Wirtschaftswissenschaftliches Zentrum
Universität Basel

Statistik
Dr. Thomas Zehrt

Bedingte Wahrscheinlichkeit

Motivation

Die bedingte Wahrscheinlichkeit eines Ereignisses A bei einem gegebenen Ereignis B (Schreibweise: $P(A|B)$) ist die Wahrscheinlichkeit von A, wenn wir die Menge aller zulässigen Ausgänge auf B einschränken. Wir wissen also, dass B eingetroffen ist und somit $P(B^c) = 0$ gilt.

Leider haben viele Menschen die Neigung, bei bedingten Wahrscheinlichkeiten Bedingung (B) und bedingtes Ereignis (A) durcheinander zu bringen. Ein Beispiel aus der Presse mag das illustrieren:

> ,,Schäferhunde sind gefährlich! " Denn ,,Jeder dritte Biss geht auf das Konto dieser Rasse."

Das Faktum dieser Aussage ist die Angabe der bedingten Wahrscheinlichkeit

$$P(\text{ Schäferhund } | \text{ Biss }) \;=\; \frac{1}{3},$$

und dieser Wert scheint bei der Vielzahl von Hunderassen recht gross. Der Journalist meint aber daraus sofort ableiten zu können, dass auch die bedingte Wahrscheinlichkeit $P(\text{ Biss } | \text{ Schäferhund })$ gross sein muss, dass Schäferhunde somit gefährlich sind.

Das geht so nicht! Wir müssen hier in Betracht ziehen, dass es sehr viele Schäferhunde (im Vergleich zu anderen Hunderassen) gibt und um einen Zusammenhang zwischen beiden bedingten Wahrscheinlichkeiten herstellen zu können, brauchen wir insbesondere Informationen über den Anteil der Schäferhunde innerhalb der Menge aller Hunde.

Das Umrechnen von bedingten Wahrscheinlichkeiten will also gelernt sein.

Benötigtes Schulwissen

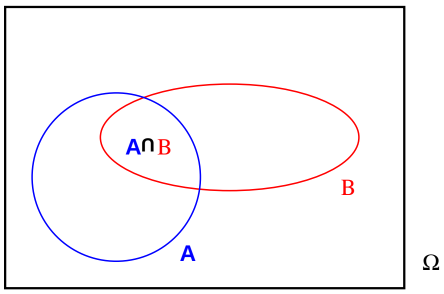
Mengenlehre

1 Grundlagen

Oft steht bei einem Zufallsexperiment schon die Information zur Verfügung, dass das Ergebnis zu einer bestimmten Teilmenge $B \subset \Omega$ des Ereignisraumes gehört.

Beispiel 1.1 *Ein Skatspieler sieht seine eigenen* 10 *Karten. Möchte er also wissen, mit welcher Wahrscheinlichkeit einer der anderen Spieler* 2 *Asse hat, so sollte er zunächst seine eigenen Asse zählen. Hat er selbst* 3 *oder* 4 *Asse, so ist die gesuchte Wahrscheinlichkeit natürlich null, hat er maximal* 2 *Asse, so ist sie positiv.*

Wir wollen nun den Begriff der bedingten Wahrscheinlichkeit zunächst für **Laplace-Experimente** einführen. Bei solchen Experimenten waren ursprünglich alle Ergebnisse $\omega \in \Omega$ gleichwahrscheinlich. Wissen wir nun, dass $\omega \in B \subset \Omega$ gilt, so ordnen wir allen Ergebnissen aus dem Komplement B^c von B die bedingte Wahrscheinlichkeit 0 zu und betrachten die Ergebnisse aus B als gleichwahrscheinlich unter der bedingten Wahrscheinlichkeit.



Für jedes Ereignis $A \subset \Omega$ lässt sich die bedingte Wahrscheinlichkeit $P(A|B)$ von A bei gegebenen B wie folgt schreiben:

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

Ausserdem gilt natürlich $P(A \cap B) = |A \cap B|/|\Omega|$ und $P(B) = |B|/|\Omega|$. Nehmen wir weiterhin noch an, dass $P(B) > 0$ so ergibt sich:

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

Definition 1.1 *Wir definieren für* **beliebige** *Wahrscheinlichkeitsräume* (Ω, P) *und für beliebige Ereignisse* B *mit* $P(B) > 0$ *die* <u>*bedingte Wahrscheinlichkeit*</u> $P(A|B)$ *von* A *bei gegebenen* B *durch*

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

Beispiel 1.2 *Wir betrachten eine Urne mit zwei weissen und drei schwarzen Kugeln und wir ziehen zwei Kugeln ohne Zurücklegen. Wir geben den weissen Kugeln die Nummern* $1, 2$ *und den schwarzen Kugeln die Nummern* $3, 4, 5$*. Der gesamte Stichprobenraum ist dann*

$$\begin{aligned}\Omega &= \{ (i,j) \ : \ i \neq j, 1 \leq i,j \leq 5 \} \\ &= \{ (1,2),(2,1),\ldots,(4,5),(5,4) \},\end{aligned}$$

und jede der möglichen Ziehungen ist gleichwahrscheinlich. Sei A *das Ereignis, dass die zweite gezogene Kugel schwarz ist und* B *das Ereignis, dass die erste gezogene Kugel weiss ist. Wie gross ist* $P(A|B)$*?*

- **Überlegung:** *Ist die erste gezogene Kugel weiss* (B)*, so sind in der Urne noch eine weisse und drei schwarze Kugeln. Ziehen wir nun eine zweite Kugel, so ist deshalb die Wahrscheinlichkeit, dass wir eine schwarze ziehen gleich* $3/4$ $(= P(A|B))$*.*

- **Nutzung der Formel:** *Wir sehen*

$$\begin{aligned}A \cap B &= \{ (1,3),(1,4),(1,5),(2,3),(2,4),(2,5) \} \\ A &= \{ (1,3),(2,3),(4,3),(5,3),(1,4),(2,4),(3,4),(5,4),(1,5),(2,5),(3,5),(4,5) \} \\ B &= \{ (1,2),(1,3),(1,4),(1,5),(2,1),(2,3),(2,4),(2,5) \}\end{aligned}$$

also $|A \cap B| = 6$, $|A| = 12$ *und* $|B| = 8$*. Somit folgt*

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{6}{8} = \frac{3}{4}.$$

- *Wir können auch* $P(B|A)$ *bestimmen, d.h. die Wahrscheinlichkeit dafür, dass die erste gezogene Kugel weiss ist unter der Bedingung, dass die zweite gezogene Kugel schwarz ist:*

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{6}{12} = \frac{1}{2}.$$

Aufgabe 1.1

Bestimmen Sie die Wahrscheinlichkeit, mit einem Laplace-Würfel eine 2 zu würfeln unter der Bedingung, dass die gewürfelte Zahl gerade ist.

Lösung: (1/3)

Aufgabe 1.2 *In einem Restaurant essen mittags gewöhnlich 40% der Gäste keine Vorspeise, 35% der Gäste keinen Nachtisch und 15% der Gäste weder Vorspeise noch Nachtisch. Wie gross ist die Wahrscheinlichkeit dafür, dass*

- *eine Gast, der keinen Nachtisch wählt, auch keine Vorspeise nimmt,*
- *eine Gast, der eine Vorspeise gewählt hat, auch noch einen Nachtisch nimmt?*

Hinweis: Nutzen Sie die Formel $(V \cup N)^c \;=\; V^c \cap N^c$.

Lösung:

Der Aufgabe entnimmt man:

$P(V^c) = 0.4$, $P(N^c) = 0.35$ und $P(V^c \cap N^c) = P((V \cup N)^c) = 0.15 = 1 - P(V \cup N)$.

Daraus folgt:

$$P(V \cap N) \;=\; P(V) + P(N) - P(V \cup N) \;=\; 0.6 + 0.65 - 0.85 \;=\; 0.4$$

Somit:

$$\begin{aligned} P(V^c|N^c) &= \frac{P(V^c \cap N^c)}{P(N^c)} = \frac{0.15}{0.35} = \frac{3}{7} \\ P(N|V) &= \frac{P(V \cap N)}{P(V)} = \frac{0.4}{0.6} = \frac{2}{3} \end{aligned}$$

2 Wichtige Sätze

2.1 Der Multiplikationssatz

In der Praxis wird meist $P(A \cap B)$ aus $P(B)$ und $P(A|B)$ berechnet, d.h. wir nutzen die Definition der bedingten Wahrscheinlichkeit in der Form $P(A \cap B) = P(B) \cdot P(A|B)$.

Aufgabe 2.1 *In einer Lostrommel befinden sich* 10 *Lose von denen genau* 6 *Nieten sind. Sie ziehen genau zwei Lose (ohne Zurücklegen). Mit welcher Wahrscheinlichkeit habe Sie genau* 2 *Nieten gezogen?*

Lösung:

Abkürzungen: $i.Z.N.$ = i. Ziehung Niete

$$P(1.Z.N. \cap 2.Z.N.) = P(1.Z.N) \cdot P(2.Z.N.|1.Z.N.) = \frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3}$$

Satz 1 (Allgemeiner Multiplikationssatz) *Sind* $A_1, A_2, \ldots, A_k \subset \Omega$ *Ereignisse mit* $P(A_1 \cap \ldots \cap A_{k-1}) > 0$*, so gilt*

$$\boxed{P(A_1 \cap \ldots \cap A_k) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \ldots \cdot P(A_k|A_1 \cap \ldots \cap A_{k-1})}$$

Beweis:

Wir führen den Beweis nur für drei Ereignisse $A_1, A_2, A_3 \subset \Omega$ mit $P(A_1 \cap A_2) > 0$ und wenden die Gleichung $P(A \cap B) = P(B) \cdot P(A|B)$ mehrfach an.

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P((A_1 \cap A_2) \cap A_3) \\ &= P(A_2 \cap A_1) \cdot P(A_3|A_1 \cap A_2) \\ &= P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2). \end{aligned}$$

$\square$

Aufgabe 2.2 *In einer Kiste mit* 100 *Überraschungseiern befinden sich genau* 15*, die eine Figur enthalten. Sie entnehmen der Kiste drei Überraschungseier (ohne Zurücklegen). Bestimmen Sie die Wahrscheinlichkeit der vier Ereignisse:* $E_i :=$ *,,genau* i *der Eier enthalten eine Figur,, für* $i = 0, 1, 2, 3$.

Lösung: ($P(E_3) = 0.00281$, $P(E_2) = 0.05519$, $P(E_1) = 0.33117$, $P(E_0) = 0.61082$)

Genauer:

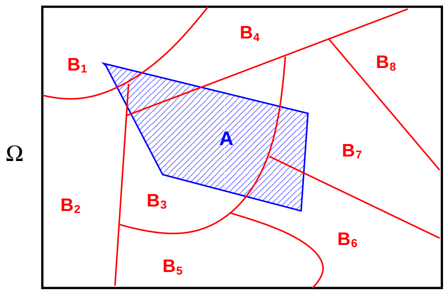
$$\begin{aligned}
P(E_2) &= P(FFF^c) + P(FF^cF) + P(F^cFF) \\
&= \quad P(1.Z.F.) \cdot P(2.Z.F.|1.Z.F.) \cdot P(3.Z.F^c.|1.Z.F. \cap 2.Z.F.) \\
&\quad + P(1.Z.F.) \cdot P(2.Z.F^c.|1.Z.F.) \cdot P(3.Z.F|1.Z.F. \cap 2.Z.F^c.) \\
&\quad + P(1.Z.F^c.) \cdot P(2.Z.F.|1.Z.F^c) \cdot P(3.Z.F.|1.Z.F^c. \cap 2.Z.F.) \\
&= \frac{15}{100} \cdot \frac{14}{99} \cdot \frac{85}{98} + \frac{15}{100} \cdot \frac{85}{99} \cdot \frac{14}{98} + \frac{85}{100} \cdot \frac{15}{99} \cdot \frac{14}{98} \\
&= 0.05519.
\end{aligned}$$

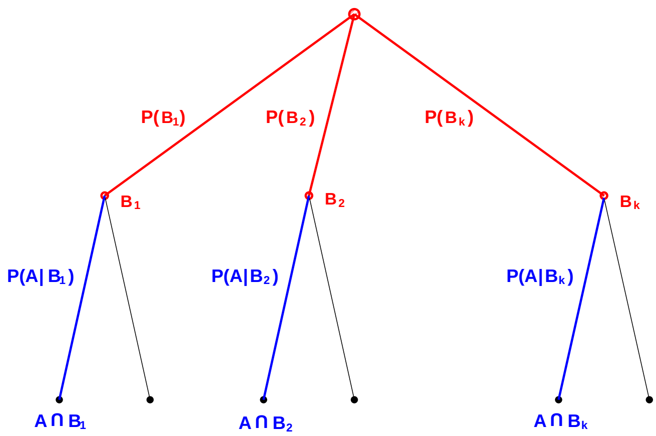
2.2 Totale Wahrscheinlichkeit und Satz von Bayes

Satz 2 *Seien $B_1, \ldots, B_k \subset \Omega$ paarweise unvereinbare Ereignisse (d.h. $B_i \cap B_j = \emptyset$ für alle $i \neq j$) deren Vereinigung ganz Ω ist. Weiterhin gelte $P(B_i) > 0$ für alle i. Dann gilt für ein Ereignis A*

1. **Formel von der totalen Wahrscheinlichkeit**

$$\boxed{P(A) \;=\; \sum_{i=1}^{k} P(A \cap B_i) \;=\; \sum_{i=1}^{k} P(B_i) \cdot P(A|B_i)}$$





2. **Formel von Bayes**

$$\boxed{P(B_i|A) \;=\; \frac{P(A \cap B_i)}{P(A)} \;=\; \frac{P(B_i)\, P(A|B_i)}{P(A)}}$$

falls $P(A) > 0$ ist.

Beweis: Es gilt:

$$\begin{aligned} P(A) &= P((A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)) \\ &= P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n) \\ &= P(B_1) \cdot P(A|B_1) + P(B_2) \cdot P(A|B_2) + \ldots + P(B_n) \cdot P(A|B_n). \end{aligned}$$

Für die Formel von Bayes nutzen wir zunächst die Identiäten

$$P(A \cap B_i) \;=\; P(B_i) \cdot P(A|B_i) \;=\; P(A) \cdot P(B_i|A).$$

Aus der zweiten Gleichung folgt dann sofort:

$$P(B_i|A) \;=\; \frac{P(B_i) \cdot P(A|B_i)}{P(A)}.$$

□

Spezialfall Ist $k = 2$, d.h. $B_1 = B$ und $B_2 = B^c$ vereinfachen sich die beiden Formeln wie folgt:

$$P(A) = P(B) \cdot P(A|B) + P(B^c) \cdot P(A|B^c)$$

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

Insbesondere können wir an der Formel von Bayes auch erkennen, wie man die bedingten Wahrscheinlichkeiten $P(B|A)$ und $P(A|B)$ umrechnen muss:

$$P(B|A) = \frac{P(B)}{P(A)} \cdot P(A|B).$$

Machen Sie sich bewusst, dass der Term $P(B)/P(A)$ sehr gross oder auch sehr klein sein könnte, i.A. wissen wir $0 < P(B)/P(A) < \infty$. Nur wenn $P(B)/P(A) \approx 1$ gilt wäre auch $P(B|A) \approx P(A|B)$.

Aufgabe 2.3 (Test für eine seltene Krankheit) *Eine Krankheit kommt bei etwa* 0.5% *der Bevölkerung vor. Ein Test zur Auffindung der Krankheit führt bei* 99% *der Kranken zu einer (positiven) Reaktion, aber auch bei* 2% *der Gesunden. Wie gross ist die Wahrscheinlichkeit, dass eine Person, bei der die Testreaktion auftritt, die Krankheit wirklich hat?*

Lösung:

Wir denken uns die Bevölkerung mit $\{1, 2, \ldots, N\}$ nummeriert. Die Menge aller Kranken sei mit B und die aller Gesunden mit B^c bezeichnet ($B \cap B^c = \emptyset$ und $B \cup B^c =$ gesamte Bevölkerung). Sicher gilt

$$|B| \approx \frac{5}{1000} N \qquad \text{und} \qquad |B^c| \approx \frac{995}{1000} N.$$

Bei zufälliger Auswahl einer Person aus der Bevölkerung ist jeder Person die Wahrscheinlichkeit $1/N$ zuzuordnen. Somit ergibt sich

$$P(B) = \frac{1}{N} \cdot \frac{5}{1000} \cdot N = \frac{5}{1000}$$

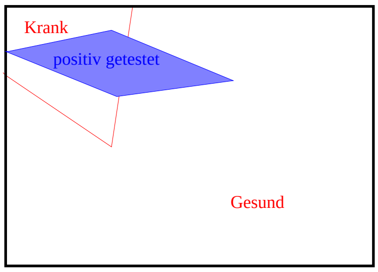
$$P(B^c) = \frac{1}{N} \cdot \frac{995}{1000} \cdot N = \frac{995}{1000}$$

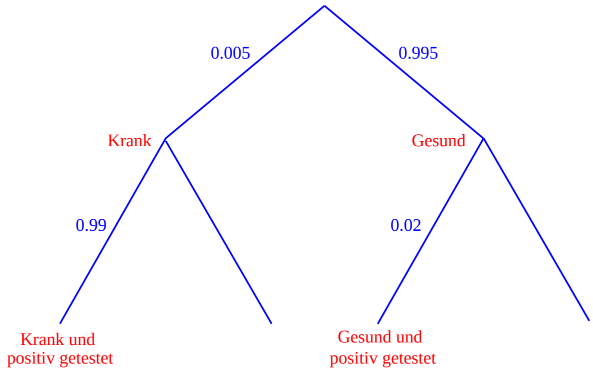
Mit A sei die Teilmenge der Personen bezeichnet, bei denen der Test zu einer (positiven) Reaktion führt. Die beiden Angaben zur Sicherheit des Testes sind bedingte Wahrscheinlichkeiten, also

$$P(A|B) = \frac{99}{100} = P(\text{ Test positiv } | \text{ krank })$$

$$P(A|B^c) = \frac{2}{100} = P(\text{ Test positiv } | \text{ gesund })$$

Natürlich sind diese beiden bedingten Wahrscheinlichkeiten ganz interessant, der Test scheint recht sicher zu arbeiten, aber eine wirklich wichtige Information wäre die Wahrscheinlichkeit dass man (wirklich) krank ist, falls der Test positiv testet.

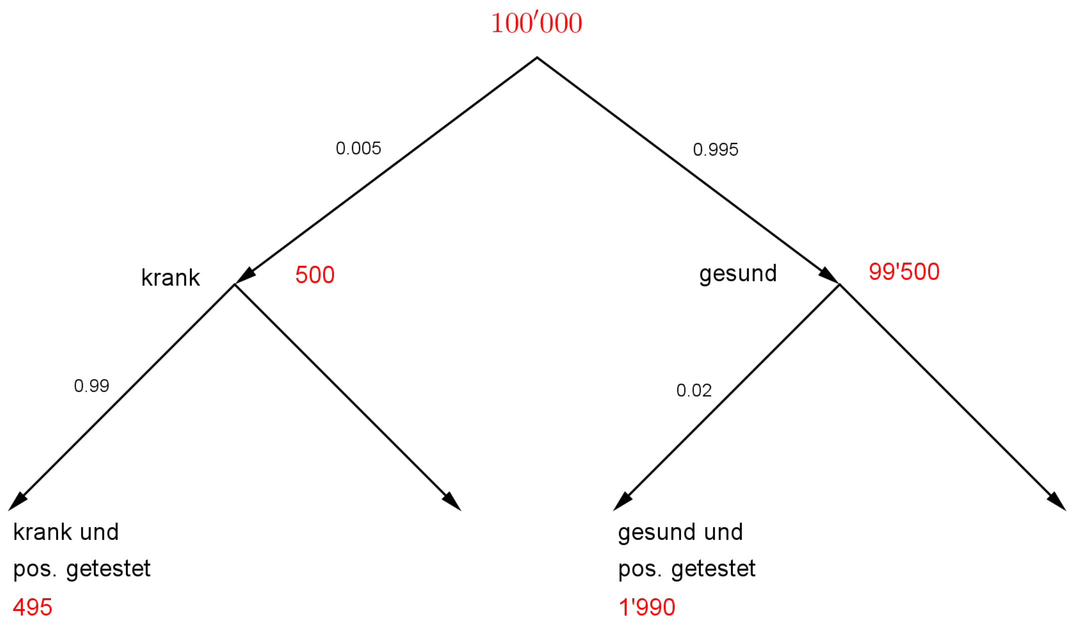




Die gesuchte Wahrscheinlichkeit ist also $P(B|A)$ und es folgt

$$\begin{aligned} P(B|A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|B^c) \cdot P(B^c)} \\ &= \frac{\frac{99}{100} \cdot \frac{5}{1000}}{\frac{99}{100} \cdot \frac{5}{1000} + \frac{2}{100} \cdot \frac{995}{1000}} = \frac{495}{2485} \approx 0.2 \end{aligned}$$

Wir wollen das an einem konkreten Zahlenbeispiel (mit 100'000 zufällig ausgewählten Personen) nochmals durchspielen.



Man erkennt: Von den insgesamt $495 + 1'990$ positiv getesteten Personen sind nur 495 wirklich krank und das entspricht einem Anteil von

$$\frac{495}{2485} \approx 0.2.$$

3 Unabhängigkeit

Es **kann** sein, dass das Eintreffen eines Ereignisses B keinen Einfluss auf die Wahrscheinlichkeit eines Ereignisses A hat: $P(A|B) = P(A)$. Falls dann noch $P(B) > 0$ gilt, so folgt aus der Definition der bedingten Wahrscheinlichkeit:

$$P(A \cap B) \quad = \quad P(A|B) \cdot P(B) \; = \; P(A) \cdot P(B)$$

Ist $P(B) > 0$ so sind die Gleichungen $P(A|B) = P(A)$ und $P(A \cap B) = P(A) \cdot P(B)$ äquivalent. Trotzdem wird zur Definition der so genannten stochastischen Unabhängigkeit meist die zweite genutzt, da sie in den beiden Ereignissen A und B ,,symmetrisch,, ist. Hinsichtlich des Verständnisses für den Begriff der stochastischen Unabhängigkeit ist aber sicher die erste Gleichung besser geeignet.

Definition 3.1 *Zwei Ereignisse A und B heissen (stochastisch) unabhängig, wenn die folgende Gleichung gilt*

$$\boxed{P(A \cap B) = P(A) \cdot P(B)}$$

Die Unabhängigkeit von zwei Ereignissen A und B drückt aus, dass A und B im **wahrscheinlichkeitstheoretischen Sinne** keinerlei Einfluss aufeinander haben. Das muss sehr genau von realen Beeinflussungen unterschieden werden. Andererseits: Zwei Ereignisse A und B können selbst dann unabhängig sein, wenn real das Eintreten von A davon abhängt, ob B geschieht.

Aufgabe 3.1 *Wir betrachten das Experiment, das aus zwei nacheinander ausgeführten Würfen eines Würfels besteht. A sei das Ereignis, dass die Summe der beiden geworfenen Augenzahlen gerade ist und B sei das Ereignis, dass die Augenzahl des zweiten Wurfes gerade ist. Sicher entscheidet hier das Ereignis B mit, ob A eintritt. Sind A und B stochastisch unabhängig?*

Lösung: Wahrscheinlich können Sie die vorkommenden Mengen auch schneller zählen (und damit die Wahrscheinlichkeiten der zugehörigen Ereignisse berechnen.

$$\begin{aligned}
\Omega &= \{\ (i,j) \mid i,j \in \{1,\ldots,6\}\ \} \text{ und } |\Omega| = 36 \\
A &= \{\ (1,1),(1,3),(1,5),(2,2),(2,4),(2,6),(3,1),(3,3),(3,5), \\
&\quad (4,2),(4,4),(4,6)(5,1),(5,3),(5,5),(6,2),(6,4),(6,6)\ \} \text{ und } |A| = 18 \\
B &= \{\ (1,2),(1,4),(1,6),(2,2),(2,4),(2,6),(3,2),(3,4),(3,6), \\
&\quad (4,2),(4,4),(4,6),(5,2),(5,4),(5,6),(6,2),(6,4),(6,6)\ \} \text{ und } |B| = 18 \\
A \cap B &= \{\ (2,2),(2,4),(2,6),(4,2),(4,4),(4,6),(6,2),(6,4),(6,6)\ \} \text{ und } |A \cap B| = 9
\end{aligned}$$

Und tatsächlich gilt (rechnen Sie das nach):

$$\underbrace{P(A)}_{=18/36} \cdot \underbrace{P(B)}_{=18/36} \quad = \quad \underbrace{P(A \cap B)}_{=9/36}$$

Definition 3.2 *Die n Ereignisse $A_1, A_2, \ldots, A_n \subset \Omega$ heissen stochastisch unabhängig, wenn für* **jede** *Auswahl $A_{i_1}, A_{i_2}, \ldots, A_{i_m}$ mit $m \leq n$ die folgende Gleichung gilt:*

$$\boxed{P(A_{i_1} \cap \ldots \cap A_{i_m}) = P(A_{i_1}) \cdot \ldots \cdot P(A_{i_m})}$$

Aufgabe 3.2 *Wieviele Gleichungen muss man überprüfen, um zu zeigen, dass drei Ereignisse A_1, A_2, A_3 unabhängig sind?*

Lösung: Sie müssen 4 Gleichungen überprüfen:

$$\begin{aligned}
P(A_1 \cap A_2) &= P(A_1) \cdot P(A_2) \\
P(A_1 \cap A_3) &= P(A_1) \cdot P(A_3) \\
P(A_2 \cap A_3) &= P(A_2) \cdot P(A_3) \\
P(A_1 \cap A_2 \cap A_3) &= P(A_1) \cdot P(A_2) \cdot P(A_3)
\end{aligned}$$

Wären (nur) die ersten drei Gleichungen korrekt, würde man die drei Ereignisse paarweise stochastisch unabhängig nennen.

4 *Die Entropie*

Die durch den amerikanischen Ingenieur C.E. Shannon (1916-2001) begründete bahnbrechende mathematische Disziplin, die heute allgemein als **Informationstheorie** bezeichnet wird, beschäftigt sich mit dem folgenden Problem:

Es bezeichne Z irgendein Zufallsexperiment mit dem zugehörigen Wahrscheinlichkeitsraum (Ω, P) mit $\Omega = \{\omega_1, \ldots, \omega_n\}$, $p_i = P(\omega_i)$ für $i = 1, \ldots, n$ und $\mathbf{p} = (p_1, \ldots, p_n)$. Bevor das Experiment ausgeführt wird herrscht **Unsicherheit** über seinen Ausgang, Wie können wir diese Unsicherheit messen?

Statt Unsicherheit gebraucht man das Wort **Entropie** (grieschisch für ,,Wendung" und ursprünglich zur Bezeichnung einer thermodynamischen Zustandsgrösse verwendet). Wir möchten also eine Zahl $H(Z)$, die Entropie des Zufallsexperimentes Z, definieren die ein Mass für die Unsicherheit sein soll. Das am wenigsten unbestimmte Zufallsexperiment ist eines, dessen Ausgang von vornherein fest steht. Einem solchen Experiment muss man natürlich die Entropie 0 (keine Unsicherheit) zuschreiben. Sicher sollte also für irgend ein Zufallsexperiment Z folgendes gelten:

$$0 \leq \quad H(Z) \quad \leq \infty.$$

Weiterhin sollte klar sein, dass die Natur des Versuches bei der Definition der Entropie nicht von Bedeutung sein sollte. Die Unsicherheit darüber, ob ein Mädchen geboren wird sollte genau so gross sein wie die bei einem Münzwurf Kopf zu werfen. Die Entropie dollte also nur von den Wahrscheinlichkeiten $\mathbf{p} = (p_1, \ldots, p_n)$ abhängen:

$$H(Z) \quad = \quad H(p_1, \ldots, p_n).$$

Nun gibt es mindestens drei Wege, um zur Definition der Entropie zu gelangen.

1. Der erste (direkte) Weg wäre, einfach die Definition

$$H(Z) \quad = \quad -\sum_{i=1}^{n} p_i \ \log_2(p_i)$$

   hinzuschreiben.

2. Der zweite (axiomatische) Weg versucht, H durch Eigenschaften (Axiome) festzulegen, die H auf jeden Fall haben sollte.

3. Der dritte (deskriptive) Weg, der von einer Interpretation von H ausgeht, ist sicher der mit Abstand beste aber auch aufwändigste. Ich möchte diesen Weg kurz skizzieren:

   - In einem esten Schritt führt man die so genannte wirkliche Entropie $H_0(Z)$ ein:
     $H_0(Z) :=$ durchschnittliche Anzahl (Erwartungswert) von ja/nein-Fragen,
     die man stellen muss, um den tatsächlichen Ausgang von Z zu erfahren,
     wenn man eine optimale Fragestrategie verwendet.
   - Dann führen wir Z unabhängig k-mal nacheinander aus und bezeichnen dieses neue Gesamtexperiment mit Z^k.

- Wir definieren nun die Entropie von Z durch:

$$H(Z) \;=\; \lim_{k\to\infty} \frac{1}{k}\, H_0(Z^k).$$

- Zum Schluss wird bewiesen, dass folgendes gilt:

$$H(Z) \;=\; \lim_{k\to\infty} \frac{1}{k}\, H_0(Z^k) \;=\; -\sum_{i=1}^{n} p_i\, \log_2(p_i)$$

Auf diesem Weg erhält man am Ende den folgenden Satz.

Satz 3 (1. Hauptsatz der Informationstheorie) *Für ein Zufallsexperiment Z mit Wahrscheinlichkeitsraum (Ω, P) mit $\Omega = \{\omega_1, \ldots, \omega_n\}$, $p_i = P(\omega_i)$ für $i = 1, \ldots, n$ ist die Entropie durch*

$$H(Z) \;=\; \lim_{k\to\infty} \frac{1}{k}\, H_0(Z^k).$$

wohldefiniert (d.h. der Grenzwert existiert, was nicht offensichtlich ist) und es gilt

$$H(Z) \;=\; H(p_1, \ldots, p_n) \;=\; -\sum_{i=1}^{n} p_i\, \log_2(p_i).$$

Aufgabe 4.1

1. *Berechnen Sie $H(0,1)$, $H(\frac{1}{2}, \frac{1}{2})$, $H(\frac{1}{3}, \frac{2}{3})$, $H(1, 0, \ldots, 0)$ und $H(\frac{1}{n}, \ldots, \frac{1}{n})$.*
2. *Diskutieren und skizzieren Sie die Funktion $f(p) = H(p, 1-p)$ auf dem Intervall $[0,1]$. Für welche p ist die Funktion (Entropie) extremal?*

5 Übungsaufgaben

1. Es werde mit zwei Würfeln einmal gewürfelt. Für die zufälligen Ereignisse

   - A: ,,Die Augensumme ist 7.,,
   - B: ,,Unter den Augenzahlen befinden sich keine 2 und keine 5.,,
   - C: ,,Eine Augenzahl ist gerade und die andere Augenzahl ist ungerade.,,

   bestimme man $P(A)$, $P(B)$, $P(C)$, $P(A|B)$, $P(B|A)$, $P(A|C)$, $P(C|A)$, $P(B|C)$, $P(C|B)$, $P(A|B \cap C)$ und $P(B|A \cap C)$.

   Sind A und B, A und C bzw. B und C jeweils stochastisch unabhängig?

2. In einer Urne befinden sich acht gelbe und vier blaue Kugeln,

   (a) Es werden gleichzeitig (zufällig) drei Kugeln gezogen. Wie gross ist die Wahrscheinlichkeit, dass es sich um zwei gelbe und eine blaue Kugel handelt?

   (b) Eine Kugel wird zufällig gezogen und durch eine Kugel der anderen Farbe ersetzt. Nun mische man den Inhalt der Urne erneut und ziehe wieder zufällig eine Kugel. Wie gross ist die Wahrscheinlichkeit, dass dies eine blaue Kugel ist?

3. Aus drei Urnen U_1, U_2 und U_3 wird zufällig eine Urne ausgewählt, wobei jede Urne dieselbe Wahrscheinlichkeit besitzt, in die Auswahl zu gelangen. Die drei Urnen enthalten weisse und schwarze Kugeln, wobei sich in Urne

   - U_1 : zwei weisse und fünf schwarze
   - U_2 : vier weisse und vier schwarze
   - U_3 : sieben weisse und vier schwarze

   Kugeln befinden. Aus der zufällig gewählten Urne wird nun eine Kugel gezogen.

   (a) Wie gross ist die Wahrscheinlichkeit, dass die gezogenen Kugel weiss ist?

   (b) Die gezogene Kugel ist schwarz. Wie gross ist die Wahrscheinlichkeit, dass sie aus Urne U_2 stammt?

4. Ein Bäcker benötigt für die Herstellung seines Spezialbrotes vier verschiedene Mehlsorten, die er von vier verschiedenen Herstellern geliefert bekommt (die alle unabhängig voneinander produzieren).

   Er kann sein Brot nur dann verkaufen, wenn alle vier Mehlsorten einwandfrei sind. Für die vier Mehlsorten gilt, dass sie mit einer Wahrscheinlichkeit von 0.1, 0.05, 0.2 bzw. 0.15 Mängel aufweisen. Wie gross ist die Wahrscheinlichkeit dafür, dass der Bäcker sein Brot nicht verkaufen kann?

5. Ein Zufallsexperiment führe zu den zwei möglichen Ereignissen A und B. A und B seien stochastisch unabhängig. Es gilt $P(B) = 0.5$ und $P(A \cap B) = 0.2$. Wie gross ist $P(A \cup B)$?

Resultate einiger Übungsaufgaben

1. $P(A) = 6/36$, $P(B) = 16/36$, $P(C) = 18/36$, $P(A|B) = 1/4$, $P(B|A) = 2/3$, $P(A|C) = 1/3$, $P(C|A) = 1$, $P(B|C) = 4/9$, $P(C|B) = 1/2$, $P(A|B \cap C) = 1/2$ und $P(B|A \cap C) = 2/3$.

   A und B sind abhängig, denn
   $$P(A \cap B) = \frac{1}{9} \neq P(A) \cdot P(B) = \frac{1}{6} \cdot \frac{4}{9}.$$
   A und C sind abhängig, denn
   $$P(A \cap C) = \frac{1}{6} \neq P(A) \cdot P(C) = \frac{1}{6} \cdot \frac{1}{2}.$$
   B und C sind unabhängig, denn
   $$P(B \cap C) = \frac{2}{9} = P(B) \cdot P(C) = \frac{4}{9} \cdot \frac{1}{2}.$$

2. (a) 0.5091

   (b) 0.361

3. (a) 0.474

   (b) 0.317

4. 0.419

5. Die Gleichung $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ gilt immer. Die Gleichung $P(A \cap B) = P(A) \cdot P(B)$ gilt hier, da in der Aufgabe die Unabhängigkeit von A und B vorausgesetzt wird! Kombiniert man beide Gleichungen erhält man:
   $$P(A \cup B) = \frac{P(A \cap B)}{P(B)} + P(B) - P(A \cap B) = 0.7.$$