Kapitel III
Bedingte Wahrscheinlichkeiten und stochastische Unabhängigkeit

In diesem Kapitel kommen wir zu einem der herausragenden Begriffe in der Stochastik, nämlich der stochastischen Unabhängigkeit, die eine Präzisierung der Vorstellung sich gegenseitig nicht beeinflussender zufälliger Phänomene beinhaltet.

10. Bedingte Wahrscheinlichkeiten

Zur Motivation der nachfolgenden Begriffsbildungen beginnen wir mit zwei Beispielen:

10.1. Beispiele. (a) ***(Ziehen mit Zurücklegen)*** Aus einer Urne mit n von $1,...,n$ numerierten Kugeln ($n \geq 2$) wird zweimal zufällig eine Kugel mit Zurücklegen gezogen, beschrieben durch ein Laplace-Experiment über der Menge

$$\Omega \;=\; \{1,...,n\}^2$$

der Mächtigkeit n^2. Sei $X_i(\omega_1, \omega_2) = \omega_i$ für $i = 1, 2$ das Ergebnis der i-ten Ziehung. Dann ist $(X_1, X_2) = \mathrm{id}_\Omega$ also Laplace-verteilt über Ω, und es erscheint anschaulich klar, daß die Kenntnis des ersten Ziehungsergebnisses keinerlei Information über das Ergebnis der zweiten Ziehung liefert.

(b) ***(Ziehen ohne Zurücklegen)*** Modifizieren wir das obige Experiment dahingehend, daß zweimal *ohne* Zurücklegen gezogen wird, so ergibt sich ein Laplace-Experiment über der Menge

$$\hat{\Omega} \;=\; [1,...,n]^2 \;=\; \{(\omega_1, \omega_2) \in \Omega : \omega_1 \neq \omega_2\}$$

der Mächtigkeit $n(n-1)$. Bezeichnen wir diesmal mit $\hat{X}_1, \hat{X}_2$ die beiden Wurfergebnisse, so liefert die Kenntnis des ersten Ziehungsergebnisses sehr wohl Information über das zweite Ziehungsergebnis: Wenn $\hat{X}_1 = i$, muß $\hat{X}_2 \neq i$ gelten.

Mit dem Begriff der "bedingten Wahrscheinlichkeit" wollen wir im folgenden Wahrscheinlichkeiten bei Vorliegen von Vorinformation erfassen.

10.2. Definition. Es seien P eine diskrete W-Verteilung über einer nichtleeren Menge Ω und $A, B \subset \Omega$ Ereignisse mit $P(B) > 0$. Dann heißt

$$P(A|B) \;\stackrel{\text{def}}{=}\; \frac{P(A \cap B)}{P(B)} \tag{10.1}$$

(elementare) bedingte Wahrscheinlichkeit von A gegeben B oder auch *unter B*.

Daß (10.1) in der Tat die "richtige" Definition ist, sieht man wie folgt: Gegeben die Information, daß das Ereignis B eingetreten ist oder eintreten wird, müssen wir den ursprünglichen *Ergebnisraum* Ω *auf* B *einschränken* und die diskrete W-Verteilung P entsprechend *renormieren*, um B nunmehr Gesamtmasse 1 zu geben. Dies geschieht per Division durch $P(B)$, wobei beachtet werde, daß dies die (relative) Massenverteilung innerhalb von B unberührt läßt. Die bedingte Wahrscheinlichkeit von A gegeben B ergibt sich nun als Wahrscheinlichkeit der Einschränkung von A auf B, also $A \cap B$, unter der renormierten W-Verteilung, was genau (10.1) entspricht. Daß hierbei die Renormierung in der Tat wieder zu einer diskreten W-Verteilung führt, zeigt Satz 10.3 weiter unten.

10.1. (Fortsetzung) (a) Für $i, j \in \{1, ..., n\}$ und $A = \{X_2 = j\}, B = \{X_1 = i\}$ erhalten wir

$$P(A|B) \;=\; \frac{P(X_1 = i, X_2 = j)}{P(X_2 = j)} \;=\; \frac{\frac{1}{n^2}}{\frac{1}{n}} \;=\; \frac{1}{n} \;=\; P(X_2 = j).$$

Die Anschauung, daß die Kenntnis von $X_1 = i$ keinerlei Einfluß auf die Wahrscheinlichkeit für das Eintreten von $X_2 = j$ hat, bestätigt sich also.

(b) Hier erhalten wir für $i, j \in \{1, ..., n\}$, $i \neq j$, und $A = \{\hat{X}_2 = j\}, B = \{\hat{X}_1 = i\}$

$$P(A|B) \;=\; \frac{P(\hat{X}_1 = i, \hat{X}_2 = j)}{P(\hat{X}_2 = j)} \;=\; \frac{\frac{1}{n(n-1)}}{\frac{1}{n}} \;=\; \frac{1}{n-1} \;\neq\; P(\hat{X}_2 = j),$$

also wiederum die Bestätigung unserer oben artikulierten Intuition.

10.3. Satz. *Es sei eine diskrete W-Verteilung über* $\Omega \neq \emptyset$ *und* $B \subset \Omega$ *ein Ereignis mit* $P(B) > 0$. *Dann bildet* $P(\cdot|B) : \mathfrak{P}(\Omega) \to [0,1]$, *definiert durch*

$$P(A|B) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(B)}, \tag{10.2}$$

eine normierte σ*-additive Mengenfunktion und folglich eine diskrete W-Verteilung über* Ω.

Beweis: Da $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = 1$, ist $P(\cdot|B)$ normiert. Für eine Folge $(A_n)_{n\geq 1}$ p.d. Teilmengen von Ω gilt außerdem

$$P\left(\sum_{n\geq 1} A_n \middle| B\right) \;=\; \frac{P(\sum_{n\geq 1} A_n \cap B)}{P(B)} \;=\; \sum_{n\geq 1} \frac{P(A_n \cap B)}{P(B)} \;=\; \sum_{n\geq 1} P(A_n|B),$$

was die σ-Additivität von $P(\cdot|B)$ beweist. ◊

Sowohl bei praktischen als auch bei theoretischen Problemen hat man häufig Kenntnis gewisser bedingter Wahrscheinlichkeiten, mit deren Hilfe man andere unbedingte Wahrscheinlichkeiten berechnen möchte. So beschäftigt man sich beispielsweise in der Marktforschung unter anderem damit, das Kaufverhalten bestimmter Bevölkerungsgruppen mittels bedingter

Wahrscheinlichkeiten anzugeben und daraus das Kaufverhalten der Gesamtbevölkerung "zusammenzusetzen". Der folgende Satz zeigt, wie man in solchen Situationen mit bedingten Wahrscheinlichkeiten rechnen kann:

10.4. Satz von der totalen Wahrscheinlichkeit. *Es sei P eine diskrete W-Verteilung über Ω und $(B_i)_{i\in I}$, $I \subset \mathbb{N}$, eine Familie p.d. Ereignisse mit $P(B_i) > 0$ für alle $i \in I$. Dann gilt für jedes Ereignis $A \subset \sum_{i\in I} B_i$*

$$P(A) = \sum_{i\in I} P(A|B_i)P(B_i). \tag{10.3}$$

Im Fall $\sum_{i\in I} B_i = \Omega$ folgt demnach (10.3) für jedes Ereignis A.

Beweis: Unter den gemachten Voraussetzungen gilt

$$\begin{aligned}\sum_{i\in I} P(A|B_i)P(B_i) &= \sum_{i\in I} \frac{P(A\cap B_i)}{P(B_i)} P(B_i)\\ &= \sum_{i\in I} P(A\cap B_i)\\ &= P\left(A \cap \sum_{i\in I} B_i\right)\\ &= P(A).\end{aligned}$$

$\diamondsuit$

Eine weitere äußerst wichtige Formel für bedingte Wahrscheinlichkeiten wollen wir vor dem Hintergrund des folgenden Beispiels angeben und beweisen:

10.5. Beispiel. Es ist bekannt, daß das Risiko, mongoloide Kinder zu gebären, für Frauen ab dem 35 Lebensjahr stark ansteigt. Dabei handelt es sich um einen genetischen Defekt, der als Trisomie 21 bezeichnet wird. Aufgrund dieses erhöhten Risikos bietet man in Deutschland schwangeren Frauen ab 35 bis zur 20. Schwangerschaftswoche die Möglichkeit, mittels eines Tests (Amniozentese, auch als Fruchtwasseruntersuchung bekannt) feststellen zu lassen, ob ihr ungeborenes Kind mongoloid ist oder nicht, um gegebenenfalls einen medizinisch indizierten Schwangerschaftsabbruch vornehmen zu lassen. Nun ist dieser Test nicht unfehlbar, wobei der Irrtum auf zwei verschiedene Weisen eintreten kann:

- Das Testergebnis weist auf Mongoloismus hin, obwohl das Kind gesund ist.
- Der Test fällt negativ aus, obwohl das Kind mongoloid ist.

Erscheint eine werdende Mutter über 35 zur Beratung, so werden ihr i.a. folgende Zahlen zur Untermauerung der Zuverlässigkeit des Testverfahrens gegeben: Wenn tatsächlich Mongoloismus vorliegt, fällt das Ergebnis mit $p\cdot 100\%$-iger Sicherheit positiv aus, im Fall eines gesunden Kindes dagegen mit $q \cdot 100\%$-iger Sicherheit. Darüberhinaus teilt man ihr normalerweise mit, daß unter n Neugeborenen von Müttern ihrer Altersgruppe durchschnittlich 1 Kind mongoloid ist[1)]

[1)] Da uns die echten Zahlen nicht bekannt sind, verwenden wir stattdessen allgemeine Variablen.

Will man die Situation in ein Modell fassen, so bietet sich als Ergebnisraum

$$\Omega \;=\; \{0,1\}^2$$

an, wobei die Interpretation eines $(\omega_1,\omega_2)\in\Omega$ lautet:

$$\begin{array}{lcl} \omega_1 = 1 & \Leftrightarrow & \text{Kind mongoloid} \\ \omega_1 = 0 & \Leftrightarrow & \text{Kind gesund} \\ \omega_2 = 1 & \Leftrightarrow & \text{Testergebnis positiv} \\ \omega_2 = 0 & \Leftrightarrow & \text{Testergebnis negativ} \end{array}$$

Es bezeichnet dann $A=\{(0,1),(1,1)\}$ das Ereignis, daß das Testergebnis positiv ausfällt, und $B=\{(1,0),(1,1)\}$ das Ereignis, daß das Kind monogoloid ist. Die obigen Zuverlässigkeitsangaben lassen sich dann wie folgt übersetzen:

$$P(A|B)=p, \quad P(A|B^c)=q \quad \text{und} \quad P(B)=\frac{1}{n}.$$

Wird nun einer Schwangeren über 35 nach Durchführung des Tests mitgeteilt, das Ergebnis sei positiv, so wird sie sich natürlich fragen, mit welcher Wahrscheinlichkeit ihr Kind bei diesem Befund tatsächlich mongoloid ist, gegeben durch $P(B|A)$. Das Problem besteht hierbei darin, daß die Rollen von A und B als bedingendem und bedingtem Ereignis im Vergleich zu den vorher genannten Größen vertauscht sind. Die im Anschluß angegebene Formel von Bayes zeigt, wie sie dennoch aus letzteren berechnet werden können.

10.6. Satz. (Formel von Bayes) *Es sei P eine diskrete W-Verteilung und $(B_i)_{i\in I}$, $I\subset\mathbb{N}$, eine Familie p.d. Ereignisse mit $P(B_i)>0$ für alle $i\in I$. Dann gilt für jedes $A\subset\sum_{i\in I}B_i$ mit $P(A)>0$ und jedes $k\in I$*

$$P(B_k|A) \;=\; \frac{P(A|B_k)P(B_k)}{\sum_{i\in I}P(A|B_i)P(B_i)}. \tag{10.4}$$

Beweis: Nach Satz 10.4 gilt $P(A)=\sum_{i\in I}P(A|B_i)P(B_i)$. Ferner ist $P(B_k\cap A)=P(B_k|A)P(A)=P(A|B_k)P(B_k)$, so daß

$$P(B_k|A) \;=\; \frac{P(B_k\cap A)}{P(A)} \;=\; \frac{P(A|B_k)P(B_k)}{\sum_{i\in I}P(A|B_i)P(B_i)}$$

für jedes $k\in I$. ♢

10.5. (Fortsetzung) Kehren wir zu unserem Beispiel vor Satz 10.6 zurück. Nach (10.4) gilt dort

$$
\begin{aligned}
P(B|A) &= \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)} \\
&= \frac{p\cdot\frac{1}{n}}{p\cdot\frac{1}{n}+q\cdot\frac{(n-1)}{n}} \qquad (10.5)\\
&= \frac{p}{p+q(n-1)}.
\end{aligned}
$$

Entsprechend ergibt sich für die bedingte Wahrscheinlichkeit, daß im Fall eines negativen Befundes das Kind tatsächlich gesund ist,

$$
\begin{aligned}
P(B^c|A^c) &= \frac{P(A^c|B^c)P(B^c)}{P(A^c|B^c)P(B^c)+P(A^c|B)P(B)} \\
&= \frac{(1-q)\cdot\frac{n-1}{n}}{(1-q)\cdot\frac{n-1}{n}+(1-p)\cdot\frac{1}{n}} \qquad (10.6)\\
&= \frac{(1-q)(n-1)}{(1-q)(n-1)+(1-p)}.
\end{aligned}
$$

Die beschriebene Situation ist typisch für die Medizin: Aufgrund einer Untersuchung muß diagnostiziert werden, ob ein Patient eine bestimmte Krankheit hat oder nicht. Der Arzt kann sich auf zwei Weisen irren:

- Er diagnostiziert die Krankheit bei einem gesunden Patienten.
- Er erkennt bei einem kranken Patienten die Krankheit nicht.

Gegeben dasselbe formale Modell wie zuvor, lautet dann die Interpretation eines $(\omega_1,\omega_2)\in\Omega$:

$$
\begin{array}{lcl}
\omega_1 = 1 & \Leftrightarrow & \text{Patient ist tatsächlich krank}\\
\omega_1 = 0 & \Leftrightarrow & \text{Patient ist gesund}\\
\omega_2 = 1 & \Leftrightarrow & \text{Diagnose lautet "krank"}\\
\omega_2 = 0 & \Leftrightarrow & \text{Diagnose lautet "gesund"}
\end{array}
$$

Es bezeichnet dann $A=\{(0,1),(1,1)\}$ das Ereignis, daß die Diagnose ”krank” gestellt wird, und $B=\{(1,0),(1,1)\}$ das Ereignis, daß der Patient tatsächlich krank ist.

Als drittes Ergebnis über das Rechnen mit bedingten Wahrscheinlichkeiten ist noch der Multiplikationssatz von Bedeutung:

10.7. Multiplikationssatz. *Es seien P eine diskrete W-Verteilung und $A_0,...,A_n$ Ereignisse mit $P(A_0\cap...\cap A_{n-1})>0$. Dann gilt*

$$
P\left(\bigcap_{j=0}^{n} A_j\right) = P(A_0)P(A_1|A_0)P(A_2|A_0\cap A_1)\cdot\ldots\cdot P(A_n|A_0\cap\ldots\cap A_{n-1}). \qquad (10.7)
$$

Beweis: Wegen $\cap_{j=0}^{n-1} A_j \subset \cap_{j=0}^{n-2} A_j \subset ... \subset A_0 \cap A_1 \subset A_0$ gilt

$$0 \; < \; P\left(\bigcap_{j=0}^{n-1} A_j\right) \; \leq \; P\left(\bigcap_{j=0}^{n-2} A_j\right) \; \leq ... \leq \; P(A_0 \cap A_1) \; \leq \; P(A_0).$$

Damit sind alle bedingten Wahrscheinlichkeiten in (10.7) definiert, und es folgt

$$\begin{aligned} & P(A_0)P(A_1|A_0)P(A_2|A_0 \cap A_1) \cdot ... \cdot P(A_n|A_0 \cap ... \cap A_{n-1}) \\ &= \; P(A_0)\frac{P(A_0 \cap A_1)}{P(A_0)} \frac{P(A_2 \cap A_1 \cap A_0)}{P(A_0 \cap A_1)} \cdot ... \cdot \frac{P(\cap_{j=0}^{n} A_j)}{P(\cap_{j=0}^{n-1} A_j)} \\ &= \; P(\cap_{j=0}^{n} A_j), \end{aligned}$$

d.h. die Behauptung. ♢

Hier ist eine typische Anwendung: Sei $(X_0, ..., X_n)$ ein Zufallsvektor und $(k_0, ..., k_n) \in \mathbb{R}^{n+1}$ mit

$$P((X_0, ..., X_{n-1}) = (k_0, ..., k_{n-1})) \; = \; P(X_0 = k_0, ..., X_{n-1} = k_{n-1}) \; > \; 0.$$

Dann erhalten wir vermöge des Multiplikationssatzes die folgende "Pfadregel", sofern wir dort $A_i = \{X_i = k_i\}$ setzen:

$$\begin{aligned} & P(X_0 = k_0, ..., X_n = k_n) \\ &= \; P(X_0 = k_0)P(X_1 = k_1|X_0 = k_0) \cdot ... \cdot P(X_n = k_n|X_{n-1} = k_{n-1}, ..., X_0 = k_0). \end{aligned}$$

Betrachten wir abschließend ein Beispiel:

10.8. Beispiel. Eine Elektrofirma stellt Elektrogeräte mit einer Betriebsspannung von 220 Volt und 110 Volt her. Die Lieferung einer Ladung von 40 Geräten in ein Land, dessen Stromnetz mit 220 Volt betrieben wird, enthält versehentlich 6 Geräte mit einer Betriebsspannung von 110 Volt. An einem bestimmten Tag sollen 5 der 40 Geräte in Betrieb genommen werden. Wie groß ist die Wahrscheinlichkeit, daß alle 5 Geräte eine Betriebsspannung von 220 Volt besitzen?

Bezeichnet A_i das Ereignis, daß das i-te Gerät mit 220 Volt betrieben wird ($i = 1, ..., 5$), so ergibt sich die gesuchte Wahrscheinlichkeit nach dem Multiplikationssatz als

$$\begin{aligned} P(A_1 \cap ... \cap A_5) \; &= \; P(A_1)P(A_2|A_1) \cdot ... \cdot P(A_5|A_1 \cap A_2 \cap A_3 \cap A_4) \\ &= \; \frac{34}{40} \cdot \frac{33}{39} \cdot \frac{32}{38} \cdot \frac{31}{37} \cdot \frac{30}{36} \; \approx \; 0.42. \end{aligned}$$

11. Stochastisch unabhängige Ereignisse

Im allgemeinen werden in einem ZE für Ereignisse A, B mit $P(B) > 0$ die bedingte Wahrscheinlichkeit $P(A|B)$ und die unbedingte Wahrscheinlichkeit $P(A)$ verschieden sein. Der Spezialfall $P(A|B) = P(A)$ ist dennoch von ganz besonderer Bedeutung und besagt, daß das Eintreten von B keinen Einfluß auf das Eintreten von A hat, d.h., daß A und B *unabhängig* voneinander eintreten. Die Beziehung $P(A|B) = P(A)$ ist offenbar äquivalent zu $P(A \cap B) = P(A)P(B)$, wobei diese Beziehung auch im Fall $P(B) = 0$ sinnvoll bleibt.

11.1. Definition. Es seien P eine W-Verteilung und A, B Ereignisse. A und B heißen *stochastisch unabhängig* (unter P), wenn $P(A \cap B) = P(A)P(B)$, und sonst *stochastisch abhängig*.

11.2. Bemerkungen. (a) Die stochastische Unabhängigkeit zweier Ereignisse A und B geht beim Übergang zu einer anderen W-Verteilung P' i.a. verloren: Seien $\Omega = \{0,1\}^2$, $A = \{(1,0),(1,1)\}$, $B = \{(0,1),(1,1)\}$, P die Laplace-Verteilung über Ω und P' gegeben durch $P'(\{(0,0)\}) = P'(\{(1,1)\}) = \frac{1}{2}$, $P'(\{(0,1)\}) = P'(\{(1,0)\}) = 0$. Dann gilt

$$P(A \cap B) \;=\; P(\{(1,1)\}) \;=\; \frac{1}{4} \;=\; \frac{1}{2} \cdot \frac{1}{2} \;=\; P(A)P(B),$$

während

$$P'(A \cap B) \;=\; P'(\{(1,1)\}) \;=\; \frac{1}{2} \;\neq\; \frac{1}{2} \cdot \frac{1}{2} \;=\; P'(A)P'(B).$$

(b) Falls $P(B) \in \{0,1\}$, folgt $P(A \cap B) = P(A)P(B)$ für alle Ereignisse A, d.h. B ist von jedem Ereignis A inklusive sich selbst unabhängig. Insbesondere sind das sichere Ereignis Ω und das unmögliche Ereignis $\emptyset$ unabhängig von jedem Ereignis A.

Definition 11.1 legt nahe, den Begriff der stochastischen Unabhängigkeit auf mehr als zwei Ereignisse zu verallgemeinern, wobei die Art und Weise allerdings nicht ganz kanonisch ist. Um ein Gefühl für die Problematik zu bekommen, betrachten wir zunächst zwei Beispiele:

11.3. Beispiele. (a) Wird ein fairer Würfel zweimal geworfen, so liegt ein Laplace-Experiment über $\Omega = \{1,2,3,4,5,6\}^2$ vor. Bezeichnen

$$A_1 \overset{\text{def}}{=} \{2,4,6\} \times \{1,2,3,4,5,6\} \quad \text{und} \quad A_2 \overset{\text{def}}{=} \{1,2,3,4,5,6\} \times \{2,4,6\}$$

die Ereignisse, im ersten bzw. zweiten Wurf eine gerade Augenzahl zu würfeln, so gilt

$$P(A_1 \cap A_2) \;=\; P(A_1)P(A_2) \;=\; \frac{1}{4},$$

also die stochastische Unabhängigkeit von A_1 und A_2. Sei nun ferner $A_3 \stackrel{\text{def}}{=} \{(i,j) \in \Omega : i+j \text{ gerade}\}$ das Ereignis, daß die Augensumme der beiden Würfe gerade ist. Dann folgt

$$P(A_3|A_1) \;=\; P(A_3|A_2) \;=\; P(A_3) \;=\; \frac{1}{2},$$

d.h., neben A_1, A_2 sind auch A_1, A_3 und A_2, A_3 unabhängig. Es gilt aber *nicht* $P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$, denn

$$P(A_1 \cap A_2 \cap A_3) \;=\; P(A_1 \cap A_2) \;=\; \frac{1}{4} \;\neq\; \frac{1}{8} \;=\; P(A_1)P(A_2)P(A_3).$$

(b) Sei (Ω, p) das durch $\Omega = \{1,2,3\}$ und $p(1) = p(2) = \frac{1}{2}$, $p(3) = 0$ definierte diskrete ZE. Sei ferner $A_i = \{i\}$ für $i = 1,2,3$. Dann gilt

$$P(A_1 \cap A_2 \cap A_3) \;=\; P(\emptyset) \;=\; 0 \;=\; P(A_1)P(A_2)P(A_3),$$

während

$$P(A_1 \cap A_2) \;=\; 0 \;\neq\; \frac{1}{4} \;=\; P(A_1)P(A_2).$$

Die beiden Beispiele lehren zweierlei:

- Sind von n Ereignissen $A_1, ..., A_n$ je zwei stochastisch unabhängig, so sichert dies noch nicht
$$P(A_1 \cap ... \cap A_n) \;=\; P(A_1) \cdot ... \cdot P(A_n).$$
- Gilt für n Ereignisse $A_1, ..., A_n$ umgekehrt
$$P(A_1 \cap ... \cap A_n) \;=\; P(A_1) \cdot ... \cdot P(A_n),$$
impliziert dies i.a. noch nicht
$$P(A_i \cap A_j) \;=\; P(A_i)P(A_j)$$
für alle $1 \leq i \neq j \leq n$.

Diese Feststellungen geben Anlaß zu folgender Definition:

11.4. Definition. Es seien P eine W-Verteilung, I eine beliebige nichtleere Indexmenge und $(A_i)_{i \in I}$ eine Familie von Ereignissen.

(a) Die A_i heißen *paarweise stochastisch unabhängig* (unter P), wenn

$$P(A_i \cap A_j) \;=\; P(A_i)P(A_j)$$

für alle $i, j \in I$ mit $i \neq j$.

(b) Die A_i heißen *stochastisch unabhängig* (unter P), wenn

$$P(A_{i_1} \cap ... \cap A_{i_n}) \; = \; P(A_{i_1}) \cdot ... \cdot P(A_{i_n})$$

für jede endliche Auswahl paarweise verschiedener $i_1, ..., i_n \in I$.

11.5. Bemerkung. Gegeben eine Familie $(A_i)_{i \in I}$ stochastisch unabhängiger Ereignisse, ist auch jede Teilfamilie $(A_i)_{i \in J}$, $J \subset I$, stochastisch unabhängig.

Da Wahrscheinlichkeiten von Ereignissen häufig über deren Komplemente berechnet werden, stellt sich die Frage, ob die Eigenschaft der stochastischen Unabhängigkeit unter der Komplementbildung erhalten bleibt. Für zwei Ereignisse rechnet man dies direkt nach: Sind A_1, A_2 stochastisch unabhängig, gilt

$$\begin{aligned} P(A_1 \cap A_2^c) &= P(A_1 \backslash (A_1 \cap A_2)) \\ &= P(A_1) - P(A_1 \cap A_2) \\ &= P(A_1) - P(A_1)P(A_2) \\ &= P(A_1)\big(1 - P(A_2)\big) \\ &= P(A_1)P(A_2^c), \end{aligned}$$

d.h. auch die Unabhängigkeit von A_1 und A_2^c. Die allgemeine Aussage bildet Inhalt des nachfolgenden Satzes:

11.6. Satz. *Eine Familie $(A_i)_{i \in I}$ von Ereignissen ist genau dann stochastisch unabhängig, wenn für jede Wahl J, K disjunkter endlicher Teilmengen von I*

$$P\left(\bigcap_{j \in J} A_j^c \cap \bigcap_{k \in K} A_k \right) \; = \; \prod_{j \in J} P(A_j^c) \prod_{k \in K} P(A_k), \tag{11.1}$$

wobei $\cap_{j \in \emptyset} A_j \stackrel{\text{def}}{=} \Omega$ und $\prod_{j \in \emptyset} P(A_j) = 1$.

Beweis: "$\Leftarrow$" klar nach Definition der stochastischen Unabhängigkeit, wenn man $J = \emptyset$ wählt.

"$\Rightarrow$" Wir zeigen die Behauptung durch Induktion über $|J|$: Falls $|J| = 0$, also $J = \emptyset$, ergibt sich (11.1) sofort aus der Definition der stochastischen Unabhängigkeit, da Komplemente gar nicht auftreten.

Induktionsschritt $n \to n+1$: (11.1) gelte für jede Wahl $J \subset I$ mit $|J| = n$ und $K \subset I$ mit $|K| < \infty$ und $J \cap K = \emptyset$. Sei nun $|J| = n+1$ und $j_0 \in J$. Dann erhalten wir vermöge der Zerlegung $(\cap_{j \in J_0} A_j^c) \cap (\cap_{k \in K} A_k) = (\cap_{j \in J} A_j^c) \cap (\cap_{k \in K} A_k) + (\cap_{j \in J_0} A_j^c) \cap (\cap_{k \in K_0} A_k)$ mit

$J_0 \overset{\text{def}}{=} J \backslash \{j_0\}$ und $K_0 \overset{\text{def}}{=} K \cup \{j_0\}$

$$\begin{aligned}
&P\Big(\bigcap_{j\in J} A_j^c \cap \bigcap_{k\in K} A_k\Big) \\
&= P\Big(\bigcap_{j\in J_0} A_j^c \cap \bigcap_{k\in K} A_k\Big) - P\Big(\bigcap_{j\in J_0} A_j^c \cap \bigcap_{k\in K_0} A_k\Big) \\
&= \Big(\prod_{j\in J_0} P(A_j^c) \prod_{k\in K} P(A_k)\Big) - \Big(\prod_{j\in J_0} P(A_j^c) \prod_{k\in K_0} P(A_k)\Big) \\
&= \Big(\prod_{j\in J_0} P(A_j^c) \prod_{k\in K} P(A_k)\Big)\big(1 - P(A_{j_0})\big) \\
&= \prod_{j\in J} P(A_j^c) \prod_{k\in K} P(A_k).
\end{aligned}$$

Die Induktionsvoraussetzung wurde in der zweiten Gleichung benutzt $(|J_0\}| = n)$. ◊

Im Fall einer endlichen Familie $(A_i)_{i\in I}$ unabhängiger Ereignisse läßt sich die Aussage des vorherigen Satzes wie folgt vereinfachen:

11.7. Satz. *Eine endliche Familie $(A_i)_{1\le i\le n}$ von Ereignissen ist genau dann stochastisch unabhängig, wenn*

$$P(B_1 \cap ... \cap B_n) \;=\; P(B_1)\cdot ... \cdot P(B_n)$$

für jede Wahl $B_i \in \{A_i, A_i^c\}$, $1 \le i \le n$, gilt.

Beweis: "⇐" ist klar nach Satz 11.6.

"⇒" Die Behauptung wird durch Induktion über n bewiesen, wobei für $n = 1$ offenbar nichts zu zeigen ist. Gegeben Ereignisse $A_1, ..., A_{n+1}$, gelte nun

$$P(B_1 \cap ... \cap B_{n+1}) \;=\; P(B_1)\cdot ... \cdot P(B_{n+1}) \tag{11.2}$$

für jede Wahl $B_i \in \{A_i, A_i^c\}$, $1 \le i \le n+1$. Sei $J \subset \{1, ..., n+1\}$ beliebig mit $|J| = n$, also $J = \{1, ..., n+1\}\backslash\{j_0\}$ für ein $j_0 \in \{1, ..., n+1\}$. Dann folgt

$$\begin{aligned}
P\Big(\bigcap_{j\in J} B_j\Big) &= P\Big(A_{j_0} \cap \bigcap_{j\in J} B_j\Big) + P\Big(A_{j_0}^c \cap \bigcap_{j\in J} B_j\Big) \\
&= P(A_{j_0}) \prod_{j\in J} P(B_j) + P(A_{j_0}^c) \prod_{j\in J} P(B_j) \;=\; \prod_{j\in J} P(B_j).
\end{aligned}$$

Nach Induktionsvoraussetzung $(|J| = n)$ sind somit $A_j, j \in J$, stochastisch unabhängig, d.h.

$$P\Big(\bigcap_{j\in K} A_j\Big) \;=\; \prod_{j\in K} P(A_j)$$

für jede Teilmenge K von J. Da nach Voraussetzung insbesondere

$$P\left(\bigcap_{j=1}^{n+1} A_j\right) = \prod_{j=1}^{n+1} P(A_j)$$

gilt (wähle $B_i = A_i$ für $i = 1, ..., n+1$ in (11.2)), erhalten wir die stochastische Unabhängigkeit der $A_i, i = 1, ..., n+1$. $\diamondsuit$

Mittels des vorherigen Satzes kann man nun für *Ereignisfolgen* $(A_n)_{n\geq 1}$ das folgende vereinfachte Kriterium für stochastische Unabhängigkeit zeigen, dessen einfachen Beweis wir dem Leser zur Übung überlassen.

11.8. Satz. *Eine Folge $(A_n)_{n\geq 1}$ von Ereignissen ist genau dann stochastisch unabhängig, wenn*

$$P(B_1 \cap ... \cap B_n) = P(B_1) \cdot ... \cdot P(B_n)$$

für alle $n \geq 1$ und jede Wahl $B_i \in \{A_i, A_i^c\}$ gilt.

11.9. Beispiel. ***(Binomialverteilung)*** [☞ 1.8] Ein Versuch, der mit Wahrscheinlichkeit θ einen bestimmten Ausgang $\mathcal{A}$ und mit Wahrscheinlichkeit $1 - \theta$ das Gegenteil $\mathcal{A}^c$ liefert, werde n-mal "unabhängig" wiederholt. Interessiert nur das Auftreten bzw. Nichtauftreten von $\mathcal{A}$, so läßt sich die Versuchsreihe vom Umfang n durch ein Element $\omega \in \Omega = \{0,1\}^n$ beschreiben, wobei "1" das Eintreten von $\mathcal{A}$ und "0" das Nichtauftreten von $\mathcal{A}$ bezeichnet. Für $k = 1, ..., n$ seien E_j^1 und E_j^0 die Ereignisse, daß $\mathcal{A}$ bzw. $\mathcal{A}^c$ im j-ten Versuch eintritt, also

$$E_j^i \stackrel{\text{def}}{=} \{0,1\}^{j-1} \times \{i\} \times \{0,1\}^{n-j} = \{\omega \in \Omega : \omega_j = i\}, \quad i = 0, 1.$$

Dann gilt $P(E_j^1) = \theta$ und $P(E_j^0) = 1 - \theta$ für jedes j. Interpretiert man die experimentelle Unabhängigkeit im Sinne stochastischer Unabhängigkeit der Ereignisse $E_1^{i_1}, ..., E_n^{i_n}$ für jede Wahl von $i_1, ..., i_n \in \{0,1\}$, so ergibt sich die Wahrscheinlichkeit für jedes Ergebnis $\omega = (\omega_1, ..., \omega_n)$ mit genau k Einsen als $p(\omega) = \theta^k(1-\theta)^{n-k}$, denn

$$\{(\omega_1, ..., \omega_n)\} = \bigcap_{j=1}^{n} E_j^{\omega_j}$$

und deshalb

$$p(\omega) = P(\{\omega\}) = P\left(\bigcap_{j=1}^{n} E_j^{\omega_j}\right) = \prod_{j=1}^{n} P(E_j^{\omega_j}) = \theta^{\sum_{j=1}^n \omega_j}(1-\theta)^{n-\sum_{j=1}^n \omega_j}.$$

Die Ereignisse $A_k \stackrel{\text{def}}{=} \{\omega \in \Omega : \sum_{i=1}^n \omega_i = k\}$, $0 \leq k \leq n$, besitzen somit die Wahrscheinlichkeiten

$$P(A_k) = \sum_{\omega \in A_k} p(\omega) = |A_k|\theta^k(1-\theta)^{n-k} = \binom{n}{k}\theta^k(1-\theta)^{n-k},$$

und für die Zufallsgröße $X : \Omega \to \{0, ..., n\}, \omega \mapsto \sum_{i=1}^n \omega_i$, erhalten wir demnach

$$P(X = k) \;=\; P(A_k) \;=\; \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

für $0 \le k \le n$, d.h., X ist binomialverteilt mit Parametern n und θ.

Als einfache Folgerung aus den beiden vorhergehenden Sätze halten wir fest:

11.10. Korollar. *Sei $I \subset \mathbb{N}$ und $(A_i)_{i \in I}$ eine Familie unabhängiger Ereignisse. Dann gilt*

$$P\left(\bigcup_{i \in I} A_i\right) \;=\; 1 - \prod_{i \in I}(1 - P(A_i)). \tag{11.3}$$

Beweis: (a) Für endliches I, o.B.d.A. $I = \{1, ..., n\}$, gilt

$$P\left(\bigcup_{i=1}^n A_i\right) \;=\; 1 - P\left(\bigcap_{i=1}^n A_i^c\right) \;=\; 1 - \prod_{i=1}^n P(A_i^c) \;=\; 1 - \prod_{i=1}^n (1 - P(A_i)),$$

wobei für die zweite Gleichheit Satz 11.7 benutzt wurde.

(b) Für unendliches I, o.B.d.A. $I = \mathbb{N}$, gilt

$$\begin{aligned} P\left(\bigcup_{i \ge 1} A_i\right) &= P\left(\lim_{n \to \infty} \bigcup_{i=1}^n A_i\right) \\ &= \lim_{n \to \infty} P\left(\bigcup_{i=1}^n A_i\right) \\ &= \lim_{n \to \infty} 1 - \prod_{i=1}^n (1 - P(A_i)) \\ &= 1 - \prod_{i \ge 1}(1 - P(A_i)), \end{aligned}$$

wobei die Stetigkeit von W-Verteilungen sowie Teil (a) verwendet wurden. $\diamondsuit$

11.11. Lemma von Borel-Cantelli. *Für jede Folge $(A_n)_{n \ge 1}$ von Ereignissen gilt:*

(a) Aus $\sum_{n \ge 1} P(A_n) < \infty$ folgt $P(\limsup_{n \to \infty} A_n) = 0$.

(b) Aus $\sum_{n \ge 1} P(A_n) = \infty$ und der Unabhängigkeit der A_n folgt $P(\limsup_{n \to \infty} A_n) = 1$.

Beweis: (a) Wegen

$$P(\limsup_{n \to \infty} A_n) \;\le\; P\left(\bigcup_{n \ge k} A_n\right) \;\le\; \sum_{n \ge k} P(A_n)$$

für alle $k \geq 1$ gilt

$$\sum_{n\geq 1} P(A_n) < \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 0.$$

(b) Seien nun $A_1, A_2, \ldots$ stochastisch unabhängig. Da P stetig von unten und oben ist und $1 - x \leq e^{-x}$ für alle $x \geq 0$, erhalten wir dann

$$\begin{aligned}
1 - P(\limsup_{n\to\infty} A_n) &= P\left(\bigcup_{n\geq 1} \bigcap_{m\geq n} A_m^c\right) \\
&= \lim_{n\to\infty} P\left(\bigcap_{m\geq n} A_m^c\right) \\
&= \lim_{n\to\infty} \lim_{k\to\infty} P\left(\bigcap_{m=n}^{n+k} A_m^c\right) \\
&= \lim_{n\to\infty} \lim_{k\to\infty} \prod_{m=n}^{n+k} (1 - P(A_m)) \\
&\leq \lim_{n\to\infty} \lim_{k\to\infty} \exp\left(-\sum_{m=n}^{n+k} P(A_m)\right).
\end{aligned}$$

Damit gilt offenbar

$$\sum_{n\geq 1} P(A_n) = \infty \quad \Rightarrow \quad P(\limsup_{n\to\infty} A_n) = 1. \qquad \diamondsuit$$

Da $\liminf_{n\to\infty} A_n = (\limsup_{n\to\infty} A_n^c)^c$, ergibt sich als direkte Folgerung:

11.12. Korollar. *Gegeben eine Folge $(A_n)_{n\geq 1}$ unabhängiger Ereignisse, gilt:*

$$(11.4) \qquad P(\liminf_{n\to\infty} A_n) = \begin{cases} 1, & \textit{falls } \sum_{n\geq 1} P(A_n^c) < \infty \\ 0, & \textit{falls } \sum_{n\geq 1} P(A_n^c) = \infty \end{cases}.$$

Für eine Folge $(A_n)_{n\geq 1}$ unabhängiger Ereignisse kann also sowohl

$$\limsup_{n\to\infty} A_n = \text{"unendlich viele der } A_n \text{ treten ein"}$$

als auch

$$\liminf_{n\to\infty} A_n = \text{"fast alle der } A_n \text{ treten ein"}$$

immer nur mit Wahrscheinlichkeit 0 oder 1 eintreten.

11.13. Beispiel. ***(Wiederholtes Werfen eines fairen Würfels)*** [☞ 9.3] Es werde ein fairer Würfel unendlich oft geworfen und der Ausgang dieses ZE's durch eine Folge $\omega =$

$(\omega_1, \omega_2, ...) \in \Omega \stackrel{\text{def}}{=} \{1,2,3,4,5,6\}^{\mathbb{N}}$ beschrieben. Dann sind die Ereignisse

$$A_n \stackrel{\text{def}}{=} \{\omega \in \Omega : \omega_n = 6\}, \quad n \geq 1,$$

stochastisch unabhängig mit $P(A_n) = \frac{1}{6}$, wobei wir allerdings nochmals darauf hinweisen (☞ auch die Bemerkungen am Ende von Abschnitt 9), daß die zugrundeliegende W-Verteilung P über Ω nicht mehr diskret ist und folglich strenggenommen hier gar nicht behandelt werden kann. Das Ereignis $A \stackrel{\text{def}}{=} \limsup_{n\to\infty} A_n$ tritt ein, wenn die "6" in unendlich vielen Versuchen gewürfelt wird. Da $\sum_{n\geq 1} P(A_n) = \sum_{n\geq 1} \frac{1}{6} = \infty$, impliziert das Borel-Cantelli-Lemma $P(A) = 1$. Bei unendlichem Werfen eines fairen Würfels tritt also die "6" mit Wahrscheinlichkeit 1 unendlich oft auf.

Fragen wir für beliebiges $k \geq 2$ nach der Wahrscheinlichkeit, daß unendlich oft k-mal in Folge eine "6" auftritt, so läßt sich auch hierfür die Antwort mittels des Borel-Cantelli-Lemmas geben. Sei dazu für $n \geq 1$

$$B_n \stackrel{\text{def}}{=} A_n \cap A_{n+1} \cap ... \cap A_{n+k-1}$$

das Ereignis ab dem n-ten Versuch k-mal in Folge eine "6" zu werfen. Dann gilt vermöge der Unabhängigkeit der A_j

$$P(B_n) = \prod_{j=0}^{k-1} P(A_{n+j}) = \frac{1}{6^k}$$

für alle $n \geq 1$, und das fragliche Ereignis lautet $B \stackrel{\text{def}}{=} \limsup_{n\to\infty} B_n$. Die B_n sind jedoch nicht stochastisch unabhängig, wie man sich leicht überlegt, wohl aber die Teilfolge $B_1, B_{k+1}, B_{2k+1}, ...$, die wir mit $B_1', B_2', ...$ bezeichnen wollen, also $B_n' \stackrel{\text{def}}{=} B_{(n-1)k+1}$ für $n \geq 1$. Es gilt nämlich

$$\begin{aligned} P\left(\bigcap_{i=1}^n B_{l_i}'\right) &= P\left(\bigcap_{i=1}^n \bigcap_{j=(l_i-1)k+1}^{l_i k} A_j\right) \\ &= \prod_{i=1}^n \prod_{j=(l_i-1)k+1}^{l_i k} P(A_i) \\ &= \prod_{i=1}^n P\left(\bigcap_{j=(l_i-1)k+1}^{l_i k} A_j\right) \\ &= \prod_{i=1}^n P(B_{l_i}') \end{aligned}$$

für jede Wahl $1 \leq l_1 < ... < l_n$ und $n \geq 1$. Setzen wir $B' \stackrel{\text{def}}{=} \limsup_{n\to\infty} B_n'$, so folgt offenbar $B' \subset B$, und weil außerdem $\sum_{n\geq 1} P(B_n') = \sum_{n\geq 1} \frac{1}{6^k} = \infty$, liefert schließlich das Lemma von Borel-Cantelli $1 = P(B') \leq P(B)$, also $P(B) = 1$. Auch das Auftreten beliebig langer Blöcke mit lauter Sechsen in einer unendlichen Würfelreihe hat somit Wahrscheinlichkeit 1.

12. Stochastisch unabhängige Zufallsvariablen

Führt man sich vor Augen, daß der Begriff "stochastische Unabhängigkeit" aus intuitiver Sicht auch für ZE bzw. die sie beschreibenden Zufallsvariablen Sinn macht, etwa beim wiederholten Werfen eines Würfels oder einer Münze, so wird klar, warum wir diesen Begriff nun in entsprechender Weise ausdehnen wollen. Im Rückblick auf die Überlegungen des vorherigen Abschnitts liegt folgende Definition nahe:

12.1. Definition. Sei P eine W-Verteilung über Ω, I eine nichtleere Indexmenge und $X_i : \Omega \to \mathcal{X}_i$ für jedes $i \in I$ eine diskrete Zufallsvariable mit Werten in einer beliebigen Menge $\mathcal{X}_i$. Dann heißt die Familie $(X_i)_{i \in I}$ *stochastisch unabhängig*, wenn

$$P(X_{i_1} \in A_1, ..., X_{i_n} \in A_n) \;=\; \prod_{k=1}^{n} P(X_k \in A_k) \tag{12.1}$$

für jedes $n \leq |I|$, jede Auswahl $i_1, ..., i_n$ paarweise verschiedener Elemente von I sowie alle $A_1 \subset \mathcal{X}_{i_1}, ..., A_n \subset \mathcal{X}_{i_n}$ gilt.

Der nächste Satz zeigt, daß die stochastische Unabhängigkeit einer Familie $(X_i)_{i \in I}$ von diskreten Zufallsvariablen bereits aus der Gültigkeit von (12.1) für Einpunktmengen $A_1, ..., A_n$ folgt.

12.2. Satz. *Unter den Voraussetzungen in 12.1 sind äquivalent:*

(a) *Die Familie $(X_i)_{i \in I}$ ist stochastisch unabhängig.*

(b) *Für alle $n \leq |I|$, paarweise verschiedenen $i_1, ..., i_n \in I$ und alle $(x_1, ..., x_n) \in \mathcal{X}_{i_1} \times ... \times \mathcal{X}_{i_n}$ gilt*

$$P(X_{i_1} = x_1, ..., X_{i_n} = x_n) \;=\; \prod_{k=1}^{n} P(X_{i_k} = x_k) \;=\; \prod_{k=1}^{n} p^{X_{i_k}}(x_k). \tag{12.2}$$

Beweis: "(a)$\Rightarrow$(b)" ist klar.

Die umgekehrte Schlußrichtung "(b)$\Rightarrow$(a)" ergibt sich durch direktes Nachrechnen von (12.1), wobei wir dies hier besonders ausführlich tun wollen, damit der Leser die genauen Argumentationsschritte wenigstens einmal penibel nachvollzieht. Gegeben irgendein $n \leq |I|$, wählen wir also beliebige paarweise verschiedene $i_1, ..., i_n \in I$ sowie $A_1 \subset \mathcal{X}_{i_1}, ..., A_n \subset \mathcal{X}_{i_n}$. Dann gilt

$$\begin{aligned} P(X_{i_1} \in A_1, ..., X_{i_n} \in A_n) \;&=\; P((X_{i_1}, ..., X_{i_n}) \in A_1 \times ... \times A_n) \\ &=\; \sum_{(x_1,...,x_n) \in A_{i_1} \times ... \times A_{i_n}} P((X_{i_1}, ..., X_{i_n}) = (x_1, ..., x_n)) \\ &=\; \sum_{(x_1,...,x_n) \in A_1 \times ... \times A_n} P(X_{i_1} = x_1, ..., X_{i_n} = x_n) \end{aligned}$$

$$
\begin{aligned}
&= \sum_{(x_1,\dots,x_n)\in A_1\times\dots\times A_n} P(X_{i_1}=x_1)\cdot\ldots\cdot P(X_{i_n}=x_n) \\
&= \sum_{x_1\in A_1}\dots\sum_{x_n\in A_n} P(X_{i_1}=x_1)\cdot\ldots\cdot P(X_{i_n}=x_n) \\
&= \left(\sum_{x_1\in A_1} P(X_{i_1}=x_1)\right)\cdot\ldots\cdot\left(\sum_{x_n\in A_n} P(X_{i_n}=x_n)\right) \\
&= \prod_{k=1}^{n} P(X_{i_k}\in A_k).
\end{aligned}
$$

Dabei beachte man:

- Die zweite Zeile gilt vermöge (1.1) für die diskrete W-Verteilung $P^{(X_{i_1},\dots,X_{i_n})}$.
- Die dritte und fünfte Zeile ist nur eine andere Schreibweise der jeweils vorherigen Zeile.
- Die vierte Zeile gilt nach Voraussetzung (12.2).
- Die vorletzte Zeile ergibt sich durch sukzessives Ausklammern.
- Die letzte Zeile folgt wiederum aus (1.1), dieses Mal für die $P^{X_{i_1}},\dots,P^{X_{i_n}}$. ♢

Der anschließende Satz ist das Analogon zu Satz 11.7 für endliche Familien diskreter Zufallsvariablen. Dabei seien weiterhin die Voraussetzungen von 12.1 unterstellt.

12.3. Satz. *Für eine endliche Familie $(X_i)_{1\le i\le n}$ diskreter Zufallsvariablen sind äquivalent:*

(a) $X_1,\dots,X_n$ *sind stochastisch unabhängig.*

(b) $P(X_1=x_1,\dots,X_n=x_n)=\prod_{k=1}^{n}P(X_k=x_k)$ *für alle* $(x_1,\dots,x_n)\in\mathcal{X}_1\times\dots\times\mathcal{X}_n$.

(c) $P(X_1\in A_1,\dots,X_n\in A_n)=\prod_{k=1}^{n}P(X_k\in A_k)$ *für alle* $A_1\subset\mathcal{X}_1,\dots,A_n\subset\mathcal{X}_n$.

Beweis: "(a)⇒(b)" folgt aus der Definition stochastischer Unabhängigkeit, während "(b)⇒(c)" durch eine analoge Rechnung wie im Beweis des vorherigen Satzes gezeigt wird. Zum Nachweis von "(c)⇒(a)" wähle $m\in\{1,\dots,n\}$, paarweise verschiedene $i_1,\dots,i_m\in\{1,\dots,n\}$ sowie $A_{i_1}\subset\mathcal{X}_{i_1},\dots,A_{i_m}\subset\mathcal{X}_{i_m}$. Setze $A_i=\mathcal{X}_i$ für $i\in\{1,\dots,n\}\backslash\{i_1,\dots,i_m\}$ (Auffüllargument). Dann folgt aus (c)

$$
\begin{aligned}
&P(X_{i_1}\in A_{i_1},\dots,X_{i_m}\in A_{i_m}) \\
&\quad= P(X_1\in A_1,\dots,X_n\in A_n) \\
&\quad= \prod_{k=1}^{n}P(X_k\in A_k) \\
&\quad= \left(\prod_{j=1}^{m}P(X_{i_j}\in A_{i_j})\right)\left(\prod_{k\in\{1,\dots,n\}\backslash\{i_1,\dots,i_m\}}P(X_k\in\mathcal{X}_k)\right) \\
&\quad= \prod_{j=1}^{m}P(X_{i_j}\in A_{i_j}),
\end{aligned}
$$

was (a) beweist. ◊

Für Folgen von Zufallsvariablen notieren wir ohne Beweis als unmittelbare Folgerung:

12.4. Korollar. *Für eine Folge $(X_n)_{n\ge 1}$ diskreter Zufallsgrößen sind äquivalent:*

(a) $(X_n)_{n\ge 1}$ *ist stochastisch unabhängig.*

(b) $P(X_1 = x_1, ..., X_n = x_n) = \prod_{k=1}^n P(X_k = x_k)$ *für alle* $(x_1, ..., x_n) \in \mathcal{X}_1 \times ... \times \mathcal{X}_n$ *und* $n \ge 1$.

(c) $P(X_1 \in A_1, ..., X_n \in A_n) = \prod_{k=1}^n P(X_k \in A_k)$ *für alle* $A_1 \subset \mathcal{X}_1, ..., A_n \subset \mathcal{X}_n$ *und* $n \ge 1$.

Die im vorherigen Abschnitt behandelte Unabhängigkeit von Ereignissen läßt sich in das Hiesige "einbetten", wenn man ein Ereignis A mit seiner Indikatorfunktion $\mathbf{1}_A$ identifiziert (☞ Beispiel 4.5).

12.5. Korollar. *Sei P eine W-Verteilung über $\Omega \ne \emptyset$ und $(A_i)_{i\in I}$ eine Familie von Ereignissen. Jeweils äquivalente Aussagen bilden dann:*

(a1) Die A_i sind stochastisch unabhängig.

(a2) Die 1_{A_i} sind stochastisch unabhängig.

sowie

(b1) Die A_i sind paarweise stochastisch unabhängig.

(b2) Die 1_{A_i} sind paarweise unkorreliert.

Beweis: "(a1)⇒(a2)" Es gilt $\mathbf{1}_A^{-1}(\{x\}) \in \{A, A^c\}$ für alle $A \subset \Omega$ und $x \in \{0,1\}$. Daher folgt unter Benutzung von Satz 11.6 für alle $n \le |I|$, $i_1, ..., i_n \in I$ und $x_1, ..., x_n \in \{0,1\}$

$$\begin{aligned} P(\mathbf{1}_{A_{i_1}} = x_1, ..., \mathbf{1}_{A_{i_n}} = x_n) &= P\left(\bigcap_{k=1}^n \underbrace{\mathbf{1}_{A_{i_k}}^{-1}(\{x_k\})}_{=A_{i_k} \text{ oder } A_{i_k}^c}\right) \\ &= \prod_{k=1}^n P(\mathbf{1}_{A_{i_k}}^{-1}(\{x_k\}) \\ &= \prod_{k=1}^n P(\mathbf{1}_{A_{i_k}} = x_k) \end{aligned}$$

und somit die stochastische Unabhängigkeit der $\mathbf{1}_{A_i}$.

"(a2)⇒(a1)" Umgekehrt gilt für jede Wahl endlicher disjunkter $J, K \subset I$

$$\begin{aligned} P\left(\bigcap_{j\in J} A_j^c \cap \bigcap_{k\in K} A_k\right) &= P\left(\bigcap_{j\in J}\{\mathbf{1}_{A_j} = 0\} \cap \bigcap_{k\in K}\{\mathbf{1}_{A_k} = 1\}\right) \\ &= \prod_{j\in J} P(\mathbf{1}_{A_j} = 0) \prod_{k\in K} P(\mathbf{1}_{A_k} = 1) \\ &= \prod_{j\in J} P(A_j^c) \prod_{k\in K} P(A_k) \end{aligned}$$

woraus nach Satz 11.6 die Unabhängigkeit der A_i folgt.

(b) Nach Teil (a) sind die A_i genau dann paarweise stochastisch unabhängig, wenn

$$P(A_i \cap A_j) \;=\; P(\mathbf{1}_{A_i} = 1, \mathbf{1}_{A_j} = 1) \;=\; E(\mathbf{1}_{A_i}\mathbf{1}_{A_j})$$

mit

$$P(A_i)P(A_j) \;=\; E\mathbf{1}_{A_i}E\mathbf{1}_{A_j}$$

für alle $i, j \in I$, $i \neq j$, übereinstimmt, wenn also

$$0 \;=\; E(\mathbf{1}_{A_i}\mathbf{1}_{A_j}) - E\mathbf{1}_{A_i}E\mathbf{1}_{A_j} \;=\; C\mathrm{ov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j})$$

für alle $i, j \in I$, $i \neq j$. ♢

Insbesondere sind also zwei Indikatorfunktionen $\mathbf{1}_A, \mathbf{1}_B$ genau dann unkorreliert, wenn sie stochastisch unabhängig sind. Daß dies für zwei quadratisch integrierbare Zufallsgrößen nicht generell gilt, zeigt:

12.6. Beispiel. Sei X Laplace-verteilt auf $\{-1, 0, 1\}$ und $Y = X^2$. Da $EX = 0$, folgt $C\mathrm{ov}(X, Y) = EXY = EX^3 = 0$, also die Unkorreliertheit von X, Y. Dagegen zeigt

$$P(X = 0, Y = 0) \;=\; P(X = 0) \;=\; \frac{1}{3} \;\neq\; \frac{1}{9} \;=\; P(X = 0)P(Y = 0),$$

daß X, Y nicht stochastisch unabhängig sind.

Es gibt aber auch eine gute Nachricht: Stochastisch unabhängige, integrierbare Zufallsgrößen sind stets paarweise unkorreliert. Eine allgemeinere Aussage liefert:

12.7. Mulitplikationssatz. *Gegeben stochastisch unabhängige, integrierbare Zufallsgrößen $X_1, ..., X_n$, ist auch deren Produkt $\prod_{k=1}^n X_k$ (und damit $|\prod_{k=1}^n X_k|$ nach Satz 6.6) integrierbar, und es gilt*

$$E\left(\prod_{k=1}^{n} X_k\right) \;=\; \prod_{k=1}^{n} EX_k. \tag{12.3}$$

Beweis: Sei $\mathcal{X}_k \subset \mathbb{R}$ der Träger von P^{X_k} für $k = 1, ..., n$. Da EX_k existiert, konvergiert

$$\sum_{x_k \in \mathcal{X}_k} x_k P(X_k = x_k)$$

nach Satz 8.6 absolut. Dies und die Unabhängigkeit der X_k beachtend, folgt

$$\begin{aligned}
\prod_{k=1}^n E|X_k| &= \prod_{k=1}^n \left(\sum_{x_k \in \mathcal{X}_k} |x_k| P(X_k = x_k) \right) \\
&= \sum_{x_1 \in \mathcal{X}_1} \ldots \sum_{x_n \in \mathcal{X}_n} |x_1 \cdot \ldots \cdot x_n| P(X_1 = x_1) \cdot \ldots \cdot P(X_n = x_n) \\
&= \sum_{x_1 \in \mathcal{X}_1} \ldots \sum_{x_n \in \mathcal{X}_n} |x_1 \cdot \ldots \cdot x_n| P(X_1 = x_1, \ldots, X_n = x_n) \\
&= \sum_{(x_1,\ldots,x_n) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n} |x_1 \cdot \ldots \cdot x_n| P((X_1, \ldots, X_n) = (x_1, \ldots, x_n)) \\
&= \sum_{(x_1,\ldots,x_n) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n} f(x_1, \ldots, x_n) P((X_1, \ldots, X_n) = (x_1, \ldots, x_n))
\end{aligned}$$

mit $f(x_1, \ldots, x_n) = \prod_{k=1}^n |x_k|$. Nach Satz 6.8 ist die letzte Zeile gleich

$$Ef(X_1, \ldots, X_n) = E\left| \prod_{k=1}^n X_k \right|$$

und somit $E|\prod_{k=1}^n X_k| < \infty$. Damit ist $\prod_{k=1}^n X_k$ integrierbar, und die obige Rechnung ohne Betragstriche liefert (12.3). $\diamondsuit$

Hier nun die schon erwähnte gute Nachricht als direkte Folgerung:

12.8. Korollar. *Es seien $X_i, i \in I$, paarweise stochastisch unabhängige, integrierbare Zufallsgrößen. Dann sind die X_i paarweise unkorreliert.*

Beweis: Für beliebige $i, j \in I, i \neq j$, gilt nach Satz 12.7

$$0 = E(X_i X_j) - EX_i EX_j = Cov(X_i, X_j). \qquad \diamondsuit$$

Für Folgen paarweise unabhängiger Zufallsgrößen erhalten wir nun als einfache Folgerung aus Satz 9.5 ein weiteres schwaches Gesetz der großen Zahlen:

12.9. Korollar. *Gegeben eine Folge $(X_n)_{n \geq 1}$ paarweise stochastisch unabhängiger, quadratisch integrierbarer Zufallsgrößen mit Erwartungswerten $\mu_n \stackrel{\text{def}}{=} EX_n$, sei*

$$\overline{X}_n \stackrel{\text{def}}{=} n^{-1} \sum_{k=1}^n X_k \quad \textit{und} \quad \overline{\mu}_n \stackrel{\text{def}}{=} n^{-1} \sum_{k=1}^n \mu_k.$$

Dann gilt

$$\overline{X}_n - \overline{\mu}_n \stackrel{P}{\to} 0 \quad (n \to \infty),$$

sofern die Varianzen der Bedingung

(12.4) $$\lim_{n \to \infty} \frac{1}{n^2} \sum_{k=1}^n Var X_k = 0$$

genügen.

Beweis: Die Behauptung ergibt sich direkt aus Satz 9.5, denn die paarweise Unabhängigkeit und damit paarweise Unkorreliertheit liefert $Var(\sum_{k=1}^n X_k) = \sum_{k=1}^n \mathrm{Var} X_k$ nach Satz 7.8. ♢

Von besonderem Interesse ist der Spezialfall, daß alle X_n dieselbe Verteilung besitzen:

12.10. Korollar. *Gegeben eine Folge $(X_n)_{n\geq 1}$ (paarweise) stochastisch unabhängiger, identisch verteilter und quadratisch integrierbarer Zufallsgrößen mit Erwartungswert μ, gilt*

$$\overline{X}_n \xrightarrow{P} \mu \quad (n \to \infty).$$

Beweis: Es genügt der Hinweis, daß die Bedingung (12.4) wegen $V\mathrm{ar}X_1 = V\mathrm{ar}X_2 = ...$ offenbar erfüllt ist. ♢

Folgen stochastisch unabhängiger und identisch verteilter Zufallsgrößen treten gerade bei der Beschreibung der wiederholten Durchführung desselben Versuchs auf, sofern die einzelnen Durchführungen sich nicht gegenseitig beeinflussen. Korollar 12.10 bestätigt demnach die intuitiv plausible Vorstellung, daß dann bei wachsender Anzahl von Versuchswiederholungen das Durchschnittsergebnis gegen den Erwartungswert pro Versuch konvergiert, und zwar im Sinne die Konvergenz nach Wahrscheinlichkeit.

Wenden wir uns als nächstes kurz der Berechnung der Verteilung der Summe von unabhängigen Zufallsgrößen zu. Wir beschränken uns dabei auf die Betrachtung von zwei Beispielen, die das Vorgehen im Fall $\mathbb{N}_0$-wertiger, also insbesondere diskreter Zufallsgrößen verdeutlichen.

12.11. Beispiel. ***(Binomialverteilungen)*** Seien X, Y zwei stochastisch unabhängige Zufallsgrößen mit einer $B(m,\theta)$- bzw. $B(n,\theta)$-Verteilung, wobei $m, n \in \mathbb{N}$ und $\theta \in (0,1)$. Gesucht ist die Verteilung von $X+Y$, deren Träger offenkundig in $\{0, ..., m+n\}$ enthalten ist. Es gilt für alle $k \in \{0, ..., m+n\}$ und unter Beachtung von $\binom{n}{k} \stackrel{\text{def}}{=} 0$ für $k < 0$

$$\begin{aligned} P(X+Y=k) &= \sum_{i=0}^{m} P(X=i, Y=k-i) \\ &= \sum_{i=0}^{k} P(X=i)P(Y=k-i) \\ &= \sum_{i=0}^{k} \left(\binom{m}{i} \theta^i (1-\theta)^{m-i} \binom{n}{k-i} \theta^{k-i} (1-\theta)^{n-k+i} \right) \\ &= \theta^k (1-\theta)^{m+n-i} \sum_{i=0}^{k} \binom{m}{i} \binom{n}{k-i} \\ &= \binom{m+n}{k} \theta^k (1-\theta)^{m+n-i}. \end{aligned}$$

$X+Y$ ist also ebenfalls binomialverteilt mit Parametern $m+n$ und θ. Die in die letzte Zeile eingehende Identität $\binom{m+n}{k} = \sum_{i=0}^{k}\binom{m}{i}\binom{n}{k-i}$ kann man natürlich per Induktion über k, aber auch mittels folgender kombinatorischer Überlegung beweisen: Bekanntlich gibt $\binom{m+n}{k}$ die Anzahl der k-elementigen Teilmengen von $\{1,2,...,m+n\}$ an. Diese lassen sich aber auch erhalten, indem man zu jedem $i \in \{0,...,k\}$ je eine i– bzw. $(k-i)$–elementige Teilmenge von $\{1,...,m\}$ bzw. $\{m+1,...,m+n\}$ wählt, wofür es $\binom{m}{i}\binom{n}{k-i}$ Möglichkeiten gibt, und diese zu einer k-elementigen Teilmenge von $\{1,...,m+n\}$ vereinigt.

12.12. Beispiel. ***(Poisson-Verteilungen)*** Die Poisson-Verteilung $Poi(\theta)$ mit Parameter $\theta > 0$ besitzt den Träger $\mathbb{N}_0$ und die Zähldichte

$$p_\theta(n) \;=\; e^{-\theta}\frac{\theta^n}{n!}, \quad n \in \mathbb{N}_0.$$

Seien nun X, Y zwei stochastisch unabhängige Zufallsgrößen mit $Poi(\theta_1)$- bzw. $Poi(\theta_2)$-Verteilung. Dann gilt

$$\begin{aligned}
P(X+Y=n) &= \sum_{k=0}^{n} P(X=k, Y=n-k) \\
&= \sum_{k=0}^{n} P(X=k)P(Y=n-k) \\
&= e^{-(\theta_1+\theta_2)}\sum_{k=0}^{n}\frac{\theta_1^k\theta_2^{n-k}}{k!(n-k)!} \\
&= \frac{e^{-(\theta_1+\theta_2)}}{n!}\sum_{k=0}^{n}\binom{n}{k}\theta_1^k\theta_2^{n-k} \\
&= e^{-(\theta_1+\theta_2)}\frac{(\theta_1+\theta_2)^n}{n!}
\end{aligned}$$

für alle $n \in \mathbb{N}_0$, wobei für die letzte Zeile offenkundig der binomische Lehrsatz benutzt wurde. $X+Y$ ist also $Poi(\theta_1+\theta_2)$-verteilt.

Als letztes wollen wir uns der Frage zuwenden, ob die stochastische Unabhängigkeit erhalten bleibt, wenn die jeweils gegebenen Zufallsvariablen durch Anwendung von Funktionen in neue überführt werden. Diese "Weiterverarbeitung" gegebener Daten tritt in vielen Situationen auf.

12.13. Satz. *Eine Familie $(X_i)_{i\in I}$ von diskreten Zufallsvariablen wie in 12.1 ist genau dann stochastisch unabhängig, wenn für jede Wahl von $\mathcal{Y}_i \neq \emptyset$ und jede Wahl von Abbildungen $f_i : \mathcal{X}_i \to \mathcal{Y}_i$, $i \in I$, die Familie $(f_i \circ X_i)_{i\in I}$ stochastisch unabhängig ist.*

Beweis: "$\Rightarrow$" Klar, da man f_i als Identität auf $\mathcal{X}_i$ für alle $i \in I$ wählen kann.

"$\Leftarrow$" Sei $(f_i)_{i\in I}$ eine beliebige Familie von Abbildungen von $\mathcal{X}$ nach $\mathcal{Y}$. Dann gilt für alle $n \leq |I|$ sowie alle paarweise verschiedenen $i_1,...,i_n \in I$ und $A_1 \subset \mathcal{Y}_{i_1},...,A_n \subset \mathcal{Y}_{i_n}$ unter

Benutzung von Satz 12.2

$$\begin{aligned} &P(f_{i_1} \circ X_{i_1} \in A_1, ..., f_{i_n} \circ X_{i_n} \in A_n) \\ &\qquad = P(X_{i_1} \in f_{i_1}^{-1}(A_1), ..., X_{i_n} \in f_{i_n}^{-1}(A_n) \\ &\qquad = P(X_{i_1} \in f_{i_1}^{-1}(A_1)) \cdot ... \cdot P(X_{i_n} \in f_{i_n}^{-1}(A_n)) \\ &\qquad = P(f_{i_1} \circ X_1 \in A_1) \cdot ... \cdot P(f_{i_n} \circ X_n \in A_n) \end{aligned} \tag{12.5}$$

und damit die Unabhängigkeit der Familie $(f_i \circ X_i)_{i \in I}$. ◊

Es bleibt noch die Frage, ob stochastische Unabhängigkeit auch dann erhalten bleibt, wenn die Zufallsergebnisse "abschnittsweise" verarbeitet werden, wobei anschaulich klar sein sollte, daß sich die Abschnitte nicht überlappen dürfen, da sonst gleiche Werte in mehrere Teile eingehen. Mit dieser Einschränkung läßt sich tatsächlich eine entsprechende Aussage beweisen. Wir gehen darauf aber nicht weiter ein, sondern verweisen auf Satz 23.5 in [1].

Literatur

[1] ALSMEYER, G. *Wahrscheinlichkeitstheorie (4. Auflage).* Skripten zur Mathematischen Statistik, Nr. 30, Universität Münster (2005).