Algorithmische Mathematik II

Dozent
Professor Dr. Patrik Ferrari

Mitschrift
Maximilian Kessler

Version
1. Juni 2021 11:06

Zusammenfassung

Bei folgenden Vorlesungsnotizen handelt es sich um (inoffizielle) Mitschriften zur Vorlesung 'Algorithmische Mathematik II', die im Sommersemester 2021 an der Universität Bonn gehalten wird. Ich garantiere weder für Korrektheit noch Vollständigkeit dieser Notizen, und bin dankbar für jegliche Art von Korrektur, sowohl inhaltlich, als auch Tippfehler.

Bemerkungen, die nicht zum eigentlichen Vorlesungsinhalt gehören, wurden mit einem * gekennzeichnet. Sie werden nach eigenem Ermessen hinzugefügt, um weitere Details oder evtl. mündliche Anmerkungen beizufügen.

Manche Umgebungen sind mit einem † versehen. Das ist dann der Fall, wenn ihr Inhalt so, oder zumindest in sehr ähnlicher Form, in der Vorlesung vorkam (unter Umständen auch mündlich), ich aber die Umgebung der Aussage geändert habe. Insbesondere also, wenn ich aus Aussagen, die einfach erwähnt werden, ein **Lemma**† mache, um sie hervorzuheben.

Weitere Informationen finden sich bei GitHub oder auf der Vorlesungshomepage

Inhaltsverzeichnis

Übersicht der Vorlesungen

Vorlesung 1
Mo 12 Apr 2021 10:16

- Es gibt einen Helpdesk, auch explizit einen nur für Studentinnen.
- Die Vorlesung wird aufgenommen, und zwar ohne Videos der Teilnehmenden sowie des Dozenten. Die Aufzeichnungen werden anschließend in Sciebo hochgeladen.
- Es gibt ein Diskussionsforum für Fragen (auf eCampus).
- Ab heute Abend, 18 Uhr (Mo 12 Apr 2021 18:00), kann man sich auf eCampus für die Übungsgruppen registrieren und dies endet am Dienstag Abend um 24 Uhr (Di 12 Apr 2021 24:00) - es wird versucht, die Studenten gleichmäßig zu verteilen.
- Falls ihr in der Warteliste landet und gewünscht ist, in der Gruppe abzugeben, schreibt eine Mail mit den gewünschten Abgabepartnern, dann kann eine gemeinse Einteilung erfolgen.
- Es gibt auch das Modul `AlmaIIb`. Registriert euch noch nicht, dies ist für den 2. Teil der Vorlesung notwendig.
- Die Abgabe der Übungsblätter erfolgt einheitlich jeden Freitag um 12 Uhr.
- Gruppenabgaben sind erlaubt, bis zu einer Größe von maximal 4 StudentInnen.
- Das 1. Blatt ist freiwillig und gibt Bonuspunkte.
- Für die Klausurzulassung werden 50% der Punkte benötigt. Von den Programmieraufgaben müssen mindestens 4 von 6 zufriedenstellend bearbeitet werden.
- Programmieraufgaben gibt es ab dem 2. Übungsblatt auf jedem 2. Blatt. Die Bearbeitungszeit beträgt dann 2 Wochen.

Einleitung

In der Vorlesung wird folgendes betrachtet werden:

Teil 1: Diskrete Stochastik
- Zufallsvariablen
- Bedingte Wahrscheinlichkeiten
- Unabhängigkeit von Variablen
- Monte-Carlo Methoden

Teil 2: Numerische Analysis
- Iterative Verfahren
- Interpolation von Daten (durch Polynome, trigonometrische Funktionen, ...)
- Numerische Verfahren für die Integration

1 Diskrete Stochastik

1.1 Einleitung

Ziel. Beschreibung von Systemen, die einen Anteil an **Zufall** haben, d.h. nicht 100% deterministisch sind.

> **Beispiel.**
> - Spiele: Kartenspiele, Glücksspiele, ...
> - Statistik: Umfragen, Versicherungen
> - Komplexe Systeme: Wettermodelle, Finanzmärkte

Was sind Quellen von Zufall?
- Zu komplexe Systeme: Zufälliges Aussehen des Gesamteffekts
- Fehlende Informationen (z.B. bei einem Kartenspiel)
- Chaotische Systeme (beispielsweise das Wetter)

- Intrinsisch unvorhersagbare Systeme (z.B. radioaktiver Zerfall)

Frage. (1) Wie modelliert man ein System mit Zufall?
(2) Wie simuliert man ein System mit Zufall? (anwendungstechnischer)
(3) Welche Voraussagen kann man machen?

Beispiel. Die **Brown'sche Bewegung**. Das System ist implizit ein Pollen mit vielen Wassermolekülen ($\sim 10^{23}$), die sich im Prinzip deterministisch bewegen.
$\Rightarrow$ Wir erhalten ein Gleichungssystem mit $(N+1) \cdot 6$ (3 Positionen, 3 Geschwindigkeit) Variablen. Dieses ist de facto unlösbar.

Was wollen wir hier eigentlich untersuchen? -> Die Bewegung des Pollens, jedoch nicht die der einzelnen Wassermoleküle.
In einer **Modellierung** ersetzt man die Stöße, die durch die Wassermoleküle entstehen, durch **zufällige Stöße**.

Diskretes Modell: Die Zeit bewegt sich in $n \in \{0, 1, 2, \ldots\}$. Sei

$$Z(n) := (\text{Position des Pollens zur Zeit } n) \in \mathbb{Z}^3.$$

OBdA setzen wir $Z(0) = 0$.
Dynamik: $Z(n+1) = Z(n) + \xi_n$, wobei wir ξ_n aus dem Ergebnis eines Würfelwurfs bestimmen werden:

$$\xi_n = \begin{cases} (1,0,0) & \text{wenn Würfel} = 1 \\ (-1,0,0) & \text{wenn Würfel} = 2 \\ (0,1,0) & \text{wenn Würfel} = 3 \\ (0,-1,0) & \text{wenn Würfel} = 4 \\ (0,0,1) & \text{wenn Würfel} = 5 \\ (0,0,-1) & \text{wenn Würfel} = 6 \end{cases}.$$

Frage. Welche Fragen können wir mit solch einem System nun beantworten? Was pasiert, wenn $n \gg 1$?

a) Typischerweise erhalten wir $|Z(n)| = O(\sqrt{n})$

b) Wenn wir die Frequenz von $[Z(n)]_i$ betrachten, (also bei welcher Koordinate in Richtung i befinden wir uns nach n Würfen) sehen wir typischerweise: Diese Verteilung sieht für $n \gg 1$ ungefähr wie die Gaussglocke aus.

Skalierung: Wir setzen nun

$$B(t) = \lim_{n \to \infty} \frac{Z(\lfloor nt \rfloor)}{\sqrt{n}}.$$

und dies ist dann die Brownsche Bewegung.
Nun möchten wir Vorhersagen treffen können:

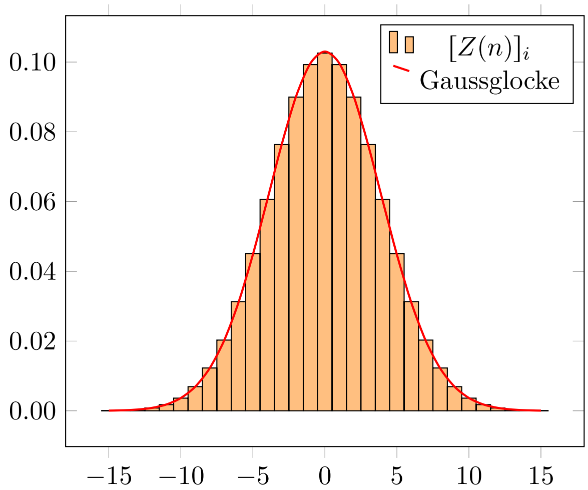


Abbildung 1: Binomialverteilung und Gaussglocke

Frage. Ist $Z(n)$ in einer gegebenen Menge A?

Das kann man (im Allgemeinen) nicht einfach mit 'Ja' oder 'Nein' beantworten. Stattdessen müssen wir folgende Frage stellen:

Frage. Wenn man $Z(n)$ beobachtet, wie häufig wird $Z(n)$ in A sein?

Diese Frage lässt sich mit einer Zahl $\in [0,1]$ beantworten.

1.2 Ereignisse und Wahrscheinlichkeiten

Wir benötigen 3 Grundelemente:

(1) Die Menge Ω von möglichen **Ergebnissen** - die Elemente von Ω heißen auch **Elementarereignisse**.

(2) Die Menge $\mathcal{F}$ der **Ereignisse** - ein Ereignis E ist eine Eigenschaft, die mit einer Teilmenge $G \subseteq \Omega$ assoziiert ist: $\omega \in G \Leftrightarrow$ Eigenschaft E ist erfüllt.

(3) Eine **Wahrscheinlichkeitsverteilung (auch W-maß)**:

$$\mathbb{P} : \mathcal{F} \to [0,1].$$

Bemerkung*. Wir werden noch sehen, dass gewisse Bedingungen für unsere Begriffe erfüllt sein müssen, dazu aber später mehr.

Beispiel. Eine Urne hat 12 nummerierte Kugeln (von 1 bis 12).

(1) Das **Zufallsexperiment** besteht darin, eine Kugel aus der Urne zu ziehen und die Zahl zu notieren, die auf der jeweiligen Kugel steht, d.h.

$$\Omega = \{1, \dots, 12\}.$$

Ein Elementarereignis ist nun z.B. gegeben durch $\omega = \{5\} \equiv 5$ (Vereinfachung der Notation).

(2) Mögliche Ereignisse sind beispielsweise:

$$\begin{aligned} A &= \text{'Die Zahl ist gerade'} \\ B &= \text{'Die Zahl ist } \leqslant 5\text{'} \\ C &= \text{'Die Zahl ist 8'} \end{aligned} \tag{1}$$

Die assoziierten Mengen sind damit

$$\begin{aligned} A &= \{2, 4, 6, 8, 10, 12\} \\ B &= \{1, 2, 3, 4, 5\} \\ C &= \{8\} \end{aligned} \tag{2}$$

(3) Für die Wahrscheinlichkeiten wird angenommen, dass jede Kugel die gleiche Chance hat, gezogen zu werden, also

$$\forall G \in \mathcal{F} : \mathbb{P}(G) = \frac{|G|}{|\Omega|}.$$

Nun erhält man damit die Wahrscheinlichkeiten

$$\mathbb{P}(A) = \frac{6}{12} = \frac{1}{2} \qquad \mathbb{P}(B) = \frac{5}{12} \qquad \mathbb{P}(C) = \frac{1}{12}.$$

Notation. $A \equiv \{\omega \in \Omega \mid \omega \in A\} \equiv \{\omega \in A\} \equiv \{A \text{ tritt ein}\}$

Vorlesung 2
Mi 14 Apr 2021 10:17

Wir kennen nun die Grundbegriffe $\Omega, \mathcal{F}, \mathbb{P}$ zur Beschreibung von Zufallsexperimenten. Diese wollen wir uns nun genauer ansehen:

Frage. Welche Struktur muss $\mathcal{F}$ besitzen?

Seien $A, B \in \mathcal{F}$. Dann kann man das Ereignis $A \cap B$ betrachten, also das beide der Eigeschaften eintreten. Genauso sollte

$$A^c := \Omega \backslash A.$$

, das **Komplement von A**, bzw. das **Gegenereignis** von A ebenfalls in $\mathcal{F}$ sein. Aus diesen beiden Eigenschaften folgt bereits, dass

$$A \cup B = (A^c \cap B^c)^c.$$

ebenfalls in $\mathcal{F}$ sein muss.

Eine Menge $\mathcal{F}$ mit solchen Eigenschaften heißt **Algebra**.

Notation†. Seien nun A, B und $(A_i)_{i \in I}$ Ereignisse, wobei I endlich oder abzählbar ist. Dann notieren wir die folgenden Ereignisse wie folgt:

a) $A \cup B$: $\omega \in A \cup B \Leftrightarrow \omega \in A \vee \omega \in B$, d.h. $A \cup B$ tritt ein, genau dann, wenn A eintritt oder B eintritt.
b) $\bigcup_{i \in I} A_i$: $\omega \in \bigcup_{i \in I} A_i$, wenn es ein $i \in I$ gibt, sodass $\omega \in A_i$ ist.
c) $A \cap B$: $\omega \in A \cap B \Leftrightarrow$ A und B treten ein.
d) $\bigcap_{i \in I} A_i$: $\omega \in \bigcap_{i \in I} A_i \Leftrightarrow \forall i \in I: \ A_i$ tritt ein.

(e) $A = \varnothing$ ist das Ereignis, welches nie eintritt.
$A = \Omega$ ist das Ereignis, welches immer eintritt.

Definition 1.1 (σ-Algebra)**.** Eine **σ-Algebra** ist eine nicht leere Menge $\mathcal{F}$ von Teilmengen von Ω mit den Eigenschaften:
(a) $\Omega \in \mathcal{F}$
(b) $\forall A \in \mathcal{F}\colon A^c \in \mathcal{F}$.
(c) Falls $(A_i)_{i\in I} \in \mathcal{F}$, dann auch $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.
Wir nennen $(\Omega, \mathcal{F})$ dann einen **Messraum**.

Lemma 1.2. Sei $\mathcal{F}$ eine σ-Algebra, dann gilt:
(a) $\varnothing \in \mathcal{F}$
(b) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$ und $A \cap B \in \mathcal{F}$.
(c) $(A_i)_{i\in I} \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$.

Beweis. (a) $\varnothing = \Omega^c \in \mathcal{F}$ nach Eigenschaften (a) und (b) aus der Definition.
(b) $A \cup B = A \cup B \cup \varnothing \cup \varnothing \ldots \in \mathcal{F}$ nach Eigenschaften (b) und (c).
$A \cap B = (A^c \cup B^c)^c \in \mathcal{F}$
(c) $\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c\right)^c \in \mathcal{F}$ nach (b) und (c). □

Wir haben nun $(\Omega, \mathcal{F})$ näher untersucht, es fehlt noch $\mathbb{P}$.

Frage. Welche Eigenschaften soll $\mathbb{P}$ (das Wahrscheinlichkeitsmaß bzw. die Wahrscheinlichkeitsverteilung) besitzen?

Seien $A, B \in \mathcal{F}$ mit $A \cap B = \varnothing$, d.h. A und B können nicht gleichzeitig eintreten. Dann fordern wir

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) \quad \text{(endliche Additivität).}$$

Dazu soll gelten, dass $\Omega \in \mathcal{F}$ immer eintritt, also $\mathbb{P}(\Omega) = 1 \equiv 100\%$ (Normierung).

Definition 1.3 (Wahrscheinlichkeitsverteilung)**.** Sei $(\Omega, \mathcal{F})$ ein Messraum. Eine Abbildung $\mathbb{P} : \mathcal{F} \to \mathbb{R}_+$ ist eine **Wahrscheinlichkeitsverteilung** auf $(\Omega, \mathcal{F})$, falls
(1) $\mathbb{P}(\Omega) = 1$
(2) Sind $(A_i)_{i\in I} \in \mathcal{F}$ paarweise disjunkt, so ist:

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i) \quad (\sigma\text{-Additivität}).$$

Bemerkung*. Die Definition macht implizit Gebrauch davon, dass die linke Seite überhaupt definiert ist. Dies folgt jedoch daraus, dass $\mathcal{F}$ eine σ-Algebra ist.

Definition 1.4 (Wahrscheinlichkeitsraum)**.** Ein **Wahrscheinlichkeitsraum** $(\Omega, \mathcal{F}, \mathbb{P})$ besteht aus einer Menge Ω, einer σ-Algebra $F \subseteq \mathbb{P}(\otimes)$ und einem Wahrschenilichkeitsmass $\mathbb{P}$ auf $(\Omega, \mathcal{F})$

Lemma 1.5. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Dann gilt

a) $\mathbb{P}(\varnothing) = 0$

b) $\forall A, B \in \mathcal{F}$ mit $A \cap B = \varnothing$ ist

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

c) $\forall A, B \in \mathcal{F}$ mit $A \subseteq B$ ist

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(A) + \mathbb{P}(B \backslash A) \\ \mathbb{P}(A^c) &= 1 - \mathbb{P}(A) \\ \mathbb{P}(A) &\leqslant \mathbb{P}(B) \leqslant 1 \end{aligned}$$

d) $\forall A, B \in \mathcal{F}$ ist

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ &\leqslant \mathbb{P}(A) + \mathbb{P}(B) \end{aligned}$$

e) Wenn $A_n \nearrow A$,d.h. $A_1 \subseteq A_2 \subseteq \ldots$ mit $\bigcup_{i=1}^{\infty} A_i = A$ (monotone Konvergenz von Mengen) oder $A_n \searrow A$ (d.h. $A_1 \supseteq A_2 \supseteq \ldots$ mit $\bigcap_{i=1}^{\infty} A_i = A$), so ist

$$\lim_{n \to \infty} \mathbb{P}(A_n) = \mathbb{P}\left(\lim_{n \to \infty} A_n\right) = \mathbb{P}(A).$$

Beweis. a) Wir wissen:

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \varnothing \cup \varnothing \cup \varnothing \ldots) = \mathbb{P}(\Omega) + \mathbb{P}(\varnothing) + \mathbb{P}(\varnothing) + \ldots.$$

Das Subtrahieren von $\mathbb{P}(\Omega) = 1$ liefert dann $\mathbb{P}(\varnothing) = 0$.

b) Sei $A \cap B = \varnothing$, dann ist:

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A \cup B \cup \varnothing \cup \varnothing \cup \ldots) \\ &\overset{\sigma-\text{Additivität}}{=} \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(\varnothing) + \mathbb{P}(\varnothing) + \ldots \\ &= \mathbb{P}(A) + \mathbb{P}(B) \end{aligned}$$

c) Sei $A \subseteq B$. Dann ist $B = A \cup (B \backslash A)$ eine disjunkte Vereinigung, also erhalten wir

$$\mathbb{P}(B) = \mathbb{P}(A) + \underbrace{\mathbb{P}(B \backslash A)}_{\geqslant 0} \geqslant \mathbb{P}(A).$$

Mit $B = \Omega$ ergibt sich $1 = \mathbb{P}(A) + \mathbb{P}(A^c)$.

d) Es gilt

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}((A \cup B) \backslash A) \\ &= \mathbb{P}(A) + \mathbb{P}(B \backslash (A \cap B)) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \underbrace{\mathbb{P}(A \cap B)}_{\geqslant 0} \qquad (3) \\ &\geqslant \mathbb{P}(A) + \mathbb{P}(B) \end{aligned}$$

(e) Übung

$\square$

Korollar 1.6 (Einschluss-Ausschluss-Prinzip)**.** Seien $A_1, \ldots, A_n \in \mathcal{F}$. Dann gilt

$$\mathbb{P}(A_1 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n} \mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}).$$

Beweis. Per Induktion: der Induktionsanfang lautet $\mathbb{P}(A_1) = \mathbb{P}(A_1)$ und ist offensichtlich wahr.

Die Aussage gelte nun für ein $n \in \mathbb{N}$, damit erhält man

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mathbb{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}\right) \\
&= \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) \\
&= \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbb{P}(A_{n+1}) - \mathbb{P}\left(\bigcup_{i=1}^{n} \underbrace{(A_i \cap A_{n+1})}_{=:\tilde{A}_i}\right) \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i < \ldots < i_k \leqslant n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) + \mathbb{P}(A_{n+1}) \\
&\qquad - \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathbb{P}(\underbrace{\tilde{A}_{i_1} \cap \ldots \cap \tilde{A}_{i_k}}_{A_{i_1} \cap \ldots \cap A_{i_k} \cap A_{n+1}})
\end{aligned} \tag{4}$$

Andererseits gilt aber auch:

$$\begin{aligned}
&\sum_{k=1}^{n+1} (-1)^{k-1} \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n+1} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant {\color{red}n}} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k}) && \Big\}\text{Terme mit } i_k \leqslant n \\
&+ \underbrace{\sum_{k=2}^{n+2} (-1)^{k-1} \sum_{1 \leqslant i_1 < \ldots < i_{k-1} \leqslant n} \mathbb{P}(A_i \cap \ldots \cap A_{i_{k-1}} \cap A_{n+1})}_{\overset{l:=k-1}{=} - \sum_{l=1}^{n} (-1)^{l-1} \sum_{1 \leqslant i_1 < \ldots < i_l \leqslant n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_l} \cap A_{n+1})} && \Big\}\text{Terme mit } i_k = n+1 \text{ und } k \geqslant 2 \\
&+ \mathbb{P}(A_{n+1}) && \Big\}\text{Terme mit } i_k = n+1 \text{ und } k = 1
\end{aligned}$$

und damit sieht man, dass die beiden Ausdrücke übereinstimmen, also ist der Induktionsschritt erbracht. $\square$

1.3 Diskrete Verteilungen

- Sei nun Ω endlich oder abzählbar.

- Falls wir $\mathcal{F}$ nicht explizit angeben, wird $\mathcal{F} = \mathcal{P}(\Omega)$ gewählt, d.h.

$$\mathrm{Card}(\mathcal{P}(\Omega)) \equiv |\mathcal{P}(\Omega)| = 2^{|\Omega|}.$$

Beispiel (Münzwurf)**.** Es sei $\Omega = \{K, Z\}$, wobei K für Kopf stehe und Z für Zahl. Dann ist

$$\mathcal{F} = \{\{K\}, \{Z\}, \{Z, K\}, \varnothing\}.$$

Sei $p \in [0,1]$ die Wahrscheinlichkeit, dass man Kopf erhält. Da $\mathbb{P}$ für alle Elemente aus $\mathcal{F}$ definiert sein muss, erhalten wir

$$\mathbb{P}(\varnothing) = 0 \qquad \mathbb{P}(K) = p, \qquad \mathbb{P}(Z) = \mathbb{P}(K^c) = 1-p \qquad \mathbb{P}(\{Z, K\}) = \mathbb{P}(\Omega) = 1.$$

Frage (Charakterisierung von diskreter Wahrscheinlichkeit)**.** Was müssen wir fordern, damit es $\mathbb{P}$ auf $\mathcal{P}(\Omega)$ gibt?

Beispiel. $\Omega = \{1, 2, \ldots, 10\}$ würde genügen, da dann $|\mathcal{P}(\Omega)| = 2^{|\Omega|} = 2^{10} = 1024$ endlich (diskret) wäre.

Vorlesung 3
Mo 19 Apr 2021 10:23

Wir stellen fest, dass es im letzten Beispiel auch genügt hätte, $\mathbb{P}(\{k\})$ für $k = 1, \ldots, 10$ anzugeben. Daraus ergibt sich:

Satz 1.7. Sei Ω abzählbar.

a) Sei $p(\omega) \in [0,1], \omega \in \Omega$, sodass

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

Dann ist $\mathbb{P}$ definiert durch:

$$\mathbb{P} : \left| \begin{array}{ccc} \mathcal{P}(\Omega) & \longrightarrow & [0,1] \\ A & \longmapsto & \sum_{\omega \in A} p(\omega) \end{array} \right.$$

eine Wahrscheinlichkeitsverteilung auf $(\Omega, \mathcal{P}(\varnothing))$.

b) Jede Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $(\Omega, \mathcal{P}(\Omega))$ hat obige Form, wobei $p(\omega) = \mathbb{P}(\{\omega\})$.

Bemerkung. $p : \varnothing \to [0,1]$ heißt Massenfunktion der Wahrscheinlichkeitsverteilung $\mathbb{P}$.

Warnung. Der Satz gilt nicht für Ω überabzählbar.

Bemerkung. Sei A abzählbar und $p(\omega) \geqslant 0$ für $\omega \in A$. Dann definiert

$$\sum_{\omega \in A} p(\omega) := \sum_{k \geqslant 1} p(\omega_k).$$

mit einer beliebigen Abzählung $\omega_1, \omega_2, \ldots, \omega_k$ von A eine wohldefinierte Summe der $p(\omega)$. Es ist wichtig, dass hier $p(\omega) \geqslant 0$, sonst wäre obiges nicht wohldefiniert.

Lemma 1.8. Sei A abzählbar.

a Sei $p(\omega) \in [0,1]$ für alle ω. Dann ist

$$\sum_{\omega \in A} p(\omega) \in [0, \infty].$$

wohldefiniert. Setzen wir

$$\mathbb{P}(A) := \sum_{\omega \in A} p(w).$$

so gilt

$$P(A) = \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F).$$

und $P(A) \leqslant P(B)$ für $A \subseteq B$.

b Ist $A = \bigsqcup_{k=1}^{\infty} A_k$ eine disjunkte Vereinigung, so ist

$$P(A) = \sum_{k=1}^{\infty} P(A_k).$$

Beweis. a Sei $\omega_1, \omega_2, \ldots$ eine beliebige Abzählung von A. Dann ist die Funktion

$$n \longmapsto \sum_{k=1}^{n} p(\omega_k).$$

monoton wachsend. Also ist

$$\sum_{k=1}^{\infty} p(\omega_k) := \lim_{n \to \infty} \sum_{k=1}^{n} p(\omega_k) = \sup_{n \in \mathbb{N}} \sum_{k=1}^{n} p(\omega_k) \in [0, \infty].$$

wohldefiniert.

Wir wollen nun noch zeigen, dass

$$P(A) := \sum_{k=1}^{\infty} p(\omega_k) \overset{!}{=} \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F).$$

Die Ungleichung '$\leqslant$' folgt sofort, da mit $F_n := \{\omega_1, \ldots, \omega_k\}$ folgt

$$\sum_{k=1}^{n} p(\omega_k) = \sum_{\omega \in F_n} p(\omega) = P(F_n) \leqslant \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F).$$

Also ergibt sich im Limes genau wie gewünscht

$$P(A) = \lim_{n \to \infty} \sum_{k=1}^{n} p(\omega_k) \leqslant \lim_{n \to \infty} \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F) = \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F).$$

Für '$\geqslant$' stellen wir fest, dass es für jedes $F \subseteq A$ endlich ein $n \in \mathbb{N}$ gibt, sodass $F \subseteq \{\omega_1, \ldots, \omega_n\}$ und somit gilt

$$P(F) = \sum_{\omega \in F} P(\omega) \leqslant \sum_{k=1}^{n} p(\omega_k) \leqslant \sum_{k=1}^{\infty} p(\omega_k) = P(A).$$

Damit ist das Supremum der $P(F)$ für $F \subseteq A, |F| < \infty$ durch $P(A)$ beschränkt.
Für die letzte Behauptung sehen wir mit $A \subseteq B$ leicht, dass

$$P(A) = \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F) \leqslant \sup_{\substack{F \subseteq B \\ |F| < \infty}} = P(B).$$

b (σ-Additivität) Wir unterscheiden zwischen zwei Fälle:

1) Falls $|A| < \infty$, so ist $A = \bigsqcup_{k=1}^{n} A_k$ für ein n, und somit gilt

$$\begin{aligned} P(A) &= \sum_{l=1}^{|A|} p(\omega_l) = \sum_{l=1}^{|A|} \sum_{k=1}^{n} p(\omega_l) \mathbb{1}_{A_k}(\omega_l) \\ &= \sum_{k=1}^{n} \sum_{l=1}^{|A|} p(\omega_l) \mathbb{1}_{A_k}(\omega_l) = \sum_{k=1}^{n} P(A_k) \end{aligned} \tag{5}$$

2) Sei nun $|A| = \infty$. Wir zeigen zunächst '$\leqslant$'. Für ein endliches $F \subseteq A$ ist

$$F = \bigcup_{k=1}^{\infty} (F \cap A_k).$$

eine disjunkte Vereinigung mit endlich vielen Termen, also ist

$$P(F) = \sum_{k=1}^{\infty} P(F \cap A_k) \leqslant \sum_{k=1}^{\infty} P(A_k).$$

Somit liefert das Supremum über beide Seiten, dass

$$P(A) = \sup_{\substack{F \subseteq A \\ |F| < \infty}} P(F) \leqslant \sum_{k=1}^{\infty} P(A_k).$$

gilt. Wir zeigen nun '$\geqslant$'.
Idee. Wir können $P(A_k) = \sup_{\substack{F_k \subseteq A_k \\ |F_k|]\infty}} P(F_k)$ schreiben und 'optimieren' nun jedes einzelne F_k.
Seien also $F_k \subseteq A_k$ jeweils endlich. Dann ist $F_k \cap F_l \subseteq A_k \cap A_l = \varnothing$, also sind auch die F_k paarweise disjunkt, und wir damit

$$\sum_{k=1}^{n} P(A_k) = \sum_{k=1}^{n} \sup_{\substack{F_k \subseteq A_k \\ |F_k| < \infty}} P(F_k) = \sup_{\substack{F_1 \subseteq A_1 \\ abs F_1 < \infty}} \dots \sup_{\substack{F_k \subseteq A_k \\ |F_k| < \infty}} \sum_{k=1}^{n} P(F_k) \tag{6}$$

Damit ist

$$\sum_{k=1}^{n} P(F_k) = P\left(\bigcup_{k=1}^{n} F_k\right) \leqslant P\left(\bigcup_{k=1}^{\infty} F_k\right) \leqslant P\left(\bigcup_{k=1}^{\infty} A_k\right) \stackrel{\text{def}}{=} P(A).$$

Setzen wir dies nun in die rechte Seite von (1) ein, so ergibt sich

$$\sum_{k=1}^{n} P(A_k) \leqslant P(A) \qquad \Rightarrow \qquad \sum_{k=1}^{\infty} P(A_k) \leqslant P(A).$$

□

Beweis von Satz 1.7. (a) Es gilt

$$\sum_{\omega \in \Omega} p(\omega) = P(\Omega) = 1.$$

nach Voraussetzung. Die σ-Additivität folgt nun aus Lemma 1.8. Deswegen ist $P(A)$ eine Wahrscheinlichkeitsverteilung.

(b) Da P σ-additiv ist, ist $\forall A \subseteq \Omega$:

$$PA) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\}).$$

und dies hat genau die angegebene Form mit $p(\omega) := P(\{\omega\})$

$\square$

1.4 Die Gleichverteilung

Sei Ω endlich ($\neq \emptyset$) und betrachte σ-Algebra $\mathcal{F} = \mathcal{P}(\Omega)$.

Die **Gleichverteilung** ist die Wahrscheinlichkeitverteilung, die ein uniformes "Gewicht"(Massenfunktion) auf die Elementarereignisse verteilt:

$$\forall \omega \in \Omega : p(w) = \mathbb{P}\left(\{\omega\}\right) = \frac{1}{|\Omega|}.$$

Aus Satz 1.7 folgt dann bereits, dass $\forall A \subseteq \Omega$:

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega) = \frac{|A|}{|\Omega|}.$$

Beispiel. (a) Betrachte n Würfe eines fairen Würfels. In diesem Fall ist $\Omega = \{1, 2, 3, 4, 5, 6\}^n = \{\omega = (\omega_1, \dots, \omega_n) \mid \omega_k \in \{1, \dots, 6\}\}$ und somit $|\Omega| = 6^n$ und die Gleichverteilung ist gegeben durch

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

(b) (Zufällige Permutationen).

- Eine Permutation $\sigma \in \mathfrak{S}_n$ von $\{1, \dots, n\}$ ist eine Abbildung von $\{1, \dots, n\}$ nach $\{1, \dots, n\}$, die bijektiv ist. Oft schreiben wir
$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}.$$
und meinen damit $\sigma(1) = 4$, $\sigma(2) = 3$, $\sigma(3) = 1$, $\sigma(4) = 2$. Manschmal wird auch
$$\sigma = (4, 3, 1, 2) = (\sigma_1, \sigma_2, \sigma_3, \sigma_4).$$
geschrieben.
- Sei $\Omega = \mathfrak{S}_n$ die Menge aller Permutationen von $\{1, \dots, n\}$. Dann ergibt sich
$$|\mathfrak{S}_n| = n!.$$

Also ergibt sich für die Gleichverteilung eine Wahrscheinlichkeit von

$$\mathbb{P}(\sigma) = \frac{1}{n!} \quad \forall \sigma \in \mathfrak{S}_n.$$

Beispiel. Sei N die Anzahl von Karten eines Kartenspiels, die gut gemischt sind, d.h. jede Reihenfolge ist gleich wahrscheinlich.

(1) Was ist die Wahrscheinlichkeit, dass die k-te Karte an der l. Stelle ist? D.h, was ist:

$$\mathbb{P}(\{\omega \in \mathfrak{S}_n \mid \omega(k) = l\}).$$

Lösung. Es ergibt sich

$$\mathbb{P}(\{\omega \in \mathfrak{S}_n \mid \omega(k) = l\}) = \frac{|\{\omega \in \mathfrak{S}_n \mid \omega(k) = l\}|}{|\Omega|} = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

□

(2) Was ist die Wahrscheinlichkeit, dass eine Karte 'auf ihrer Stelle' ist? D.h., was ist?

$$\mathbb{P}(\{\omega \mid \exists k\colon \omega(k) = k\}).$$

Lösung. Definiere die Ereignisse $A_k := \{\omega(k) = k\}$. Diese sind nicht disjunkt für verschiedene k. Es ergibt sich:

$$\begin{aligned}
\mathbb{P}(\exists k\colon \omega(k) = k) &= \mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \underbrace{\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k})}_{=\frac{(n-k)!}{n!}} \\
&= \sum_{k=1}^{n} (-1)^{k-1} \frac{(n-k)!}{n!} \underbrace{\sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} 1}_{=\binom{n}{k}} \\
&= \sum_{k=1}^{n} (-1)^{k-1} \frac{(n-k)!}{n!} \cdot \frac{n!}{(n-k)!k!} \\
&= -\sum_{k=1}^{n} \frac{(-1)^k}{k!} \\
&= 1 - \frac{1}{e} + \sum_{k=n+1}^{\infty} \frac{(-1)^k}{k!}
\end{aligned} \tag{7}$$

Für $n \to \infty$ geht das gegen $1 - \frac{1}{e} \in (0,1)$. □

Bemerkung*. Wir verwenden hier die Exponentialreihe, d.h.

$$e^x := \exp(x) := \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

Diese konvergiert absolut und auf ganz $\mathbb{R}$, insbesondere für $x = -1$ gegen $\frac{1}{e}$

1.5 Die empirische Verteilung

Diese wird aus den Beobachtungen definiert.

Definition (Empirische Verteilung)**.** Seien $x_1, x_2, \ldots, x_n \in \Omega$ n Beobachtungen. Setze

$$N(A) := |\{k \in \{1, \ldots, n\} \mid x_k \in A\}|.$$

Dann definiert

$$\mathbb{P}(A) = \frac{N(A)}{n}.$$

die **empirische Häufigkeit** von A. $\mathbb{P}$ nennen wir die **empirische Verteilung**. Zudem ist

$$p(\omega) = \frac{N(\{\omega\})}{n}.$$

die **relative Häufigkeit** von $\omega \in \Omega$.

Beispiel. Die empirische Verteilung von n Zufallswürfeln eines Würfels wird gegeben durch $x_1, \ldots, x_n \in \{1, \ldots, 6\}$. Die Plots für $p_k := \frac{N(k)}{n}$ für verschiede n sehen wie folgt aus:

Plots einfügen

Vorlesung 4
Mi 21 Apr 2021 10:15

1.6 Zufallsvariablen

Wir werden Funktionen der Ergebnisse betrachten:

Definition 1.9 (Diskeret Zufallsvariable)**.** Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine **diskrete Zufallsvariable** ist eine <u>messbare</u> Abbildung

$$X : \Omega \longrightarrow \mathcal{S}.$$

mit $\mathcal{S}$ abzählbar (denke: 'diskret').
Messbar bedeutet hierbei, dass

$$\forall s \in \mathcal{S} \colon X^{-1}(s) = \{\omega \in \Omega \mid X(\omega) = s\} \in \mathcal{F}.$$

Notation. Wir schreiben auch kurz:

$$X^{-1}(s) = \{X(\omega) = s\} = \{X = s\}.$$

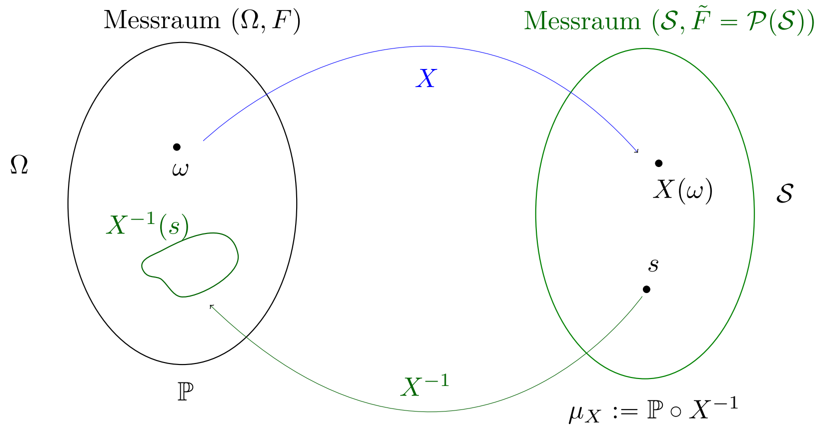


Abbildung 2: Diskrete Zufallsvariable

Definition 1.10. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathcal{S}$ eine diskrete Zufallsvariable.
Die **Verteilung von X** ist die Wahrscheinlichkeitverteilung μ_X auf $\mathcal{S}, \mathcal{P}(\mathcal{S}) = \tilde{F}$, s.d. $\forall B \in \tilde{F} : \mu_X(B) := \mathbb{P}(X^{-1}(B))$.
μ_X hat eine **Massenfunktion**

$$p_X(s) := \mathbb{P}.$$

Beispiel (Werfen von n Münzen)**.** Betrachte folgende Situation:

- Sei $\Omega = \{\omega = (\omega_1, \ldots, \omega_n \mid \omega_i \in \{0,1\} \text{ für } 1 \leqslant i \leqslant n\}$, wobei

$$\omega_k = \begin{cases} 0 & \text{falls } k\text{-ter Wurf ist Zahl} \\ 1 & \text{falls } k\text{-ter Wurf ist kopf} \end{cases}.$$

- Und sei $\mathcal{F} = \mathcal{P}(\Omega)$ und $\mathbb{P}$ die Gleichverteilung.

(1) Setze

$$X_k : \left| \begin{array}{ccc} \Omega & \longrightarrow & \mathcal{S} = \{0,1\} \\ \omega & \longmapsto & \omega_k \end{array} \right.$$

für $k = 1, \ldots, n$. Dies ist eine diskrete Zufallsvariable mit Verteilung μ_{X_k} mit

$$p_{X_k}(s) = \mathbb{P}(X_k = s) = \frac{2^{n-1}}{2^n} = \frac{1}{2}.$$

Wir sehen also, dass X_k gleichverteilt ist.

(2) Definiere

$$Y : \left| \begin{array}{ccc} \Omega & \longrightarrow & \mathcal{S} := \{0, 1, \ldots, n\} \\ \omega & \longmapsto & \omega_1 + \ldots + \omega_n \end{array} \right.$$

d.h.

$$Y(\omega) = \# \{\text{geworfene Köpfe}\}.$$

Es hat nun μ_Y die Massenfunktion:

$$p_Y(k) = \frac{1}{2^n} |\{\omega \mid w_1 + \ldots + \omega_n = k\}| = \frac{\binom{n}{k}}{2^n}.$$

Diese Verteilung sieht wie folgt aus:

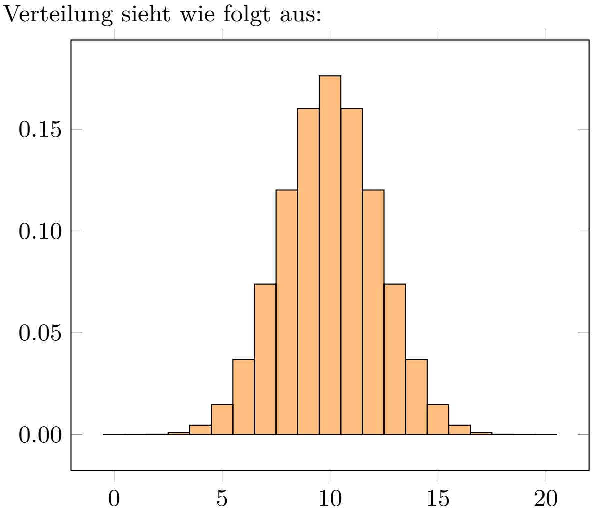


Abbildung 3: Massenfunktion $p_Y(k)$

Diese sind Sonderfälle der Bernoulliverteilung und der Binomialverteilung.

1.6.1 Die Bernoulli-Verteilung

Definition 1.11 (Bernoulli-Verteilung)**.** Sei $p \in [0, 1]$ gegeben. Die Wahrscheinlichkeitsverteilung auf $\{0, 1\}$ mit Massenfunktion

$$p(k) = \begin{cases} p & k = 1 \\ 1 - p & k = 0 \end{cases}.$$

heißt **Bernoulli-Verteilung mit Parameter** p.

Notation. Wir notieren $\mathrm{Ber}(p)$ für die Bernoulli-Verteilung mit Parameter p.

Beispiel. a Sei eine Münze gegeben, die mit Wahrscheinlichkeit p Kopf zeigt. Hier ist

$$\Omega = \{\text{Zahl}, \text{Kopf}\} \qquad \mathbb{P}(\text{Kopf}) = p = 1 - \mathbb{P}(\text{Zahl}).$$

Sei

$$X(\omega) = \begin{cases} 1 & \omega = \text{Kopf} \\ 0 & \omega = \text{Zahl} \end{cases}.$$

Dann ist $\mathbb{P}(X = 1) = p$ und $X \sim \mathrm{Ber}(p)$.

Notation. Wir schreiben $X \sim \mathrm{Ber}(p)$, wenn X die Verteilung $\mathrm{Ber}(p)$ hat.

(b) In einer Urne befinden sich n blaue Kugeln und m rote Kugeln. Wir ziehen eine Kugel aus der Urne (Annahme: Gleichverteilung). Dann ist

$$\Omega = \{\omega = (\omega_1, \dots, \omega_{n+m}) \mid \omega_i \in \{\text{blau}, \text{rot}\} \text{ mit } n \text{ mal blau}\}.$$

Setze $\mathcal{F} = \mathcal{P}(\Omega)$ und wähle $\mathbb{P}$ als die Gleichverteilung. Betrachte

$$X(\omega) = \begin{cases} 1 & \text{falls } \omega_i \text{ ist blau} \\ 0 & \text{sonst} \end{cases}.$$

Diese hat also die Verteilung

$$\mathbb{P}(X = 1) = \frac{\binom{m+n-1}{n-1}}{\binom{m+n}{n}} = \frac{(m+n-1)!}{(n-1)!m!} \frac{n!m!}{(m+n)!} = \frac{n}{m+n}.$$

Also ist $X \sim \mathrm{Ber}\left(\frac{n}{n+m}\right)$.

1.6.2 Die Binomial-Verteilung

Definition 1.12 (Binomial-Verteilung)**.** Seien $n \in \mathbb{N}$ und $p \in [0,1]$ gegeben. Die Wahrscheinlichkeitsverteilung auf $\{0, 1, \dots, n\}$ mit Massenfunktion

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

für $k = 0, \dots, n$ heißt **Binomialverteilung mit Parametern n und p**.

Notation. Wir notieren $\mathrm{Bin}(n,p)$ für die Binomialverteilung mit Parametern n und p.

Beispiel (Ziehen mit Zurücklegen)**.**

- Seien m Kugeln in einer Urne, davon $p \cdot m \in \mathbb{N}$ weiße Kugeln und $(1-p)m$ schwarze Kugeln.
- Wir ziehen eine Kugel, notieren uns die Farbe und legen sie wieder zurück.
- Wir mischen die übrigen Kugeln wider gut.
- Wir wiederholen die vorherigen Schritte, bis wir n Ziehungen durchgeführt haben.
- Dies modellieren wir durch

$$\Omega = \{0,1\}^n.$$

wobei $\omega = (\omega_1, \dots, \omega_n) \in \Omega$ gegeben ist durch

$$\omega_i = \begin{cases} 1 & \text{falls Farbe der } i\text{-ten Kugeln weiß ist} \\ 0 & \text{sonst} \end{cases}.$$

- Sei nun $X(\omega) = \sum_{k=1}^n \omega_k = \#\{\text{weiße Kugeln}\}$.
Dann behaputen wir, dass $X \sim \mathrm{Bin}(n,p)$. In der Tat gilt:

$$\begin{aligned} \frac{|\omega \in \Omega \mid X(\omega) = l|}{|\Omega|} &= \frac{\binom{n}{l} \cdot (pm)^l \, ((1-p)m)^{n-l}}{m^n} \\ &= \frac{\binom{n}{l} p^l (1-p)^{n-l} \cdot m^n}{m^n} = \binom{n}{l} p^l (1-p)^{n-l} \end{aligned} \tag{8}$$

Bemerkung. Wir haben hier den Begriff der **Unabhängigkeit** genutzt, den wir nun genauer kennenlernen wollen.

Definition 1.13 (Unabhängige Ereignisse)**.** Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Die Ereignisse $E_1, E_2, \dots, E_n$ heißen **unabhängig**, falls

$$\mathbb{P}(E_{i_1} \cap E_{i_2} \cap \ldots \cap E_{i_k}) = \prod_{l=1}^{k} \mathbb{P}(E_{i_l}).$$

für alle $2 \leqslant k \leqslant n$ und $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$.

Beispiel.

- Betrachte zwei Würfelwürfe, d.h. $\Omega = \{1, \dots, 6\}^2$ und notiere $\omega = (\omega_1, \omega_2)$. Dann können wir

$$E_1 = \{\omega_1 = 3\} \qquad E_2 = \{\omega_2 \geqslant 4\}.$$

betrachten. Wir rechnen nach, dass

$$\mathbb{P}(\omega_1 = 3 \cap \omega_2 \geqslant 4) = \frac{3}{36} = \frac{1}{6} \cdot \frac{3}{6} = \mathbb{P}(\omega_1 = 3) \cdot \mathbb{P}(\omega_2 \geqslant 4).$$

also sind die beiden Ereignisse unabhängig voneinander. Das macht auch semantisch Sinn, weil wir durch das Ergebnis des ersten Würfelwurfs keine Informationen über das Ergebnis des zweiten Würfelwurfs erhalten.
- Falls $E_1, E_2, \dots, E_n$ unabhängige Ereignisse sind mit $\mathbb{P}(E_i) = p$ für $1 \leqslant i \leqslant n$, dann gilt

$$\mathbb{P}(\text{genau } k \text{ der Ereignisse treten ein}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Dies rechnen wir nach. Setze hierzu

$$A_{(i_1, \dots, i_k)} = \{\omega \in \Omega \mid E_{i_1}, \dots, E_{i_k} \text{ treten ein, die anderen nicht}\}$$

Dann ist

$$\tilde{A} = \bigsqcup_{1 \leqslant i_1 < \ldots < i_k \leqslant n} A_{i_1,\ldots,i_k}.$$

eine disjunkte Vereinigung, also erhalten wir

$$\begin{aligned}
\mathbb{P}(\tilde{A}) &= \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathbb{P}(A_{(i_1,\ldots,i_k})\\
&\overset{\text{unabhängig}}{=} \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \prod_{j \in \{i_1,\ldots,i_k\}} \mathbb{P}(E_j) \cdot \prod_{l \notin \{i_1,\ldots,i_k\}} \mathbb{P}(E_l^c)\\
&= \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} p^k (1-p)^{n-k}\\
&= \binom{n}{k} p^k (1-p)^{n-k}
\end{aligned} \tag{9}$$

Bemerkung. Streng genommen haben wir in der letzten Rechnung verwendet, dass mit $E_1, \ldots, E_n$ unabhängig auch $F_1, \ldots, F_n$ für $F_i = E_i$ oder $F_i = E_i^c$ unabhängig voneinander sind. Dies müssten wir noch einmal nachrechnen. Für den Fall $n = 2$ ist z.B:

$$\mathbb{P}(E_1 \cap E_2^c) + \mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1 \cap (E_2 \cup E_2^c) = \mathbb{P}(E_1).$$

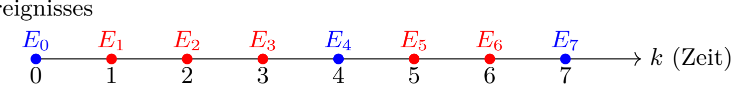
Also ergibt sich

$$\mathbb{P}(E_1 \cap E_2^c) = \mathbb{P}(E_1) - \mathbb{P}(E_1)\mathbb{P}(E_2) = \mathbb{P}(E_1)(1 - \mathbb{P}(E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2^c).$$

wie zu zeigen war.

1.6.3 Die Poisson-Verteilung

Betrachte Ereignisse $E_1, \ldots, E_n$, die unabhängig voneinander sind und jeweils Wahrscheinlichkeit p haben. Wir skizzieren dies mit • für ein eingetretenes Ereignis und • für das Nicht-Eintreten des entsprechenden Ereignisses

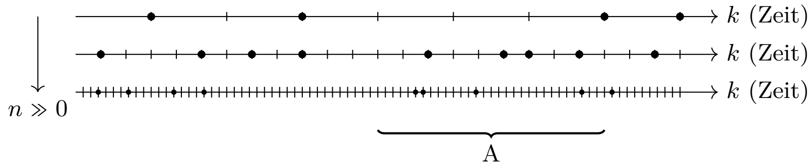


Frage. Was passiert, wenn $n \gg 1$, d.h. wie viele Erfolge werden wir (ca.) unter diesen n Ereignissen haben?

Typischerweise haben wir dann $\mathcal{O}(pn)$ Erfolge in $E_1, \ldots, E_n$. Die Erfolgswahrscheinlichkeit p hänge nun von n ab, d.h. $p = p(n)$. Wir wollen den Erwartungswert $p \cdot n$ festhalten und n groß werden lassen, d.h. sei $\lambda \in (0, \infty)$ so, dass

$$\lim_{n \to \infty} p(n) \cdot n = \lambda.$$

Wir können uns das Ganze so vorstellen, dass wir in einem Zeitintervall von 1 erwarten, dass λ Ereignisse eintreten und wir nun mit einem kleinen Zeitintervall $\delta = \frac{1}{n}$ für große n das kontinuierliche Zeitintervall durch n unabhängige Ereignisse ersetzen, die jeweils mit Wahrscheinlichkeit $\frac{\lambda}{n}$ eintreten und uns nun fragen, wie wahrscheinlich es also ist, dass wir eine gewisse Anzahl an Ereignissen im Zeitintervall beobachten.



Es stellt sich also

Frage. Was ist für ein Zeitintervall A die Wahrscheinlichkeit, dass wir im Intervall genau k Ereignisse beobachten, d.h. was ist

$$\lim_{n\to\infty} \mathbb{P}(\exists k \text{ Erfolge in } A).$$

?

- Sei $p = p(n)$ sodass $\lim\limits_{n\to\infty} pn = \lambda \in (0,\infty)$.
- Wähle Zeiteinheit $\delta = \frac{1}{n}$.

Die Antwort gibt der folgende Satz.

Satz 1.14. Sei $\lambda \in (0,\infty)$. Dann ist

$$\lim_{n\to\infty} \operatorname{Bin}\left(n, \frac{\lambda}{n}\right)(k) = \frac{e^{-\lambda}\lambda^k}{k!}.$$

für $k = 0, 1, 2, \ldots$

Beweis. Sei k fest. Dann ist

$$\begin{aligned}
\operatorname{Bin}\left(n, \frac{\lambda}{n}\right)(k) &= \frac{n!}{(n-k)^k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\
&= \underbrace{\frac{n(n-1)\cdot\ldots(n-k+1)}{n^k}}_{\to 1}\frac{\lambda^k}{k!}\underbrace{\left(1-\frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}}\underbrace{\left(1-\frac{\lambda}{n}\right)^{-k}}_{\to(1-0)^{-k}=1} \\
&= \frac{\lambda^k}{k!}\cdot e^{-\lambda}
\end{aligned} \tag{10}$$

□

Definition 1.15 (Poisson-Verteilung)**.** Sei $\lambda \in (0,\infty)$ fest. Die Wahrscheinlichkeitsverteilung auf $\{0,1,2,\ldots\}$ mit Massenfunktion

$$p(k) = \frac{e^{-\lambda\lambda^k}}{k!}.$$

heißt **Poisson-Verteilung mit Parameteter λ**.

Notation. Wir schreiben auch $\operatorname{Poi}(\lambda)$ für die Poisson-Verteilung zum Parameter λ.

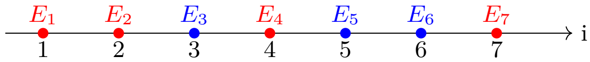
Vorlesung 5
Mo 26 Apr 2021 10:17

1.6.4 Die geometrische Verteilung

- Seien $E_1, E_2, \ldots$ unabhängige Ereignisse mit

$$\mathbb{P}(E_i \text{ tritt ein}) = 1 - q$$
$$\mathbb{P}(E_/ \text{ tritt nicht ein}) = q^{\cdot}$$

Wir verwenden wieder $\color{blue}\bullet$ um das Eintreteten eines Ereignisses zu notieren, und $\color{red}\bullet$ für das Gegenteil.



Setze nun $T_0 = 0$ sowie $T_{l+1} = \min\{i > T_l \colon E_i \text{ tritt ein}\}$ als die Eintrittszeitpunkte des k-ten Erfolgs. Dann definieren

$$N_l := T_l - T_{l-1} - 1 = \#\{{\color{red}\bullet} \text{ zwischen } T_{l-1} \text{ und } T_l\}.$$

die Abstände zwischen den jeweiligen Erfolgen. Es ergibt sich:

$$\begin{aligned}
\mathbb{P}(N_l = k) &= \mathbb{P}(N_1 = k) = \mathbb{P}(T_1 = k + 1) \\
&= \mathbb{P}\left(\bigcap_{i=1}^{k} \{E_i \text{ tritt nicht ein}\} \cap \{E_{k+1} \text{ tritt ein}\}\right) \\
&= \left(\prod_{i=1}^{k} \mathbb{P}(E_i \text{ tritt nicht ein})\right) \cdot \mathbb{P}(E_{k+1} \text{ tritt ein}) \\
&= (1-q)q^k
\end{aligned} \tag{11}$$

für $k \geqslant 0$. Das motiviert nun

Definition 1.16 (Geometrische Verteilung)**.** Sei $q \in [0, 1)$. Die Wahrscheinlichkeitsverteilung auf $\{0, 1, 2, \ldots\}$ mit Massenfunktion

$$p(k) = (1-q)q^k.$$

heißt **geometrische Verteilung** mit Parameter q.

Notation. Wir notieren hierfür $\mathrm{Geo}(q)$

Warnung. Eine andere Version setzt $p(k) = (1-q)q^{k-1}$, wobei $k = 1, 2, \ldots$. Hier ist nur der Index verschoben. Wenn nicht anders genannt, ist für uns aber immer obige Definition gemeint.

1.7 Simulation von Gleichverteilung

Frage. Wie werden Zufallsvariablen simuliert?

Typischerweise benutzen wir folgende Situation:
Input Zahl(en), z.B. Redinerzeit
Output 'Zufällige Zahl' in $\{0, \ldots, m-1\}$

1.7.1 Lineare Kongruenzgeneratoren (LCG)

Startwert $x_0 \in \mathbb{N}$ gegeben.
Parameter $a, c, m \in \mathbb{N}$

Schritt Setze $x_{n+1} := (a \cdot x_n + c) \mod m$.
Dieses Vorgehen produziert eine scheinbar zufällige Folge.

Beispiel. Es gibt folgende LCGs:

	m	a	c
ZX81	$2^{16}+1$	75	0
RANDN	2^{31}	65539	0
Marsaglia	2^{32}	69069	1

Beispiel (Eine schlechte Wahl)**.** Wenn wir $a = 4, c = 1, m = 31$ wählen sowie $x_0 = 3$, so erreichen wir Periode 9, und somit werden nicht alle Zahlen erreichen / generieren.

Plot einfügen

Lemma 1.17 (Knuth)**.** Die Periode eines LCG ist gleich m, genau dann, wenn

(a) c und m haben keine gemeinsamen Primfaktoren
(b) Jeder Primfaktor von m ist ein Teiler von $a - 1$
(c) Falls $4 \mid m$, dann $4 \mid a - 1$.

Beweis. Kein Beweis. □

Beispiele von LCG's einfügen

1.7.2 Zufallsvariablen aus $[0, 1)$

- Sei $(x_n)_{n\geqslant 1}$ eine Folge von (Pseudo)zufallszahlen aus $\{0, 1, \ldots, m-1\}$. Dann ist
$$u_n := \left(\frac{x_n}{m}\right)_{n\geqslant 1}.$$
eine Folge von Pseudozahlen in $[0, 1)$. Gut ist aber nur der Fall, wenn $m \approx 10^N$, wobei N = Rechnergenauigkeit, d.h. #Ziffern.

1.7.3 Zufallspermutationen

Frage. Wie erzugt man eine gleichverteilte Permutation von $\{1, \ldots, N\}$?

Algorithmus 1.18 : Zufallspermutationen

Eingabe : Möglichkeit, aus endlicher Menge gleichverteilt zufällige Zahlen zu ziehen
Ausgabe : Eine zufällige Permutation von $\{1, \ldots, N\}$

Setze $\sigma_0 := \{1, \ldots, N\}$
for $i = 1$ **to** $n - 1$ **do**
  wähle $k \in \{i, \ldots, N\}$ gleichverteilt
  Setze $\sigma_k := \sigma_{k-1} \circ \tau_{i,k}$

Lemma 1.19. Der Algorithmus erzeugt eine zufällige gleichverteilte Permutation.

Beweis. Der Algorithmus benutzt eine Gleichverteilung auf

$$\Omega_n := \{1, \ldots, N\} \times \{2, \ldots, n\} \times \{n-1, n\}.$$

Für $\omega = (w_1, \ldots, w_{N-1}) \in \Omega_N$ ist

$$\sigma(\omega) = \tau_{N-1,\omega_{N-1}} \circ \ldots \circ \tau_{1,w} \circ \underbrace{(1, \ldots, N)}_{\sigma_0}.$$

Es genügt also zu zeigen, dass $\sigma : \Omega_N \to \mathcal{S}_N$ eine Bijektion ist. Wir sehen:

a) $|\Omega_N| = |\mathcal{S}_N| = N!$

b) Sei $w \neq \tilde{\omega}$ und setze $k = \min\{j \mid \omega_j \neq \tilde{\omega}_j\}$. Dann ist $\sigma(\omega)_k \neq \sigma(\tilde{\omega})_k$ und somit ist die Funktion injektiv

Damit ist die Abbildung sogar bijektiv und wir sind fertig. $\square$

1.7.4 Geometrische Verteilung

Sei $X \sim \text{Geo}(q)$, d.h.

$$\mathbb{P}(X = k) = (1-q)q^k.$$

Frage. Wie simuliert man nun X?

Erzeuge zunächst $n \sim U[0,1)$ als gleichverteilte Zufallsvarable auf $[0,1)$.

Sei $T_k := \mathbb{P}(X < k)$. Falls $n \in [T_k, T_{k+1})$, dann setze $X = k$. Wir berechnen also

$$\begin{aligned} T_k &= \mathbb{P}(X < k) = 1 - \mathbb{P}(X \geqslant k) \\ &= 1 - \sum_{x \geqslant k} (1-q)q^k \\ &= 1 - q^k \end{aligned} \tag{12}$$

Wegen

$$u = 1 - q^x \Leftrightarrow x = \frac{\ln(1-u)}{\ln(q)}.$$

erhalten wir also für k den Wert

$$k := \left\lfloor \frac{\ln(1-u)}{\ln(q)} \right\rfloor = \left\lfloor \frac{\ln(\tilde{u})}{\ln(q)} \right\rfloor.$$

wobei wir ebenfalls $\tilde{u} \sim U[0,1)$ gleichverteilen.

1.8 Erwartungswert und Varianz

- Sei X eine reellwertige diskrete Zufallsverteilung. Sei

  $$X : \Omega \to \mathcal{S} \subseteq \mathbb{R}.$$

  eine diskrete Zufallsvariable, d.h. $\mathcal{S}$ abzählbar.

Definition 1.20 (Empirischer Mittelwert)**.** Seien $x_1, \ldots, x_n \in \mathcal{S}$ n Beobachtungen einer Zufallsvariable X. Der **empirische Mittel-**

wert ist durch

$$\frac{1}{n}\sum_{i=1}^{n} x_i.$$

definiert.

Ziel. Wir wollen eine Sorte von Mittelwert definieren, der nur von X abhängig ist, und nicht von den Beobachtungen.

Folgende Forderungen ergeben sich an solch einen Mittelwert:

- Falls $X(\omega) = x$ für jedes ω, dann muss der Mittelwert von X gleich x sein.
- Jeder Wert $x \in \mathcal{S}$ muss bezüglich der Massenfunktion $p_X(x)$ gewichtet sein.

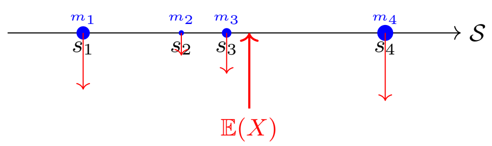
Das motiviert folgende

Definition 1.21 (Erwartungswert)**.** Der **Erwartungswert** von X bzgl. $\mathbb{P}$ ist durch

$$\mathbb{E}(X) = \sum_{s\in\mathcal{S}} s \cdot \mathbb{P}(X = s) = \sum_{s\in\mathcal{S}} s \cdot p_X(s).$$

definiert. Dies ist wohldefiniert, falls die Reiche absolut gegen einen Wert $< \infty$ konvergiert.

Stellen wir uns $\mathcal{S}$ als Balken vor und tragen an den stellen $s_i \in \mathcal{S}$ jeweils das Gewicht $m_i := p_X(s_i)$ an, so ist $\mathbb{E}(X)$ derjenige Punkt, mit dem der Balken ausbalanciert ist:



Bemerkung. Nicht alle Wahrscheinlichkeitsverteilungen besitzen einen endlichen Mittelwert, das zeigt folgendes

Beispiel. Sei X auf $\{1, 2, \ldots\}$ verteilt mit

$$\mathbb{P}_X(s) = \frac{6}{\pi^2 r}.$$

dann ergibt sich für den Erwartungswert:

$$\mathbb{E}(X) = \sum_{s\geqslant 1} s \cdot \frac{6}{\pi^2 s^2} = \frac{6}{\pi^2} \cdot \sum_{s\geqslant 1} \frac{1}{s} \to \infty.$$

Vorlesung 6
Sa 01 Mai 2021 09:18

Beispiel (Zufallsvariablen mit Werten in $\{0, 1\}$)**.** Sei $A \in \mathcal{F}$ ein Ereignis, und definiere X durch

$$X(\omega) := \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{falls } \omega \in A \\ 0 & \text{sonst} \end{cases}.$$

Dann ist

$$\mathbb{E}(X) = \mathbb{P}(A).$$

Beweis. Nach Definition ist

$$\begin{aligned}\mathbb{E}(X) &= 0 \cdot \mathbb{P}(X = 0) + 1 \cdot \mathbb{P}(X = 1)\\ &= \mathbb{P}(A)\end{aligned}$$

□

Beispiel (Binomialverteilung)**.** Sei $X \sim \operatorname{Bin}(n, p)$. Wir wollen zeigen, dass $\mathbb{E}(X) = p \cdot n$.

Beweis.

$$\mathbb{E}(X) = \sum_{k=0}^{n} \underbrace{\binom{n}{k} p^k (1-p)^{n-k}}_{=\mathbb{P}(X=k)} \cdot k.$$

Wir wollen nun allgemein, in Anlehnung an die Binomialformel, den Wert von

$$\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} \cdot k.$$

berechnen. Dazu stellen wir fest, dass

$$p \cdot \frac{d}{dp} \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = p \cdot \sum_{k=0}^{n} \binom{n}{k} p^{k-1} q^{n-k} \cdot k.$$

unser Ausdruck ist, also suchen wir

$$p \cdot \frac{d}{dp} (p+q)^n = p \cdot (p+q)^{n-1} \cdot n.$$

Nun können wir $q = 1-p$ auf beiden Seiten setzen, und somit erhalten wir

$$\mathbb{E}(X) = p \cdot (p + (1-p))^{n-1} \cdot n = p \cdot n.$$

wie gewünscht. □

Beispiel (Poisson-Verteilung)**.** Sei $X \sim \operatorname{Poi}(\lambda)$, dann behaupten wir, dass $\mathbb{E}(X) = \lambda$.

Beweis.

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k \geqslant 0} \frac{e^{-\lambda} \lambda^k}{k!} \cdot k\\ &= e^{-\lambda} \cdot \lambda \cdot \underbrace{\sum_{k \geqslant 1} \frac{\lambda^{k-1}}{(k-1)!}}_{\to e^{\lambda}}.\\ &= \lambda\end{aligned}$$

□

Bemerkung*. Diese Feststellung passt auch zur Konstruktion der Poisson-Verteilung.

Bemerkung. Oft kann man definieren

$$\psi(z) := \sum_{k\in\mathcal{S}} p(k)z^k.$$

Dann werden wir uns

$$\frac{d}{dz}\psi(z) = \sum_{k\in\mathcal{S}} p(k)kz^{k-1}$$

ansehen und bei $z = 1$ evaluieren, um $\mathbb{E}(X)$ zu berechnen.
Das ganze funktioniert, wenn X durch $\mathbb{P}(X = k) = p(k)$ verteilt ist und natürlich nur, wenn alle Objekte wohldefiniert sind.

Wir wollen auch Funktionen von X betrachten.

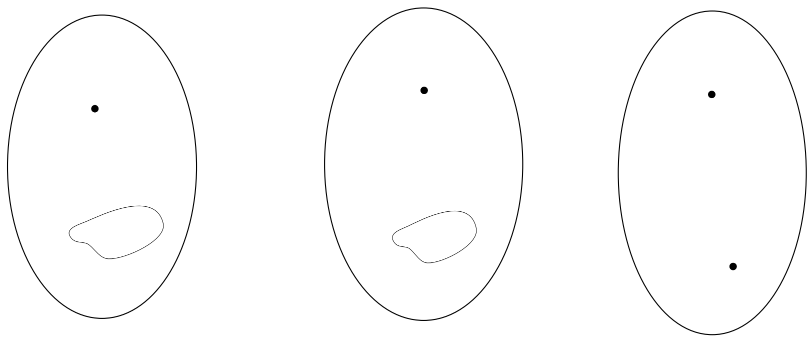


Abbildung 4: Funktionen von Zufallsvariablen

Satz 1.22 (Transformationssatz)**.** Sei $X : \Omega \to \mathcal{S}$ eine diskrete Zufallsvariable und $f : \mathcal{S} \to \mathbb{R}$ eine Funktion. Dann ist $f(X) := f \circ X : \Omega \to \mathbb{R}$ auch eine Zufallsvariable und

$$\mathbb{E}(f(X)) = \sum_{s\in\mathcal{S}} f(s)\mathbb{P}(X = s).$$

falls die Summe wohldefiniert ist.

Beweis. Messbarkeit: Es ist

$$\{f(X) = a\} = \bigcup_{s\in f^{-1}(a)} \{X = s\} \in \mathcal{F}.$$

weil

$$\{\omega \mid X(\omega) = s\} \in \mathcal{F}.$$

da X eine Zufallsvariable ist. Nach Definition ist nun

$$\begin{aligned}
\mathbb{E}(f(X)) &= \sum_{a\in f(\mathcal{S})} a\cdot\mathbb{P}(f(X)=a) \\
&= \sum_{a\in f(\mathcal{S})} a\cdot\mathbb{P}\left(\bigcup_{s\in f^{-1}(a)}\{X=s\}\right) \\
&= \sum_{a\in f(\mathcal{S})} a\cdot \sum_{s\in f^{-1}(a)}\mathbb{P}(X=s) \\
&= \sum_{a\in f(\mathcal{S})} \sum_{s\in f^{-1}(a)} f(s)\mathbb{P}(X=s) \\
&= \sum_{s\in\mathcal{S}} f(s)\mathbb{P}(X=s)
\end{aligned}$$

□

Der Erwartungswert ist LINEAR und MONOTON:

Satz 1.23 (Linearität des Erwartungswerts)**.** Seien $X_1 : \Omega \to \mathcal{S}_1$ und $X_2 : \Omega \to \mathcal{S}_2$ zwei diskrete Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Falls $\mathbb{E}(|X_1|) < \infty$ und $\mathbb{E}(|X_2|) < \infty$, dann ist $\forall \lambda_1, \lambda_2 \in \mathbb{R}$:

$$\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1\mathbb{E}(X_1) + \lambda_2\mathbb{E}(X_2).$$

Beweis. Es ist

$$|\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2)| \leqslant |\lambda_1|\mathbb{E}(|X_1|) + |\lambda_2|\mathbb{E}(|X_2|) < \infty.$$

(nach Dreiecksungleichung). Nun rechnen wir aus:

$$\mathbb{E}(\lambda_1 X_1 + \lambda_2 X_2) = \mathbb{E}(f(X_1, X_2)).$$

wobei $f(x_1, x_2) = \lambda_1 x_1 + \lambda_2 x_2$, also

$$\begin{aligned}
&= \sum_{\substack{x_1\in\mathcal{S}_1\\ x_2\in\mathcal{S}_2}} f(x_1,x_2)\mathbb{P}(X_1 = x_1 \cap X_2 = x_2) \\
&= \lambda_1 \sum_{x\in\mathcal{S}_1} x_i \sum_{x_2\in\mathcal{S}_2} \mathbb{P}(X_1 = x_1 \cap X_2 = x_2) + \text{sym.} \\
&= \lambda_1 \sum_{x\in\mathcal{S}_1} x_i\mathbb{P}(X_1 = x_1) + \text{sym.} \\
&= \lambda_1\mathbb{E}(X_1) + \lambda_2\mathbb{E}(X_2)
\end{aligned}$$

□

Korollar 1.24 (Monotonie des Erwartungswerts)**.** Seien X_1, X_2 reellwertige Zufallsvariablen mit $X_1(\omega) \leqslant X_2(\omega)$ für alle $w \in \Omega$. Dann ist

$$\mathbb{E}(X_1) \leqslant \mathbb{E}(X_2).$$

Beweis. Da $X_2(\omega) - X_1(\omega) \geqslant 0$, also ist trivialerweise $\mathbb{E}(X_2 - X_1) \geqslant 0$. Wegen der Linearität ist nun $\mathbb{E}(X_2 - X_1) = \mathbb{E}(X_2) - \mathbb{E}(X_1)$ und somit sind wir fertig. □

Beispiel. Seien $A_1, A_2, \dots, A_n \in \mathcal{F}$ mit $\mathbb{P}(A_i) = p$ für alle i. Sei $X_i := \mathbb{1}$. Dann ist $X_i \sim \text{Ber}(p)$ und $\mathbb{E}(X_i) = p$. Sei

$$\mathcal{S}_n := \sum_{i=1}^{n} X_i.$$

Dann ist

$$\mathbb{E}(\mathcal{S}_n) = \sum_{i=1}^{n} \mathbb{E}(X_i) = \sum_{i=1}^{n} p = np.$$

Das ist eine Verallgemeinerung des Falles, in dem $A_1, \dots, A_n$ unabhängig sind, weil dann $\mathcal{S}_n \sim \text{Bin}(n, p)$.

Oft interessieren wir uns auch dafür, wie weit eine Zufallsvariable von ihrem Ursprungswert entfernt ist. Ist z.B. $\mathbb{P}(X = k) = \mathbb{P}(X = -k)$, so ist $\mathbb{E}(X) = 0$, so sind immer noch folgende Fälle denkbar:

Plot einfügen

Frage. Wie weit sind die Werte vom X Mittelwert ($\mathbb{E}(X)$) entfernt?

Die Antwort liefert die sogenannte Varianz $\text{Var}(X)$:

Definition 1.25 (Varianz)**.** Sei X eine reellwertige Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{PP})$ mit $\mathbb{E}(X^2) < \infty$. Die **Varianz** von X ist durch

$$\text{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2).$$

definiert.

Lemma* (Eigenschaften der Varianz)**.** Es gilt folgendes:

i) $\text{Var}(X) \geqslant 0$, und $\text{Var}(X) = 0 \Leftrightarrow \mathbb{P}(X = \mathbb{E}(X)) = 1$, d.h. $X(\omega)$ ist in diesem Fall eine Konstante.

ii) Es ist

$$\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \qquad \text{Var}(\lambda \cdot X) = \lambda^2 \text{Var}(X.$$

iii) Die Varianz hängt nicht vom Erwartungswert ab, d.h.

$$\text{Var}(X) = \text{Var}(X + a) \qquad \forall a \in \mathbb{R}.$$

Beweis.* i) $\text{Var}(X) \geqslant 0$ ist klar, weil Quadrate nichtnegativ sind, Gleichheit gilt, wenn $X = \mathbb{E}(X)$ für jedes X, also wenn $\mathbb{P}(X = \mathbb{E}(X)) = 1$, und dann ist $X(\omega)$ konstant $\mathbb{E}(X)$.

ii) Wir rechnen nach:

$$\begin{aligned}\mathbb{E}((X - \mathbb{E}(X))^2) &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X\mathbb{E}(X)) + \mathbb{E}(\mathbb{E}(X)^2) \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X) \cdot \mathbb{E}(X) + \mathbb{E}(X)^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2\end{aligned}.$$

Für die 2. Behauptung ist

$$\text{Var}(\lambda X) = \mathbb{E}((\lambda X)^2) - \mathbb{E}(\lambda X)^2 = \mathbb{E}(\lambda^2 X^2) - (\lambda \mathbb{E}(X))^2 = \lambda^2 \text{Var}(X).$$

iii) Es ist

$$\mathrm{Var}(X+a) = \mathbb{E}((X+a-\mathbb{E}(X+a))^2) = \mathbb{E}((X-\mathbb{E}(X))^2) = \mathrm{Var}(X).$$

□

Beispiel. a) Ist $X \sim \mathrm{Ber}(p)$, so ist $\mathrm{Var}(X) = p(1-p)$.

b) Ist $X \sim \mathrm{Bin}(n,p)$, so ist $\mathrm{Var}(X) = n \cdot p(1-p)$.

c) Ist $X \sim \mathrm{Poi}(\lambda)$, so ist $\mathrm{Var}(X) = \lambda$

d) Ist $X \sim \mathrm{Geo}(q)$, so ist $\mathrm{Var}(X) = \frac{q}{(1-q)^2}$

Beweis. a) Wir benutzen $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. Nun ist

$$\mathbb{E}(X^2) = 0^2\mathbb{P}(X=0) + 1^2\mathbb{P}(X=1) = p.$$

und somit $\mathrm{Var}(X) = p - p^2 = p(1-p)$

b) Sei $X \sim \mathrm{Bin}(n,p)$, wir wissen bereits $\mathbb{E}(X) = p \cdot n$. Nun ist

$$\mathbb{E}(X^2) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} k^2.$$

Wir benutznen den gleichen Trick wie vorher nochmal, indem wir feststellen:

$$\begin{aligned}
\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} k^2 &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} k(k-1) + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} k \\
&= p^2 \frac{d^2}{dp^2}(p+q)^n + p\frac{d}{dp}(p+q)^n \\
&= p^2(p+q)^{n-1} n(n-1) + p(p+q)^{n-1} \cdot n
\end{aligned}.$$

Einsetzen von $q = 1-p$ liefert nun:

$$\mathbb{E}(X^2) = p^2 n(n-1) + p \cdot n = p^2n^2 - p^2 n + pn.$$

Damit erhalten wir schlussendlich

$$\mathrm{Var}(X) = p^2n^2 - p^2n + pn - (pn)^2 = np(1-p).$$

c) Ist $X \sim \mathrm{Poi}(\lambda)$, so wissen wir beretis $\mathbb{E}(X) = \lambda$. Nun ist

$$\begin{aligned}
\mathbb{E}(X^2) &= \sum_{k \geqslant 0} k^2 \cdot \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k\geqslant 0} k(k-1)\frac{e^{-\lambda}\lambda^k}{k!} + \sum_{k \geqslant 0} k \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \lambda^2 e^{-\lambda} \sum_{k \geqslant 2} \frac{\lambda^{k-2}}{(k-2)!} + e^{-\lambda}\lambda \sum_{k\geqslant 1} \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda^2 + \lambda
\end{aligned}.$$

Und damit ergibt sich für die Varianz:

$$\mathrm{Var}(X) = \lambda^2 + \lambda - (\lambda)^2 = \lambda.$$

□

Bemerkung. Sei $X : \Omega \to \mathcal{S}$ eine Zufallsvariable. Wir beobachten X n mal. Der Erwartungswert des Empirischen Masses ist genau der empirische Mittelwert, also für Beobachtungen $x_1, \ldots, x_n$ genau

$$m_n := \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Beispiel (Anwendung)**.** Ein alternativer Beweis des Einschluss-Ausschluss-Prinzips lässt sich mit der Linearität des erwartungswerts führen:
Es ist

$$\begin{aligned}
\mathbb{P}((A_1 \cup \ldots \cup A_n)^c) &= \mathbb{P}(A_1^c \cap \ldots \cap A_n^c) \\
&= E(1_{A^c \cap \ldots \cap A_n^c}) = (1_{A_1^c} \cdot \ldots \cdot 1_{A_n^c}) \\
&= E((1 - 1_A) \cdot \ldots \cdot (1 - 1_{A_n})) \\
&\overset{\text{Linearität}}{=} \sum_{k=0}^{n} (-1)^k \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathbb{E}(1_{A_{i_1}} \cdot \ldots \cdot 1_{A_{i_k}})^{\cdot} \\
&= \sum_{k=0}^{n} (-1)^k \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})
\end{aligned}$$

Komplementbildung liefert nun das gewünschte Ergebnis:

$$\begin{aligned}
\mathbb{P}(A_1 \cup \ldots \cup A_n) &= 1 - \mathbb{P}((A_1 \cap \ldots \cap A_n)^c) \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leqslant i_1 < \ldots < i_k \leqslant n} \mathbb{P}(A_{i_1} \cap \ldots \cap A_{i_k})^{\cdot}
\end{aligned}$$

Vorlesung 7
Mo 03 Mai 2021 10:17

2 Bedingte Wahrscheinlichkeit und Unabhängigkeit

2.1 Bedingte Wahrscheinlichkeit

Beispiel. Es werden statistische Daten über Kleinkinder gemessen: Wann sie krabbeln und wann sie laufen. Betrachten wir die Ereignisse

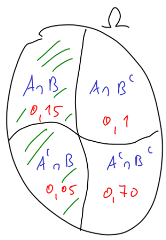
$$A = \{\text{Kind läuft vor dem 10. Monat}\} \qquad B = \{\text{Kind krabbelt vor dem 6. Monat}\}.$$

Aus den Daten geht hervor, dass $\mathbb{P}(A) = 25\%$ und $\mathbb{P}(B) = 20\%$.

> **Frage.** Sei ein Kind, das mit 6 Monaten krabbelt, gegeben. Wie hoch ist die Wahrscheinlichkeit, dass es mit 10 Monaten schon läuft?

Wir brauchen mehr Information als die obige, um die Frage beantworten zu können! Wir gehen also davon aus, dass wir sogar folgende Daten zur Verfügung haben:

$\cap$	A	A^c
B	0,15	0,05
B^c	0,10	0,70



Wir wissen, dass B eintritt, also befinden wir uns bereits im Zustandsraum $\Omega_B := \{w \in \Omega \mid w \in B\}$. Ziel ist es also, eine neue Massenfunktion $\mathbb{P}(\cdot|B)$ (auf Ω) zu definieren, die die Information '$\omega \in B$' berücksichtigt. Insbesondere muss also gelten:

$$\mathbb{P}(\omega|B) = 0 \qquad \forall\, \omega \in \Omega_B^c.$$

Zudem wollen wir, dass die Information '$\omega \in B$' dieselbe ist für alle $w \in \Omega_B$, d.h.

$$p(\omega|B) = \subseteq \cdot p(\omega) \qquad \forall\, \omega \in \Omega_B.$$

wobei $p(\omega)$ die Massenfunktion von $\mathbb{P}$ ist. Wegen Normierung ergibt sich also bereits

$$1 = \sum_{\omega\in\Omega} \mathbb{P}(\omega|B) = o \cdot \sum_{\omega\in\Omega_B} p(\omega) \quad \Leftrightarrow \quad c = \frac{1}{\mathbb{P}(B)}.$$

Also ergibt sich, dass

$$p(\omega \mid B) = \begin{cases} \frac{p(\omega)}{\mathbb{P}(B)} & \text{falls } \omega \in B \\ 0 & \text{sonst} \end{cases}.$$

Wir können das ganze so darstellen:

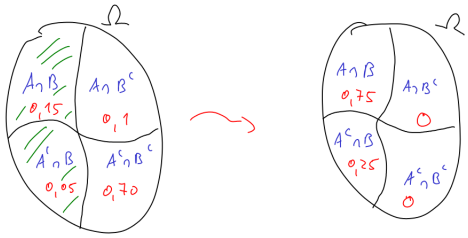


Abbildung 5: Änderung des Zustandsraums bei bedingten Wahrscheinlichkeiten

Wir erhalten also:
Antwort. Ein Kind, das mit 6 Monaten krabbelt, wird mit einer Wahrscheinlichkeit von 75% mit 10 Monaten laufen können.

Besser Skizzen machen

Definition 2.1 (Bedingte Wahrscheinlichkeit)**.** Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Seien $A, B \in \mathcal{F}$ Ereigniss mit $\mathbb{P}(B) \neq 0$. Dann definieren wir

$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

und nennen dise die **bedingte Wahrscheinlichkeit von A gegeben B**.

Bemerkung. Die Abbildung

$$\mathbb{P}(\cdot \mid B) : \left| \begin{array}{ccc} \mathcal{F} & \longrightarrow & \mathbb{R}_+ \\ A & \longmapsto & \mathbb{P}(A \mid B) \end{array} \right.$$

ist eine Wahrscheinlichkeitsverteilung auf $(\Omega, \mathcal{F})$, die wir auch die **bedingte Wahrscheinlichkeitsverteilung gegeben B** nennen.

Definition 2.2 (Bedingter Erwartungswert)**.** Sei $X : \Omega \to \mathcal{S} \subseteq \mathbb{R}$ eine (diskrete) Zufallsvariable mit Verteilung $\mathbb{P}(\cdot \mid B)$. Dann hat X den Erwartungswert

$$\sum_{s \in \mathcal{S}} s \cdot \mathbb{P}(X = s|B) =: \mathbb{E}(X \mid B).$$

Dieser heißt **bedingter Erwartungswert von X gegeben B**.

Beispiel. Wir werfen eine faire Münze N mal, dabei beobachten wir n mal das Ergebnis 'Zahl'.

> **Frage.** Was ist die Wahrscheinlichkeit, dass bei den ersten m Würfen immer 'Zahl' gefallen ist?

Ohne Weitere Informationen (dass insgesamt n mal Zahl gefallen ist) würden wir hier $\mathbb{P} \equiv \frac{1}{2^m}$ erhalten.
Betrachte nun den Zustandsraum

$$\Omega = \{\omega = (x_1, \ldots, x_N) \mid x_i \in \{0, 1\}, 1 \leqslant i \leqslant N\}.$$

wobei

$$x_k := \begin{cases} 1 & \text{falls } k\text{-ter Wurf ist 'Zahl'} \\ 0 & \text{falls } k\text{-ter Wurf ist 'Kopf'} \end{cases}.$$

und versehe ihn mit $\mathcal{F} = \mathcal{P}(\Omega)$ sowie $\mathbb{P}$ als Gleichverteilung auf Ω. Mit $X_k(\omega) := x_k$ interessieren wir uns also für

$$\mathbb{P}\left(X_1 = X_2 = \ldots = X_m = 1 \mid \sum_{k=1}^{N} X_k = n\right).$$

Nach Definition ist dies

$$
\begin{aligned}
&= \frac{\mathbb{P}\left((X_1 = \ldots = X_m = 1) \cap \left(\sum_{k=m+1}^{N} X_k = n - m\right)\right)}{\mathbb{P}\left(\sum_{k=1}^{N} X_k = n\right)} \\
&= \frac{\frac{1}{2^N}\binom{N-m}{n-m}}{\frac{1}{2^N}\binom{N}{n}} \\
&= \frac{(N-m)!}{(n-m)!(N-n)!} \cdot \frac{(N-n)!n!}{N!} \\
&= \frac{(N-m)!n!}{N!(n-m)!}
\end{aligned}
.
$$

Notation. Zur Vereinfachung der Notation schreiben wir oft

$$\mathbb{P}(X_1 = X_2 = a) = \mathbb{P}(\{\omega \in \Omega \mid X_1(\omega) = a\} \cap \{\omega \in \Omega \mid X_2(\omega) = a\}).$$

sowie

$$\mathbb{P}(X_1 = a, X_2 = b) = \mathbb{P}(\{\omega \in \Omega \mid X_1(\omega) = a\} \cap \{\omega \in \Omega \mid X_2(\omega) = b\}).$$

Wir haben gerade aus einer Wahrscheinlichkeitsverteilung $\mathbb{P}$ die Verteilung $\mathbb{P}(\cdot \mid B)$ gewonnen. Das ganze geht auch umgekehrt:

Satz 2.3. Sei $\Omega = \bigcup_{k \in I} H_k$ eine disjunkte Zerlegung von Ω in (abzählbar viele) Ereignisse $H_k, k \in I$, wobei $\mathbb{P}(H_k) \neq 0$. Dann ist $\forall A \in \mathcal{F}$:

$$\mathbb{P}(A) = \sum_{k \in I} \mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k).$$

Beweis. $\forall A \in \mathcal{F}$ ist

$$A = A \cap \bigcup_{k \in I} H_k = \bigsqcup_{k \in I} (A \cap H_k).$$

eine disjunkte Vereinigung. Also folgt aus σ-Additivität, dass

$$
\begin{aligned}
\mathbb{P}(A) &= \sum_{k \in I} \mathbb{P}(A \cap H_k) \\
&= \sum_{\substack{k \in I \\ \mathbb{P}(H_k) \neq 0}} \mathbb{P}(A \cap H_k) \\
&= \sum_{\substack{k \in I \\ \mathbb{P}(H_k) \neq 0}} \mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k)
\end{aligned}
$$

□

Beispiel. Eine Urne A enthält 2 rote und 3 blaue Kugeln. In Urne B liegen umgekehrt 3 rote und nur 2 blaue Kugeln. Wir gehen davon aus, dass die Urnen immer gut gemischt sind. Nun machen wir Folgendes:

1. Wir ziehen eine Kugel K_1 aus A und legen sie in B

(2) Wir ziehen eine Kugel K_2 aus B und lg

Frage. Was ist die Wahrscheinlichkeit, dass K_2 rot ist?

Wir erhalten nun

$$\begin{aligned}\mathbb{P}(K_2 \text{ ist rot}) &= \mathbb{P}(K_2 \text{ rot} \mid K_1 \text{ rot}) \cdot \mathbb{P}(K_1 \text{ rot}) \\ &+ \mathbb{P}(K_2 \text{ rot} \mid K_1 \text{ blau}) \cdot \mathbb{P}(K_1 \text{ blau}). \\ &= \frac{4}{6} \cdot \frac{2}{5} + \frac{3}{6} \cdot \frac{3}{5} = \frac{17}{30}\end{aligned}$$

Graphik einfügen

2.2 Baye'sche Regel

In der Baye'schen Statistik ist $\mathbb{P}(H_k)$ auch die **a-priori-Einschätzung** der Wahrscheinlichkeit einer Hypothese H_k, das könnte z.B. sein

$H_k = \{\text{Die Unfallskosten pro Jahr liegen im Bereich } [100k, 100(k+1))\}.$

Aus statistischen Daten weiß man, dass ein Ereignis $A \in \mathcal{F}$ mit einer Wahrscheinlichkeit $\mathbb{P}(A) \neq 0$ eintritt, also z.B.

$$A =' \text{Es handelt sich um einen Auffahrunfall}'.$$

Dazu kennt man $\mathbb{P}(A \mid H_k)$. Falls A eintritt, werden die Versicherungskosten neu berechnet, auf der Basis von

$$\mathbb{P}(H_k \mid A).$$

Dies nennt man dann auch **a-posteriori-Verteilung** von H_k.

Korollar 2.4 (Baye'sche Regel)**.** Für $A \in \mathcal{F}$ mit $\mathbb{P}(A) \neq 0$ gilt

$$\mathbb{P}(H_k \mid A) = \frac{\mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k)}{\sum\limits_{\substack{l \in I \\ \mathbb{P}(H_l) \neq 0}} \mathbb{P}(A \mid H_l) \cdot \mathbb{P}(H_l)}.$$

Beweis. Es ist

$$\mathbb{P}(H_k \mid A) = \frac{\mathbb{P}(H_k \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \mid H_k) \cdot \mathbb{P}(H_k r}{\mathbb{P}(A)}.$$

Aus Satz 2.3 erhalten wir nun aber genau

$$\mathbb{P}(A) = \sum_{\substack{l \in I \\ \mathbb{P}(H_l) \neq 0}} \mathbb{P}(A \mid H_l) \cdot \mathbb{P}(H_l).$$

und wir sind fertig. □

Beispiel. Eine Krankheit K tritt selten auf, mit einer Häufigkeit von 10^{-4}, also bei 10 von 100.000 Menschen. Ein Test zur Erkennung der Krankheit ist positiv (+) bei 96% der Kranken und 0,1% der Gesunden.
Der Test liefert also 0,1% falsch positive und 4% falsch negative Ergebnisse.

Frage. Was ist die Wahrscheinlichkeit, dass jemand krank ist, sofern er positiv getesten wurde?

- Die a-priori-Wahrscheinlichkeit beträgt $\mathbb{P}(k) = 10^{-4}$ sowie $\mathbb{P}(K^c) = 1 - 10^{-4}$.
- Wir kennen die bedingten Wahrscheinlichkeiten $\mathbb{P}(+ \mid K) = 0,96$ sowie $\mathbb{P}(+ \mid K^c) = 0,001$.
- Als A-posteriori Wahrscheinlichkeit erhalten wir nun:

$$\begin{aligned}\mathbb{P}(K \mid +) &= \frac{\mathbb{P}(+ \mid K) \cdot \mathbb{P}(K)}{\mathbb{P}(+ \mid K) \cdot \mathbb{P}(K) + \mathbb{P}(+ \mid K^c) \cdot \mathbb{P}(K^c)} \\ &= \frac{0,96 \cdot 10^{-4}}{0,96 \cdot 10^{-4} + 0,001 \cdot (1 - 10^{-4})} \\ &\approx 9,6\%\end{aligned}$$

Antwort. Die Wahrscheinlichkeit, dass man krank ist, wenn man positiv getestet ist, beträgt also (nur) $9,6\%$.

Bemerkung*. Ein Test hat üblicherweise eine Sensitivität und ein Spezifität. Die Sensitivität gibt an, welcher Anteil der tatsächlich infizierten positiv getestet werden. Die Spezifität gibt an, welcher Anteil der gesunden Menschen auch negativ gestetest wird.

Beispiel (Aktuelle Corona-Zahlen)**.** Bei den aktuellen Schnelltestes gibt es (in etwa) eine falsch-positiven Rate von 2%, also $\mathbb{P}(B \mid K^c) = 2\%$, und eine falsch-negativen Rate von 20%, also $\mathbb{P}(- \mid K) = 20\%$. Bei einer Inzidenz von 150-200 pro 100.000 Einwohner pro Woche, einer Dunkelziffer nah bei 2 ergibt sich eine Schätzung der aktuell infizierten von

$$\mathbb{P}(K) \in [0.005, 0.01].$$

(Zum Vergleich: Die aktuell gemeldeten positiven Fälle liegen bei $0,0035$).
Nun können wir wieder berechnen:

$$\begin{aligned}\mathbb{P}(K \mid +) &= \frac{\mathbb{P}(+ \mid K) \cdot \mathbb{P}(K)}{\mathbb{P}(+ \mid K) \cdot \mathbb{P}(K) + \mathbb{P}(+ \mid K^c) \cdot \mathbb{P}(K^c)} \\ &= \frac{0,8 \cdot \mathbb{P}(K)}{0,8\mathbb{P}(K) + 0,02(1 - \mathbb{P}(K)}\end{aligned}$$

Wir erhalten

(a) Mit $\mathbb{P}(K) = 0,005$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid +) \approx 17\%$
(b) Mit $\mathbb{P}(K) = 0,01$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid +) \approx 29\%$
(c) Mit $\mathbb{P}(K) = 0,001$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid +) \approx 3,8\%$

> **Frage.** Lohnt es sich die Schnelltests in den Schulen zu machen? Also: Was ist $\mathbb{P}(K \mid -)$

Auch das lässt sich mit der gleichen Formel beantworten, mit $\mathbb{P}(- \mid K) = 0,2$ und $\mathbb{P}(- \mid K^c) = 0,98$ erhalten wir

(a) Für $\mathbb{P}(K) = 0,005$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid -) \approx 0,1\%$

b Für $\mathbb{P}(K) = 0,01$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid -) \approx 0,2\%$
c Für $\mathbb{P}(K) = 0,001$ eine Wahrscheinlichkeit von $\mathbb{P}(K \mid -) \approx 0,02\%$

Vorlesung 8
Mi 05 Mai 2021 10:11

2.3 Mehrstufige Modelle

Sei eine Folge von n Zufallsexperimenten in den Wahrscheinlichkeitsräumen $\Omega_1, \Omega_2, \ldots, \Omega_n$ gegeben. Wir definieren ein **n-stufiges Zufallsexperiment** durch

- $\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n = \{\omega = (\omega_1, \ldots, \omega_n) \mid \omega_k \in \Omega_k, 1 \leqslant k \leqslant n\}$
- $\mathcal{F} = \mathcal{P}(\Omega)$.
- Definiere die Zufallsvariablen
$$X_k(\omega) = \omega_k \qquad 1 \leqslant k \leqslant n.$$
Den Index k interpretieren wir hierbei als Zeit. $k \mapsto X_k$ ist eine Trajektion von $X = (X_1, \ldots, X_n)$?
- $\mathbb{P} =?$. Wir konstruieren $\mathbb{P}$ auf $(\Omega, \mathcal{F})$ mit
  a Der Anfangsverteilung $\mathbb{P}(X_1 = x_1) := p_1(x_1)$ für alle $x_1 \in \Omega_1$.
  b Den bedingten Verteilungen
$$\mathbb{P}(X_k = x_k \mid X_1 = x_1, X_2 = x_2, \ldots, X_{k-1} = x_{k-1}) =: p_k(x_k \mid x_1, \ldots, x_{k-1}).$$
  für alle $x_l \in \Omega_l$, $1 \leqslant l \leqslant k-1$, sodass $\mathbb{P}(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) \neq 0$.

i

Bemerkung*. Man kann das Allgemeiner machen, indem wir $\mathcal{F}$ als die Produktsigmalalgebra der $\mathcal{F}_i$ wählen.

Satz 2.5. Sei $p_i(\cdot)$ die Massenfunktion einer Wahrscheinlichkeitsverteilung auf Ω_i und $p_k(\cdot \mid x_1, \ldots, x_{k-1})$ für alle $1 \leqslant k \leqslant n$ mit $x_1 \in \Omega_1, \ldots, x_{k-1} \in \Omega_{k-1}$ eine Massenfunktion auf Ω_k.
Dann existiert eine eindeutige Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $(\Omega, \mathcal{F})$, sodass
a $\mathbb{P}(X_1 = x_1) = p_1(x_1) \quad \forall x_1 \in \Omega_1$
b $\mathbb{P}(X_k = x_k) \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) = p_k(x_k \mid x_1, \ldots, x_{k-1})$
Die Wahrscheinlichkeitsverteilung $\mathbb{P}$ hat die Massenfunktion
$$p(x_1, \ldots, x_n) = p_1(x_1) p_2(x_2 \mid x_1) \cdot \ldots \cdot p_n(x_n \mid x_1, \ldots, x_{n-1}).$$

Beweis. 1) Nimm zunächst an, dass solch ein Maß existiert, wir zeigen die letzte Aussage. Sei $\mathbb{P}$ sodass a und b erfüllt sind. Dann ist
$$\forall 1 \leqslant k \leqslant n \colon \mathbb{P}(X_1 = x_1, \ldots, X_k = x_k) = p(x_1, \ldots, x_k).$$

- Für $k = 1$ gilt das (aus a).
- Falls es für $k-1$ gilt, so haben wir die Fälle
- $p(x_1, \ldots, x_{k-1}) = 0$, dann ist $0 = 0$ wahr.
- Falls $p(x_1, \ldots, x_{k-1}) \neq 0$, so ist
$$\begin{aligned}&\mathbb{P}(X_1 = x_1, \ldots, X_k = x_k)\\ &= \mathbb{P}(X_k = x_k \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) \cdot \mathbb{P}(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) \quad .\\ &= p_1(x_1) \cdot \ldots p_{k-1}(x_{k-1} \mid x_1, \ldots, x_{k-2} r \cdot p_k(x_k \mid x_1, \ldots, x_{k-1}) = p(x_1, \ldots, x_k)\end{aligned}$$

Normierung: $\forall x \in \Omega, x = (x_1, \ldots, x_n)$ mit $x_k \in \Omega_k$ ist

$$\sum_{x \in \Omega} p(x) = \sum_{x_1 \in \Omega_1} \ldots \sum_{x_n \in \Omega_n} p(x_1, \ldots, x_n)$$
$$= \sum_{x_1 \in \Omega_1} p(x_1) \sum_{x_2 \in \Omega_2} p(x_2 \mid x_1) \ldots \sum_{x_n \in \Omega_n} p(x_n \mid x_1, \ldots, x_{n-1}) = 1^{\cdot}$$

Für Eigenschaft b ist

$$\mathbb{P}(X_1 = x_1, \ldots, X_k = x_k) = \sum_{x_{k+1} \in \Omega_{k+1}} \ldots \sum_{x_n \in \Omega_n} p(x_1, \ldots, x_n)$$
$$= p_1(x_1) \ldots p_{k-1}(x_{k-1} \mid x_1, \ldots, x_{k-2}) p_k(x_k \mid x_1, \ldots, x_{k-1})$$
$$= \mathbb{P}(X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) p_k(x_k \mid x_1, \ldots, x_{k-1})$$

Also erhalten wir

$$\mathbb{P}(X_k = x_k \mid X_1 = x_1, \ldots, X_{k-1} = x_{k-1}) = p_k(x_k \mid x_1, \ldots, x_{k-1}).$$

□

Beweis nochmal sortieren

Anmerkung. Mir ist noch nicht klar, wo wir im Beweis des Satzes jetzt gezeigt haben wollen, dass solch ein Wahrscheinlichkeitsmaß existiert, das muss ich noch ausarbeiten.

Bemerkung. Falls $p_k(x_k \mid x_1, \ldots, x_{k-1})$ nur eine Funktion von $x_{k-1}, \ldots, x_{k-m-1}$, dann sagen wir, dass unser Modell ein Gedächtnis von m Schritten hat.

2.3.1 Produktmodelle

Falls $p_k(x_k \mid x_1 = \ldots = x_{k-1}) = p_k(x_k)$, d.h. x_k hängt nicht von den Werten $x_1, \ldots, x_{k-1}$ ab. Dann erhalten wir aus Satz 2.5, dass

$$p(x_1, \ldots, x_n) = \prod_{k=1}^{n} p_k(x_k).$$

Definition 2.6 (Produktmodell)**.** Die Wahrscheinlichkeitsverteilung $\mathbb{P}$ auf $\Omega = \Omega_1 \times \ldots \times \Omega_n$ mit Massenfunktion

$$p(x_1, \ldots, x_n) = \prod_{k=1}^{n} p_k(x_k).$$

heißt **Produkt von** $\mathbb{P}_1, \ldots, \mathbb{P}_n$. ($\mathbb{P}_k$ hat Massenfunktion p_k).

Notation. Wir schreiben $\mathbb{P} = \mathbb{P} \otimes \mathbb{P}_2 \otimes \ldots \otimes \mathbb{P}_n$, wenn $\mathbb{P}$ das Produkt von $\mathbb{P}_1, \ldots, \mathbb{P}_n$ ist.

Beispiel. Seien n unabhängige 0-1-Experimente mit Erfolgswahrscheinlichkeit p gegeben. Also

$$\Omega_1 = \ldots = \Omega_n = \{0, 1\}.$$

und $p_k(1) = p = 1 - p_k(0)$ für $k = 1, \ldots, n$. dann ist $p_k(x) =$

$(1-p)\left(\frac{p}{1-p}\right)^x$ für $x \in \{0,1\}$. Die entstehende Verteilung

$$p(x_1,\ldots,x_n) = (1-p)^n \prod_{k=1}^{n} \left(\frac{p}{1-p}\right)^{x_k}.$$

ist die **n-dimensionale Bernoulli-Verteilung mit Parameter p**
.

Satz 2.7. Sei $(\Omega, \mathcal{F}, \mathbb{P})$ ein Produktmodell. Dann ist für beliebige Ereignisse $A_k \subseteq \Omega_k$, $k = 1,\ldots,n$:

$$\mathbb{P}(A_1 \times \ldots \times A_n) = \prod_{k=1}^{n} \mathbb{P}_k(A_k).$$

und $\mathbb{P}(\tilde{A}_k) = \mathbb{P}_k(A_k)$, wobei

$$\tilde{A}_k := \Omega_1 \times \ldots \times \Omega_{k-1} \times A_k \times \Omega_{k+1} \times \ldots \times \Omega_n.$$

Deswegen sind $\tilde{A}_1,\ldots,\tilde{A}_n$ unabhängige Ereignisse.

Beweis. Es ist

$$\begin{aligned}
\mathbb{P}(A_1 \times \ldots \times A_n) &= \mathbb{P}((X_1,\ldots,X_n) \in A_1 \times \ldots \times A_n)) \\
&= \sum_{(x_1,\ldots,x_n) \in A_1 \times \ldots \times A_n} p(x_1,\ldots,x_n) \\
&= \sum_{x_1 \in A_1} \ldots \sum_{x_n \in A_n} p_1(x_1) \cdot \ldots p_n(x_n) \\
&= \prod_{k=1}^{n} \sum_{x_k \in A_k} p_k(a_k) \\
&= \prod_{k=1}^{n} \mathbb{P}(A_k)
\end{aligned}$$

Es ergibt sich leicht

$$\begin{aligned}
\mathbb{P}(\tilde{A}_k) &= \mathbb{P}(X_k \in A_k, X_l \in \Omega_l \ \forall l \neq k) \\
&= \left(\prod_{l \neq k} \underbrace{\mathbb{P}_l(X_l \in \Omega_l)}_{=1}\right) \cdot \mathbb{P}_k(X_k \in A_k) \\
&= \mathbb{P}k(A_k)
\end{aligned}$$

Damit ergibt sich schlussendlich für beliebiges $I = \{i_1,\ldots,i_l\}$:

$$\begin{aligned}
\mathbb{P}(\tilde{A}_{i_1} \cap \ldots \cap \tilde{A}_{i_l}) &= \mathbb{P}\left(\bigcap_{i \in I} \{X_i \in A_i\} \cap \bigcap_{j \neq I} \{x_j \in \Omega_j\}\right) \\
&= \prod_{i \in I} \mathbb{P}_i(A_i) \cdot \prod_{j \neq I} \underbrace{\mathbb{P}_j(\Omega_j)}_{=1} \\
&= \prod_{i \in I} \mathbb{P}_i(A_i) \\
&\overset{\mathbb{P}(A_i)=\mathbb{P}(\tilde{A}_i)}{=} \prod_{k=1}^{l} \mathbb{P}(\tilde{A}_{i_k})
\end{aligned}$$

□

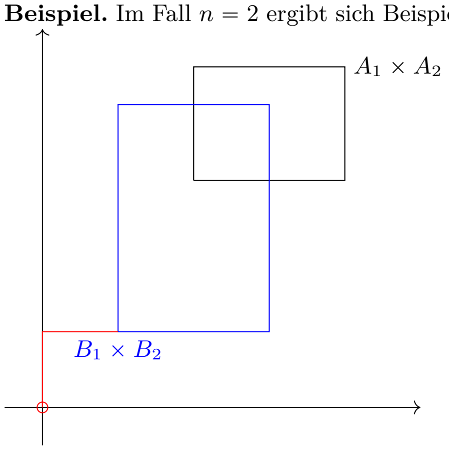
Bemerkung.
- Es ist $\Omega = \Omega_1 \times \ldots \times \Omega_n$
- Eigentlich müssen wir $(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)$ als entsprechende Wahrscheinlichkeitsräume betrachten, wir unterdrücken aber oft die Notation $\mathcal{F}_i, \mathbb{P}_i$

Im Allgemeinen setzen wir

$$\mathcal{F} = \mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_n.$$

wobei $\mathcal{F}$ dann die σ-Algebra ist, die von $A_1 \times \ldots \times A_n$ mit $A_i \in \mathcal{F}_i$ erzeugt ist. Im Spezialfall $F_i = \mathcal{P}(\Omega_i)$ ergibt sich insbesondere wieder der uns bekannte Fall $\mathcal{F} = \mathcal{P}(\Omega)$.
Für Produktmodelle erhalten wir also $\Omega = \Omega_1 \times \ldots \times \Omega_n$, $\mathcal{F} = \mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_n$ sowie $\mathbb{P} = \mathbb{P}_1 \otimes \ldots \otimes \mathbb{P}_n$.
Beachte, dass $\mathcal{F}_1 \otimes \ldots \otimes \mathcal{F}_n \neq \mathcal{F}_1 \times \ldots \times \mathcal{F}_n$ im Allgemeinen, wie das folgende Beispiel zeigt.

Beispiel. Im Fall $n = 2$ ergibt sich Beispielsweise folgende Situation:



Vorlesung 9
Mo 10 Mai 2021 10:15

2.3.2 Markovketten (MK)

- Setze $X = (X_0, X_1, X_2, \ldots, X_n)$. Die Zeit beginnt hier bei $k = 0$.
- Betrachte $\Omega_k = \mathcal{S}$ für festes $\mathcal{S}$, also

$$\Omega = \mathcal{S}^{n+1} = \{(x_0, \ldots, x_n) \mid x_i \in \mathcal{S}, 0 \leqslant i \leqslant n\}.$$

Definition 2.8 (Markovkette)**.** Eine **Markovkette** (abgekürzt: MK) ist ein mehrstufiges Modell mit der Eigenschaft

$$p_k(x_k \mid x_0, \ldots, x_{k-1}) = p_k(x_k \mid x_{k-1}).$$

Frage. Sei $\mathcal{S}$ abzählbar. Wie beschreibt man die Übergänge von X_k nach X_{k+1}?

Definition 2.9 (Sotchastische Matrix)**.** Eine Matrix $P = [\mathbb{P}(x,y)]_{x,y\in\mathcal{S}}$ mit den Eigenschaften
(a) $\forall x\,\forall y\colon \mathbb{P}(x,y) \geqslant 0$
(b) $\forall x \in \mathcal{S}\colon \sum_{y\in\mathcal{S}} \mathbb{P}(x,y) = 1$
heißt **stochastische Matrix**.

Bemerkung*. Beachte, dass die Matrix in obiger Definition nicht zwingend endlich sein muss, Definitionen verallgemeinern sich kanonisch. Wir fordern aber, dass $\mathcal{S}$ abzählbar ist.

Lemma 2.10. Die Matrix P_k mit den Einträgen

$$P_k(x,y) = p_k(Y \mid X) \quad \forall x, y, \in \mathcal{S}.$$

ist eine stochastische Matrix.

Beweis. Offenbar ist $p_k(Y \mid X) \geqslant 0$. Zudem

$$\begin{aligned}\sum_{y\in\mathcal{S}} P_k(x,y) &= \sum_{y\in\mathcal{S}} \mathbb{P}(X_k = Y \mid X_{k-1} = x)\\ &= \mathbb{P}\left(\bigcup_{y\in\mathcal{S}} \{X_k = y\} \mid X_{k-1} = x\right)\\ &= \mathbb{P}(\Omega_k \mid X_{k-1} = x)\\ &= 1\end{aligned}$$

weil es sich bei $\mathbb{P}(\cdot \mid X_{k-1} = x)$ um eine Wahrscheinlichkeitsverteilung handelt, und $\mathcal{S} = \bigsqcup_{y\in\mathcal{S}} \{y\}$ eine disjunkte Vereinigung ist. □

Bemerkung. P_k ist eine sogenannte **Übergangsmatrix**. Sie beschreibt den Übergang der Markovkette von Ω_k nach Ω_{k+1}

Die Massenfunktion einer Markovkette ist

$$p(x_0, x_1, \ldots, x_n) = p_0(x_0) \cdot P_1(x_0, x_1) \cdot \ldots \cdot P_n(x_{n-1}, x_n).$$

wobei p_0 die sogenannte **Anfangsverteilung** ist.

Bemerkung. Falls $P_k = \mathbb{P}$, d.h. die Übergangsmatrixk hängt nicht von k ab, dann heißt die Markovkette (zeitlich) **homogen**.

Bemerkung. Seien P, Q zwei stochastische Matrizen. Dann ist auch $P \cdot Q$ eine stochastische Matrix, wobei

$$(P \cdot Q)(x,y) = \sum_{z\in\mathcal{S}} P(x,z) \cdot Q(z,y).$$

Frage. Was ist
1) $\mathbb{P}(X_n = x)$
2) $\lim\limits_{n\to\infty} \mathbb{P}(X_n = x)$ (Existiert dieser überhaupt?)
3) Ist $\lim\limits_{n\to\infty} \mathbb{P}(X_n = x)$ von x_0 abhängig?

Satz 2.11 (Massenfunktion in Markovketten)**.** Sei μ_0 der Zeilenvektor mit Elementen $p_0(x), x \in \mathcal{S}$. Seien dazu $P_1, P_2, \ldots, P_n$ die Übergangsmatrizen einer Markovkette $X = (X_0, X_1, \ldots, X_n)$ auf $\mathcal{S}$. Dann hat die Wahrscheinlichkeit von X_n die Massenfunkiton

$$\mu_n(x) := \mathbb{P}(X_n = x) = (\mu_0 \cdot P_1 \cdot \ldots \cdot P_n)(x) \quad \forall x \in \mathcal{S}.$$

Beweis.

$$\begin{aligned}\mathbb{P}(X_n = x) &= \sum_{x_0,\ldots,x_{n-1}\in\mathcal{S}} \mathbb{P}(X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = x)\\ &= \mu_0(x_0)P_1(x_0,x_1)\ldots P_n(x_{n-1},x)\end{aligned}$$

$\square$

Beispiel. (a) Produktmodelle sind Markovketten.

(b) Irrfahrten (eng: 'Random Walk') auf $\mathbb{Z}^d$ sind Markovketten:
Sei $\mathcal{S} = \mathbb{Z}^d$ mit $d \in \mathbb{N}$ fest. Die (symmetrische) Irrfahrt ist eine homogene Markovkette mit

$$P_k(x,y) = P(x,y) = \begin{cases} \frac{1}{2d} & \text{falls } \|x-y\| = 1 \\ 0 & \text{sonst} \end{cases}.$$

In jedem Schritt bewegen wir uns also auf einen benachbarten Gitterpunkt.

(c) **Urnenmodell von Ehrenfest**. Wir haben ein System mit N Teilchen, die auf zwei Urnen A, B verteilt sind. Zu jedem Zeitpunkt $t \in \mathbb{N}$ welchselt eine zufällig ausgewählte Kugel die Urne.
In der Makroskopischen Modellierung sei $n_A :=$ #Teilchen in A. Dann ist

$$\rho_A := \frac{n_A}{N} \in \left\{0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1\right\} =: \mathcal{S}.$$

und die Markovkette, die die Zeitentwicklung von ρ_A beschreibt, hat die Übergangsmatrix

$$P(x,y) = \begin{cases} x & \text{falls } y = x - \frac{1}{N} \\ 1-x & \text{falls } y = x + \frac{1}{N} \\ 0 & \text{sonst} \end{cases}.$$

Die Fälle spiegeln wieder, dass wir ein Teilchen aus A bzw. B gezogen haben, wobei der dritte Fall dijenigen y abdeckt, die wir nicht erreichen können, weil sich n_A immer nur um $\pm_1$ ändert.

In der Mikroskopischen Modellierung nummerieren wir die Teilchen, d.h.

$$\mathcal{S} = \{0,1\}^n = \{\sigma = (\sigma_1, \dots, \sigma_N) \mid \sigma_k \in \{0,1\}, 1 \leqslant k \leqslant N\}.$$

wobei

$$\sigma_k = \begin{cases} 1 & \text{falls Teilchen } k \text{ ist in } A \\ 0 & \text{falls Teichen } k \text{ ist in } B \end{cases}.$$

Die Dynamik von $\sigma \in \mathcal{S}$ wir durch die Markovkette beschrieben:

$$P(\sigma, \tilde{\sigma}) = \begin{cases} \frac{1}{N} & \text{falls } \sum_{k=1}^N |\sigma_k - \tilde{\sigma}_k| = 1 \\ 0 & \text{sonst} \end{cases}.$$

Bemerkung. Die Notation $\sum_{k=1}^N |\sigma_k - \tilde{\sigma_k}| = 1$ ist nur eine kurze (elegante) Möglichkeit auszudrücken, dass sich σ_k nach $\tilde{\sigma}_k$ in nur einem Eintrag ändert.

Es ergibt sich also eine Irrfahrt auf dem Hyperwürfel $\{0,1\}^N$.

Frage. Was ist $P(X_l = y \mid X_k = x)$ für $k < l$ beliebig?

Intuitiv sollten wir hier $P_{k+1} \cdot \ldots \cdot P_l(x_k, x_l)$ erhalten, indem wir die Übergangsmatrizen multiplizieren.

Satz 2.12 (Markoveigenschaft)**.** Sei X eine Markovkette auf $\mathcal{S}$. Für alle $0 \leqslant k \leqslant l \leqslant n$ und $x_0, \dots, x_l \in \mathcal{S}$ mit $\mathbb{P}(X_0 = x_0, \dots, X_k = x_k) \neq 0$ ist

$$\begin{aligned} \mathbb{P}(X_l = x_l \mid X_0 = x_0, \dots, X_k = x_k) &= \mathbb{P}(X_l = x_l \mid X_k = x_k) \\ &= (P_{k+1} \cdot \ldots \cdot P_l)(x_k, x_l) \end{aligned} \tag{13}$$

Diese Eigenschaft (erstes Gleichheitszeichen) heißt auch **Markov-Eigenschaft**.

Beweis. Es ergibt sich

$$\begin{aligned}
&P(X_l = x_l \mid X_0 = x_0, \dots, X_k = x_k) \\
\overset{\text{Def.}}{=}\ & \frac{P(X_0 = x_0, \dots, X_k = X_k, X_l = x_l)}{P(X_0 = x_0, \dots, X_k = X_k)} \\
=\ & \frac{\sum_{\substack{x_{k+1},\dots,x_{l-1} \\ x_{l+1},\dots,x_n \in \mathcal{S}}} p_0(x_0) P_1(x_0, x_1) \cdot \ldots \cdot P_k(x_{k-1}, x_k) \cdot P_{k+1}(x_k, x_{k+1}) \cdot \ldots \cdot P_l(x_{l-1}, x_l) \cdot P_{l+1}(x_l, x_{l+1}) \cdot \ldots \cdot P_{(}}{\sum_{x_{k+1},\dots,x_n \in \mathcal{S}} \text{analog}} \\
=\ & \frac{\sum_{x_{k+1},\dots,x_{l-1} \in \mathcal{S}} p_0(x_0) \cdot \ldots \cdot P_k(x_{k-1}, x_k) \cdot P_{k+1}(x_1, x_{k+1}) \cdot P_l(x_{l-1}, x_l)}{p_0(x_0) \cdot \ldots \cdot P_k(x_{k-1}, x_k)} \\
=\ & \sum_{x_{k+1},\dots,x_{l-1} \in \mathcal{S}} P_{k+1}(x_k, x_{k+1}) \cdot P_{l(x_{l-1}, x_l)} \\
=\ & (P_{k+1} \cdot \ldots \cdot P_l)(x_k, x_l)
\end{aligned}$$

Das zeigt die erste Gleichheit. Dazu ist

$$\begin{aligned}
&\mathbb{P}(X_l = x_l \mid X_k = x_k)\\
&= \frac{\mathbb{P}(X_l = x_l, X_k = x_k)}{\mathbb{P}(X_k = x_k)}\\
&= \frac{\sum_{x_{k+1},\dots,x_{l-1}\in\mathcal{S}} (\mu_0 P_1 \dots P_k)(x_k)\cdot P_{k+1}(x_k, x_{k+1}\cdot P_l(x_{l-1}, x_l)}{(\mu_0 P_1 \cdot \dots \cdot P_k)(x_k)}\\
&= (P_{k+1}\cdot \dots \cdot P_l)(x_k, x_l)
\end{aligned}$$

□

Beispiel. Ist $\mathcal{S} = \{1,\dots,N\}$, so ergibt sich mit $\mu_0 = (\mu_0(1),\dots,\mu_0(N))$ ...

$$= (\sum_k \mu_0(k)P(k,1), \sum_k \mu_0(k)P(k,2),\dots) = ((\mu P)(1), (\mu_0 P)(2),\dots).$$

Beispiel fertig

Vorlesung 10
Mi 12 Mai 2021 10:16

Beispiel. Sei $\mathcal{S} = \{0,1\}$ und $\alpha,\beta \in (0,1)$. Setze

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

Oft machen wir eine graphische Darstellung: Wir wollen also P^n ausrechnen, um das Verhalten der Markovkette zu studieren.

(a) Man könnte nun P diagonalisieren, also $P = U\Lambda U^{-1}$, wobei A Diagonalform hat, so ist $P^n = U\Lambda^n U^{-1}$, und die Potenzen von Λ können wir leicht berechnen.

(b) P^n ist stochastisch, also ist auch $P^n(0,0) + P^n(0,1) = 1$, $P^n(1,1) + P^n(1,0) = 1$, weil P^n stochastisch ist. Wir wollen uns nun, welchen Wert $P^n(0,0)$ hat, dazu machen wir den Rekursion Ansatz:
   - Sind wir nach $n-1$ Schritten wieder bei der 0, so müssen wir von 0 nach 0 gehen
   - Sind wir nach $n-1$ Schritten bei der 1, so müssen wir im n-ten von 1 nach 0 gehen

   Also ergibt sich

$$\begin{aligned}
P^n(0,0) &= P^{n-1}(0,0)\cdot P(0,0) + P^{n-1}(0,1)\cdot P(1,0)\\
&= P^{n-1}(0,0)(1-\alpha) + \underbrace{P^{n-1}(0,1)}_{=1-P^{n-1}(0,0)}\cdot\beta\\
&= P^{n-1}(0,0)(1-\alpha-\beta) + \beta
\end{aligned}$$

   Als Rekursion suhcne wir also eine Funktion $n \mapsto f(n)$, sodass

$$\begin{cases} f(n) &= \beta + (1-\alpha-\beta)f(n-1) \\ f(1) &= 1-\alpha \end{cases}.$$

   Als Lösung ergibt sich (Theorie der linearen Rekursionen!):

$$f(n) = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n.$$

Damit ergibt sich

$$P^n = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix} + (1-\alpha-\beta)^n \begin{pmatrix} \frac{\alpha}{\alpha+\beta} & \frac{-\alpha}{\alpha+\beta} \\ \frac{-\beta}{\alpha+\beta} & \frac{\beta}{\alpha+\beta} \end{pmatrix}.$$

Wir beschränken uns auf $\alpha, \beta \in (0,1)$. Da $(1-\alpha-\beta)^n \to 0$ exponentiell schnell, ergibt sich

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \\ \frac{\beta}{\alpha+\beta} & \frac{\alpha}{\alpha+\beta} \end{pmatrix}.$$

Wir stellen fest, dass diese Matrix Rang 1 hat. Ebenso ergibt sich

$$\lim_{n\to\infty} \mu_0 P^n = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right).$$

und dieser ist von μ_0 unabhängig.

Grafik

2.4 Unabhängige Zufallsvariablen

2.4.1 Unabhängige Ereignisse

Definition 2.13 (Unabhängigkeit)**.** Eine Familie von Ereignissen $\{A_k\}_{k\in I}$ heißt unabhängig, falls

$$P(A_{i_1} \cap \ldots \cap A_{i_n}) = \prod_{k=1}^{n} P(A_{i_k}).$$

für alle verschiedenen $i_1, \ldots . i_n \in I$ und für alle $n \leqslant |I|$.

Seien $A, B \in \mathcal{F}$, sodass $\mathbb{P}(A) \neq 0, \mathbb{P}(B) \neq 0$. Dann sind A und B unabhängig gwd

$$\mathbb{P}(A \mid B) = \mathbb{P}(A) \Leftrightarrow \mathbb{P}(B \mid A) = \mathbb{P}(B).$$

Warnung. Falls $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)$ für $i \neq j \in I$, so folgt noch nicht, dass $\{A_i\}_{i\in I}$ unabhängig ist!

Beispiel. Betrachte 3 faire Münzen, also

$$\Omega = \{0,1\}^3 f \{(\omega_1, \omega_2, \omega_3) \mid \omega_i \in \{0,1\} \text{ für } i = 1,2,3,\}.$$

wobei

$$\omega_i = \begin{cases} 0 & \text{Münze } i \text{ ist Kopf} \\ 1 & \text{Münze } i \text{ ist Zahl} \end{cases}.$$

Wir setzen $\mathcal{F} = \mathcal{P}(\Omega)$ und $\mathbb{P}$ als die Gleichverteilung und betrachten die Ereignisse

$$A_1 = \{\omega_1 = \omega_2\} \qquad A_2 = \{\omega_2 = \omega_3\} \qquad A_3 = \{\omega_3 = \omega_1\}.$$

und erhalten unmittelbar

$$\mathbb{P}(A_i) = \frac{4}{8} = \frac{1}{2} \qquad \mathbb{P}(A_i \cap A_j) = \frac{2}{8} = \frac{1}{4} = \mathbb{P}(A_i) \cdot \mathbb{P}(A_j)\ \forall i \neq j.$$

allerdings ist auch

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{2}{8} \neq \mathbb{P}(A_1)\mathbb{P}(A_2)\mathbb{P}(A_3).$$

also sind die Ereignisse nicht unabhängig, allerdings paarweise unabhängig. Dies liegt daran, dass $A_1 \wedge A_2 \Rightarrow A_3$ etc.

2.4.2 Unabhängige Zufallsvariablen

Skizze

Definition 2.14 (Gemeinsame Verteilung)**.** Seien $X_k : \Omega \to \mathcal{S}_k$ für $1 \neq k \leqslant n$ diskeret Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Die Verteilung $\mu_{X_1,\ldots,X_n}$ der Zufallsvektoren $(X_1, \ldots, X_n)$ mit Massenfunktion

$$p_{X_1,\ldots,X_n}(a_1, \ldots, a_n) := \mathbb{P}(X_1 = a_1, \ldots, X_n = a_n).$$

heißt die **gemeinsame Verteilung** von $X_1, \ldots, X_n$.

Definition 2.15 (Unabhängigkeit)**.** Die diskreten Zufallsvariablen $X_1, \ldots, X_n$ heißen **unabhängig**, falls

$$\mathbb{P}(X_1 = a_1, \ldots, X_n = a_n) = \prod_{k=1}^{n} \mathbb{P}(X_k = a_k).$$

für alle $a_1 \in \mathcal{S}_1 \ldots, a_n \in \mathcal{S}_n$.

Satz 2.16. Die folgenden Aussagen sind äquivalent

a) Die Zufallsvariablen $X_1, \ldots, X_n$ sind unabhängig

b) $p_{X_1,\ldots,X_n}(a_1, \ldots, a_n) = \prod_{k=1}^{n} p_{X_k}(a_k)$ für alle $a_k \in S_k, 1 \leqslant k \leqslant n$.

c) $\mu_{X_1,\ldots,X_n} = \mu_{X_1} \otimes \ldots \otimes \mu_{X_n}$ (Produktmass)

d) Die Ereignisse $\{X_1 \in A_1\}, \ldots, \{X_k \in A_k\}$ sind unabhängig für alle $A_1 \subseteq \mathcal{S}_1, \ldots, A_n \subseteq \mathcal{S}_n$

e) Die Ereignisse $\{X_1 = a_1\}, \ldots, \{X_n = a_n\}$ sind unabhängig für $a_1 \in \mathcal{S}_1, \ldots, a_n \in \mathcal{S}_n$

Beweis.

- a $\Leftrightarrow$ b folgt sofort aus Definition 2.14.
- (b) $\Leftrightarrow$ c folgt aus Definition 2.6, da $\mathbb{P} = \mathbb{P}_1 \otimes \mathbb{P}_2$ genau bedeutet, dass für $A_1 \subseteq \Omega_1, A_2 \subseteq \Omega_2$

  $$\mathbb{P}(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2).$$

  gilt
- c $\Rightarrow$ d Seien $1 \leqslant i_1 < i_2 < \ldots < i_k \leqslant n$ und $A_{i_k \subseteq \mathcal{S}_{i_k}}$ für $1 \leqslant k \leqslant n$. Für $i \notin \{i_1, \ldots, i_k\}$ setzen wir $A_i =: \mathcal{S}_i$, dann erhalten wir

$$\begin{aligned}\mathbb{P}(X_{i_1} \in A_{i_1}, \dots, X_{i_k} \in A_{i_k}) &= \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n) \\ &= \mu_{X_{i_1},\dots,X_{i_k}}(A_1 \times \dots \times A_n) \\ &\overset{\text{Satz 2.5}}{=} \prod_{k=1}^{n} \mu_{X_k}(A_k) = \prod_{k=1}^{m} \mathbb{P}(X_{i_k} \in A_{i_k})\end{aligned}$$

- (d) $\Rightarrow$ (e) Wähle $A_i = \{a_/\}$, dann sind wir sofort fertig.
- (e) $\Rightarrow$ (a). Das folgt sofort aus Definition 2.13.

□

Gleichung nochmal durchgehen

2.4.3 Die Kovarianz

Seien $X_1, \dots, X_n$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Wir wissen bereits, dass

$$\mathbb{E}(X_1 + \dots + X_n) = \sum_{k=1}^{n} \mathbb{E}(X_k).$$

nach Linearität des Erwartungswerts.

Frage. Was ist allerdings $\operatorname{Var}(X_1 + \dots + X_n)$?

Definition 2.17 (Kovarinz)**.** Seien X, Y zwei Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann ist

$$\operatorname{Cov}(X, Y) := \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y).$$

die **Kovarianz** von X und Y.

Bemerkung. Für $X = Y$ ist $\operatorname{Cov}(X, X) = \operatorname{Var}(X)$.

Lemma 2.18. Seien X_1, X_2 zwei unabhängige Zufallsvariablen. Dann ist

$$\operatorname{Cov}(X_1, X_2) = 0.$$

Warnung. Die Umkehrung des Lemmas gilt nicht!

Beweis von Lemma 2.18. Wir erhalten

$$\begin{aligned}\mathbb{E}(X_1 \cdot X_2) &= \sum_{\substack{a_1 \in \mathcal{S}_1 \\ a_2 \in \mathcal{S}_2}} \mathbb{P}(X_1 = a_1, X_2 = a_2) a_1 a_2 \\ &= \sum_{\substack{a_1 \in \mathcal{S}_1 \\ a_2 \in \mathcal{S}_2}} \mathbb{P}(X_1 = a_1)\mathbb{P}(X_2 = a_2) a_1 a_2 \\ &= \sum_{a_1 \in \mathcal{S}_1} a\mathbb{P}(X_1 = a_1) \underbrace{\sum_{a_2 \in \mathcal{S}_2} a_2 \mathbb{P}(X_2 = a_2)}_{=\mathbb{E}(X_2)} \\ &= \mathbb{E}(X_1) \sum_{a_1 \in \mathcal{S}_1} a_1 \mathbb{P}(X_1 = a_1) \\ &= \mathbb{E}(X_2)\mathbb{E}(X_1)\end{aligned}$$

□

Definition 2.19 (Unkorreliert)**.** Zwei Zufallsvariablen X_1, X_2 sind **unkorreliert**, falls $\mathrm{Cov}(X_1, X_2) = 0$.

Warnung. Wir haben also Unabhängigkeit $\Rightarrow$ Unkorreliertheit, aber nicht die Umkehrung.

Lemma 2.20. Seien $X_1, \ldots, X_n$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann ist

$$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_{k=1}^{n} \mathrm{Var}(X_k) + 2 \sum_{1 \leqslant i < j \leqslant n} \mathrm{Cov}(X_i, X_j).$$

Korollar†. Für unabhängige (oder auch unkorrelierte) Zufallsvariablen $X_1, \ldots, X_n$ ist

$$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_{k=1}^{n} \mathrm{Var}(X_k).$$

Beweis.* Einsetzen in Lemma 2.20 und verwenden von $\mathrm{Cov}(X_i, X_j) = 0$ für $i \neq j$ liefert sofort das Ergebnis. □

Beispiel. Seien $X_1, \ldots, X_n \sim \mathrm{Ber}(p)$ unabhängige Zufallsvariablen. Wir wissen bereits, dass $X = X_1 + \ldots + X_n \sim \mathrm{Bin}(n, p)$ binomialverteilt ist. Also ergibt sich auch für die Varianz:

$$\mathrm{Var}(X) = n \cdot \mathrm{Var}(X_1) = np(1-p).$$

- ohne eine lange Rechnung machen zu müssen, wie wir es bisher taten.

Beweis von Lemma 2.20.

$$\begin{aligned}
\mathrm{Var}(X_1 + \ldots + X_n) &= \mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)^2\right) - \left(\sum_{i=1^n} \mathbb{E}(X_i)\right)^2 \\
&= \mathbb{E}\left(\sum_{k,l=1}^{n} X_k \cdot X_l\right) - \sum_{k,l=1}^{n} \mathbb{E}(X_k)\mathbb{E}(X_l) \\
&= \sum_{k=1}^{n} \left(\mathbb{E}(X_k^2) - (\mathbb{E}(X_k))^2\right) + 2 \sum_{1 \leqslant k < l \leqslant n} (\mathbb{E}(X_k \cdot X_l) - \mathbb{E}(X_k) \cdot \mathbb{E}(X_l)) \\
&= \sum_{k=1}^{n} \mathrm{Var}(X_k) + 2 \sum_{k=1}^{n} \mathrm{Cov}(X_k, X_l)
\end{aligned}$$

□

Frage. Wenn $X_1, \ldots, X_n$ Zufallsvariablen sind, kennen wir bereits ihren Erwartungswert als ungefähre Näherung für ihre Summe. Genauer wollen wir aber wissen, was

$$\mathbb{P}(|(X_1 + \ldots + X_n) - (\mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n))| \geqslant \varepsilon).$$

ist, für gegebene ε (d.h. wie hoch ist die Wahrscheinlichkeit, dass das tatsächliche Ergebnis unseren Zufallsexperminets um mindestens ε vom Erwartungswert abweicht?). Insbesondere interessieren wir uns darüber, wann dieser Wert klein ist (dann ist der Erwartungswert 'zuverlässiger')

Vorlesung 11
Mo 17 Mai 2021 10:15

2.4.4 Abschätzung von Abweichungen

Seien $X_1, X_2, \ldots, X_n \sim \text{Ber}(p)$ unabhängige Zufallsvariablen. wobei $p \neq 0, 1$. Dann ist $\mathcal{S}_n := X_1 + \ldots + X_n \sim \text{Bin}(n, p)$. Bekannt ist für uns bereits

$$\mathbb{E}\left(\frac{\mathcal{S}_n}{n}\right) = p, \text{Var}\left(\frac{\mathcal{S}_n}{n}\right) = \frac{1}{n^2}\text{Var}(\mathcal{S}_n) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}.$$

Binomialverteilung für große n

Frage. Was ist $\mathbb{P}\left(\left|\frac{\mathcal{S}_n}{n} - p\right| \geqslant \varepsilon\right)$, bzw ist das $\leqslant$? für $n \gg 1$.

Satz 2.21 (Ungleichung von Tchebishev)**.** Sei X eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann gilt $\forall \varepsilon > 0$:

$$\mathbb{P}\left(|X - \mathbb{E}(X)| \geqslant \varepsilon\right) \leqslant \frac{\text{Var}(X)}{\varepsilon^2}.$$

Beweis. Wir stellen fest, dass beide Seiten unabhängig vom Milltwert sind, also können wir $\mathbb{E}(X) = 0$ voraussetzen, indem wir $Y := X - \mathbb{E}(X)$ als Zufallsvariable betrachten. Wir überlegen uns nun:

$$\begin{aligned}
\mathbb{P}(|X| \geqslant \varepsilon) &= \mathbb{E}\left(\mathbb{1}_{|X| \geqslant \varepsilon}\right) \\
&\underset{(1)}{\leqslant} \mathbb{E}\left(\mathbb{1}_{|X| \geqslant \varepsilon}\frac{X^2}{\varepsilon^2}\right) \\
&\leqslant \frac{\mathbb{E}(X^2)}{\varepsilon^2} \\
&= \frac{\text{Var}(X)}{\varepsilon^2}
\end{aligned}$$

Man beachte, dass wir hier bei (1) einfach benutzen, dass $1 \leqslant \frac{X^2}{\varepsilon^2}$, denn $|X| \geqslant \varepsilon$. □

Wir lernen nun eine Verallgemeinerung kennen:

Lemma 2.22. Sei X enie Zufallsvariable, $f : \mathbb{R} \to \mathbb{R}_+$ eine monoton wachsende Zufallsvariable. Dann ist

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}(f(X))}{f(a)} \qquad \forall a \in \mathbb{R}.$$

Beweis. Völlig analog:

$$\begin{aligned}\mathbb{P}(X \geqslant a) &= \mathbb{E}(\mathbb{1}_{X \geqslant a}) \\ &\leqslant \mathbb{E}\left(\mathbb{1}_{X \geqslant a} \frac{f(X)}{f(a)}\right) \\ &\leqslant \frac{1}{f(a)} \mathbb{E}(f(X))\end{aligned}$$

□

Beispiel. Wir erhalten zum Beispiel die **Markov-Ungleichung**:

$$\mathbb{P}(|X| \geqslant \varepsilon) \leqslant \frac{\mathbb{E}(|X|)}{\varepsilon}.$$

Korollar 2.23. Sei X eine Zufallsvariable auf $(\Omega, \mathcal{F}, \mathbb{P})$. Dann ist

$$\mathbb{P}(X \geqslant a) \leqslant \inf_{\lambda \geqslant 0} e^{-\lambda a} \mathbb{E}(e^{\lambda x}).$$

Beweis. Verwende das Lemma mit $f(x) = e^{\lambda x}$ für alle $\lambda \geqslant 0$, dann sind wir schon fertig. □

Satz 2.24. Betrachte wieder das Setting zu Beginn des Kapitels. Für alle $\varepsilon > 0$ ist

$$\mathbb{P}\left(\frac{\mathcal{S}_n}{n} \geqslant p + \varepsilon\right) \leqslant e^{-c\varepsilon^2 n}.$$

und

$$\mathbb{P}\left(\frac{\mathcal{S}_n}{n} \leqslant p - \varepsilon\right) \leqslant e^{-c\varepsilon^2 n}.$$

für ein $0 < c = c(p)$, das nur von p abhängt. Also ist

$$\mathbb{P}\left(\left|\frac{\mathcal{S}_n}{n} - p\right| \geqslant \varepsilon\right) \leqslant 2 \cdot e^{-c\varepsilon^2 n}.$$

Bemerkung. Man kann $c = 2$ zeigen.

Die Abschätzung konvergiert gegen 0 exponentiell schnell in n für $n \to \infty$. Wäs würden wir mit Tchebishev erhalten?

$$\mathbb{P}\left(\left|\frac{\mathcal{S}_n}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{1}{\varepsilon^2} \operatorname{Var}\left(\frac{\mathcal{S}_n}{n}\right) = \frac{p(1-p)}{n \cdot \varepsilon^2}.$$

Das geht auch gegen 0, aber eben nicht exponentiell (sondern nur quadratisch / polynomiell), also nicht wirklich schnell.

Bemerkung. Es gilt sogar

$$\mathbb{P}\left(\left|\frac{\mathcal{S}_n}{n} - p\right| \geqslant \frac{X}{\sqrt{n}}\right) \to 2 \cdot \frac{1}{2\pi\gamma^2} \int_x^\infty e^{-y^2/(2\sigma^2)}\, \mathrm{dy}\,.$$

aber das soll nicht Teil dieser Vorlesung sein.

Beweis. Stellen wir zunächst fest, dass wir mit $X := \frac{\mathcal{S}_n}{n} - p$

$$\begin{aligned}
\mathbb{P}\left(\frac{\mathcal{S}_n}{n} \geqslant p + \varepsilon\right) &= \mathbb{P}(X \geqslant \varepsilon) \\
&\leqslant \inf_{\lambda > 0} e^{-\lambda\varepsilon} \mathbb{E}\left(e^{\lambda\left(\frac{\mathcal{S}_n}{n} - p\right)}\right) \\
&\overset{\lambda := \mu n}{=} \inf_{\mu > 0} e^{-\mu n(p+\varepsilon)} \underbrace{\mathbb{E}(e^{\mu \mathcal{S}_n})}_{= \psi(e^\mu)}
\end{aligned}$$

Setzen wir nun $\psi(z) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} z^k = (p \cdot z + 1 - p)^n$, so erhalten wir weiter:

$$\mathbb{P}\left(\frac{\mathcal{S}_n}{n} \geqslant p + \varepsilon\right) = \inf_{\mu > 0} e^{-n[\underbrace{\mu(p+\varepsilon) - \ln[pe^\mu + 1 - p]}_{:= I(\mu, p)}]}$$

Wir untersuchen nun $I(\mu, p)$ weiter:

- $I(0, p) = 0$
- $\frac{d}{d\mu} I(\mu, p) = p + \varepsilon - \frac{pe^\mu}{1 - p + pe^\mu}$ und diese ist 0 genau dann, wenn $e^\mu = \frac{(1-p)(p+\varepsilon)}{p(1-p-\varepsilon)} =: e^{\mu^*}$
- Es ist

$$\begin{aligned}
I(\mu^*, p) &= (p+\varepsilon) \ln\left(\frac{p+\varepsilon}{p}\right) + (1 - p - \varepsilon) \ln\left(\frac{1-p-\varepsilon}{}\right) \\
&\overset{\text{Taylor}}{=} \frac{\varepsilon^2}{2p(1-p)} + O(\varepsilon^3)
\end{aligned}$$

Also erhalten wir für kleine ε und $p \in (0, 1)$, dass

$$I(\mu^*, p) \geqslant 2\varepsilon^2 + O(\varepsilon^3) \geqslant \varepsilon^2.$$

und damit

$$\mathbb{P}\left(\frac{\mathcal{S}_n}{n} \geqslant p + \varepsilon\right) \leqslant e^{-n\varepsilon^2}.$$

Setze nun $X = -\frac{\mathcal{S}_n}{n} + p$, dann erhalten wir direkt die zweite Abschätzung, die ir zeigen wollten.

□

Details des Beweises aufschreiben.

Bemerkung. Wir hätten das ganze auch so schreiben können:

$$\begin{aligned}
\mathbb{E}(e^{\mu \mathcal{S}_n}) &= \mathbb{E}(e^{\mu X_1} \cdot \ldots \cdot e^{\mu X_n}) \\
&\overset{zz.}{=} \prod_{k=1}^n \mathbb{E}(e^{\mu X_k})
\end{aligned}$$

Lemma 2.25. Seien $X_1 : \Omega \to \mathcal{S}_1$, $X_2 \colon \Omega \to \mathcal{S}_2$ zwei Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$. Falls X_1 und X_2 unabhängig sind, so sind für alle $f_1 \colon \mathcal{S} \to \mathbb{R}$ und $f_2 \colon \mathcal{S}2 \to \mathbb{R}$ auch $f_1(X_1)$ und $f_2(X_2)$ zwei unabhängige Zufallsvariablen.

Beweis.

$$\begin{aligned}
\mathbb{P}(f_1(X_1) = a_1, f_2(X_2) = a_2) &= \mathbb{P}(X_1 \in f_1^{-1}(a_1), X_2 \in f_2^{-1}(a_2)) \\
&\overset{X_1, X_2 \text{ unabhängig}}{=} \mathbb{P}(X_1 \in f_1^{-1}(a_1)) \cdot \mathbb{P}(X_2 \in f_2^{-1}(a_2)) \\
&= \mathbb{P}(f_1(X_1) = a_1) \cdot \mathbb{P}(f_2(X_2) = a_2)
\end{aligned}$$

□

2.4.5 Das schwache Gesetz der großen Zahlen

Seien X_1,X_2,... Ergebnisse von Experimenten mit $X_k \sim \mu$ für jedes k. Setze

$$M_n(\omega) := \frac{1}{n} \sum_{k=1}^{n} X_k(\omega).$$

als den empirischen Mittelwert.

Frage. Unter welchen Bedingungen gilt für große n, dass $M_n - \mathbb{E}(M_n)$ klein ist?

Extremfälle:

a Sei $X_1 = X_2 = X_3 = \ldots$ für alle ω. Dann ist $M_n - \mathbb{E}(M_n) = X_1 - \mathbb{E}(X_1)$, und das geht nicht gegen 0.

b Falls die Zufallsvariablen paarweise unabhängig (aber gleich verteilt) sind, so erhalten wir

$$\mathbb{P}\left(|M_n - \mathbb{E}(M_n)| \geqslant \varepsilon\right) \leqslant \frac{\operatorname{Var}(X_1)}{n\varepsilon^2} \to 0.$$

Ein allgemeineres Resultat bietet uns:

Satz 2.26 (Schwaches Gesetz der großen Zahlen)**.** Seien $X_1, X_2, \ldots$ Zufallsvariablen auf $(\Omega, \mathcal{F}, \mathbb{P})$ mit $\mathbb{E}(X_k^2) < \infty$ für alle k, sodass

a $\operatorname{Cov}(X_k, X_l) = 0$ für $k \neq l$

b $\nu := \sup_{k \geqslant 1} \operatorname{Var}(X_k) < \infty$

Sei $M_n = X_1 + \ldots + X_n$, so ist für jedes $\varepsilon > 0$:

$$\mathbb{P}\left(\left| \frac{\mathcal{M}_n}{n} - \frac{\mathbb{E}(\mathcal{M}_n)}{n} \right| \geqslant \varepsilon \right) \leqslant \frac{\nu}{\varepsilon^2 n} \to 0.$$

Notation†. Wir schreiben auch $X_k \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ für $\mathbb{E}(X_k)^2 < \infty$.

Beweis. Mit Tchebishev erhalten wir, dass

$$\begin{aligned}
\text{LHS} &\leqslant \frac{\operatorname{Var}(M_n)}{\varepsilon^2 n^2} \\
&\overset{lemma20}{=} \frac{\operatorname{Var}(X_1) + \ldots + \operatorname{Var}(X_n)}{\varepsilon^2 n^2} \\
&\leqslant \frac{\nu}{e^2 n}
\end{aligned}$$

□

Namen und label für Sätze hinzufügen

Vorlesung 12
Mo 31 Mai 2021 10:16

3 Simulationsverfahren und Monte Carlo Methode

3.1 Simulation von Zufallsvariablen

Definition 3.1. a U ist eine reellwertige Zufallsvariable, falls

$$\{\omega \in \Omega \mid U(\omega) \leqslant x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}.$$

b $U : \Omega \to [0,1]$ ist gleichverteilt auf $[0,1]$, falls

$$\mathbb{P}(U \leqslant x) = x \qquad \forall x \in [0,1].$$

Eine Familie von rellwertigen Zufallsvariablen $\{U_k\}_{k \in I}$ heißen unabhängig, falls $\forall x_1 x_2, \ldots, \in \mathbb{R} : \{U_k \leqslant x_k\}_{k \in I}$ unabhängig sind.

Notation. Wir schreiben hierfür $U \sim U([0,1])$ oder auch $U \sim \mathrm{Unif}([0,1])$

- Sei $\mathcal{S} = \{a_1, a_2, \ldots\}$ ein diskreter Zustandsraum (d.h. abzählbar) und μ eine Wahrscheinlichkeitsverteilung auf $\mathcal{S}$ mit $p_k = \mu(a_k)$, also $\sum_{k=1}^{\infty} p_k = 1$. Setze nun $s_0 := 0$, $s_n := \sum_{k=1}^{n} p_k$ für $n \geqslant 1$.

Lemma 3.2. Sei $U \sim \mathrm{Unif}([0,1])$ uniform verteilt und $X(\omega) := a_n$ falls $U(\omega) \in (s_{n-1}, s_n]$. Dann ist $X \sim \mu$.

Beweis. Für $n \geqslant 1$ ist

$$\begin{aligned}
\mathbb{P}(X = a_n) &= \mathbb{P}(s_{n-1} < U(\omega) \leqslant s_n) \\
&= \mathbb{P}(U(\omega \leqslant s_n)) - \mathbb{P}(U(\omega \leqslant s_{n-1})) \\
&= s_n - s_{n-1} \\
&= p_n
\end{aligned}$$

□

Mündliche Anmerkung. Wir können mit diesem Lemma theoretisch schon für beliebige diskrete Verteilungen einen enstprechenden Algorithmus schreiben, das ist aber unter Umständen äußerst unpraktikabel, wenn $\mathcal{S}$ sehr groß ist oder μ keine tolle Form hat, wie wir gleich sehen werden:

Algorithmus 3.3 : Simulation von μ

Eingabe : $p_1, p_2, \ldots$
Ausgabe : Eine Zufallsvariable X mit $X \sim \mu$

$n = 1$
$s = p_1$
erzeuge $u \sim \mathrm{Unif}([0,1])$
while $n > s$ **do**
  $n = n + 1$
  $s = s + p_n$
return a_n

Frage. Wie viele Schritte brauchen wir?

$$\mathbb{E}(\#\text{'Schritte'}) = \sum_{n \geqslant 1} n \cdot p_n.$$

Mündliche Anmerkung. Abgesehen von potentiell sehr langer Rechendauer, kommen hier auch noch numerische Ungenauigkeiten / Probleme vor.

3.2 Acceptance-Rejection-Verfahren

- Sei μ eine Wahrscheinlichkeitsverteilung mit Massenfunktion p und ν eine Wahrscheinlichkeitsverteilung mit Massenfunktion q, wobei wir ν simulieren wollen.
- Nehmen wir an, dass $\exists c \in [1, \infty)$, sodass

$$\begin{aligned} & p(x) \leqslant c \cdot q(x) && \forall x \in \mathcal{S} \\ \Leftrightarrow \quad & 0 \leqslant \frac{p(x)}{c \cdot q(x)} && \forall x \in \mathcal{S} \end{aligned}$$

Algorithmus 3.4 :

Eingabe : $p(x), q(x)$ für $x \in \mathcal{S}$, Konstante c
Ausgabe : Zufallsvariable X mit $X \sim \mu$

repeat:
erzeuge $x \sim \nu$
erzeuge $u \sim \mathrm{Unif}([0,1])$
until ($\frac{p(x)}{c \cdot q(x)} \geqslant u$)
return x.

Wir nehmen also den Vorschlag $X = x$ mit der Wahrscheinlichkeit $\frac{p(x)}{c \cdot q(x)}$ an.

Mündliche Anmerkung. Die Erzeugung von $x \sim \nu$ ist im Algorithmus nicht genauer beschrieben. Der Algorithmus ist also nur dann (sinnvoll) anwendbar, wenn wir $x \sim \nu$ mit einem anderen Algorithmus sinnvoll schnell erzeugen können.
Wir gehen ebenfalls davon aus, dass wir $u \sim \mathrm{Unif}([0,1])$ mit einem Pseudo-Zufallsgenerator mit dem Computer erzeugen können.

Seien $X_1, X_2, \ldots \sim \nu$ die Vorschläge, die wir erhalten, und seien $U_1, U_2, \ldots \sim \mathrm{Unif}([0,1])$ die uniformen Zufallsvariablen. Sei

$$T := \min \left\{ n \geqslant 1 \mid \frac{p(X_n)}{c \cdot q(X_n)} \geqslant U_n \right\}.$$

der erste Zeitpunkt, bei dem die entsprechende Ungleichung gilt, d.h. wenn wir unsere Schleife abbrechen. Also ist $X_T(\omega) := X_{T(\omega)}(\omega)$ der Output unseres Algorithmus.

Satz 3.5. a $X_T \sim \mu$, dh. der Algorithmus ist korrekt.
b $T - 1 \sim \mathrm{Geo}(\frac{1}{c})$, also $\mathbb{E}(T) = c$.

Beweis. Die Ereignisse

$$A_n = \left\{ \frac{p(X_n)}{c \cdot q(X_n)} \geqslant U_n \right\} \qquad n = 1, 2, \ldots.$$

sind unabhängig, also ist

$$\begin{aligned} \mathbb{P}(T = n) &= \mathbb{P}(A_1^c \cap \ldots \cap A_{n-1}^c \cap A_n) \\ &\overset{\text{unabhängig}}{=} \mathbb{P}(A_1) \cdot (1 - \mathbb{P}(A_1))^{n-1} \end{aligned}$$

Nun ist

$$\begin{aligned} \mathbb{P}(A_1) &= \sum_{a \in \mathcal{S}} \mathbb{P}(\left\{ U_1 \leqslant \frac{p(a)}{c \cdot q(a)} \right\} \cap \{X_1 = a\}) \\ &= \sum_{a \in \mathcal{S}} \frac{p(a)}{c \cdot q(a)} \cdot q(a) \\ &= \frac{1}{c} \end{aligned}$$

Also folgt bereits Teil b.
Es ist

$$\begin{aligned} \mathbb{P}(X_T) = a) &= \sum_{n \geqslant 1} \mathbb{P}(X_T = a, T = n) \\ &= \sum_{n \geqslant 1} \mathbb{P}(A_1^c \cap \ldots \cap A_n \cap \{X_n = a\}) \\ &= \sum_{n \geqslant 1} \left(1 - \frac{1}{c}\right)^{n-1} \frac{p(a)}{c} \\ &= p(a) \end{aligned}$$

□

Beispiel. Sei $Y \sim \nu = \mathrm{Geo}(q)$ geometrische verteilt.

- In Kapitel 1.7.4 haben wir gesehen, dass wir $u \sim \mathrm{Unif}([0,1])$ erzeugen, und dann

$$Y = \left\lfloor \frac{\ln(1-u)}{\ln(q)} \right\rfloor.$$

setzen können, um Y zu simulieren, und dann eben $\mathbb{P}(Y = k) = (1-q)q^k$ für $k = 0, 1, \ldots$.

- Sei $X \sim \mu = \mathrm{Poi}(\lambda)$, d.h.

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

- Sei

$$\mathcal{R}(k) := \frac{e^{-\lambda}\lambda^k}{k!} \cdot \frac{1}{(1-q)q^k}.$$

- Wähle $q = \frac{\lambda}{\lambda+1}$, denn dann ist $\lambda = \frac{q}{1-q}$, und die beiden Verteilungen erhalten den gleichen Mittelwert, also ist

$$\mathcal{R}(k) = \frac{e^{-\lambda}(\lambda+1)^{k+1}}{k!}.$$

Wegen

$$\mathcal{R}(k+1) - \mathcal{R}(k) = \mathcal{R}(k)\left(\frac{1+\lambda}{1+k} - 1\right)\begin{cases} > 0 & k < \lambda \\ < 0 & k > \lambda \end{cases}.$$

Veranschaulichung des maximalen k

erhalten wir

$$\max_{k \geqslant 0} \mathcal{R}(k) \leqslant \frac{e^{-\lambda}(1+\lambda)^{\lfloor\lambda\rfloor+1}}{\lfloor\lambda\rfloor!} =: c.$$

Wir können nun also $\mathrm{Poi}(\lambda)$ mit Algorithmus 3 simulieren, weil wir ein entsprechendes c gefunden haben.

3.3 Monte-Carlo-Verfahren

Ziel. es ist

$$\mathbb{P}_\mu(f) := \sum_{x \in \mathcal{S}} f(x)\mu(x). \tag{1}$$

für $|S| \gg 1$ zu berechnen.
Beweisstrategie. Man simuliert eine Folge von Zufallsvariablen oder eine Markovkette mit $\mu = \mu P$.

Beispiel. Sei $\mathcal{S} = \{-1, +1\}^{100}$, also $|\mathcal{S}| = 2^{100} \approx 10^{30}$. Das könnte z.B. dann der Fall sein, wenn wir bei einem 10×10-Gitter in jedem Punkt eine Markierung haben oder nicht.

Mündliche Anmerkung. Wenn wir nun den Erwartungswert einer Funktion über diesem Zustandsraum berechnen wollen, so ergeben sich Probleme, da wir einerseits absurd viele Summanden haben, aber auch numerische, denn μ ist für jeden einzelnen Zustand verschwindend klein. Unsere Formel (1) stimmt also, wir haben aber keine Chance, diese auszurechnen.

- Sei $\mathcal{S}$ abzählbar und μ eine Wahrscheinlichkeitsverteilung auf $\mathcal{S}$.
- Sei f eine Funktion $\mathcal{S} \to \mathbb{R}$, sodass

$$\mathbb{E}_\mu(f^2) = \sum_{x \in \mathcal{S}} (f(x))^2 \mu(x) < \infty.$$

Wir wollen z.B. $\theta := \mathbb{E}_\mu(f)$ bestimmen. Als Schätzung definieren wir

$$\theta_n := \frac{1}{n} \sum_{k=1}^{n} f(X_k).$$

wobei X_1, X_2,... Zufallsvariablen mit Verteilung μ sind. Falls $X_1, X_2, \ldots$ unabhängig:

Satz 3.6. $\forall \varepsilon > 0$ ist $\mathbb{P}(|\theta_n - \theta| \geqslant \varepsilon) \leqslant$

$$\leqslant \frac{\operatorname{Var}_\mu(f^2)}{n\varepsilon^2} \leqslant \frac{\mathbb{E}_\mu(f^2)}{n\varepsilon^2} \to 0.$$

Beweis. Tchebishev Ungleichung, siehe Gesetz der großen Zahlen. □

Beispiel. 1 Schätzung von $\theta = \int_0^1 f(x)\,\mathrm{d}x$.
- Seien $u_1, u_2, \ldots \sim \operatorname{Unif}([0,1])$, dann setze

$$\theta_n := \frac{1}{n} \sum_{k=1}^{n} f(u_k).$$

und diese ist eine Schätzung des entsprechenden Integrals mit

$$\sqrt{\mathbb{E}(|\theta_n - \theta|^2} = O\left(\frac{1}{n}\right).$$

2 Schätzung einer Wahrscheinlichkeitsverteilung.
- Sei $\mathcal{S}$ ein diskreter Zustandsraum und $B \subseteq \mathcal{S}$. Gesucht ist $p := \mu(B)$ (unbekannt).
- Es ist $\mu(B) = \mathbb{E}_\mu(\mathbb{1}_B)$, also setze

$$p_n := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_B(X_k).$$

wobei $X_1, X_2, \ldots \sim \mu$ (unabhängige) Zufallsvariablen sind und erhalte damit eine Näherung.

Frage. Wie gut ist diese Annäherung?

$$\mathbb{P}(|p_n - p| \geqslant \varepsilon) \leqslant \frac{1}{\varepsilon^2} \operatorname{Var}(p_n) = \frac{1}{n\varepsilon^2} \operatorname{Var}(\mathbb{1}_B) = \frac{p(1-p)}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2}.$$

Gegeben ε wählen wir n also groß genug, sodass $\frac{1}{4n\varepsilon^2} \leqslant$ angegebene Fehlerschranke.

Mündliche Anmerkung. Wir können nun, wenn wir $p := \mu(B)$ bestimmen wollen, also für ein gewünschtes ε ein geeignetes n wählen, sodass wir p approximieren und mit höchstens der gewünschten Wahrscheinlichkeit um mindestens ε von der echten Wahrscheinlichkeit abweichen.

C Stichwortverzeichnis