$.

Sei $n \geq \max(n_b, n_0)$, dann gilt

$$\left| \frac{1}{b_n} - \frac{1}{b} \right| = \frac{|b_n - b|}{|b_n| \cdot |b|} = \frac{1}{|b_n| \cdot |b|} \cdot |b_n - b|$$

Es gilt, dass $|b_n| \cdot |b| > |b| \cdot \frac{|b|}{2}$. Somit ist auch $\frac{1}{|b_n| \cdot |b|} < \frac{2}{|b|^2}$. Es folgt daher

$$\frac{1}{|b_n| \cdot |b|} \cdot |b_n - b| < \frac{2}{|b|^2} \cdot \frac{\varepsilon |b|^2}{2} = \varepsilon$$

Somit ist $\lim_{n \rightarrow \infty} \left| \frac{a}{b_n} \right| = \frac{1}{b}$ gezeigt und (iv) folgt aus (iii).

□

Nicht nur bei den Grenzwerten von Folgen gibt es bestimmte Rechenregeln, auch beim Rechnen mit konvergenten Reihen sind Rechenregeln festgelegt.

Satz 3 .

Es seien $(a_n)_{n \in \mathbb{N}}$ und $(b_n)_{n \in \mathbb{N}}$ konvergente Folgen, wobei $a_n \leq b_n$, dann gilt für die Grenzwerte

$$\lim_{n \rightarrow \infty} a_n \leq \lim_{n \rightarrow \infty} b_n.$$

Beweis

Es sei $s_n := b_n - a_n$, dann genügt es zu zeigen, dass $\lim_{n \rightarrow \infty} s_n \geq 0$, weil s_n nach Satz 2 konvergiert und aus der Voraussetzung folgt auch $s_n \geq 0, \forall n \in \mathbb{N}$. Man nimmt an, dass $\lim_{n \rightarrow \infty} s_n = -\varepsilon$ für $\varepsilon > 0$. Daraus würde folgen, dass es ein n_0 gibt, sodass für alle $n \geq n_0$ gilt

$$|s_n - (-\varepsilon)| = |s_n + \varepsilon| < \varepsilon.$$

Dies würde bedeuten, dass $s_n < 0$, was ein Widerspruch zur Annahme wäre. □

Satz 4 Rechenregeln bei konvergenten Reihen

Es seien die Reihen $\sum_{n=1}^{\infty} a_n$ und $\sum_{n=1}^{\infty} b_n$ konvergent und $c \in \mathbb{R}$ eine Konstante. Dann gelten folgende Rechenregeln.

(i)

$$\sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} (a_n + b_n) \text{ und}$$

$$\sum_{n=1}^{\infty} a_n - \sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} (a_n - b_n), \text{ wobei}$$

$$\sum_{n=1}^{\infty} (a_n + b_n) \text{ und } \sum_{n=1}^{\infty} (a_n - b_n) \text{ ebenso konvergent sind.}$$

(ii) $\sum_{n=1}^{\infty} c \cdot a_n = c \cdot \sum_{n=1}^{\infty} a_n$

(iii) Sei $a_n \leq b_n, \forall n \in \mathbb{N}$, dann gilt $\sum_{n=1}^{\infty} a_n \leq \sum_{n=1}^{\infty} b_n$.

Beweis

Für die Beweise wird Bezug auf Satz 2 genommen.

Man definiere für die Beweise $x_n = \sum_{i=1}^n a_i, y_n = \sum_{i=1}^n b_i$ und $z_n = \sum_{i=1}^n (a_i + b_i)$.

(i) Nach Satz 2 gilt für den Grenzwert (Limes) der Folge $x_n + y_n$, dass

$$\lim_{n \rightarrow \infty} x_n + \lim_{n \rightarrow \infty} y_n = \lim_{n \rightarrow \infty} (x_n + y_n) = \lim_{n \rightarrow \infty} z_n.$$

Somit gilt auch Satz 4 (i). □

(ii) Sei $c \in \mathbb{R}$ eine Konstante, dann gilt nach Satz 2, dass

$$\lim_{n \rightarrow \infty} c \cdot x_n = c \cdot \lim_{n \rightarrow \infty} x_n = c \cdot \sum_{k=1}^{\infty} a_k.$$

□

(iii) Es folgt wegen $a_i \leq b_i$, dass für alle $i \in \mathbb{N} : x_n \leq t_n, \forall n \in \mathbb{N}$.

Nach Satz 3 ergibt sich somit die Behauptung $\sum_{n=1}^{\infty} a_n \leq \sum_{n=1}^{\infty} b_n$.

□

Definition 4 Cauchy-Folge

Eine Folge in \mathbb{R} nennt man Cauchy-Folge, falls für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert, sodass

$$|a_n - a_m| < \varepsilon, \quad \forall m, n \geq n_0$$

gilt.

Satz 5 Cauchy-Kriterium

Sei $\sum_{n=1}^{\infty} a_n$ eine Reihe. Das Cauchy-Kriterium besagt, dass die Reihe genau dann konvergiert, wenn es für jedes $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ gibt, sodass gilt:

$$\left| \sum_{i=n+1}^m a_i \right| < \varepsilon, \quad \forall n, m \in \mathbb{N}, \text{ mit } m > n > n_0.$$

Beweis

Seien $s_n = \sum_{n=1}^{\infty} a_n$ und $s_m = \sum_{m=1}^{\infty} a_m$ und $m, n \in \mathbb{N}$ mit $n < m$. Es gilt somit

$$|s_m - s_n| = |a_{n+1} + a_{n+2} + a_{n+3} + \dots + a_m| = \left| \sum_{i=n+1}^m a_i \right|.$$

□

Diese Behauptung folgt aus dem Cauchy-Kriterium für Folgen, welches besagt, dass eine reelle Folge (a_n) genau dann konvergiert, wenn sie eine Cauchy Folge ist.

Definition 5 absolute Konvergenz

Neben der Konvergenz von Reihen existiert noch die absolute Konvergenz. Man sagt, dass eine Reihe $\sum_{n=1}^{\infty} a_n$ absolut konvergiert, wenn die Reihe $\sum_{n=1}^{\infty} |a_n|$ konvergiert.

Satz 6 Dreiecksungleichung

Es sei $\sum_{n=1}^{\infty} a_n$ eine absolut konvergente Reihe, dann gilt $\left| \sum_{n=1}^{\infty} a_n \right| \leq \sum_{n=1}^{\infty} |a_n|$.

Beweis

Für den Beweis der Dreiecksungleichung wird das Cauchy-Kriterium verwendet. Gegeben sei ein $\varepsilon > 0$. Wegen der absoluten Konvergenz gibt es ein $n_0 \in \mathbb{N}$, sodass $\sum_{i=n+1}^m |a_i| < \varepsilon$ für alle $m, n \in \mathbb{N}$ gilt, mit $m > n > n_0$. Es gilt dann durch die Dreiecksungleichung

$$\left| \sum_{i=n+1}^m a_i \right| = |a_{n+1} + a_{n+2} + \dots + a_m| \geq |a_{n+1}| + |a_{n+2}| + \dots + |a_m| = \sum_{i=n+1}^m |a_i| < \varepsilon.$$

Aufgrund des Cauchy-Kriteriums ist die Reihe konvergent.

Man definiere $s_n := \sum_{i=1}^n a_i$ und $x_n := \sum_{i=1}^n |a_i|$ für alle $n \in \mathbb{N}$. Durch die Dreiecksungleichung erhält man $|s_n| \leq x_n$ für alle n und es gilt somit auch

$$\left| \sum_{i=1}^{\infty} a_i \right| = \left| \lim_{n \rightarrow \infty} s_n \right| = \lim_{n \rightarrow \infty} |s_n| \leq \lim_{n \rightarrow \infty} x_n = \sum_{i=1}^{\infty} |a_i|.$$

□

Um zu entscheiden, ob eine unendliche Folge konvergent oder divergent ist, wird das Majorantenkriterium angewendet.

Es sei $S = \sum_{n=0}^{\infty} a_n$ eine unendliche Reihe mit reellen oder auch komplexen Summanden a_n .

Falls eine konvergente unendliche Reihe $T = \sum_{n=0}^{\infty} b_n$ (mit positiven Summanden) mit dem Kriterium $|a_n| \leq b_n$ existiert, dann nennt man die Reihe S *absolut konvergent*.

Man definiert außerdem die Reihe T als *Majorante von S* .

Satz 7 Majorantenkriterium

Es sei $|a_k| \leq b_k$ für alle $k \in \mathbb{N}$. Ist die Reihe $\sum_{k=1}^{\infty} b_k$ konvergent, dann ist die Reihe $\sum_{k=1}^{\infty} a_k$ absolut konvergent.

Die Reihe $\sum_{k=1}^{\infty} b_k$ nennt man somit die Majorante der Reihe $\sum_{k=1}^{\infty} a_k$.

Beweis

Sei die Reihe $\sum_{i=1}^{\infty} b_i$ konvergent, dann gibt es laut Definition für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$, sodass

gilt $\sum_{i=n}^{\infty} b_i < \varepsilon$, $\forall m \geq n > N$.

Weiters folgt wegen der Dreiecksungleichung und $|a_i| \leq b_i$, dass

$$\left| \sum_{i=i}^m a_i \right| \leq \sum_{i=n}^m |a_i| \leq \sum_{i=n}^m b_i < \varepsilon,$$

woraus die absolute Konvergenz von $\sum_{i=1}^{\infty} a_i$ folgt. \square

Manchmal ist es notwendig Reihen umzuordnen. Dies ist - wie man in Folge sehen wird - im Falle einer endlichen Reihe kein Problem. Bei unendlichen Reihen braucht es jedoch den Umordnungssatz.

Definition 6 *Umordnung von konvergenten Reihen*

Es sei $\phi : \mathbb{N} \rightarrow \mathbb{N}$ eine bijektive Abbildung, dann nennt man die Reihe $\sum_{n=1}^{\infty} a_{\phi(n)}$ Umordnung der Reihe $\sum_{n=1}^{\infty} a_n$.

Bemerkung

Es ist möglich, dass die Umordnung einer konvergenten Reihe einen anderen Grenzwert hat, als die ursprüngliche Reihe.

Im Folgenden wird der große Umordnungssatz für Reihen beschrieben und bewiesen. Dieser wird für Beweise in späteren Kapiteln verwendet werden.

Satz 8 *Großer Umordnungssatz bei Reihen*

Sei $(T_i)_{i \in \mathbb{N}}$ eine Familie von paarweise disjunkten Teilmengen von \mathbb{N} mit $\bigcup_{i=1}^{\infty} T_i = \mathbb{N}$, wobei für jedes i mit $T_i \neq \emptyset$ die Elemente von T_i beliebig angeordnet sind.

Außerdem sei $(a_n)_{n \geq 1}$ eine reelle Folge, für die gilt $\sum_{n=1}^{\infty} |a_n| < \infty$ und

$$b_i := \begin{cases} 0, & \text{falls } T_i = \emptyset \\ \sum_{n \in T_i} a_n, & \text{falls } T_i \neq \emptyset. \end{cases}$$

Unter diesen Voraussetzungen gilt der große Umordnungssatz für Reihen:

$$\sum_{n=1}^{\infty} a_n = \sum_{i=1}^{\infty} b_i.$$

Beweis

Der Beweis beschränkt sich auf die Bedingung $a_n \geq 0, n \geq 1$.

Aufgrund des Majorantenkriteriums folgt, dass für eine unendlich große Menge T_i die Reihe $\sum_{n \in T_i} a_n$ konvergiert.

(i) Es sei $\varepsilon > 0$ und es existiert für $i \geq 1$ ein $T'_i \subseteq T_i$, wobei $|T'_i| < \infty$ mit

$$b_i < \sum_{k \in T'_i} a_k + \frac{\varepsilon}{2^i}.$$

Daraus folgt, dass für $m \in \mathbb{N}$ gilt

$$\sum_{i=1}^m b_i \leq \sum_{k \in \bigcup_{i=1}^m T_i} a_k + \varepsilon \leq \sum_{n \in T} a_n + \varepsilon.$$

Damit wäre $\sum_{n \in T} a_n \geq \sum_{i=1}^{\infty} b_i$ bewiesen. □

(ii) Für jedes n gibt es ein m , sodass $\{1, 2, \dots, n\} \subseteq \bigcup_{i=1}^m T_i$. Daher ist $\sum_{k=0}^n a_k \leq \sum_{i=1}^m b_i \leq \sum_{i=1}^{\infty} b_i$.

Lässt man $n \rightarrow \infty$ gehen, so ist deshalb $\sum_{n=1}^{\infty} a_n \leq \sum_{i=1}^{\infty} b_i$.

Somit ist die Gleichung $\sum_{n=1}^{\infty} a_n = \sum_{i=1}^{\infty} b_i$ bewiesen. □

1.2 Elementare Stochastik

Definition 7 Wahrscheinlichkeitsräume

Unter Wahrscheinlichkeitsraum versteht man ein Tripel $(\Omega, \mathcal{A}, \mathbb{P})$. Ω ist dabei eine Menge und \mathcal{A} eine σ -Algebra zu Ω . \mathbb{P} bezeichnet man als das Wahrscheinlichkeitsmaß auf (Ω, \mathcal{A}) .

Definition 8 σ -Algebra

Sei Ω eine Menge und $\mathcal{P}(\Omega)$ die dazugehörige Potenzmenge. Ein beliebiges Mengensystem (eine Menge von Teilmengen) $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ nennt man σ -Algebra, falls dieses Mengensystem folgende drei Eigenschaften erfüllt:

(i) $\Omega, \emptyset \in \mathcal{A}$

Die Grundmenge und die leere Menge sind in der σ -Algebra enthalten.

(ii) $A \in \mathcal{A}$ wenn $A^c := \Omega \setminus A \in \mathcal{A}$

Die Komplemente aller bereits enthaltenen Mengen der σ -Algebra sind ebenso in der σ -Algebra enthalten.

(iii) Wenn $A_i \in \mathcal{A}$ für $i \in \mathbb{N}$, dann ist $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$

Eine abzählbare Vereinigung von Elementen der σ -Algebren liegt wieder in der σ -Algebra.

Eine σ -Algebra \mathcal{A} auf der Menge Ω ist ein System aus Teilmengen der Potenzmenge, das die möglichen Ereignisse eines Zufallsexperiments beschreibt.

Definition 9 Borelmengen und Borel- σ -Algebren

Es sei $\mathcal{B}(\mathbb{R})$ die kleinste σ -Algebra auf \mathbb{R} , die alle Intervalle enthält. Für ein Intervall $[a, b]$ sei $\mathcal{B}([a, b])$ die kleinste σ -Algebra auf $[a, b]$, die alle Teilintervalle von $[a, b]$ enthält.

Diese kleinsten σ -Algebren $\mathcal{B}(\mathbb{R})$ und $\mathcal{B}([a, b])$ nennt man Borel- σ -Algebren. Die enthaltenen Elemente nennt man Borelmengen.

Um sich die Bedeutung des Tupels (Ω, \mathcal{A}) besser vorstellen zu können wird ein Beispiel angeführt.

Beispiel

Angenommen man zieht einmal aus einer Urne mit zwei roten, einer blauen und einer grünen Kugel. Dann ergibt sich für $\Omega = \{R, B, G\}$. Die σ -Algebra \mathcal{A} gibt die Menge aller Ereignisse an, die man aus der Ergebnismenge Ω bilden kann. Man kann sich zum Beispiel die Wahrscheinlichkeit für jede Farbe einzeln berechnen, dafür benötigt man die Ereignismengen $\{R\}$, $\{B\}$ und $\{G\}$. Man kann sich aber auch fragen, wie wahrscheinlich es ist rot oder blau zu ziehen. Dabei ergibt sich die Ereignismenge $\{R, B\}$. Jede Kombination von zwei Farben ist eine Ereignismenge, also auch $\{R, G\}$ und $\{B, G\}$. Natürlich sind die Ergebnismenge $\Omega = \{R, B, G\}$ selbst und die leere Menge \emptyset auch eine Ereignisse.

Somit gilt für die σ -Algebra \mathcal{A} , dass

$$\mathcal{A} = \{\emptyset, \{R\}, \{B\}, \{G\}, \{R, B\}, \{R, G\}, \{B, G\}, \{R, B, G\}\}$$

Bemerkung

Für die σ -Algebra gilt, dass

(i) bei einer diskreten Ergebnismenge Ω die Potenzmenge \mathcal{P} von Omega als σ -Algebra genommen wird.

(ii) bei einer Ergebnismenge Ω , die Teilmenge aus \mathbb{R} oder sogar \mathbb{R} selbst ist, die Borel- σ -Algebra \mathcal{B} verwendet wird.

Definition 10 Wahrscheinlichkeitsmaß

Es sei \mathcal{A} eine σ -Algebra auf Ω . Die Abbildung $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ nennt man Wahrscheinlichkeitsmaß, falls $\mathbb{P}(\Omega) = 1$ ist und für alle Folgen (A_n) paarweise disjunkter Mengen in \mathcal{A} gilt, dass

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n).$$

Beispiel fairer Würfel

Angenommen ein fairer Würfel wird geworfen. Dabei ergibt sich für den Ergebnisraum $\Omega = \{1, 2, 3, 4, 5, 6\}$, da alle Augenzahlen von 1 bis 6 je auf einer der 6 Würfelflächen abgebildet ist. Die σ -Algebra \mathcal{A} besteht aus der Potenzmenge von Ω . Somit gilt $\mathcal{A} = \mathcal{P}(\Omega)$. Man definiert nun das Wahrscheinlichkeitsmaß \mathbb{P} wie folgt. Jede Augenzahl auf dem Würfel ist gleichwahrscheinlich. Somit gilt für $\omega = \{1, \dots, 6\}$ das Wahrscheinlichkeitsmaß

$$\mathbb{P}(\{\omega\}) := \frac{1}{6}.$$

Definition 11 Axiome von Kolmogorow

Andrei Kolmogorow entwickelte in den 1930er Jahren die axiomatische Begründung der Wahrscheinlichkeitstheorie. Demnach muss ein Wahrscheinlichkeitsmaß folgende drei Axiome erfüllen.

(i) Jedem Ereignis $A \in \mathcal{A}$ wird eine reelle Zahl $0 \leq \mathbb{P}(A) \leq 1$ als Wahrscheinlichkeit des Ereignisses A zugeordnet.

(ii) Das Ereignis $\Omega \in \mathcal{A}$ hat immer die Wahrscheinlichkeit $\mathbb{P}(\Omega) = 1$ und ist somit ein sicheres Ereignis.

(iii) Man berechnet die Wahrscheinlichkeit von Vereinigungen disjunkter Ereignisse $A_i \cap A_j = \emptyset$, $i \neq j$, indem man die Wahrscheinlichkeiten der einzelnen Ereignisse summiert. Also gilt

$$\mathbb{P}(A_1 \cup A_2 \cup \dots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots$$

Dies bezeichnet man auch als σ -Additivität.

Wichtige Beispiele zu Wahrscheinlichkeitsräumen sind die Laplace- und Bernoulliräume.

Beispiele

Laplaceräume:

In Laplaceräumen ist die Menge Ω endlich. Es gilt, dass allen Elementarereignissen die gleiche Wahrscheinlichkeit zugeordnet wird. Ein Laplace-Experiment beruht auf dem Laplace-Raum. Das Laplace-Experiment hat endlich viele Ausgänge, die gleichwahrscheinlich sind. Beispiele dafür sind Würfe mit einem fairen Würfel oder einer ungezinkten bzw. unmanipulierten Münze. Bei Laplace-Experimenten gilt allgemein für die einzelnen Ausgänge ω die Wahrscheinlichkeit

$$\mathbb{P}(\omega) = \frac{1}{|\Omega|}.$$

Für ein spezielles Ereignis A gilt die Wahrscheinlichkeit

$$\mathbb{P}(A) = \frac{\text{Anzahl der Ausgänge mit Ereignis } A}{\text{Anzahl aller möglichen Ausgänge}} = \frac{|A|}{|\Omega|}$$

Im Folgenden wird ein praktisches Beispiel mit einem fairen Würfel beschrieben. Es soll die Wahrscheinlichkeit für einen Wurf mit dem Ausgang $A = \text{„ungerade Zahl“}$ berechnet werden. Der Würfel hat 3 Seiten mit einer ungeraden Zahl. Somit gilt für die Menge der gewünschten Ereignisse $|A| = 3$. Insgesamt hat der Würfel 6 Seiten, somit ist die Ergebnismenge $|\Omega| = 6$. Setzt man in die Formel ein ergibt sich

$$\mathbb{P}(A = \text{„ungerade Zahl“}) = \frac{|A|}{|\Omega|} = \frac{3}{6} = \frac{1}{2}.$$

Es ergibt sich eine Wahrscheinlichkeit von 50% für einen Wurf mit ungerader Zahl.

Bernoulliräume:

In Bernoulliräumen gilt $\Omega = \{0, 1\}$. Für die Angabe des Wahrscheinlichkeitsmaßes ist es ausreichend die Wahrscheinlichkeit für „Erfolg“ bzw. „Misserfolg“ zu kennen, da es nur zwei mögliche Ereignisse gibt und die Wahrscheinlichkeiten dafür sich auf 1 ergänzen. Es gilt somit $\mathbb{P}(\{0\}) + \mathbb{P}(\{1\}) = 1$.

Ein praktisches Beispiel für ein Bernoulli-Experiment wäre folgendes. Angenommen man hat eine Urne mit 8 Bällen, wobei ein Ball rot und die restlichen grün sind. Man möchte sich nun die Wahrscheinlichkeit für das Ziehen des roten Balls („Erfolg“) berechnen. Es ergibt sich für die Wahrscheinlichkeiten

$$\begin{aligned} \mathbb{P}(\text{rot, „Erfolg“} = 1) &= \frac{1}{8} \\ \mathbb{P}(\text{grün, „Misserfolg“} = 0) &= \frac{7}{8} \end{aligned}$$

Aus Bernoulliräumen kann man noch weitere Verteilungen ableiten, wie zum Beispiel die geometrische und die hypergeometrische Verteilung. Außerdem auch die Binomial- und Poissonverteilung.

Definition 12 Dichtefunktion

Sei Ω ein Teilintervall von \mathbb{R} und $f : \Omega \rightarrow \mathbb{R}$ eine stetige nichtnegative Funktion, dessen Integral gleich 1 ist. Für ein Teilintervall E von Ω definiert man $\mathbb{P}(E) := \int_E f(x) dx$. Dann kann man \mathbb{P} zu einem Wahrscheinlichkeitsmaß auf $\mathcal{B}(\Omega)$ erweitern. Man nennt f die Dichtefunktion dieses Wahrscheinlichkeitsmaßes.

Definition 13 Zufallsvariable

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \rightarrow \mathbb{R}$ eine Abbildung. Diese Abbildung X nennt man Zufallsvariable, falls $X^{-1}((-\infty, x]) \in \mathcal{A}$ für alle $x \in \mathbb{R}$ gilt.

(i) Eine diskrete Zufallsvariable X nimmt nur endlich viele oder abzählbar unendlich viele Werte an.

(ii) Man nennt X eine stetige Zufallsvariable, falls es eine Dichtefunktion f gibt, sodass

$$P(\{\omega \in \Omega : X(\omega) \in E\}) = \int_E f(x) dx$$

für alle Teilintervalle $E \subset \mathbb{R}$ gilt. Eine stetige Zufallsvariable nimmt stets überabzählbar viele Werte an.

Bemerkung

Um die Wahrscheinlichkeitsverteilung einer diskreten bzw. stetigen Zufallsvariable zu beschreiben, erstellt man *Verteilungsfunktionen*. Für diskrete Zufallsvariablen können auch *Wahrscheinlichkeitsfunktionen* verwendet werden. Ebenso können für stetige Zufallsvariablen *Dichtefunktionen* für die Beschreibung verwendet werden.

Im Folgenden wird nur die Verteilungsfunktion definiert, da diese sowohl für diskrete als auch für stetige Zufallsvariablen verwendet werden kann.

Grundsätzlich enthalten sowohl die Wahrscheinlichkeits- als auch die Dichtefunktion die gleichen Informationen wie die Verteilungsfunktion. Der einzige Unterschied ist die Darstellung dieser Informationen.

Definition 14 Verteilungsfunktion von X

Es sei X eine reelle Zufallsvariable mit Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, dann nennt man die Funktion

$$F_X(x) = \mathbb{P}(X \leq x), \quad F : \mathbb{R} \rightarrow \mathbb{R}$$

Verteilungsfunktion von X .

Bemerkung

Bei stetigen Zufallsvariablen verwendet man im Allgemeinen für die Darstellung die Dichtefunktion $f(x)$, welche sich aus der Ableitung der Verteilungsfunktion ergibt.

Beispiele: Wahrscheinlichkeitsverteilungen

diskrete Verteilungen

(i) Binomialverteilung

Eine Zufallsvariable $X \sim B(n, p)$ mit den Parametern $n \in \mathbb{N}$ (Versuchszahl) und $p \in [0, 1]$ (Wahrscheinlichkeit für Erfolg) nennt man binomialverteilt mit der Verteilungsfunktion

$$F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}.$$

Dabei steht der Ausdruck $\binom{n}{k}$ für den Binomialkoeffizienten, welcher mit der folgenden Formel berechnet werden kann

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

Ein praktisches Beispiel für die Binomialverteilung ist folgendes. Bei 4-maligem Wurf eines fairen Würfels möchte man sich die Wahrscheinlichkeit für $A =$ „höchstens 2 Mal Augenzahl 6“ berechnen. Die Zufallsvariable X gibt die Anzahl der Treffer an. Für die Wahrscheinlichkeit ergibt sich durch die Verteilungsfunktion der Binomialverteilung

$$F_X(2) = \sum_{k=0}^2 \binom{4}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{4-k} \approx 0,9838$$

Die Wahrscheinlichkeit bei 4 Würfeln „höchstens 2 Mal Augenzahl 6“ zu würfeln beträgt ungefähr 98,38%.

(ii) Hypergeometrische Verteilung

Eine Zufallsvariable $X \sim H(N, M, n)$ mit den Parametern N (Anzahl von Elementen einer Grundmenge), $M \leq N$ (Anzahl von bestimmten Elementen der Grundmenge) und $n \leq N$ (Stichprobenumfang) nennt man hypergeometrisch verteilt mit der Verteilungsfunktion

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{k=0}^x \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}.$$

Die Verteilung sagt aus, mit welcher Wahrscheinlichkeit k Elemente mit der gesuchten Eigenschaft in der Stichprobe enthalten sind.

Die Hypergeometrische Verteilung verwendet man so wie die Binomialverteilung bei Zufallsexperimenten mit nur 2 möglichen Ausgängen (Erfolg, Misserfolg).

Ein typisches Beispiel für die Hypergeometrische Verteilung ist das Lotto Spiel. Man möchte sich die Wahrscheinlichkeit für einen *Sechser* im Lotto (6 Richtige aus 49 Zahlen) berechnen. Man setzt für $N = 49$ und für $M = 6$. Es wird 6 Mal gezogen, also ist $n = 6$. Es ergibt sich

$$F_X(6) = \mathbb{P}(X \leq 6) = \sum_{k=0}^6 \frac{\binom{6}{k} \binom{49-6}{6-k}}{\binom{49}{6}} \approx 0,000000715.$$

Die Wahrscheinlichkeit einen *Sechser* im Lotto zu haben liegt also bei ungefähr 0,00000715%.

(iii) Poisson-Verteilung

Eine Zufallsvariable $X \sim P(X = k)$ mit dem Parameter $\lambda > 0$ (Erwartungswert und Varianz) nennt man poisson verteilt mit der Verteilungsfunktion

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Die Verteilung gibt an, mit welcher Wahrscheinlichkeit $k \in \mathbb{N}$ eintritt, wenn λ bekannt ist. Die Poisson Verteilung verwendet man, wenn die Häufigkeit der Ausgänge eines Zufallsexperiments über einen Zeitraum betrachtet wird.

Ein Beispiel für die Anwendung der Poisson Verteilung wäre folgendes. Man möchte sich berechnen, mit welcher Wahrscheinlichkeit zwischen 13 : 00 und 13 : 15 genau 12 Kunden

die Bäckerei betreten. Man nimmt an, dass im Schnitt 10 Personen die Bäckerei zwischen 13 : 00 und 13 : 15 betreten, also ist $\lambda = 10$ in diesem Fall.

$$\mathbb{P}(X = 12) = \frac{10^{12}}{12!} e^{-10} = 0,0948$$

Es betreten mit Wahrscheinlichkeit 9,5% zwischen 13 : 00 und 13 : 15 genau 12 Kunden die Bäckerei.

stetige Verteilungen

(i) Normalverteilung

Die Normalverteilung ist auf ganz \mathbb{R} mit den Parametern $\mu \in \mathbb{R}$ und $\sigma > 0$ definiert. Die Dichtefunktion ist gegeben durch

$$f(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Ein typisches Beispiel mit Normalverteilung wäre folgendes. Man möchte beispielsweise wissen, mit welcher Wahrscheinlichkeit eine Kerze zwischen 40 und 50 Stunden brennt.

(ii) Stetige Gleichverteilung

Definiert ist die Gleichverteilung auf das Intervall $[a, b]$ mit der dazugehörigen Dichtefunktion $f(x) := 1/(b - a)$.

Bei der stetigen Gleichverteilung besteht in einem bestimmten Intervall eine konstante Wahrscheinlichkeitsdichte. Damit ist gemeint, dass jeder Wert in diesem Intervall gleich wahrscheinlich ist mit den übrigen Werten des Intervalls.

Beispielsweise ist die Wartezeit auf den Bus, der jede halbe Stunde fährt stetig gleichverteilt. Geht man zu einem beliebigen Zeitpunkt zur Busstation, so ist jede mögliche Wartezeit im Intervall $[0, 30]$ gleich wahrscheinlich.

(iii) Exponentialverteilung

Die Dichtefunktion der Exponentialverteilung als $f(x) := \lambda \cdot e^{-\lambda x}$ auf \mathbb{R}^+ mit dem Parameter $\lambda > 0$ definiert. Charakteristisch für diese Wahrscheinlichkeitsverteilung ist das gedächtnislose Warten.

1.3 Maßtheorie

Für weiteren Verlauf der Arbeit ist es von Bedeutung zu wissen, dass man sich hierbei nicht auf diskrete Räume, wie es in der elementaren Stochastik ist, beschränken kann. Man benötigt den Maßbegriff, da die Räume komplizierter sind.

Vorerst gilt es die verschiedenen Bedeutungen einer σ -Algebra zu klären, da diese im Folgenden besonders bedeutsam sind. Man unterscheidet zwischen den von Mengensystemen erzeugten σ -Algebren und den von Zufallsvariablen erzeugten σ -Algebren.

σ -Algebra - vom Mengensystem erzeugt

Es sei \mathcal{B} ein Mengensystem auf der Menge M . Dann schreibt man $\sigma(\mathcal{B})$ für die kleinste σ -Algebra auf M , die \mathcal{B} enthält. Man nennt $\sigma(\mathcal{B})$ die von \mathcal{B} erzeugte σ -Algebra.

σ -Algebra - von Zufallsvariable erzeugt

Sei X eine Zufallsvariable mit Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Für die Borelmengen \mathcal{B} , die nicht notwendig echte Teilmengen von \mathbb{R} sind, gilt $X^{-1}(\mathcal{B}) \in \mathcal{A}$. Man bezeichnet die kleinste σ -Algebra, die alle Mengen der Form $X^{-1}(\mathcal{B})$ für Borelmengen \mathcal{B} enthält mit $\sigma(X)$, und nennt diese die von X erzeugte σ -Algebra.

Seien X_0, \dots, X_n Zufallsvariablen mit Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Man bezeichnet $\sigma(X_0, \dots, X_n)$ als kleinste σ -Algebra, die $\sigma(X_0), \dots, \sigma(X_n)$ enthält.

Beispiele:

(i) Bei konstantem X ergibt sich $\sigma(X)$ als die triviale σ -Algebra $\{\emptyset, \Omega\}$.

(ii) Gegeben sei $X(n) := n^2$ mit $n \in \mathbb{Z}$, so ergeben alle symmetrische Teilmengen von \mathbb{Z} die kleinste σ -Algebra $\sigma(X)$.

Die folgenden Sätze sind für das nächste Kapitel sehr bedeutsam.

Satz 9 .

Es seien Y_1, \dots, Y_n Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$.

(i) Ist $g : \mathbb{R}^n \rightarrow \mathbb{R}$ eine Borelfunktion und definiert man $X : \Omega \rightarrow \mathbb{R}$ durch

$$\omega \mapsto g \circ (Y_1, \dots, Y_n)(\omega) = g(Y_1(\omega), \dots, Y_n(\omega)),$$

so ist X eine Zufallsvariable, für die $\sigma(X) \subset \sigma(Y_1, \dots, Y_n)$ gilt.

(ii) Umgekehrt gilt das auch: Ist $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable, die der Bedingung $\sigma(X) \subset \sigma(Y_1, \dots, Y_n)$ genügt, so gibt es eine Borelfunktion g mit der Eigenschaft $X = g \circ (Y_1, \dots, Y_n)$.

Der Beweis zu Satz 9 ist für die Arbeit nicht von Bedeutung. Man findet ihn zum Nachlesen im Buch *Markovprozesse und stochastische Differentialgleichungen* von Ehrhard Behrends [2].

Definition 15 Banachraum

Ein Banachraum bezeichnet einen vollständig normierten Raum $(X, \|\cdot\|)$. Dies ist ein Vektorraum X über dem Körper \mathbb{K} mit einer Norm $\|\cdot\|$, in dem jede Cauchy-Folge aus Elementen von X in der von der Norm induzierten Metrik $d(x, y) = \|x - y\|$ konvergiert.

Bemerkung Bochner-Integral

Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein σ -endlicher und vollständiger Maßraum und $(B, \|\cdot\|)$ ein Banachraum. Das Bochner-Integral $v(E) = \int_E Y d\mathbb{P}$ einer Funktion $Y : \Omega \rightarrow \mathbb{R}^+$ ist eine allgemeinere Form des Lebesgue-Integrals auf Funktionen mit Werten aus einem Banachraum.

Satz 10 von Radon-Nikodym

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Angenommen es gibt ein weiteres endliches Maß $v : \mathcal{A} \rightarrow [0, 1]$ für das ebenso aus $\mathbb{P}(E) = 0$ stets folgt, dass $v(E) = 0$, dann existiert eine Zufallsvariable $Y : \Omega \rightarrow \mathbb{R}^+$, sodass für alle $E \in \mathcal{A}$ gilt, dass

$$v(E) = \int_E Y d\mathbb{P}.$$

Der Satz von Radon-Nikodym gibt an, wann ein Maß mittels Lebesgue-Integral einer Funktion dargestellt werden kann und ist somit sowohl für die Wahrscheinlichkeitstheorie, als auch für die Maßtheorie von Bedeutung.

Der Beweis von Satz 10 ist im Buch *Markovprozesse und stochastische Differentialgleichungen* von Ehrhard Behrends [2] zu finden. Für die Arbeit reicht es, den Beweis für die Konvergenz von absolut konvergenten Reihen zu beweisen. In Kapitel 1.1 wurde dies im Beweis des Majorantenkriteriums bereits gezeigt.

Definition 16 Urbild

Gegeben sei eine Abbildung $f : \Omega \rightarrow \Omega'$. Das Urbild von $B \subset \Omega'$ unter f ist

$$f^{-1}[B] = \{\omega \in \Omega \mid f(\omega) \in B\}.$$

Bemerkung

Das Urbild ist nicht die Umkehrabbildung. Das Urbild existiert immer, egal ob f injektiv bzw. surjektiv ist. Die Umkehrabbildung existiert nur, wenn f bijektiv ist.

Definition 17 messbare Abbildung

Es seien (Ω, \mathcal{A}) und (Z, \mathcal{Z}) zwei messbare Räume. Eine Abbildung $f : \Omega \rightarrow Z$ nennt man messbar, bzw. \mathcal{A} - \mathcal{Z} -messbar, falls gilt:

$$B \in \mathcal{Z} \Rightarrow f^{-1}(B) \in \mathcal{A}.$$

Falls also die Urbilder der messbaren Mengen messbar sind.

Definition 18 Indikatorfunktion

Sei $A \in \mathcal{A}$, dann nennt man die Funktion

$$I_A(\omega) = \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{falls } \omega \notin A. \end{cases}$$

Indikatorfunktion der Menge A .

Definition 19 *primitive Funktion*

Man nennt eine Funktion der Form

$$f = \sum_{i=1}^{\infty} a_i I_{A_i}$$

primitive Funktion bzw. Treppenfunktion, wobei gilt $a_i \in \mathbb{R}$, $A_i \in \mathcal{A}$, $A_i \cap A_j = \emptyset$ mit $i \neq j$ ($i, j = 1, 2, \dots$).

Bemerkung

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und X eine Zufallsvariable und eine messbar reellwertige Funktion auf den Grundraum Ω . Dann gilt für jede Borel-Menge \mathcal{B} , dass

$$[\omega : X(\omega) \in \mathcal{B}] = X^{-1}(B) \in \mathcal{A}.$$

Außerdem ist jede nichtnegative diskrete Zufallsvariable X als

$$X(\omega) = \sum_{k=1}^{\infty} a_k I_{A_k}(\omega)$$

darstellbar. Dabei sind a_k mit $k = 1, 2, \dots$ die für die X möglichen Werte gemeint und die Mengen $A_k = \{\omega : X(\omega) = x_k\}$ sind paarweise disjunkt. Für die Vereinigung der Mengen gilt $\bigcup_{k=1}^{\infty} A_k = \Omega$.

Definition 20 .

(i) Sei $X(\omega) = \sum_{k=1}^{\infty} a_k I_{A_k}(\omega)$ eine nichtnegative primitive Funktion, dann wird die Zahl

$$\int X d\mathbb{P} = \int X(\omega) \mathbb{P}(d\omega) = \sum_{k=1}^{\infty} a_k \mathbb{P}(A_k)$$

das Integral dieser nicht primitiven Funktion genannt, falls

$$\sum_{k=1}^{\infty} a_k \mathbb{P}(A_k) < +\infty,$$

ansonsten $+\infty$.

(ii) Sei $X(\omega)$ eine nichtnegative primitive Funktion über $A \in \mathcal{A}$, dann wird der Ausdruck

$$\int_A X d\mathbb{P} = \int X(\omega) I_A(\omega) \mathbb{P}(d\omega)$$

als Integral dazu bezeichnet.

Bemerkung

Sei $X(\omega) = I_A(\omega)$, dann gilt für das Integral

$$\int I_A(\omega) \mathbb{P}(d\omega) = \mathbb{P}(A).$$

Das Integral $\int X(\omega) \mathbb{P}(d\omega)$ wird in der Wahrscheinlichkeitstheorie auch *Erwartungswert* $\mathbb{E}(X)$ der Zufallsvariable X genannt.

Definition 21 .

Man nennt die Zahl

$$\mathbb{E}(X) = \int X \, d\mathbb{P} = \lim_{n \rightarrow \infty} \int_{\{X \leq n\}} X \, d\mathbb{P}$$

das Integral der messbaren Funktion $X(\omega) \geq 0$.

Bemerkung

Das Integral $\mathbb{E}(X)$ existiert, wenn die oben definierte Zahl endlich ist. Unter der Bedingung, dass das Integral von $\mathbb{E}(X)$ existiert, nennt man die Funktion X *integrierbar*.

Falls die oben definierte Zahl unendlich ist, so nennt man die Funktion X *quasi-integrierbar*.

Falls die Funktion $X(\omega)$, sowohl negative wie auch positive Werte annehmen kann, so definiert man

$$\mathbb{E}(X) = \int X \, d\mathbb{P} = \int X^+ \, d\mathbb{P} - \int X^- \, d\mathbb{P} = \mathbb{E}(X^+) - \mathbb{E}(X^-),$$

wenn eines der Integrale $\mathbb{E}(X^+)$ bzw. $\mathbb{E}(X^-)$ existiert.

Falls beide Integrale $\mathbb{E}(X^+)$ und $\mathbb{E}(X^-)$ existieren, so sagt man, dass das Integral $\mathbb{E}(X)$ existiert und die Funktion X integrierbar ist.

Die Funktion X ist quasi-integrierbar, wenn $\mathbb{E}(X)$ eindeutig bestimmt - also $\mathbb{E}(X^+)$ oder $\mathbb{E}(X^-)$ endlich ist - aber nicht endlich ist.

Eigenschaften

Seien die Mengen $A_j \in \mathcal{A}$ paarweise disjunkt und sei $\bigcup_{j=1}^{\infty} A_j = \Omega$, dann gilt

(i) $\int X \, d\mathbb{P} = \sum_{j=1}^{\infty} \int_{A_j} X \, d\mathbb{P}.$

(ii) $\int (X + Y) \, d\mathbb{P} = \int X \, d\mathbb{P} + \int Y \, d\mathbb{P}.$

(iii) Es sei a eine beliebige Konstante, dann gilt $\int aX \, d\mathbb{P} = a \int X \, d\mathbb{P}.$

(iv) $X \leq Y \Rightarrow \int X \, d\mathbb{P} \leq \int Y \, d\mathbb{P}.$

Beweis

Die Beweise werden teilweise aus dem Buch *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie* [3] von H. Bauer zitiert.

(i) Hier ist es ausreichend den Beweis für nichtnegative Funktionen $X(\omega)$ durchzuführen, da die Eigenschaft für primitive Funktionen offensichtlich ist wegen

$$\int X \, d\mathbb{P} = \sum_k a_k \mathbb{P}(X = a_k) = \sum_{j=1}^{\infty} \sum_k a_k \mathbb{P}(\{X = a_k\} \cap A_j).$$

Da es sich hier um unendliche Reihen handelt wird die Umordnung der Summen durch den großen Umordnungssatz bei Reihen (Satz 8) begründet.

Es gilt im allgemeinem Fall

$$\int X \, d\mathbb{P} = \lim_{n \rightarrow \infty} \int X_n \, d\mathbb{P} = \lim_{n \rightarrow \infty} \sum_{j=1}^{\infty} \int_{A_j} X_n \, d\mathbb{P} = \sum_{j=1}^{\infty} \lim_{n \rightarrow \infty} \int_{A_j} X_n \, d\mathbb{P} = \sum_{j=1}^{\infty} \int_{A_j} X \, d\mathbb{P}$$

Das Vertauschen der Summe und des Grenzwertes ist hier erlaubt und wird wie folgt argumentiert. Man setzt $c = \sup X(\omega)$. Dann gilt für $N \rightarrow \infty$ gleichmäßig in n , dass

$$\sum_{j=N}^{\infty} \int_{A_j} X_n \, d\mathbb{P} \leq \sum_{j=N}^{\infty} \int_{A_j} X \, d\mathbb{P} \leq c \mathbb{P}\left(\bigcup_{j=N}^{\infty} A_j\right) \rightarrow 0.$$

□

(ii) Der Beweis für die zweite Eigenschaft ergibt sich aus den folgenden Bemerkungen, die aus dem Buch *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie* [3] von H. Bauer zitiert werden.

1. Die Eigenschaft ist für primitive Funktionen trivial. Wenn $X \geq 0, Y \geq 0$, dann gilt die Eigenschaft, weil, wie in (i) begründet, Summe und Grenzübergänge vertauscht werden können.

2. Sei $X = X_1^+ - X_1^-$, wobei X_1^+ und X_1^- nichtnegativ und beschränkt sind, so gilt

$$\int X \, d\mathbb{P} = \int X_1^+ \, d\mathbb{P} - \int X_1^- \, d\mathbb{P}.$$

In der Tat gilt $X_1^+ \geq X^+ = \max(0, X)$, $X_1^- \geq X^- = \max(0, -X)$. Also existiert eine messbare Funktion $Z \geq 0$ derart, dass $X_1^{\pm} = X^{\pm} + Z$ ist, und es gilt

$$\begin{aligned} \int X \, d\mathbb{P} &= \int X^+ \, d\mathbb{P} - \int X^- \, d\mathbb{P} + \int Z \, d\mathbb{P} - \int Z \, d\mathbb{P} \\ &= \int X_1^+ \, d\mathbb{P} - \int X_1^- \, d\mathbb{P}. \end{aligned}$$

Im allgemeineren Falle gilt

$$\begin{aligned} \int (X + Y) \, d\mathbb{P} &= \int [(X^+ + Y^+) - (X^- + Y^-)] \, d\mathbb{P} = \int (X^+ + Y^+) \, d\mathbb{P} - \int (X^- + Y^-) \, d\mathbb{P} \\ &= \int X^+ \, d\mathbb{P} - \int X^- \, d\mathbb{P} + \int Y^+ \, d\mathbb{P} - \int Y^- \, d\mathbb{P} = \int X \, d\mathbb{P} + \int Y \, d\mathbb{P}. \end{aligned}$$

□

Die Beweise für die Eigenschaften (iii) und (iv) sind trivial.

Bemerkung

Die eben oben angeführten Eigenschaften zu den Integralen (i) - (iv) können ebenso als Eigenschaften von Erwartungswerten angeschrieben werden. Im Kapitel *Erwartungswert* werden diese näher angeführt.

Aus den bisherigen Eigenschaften lassen sich noch folgende weitere Eigenschaften ableiten. [3]

(v) $|\mathbb{E}(X)| \leq \mathbb{E}(|X|)$, da $|\mathbb{E}(X)| \leq \mathbb{E}(X^+) + \mathbb{E}(X^-) = \mathbb{E}(|X|)$.

(vi) Aus $C_1 \leq X \leq C_2$ folgt $C_1 \leq \mathbb{E}(X) \leq C_2$.

(vii) Aus $X \geq 0$ und $\mathbb{E}(X) = 0$ folgt $\mathbb{P}(X = 0) = 1$.

(viii) Aus $\mathbb{P}(X = Y) = 1$ und der Quasiintegrierbarkeit von X folgt die Gleichung $\mathbb{E}(X) = \mathbb{E}(Y)$.

(ix) Aus $\int_A X d\mathbb{P} = 0$, für alle $A \in \mathcal{A}$ folgt $X = 0$.

Die dazugehörigen Beweise sind ebenso im Buch *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie* [3] von H. Bauer zu finden.

Definition 22 Konvergenzarten

Es seien X, X_1, X_2, \dots Zufallsvariablen im metrischen Raum (E, d) .

(i) (X_n) konvergiert fast sicher gegen X , wenn gilt

$$\lim_{n \rightarrow \infty} d(X_n, X) = 0.$$

Notation: $X_n \xrightarrow[n \rightarrow \infty]{f.s.} X$

(ii) (X_n) konvergiert stochastisch gegen X , wenn für alle $\varepsilon > 0$ gilt, dass

$$\lim_{n \rightarrow \infty} \mathbb{P}(d(X_n, X) > \varepsilon) = 0.$$

Notation: $X_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} X$

Satz 11 .

(i) Es gilt, dass aus der fast-sicheren Konvergenz die stochastische Konvergenz folgt.

(ii) Cauchy-Kriterium:

Es gilt, dass die fast-sichere Konvergenz äquivalent ist zu

$$\lim_n \mathbb{P}\left(\bigcup_{k \geq n} \{|Z_k - Z| > \varepsilon\}\right) = 0.$$

Für den Beweis von Satz 11 benötigt man folgendes Korollar.

Korollar 1 .

Wenn $Z_n \xrightarrow{\mathbb{P}} Z$, dann existiert eine Teilfolge $Z_{n(j)}$, die fast sicher gegen Z konvergiert.

Beweis

Man wählt $n(j)$ so, dass $\mathbb{P}(|Z_{n(j)} - Z| \geq \frac{1}{j}) \leq \frac{1}{j^2}$. Aus der Behauptung (ii) von Satz 11 ergibt sich

$$\lim_n \mathbb{P} \left(|Z_{n(j)} - Z| \geq \frac{1}{j} \right) = 0.$$

□

Beweis von Satz 11

Die Aussagen (i) und (ii) werden in einem gemeinsamen Beweis gezeigt.

Es gilt

$$1 = \mathbb{P}(\lim_n Z_n = Z) = \mathbb{P} \left(\bigcap_{\varepsilon > 0} \bigcup_{n \geq 1} \bigcap_{k \geq n} \{|Z_k - Z| \leq \varepsilon\} \right).$$

Für die Mengen $\bigcap_{k \geq n} \{|Z_k - Z| \leq \varepsilon\}$ gilt, dass sie monoton wachsend sind in n und ε . Dies ist für den Verlauf des Beweises von Nutzen. \mathbb{P} ist σ -stetig, somit folgt für alle $\varepsilon > 0$, dass

$$1 = \mathbb{P} \left(\bigcup_{n \geq 1} \bigcap_{k \geq n} \{|Z_k - Z| \leq \varepsilon\} \right).$$

Außerdem folgt aus der σ -Stetigkeit, dass $\mathbb{P} \left(\bigcap_{\varepsilon > 0} \bigcup_{n \geq 1} \bigcap_{k \geq n} \{|Z_k - Z| \leq \varepsilon\} \right)$ äquivalent zu

$$\lim_n \mathbb{P} \left(\bigcap_{k \geq n} \{|Z_k - Z| \leq \varepsilon\} \right) = 1$$

ist, für alle $\varepsilon > 0$. Daraus folgt wiederum, dass

$$0 = \lim_n \mathbb{P} \left(\bigcap_{k \geq n} \{|Z_k - Z| \geq \varepsilon\} \right).$$

Somit wäre die Behauptung (ii) bewiesen.

Es gilt außerdem

$$\lim_n \mathbb{P} \left(\bigcap_{k \geq n} \{|Z_k - Z| \leq \varepsilon\} \right) \geq \lim_n \mathbb{P}(|Z_n - Z| > \varepsilon)$$

woraus die stochastische Konvergenz folgt.

□

Satz 12 *Der dominierten Konvergenz (Lebesgue)*

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Sei X_n eine Folge von Zufallsvariablen, die für $n \rightarrow \infty$ fast überall bzw. stochastisch gegen X konvergiert. Sei $|X_n| \leq Y$. Es folge aus der Existenz für $\mathbb{E}Y$ die Existenz von $\mathbb{E}X$ und

$$\mathbb{E}X = \lim_{n \rightarrow \infty} \mathbb{E}X_n.$$

Beweis

Dieser Beweis wird aus dem Buch *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie* [3] von Bauer H. zitiert.

Das Integral $\mathbb{E}X$ existiert, da $|X| \leq Y$ fast überall und

$$\mathbb{E}|X| = \lim_{n \rightarrow \infty} \int_{|X| \leq n} |X| d\mathbb{P} \leq \lim_{n \rightarrow \infty} \int_{|X| \leq n} Y d\mathbb{P} \leq \mathbb{E}Y < \infty$$

ist. Aus der Konvergenz $X_n \xrightarrow{f.s.} X$, folgt die stochastische Konvergenz, das heißt die Gültigkeit der Beziehung $\mathbb{P}(|X_n - X| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$ für jedes $\varepsilon > 0$.

Es gilt für jedes $\delta > 0$, für genügend große n , dass

$$\int_{|X - X_n| > \varepsilon} Y < \delta.$$

Daraus ergibt sich

$$\begin{aligned} |\mathbb{E}X_n - \mathbb{E}X| &\leq \int_{|X - X_n| \leq \varepsilon} |X_n - X| + \int_{|X - X_n| > \varepsilon} |X_n - X| \\ &\leq \varepsilon + 2 \int_{|X - X_n| > \varepsilon} Y \leq \varepsilon + 2\delta \end{aligned}$$

□

Der Satz von der dominierten Konvergenz (auch Satz von der majorisierenden Konvergenz genannt) zählt zu einem der wichtigsten Grenzwertaussagen der Maßtheorie. Er bringt das Entscheidungskriterium zur Vertauschung der Bildung von Integral und Grenzwert.

1.4 Bedingte Erwartung

Bevor man über den bedingten Erwartungswert sprechen kann, gilt es noch die zwei Ausdrücke *bedingte Wahrscheinlichkeit* und *Erwartungswert* zu begreifen.

1.4.1 Bedingte Wahrscheinlichkeit

Um den Begriff der bedingten Wahrscheinlichkeit besser zu veranschaulichen, stellt man sich folgendes wiederholbares Zufallsexperiment vor.

Das Zufallsexperiment ist durch den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definiert. Es sei bekannt, dass das Ereignis $B \in \mathcal{A}$ eingetreten ist. Somit gilt $\omega \in B$. Im Folgenden nennt man diese Information *Bedingung* B . Es sei $A \in \mathcal{A}$ ebenso ein Ereignis. Nun möchte man wissen, wie hoch die Wahrscheinlichkeit für das Eintreten von A ist, unter der Bedingung, dass B

eingetreten ist. Man berechnet also dafür die Wahrscheinlichkeit für das Ereignis A unter der Bedingung B .

Notation

Man schreibt

$$\mathbb{P}(A|B)$$

für die Wahrscheinlichkeit von A unter der Bedingung B .

Definition 23 [4]

Es seien, $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $B \in \mathcal{A}$ ein Ereignis mit $\mathbb{P}(B) > 0$. Dann heißt

$$P(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

für die bedingte Wahrscheinlichkeit von A unter der Bedingung B .

Dass durch

$$\mathbb{P}_B(A) := \mathbb{P}(A|B), A \in \mathcal{A},$$

definierte Wahrscheinlichkeitsmaß auf \mathcal{A} heißt bedingte Verteilung von \mathbb{P} unter der Bedingung B .

Eigenschaften - bedingte Wahrscheinlichkeit

Für die Wahrscheinlichkeit von A unter der Bedingung B , $\mathbb{P}(A|B)$ gelten folgende Eigenschaften.

- (i) $0 \leq \mathbb{P}(A|B) \leq 1$
- (ii) $\mathbb{P}(A|B) = 1$, falls $B \subseteq A$
- (iii) $\mathbb{P}(A|B) = 0$, falls $A \cap B = \emptyset$

Beweis der Eigenschaften

Für die Beweise der Eigenschaften der bedingten Wahrscheinlichkeit wird Bezug zu den Axiomen von Kolmogorow genommen.

- (i) Es gilt $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$, also ist $\mathbb{P}(A|B)$ ein Quotient zweier Wahrscheinlichkeiten, für die jeweils Axiom (i) gilt, also $\mathbb{P}(A \cap B) \geq 0$. Somit gilt auch für $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geq 0$. Daraus ergibt sich $\mathbb{P}(A \cap B) \geq 0$ und $\mathbb{P}(B) \geq 0$. Außerdem sind $A \cap B$ und $B \setminus A$ disjunkt und die Vereinigung somit B . Somit gilt nach Axiom (iii), dass

$$\mathbb{P}(A \cap B) = \mathbb{P}(B) - \mathbb{P}(B \setminus A).$$

Weil $\mathbb{P}(A \cap B) \geq 0$ folgt $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ und somit $\mathbb{P}(A|B) \leq 1$. □

(ii) Diese Eigenschaft folgt direkt aus

$$\mathbb{P}(B|B) = \frac{\mathbb{P}(B \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1.$$

□

(iii) Es gilt $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = 0$, da $A \cap B = \emptyset$.

□

1.4.2 Erwartungswert

Einer der grundlegenden Kenngrößen von Verteilungen ist der *Erwartungswert*. Um die Definition besser zu verstehen, stellt man sich ein Zufallsexperiment in Form eines Glücksspiels vor, welches durch den Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definiert ist. Der Grundraum $\Omega = \{\omega_1, \dots, \omega_s\}$ erfasst alle möglichen Ausgänge dieses Glücksspiels. \mathcal{A} ist die Potenzmenge von Ω . Das Ergebnis ω_j tritt mit der Wahrscheinlichkeit p_j auf. Daraus ergibt sich außerdem $p_1 + \dots + p_s = 1$. Die Wahrscheinlichkeiten aller einzelnen Ereignisse summieren sich wie üblich auf 1 und es entsteht ein endlicher Wahrscheinlichkeitsraum mit

$$\mathbb{P}(A) := \sum_{j:\omega_j \in A} p_j, \quad A \in \mathcal{A}.$$

Bei dem Spiel bekommt man beim Ergebnis ω_j genau $X(\omega_j)$ Euro ausbezahlt. Wiederholt die/der Spieler/in das Glücksspiel n -mal und das Ergebnis ω_j tritt h_j -mal auf, so gilt ($h_j \geq 0, h_1 + \dots + h_s = n$) und der Gesamtgewinn nach n Spielen beträgt $\sum_{j=1}^s X(\omega_j) \cdot h_j$ Euro.

Für den durchschnittlichen Gewinn erhält man $\sum_{j=1}^s \frac{X(\omega_j) \cdot h_j}{n}$ Euro. Der Quotient h_j/n nähert sich durch die Stabilisierung von relativen Häufigkeiten bei steigendem n der Wahrscheinlichkeit $\mathbb{P}(\{\omega_j\})$. Somit ergibt sich für den langfristig erwarteten Gewinn pro Glücksspiel

$$\sum_{j=1}^s X(\omega_j) \cdot \mathbb{P}(\{\omega_j\}) \text{ Euro.}$$

Mathematisch gesehen steht diese Summe für den *Erwartungswert* von der Zufallsvariablen X mit Abbildung auf Ω .

Definition 24 .

Sei $X : \Omega \rightarrow \mathbb{R}$ eine diskrete Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, die die Werte $(\alpha_j)_{j \in J}$ annimmt.

Falls $\sum_{j \in J} |\alpha_j| \mathbb{P}(X = \alpha_j) < \infty$, dann sagen wir, dass der Erwartungswert $\mathbb{E}(X)$ von X existiert.

$$\mathbb{E}(X) := \mathbb{E}_{\mathbb{P}}(X) := \sum_{j \in J} \alpha_j \mathbb{P}(X = \alpha_j)$$

nennt man Erwartungswert von X (bezüglich \mathbb{P}).

Eigenschaften - Rechenregeln

Seien X und Y zwei Zufallsvariablen mit je einem existierenden Erwartungswert. Sei $a \in \mathbb{R}$. Dann existiert auch die Summe der Erwartungswerte $X + Y$ und das Produkt aX . Es gelten dafür folgende Eigenschaften:

- a) Homogenität $\mathbb{E}(aX) = a\mathbb{E}$,
- b) Additivität: $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,
- c) $\mathbb{E}(1_A) = \mathbb{P}(A)$, $A \in \mathcal{A}$,
- d) Monotonie: $X \leq Y \Rightarrow \mathbb{E}(X) \leq \mathbb{E}(Y)$,
- e) Dreiecksungleichung: $|\mathbb{E}(X)| \leq \mathbb{E}|X|$.

1.4.3 Bedingter Erwartungswert

Beim bedingten Erwartungswert geht es sowohl in der Wahrscheinlichkeitstheorie, als auch in der Statistik um den zu erwartenden Wert einer Zufallsvariablen, welche weitere Informationen zum Ausgang des Zufallsexperiments voraussetzt. Die Bedingung kann zum Beispiel ein bestimmtes Eintreten eines Ereignisses sein. Außerdem kann man als Bedingung auch Werte einer zusätzlichen Zufallsvariable setzen. Diese zusätzliche Information stellt man sich als Unterraum des ursprünglichen Ereignisraums vor.

Besonders die neuere Stochastik beschäftigt sich viel mit bedingten Erwartungswerten. Unter anderem werden diese bei Untersuchungen von stochastischen Prozessen angewandt.

Erklärung der Funktionsweise

Man bildet den bedingten Erwartungswert in gewisser Hinsicht mittels einer "Glättung" der Zufallsvariable auf einer Teilmenge der σ -Algebra. Durch σ -Algebren werden die zur Verfügung gestellten Informationen modelliert. Die geglättete Zufallsvariable, die man auf einer Teilmenge der σ -Algebra messen kann, enthält durch die "Glättung" weniger Information zum Ausgang des Zufallsexperiments. Die bedingte Erwartung reduziert also den Informationsgehalt über die Zufallsvariable und kreiert eine besser messbare Zufallsvariable.

Im Vergleich dazu kann man den Erwartungswert betrachten. Der Erwartungswert einer Zufallsvariablen reduziert die Information der Zufallsvariablen genau auf einen einzelnen Wert - den Erwartungswert.

Definition 25 *bedingter Erwartungswert*

Sei $B \subset \Omega$ mit $\mathbb{P}(B) > 0$. Sei Ω endlich und die Potenzmenge von Ω sei die σ -Algebra. Dann nennt man

$$\mathbb{E}(X|B) := \frac{1}{\mathbb{P}(B)} \sum_{\omega \in B} X(\omega) \mathbb{P}(\{\omega\})$$

den bedingten Erwartungswert von X unter der Bedingung B .

Beispiel

Sei X_j die Augenzahl des j -ten Wurfs und $j = 1, 2$.

Gesucht ist der Erwartungswert, dass der erste Wurf geworfen wird unter der Bedingung, dass die Augensumme der zwei Würfe größer gleich 9 ist, also $\mathbb{E}(X_1|\{X_1 + X_2 \geq 9\})$.

$$\Omega := \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$$

$$B := \{X_1 + X_2 \geq 9\}$$

$$B = \{(3, 6), (4, 5), (5, 4), (6, 3), (4, 6), (5, 5), (6, 4), (5, 6), (6, 5), (6, 6)\}$$

$$\mathbb{P}(B) = \frac{10}{36}$$

$$\mathbb{P}(\{\omega\}) = \frac{1}{36}, \omega \in \Omega$$

$$\mathbb{E}(X_1|B) = \mathbb{E}(X_1|X_1 + X_2 \geq 9)$$

$$\mathbb{E}(X_1|B) = \frac{1}{10/36} \cdot \frac{1}{36} \cdot (3 + 4 + 5 + 6 + 4 + 5 + 6 + 5 + 6 + 6)$$

$$\mathbb{E}(X_1|B) = 5$$

Es ergibt sich ein Erwartungswert von 5. Das bedeutet, dass 5 der erwartete Wert für den 1. Wurf ist, wenn die Augensumme der beiden Würfe größer oder gleich 9 ist.

Definition 26 .

Sei $\omega \in \Omega$. Sei X eine Zufallsvariable mit $\mathbb{E}(X) < \infty$. Sei $Z : \Omega \rightarrow \mathbb{R}^k$ ein diskreter Zufallsvektor, der die Vektoren z_1, z_2, \dots annimmt, wobei gilt $z_j \in \mathbb{R}^k$ und $P(Z = z_j) > 0$. Seien X, Z stochastisch unabhängig. Dann folgt

$$\mathbb{E}(X|Z)(\omega) := \begin{cases} \mathbb{E}(X|Z = Z(\omega)), & \text{falls } Z(\omega) \in \{z_1, z_2, \dots\}, \\ 0, & \text{sonst} \end{cases}$$

Die Zufallsvariable $\mathbb{E}(X|Z)$ heißt bedingte Erwartung von X bei gegebenem Z .

Bemerkung

(i) $\mathbb{E}(X|Z)(\omega), \omega \in \Omega$, hängen nur von $Z(\omega)$ ab.

(ii) $\mathbb{E}(X|Z)$ ist als Funktion auf Ω konstant auf den Mengen $\{Z = z_j\}, j \geq 1$.

(iii) $\mathbb{E}[\mathbb{E}(X|Z)] = \sum_{j \geq 1} \mathbb{E}(X|Z = z_j) \mathbb{P}(Z = z_j)$

$$\mathbb{E}[\mathbb{E}(X|Z)] = \sum_{j \geq 1} \sum_{\omega: Z(\omega)=z_j} X(\omega) \mathbb{P}(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\{\omega\}) = \mathbb{E}(X)$$

Beispiel

Sei X_i die Augenzahl des i -ten Wurfs und $i = 1, 2$. Außerdem sei $M := \max(X_1, X_2)$.

$$\mathbb{P}(M = k | X_1 = j) = \begin{cases} \frac{j}{6}, & \text{falls } k = j, \\ \frac{1}{6}, & \text{falls } k > j \end{cases}$$

$$\mathbb{E}(M | X_1 = j) = j \cdot \frac{j}{6} + \sum_{k=j+1}^6 k \cdot \frac{1}{6} = \frac{1}{6} \cdot (j^2 + 21 - \frac{j(j+1)}{2})$$

$$\mathbb{E}(M | X_1 = j) = 3.5 + \frac{j(j-1)}{12}, \quad j = 1, \dots, 6$$

$$\mathbb{E}(M | X_1) = 3.5 + \frac{X_1(X_1 - 1)}{12}$$

$$\mathbb{E}(M | X_1 = 1) = 3.5$$

$$\mathbb{E}(M | X_1 = 2) = 3.67$$

$$\mathbb{E}(M | X_1 = 3) = 4$$

$$\mathbb{E}(M | X_1 = 4) = 4.5$$

$$\mathbb{E}(M | X_1 = 5) = 5.17$$

$$\mathbb{E}(M | X_1 = 6) = 6$$

1.5 Stochastische Prozesse

Man spricht in der Mathematik von einem *stochastischem Prozess* (Zufallsprozess), wenn man zeitlich geordnete und zufällige Vorgänge beschreibt. Stochastische Prozesse bilden die Grundlage in der stochastischen Analysis. Einfache stochastische Prozesse existieren in der Forschung schon lange. Die endgültige und heutzutage verwendete Theorie wurde jedoch erst im 20. Jahrhundert, insbesondere von Paul Lévy und Andrei Kolmogorow entwickelt.

Definition 27 messbarer Raum

Ein messbarer Raum ist ein Dupel (Ω, \mathcal{A}) , das aus einer nichtleeren Menge Ω sowie einer σ -Algebra \mathcal{A} über Ω besteht. Man nennt die Mengen aus \mathcal{A} messbar.

Definition 28 Zustandsraum

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Sei (Z, \mathcal{Z}) ein Raum mit einer σ -Algebra und T eine Indexmenge mit $T \in \{\mathbb{N}_0, \mathbb{R}_+\}$, die bei Anwendung meist die Menge der Zeitpunkte darstellt. Man nennt eine Abbildung

$$X : \Omega \times T \rightarrow Z, (\omega, t) \mapsto X_t(\omega)$$

einen stochastischen Prozess mit den Zufallsvariablen $X_t : \Omega \rightarrow Z, t \in T$, falls $X_t : \omega \mapsto X_t(\omega)$, für alle $t \in T$ eine \mathcal{A} - \mathcal{Z} -messbare Abbildung ist.

Man nennt die Menge Z auch Zustandsraum des Prozesses, da er alle Werte, welche der Prozess annehmen kann, enthält.

Bemerkung

Stochastische Prozesse werden mittels ihrer Struktur in der Abhängigkeit eingeteilt. Diese Einteilung erfolgt meist durch den bedingten Erwartungswert. Dazu gehören folgende zwei Einteilungen.

Markow-Prozesse

Markow-Prozesse haben ein kurzes Gedächtnis. Die Wahrscheinlichkeit, dass ein Zustand angenommen wird, hängt von dem vorherigen Zustand ab, jedoch nicht vom kompletten vergangenen Prozess.

Martingale

Martingale können ein faires Spiel modellieren. Falls man zu einem bestimmten Zeitpunkt einen Betrag bereits erspielt bzw. gewonnen hat, so ist genau dieser Betrag der Erwartungswert für zukünftige Gewinne.

1.5.1 Stochastische Matrix - Übergangsmatrizen

Bei *Stochastischen Matrizen* (Übergangsmatrizen) geht es um die Vorstellung eines abstrakten Systems. Durch Stochastische Matrizen ist es möglich, Schritt für Schritt von einem in den nächsten Zustand zu rechnen.

Eigenschaften - Stochastische Matrix

Eine Stochastische Matrix hat folgende Eigenschaften.

(i) Die Stochastische Matrix ist immer quadratisch. Die Anzahl ihrer Spalten ist gleich der Anzahl der Zeilen.

(ii) Die Einträge der Stochastischen Matrix dürfen nicht negativ sein.

(iii) Die Einträge der Stochastischen Matrix sind Zahlenwerte aus der Menge $[0,1]$.

(iv) (1) Ergibt die Zeilensumme der Stochastischen Matrix immer genau 1, so nennt man diese *zeilenstochastische Matrix*.

(2) Ergibt die Spaltensumme der Stochastischen Matrix immer genau 1, so nennt man diese *spaltenstochastische Matrix*.

Bemerkung

In der Stochastik verwendet man Stochastische Matrizen, um Übergangswahrscheinlichkeiten von Markow-Ketten darzustellen. Daher werden diese auch *Übergangsmatrizen* genannt.

Beispiel

$$(i) A = \begin{pmatrix} 0.2 & 0.1 & 0.4 \\ 0.6 & 0.1 & 0.3 \\ 0.2 & 0.8 & 0.3 \end{pmatrix}$$

Dabei ist die Matrix A eine spaltenstochastische Matrix.

$$(ii) B = \begin{pmatrix} 0.2 & 0.2 & 0.6 \\ 0.8 & 0.1 & 0.1 \\ 0.2 & 0.5 & 0.3 \end{pmatrix}$$

Dabei ist die Matrix B eine zeilenstochastische Matrix.

1.5.2 Filtration

Bei stochastischen Prozessen spricht man oftmals von Filtrierungen. Diese sind eine Gruppe von σ -Algebren. Eine Filtrierung modelliert verfügbare Informationen, die zu unterschiedlichen Zeitpunkten verfügbar sind zu einem Zufallsprozess.

Definition 29 Filtration

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $\tau \subset \mathbb{R}$ eine Indexmenge. Sei außerdem eine Unter- σ -Algebra \mathcal{F}_n von \mathcal{A} für jedes $n \in \tau$ gegeben.

Dann nennt man die Familie von σ -Algebren $\mathbb{F} = (\mathcal{F}_n)_{n \in \tau}$ Filtration bzw. Filtrierung in \mathcal{A} oder auf $(\Omega, \mathcal{A}, \mathbb{P})$, falls diese steigend geordnet ist.

Dafür muss für alle $s, n \in \tau$ gelten, dass $\mathcal{F}_s \subseteq \mathcal{F}_n$ mit $s \leq n$.

Falls $\mathbb{F} = (\mathcal{F}_n)_{n \in \tau}$ eine Filtrierung ist, so nennt man $(\Omega, \mathcal{A}, (\mathcal{F}_n)_{n \in \tau}, \mathbb{P})$ einen gefilterten Wahrscheinlichkeitsraum.

Beispiel

Sei $(\mathbb{Z}, \mathcal{P}(\mathbb{Z}), \mathbb{P})$ ein Wahrscheinlichkeitsraum, wobei die Grundmenge \mathbb{Z} abzählbar ist. Eine mögliche Filtration wäre zum Beispiel

$$\mathcal{F}_n := \sigma(\mathcal{P}(\{-n, \dots, n\})).$$

Dieses Beispiel modelliert Informationen, die angeben, dass man sich bis zum n -ten Zeitpunkt bis zu n Schritte vom ursprünglichen Ort fortbewegt hat. Solch eine Filtration würde man für einen *Random Walk* verwenden.

Beispiel

Es wird eine Münze T -Mal geworfen. Es ergibt sich daher ein Grundraum $\Omega = \{0, 1\}^T$ und $\mathcal{F} = 2^\Omega$.

Es sind die ersten t Münzwürfe bis zum Zeitpunkt t bekannt, aber die Ausgänge der übrigen $T - t$ Münzwürfe nicht. Diese werden durch die Filtration \mathcal{F}_t modelliert. Es gilt dabei

$$\mathcal{F}_t = 2^{\{0,1\}^t} \times \{0,1\}^{T-t} := \{A \in 2^\Omega \mid A = A_1 \times \{0,1\}^{T-t} \text{ f\"ur ein } A_1 \in 2^{\{0,1\}^t}\}$$

Dabei steht $2^{\{0,1\}^t}$ für die Potenzmenge von $\{0,1\}^t$.

Setzt man beispielsweise $T = 2$ und $t = 1$, dann ergibt sich

$$\mathcal{F}_1 = \{\emptyset, \{(0,0), (0,1)\}, \{(1,0), (1,1)\}, \{(0,0), (0,1), (1,0), (1,1)\}\}.$$

Bemerkung

Zusammenfassend kann man sagen, dass eine Adaption eines stochastischen Prozesses $(X_n)_{n \in \tau}$ an eine Filtration $(\mathcal{F}_n)_{n \in \tau}$ bedeutet, dass bekannt ist, wie die Funktion $s \mapsto X_s(\omega)$ in $[0, n]$ zu einem Zeitpunkt n verläuft. Dabei ist $\omega \in \Omega$ beliebig und unbekannt.

Definition 30 .

Eine Filtration wird Filtration der vollständigen Information genannt, falls die Festlegung von $\mathcal{F}_n := \mathcal{A}$ für alle möglichen $n \in \tau$ gegeben ist. Somit ist zu allen Zeitpunkten t eine vollständige Information bekannt.

Definition 31 .

Es sei $X = (X_n)_{n \in \mathbb{N}_0}$ ein stochastischer Prozess. Dann ist die natürliche Filtration von X definiert als $\mathcal{F}_n^X = \sigma(X_0, \dots, X_n), n \in \mathbb{N}_0$.

1.5.3 Stoppzeit

Definition 32 Stoppzeit

Es sein $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(\mathcal{F}_n)_{n \in T}$ eine beliebige Filtration. Es gilt $T = \mathbb{N}_0$.

Sei $\tau : \Omega \rightarrow T \cup \{+\infty\}$ eine beliebige Zufallsvariable über den Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. Man nennt τ eine Stoppzeit bezüglich der Filtration \mathcal{F}_n , falls

$$\{\omega \in \Omega : \tau(\omega) \leq n\} \in \mathcal{F}_n, \quad \forall n \in [0, +\infty).$$

Bemerkung

1) Man interpretiert $0, 1, 2, 3, \dots$ als *Zeitpunkte*. Zu einem Zeitpunkt n ist es möglich, das Eintreten oder Nichteintreten von Ereignissen aus \mathcal{F}_n zu beobachten. Man kann sagen, dass eine Filtration einen Informationsgewinn bei fortschreitender Zeit modelliert.

2) Betrachtet man zum Beispiel ein Glücksspiel, so kann man dieses zu jedem zufallsabhängigen Zeitpunkt anhalten und beenden. Wegen der Bedingung $\tau = n \in \mathcal{F}_n$ fließt nur die Information bis zum Zeitpunkt n , an dem gestoppt wurde, ein.

3) Es können ebenso auch nur endlich viele Zufallsvariablen X_0, X_1, \dots, X_k $\mathbb{F} := (\mathcal{F}_n)_{n \geq 0}$ vorhanden sein. In diesem Fall ist die Stoppzeit $\tau : \Omega \rightarrow \{0, 1, \dots, k\}$, mit $\tau = n \in \mathcal{F}_n$ für $0 \leq n \leq k$.

Satz 13 .

Die Abbildung $\tau : \Omega \rightarrow \mathbb{N}_0 \cup \{\infty\}$ ist genau dann eine Stoppzeit, falls gilt

$$\{\tau \leq n\} \in \mathcal{F}_n$$

für alle $n \in \mathbb{N}_0$.

Beweis

Aus der Definition folgt, dass $\{\tau = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ für jedes $k \leq n$ und somit auch

$$\{\tau \leq n\} = \bigcup_{k=0}^n \{\tau = k\} \in \mathcal{F}_n.$$

In die umgekehrte Richtung folgt aus $\{\tau \leq n\} \in \mathcal{F}_n$ für alle $n \in \mathbb{N}_0$, dass

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\} \in \mathcal{F}_n.$$

□

Bemerkung

Falls τ_1 und τ_2 Stoppzeiten zur Filtration (\mathcal{F}_n) sind, dann sind auch

$$\tau_1 + \tau_2, \max(\tau_1, \tau_2), \min(\tau_1, \tau_2)$$

Stoppzeiten zur Filtration (\mathcal{F}_n) .

Beispiele

(i) *Ersteintrittszeit:*

Gegeben sei die Folge $(X_n)_{n \geq 0}$ mit (Ω', \mathcal{A}') -wertigen Zufallsvariablen, wobei $A' \in \mathcal{A}'$. Dann definiert man die Eintrittszeit in die Menge A' mit

$$\tau := \inf\{n \geq 0 : X_n \in A'\}.$$

Dabei ist $\inf \emptyset := \infty$. Diese definierte Eintrittszeit in die Menge A' ist eine Stoppzeit zur natürlichen Filtration \mathbb{F}^X (von X erzeugte Filtration). Dabei gilt für jedes $n \in \mathbb{N}_0$, dass

$$\{\tau = n\} = \{X_n \in A'\} \cap \bigcap_{j=0}^{n-1} \{X_j \notin A'\} \in \sigma(X_0, \dots, X_n).$$

(ii) *feste Stoppzeit*

Für die feste Stoppzeit $\tau(\omega) := c, \omega \in \Omega$ bei vorgegebenen $c \in \mathbb{N}_0$ gilt, dass

$$\{\tau = n\} = \begin{cases} \Omega, & \text{falls } n = c \\ \emptyset, & \text{falls } n \neq c. \end{cases}$$

(iii) *Keine Stoppzeit*

Ein Beispiel für keine Stoppzeit im Allgemeinen ist

$$\tau := \inf\{n \geq 0 : X_n = X_{n+1}\}.$$

Bemerkung

Auf den ersten Blick wäre das bei einer konstanten Folge eine Stoppzeit. Doch definiert man X_n als Würfelwurf, so würde es bedeuten, dass man stoppt, sobald man nacheinander die gleiche Augensumme würfelt. Jedoch kann man zum Zeitpunkt n noch nicht wissen, ob man danach die gleiche Augensumme würfelt, somit ist dies keine Stoppzeit.

Definition 33 .

Sei τ eine Stoppzeit zur Filtration (\mathcal{F}_n) . Dann nennt man das folgende Mengensystem

$$\mathcal{A}_\tau := \{A \in \mathcal{A} : A \cap \{\tau \leq n\} \in \mathcal{F}_n, \forall n \in \mathbb{N}_0\}$$

σ -Algebra der τ -Vergangenheit.

Satz 14 .

Sei τ eine endliche Stoppzeit zur Filtration (\mathcal{F}_n) und sei $(X_n)_{n \geq 0}$ eine Folge mit reell adaptierten Zufallsvariablen.

Dann ist die wie folgt definierte Abbildung

$$X_\tau : \Omega \rightarrow \mathbb{R}, \omega \rightarrow X_\tau(\omega) := \begin{cases} X_{\tau(\omega)}(\omega), & \text{falls } \tau(\omega) < \infty, \\ 0, & \text{falls } \tau(\omega) = \infty \end{cases}$$

\mathcal{A}_τ -messbar.

Beweis [4]

B ist eine beliebige Borelmenge und für alle $n \in \mathbb{N}_0$ gilt

$$\{X_\tau \in B\} \cap \{\tau \leq n\} = \bigcup_{k=0}^n (\{X_k \in B\} \cap \{\tau = k\}) \in \mathcal{F}_n,$$

da nach der Definition von \mathcal{A}_τ folgt, dass $\{X_\tau \in B\} = X_\tau^{-1}(B)$. □

Beispiel

Gegeben ist eine Urne mit 5 Kugeln. 3 von den Kugeln haben den Wert -1 und 2 Kugeln haben den Wert +1. Es wird ohne Zurücklegen gezogen. Es ist möglich jederzeit zu stoppen. Der Gewinn rechnet sich aus der Summe der gezogenen Kugeln.

Frage: Gibt es eine Regel, wann man stoppen soll, sodass der erwartete Gewinn positiv ist?

Mögliche Strategie:

$$\Omega := \{\omega := (a_1, \dots, a_5) \in \{-1, +1\}^5 : \sum_{j=1}^5 a_j = -1\}.$$

\mathbb{P} ist die Gleichverteilung auf Ω .

Seien $X_j(\omega) := a_j$ die Projektions-Abbildungen.

$$\tau(\omega) := \begin{cases} 1, & \text{falls } X_1(\omega) = 1, \\ 2, & \text{falls } X_1(\omega) = -1, X_2(\omega) = 1, \\ 4, & \text{falls } X_1(\omega) = X_2(\omega) = -1, X_3(\omega) = X_4(\omega) = 1, \\ 5, & \text{falls } X_1(\omega) = X_2(\omega) = -1, X_3(\omega) \cdot X_4(\omega) = -1. \end{cases}$$

Die Summe S_τ , die sich nach dem Stoppen ergibt, ist der Gewinn. Dabei gilt $S_n := \sum_{j=1}^n X_j$. Es ergibt sich dabei folgender Erwartungswert:

$$\mathbb{E}(S_\tau) = 1 \cdot \mathbb{P}(S_\tau = 1) + (-1) \cdot \mathbb{P}(S_\tau = -1)$$

$$\mathbb{E}(S_\tau) = 1 \cdot \frac{2}{5} + (-1) \cdot \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{2}{3} = \frac{1}{5}.$$

Man erhält bei dieser Stoppzeit einen positiven erwarteten Gewinn.

Beispiel

Bei wiederholtem Würfelwurf werden nach jedem Wurf die Augenzahlen addiert und auf ein Punktekonto gelegt. Falls jedoch eine Sechs gewürfelt wird, wird das Spiel beendet und der Punktestand auf Null gesetzt.

Das Spiel kann jederzeit gestoppt werden solange keine Sechs gewürfelt wird.

Welche Stoppstrategie ergibt sich bei mehrmaliger Wiederholung des Spiels?

Die Frage könnte ebenso lauten: Mit welcher Stoppstrategie maximiert man den Erwartungswert?

Bemerkung: Die Strategie hängt nicht davon ab, wie oft man keine Sechs würfelt. Bei diesem Spiel hat man Gedächtnislosigkeit, da ein Wurf nicht vom vorherigen abhängt.

Die Strategie kann sich somit nur auf den bereits erzielten Gewinn beziehen. Je nachdem ob dieser hoch bzw. kostbar ist, muss man entscheiden, ob man stoppt oder weiterspielt.

Angenommen man hat den Punktestand k , dann ist X_k der Punktestand, den ich gedanklich nach einem weiteren Wurf erhalte.

Welche Werte kann X_k annehmen?

$X_k \in \{k+1, k+2, k+3, k+4, k+5, 0\}$, dabei hat jeder Wert die Wahrscheinlichkeit $1/6$.

Bemerkung: Die weitere Überlegung dazu wäre folgende. Man sollte nur weiterwerfen, wenn der erwartete Gewinn nach dem nächsten Wurf größer ist als der Punktestand den man bereits hat (k).

$$\mathbb{E}(X_k) = \frac{1}{6} \sum_{j=1}^5 (k+j)$$

$$\mathbb{E}(X_k) = \frac{5k+15}{6} > k$$

$$5k+15 > 6k$$

$$15 > k$$

Bemerkung: Falls der bisherige Punktestand größer als 15 ist, so sollte man das Spiel stoppen.

Man definiert G als zufälligen Spielgewinn mit dieser Strategie. Gesucht ist nun der Erwartungswert unter der Bedingung, dass ein Punktestand von k bereits vorliegt, also $\mathbb{E}_k(G)$. Es wird der Erwartungswert zu Beginn gesucht, also für $k=0$, somit $\mathbb{E}_0(G) = \mathbb{E}(G)$.

Bemerkung: Wir kennen den Erwartungswert für gewisse k bereits, somit arbeiten wir uns zurück zu $k=0$.

Aus der obigen Stoppregel folgt, dass wenn $\mathbb{E}_k(G) = k$ und $k \in \{15, 16, 17, 18, 19\}$ ist, das Spiel gestoppt werden sollte.

Wenn $k \leq 14$ ist, wird das Ergebnis x des nächsten Wurfs betrachtet.

Aus der Formel des totalen Erwartungswertes folgt für $\mathbb{E}_k(G)$, dass

$$\mathbb{E}_k(G) = \sum_{j=1}^6 \mathbb{E}_k(G|X=j)\mathbb{P}(X=j).$$

Es ist wichtig zu wissen, dass $\mathbb{E}_k(G|X=6)$ gilt. Außerdem gilt $\mathbb{E}_k(G|X=j) = \mathbb{E}_{k+j}(G)$ für $j \leq 5$.

Daraus folgt

$$\mathbb{E}_k(G) = \frac{1}{6} \sum_{j=1}^5 \mathbb{E}_{k+j}(G).$$

Bemerkung: Nun kann man $\mathbb{E}(G)$ mit Rückwärts Induktion berechnen.

$$\mathbb{E}_{14}(G) = \frac{1}{6} \cdot (\mathbb{E}_{15}(G) + \mathbb{E}_{16}(G) + \mathbb{E}_{17}(G) + \mathbb{E}_{18}(G) + \mathbb{E}_{19}(G))$$

$$\mathbb{E}_{14}(G) = \frac{1}{6} \cdot (15 + 16 + 17 + 18 + 19) = \frac{85}{6} \approx 14,17.$$

$$\mathbb{E}_{13}(G) = \frac{1}{6} \cdot \left(\frac{85}{6} + 15 + 16 + 17 + 18\right) = \frac{481}{36} \approx 13,36.$$

Schlussendlich ergibt sich für $\mathbb{E}(G) \approx 6,15$.

Bemerkung: Müsste man bei dem Spiel einen Einsatz zahlen, so wäre das Spiel bei einem Einsatz von unter 6,15 Euro vorteilhaft. Bei einem höheren Einsatz wäre das Spiel eher unvorteilhaft für die/den Spieler/in.

2 Markovs Entscheidungstheorie zum optimalen Stoppen

In diesem Kapitel wird das Problem des optimalen Stoppens anhand von zwei Beispielen untersucht und dargestellt. Die Beispiele beruhen auf Zufallsspaziergängen, die jederzeit gestoppt werden können. Es ergeben sich viele Stoppmöglichkeiten, jedoch soll die optimale Variante, die den maximalen Gewinn bringt, gezeigt werden.

Beispiel 2.1

Man stellt sich das Spiel (den Zufallsspaziergang) mit folgenden Regeln vor.

Es gibt ein Spielfeld mit den Feldern $0, 1, 2, \dots$ und der Spielstein der Spielerin bzw. des Spielers steht auf dem Feld 0. Die/Der Spieler/in würfelt die Anzahl der Felder, die er vorrückend darf. Wichtig ist, dass die/der Spieler/in nach jedem Wurf die Möglichkeit hat aufzuhören. Falls die/der Spieler/in aufhört, erhält dieser eine Geldsumme ausbezahlt, die das Tausendfache der Zahl des Feldes, auf dem er steht beträgt. Falls die/der Spieler/in über die Zahl 5 hinauskommt, erhält er nichts.

Problem: Wann ist der richtige Zeitpunkt, mit dem Spiel aufzuhören, um mit maximalem Gewinn auszusteigen? Mathematisch gesagt: Wann ist der Erwartungswert des Gewinns maximal?

Lösungsansatz:

Sei i das Feld, auf dem sich die/der Spieler/in befindet.

- Die/Der Spieler/in sollte das Spiel definitiv spielen, da man auf Position $i = 0$ ohne Gewinn aussteigt. In Startposition gewinnt die/der Spieler/in mit Wahrscheinlichkeit $5/6$ und hat er auf jeden Fall einen positiven erwarteten Gewinn.

- Der Fall $i = 5$ ist trivial. Würfelt die/der Spieler/in gleich zu Beginn die Augenzahl 5, so sollte er aufhören. Der nächste Wurf würde garantiert zu einem Verlust führen. Die/Der Spieler/in steigt mit 5.000 Euro aus dem Spiel aus.

- Fall $i = 4$ bietet zwei Optionen.

Angenommen die/der Spieler/in entscheidet sich zu stoppen, dann verlässt er das Spiel bei 4.000 Euro.

Spielt die/der Spieler/in weiter, so landet er mit einer Wahrscheinlichkeit von $1/6$ auf dem Feld 5. Somit würde die/der Spieler/in 5.000 Euro erhalten.

Die Wahrscheinlichkeit, dass die/der Spieler/in ohne Geld das Spiel verlässt liegt bei $5/6$. Somit würde der erwartete Gewinn bei $\frac{5.000}{6}$ Euro liegen, was definitiv weniger ist als die 4.000 Euro.

In diesem Fall ist es sicherer, nicht weiterzuspielen.

- Fall $i = 3$ ist ähnlich zu dem oberen.

Falls die/der Spieler/in stoppt, erhält er 3.000 Euro. Spielt die/der Spieler/in weiter, so verliert mit Wahrscheinlichkeit $4/6 = 2/3$ und erhält nichts. Der erwartete Gewinn würde bei

$\frac{4.000+5.000}{6} = 1.500$ Euro liegen und somit weniger als 3.000 Euro betragen.

Auch in diesem Fall ist es sicherer, nicht weiterzuspielen.

- Bei Fall $i = 2$ tritt ein interessanter Fall ein, da es hier egal ist, ob man weiterspielt oder stoppt.

Falls die/der Spieler/in stoppt, erhält er 2.000 Euro.

Spielt die/der Spieler/in das Spiel weiter, so verliert er mit Wahrscheinlichkeit $3/6 = 1/2$ und erhält nichts. Beim Weiterspielen liegt der zu erwartende Gewinn bei

$$\frac{3.000 + 4.000 + 5.000}{6} = 2.000 \text{ Euro.}$$

In diesem Fall ist es also egal, wie sich die/der Spieler/in entscheidet.

- Fall $i = 1$ ist so gut wie klar.

Falls die/der Spieler/in stoppt, erhält er 1.000 Euro.

Spielt die/der Spieler/in noch einmal weiter, so verliert er mit Wahrscheinlichkeit $2/6 = 1/3$ und erhält nichts. Er gewinnt jedoch mit einer Wahrscheinlichkeit von $4/6 = 2/3$ und erhält einen zu erwartenden Gewinn von

$$\frac{2.000 + 3.000 + 4.000 + 5.000}{6} \approx 2.333 \text{ Euro.}$$

Zusammenfassend kann gesagt werden, dass die/der Spieler/in das Spiel weiterspielen sollte, falls der erste Wurf 1 oder 2 ist. Ist die Augensumme des ersten Wurfs größer als 2, so sollte das Spiel gestoppt werden.

Beispiel 2.2

Auch in diesem Beispiel geht es um ein Würfelspiel mit folgenden Regeln.

Die/Der Spieler/in darf bis zum ersten 6er würfeln und wenn er möchte, das Spiel auch vorher stoppen. Falls gestoppt wird, erhält die/der Spieler/in einen Gewinn, der das Tausendfache der zuletzt gewürfelten Augensumme beträgt. Würfelt die/der Spieler/in die Augensumme 6, so endet das Spiel.

Problem: Welche Spielstrategie wäre bei diesem Spiel die optimale? Wann sollte die/der Spieler/in aufhören zu würfeln? Wie ist $i_0 \in \{1, \dots, 5\}$ zu wählen?

Strategie: Stoppe das Spiel, falls der letzte Wurf die Augensumme aus $i_0, \dots, 5$ hat. Spiele bei Würfeln mit anderer Augensumme weiter.

Lösungsansatz:

Es sei M_{i_0} der beim Weiterspielen maximal zu erreichende Erwartungswert. Dabei ist $i_0 \in \{1, \dots, 5\}$.

- Fall $i_0 = 5$. Stoppt die/der Spieler/in hier das Spiel, so erhält er einen Gewinn von 5.000 Euro.

Spielt die/der Spieler/in weiter, so wird das Spiel mit Wahrscheinlichkeit $4/6 = 2/3$ fortgesetzt und die/der Spieler/in erhält M_5 . Die/Der Spieler/in erhält mit Wahrscheinlichkeit $1/6$ den Wurf 5 und erhält 5.000 Euro. Somit folgt die Gleichung

$$M_5 = \frac{2}{3} M_5 + \frac{1}{6} 5000 \implies M_5 = 2500.$$

- Fall $i_0 = 4$. Die/Der Spieler/in erhält beim Stoppen des Spieles 4.000 Euro. Sollte die/der Spieler/in das Spiel fortsetzen so, führt es mit analoger Überlegung von oben zur Gleichung

$$M_4 = \frac{1}{2} M_4 + \frac{1}{6} (4000 + 5000) \implies M_4 = 3000.$$

- Fall $i_0 = 3$ führt zur Gleichung

$$M_3 = \frac{1}{3} M_3 + \frac{1}{6} (3000 + 4000 + 5000) \implies M_3 = 3000,$$

falls die/der Spieler/in weiterspielt. Stoppt sie/er das Spiel, so erhält er ebenso 3.000 Euro.

- Fall $i_0 = 2$ liefert die Gleichung

$$M_2 = \frac{1}{6} M_2 + \frac{1}{6} (2000 + 3000 + 4000 + 5000) \implies M_2 = 2800,$$

falls die/der Spieler/in weiterspielt. Somit ist das Stoppen des Spiels hier definitiv ungünstiger, da die/der Spieler/in beim Stoppen 2.000 Euro erhält.

- Fall $i_0 = 1$ ergibt die Gleichung

$$M_1 = \frac{1000 + 2000 + 3000 + 4000 + 5000}{6} = \frac{15000}{6} \implies M_1 = 2500,$$

falls die/der Spieler/in weiterspielt. Beim Stoppen würde die/der Spieler/in 1.000 Euro erhalten. Somit ist das Stoppen hier ebenso ungünstiger als das Weiterspielen.

Zusammenfassend ergibt sich, dass die/der Spieler/in das Spiel bei den Augenzahlen $\{3, 4, 5, 6\}$ stoppen und bei 1 und 2 weiterspielen soll.

2.1 Genauere Problemstellung

Die Lösungswege der beiden Beispiele waren unterschiedlich. Der Inhalt dieses Kapitels bezieht sich auf eine allgemeine optimale Strategie für derartige Problemstellungen.

Definition 34 *Es sei $X_0, X_1, \dots : \Omega \rightarrow S$ eine Folge von Zufallsvariablen, welche über dem selben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definiert sind. Die Zufallsvariablen nehmen ihre Werte aus dem Zustandsraum $S = \{1, \dots, l\}$ an.*

$(X_n)_{n \geq 0}$ nennt man *homogene Markov-Kette mit Anfangsverteilung*

$\alpha = (\alpha_1, \dots, \alpha_l)$ und *Übergangsmatrix* $P = (p_{ij})$, wenn für beliebige

$n = 0, 1, 2, \dots$ und $i_0, i_1, \dots, i_n \in S$ gilt:

$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}$$

für $n = 0, 1, 2, \dots$ und $i_0, i_1, \dots, i_n \in S$.

Auf einem endlichen Zustandsraum S sei eine homogene Markovkette mittels stochastischer Matrix P definiert, wobei in x_0 gestartet wird. Den dazu veranlassten Markovprozess bezeichnen wir mit $(X_n)_{n \geq 0}$. Die X_n seien dabei auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definiert.

Außerdem definieren wir $u : S \rightarrow \mathbb{R}$ die Funktion als *Gewinnfunktion*. Während der Prozess im Laufen ist, kann man ihn jederzeit stoppen und erhält bei Stand x genau $u(x)$ Euro als Gewinn bei positiver Zahl, oder auch Verlust, falls die Zahl negativ ist. Im folgenden wenden wir dieses Wissen an den oben beschriebenen Beispielen an.

zu Beispiel 2.1:

In diesem Beispiel kann man S als $\{0, 1, 2, 3, 4, 5, 6\}$ wählen und folgende stochastische Matrix M_1 aufstellen.

$$M_1 = \begin{pmatrix} 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{2}{6} \\ 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{3}{6} \\ 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{1}{6} & \frac{4}{6} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{5}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Es ergibt sich für die Gewinnfunktion $u(i) := 1000 \cdot i$, $i < 6$. Dabei ist $u(6) = 0$.

zu Beispiel 2.2:

Hier wird $s = \{1, 2, 3, 4, 5, 6\}$ gewählt, und es ergibt sich folgende stochastische Matrix M_2 .

$$M_2 = \begin{pmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Die Gewinnfunktion u ist hier dieselbe wie oben.

Bemerkung: Die ersten fünf Spalten bestehen aus $\frac{1}{6}$, da die Augensumme im nächsten Wurf (bzw. der Gewinn nach dem nächsten Wurf) unabhängig vom vorherigen Wurf ist. Die letzte Spalte besteht bis auf den letzten Eintrag aus Nullern - hat man bereits 6 gewürfelt, so ist die Wahrscheinlichkeit beim nächsten Wurf für einen positiven Gewinn (den man bei Augensumme ≤ 5 erhält) gleich Null und die Wahrscheinlichkeit für keinen Gewinn (den man bei Augensumme = 6 erhält) gleich 1.

Fragestellung:

Wie muss man vorgehen, um die Gewinnerwartung zu maximieren bzw. zumindest dem optimalen Wert sehr nahe zu sein?

Anders und genauer formuliert lässt es sich folgendermaßen ausdrücken.

Es sei $(F_n)_{n \geq 0}$ eine natürliche Filtration zu (X_n) . Finde eine Stoppzeit $\tau : \Omega \rightarrow \mathbb{N}_0 \cup \{\infty\}$, sodass gilt:

- Für τ gilt, dass τ fast sicher endlich ist.

- Der Erwartungswert von $u(X_\tau)$ ist (zumindest nahezu) optimal, falls $X_\tau : \Omega \rightarrow \mathbb{R}$ durch $\omega \mapsto X_{\tau(\omega)}(\omega)$ definiert ist.

2.2 Die Lösung der optimalen Stoppzeit

Um zur optimalen Lösung des Stoppproblems zu kommen, braucht es noch folgendes Resultat, welches auch bewiesen wird.

Definition 35 [5]

Sei $P = (P_{ab})_{a,b \in S}$ eine stochastische Matrix und $A \subset S$.

Dann heißt $h : S \rightarrow \mathbb{R}$ superharmonisch auf A , wenn gilt, dass

$$\sum_{b \in S} P_{ab} h(b) \leq h(a), \quad a \in A.$$

Falls bei der Ungleichung immer eine Gleichheit besteht, so heißt h harmonisch in A . Falls $A = S$, dann heißt h superharmonisch bzw. harmonisch.

Definition 36 .

Die Funktion $\hat{f} : S \rightarrow \mathbb{R}$,

$$\hat{f}(x) := \inf\{u(x) \mid u \text{ superharmonisch}, u \geq f\}$$

ist für jedes $f : S \rightarrow \mathbb{R}$ superharmonisch und heißt superharmonische Majorante von f .

Ist die Auszahlungsfunktion superharmonisch, so ist längeres Warten immer ungünstiger als kürzeres Warten.

Die Funktion $u : S \rightarrow \mathbb{R}$ bezeichne man als Auszahlungsfunktion. Befindet man sich bei x , so erhält man $u(x)$ ausbezahlt. Die Auszahlung hat den Erwartungswert

$$\sum_{y \in S} p_{xy} u(y),$$

falls man den ersten Schritt auf jeden Fall macht.

Die Funktion $u : S \rightarrow \mathbb{R}$ heißt superharmonisch (bezüglich (p_{xy})), falls das Warten ungünstig ist, also falls für alle x gilt

$$\sum_{y \in S} p_{xy} u(y) \leq u(x).$$

Es ist trivial, dass jede konstante Funktion somit superharmonisch ist.

Lemma 1 .

Ist $E \in \sigma(X_0, \dots, X_{n-1})$, so gilt

$$\mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, E) = p_{x_{n-1}, x_n} \mathbb{P}(X_{n-1} = x_{n-1}, E).$$

Beweis:

Sei E der Form $\{X_0 = y_0, \dots, X_{n-1} = y_{n-1}\}$, dann ist es trivial, dass die Aussage richtig ist, falls $x_{n-1} \neq y_{n-1}$, da beide Seiten der Gleichung Null ergeben.

Tritt dieser Fall nicht ein, so hat man eine Umformulierung der Markoveigenschaft:

$$\begin{aligned} \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}, X_{n-2} = y_{n-2}, \dots, X_0 = y_0) &= \\ &= \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}) = p_{x_{n-1}, x_n}. \quad \square \end{aligned}$$

Mit folgendem Lemma wird verdeutlicht, dass längere Stoppzeiten ungünstiger sind als kürzere. Zuerst wird ein Spezialfall gezeigt, bei welchem die Stoppzeiten nahe zueinander liegen.

Lemma 2 .

Es sei $x_0 \in S$ und man betrachte den Prozess, der bei x_0 startet. Falls $u : S \rightarrow \mathbb{R}$ superharmonisch ist und τ, σ fast sicher endliche Stoppzeiten mit $\tau \leq \sigma \leq \tau + 1$, so gilt

$$\mathbb{E}(u(X_\tau)) \geq \mathbb{E}(u(X_\sigma)).$$

Beweis:

$$\begin{aligned}\mathbb{E}(u(X_\sigma)) &= \sum_{x \in S, k \in \mathbb{N}_0} \mathbb{P}(\sigma = k, X_k = x) u(x) = \\ &= \sum_{x \in S, k \in \mathbb{N}_0} \mathbb{P}(\sigma = \tau = k, X_k = x) u(x) + \sum_{x \in S, k \in \mathbb{N}} \mathbb{P}(\sigma = k, \tau = k - 1, X_k = x) u(x) =: A\end{aligned}$$

Betrachtet man die Mengen $\{\sigma = k, \tau = k - 1, X_k = x\}$ so kann man sie in die disjunkten Teilmengen $\{\sigma = k, \tau = k - 1, X_{k-1} = y, X_k = x\}$ für $y \in S$ zerlegen.

Die Menge $\{\sigma = k, \tau = k - 1, X_{k-1} = y\}$ liegt in $\sigma(X_0, \dots, X_{k-1})$, da die Menge $\{\sigma = k, \tau = k - 1, X_{k-1} = y\}$ als die Menge $\{\sigma \geq k, \tau = k - 1, X_{k-1} = y\}$ geschrieben werden kann.

Angemerkt sei auch, dass die Mengen $\{\sigma \geq k\}$ und $\{\sigma \leq k - 1\}$ komplementär sind.

Wegen Lemma 1 ergibt sich:

$$\mathbb{P}(\{\sigma = k, \tau = k - 1, X_{k-1} = y, X_k = x\}) = \mathbb{P}(\{\sigma = k, \tau = k - 1, X_{k-1} = y\}) p_{y,x}.$$

Mit dieser Aussage und der Annahme, dass u eine superharmonische Funktion ist, wird die Rechnung nun fortgesetzt.

$$\begin{aligned}A &= \sum_{x \in S, k \in \mathbb{N}_0} \mathbb{P}(\sigma = \tau = k, X_k = x) u(x) + \sum_{x, y \in S, k \in \mathbb{N}} \mathbb{P}(\sigma = k, \tau = k - 1, X_{k-1} = y) p_{y,x} u(x) \\ &\leq \sum_{x \in S, k \in \mathbb{N}_0} \mathbb{P}(\sigma = \tau = k, X_k = x) u(x) + \sum_{y \in S, k \in \mathbb{N}} \mathbb{P}(\sigma = k, \tau = k - 1, X_{k-1} = y) u(y)\end{aligned}$$

Im nächsten Schritt wird in der Zweiten Summe der Laufindex k umgeändert zu $k - 1$ und die Bezeichnung y wird auf x geändert.

$$A \leq \sum_{x \in S, k \in \mathbb{N}_0} \mathbb{P}(\sigma = \tau = k, X_k = x) u(x) + \sum_{x \in S, k \in \mathbb{N}_0} \mathbb{P}(\sigma = k + 1, \tau = k, X_k = x) u(x).$$

Die Mengen $\{\tau = \sigma = k\}$ und $\{\tau = k, \sigma = k + 1\}$ ergeben die disjunkte Vereinigung $\{\tau = k\}$. Somit gilt:

$$A \leq \sum_{x \in S, k \in \mathbb{N}_0} \mathbb{P}(\tau = k, X_k = x) u(x) = \mathbb{E}(u(X_\tau)).$$

Damit ist gezeigt, dass

$$\mathbb{E}(u(X_\sigma)) \leq \mathbb{E}(u(X_\tau))$$

gilt. □

Um zu zeigen, dass das längere Warten bei einer superharmonischen Funktion *in jedem Fall* schlechter ist als kürzeres Warten, betrachten wir folgendes Lemma.

Lemma 3 .

Sei $x_0 \in S$ und man betrachte x_0 als Start des Prozesses. Sei $u : S \rightarrow \mathbb{R}$ eine superharmonische Funktion und τ, σ die fast sicher endlichen Stoppzeiten. Falls $\tau \leq \sigma$, dann gilt

$$\mathbb{E}(u(X_\sigma)) \leq \mathbb{E}(u(X_\tau)).$$

Beweis:

Die Stoppzeiten τ_n seien für n in \mathbb{N} definiert durch

$$\tau_n := \min\{\sigma, \max\{\tau, n\}\}.$$

Nun gilt

$$\tau_n \leq \tau_{n+1} \leq \tau_n + 1$$

und auch wegen dem obigen Lemma

$$\mathbb{E}(u(X_{\tau_n})) \geq \mathbb{E}(u(X_{\tau_{n+1}})).$$

Die τ_n konvergieren fast sicher punktweise gegen σ . Alle Funktionen, die auftreten sind durch $\max_x |u(x)|$ beschränkt. Daraus und aus dem Satz von der dominierten Konvergenz von Lebesgue (siehe unten) folgt, dass die $\mathbb{E}(u(X_{\tau_n}))$ gegen $\mathbb{E}(u(X_\sigma))$ konvergieren. Weil $\tau_0 = \tau$ ist die Aussage somit bewiesen. \square

3 Das Stopp-Problem

In den 1950er Jahren wurde das SekretärInnen Problem bekannt und weckte bei einigen MathematikerInnen Interesse, das Problem zu studieren. Bis heute entstanden dabei unterschiedliche Fassungen der Problemdarstellung.

Problemdarstellung 3.1

Das *SekretärInnenproblem* ist folgendes Optimierungsproblem. Gegeben sind n Güter, die einer/m Spieler/in in beliebiger Reihenfolge präsentiert werden. Die Zahl $n \in \mathbb{N}$ ist bekannt. Alle $n!$ Reihenfolgen sind gleich wahrscheinlich. Die Präsentation des r -ten Guts ist der r -te Schritt. Die/Der Spieler/in ist risikoneutral.

Die Güter haben alle unterschiedlichen Wert für den/die Spieler/in. Sobald ihm das r -te Gut präsentiert wird, kennt er das Ranking unter den ersten r Gütern, aber er weiß nichts über das Ranking der weiteren $n-r$ Güter.

Er muss nach einem Schritt das zuletzt präsentierte Gut wählen. Dann endet das Spiel. Er hat gewonnen, wenn das gewählte Gut das beste aller n Güter ist. Sonst hat er verloren. Mit welcher Strategie kann er die Wahrscheinlichkeit zu gewinnen maximieren? Wie hoch ist sie dann ungefähr?

Mathematische Voraussetzungen

Folgende sechs Voraussetzungen soll das SekretärInnenproblem aufweisen:

1. Es handelt sich nur um eine vakante Stelle, die zu vergeben ist. Man wählt also maximal eine/n BewerberIn aus.
2. Die genaue Anzahl der BewerberInnen ist bekannt. Wahlweise ist auch der Zeitraum bekannt, in dem eine Entscheidung getroffen werden muss. Dies kann zum Beispiel der Fall sein, wenn man Immobilien verkaufen will und diese über mehrere Wochen in Zeitungen inseriert.
3. Die einzelnen BewerberInnen wurden vor dem Gespräch nicht sortiert bzw. vorher klassifiziert.
4. BewerberInnen können, verglichen mit anderen BewerberInnen, nicht gleich eingestuft werden. Letztendlich ist eine Klassifizierung der Bewerber möglich.
5. Sollte man sich gegen eine/n Bewerber/in entschieden haben, ist es nicht möglich ihn erneut zu befragen oder sogar einzustellen. Die Entscheidung ist definitiv.
6. Der Auswählende ist nur mit dem beste/n Bewerber/in zufrieden und will auch nur diesen einstellen.

3.1 Trefferwahrscheinlichkeit

Für die Bestimmung der Trefferwahrscheinlichkeit des SekretärInnen-Problems durch Kombinatorik und Wahrscheinlichkeitsrechnung werden zuerst die dafür verwendeten Variablen

definiert.

Bemerkung

Hier ist wichtig anzumerken, dass vorerst angenommen wird, dass es mehrere akzeptable Güter unter den Gesamtgütern gibt.

n ...Anzahl der gesamten Güter

k ...Anzahl der bereits abgewiesenen Güter (bestimmen den Maßstab)

m ...Anzahl der akzeptablen Güter

α ...unbekannte Position des besten Gutes

K ...Testmenge: Menge der Positionen, der bereits abgewiesenen Güter $K = \{1, 2, \dots, k\}$.

M ...Menge der akzeptablen Güter $|M| = m$

Um die Trefferwahrscheinlichkeit berechnen zu können, ist eine Unterscheidung, ob das beste Gut unter den frühen abgewiesenen Gütern ($\alpha \in K$) ist oder nicht ($\alpha \notin K$) vorteilhaft.

Fall 1: $\alpha \in K$

In diesem Fall gilt die Wahrscheinlichkeit $\mathbb{P}(\text{Fall 1}) = \frac{k}{n}$. Ein Treffer ist hier jedoch unmöglich, da das beste Gut bereits abgewiesen wurde.

Fall 2: $\alpha \notin K$

Aus dem Fall 1 ergibt sich hier die Wahrscheinlichkeit

$$\mathbb{P}(\text{Fall 2}) = 1 - \mathbb{P}(\text{Fall 1}) = 1 - \frac{k}{n}.$$

Nun wird wieder in zwei Fällen unterschieden, indem in einem Fall unter den k Gütern ein akzeptables Gut dabei ist ($K \cap M \neq \emptyset$) und im anderen Fall nicht ($K \cap M = \emptyset$).

Fall 2.1: $K \cap M = \emptyset$

Der Fall, dass keiner der akzeptablen Güter in der Menge K enthalten ist, hat die Wahrscheinlichkeit

$$\mathbb{P}(\text{Fall 2.1}) = \frac{\binom{n-k-1}{m-1}}{\binom{n-1}{m-1}}.$$

Der Zähler steht für die Anzahl der Kombinationen, bei welchen die akzeptablen Gütern die Positionen $k+1, k+2, \dots, n$ annehmen. Hierbei ist bereits klar, dass das beste Gut dort steht. Der Nenner gibt alle möglichen Kombinationen (mit Ausnahme des besten Gutes) an.

Es wird hier außerdem bezüglich der Position j des ersten akzeptablen Gutes unterschieden. Für jedes $j = k+1, \dots, n-m+1$ gibt es einen untergeordneten Fall.

Fall 2.1.j: $j = \min(M)$

Die Wahrscheinlichkeit, dass einer der akzeptablen Güter die Position j hat ist $\frac{m}{n-k}$. Die Kombinationen der Plätze für die restlichen akzeptablen Güter mit Position $> j$ ergibt $\binom{n-j}{m-1}$. Für die Kombinationen der restlichen Güter ergibt sich die Anzahl $\binom{n-k-1}{m-1}$.

Die Wahrscheinlichkeit für den Fall 2.1.j ist

$$P(\text{Fall 2.1.j}) = \frac{m}{n-k} \cdot \frac{\binom{n-j}{m-1}}{\binom{n-k-1}{m-1}} = \frac{\binom{n-j}{m-1}}{\binom{n-k}{m}}$$

Der Quotient gibt die Wahrscheinlichkeit an, dass die restlichen akzeptablen Güter ihre Position oberhalb von j haben, mit der Bedingung, dass kein Gut davon in der Testmenge liegt.

Es kommt zu einem Treffer, falls das beste derjenigen Güter, die vor Position j liegen, in der Testmenge enthalten ist. Die Wahrscheinlichkeit für diesen Fall ist $\frac{k}{j-1}$.

Fall 2.2: $K \cap M \neq \emptyset$

Für die Wahrscheinlichkeit dieses Falls gilt

$$P(\text{Fall 2.2}) = 1 - P(\text{Fall 2.1}).$$

In der Testmenge ist bereits ein akzeptables Gut, somit wird die/der Spieler/in spätestens mit dem besten Gut (welches außerhalb der Testmenge liegt) ein akzeptables Gut finden.

Für die Trefferwahrscheinlichkeit ergibt sich die Formel

$$P = P(\text{Fall 2}) \cdot \left(P(\text{Fall 2.2}) + P(\text{Fall 2.1}) \cdot \sum_{j=k+1}^{n-m+1} P(\text{Fall 2.1.j}) \cdot \frac{k}{j-1} \right).$$

Somit ergibt sich durch das Einsetzen der oben errechneten bedingten Wahrscheinlichkeiten:

$$P = \frac{n-k}{n} - \frac{\binom{n-k}{m}}{\binom{n}{m}} + \frac{k}{\binom{n}{m}} \sum_{j=k+1}^{n-m+1} \frac{\binom{n-j}{m-1}}{j-1}$$

Bemerkung

Bei den bisherigen Schritten gab es mehrere akzeptable Güter. Beim SekretärInnen-Problem geht es jedoch darum, das beste Gut zu treffen. Dafür muss man den Fall $m = 1$ betrachten.

Spezialfall $m=1$

Im Spezialfall $m = 1$ ändert sich die Formel der Trefferwahrscheinlichkeit.

$$P = \frac{1}{n} \sum_{j=k+1}^n \frac{k}{j-1} = \frac{k}{n} \sum_{i=k}^{n-1} \frac{1}{i}$$

Die Summe

$$\sum_{i=k}^{n-1} \frac{1}{i}$$

kann man auch als Riemannsche Summe von einem Integral sehen. Dadurch ergibt sich folgende Näherungsdarstellung:

$$\sum_{i=k}^{n-1} \frac{1}{i} \approx \int_k^n \frac{1}{x} dx = \ln n - \ln k = \ln \frac{n}{k}$$

Somit erhält man für die Trefferwahrscheinlichkeit im Spezialfall $m = 1$:

$$P \approx -\frac{k}{n} \cdot \ln \frac{k}{n}$$

Bemerkung

Es ergibt sich ein Maximum für den Wert $\frac{k}{n} = \frac{1}{e}$. Somit folgt, dass die/der Spieler/in die besten Chancen hat das beste Gut zu wählen, wenn er nach ca. 37% der Güter das Gut wählt, welches die bereits abgewiesenen Güter übertrifft. Somit liegt die Wahrscheinlichkeit das beste Gut zu wählen bei etwa 37%.

In dieser Arbeit werden nur Strategien für Szenarien mit festem N betrachtet und erläutert. Es existieren jedoch auch Strategien, die man bei unbekanntem N anwenden kann.

Sei $N = n, n \in \mathbb{N}/0$ bekannt und fest gewählt. Es folgen Definitionen einer Klasse von Strategien.

3.2 Strategie von Lindley (1961)

Lindley veröffentlichte 1961 eine Lösung zu dem Problem 3.1. Er verwendet dabei die oben genannten sechs mathematischen Voraussetzungen. Das Szenario dreht sich jedoch bei ihm nicht um die Einstellung der/des besten BewerberIn, sondern um die Wahl der richtigen Ehefrau.

Definition 37 .

Wenn die/der Spieler/in die ersten $r - 1$ Güter nicht nimmt, und dann ab dem r -te Gut das nächste Gut nimmt, das besser als eines der ersten $r - 1$ Güter ist, so wird diese Spielstrategie σ_r , $r \in \{1, \dots, n\}$ genannt.

Die Wahrscheinlichkeit, dass die/der Spieler/in mit der Strategie σ_r gewinnt, bezeichnet man mit U_r .

Folgendes Lemma zeigt Aussagen zur Wahrscheinlichkeit mit der Strategie σ_r zu gewinnen.

Lemma 4 [9] *Wahrscheinlichkeit der Strategie σ_r*

$$U_1 = \frac{1}{n}$$

$$U_r = \frac{r-1}{n} \cdot \sum_{j=r}^n \frac{1}{j-1}, \text{ falls } r \geq 2 \text{ ist.}$$

Der folgende Beweis dazu enthält Aussagen aus der Vorlesung zur Spieltheorie von Claus Hertling [9].

Beweis:

Das j -te Gut ist mit Wahrscheinlichkeit $\frac{1}{n}$ das beste. Für den Fall $j = 1$ gilt $U_1 = \frac{1}{n}$. Es gilt, dass $r \geq 2$ ist.

$$U_r = \sum_{j=r}^n P(\text{das } j\text{-te Gut wird mit Strategie } \sigma_r \text{ gewählt und ist das beste})$$

$U_r = \sum_{j=r}^n P(j\text{-te Gut ist das beste}) \cdot P(\text{das beste Gut } 1, \dots, j-1 \text{ befindet sich unter den Gütern } 1, \dots, r-1)$

$$U_r = \sum_{j=r}^n \frac{1}{n} \cdot \frac{r-1}{j-1}$$

$$U_r = \frac{r-1}{n} \sum_{j=r}^n \frac{1}{j-1} \quad \square$$

Die Wahrscheinlichkeit mit dieser Strategie zu gewinnen, lässt sich für kleine n gut berechnen. Bei größerem n kann das maximale U_r von r abgeschätzt werden, was jedoch nicht gemacht wird, da Lindley (1961) einen weiteren Ansatz für große n gefunden hat.

Im r -ten Schritt kann die/der Spieler/in noch entscheiden, mit welcher von den Strategien σ_j , $j \geq r$ er spielen möchte. Dabei wäre die Wahrscheinlichkeit das beste Gut zu wählen

$$V_r := \max(U_j | j \geq r).$$

Folgendes Theorem enthält mathematische Aussagen aus der Vorlesung zur Spieltheorie von Claus Hertling [9].

Theorem 1 [9]

(a)

$$V_r = \frac{r-1}{r} \cdot V_{r+1} + \frac{1}{r} \cdot \max\left(\frac{r}{n}, V_{r+1}\right)$$

(b) Es existiert ein R , sodass

$$V_1 = V_2 = \dots = V_R > V_{R+1} > V_{R+2} > \dots > V_n.$$

Maximales R mit

$$\sum_{j=R}^n \frac{1}{j-1} \geq 1$$

Es gilt

$$U_j = V_j \text{ für } j \geq R,$$

$$U_j \leq V_j \text{ für } j < R.$$

(c) Seien $\sigma_j, j \in 1, \dots, n$ Strategien, so ist die Strategie σ_R die optimale.

(d) Es gilt

$$R \in \left(\frac{n}{e}, \frac{n}{e} + (2 - e^{-1}) \right)$$

Das Intervall besteht aus einer bzw. zwei natürlichen Zahlen. Es gilt somit, dass das Intervall, falls es aus einer natürlichen Zahl besteht, eindeutig bestimmt ist.

Es gilt auch, da $U_R = \frac{R}{n}$, dass

$$\lim_{n \rightarrow \infty} \frac{R}{n} = \frac{1}{e}$$

$$\lim_{n \rightarrow \infty} U_R = \frac{1}{e}$$

Beweis:

(a) Der erste Summand $V_r = \frac{r-1}{r} \cdot V_{r+1}$ der Formel folgt aus dem Fall "das r -te Gut ist nicht das beste unter den ersten r Gütern". $\frac{r-1}{r}$ ist die Wahrscheinlichkeit dafür, dass das r -te Gut unter den ersten r Gütern nicht das beste ist. In diesem Fall geht die/der Spieler/in weiter zum $r+1$ -ten Gut.

Der zweite Summand $\frac{1}{r} \cdot \max(\frac{r}{n}, V_{r+1})$ der Formel folgt aus dem Fall "das r -te Gut ist das beste unter den ersten r Gütern". $\frac{1}{r}$ ist die Wahrscheinlichkeit dafür, dass das r -te Gut unter den ersten r Gütern das beste ist. In diesem Fall bleibt der Spielerin bzw. dem Spieler die Entscheidung, ob er das Gut wählt oder ob er weiterspielt und zum $r+1$ -ten Gut übergeht. Betrachtet man das Ereignis "das r -te Gut ist auch das beste aller n Güter, wenn das r -te Gut das beste Gut unter den ersten r Gütern ist", dann erhält man die bedingte Wahrscheinlichkeit

$$\frac{P(\text{das beste aller } n \text{ Güter})}{P(\text{das beste der ersten } r \text{ Güter})} = \frac{1/n}{1/r} = \frac{r}{n}.$$

□

(b) *Induktionsbeweis:*

Es sei die maximale Zahl $R \in 1, \dots, n$ mit

$$\sum_{j=R}^n \frac{1}{j-1} \geq 1.$$

Induktionsannahme: Es gilt für $k+1 > R$, dass

$$U_{k+1} = V_{k+1} > U_{k+2} = V_{k+2} > \dots > U_n = V_n.$$

Induktionsanfang: $V_n = U_n = \frac{1}{n}$.

Induktionsschritt: Es gelte auch $k \geq R$. Nach Lemma 1 gilt

$$U_{k+1} = \frac{k}{n} \cdot \sum_{j=k+1}^n \frac{1}{j-1}.$$

Außerdem wird (a) und $V_{k+1} = U_{k+1}$ verwendet.

$$\begin{aligned} V_k &= \frac{k-1}{k} \cdot V_{k+1} + \frac{1}{k} \cdot \max\left(\frac{r}{n}, V_{r+1}\right) \\ V_k &= \frac{k-1}{n} \cdot \sum_{j=k+1}^n \frac{1}{j-1} + \frac{1}{n} \cdot \max\left(1, \sum_{j=k+1}^n \frac{1}{j-1}\right) \\ V_k &= \frac{k-1}{n} \cdot \sum_{j=k+1}^n \frac{1}{j-1} + \frac{1}{n} \\ V_k &= \frac{k-1}{n} \cdot \sum_{j=k}^n \frac{1}{j-1} \\ V_k &= U_k \end{aligned} \tag{1}$$

Außerdem wird durch die Rechnung auch

$$V_k > \frac{k-1}{k} \cdot V_{k+1} + \frac{1}{k} \cdot V_{k+1} = V_{k+1}$$

gezeigt.

Durch diesen Induktionsbeweis werden Aussagen für $j \geq R$ in (b) gezeigt.

Aus $\sum_{j=R}^n \frac{1}{j-1} \geq 1$ folgt $\frac{R-1}{n} \leq V_R$. Daher gilt $V_{R-1} = V_R$.

Für $j \leq R$ folgt induktiv

$$\frac{j-1}{n} \leq V_R = V_j$$

und

$$V_{j-1} = V_j,$$

und somit auch

$$V_1 = V_2 = \dots = V_R.$$

$U_j \leq V_j$ ist wegen der Definition von V_j trivial. □

(c) Folgt direkt aus (b).

(d) R ist durch $\sum_{j=R+1}^n \frac{1}{j-1} < 1 \leq \sum_{j=R}^n \frac{1}{j-1}$ bestimmt.

Man kann die Summen mithilfe von Integralen abschätzen. Dazu deutet man sie als Ober- bzw. Untersumme eines Integrals wie folgt.

$$\sum_{j=R}^{n-1} \frac{1}{j} > \int_R^n \frac{1}{x} dx = \log n - \log R, \quad (2)$$

$$\sum_{j=R-1}^{n-1} \frac{1}{j} < \int_{R-2}^{n-1} \frac{1}{x} dx = \log(n-1) - \log(R-2). \quad (3)$$

Man bekommt

$$\log n - \log R < 1 < \log(n-1) - \log(R-2)$$

und somit

$$\frac{n}{e} < R < \frac{n-1}{e} + 2 = \frac{n}{e} + (2 - e^{-1}).$$

Somit ist die Wahrscheinlichkeit, dass man mit Strategie σ_R gewinnt

$$U_R = \frac{R-1}{n} \cdot \sum_{j=R}^n \frac{1}{j-1}.$$

Für große n kann man sagen, dass die Wahrscheinlichkeit bei $U_R \approx \frac{R-1}{n} \approx \frac{1}{e}$ liegt. □

Beispiel 3.2.1

Es werden $n = 10$ Güter geboten, die unterschiedlich gewertet sind und deren Werte vor der Präsentation noch unbekannt sind. Man bezeichne die Güter mit (G_1, \dots, G_{10}) . Die Güter werden in einer zufälligen Reihenfolge präsentiert. Die Gelegenheit ein Gut zu wählen, kann nur bei der Präsentation ergriffen werden, sobald es abgelehnt wird, kann nicht mehr zurückgegriffen werden.

Das Ziel ist es - wie bei den bisherigen Beispielen - das beste Gut zu wählen.

Man bezeichne A als das Ereignis *Es wird das beste Gut ausgewählt*. Die Wahrscheinlichkeit, dass A eintritt, ist von der Auswahlstrategie abhängig.

Gesucht ist also eine Strategie, mit der $p(A)$ möglichst groß ist. Die einzig gute Wahl, ist die des besten Gutes.

Es werden zwei Strategien näher betrachtet.

Strategie 1:

Wähle willkürlich ein Gut aus.

Wenn die Wahl willkürlich getroffen wird, so ergibt sich für die Wahrscheinlichkeit $P(A) = \frac{1}{10}$. Dies folgt, weil jedes der 10 Güter mit der gleichen Wahrscheinlichkeit das beste ist.

Strategie 2:

Beobachte eine Anzahl an Gütern und merke dir das beste Gut dieser Anzahl (nenne es Maximum). Wähle nun aus den nachfolgenden Gütern das erste Gut, das besser ist, als das Maximum.

Bei dieser Strategie kommt die Frage auf, wie lange man wartet bzw. wie viele Güter man ziehen lässt. Im Folgenden wird diese Wartezeit *Beobachtungszeit* genannt.

Ist die Beobachtungszeit zu lang, so ist die Wahrscheinlichkeit höher, dass man damit das beste Gut ziehen lässt und keines der weiteren Güter besser ist.

Angenommen man wählt als Beobachtungszeit 5 Güter. Das heißt, dass man die ersten 5 Güter (G_1, G_2, G_3, G_4, G_5) ziehen lässt und sich das beste davon merkt (das Maximum).

Falls einer der übrigen Werte ($G_6, G_7, G_8, G_9, G_{10}$) der größte ist, so tritt Ereignis A ein, und man wählt das beste Gut.

Man definiert A_k als das Ereignis: G_k ist der größte Wert und G_k wird ausgewählt. Dabei ist $k = 6, \dots, 10$.

Daraus folgt: $A = A_6 \cup A_7 \cup A_8 \cup A_9 \cup A_{10}$

und weil die Ereignisse A_k disjunkt sind, folgt somit auch :

$$P(A) = P(A_6) + P(A_7) + P(A_8) + P(A_9) + P(A_{10})$$

Allgemein:

$$P(A_k) = P(G_k \text{ wird genommen} \mid G_k \text{ ist bestes Gut}) \cdot P(G_k \text{ ist bestes Gut})$$

$$G_k \text{ ist das beste Gut mit } P(G_k \text{ ist bestes Gut}) = \frac{1}{10}.$$

G_k wird laut Strategie genommen, wenn keines der Güter G_6, \dots, G_{k-1} besser ist, als das Maximum der Güter G_1, \dots, G_5 . Dieses Ereignis kommt in 5 aus $k - 1$ Möglichkeiten vor.

$$P(G_k \text{ wird genommen} \mid G_k \text{ ist bestes Gut}) = \frac{5}{k-1}$$

$$\text{Es ergibt sich also } P(A_k) = \frac{5}{k-1} \cdot \frac{1}{10} = \frac{1}{2k-2}.$$

Nun folgt die Berechnung der einzelnen Wahrscheinlichkeiten.

$$P(A_6) = P(G_6 \text{ wird genommen} \mid G_6 \text{ ist bestes Gut}) \cdot P(G_6 \text{ ist bestes Gut})$$

$$P(A_6) = 1 \cdot \frac{1}{10}$$

$$P(A_7) = P(G_7 \text{ wird genommen} \mid G_7 \text{ ist bestes Gut}) \cdot P(G_7 \text{ ist bestes Gut})$$

$$P(A_7) = \frac{5}{6} \cdot \frac{1}{10}$$

$$P(A_8) = P(G_8 \text{ wird genommen} \mid G_8 \text{ ist bestes Gut}) \cdot P(G_8 \text{ ist bestes Gut})$$

$$P(A_8) = \frac{5}{7} \cdot \frac{1}{10}$$

$$P(A_9) = P(G_9 \text{ wird genommen} \mid G_9 \text{ ist bestes Gut}) \cdot P(G_9 \text{ ist bestes Gut})$$

$$P(A_9) = \frac{5}{8} \cdot \frac{1}{10}$$

$$P(A_{10}) = P(G_{10} \text{ wird genommen} \mid G_{10} \text{ ist bestes Gut}) \cdot P(G_{10} \text{ ist bestes Gut})$$

$$P(A_{10}) = \frac{5}{7} \cdot \frac{1}{10}$$

Somit erhält man insgesamt $P(A) = \frac{1}{10} + \frac{1}{12} + \frac{1}{14} + \frac{1}{16} + \frac{1}{18} \approx 0,373$

Bei einer Beobachtungszeit von 5 Gütern, wählt man mit einer Wahrscheinlichkeit von 37% das beste Gut.

In einem der folgenden Kapitel wird dieses Beispiel nochmals aufgefasst. Dabei wird gezeigt, wie viele Güter man im allgemeinen Fall beobachten sollte, um danach das beste aller Güter mit höchster Wahrscheinlichkeit zu wählen.

3.3 Odds-Strategie / Bruss-Algorithmus

Es gibt einige verschiedene Varianten des SekretärInnenproblems. In diesem Unterkapitel wird die Variante und die dazugehörige Lösung von Franz Thomas Bruss aus dem Jahr 1984 näher erläutert. Die dazu benötigten Begriffe und Definitionen wurden in Kapitel 2 bereits erwähnt. Im Folgenden wird die Variante des Sekretärinnenproblems von Franz Thomas Bruss geschildert.

Franz Thomas Bruss beginnt seine Suche nach einer Strategie zur maximalen Erfolgswahrscheinlichkeit für ein festes n mit der *Odds-Strategie* bzw. dem *Bruss-Algorithmus*.

Definition 38 .

Es seien X_1, X_2, \dots, X_n eine Folge unabhängiger Bernoulli-ZV. Die Zufallsvariablen X_k haben eine Erfolgswahrscheinlichkeit p_k . Trivial ist, dass $q_k = 1 - p_k$ die Wahrscheinlichkeit ist, dass kein Erfolg eintritt. Es ergibt sich dadurch der Quotient

$$r_k = \frac{p_k}{q_k}$$

welcher auch der Strategie ihren Namen gibt (Quotient im englischen Odds).

Sei $R(k) := \sum_{i=k}^n r_i$ die Summe der Quotienten (Odds) rückwärts aufsummiert.

Die Ereignisse werden der Reihe nach beobachtet und die Wahrscheinlichkeit, beim letzten Erfolg zu stoppen, wird mit folgender Strategie maximiert:

Stoppe wenn für k gilt, dass $k \geq s$ und $X_k = 1$, dabei ist

$$s = \begin{cases} \sup\{k \in \{1, \dots, n\} : R(k) \geq 1\}, & R(1) \geq 1 \\ 0 & \text{sonst} \end{cases}$$

Um diese Strategie anwenden zu können, wird ein Stoppindex s benötigt, der den Index k stoppt. Dieser Stoppindex s stoppt, sobald die Summe der Quotienten 1 erreicht oder übertraffen hat.

Sei $Q(s) = \prod_{i=s}^n q_i$ die Wahrscheinlichkeit, dass unter den in Frage kommenden Zufallsvariablen keine erfolgreiche dabei ist.

Für die *Erfolgswahrscheinlichkeit* V dieser Strategie gilt

$$V = R(s) \cdot Q(s) = \sum_{i=s}^n r_i \cdot \prod_{j=s}^n q_j.$$

Bei $R(1) \geq 1$ gilt dann $V > 1/e$.

Die Definition wird nun im folgenden Beispiel angewendet und verdeutlicht.

Beispiel

Es wird das SekretärInnen Problem mit bekannter BewerberInnen Anzahl als Beispiel herangezogen.

Angenommen ein/e Chef/in hat 20 BewerberInnen in beliebiger Reihenfolge. Es gelten die Voraussetzungen des typischen SekretärInnen Problems. Die/Der Chef/in schaut sich die BewerberInnen an, und entscheidet ob diese/r gewählt ist oder nicht. Die BewerberInnen können bei Ablehnung nicht mehr zurückgerufen werden. Sobald die/der Chef/in sich für eine/n Bewerber/in entschieden hat, kommen keine weiteren Bewerber/innen in Frage.

Weil die/der Chef/in 20 BewerberInnen hat, gilt $n = 20$.

Ein/e Bewerber/in ist für die/den Chef/in dann ein Erfolg, wenn sie/er besser ist, als alle BewerberInnen davor. Somit ergeben sich die Erfolgswahrscheinlichkeiten

$$p_1 = 1, p_2 = \frac{1}{2}, \dots, p_{20} = \frac{1}{20}.$$

Also allgemein $p_k = \frac{1}{k}$.

Daraus und aus $q_k = 1 - p_k$ folgt $q_k = \frac{k-1}{k}$.

Ebenso folgt für $r_k = \frac{p_k}{q_k} = \frac{1}{k-1}$.

Es gilt nun den Stoppindex $s = 7$ herauszufinden. Dabei hilft die folgende Tabelle.

Die Tabelle 1 zeigt zur Übersicht gerundete Werte, gerechnet wurde jedoch mit ungerundeten Werten.

Bewerber/in	p_k	q_k	r_k	R_k
1	1	0	-	-
2	0.5	0.5	1	3.53
3	0.33	0.67	0.49	2.53
4	0.25	0.75	0.33	2.04
5	0.2	0.8	0.25	1.71
6	0.17	0.83	0.2	1.46
7	0.14	0.86	0.16	1.26
8	0.13	0.87	0.15	1.1
9	0.11	0.89	0.12	0.95
10	0.1	0.9	0.11	0.83
11	0.09	0.91	0.1	0.72
12	0.08	0.92	0.09	0.62
13	0.08	0.92	0.09	0.53
14	0.07	0.93	0.08	0.44
15	0.07	0.93	0.08	0.36
16	0.06	0.94	0.06	0.28
17	0.06	0.94	0.06	0.22
18	0.06	0.94	0.06	0.16
19	0.05	0.95	0.05	0.1
20	0.05	0.95	0.05	0.05

Tabelle 1: Berechnung der gerundeten Werte

Anhand der Tabelle ist zu erkennen, dass R_k bei Bewerber/in 8 den Wert 1 überschreitet. Daraus lässt sich für den Stoppindex $s = 8$ folgern.

Das bedeutet, dass die/der Chef/in die ersten 8 BewerberInnen abwarten sollte und die/den erste/n BewerberIn nehmen soll, die/der besser ist, als alle bisherigen. Dabei ergibt sich eine Erfolgswahrscheinlichkeit von

$$V = R(s) \cdot Q(s) = 1.1016 \cdot 0.315 = 0.34650216.. \approx 0.35.$$

Die/Der Chef/in wählt in 35% aller Fälle die/den besten Bewerber/in.

3.4 x-Strategie

Definition 39 *x-Strategie*

Es sei $x \in [0, \tau]$. Eine Strategie nennt man x-Strategie, falls im Zeitraum $[0, x]$ ausschließlich beobachtet wird und erst danach das Gut genommen und das Spiel beendet wird.

Bemerkung

Zur Berechnung der Erfolgswahrscheinlichkeit V der x -Strategie werden die ersten $k + 1$ Ränge betrachtet, mit $1 \leq k \leq n - 1$.

Das Gut mit dem ranghöchsten Wert $k + 1$ liegt wegen der Gleichverteilung mit Wahrscheinlichkeit x im Intervall $[0, x]$. Für die Wahrscheinlichkeit, dass die übrigen k -höchsten im Intervall $(x, 1]$ liegen, gilt $(1 - x)^k$.

3.5 1/e-Strategie

Wie bereits erwähnt, strebt $p_{n,optimal}$ gegen $1/e$. Es ergibt sich somit das $1/e$ -Gesetz zum zeitstetigen Modell von Franz Thomas Bruss, welches von Verteilungsfunktionen von N unabhängig ist.

3.6 Die 37-Prozent-Regel - Geoffrey Miller

Geoffrey Miller beschreibt in seinem Buch *The Mating Mind* [6] die 37-Prozent-Regel, die seiner Meinung nach die Lösung des SekretärInnen Problems sein soll. Er beschreibt statt dem oben genannten Sekretärinnenproblem ein Verfahren zur Partnersuche. Die Regel von Miller knüpft stark an die $1/e$ -Strategie von Bruss an, wie im Folgenden beschrieben wird. Es besteht Verwechslungsgefahr, da das $1/e$ -Gesetz bei bekannter Anzahl von n Kandidatinnen bzw. Gütern gilt. Miller geht in seinem *Take a Dozen*-Verfahren von einer unbekanntenen Menge n aus.

Die 37-Prozent-Regel besagt, dass die/der Chef/in bzw. Spieler/in die ersten 37% der Bewerberinnen bzw. Güter nicht nimmt und nach den 37% die erste Bewerberin/das erste Gut nimmt, die/das besser ist, als die/das beste von den bisherigen.

zu Beispiel 3.2.1

An dieser Stelle wird an das *Beispiel 3.2.1* angeknüpft.

Sei i die Anzahl der Güter, die beobachtet werden.

Es sei analog wie oben definiert:

$$P(A) = \sum_{k=i+1}^n P(A_k)$$

$$P(A_k) = P(G_k \text{ wird genommen} \mid G_k \text{ ist bestes Gut}) \cdot P(G_k \text{ ist bestes Gut})$$

$$P(A_k) = \frac{i}{k-1} \cdot \frac{1}{n}$$

$$P(A) = \sum_{k=i+1}^n \frac{i}{k-1} \cdot \frac{1}{n} = \frac{i}{n} \cdot \sum_{k=i}^{n-1} \frac{1}{k} =: P_i$$

Nun wird allgemein untersucht, für welches i die Wahrscheinlichkeit P_i maximal ist.

Im Beispiel von oben gilt $n = 10$. Hier ist die Antwort durch eine Tabelle (siehe Tabelle 2) mit den berechneten Werten ersichtlich.

Es ergibt sich folgende Strategie für $n = 10$ Güter.

Beobachte die ersten drei Güter G_1, G_2, G_3 und bestimme ihr Maximum. Wähle aus den darauffolgenden Gütern das erste Gut, welches besser ist als das Maximum.

Mit dieser Strategie wählt man mit nahezu 40% das beste aller Güter. Es besteht jedoch auch das Risiko von 30%, dass man das beste Gut nicht erhält, da es unter den ersten drei Beobachtungsgütern sein könnte.

Allgemein:

k	$P(A_k)$
1	0,283
2	0,366
3	0,399
4	0,398
5	0,373
6	0,327
7	0,265
8	0,189
9	0,1
10	0

Tabelle 2: Berechnung zu Beispiel 3.2.1

Nach dem obigen Beispiel wird nun allgemein untersucht, für welches i die Wahrscheinlichkeit P_i maximal ist.

Sei dabei $P_{i-1} < P_i$ und $P_{i+1} < P_i$.

Aus der ersten Ungleichung ergibt sich:

$$(i-1) \cdot \left(\frac{1}{i-1} + \frac{1}{i} + \dots + \frac{1}{i-1} \right) < i \cdot \left(\frac{1}{i} + \frac{1}{i+1} + \dots + \frac{1}{i-1} \right)$$

Nach einfachen Rechenschritten erhält man folgende äquivalente Ungleichung.

$$\frac{1}{i} + \frac{1}{i+1} + \dots + \frac{1}{i-1} > 1$$

Ebenso folgt aus der zweiten Ungleichung und einfachen Umformungen:

$$\frac{1}{i+1} + \frac{1}{i+2} + \dots + \frac{1}{i-1} < 1$$

Es ergibt sich somit für jedes n genau ein i , für das gilt

$$\frac{1}{i+1} + \frac{1}{i+2} + \dots + \frac{1}{n-1} < 1 < \frac{1}{i} + \frac{1}{i+1} + \dots + \frac{1}{n-1}$$

Für das obige Beispiel $n = 10$ erhält man für $i = 3$.

Bemerkung

Für große n ergibt sich ein interessanter Sachverhalt, der mit Hilfe der Analysis bewiesen werden kann.

(i) P_i ist maximal mit $i = \frac{n}{e} \approx 0,368 \cdot n$

(ii) Das Maximim von P_i liegt ungefähr bei $\frac{1}{e} \approx 0,368$

Daraus kann man schließen, dass man ungefähr 37% der Güter beobachten soll. In ungefähr 37% der Fälle wählt man das beste Gut.

Fazit

Um sich mit dem *Stoppproblem* bzw. *Heiratsproblem* mathematisch auseinandersetzen zu können, bedarf es eines grundlegenden Wissens über Folgen, Reihen, elementare Stochastik und weitere mathematische Grundlagen, wie die Maßtheorie und Stochastische Prozesse.

Für die Stoppptheorie spielt die Markovsche Entscheidungstheorie zum optimalen Stoppen eine wichtige Rolle. Zudem wird in der vorliegenden Arbeit darauf eingegangen, wie man zur Lösung einer optimalen Stoppzeit mittels Markovprozessen gelangt.

Abschließend wird das SekretärInnenproblem als Optimierungsproblem vorgestellt und drei Strategien dazu erläutert. Die Strategie von Lindley bezieht sich auf die gleichen Voraussetzungen wie beim SekretärInnenproblem, jedoch nennt er das Szenario Heiratsproblem. Seine Strategie ist es die ersten $r - 1$ Güter bzw. SekretärInnen abzulehnen und erst ab dem r -ten Schritt das erste bessere Gut zu wählen. Die Wahrscheinlichkeit, mit dieser Strategie bei einer großen Anzahl von Gütern zu gewinnen, liegt bei ungefähr $\frac{1}{e}$.

Bruss veröffentlichte die Odds-Strategie, die sich auf den Quotienten $r_k = \frac{p_k}{q_k}$ bezieht. Bei dieser Strategie gilt es, den besten Stoppzeitpunkt für das Beobachten der Güter herauszufinden. Nach dem Stoppzeitpunkt wird das erste Gut gewählt, welches besser ist als das beste unter den bereits beobachteten Gütern. Hat man den Stoppindex mittels vorgegebenen Algorithmus gefunden, so wählt man mit einer Wahrscheinlichkeit von ungefähr 35% das beste Gut.

Man betrachtet die $1/e$ -Strategie als eine optimale Strategie aus der Klasse von x -Strategien. Ob die $1/e$ -Strategie die optimalste unter allen Strategien ist, steht noch in Frage. Für ein festes $N = n$ ist bekannt, dass die x_n -Strategie die optimalste ist.

Hat man keinerlei Information über N , so könnte die $1/e$ -Strategie die optimalste sein. Es kommt aber auch die Fragestellung auf, ob es überhaupt eine optimale Strategie geben kann, wenn es keinerlei Informationen über N gibt. Bruss selbst und auch andere Forschenden in seinem Kreis zweifelten an der Korrektheit und somit an der Deutlichkeit der Fragestellung. Die Diskussion darüber wird in der Arbeit jedoch nicht näher behandelt.

Literatur

- [1] Thorsten Pampel. Reihen, pages 55–62. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
- [2] Ehrhard Behrends. Markovprozesse und stochastische Differentialgleichungen. Springer Verlag, 2013.
- [3] Heinz Bauer. Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie. De Gruyter, Berlin, Boston, 1974.
- [4] Norbert Henze. Stochastik: eine Einführung mit Grundzügen der Maßtheorie: inkl. zahlreicher Erklärvideos. Springer-Verlag, 2019.
- [5] Götz Kersting and Anton Wakolbinger. Stochastische Prozesse. Springer, 2014.
- [6] Geoffrey Miller. The mating mind: How sexual choice shaped the evolution of human nature. Anchor, 2011.
- [7] Thomas S Ferguson. Who solved the secretary problem? Institute of Mathematical Statistics, 4(3):282–289, 1989.
- [8] Denis V. Lindley. Dynamic programming and decision theor. Journal of the Royal Statistical Society: Series C (Applied Statistics), 10(1):39–51, 1961.
- [9] Claus Hertling. Spieltheorie ii 2-stündige vorlesung im fss 2018 mannheim. Universität Mannheim, 2018.
- [10] Harald Luschgy. Martingale in diskreter Zeit: Theorie und Anwendungen. Springer-Verlag, 2012.
- [11] Harald Luschgy and Harald Luschgy. Stoppzeiten und lokale martingale. Martingale in diskreter Zeit: Theorie und Anwendungen, pages 35–64, 2013.
- [12] Kordecki Wojciech. Secretary problem: Graphs, matroids and greedoids. Operations Research Forum 2: 63, 2021.
- [13] Klaus D Schmidt. Maß und Wahrscheinlichkeit. Springer-Verlag, 2011.
- [14] Andrey Kolmogoroff. Grundbegriffe der wahrscheinlichkeitsrechnung. Springer, 1933.
- [15] Goran Peskir and Albert Shiryaev. Optimal stopping and free-boundary problems. Springer, 2006.
- [16] Franz Thomas Bruss. A unified approach to a class of best choice problems with an unknown number of options. The Annals of Probability, pages 882–889, 1984.
- [17] Norbert Henze. Der große Umordnungssatz für Reihen. Karlsruher Institut für Technologie, 2019.
- [18] Burkhard Lenze. Folgen und Reihen, pages 93–126. Springer Fachmedien Wiesbaden, Wiesbaden, 2020.