1 Elementare Wahrscheinlichkeitstheorie

1.1 Zufallsexperiment

Definition 1.1. Ein **Zufallsexperiment** ist ein "Vorgang", der im Prinzip beliebig oft unter identischen Randbedingungen wiederholt werden kann.

Damit ist natürlich jedes naturwissenschaftliche Experiment auch ein Zufallsexperiment (was ja auch stimmt: Bei jeder Messung werden Fehler gemacht!). Interessant sind für uns insbesondere solche Experimente, deren Ergebnisse aufgrund der Komplexität der Randbedingungen nicht mit Sicherheit vorhergesagt werden können. Die klassischen Beispiele sind der Münzwurf und der Würfelwurf.

Definition 1.2. Die Menge der möglichen Ergebnisse eines Zufallsexperimentes heißt der **Ereignisraum** Ω.

Beispiel 1.3.

1. Im Fall des Würfels ist $\Omega = \{1, 2, 3, 4, 5, 6\}$.
2. Beim Lotto "6 aus 49" ist $\Omega = \{A \subset \{1, \ldots, 49\} : |A| = 6\}$.
3. Beim Münzwurf gilt $\Omega = \{\text{Kopf, Zahl}\}$.
4. Wir können auch eine Reißzwecke werfen und die möglichen Ausgänge sind dann "bleibt auf dem Kopf liegen (Stachel zeigt nach oben)" oder "fällt auf die Seite". Hier ist von vornherein nicht klar, wie "wahrscheinlich" welcher Ausgang des Zufallsexperimentes ist: Man muss einfach oft genug eine Reißzwecke (oder einmal ganz viele) werfen und die "relativen Häufigkeiten" auszählen.
5. Wenn wir mit zwei Würfeln werfen und als Ausgang des Experimentes die Summe der Augenzahlen nehmen, dann haben wir ein Zufallsexperiment mit $\Omega = \{2, 3, \ldots, 12\}$.

Definition 1.4. Teilmengen des Ereignisraums Ω heißen **Ereignisse**. Sprechweise: Ist $A \subseteq \Omega$ und liegt das Ergebnis eines Zufallsexperimentes in A, so sagt man auch, das Ereignis A ist eingetreten.

Spezielle Ereignisse A bekommen spezielle Namen:

1. $|A| = 1$: Elementarereignis.
2. $A = \Omega$: Sicheres Ereignis
3. $A = \{\ \}$: Unmögliches Ereignis.
4. Sind A und B zwei Ereignisse mit $A \cap B = \{\ \}$, dann heißen die Ereignisse **disjunkt** oder **unvereinbar**.

5. Das Ereignis $\Omega \setminus A$ heißt das **Gegenereignis** von A. Bezeichnung auch $\overline{A}$.

Beispiel 1.5. Angenommen, wir würfeln mit zwei Würfeln gleichzeitig. Dann ist

$$\Omega = \{A \subseteq \{1,2,3,4,5,6\} \; : \; 1 \leq |A| \leq 2\}.$$

und $|\Omega| = \binom{6}{2} + \binom{6}{1} = \frac{6 \cdot 5}{2} + 6 = 15 + 6 = 21$, wobei die einelementigen Mengen gerade die Situation beschreiben, dass beide Würfel dieselbe Zahl zeigen. Wenn wir aber zweimal nacheinander würfeln, wäre ein angemessener Ereignisraum, um dieses Zufallsexperiment zu beschreiben,

$$\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}.$$

eine Menge mit 36 Elementen. Ein mögliches Ereignis wäre in beiden Fällen das Werfen eines Paschs. Im ersten Fall wären das die einelementigen Teilmengen, im zweiten Fall $\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$.

Da Ereignisse Teilmengen von Ω sind, kann man sie mengentheoretisch verknüpfen (Vereinigung, Schnitt, Komplement). Das Komplement haben wir schon betrachtet (Gegenereignis). Die Vereinigung $A \cup B$ beschreibt anschaulich das Ereignis "A oder B", und $A \cap B$ das Ereignis "A und B".

1.2 Wahrscheinlichkeitsraum

Beachten Sie, dass wir bislang noch gar nicht von Wahrscheinlichkeiten gesprochen haben. Das kommt erst jetzt: Dazu nehmen wir an, dass $|\Omega| < \infty$, also endlich ist, oder aber zumindest abzählbar unendlich, d.h. $|\Omega| = |\mathbb{N}|$. Im ersten Fall können wir die Ereignisse einfach als $\Omega = \{\omega_1, \ldots, \omega_n\}$ schreiben, im letzteren Fall wenigstens noch abzählen, also $\Omega = \{\omega_i \; : \; i \in \mathbb{N}\}$.

Definition 1.6. Sei Ω eine endliche oder abzählbar unendliche Menge, und $P : 2^\Omega \to \mathbb{R}$ eine Abbildung. Das Paar (Ω, P) heißt **Wahrscheinlichkeitsraum**, wenn gilt:

(K1) $P(A) \geq 0$ für alle $A \subset \Omega$.

(K2) $P(\Omega) = 1$.

(K3) Sind A_i, $i = 1, \ldots$ paarweise disjunkte Ereignisse, so gilt

$$P(A_1 \cup A_2 \cup \ldots) = P(A_1) + P(A_2) + \ldots.$$

Die Abbildung P heißt **Wahrscheinlichkeitsmaß**. Der Wert $P(A)$ heißt die Wahrscheinlichkeit für das Ereignis A.

Mit diesem Axiomensystem umgeht man das Problem, was eigentlich eine Wahrscheinlichkeit ist: Wahrscheinlichkeit ist einfach etwas, was (K1), (K2) und (K3) erfüllt. Wir sprechen im Fall der Definition 1.6 von einem diskreten Wahrscheinlichkeitsraum (weil $|\Omega|$ höchstens abzaählbar unendlich ist). Im Fall überabzählbarer Ereignisräume kann man nicht mehr sinnvoll jedem Ereignis (also jeder Teilmenge von Ω) eine W.K. zuordnen. Wir wollen dieses Problem hier nicht vertiefen; man sollte sich allerdings merken, dass es für überabzählbare Mengen Ω kein P gibt, das (K1), (K2), (K3) erfüllt und jeder Teilmenge von Ω einen Wert aus $\mathbb{R}$ zuordnet. In der Praxis genügt es aber, wenn wir allen Mengen A, die wir durch abzählbare Vereinigung und Schnitte und Komplementbildung aus abgeschlossenen Intervallen erhalten eine W.K. $P(A)$ zuordnen können, so dass (K1), (K2) und (K3) gilt.

Wir werden aber zunächst nur diskrete Wahrscheinlichkeitsräume behandeln; dort treten die gerade angesprochenen Probleme nicht auf, und wir können jeder Teilmenge des Ereignisraumes eine W.K. zuordnen, insbesondere auch den einelementigen Teilmengen. Es ist klar, dass den Wahrscheinlichkeiten dieser Elementarereignisse eine besondere Bedeutung zukommt. Kennt man nämlich $P(\{\omega\})$ (ab jetzt meistens nur mit $P(\omega)$ bezeichnet) für alle $\omega \in \Omega$, dann kennt man bereits das ganze Wahrscheinlichkeitsmaß (im abzählbaren Fall, also in dem Fall, den wir diskret genannt haben).

1.3 Laplace-Experiment

Das einfachste Wahrscheinlichkeitsmaß ist das folgende: Ω ist eine endliche Menge, und $P(\omega) = \frac{1}{|\Omega|}$. Man sieht rasch, dass in dem Fall gilt:

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

lax gesprochen: Anzahl günstiger Ausgänge geteilt durch die Anzahl möglicher Ausgänge eines Zufallsexperimentes.

Um dieses Wahrscheinlichkeitsmaß mit einem Zufallsexperiment in Verbindung zu bringen, müssen wir nun doch von einer naiven Vorstellung von Wahrscheinlichkeit ausgehen: W.K. gibt an, wie oft ein Ereignis ungefähr eintritt, wenn wir das Zufallsexperiment nur oft genug wiederholen.

Unter einem **Laplace-Experiment** verstehen wir nun ein Zufallsexperiment, bei dem das Ereignis A mit W.K. $\frac{|A|}{|\Omega|}$ eintritt.

Beispiel 1.7. Die Wahrscheinlichkeit, beim Würfeln mit einem Würfel eine bestimmte Zahl zu würfeln, ist 1/6, es liegt also ein Laplace-Experiment vor und so gilt z.B., dass die Wahrscheinlichkeit, eine gerade Zahl zu würfeln, 1/2 ist.

Beispiel 1.8. Mit welcher Wahrscheinlichkeit erhält man beim Würfeln mit zwei nicht unterscheidbaren Würfeln einen Pasch? Unser Ereignisraum hat hier die Mächtigkeit 21, und es gibt 6 Möglichkeiten für einen Pasch. Allerdings ist die Wahrscheinlichkeit nicht $\frac{6}{21}$, sondern $\frac{6}{36}$. Ähnlich ist die W.K., dass ein Würfel eine 1, ein anderer eine 2 zeigt, nicht $\frac{1}{21}$ sondern $\frac{2}{36}$. Es gibt also 15 Elementarereignisse, die jeweils mit Wahrscheinlichkeit $\frac{1}{18}$ eintreten, und 6 Elementarereignisse, die jeweils mit W.K. $\frac{1}{36}$ eintreten (die Paschs), in der Summe gibt das 1. Es handelt sich hierbei also nicht um ein Laplace-Experiment.

1.4 Beispiel eines unendlichen Ereignisraumes

Hier sei $\Omega = \{0, 1, \ldots\}$ und $\lambda > 0$. Die Wahrscheinlichkeit für das Eintreten eines Elementarereignisses $\{i\}$ sei

$$P(i) = \frac{\lambda^i}{i!} e^{-\lambda}$$

und

$$P(A) = \sum_{i \in A} P(i).$$

Offenbar gilt (K1) und (K3). Nicht ganz klar ist (K2): Benutze den bekannten Grenzwert

$$\sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{\lambda},$$

also

$$P(\Omega) = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} e^{-\lambda} = 1.$$

Wir werden diesem Wahrscheinlichkeitsmaß noch unter dem Stichwort "Poisson-Verteilung" begegnen. es handelt sich bei diesem Modell wieder nicht um ein Laplace-Experiment.

1.5 Elementare Eigenschaften eines Wahrscheinlichkeitsmaßes

Satz 1.9. *Sei (Ω, P) ein Wahrscheinlichkeitsraum (d.h. Ω ist höchstens abzählbar unendlich). Dann gilt:*

1. $P(\Omega \setminus A) = 1 - P(A)$.
2. $P(\{\ \}) = 0$.
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
4. *Wenn $A \subseteq B$, dann gilt $P(A) \leq P(B)$.*

Es ist wichtig zu bemerken, dass diese Eigenschaften aus den Axiomen in Definition 1.6 folgen, also beweisbar sind. Sie sind nicht Bestandteil des Axiomensystems.

1.6 Geburtstagsparadoxon: Anwendung in der Kryptographie

Angenommen, jedes Dokument wird mit einem sogenannten **Hashwert**, der eine Zahl h mit $1 \leq h \leq N$ ist, "unterschrieben". Diese Zahl berechnet sich aus dem Dokument (ist also so etwas wie eine Prüfziffer). Wie groß ist die W.K., dass bei n unterzeichneten Dokumenten alle verwendeten Hashwerte verschieden sind? Warum ist das wichtig? Hashwerte werden zur digitalen Signatur benutzt, d.h. eine Person (nennen wir sie Alice) unterzeichnet ein Dokument so, dass sie erst den Hashwert berechnet und dann den Hashwert digital unterschreibt. Der Empfänger, sagen wir Bob, kann überprüfen, ob die Signatur korrekt ist (dazu sagen wir hier nichts), und wenn ja, überprüft er, ob der Hashwert zum Dokument passt (denn sonst hätte unterwegs ja jemand das Dokument austauschen können!). Nun könnte aber ein "bad guy", sagen wir Eve, in betrügerischer Absicht ganz viele harmlose Varianten desjenigen Dokumentes erzeugen, das Alice unterschreiben soll. Gleichzeitig erzeugt Eve viele Varianten eines gefälschten Dokumentes, zusammen mit den Hashwerten. Wenn sie ein Paar (S, T) mit identischen Hashwerten gefunden hat, wobei S das richtige und T das gefälschte Dokument ist, legt sie S und den Hashwert von S Alice zur Unterschrift vor, und dann tauscht sie beim Senden an Bob den Text S durch T aus. Weil der Hashwert von T mit dem von S übereinstimmt, erkennt Bob die betrügerische Absicht von Eve nicht.

Wie groß ist also die W.K. für eine Kollision, wenn man N Hashwerte hat, d.h. wie leicht ist es für Eve, zwei Dokumente mit gleichem Hashwert zu finden?

Es werden n Dokumente zufällig gewählt, und dann sind auch die Hashwerte zufällig. Wir können uns das als ein Urnenexperiment vorstellen, wobei in der Urne N verschiedene Kugeln liegen (die Hashwerte) und es werden, mit Zurücklegen, n gezogen. Dann ist die Anzahl "günstiger" Ausgänge ohne Kollision

$$N(N-1)(N-2)\cdots(N-n+1) = \frac{N!}{(N-n)!}.$$

Die Anzahl aller möglichen Ziehungen ist N^n. Das liefert die gesuchte W.K.

$$\begin{aligned} P &= \frac{N!}{(N-n)!N^n} = \frac{N}{N} \cdot \frac{N-1}{N} \cdots \frac{N-n+1}{N} \\ &= \left(1 - \frac{0}{N}\right)\left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\cdots\left(1 - \frac{n-1}{N}\right). \end{aligned}$$

Ist n/N klein, benutzen wir $e^x \approx 1 + x$ und erhalten

$$P \approx e^{[-0-1-2-\ldots-(n-1)]/N} = e^{-\frac{n(n-1)}{2N}}.$$

Wann ist dieser Wert z.B. $1/2$? Er ist $1/2$ wenn

$$-\frac{n(n-1)}{2N} = \ln(\frac{1}{2}) = -\ln(2),$$

also

$$n^2 - n = 2N\ln(2)$$

gilt, d.h.

$$n = \frac{1}{2} + \sqrt{\frac{1}{4} + 2N\ln(2)} \approx \sqrt{2\ln(2)N}$$

für große N. Für $N = 365$ erhalten wir $n \approx 23$, also in einer Gruppe von 23 Menschen gibt es mit W.K. etwa $\frac{1}{2}$ zwei, die am selben Tag Geburtstag feiern (daher der Name Geburtstagsparadoxon).

1.7 Bedingte Wahrscheinlichkeiten

Definition 1.10. Zwei Ereignisse A und B heißen **unabhängig**, wenn

$$P(A \cap B) = P(A) \cdot P(B)$$

gilt.

Beachten Sie den Unterschied zu unvereinbaren Ereignissen A und B (da gilt $P(A \cap B) = 0$.

Anschaulich soll bedingte Wahrscheinlichkeit folgendes bedeuten: Wir wollen $P(B)$ bestimmen, wenn wir schon wissen, dass A eingetreten ist.

Definition 1.11. Es seien A und B Ereignisse eines Ereignisraums Ω, auf dem ein Wahrscheinlichkeitsmaß P definiert ist. Ferner sei $P(B) \neq 0$. Dann definieren wir

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

Gesprochen: P von A gegeben B. Wir nennen $P(A|B)$ die **bedingte** Wahrscheinlichkeit.

Bemerkung 1.12. Es gilt

$$\begin{aligned} P(A \cap B) &= P(B) \cdot P(A|B) \\ &= P(A) \cdot P(B|A). \end{aligned}$$

Satz 1.13 (Satz von Bayes)**.**

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

sofern $P(A), P(B) \neq 0$.

Wenn zwei Ereignisse A und B unabhängig voneinander sind, gilt

$$\begin{aligned} P(A|B) &= P(A) \\ P(B|A) &= P(B). \end{aligned}$$

Manchmal nennt man $P(A)$ die "a-priori" Wahrscheinlichkeit und $P(A|B)$ die "a-posteriori"-Wahrscheinlichkeit. Das Problem, bedingte Wahrscheinlichkeiten vernünftig zu interpretieren, ist folgendes: Von Wahrscheinlichkeiten reden wir, wenn ein Zufallsexperiment durchgeführt wird, und die W.K. sagt dann etwas aus über das Ergebnis eines Experimentes, das in der Zukunft liegt, also erst ausgeführt wird. Wenn uns jemand nach Durchführung des Zufallsexperimentes ein wenig Informationen gibt (also sagt, das Ereignis B sei eingetreten), haben wir ja gar kein Zufallsexperiment mehr! Man muss vielmehr **vor** der Durchführung des Experimentes vereinbaren, ob die Information B weitergegeben wird. Wir schauen uns also all die Ausgänge eines Zufallsexperimentes an, die eintreten, wenn auch B eingetreten ist und auch wirklich nur dann! Wir erhalten so einen neuen Wahrscheinlichkeitsraum:

Satz 1.14. *Sei (Ω, P) ein Wahrscheinlichkeitsraum und sei B ein Ereignis mit $P(B) \neq 0$. Dann definiert $P(A|B)$ für $A \subseteq \Omega$ ein Wahrscheinlichkeitsmaß auf Ω. Dieses neue Maß wird auch $P_B(A)$ bezeichnet.*

Beispiel 1.15. Angenommen wir würfeln mit einem Würfel und B sei das Ereignis "es wird eine gerade Zahl gewürfelt". Dann ist

$$\begin{aligned} P(\{1\}|B) &= 0 \\ P(\{2\}|B) &= 1/3 \\ P(\{3\}|B) &= 0 \\ P(\{4\}|B) &= 1/3 \\ P(\{5\}|B) &= 0 \\ P(\{6\}|B) &= 1/3. \end{aligned}$$

Satz 1.16 (Satz von der totalen Wahrscheinlichkeit)**.** *(Ω, P) sei ein Wahrscheinlichkeitsraum, $A_i \subseteq \Omega$ mit $1 \leq i \leq m$ paarweise disjunkte Ereignisse mit $\Omega = \bigcup_{i=1}^m A_i$. Ferner sei $B \subseteq \Omega$ sowie $P(B), P(A_i) \neq 0$ für $i = 1, \dots, m$. Dann gilt*

$$P(B) = \sum_{i=1}^{m} P(A_i) \cdot P(B|A_i) = \sum_{i=1}^{m} P(A_i \cap B)$$

sowie (Satz von Bayes*)*

$$P(A_k|B) = \frac{P(A)}{P(B)} \cdot P(B|A_k) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^m P(A_i) \cdot P(B|A_i)}.$$

Beispiel 1.17. Nehmen wir an, wir leben in einem Land, in dem jede Familie genau zwei Kinder hat, jeweils 1/4 der Familien haben zwei Jungen und zwei Mädchen, und bei jeweils 1/4 der Familien ist die Verteilung mj und jm, wobei im ersten Fall das Mädchen (m) das erstgeborene Kind ist, im zweiten Fall der Junge (j). Es gilt hier also

$$\Omega = \{mm, jj, mj, jm\}$$

und wir haben, wenn eine Familie zufällig ausgewählt wird, ein Laplace-Experiment. Nun nehmen wir an, dass in dem Land einem Besucher stets (falls möglich) zuerst die Tochter vorgestellt wird. Wenn man also eine Familie besucht und es wird eine Tochter vorgestellt, erhält man Teilinformationen, nämlich man weiß, dass das Ereignis

$$B = \{mj, jm, mm\}$$

eingetreten ist. Das liefert

$$\begin{aligned} P_B(mm) &= 1/3 \\ P_B(mj) &= 1/3 \\ P_B(jm) &= 1/3 \\ P_B(jj) &= 0. \end{aligned}$$

Das bedeutet, mit Wahrscheinlichkeit 2/3 handelt es sich bei der Familie um eine mit einem Mädchen und einem Jungen.

Nun ändern sich die Traditionen in dem Land, und einem Besucher wird stets der/die Erstgeborene vorgestellt. Angenommen, dann wird uns eine Tochter vorgestellt. Das Ereignis C wäre dann

$$C = \{mm, mj\}$$

und wir erhalten

$$\begin{aligned} P_C(mm) &= 1/2 \\ P_C(mj) &= 1/2 \\ P_C(jm) &= 0 \\ P_C(jj) &= 0. \end{aligned}$$

Damit ist nun die Wahrscheinlichkeit für eine Familie mit einem Mädchen und einem Jungen genau 1/2.

Nun ändern sich die Traditionen noch mehr und die Eltern schnappen sich irgendein Kind, das sie dem Besucher als erstes vorstellen. Angenommen, das ist ein Mädchen. Wie groß ist jetzt die W.K., dass das zweite Kind ein Junge ist? Oft wird hier so argumentiert wie im ersten Fall. Andererseits hat man intuitiv das Gefühl, die W.K. für das Geschlecht des zweiten Kindes sollte unabhängig sein vom Geschlecht eines zufällig beobachteten Kindes. Wie können wir hier vorgehen? Gehen wir zu einem gößeren Ereignisraum über:

$$\Omega_e := \Omega \times \Omega_v,$$

wobei $\Omega_v = \{j, m\}$ das Geschlecht des Kindes ist, das vorgestellt wird. Wir erhalten

$$\begin{aligned} P(mm,m) &= 1/4 \\ P(mm,j) &= 0 \\ P(mj,m) &= 1/8 \\ P(mj,j) &= 1/8 \\ P(jm,m) &= 1/8 \\ P(jm,j) &= 1/8 \\ P(jj,m) &= 0 \\ P(jj,j) &= 1/4. \end{aligned}$$

Die Wahrscheinlichkeit, dass uns ein Mädchen vorgestellt wird (nennen wir das Ereignis D) ist $1/2$. Wenn A das Ereignis bezeichnet "Ein Kind ist ein Junge", so ist die Wahrscheinlichkeit $P(A \cap D) = 1/4$, somit $P(A|D) = 1/2$.

Beispiel 1.18. In einer Urne liegen je zwei rote, schwarze und blaue Kugeln. Es wird vereinbart, dass eine Person I nacheinander ohne Zurücklegen Kugeln aus der Urne zieht und einer anderen Person mitteilt, wann erstmals eine blaue Kugel gezogen wird. Diese Person II sieht aber nicht, welche Kugeln gezogen wurden. Angenommen die dritte Kugel ist blau (erstmals!). Mit welcher W.K. wurden dann in den ersten beiden Ziehungen die beiden roten Kugeln gezogen? Ohne die Zusatzinformation ist die W.K., in den ersten beiden Ziehungen die beiden roten Kugeln zu ziehen,

$$p = \frac{1}{3} \cdot \frac{1}{5} = \frac{1}{15}.$$

In der Vorlesung zeigen wir, dass die W.K., dass die ersten beiden Kugeln rot sind, $1/6$ ist unter der Annahme, dass die 3. Kugel blau ist.

Beispiel 1.19. In einer Bevölkerungsgruppe seien 0.1% der Bevölkerung mit einem Virus infiziert, der Rest ist nicht infiziert. Ein Test habe eine zuverlässigkeit von 99%, d.h. er liefert in 1% der Fälle ein falsches Ergebnis. Mit welcher W.K. ist ein positiv auf das Virus getesteter Mensch in Wirklichkeit gesund (also nicht infiziert). Wir können das mit dem Satz von der totalen W.K. machen: Jemand kann gesund oder krank (infiziert) sein (G, K) und jemand kann positiv (auf Virus) p oder negativ (gesund) n getestet werden. $P(\cdot)$ bezeichne die W.K. für diese Ereignisse. Dann gilt

$$P(G|p) = \frac{P(G) \cdot P(p|G)}{P(G) \cdot P(p|G) + P(K) \cdot P(p|K)} = \frac{0.999 \cdot 0.01}{0.999 \cdot 0.01 + 0.001 \cdot 0.99} \approx 0.9098\ldots$$

Also: Die meisten positiv getesteten Menschen sind gesund!

1.8 Stoppstrategie

Sie suchen den besten Partner/in fürs Leben! Oder aber Sie sind Personalchef und suchen unter einer Menge von n Bewerbern den Besten. Nun können Sie sich nicht alle Bewerber/innen anschauen, genauso wenig wie Sie alle möglichen Partner/innen ausprobieren können. Sie können aber alle Kandidaten, die sie sich genauer anschauen, miteinander vergleichen und sagen, wer besser ist. Es gibt aber a priori kein Maß, was eigentlich ein guter Bewerber ist (oder was ein guter Lebenspartner ist). Eine mögliche Strategie wäre: Sie schauen sich j Bewerber/innen an, dann haben Sie also Marktanalyse gemacht. Danach nehmen Sie den nächsten Bewerber, der besser ist als der Beste in der Referenzmenge der ersten j. Die Frage ist: Wie sollte man j wählen (sicherlich in Abhängigkeit von n), um die W.K. P_j, den besten Bewerber zu finden, zu maximieren?

Dazu betrachten wir folgende Ereignisse:

$$
\begin{aligned}
A_k \quad &: \quad \text{Bewerber } k \text{ wird angenommen.} \\
B_k \quad &: \quad \text{Bewerber } k \text{ ist der Beste aller Bewerber.}
\end{aligned}
$$

Wir erhalten folgende W.K.:

$$
\begin{aligned}
P(B_k) &= \frac{1}{n}, \\
P(A_k|B_k) &= 0 \text{ für } k \leq j, \\
P(A_k|B_k) &= \frac{j}{k-1} \text{ für } k > j.
\end{aligned}
$$

Kurz zur letzten Gleichung: Wenn der k-te Bewerber angenommen wird und dieser auch der Beste ist, muss der beste Bewerber unter den ersten $k-1$ unter den ersten j gewesen sein, denn sonst hätte man ja schon vor dem k-ten Bewerber gestoppt und wäre gar nicht in die Verlegenheit gekommen, sich den k-ten anzuschauen. Wir erhalten

$$
P_j = \sum_{k=1}^{n} P(A_k \cap B_k) = \sum_{k=1}^{n} P(B_k) P(A_k|B_k),
$$

also

$$
\begin{aligned}
P_j &= \frac{1}{n} \left(\frac{j}{j} + \frac{j}{j+1} + \ldots + \frac{j}{n-1} \right) \\
&= \frac{j}{n} \left(\frac{1}{j} + \ldots + \frac{1}{n-1} \right).
\end{aligned}
$$

Benutze nun

$$
\sum_{k=j}^{n-1} \frac{1}{k} \approx \int_j^n \frac{1}{x} \, \mathrm{d}x = \ln\left(\frac{n}{j}\right)
$$

um

$$
P_j \approx -x \ln(x)
$$

mit $x = j/n$ zu erhalten. Diese Funktion nimmt ihr Maximum für $x = 1/e$ an, und auch die W.K. P_j ist dann $1/e$.

1.9 Mehrstufige Experimente

Sehr häufig werden Zufallsexperimente als mehrstufige Experimente durchgeführt, sagen wir m Experimente. Die möglichen Ausgänge der i-ten Stufe seien in dem Ereignisraum Ω_i zusammengefasst. Ein Ereignis ist also ein m-Tupel

$$(\omega_1, \ldots, \omega_m) \in \Omega_1 \times \ldots \times \Omega_m.$$

Wir müssen die Wahrscheinlichkeiten der Elementarereignisse bestimmen. Das ist dann ganz einfach, wenn die Experimente, die zu den Ereignisräumen $\Omega_1, \ldots, \Omega_m$ gehören, unabhängig voneinander sind. In dem Fall gilt

$$P(\omega_1, \ldots, \omega_m) = \prod_{i=1}^{m} P_i(\omega_i),$$

wobei $P_i(\omega_i)$ die W.K. angibt, dass in der i-ten Stufe das Ereignis ω_i eintritt.

Sehr oft ist es aber so, dass die W.K. in der zweiten Stufe davon abhängt, was in der ersten Stufe passiert ist:

Beispiel 1.20. In einer Urne liegen eine rote und drei schwarze Kugeln. Nun wird in der ersten Stufe des Experimentes eine Kugel gezogen, die Farbe notiert und die Kugel sowie eine weitere Kugel derselben Farbe zurückgelegt. Wir haben $\Omega_1 = \{R, S\} = \Omega_2$. Gesucht ist $P(R, R), P(R, S), P(S, R), P(S, S)$. Es ist $P_1(R) = 1/4$ und $P_1(S) = \frac{3}{4}$. Man nennt das oft auch die Startverteilung des mehrstufigen Experimentes. Offenbar gilt

$$\begin{aligned}
P(R, R) &= \frac{1}{4} \cdot \frac{2}{5} = \frac{2}{20} \\
P(R, S) &= \frac{1}{4} \cdot \frac{3}{5} = \frac{3}{20} \\
P(S, R) &= \frac{3}{4} \cdot \frac{1}{5} = \frac{3}{20} \\
P(S, S) &= \frac{3}{4} \cdot \frac{4}{5} = \frac{12}{20}.
\end{aligned}$$

Die Wahrscheinlichkeiten beim Übergang von Stufe $j-1$ zu Stufe j hängen also davon ab, was “vorher” passiert ist. Wir nennen

$$P_j(\omega_j | \omega_1, \ldots, \omega_{j-1})$$

die **Übergangswahrscheinlichkeit**, das ist die W.K. für das Eintreten von ω_j unter der Voraussetzung, dass in den vorhergehenden Stufen des mehrstufigen Experimentes die Ereignisse $\omega_1, \ldots, \omega_{j-1}$ eingetreten sind. Offenbar gilt

$$P(\omega_1, \ldots, \omega_m) = P(\omega_1) \cdot P(\omega_2 | \omega_1) \cdot P(\omega_3 | \omega_1, \omega_2) \cdots P(\omega_m | \omega_1, \ldots, \omega_{m-1}).$$

Man macht sich so etwas oft an einem Baumdiagramm klar.

Beispiel 1.21. Wir würfeln mit einem Würfel. Danach nehmen wir eine Münze und werfen diese so oft, wie beim vorhergehenden Würfelwurf Augen angezeigt wurden. Angenommen, jemand teilt uns mit, dass bei all diesen Münzwürfen als Ergebnis Kopf herauskam. Mit welcher Wahrscheinlichkeit war dann das Ergebnis der Würfelwurfs eine $1, 2, 3, 4, 5$ oder 6. Es stellt sich heraus, dass diese W.K. 32/63, 16/63, 8/63, 4/63, 2/63 und 1/63 sind.

Beispiel 1.22 (*Das Ziegenproblem)*. In einer Fernsehshow wird hinter einer von drei verschlossenen Türen ein Fahrrad, hinter den anderen beiden eine Ziege versteckt. Dann darf die Kandidatin eine der Türen öffnen: Ist dahinter das Rad, darf sie es behalten, andernfalls darf sie nix behalten, nicht mal die Ziege! Nun deutet die Kandidatin erst einmal auf eine Tür. Der Moderator sagt daraufhin, hinter einer der anderen beiden Türen sei ja eh eine Ziege, und das weiß doch jeder, deshalb könne er doch auch gefahrlos eine Tür, hinter der eine Ziege steht, öffnen (wir gehen davon aus, dass der Moderator weiß, hinter welcher Tür das Fahrrad steht). Der Moderator gibt der Kandidatin nun die Möglichkeit, sich neu zu entscheiden, welche Tür sie öffnen will. Frage: Lohnt es sich für die Kandidatin, sich neu zu entscheiden? Die Antwort ist ja, wie wir in der Vorlesung zeigen. Die W.K., das Fahrrad zu gewinnen, steigt durch einen Wechsel der Entscheidung auf 2/3!