4 Diskrete Wahrscheinlichkeitsverteilungen

4.1 Wahrscheinlichkeitsräume, Ereignisse und Unabhängigkeit

Definition: Ein *diskreter Wahrscheinlichkeitsraum* ist ein Paar $(\Omega, \Pr)$, wobei Ω eine endliche oder abzählbar unendliche Menge von elementaren Ereignissen und $\Pr : \Omega \longrightarrow [0,1]$ eine *Wahrscheinlichkeitsverteilung* ist, die jedem $\omega \in \Omega$ seine *Wahrscheinlichkeit* $\Pr(\omega)$ zuordnet. Außerdem wird gefordert, daß $\sum_{\omega\in\Omega} \Pr(\omega) = 1$ ist. Wenn $\Omega = n$ endlich ist für alle $\omega \in \Omega$ auch $\Pr(\omega) = \frac{1}{n}$ gilt, dann wird Pr eine Gleichverteilung genannt.

Beispiele:

1) Für einen (fairen) Würfel ist $\Omega = \{1,2,3,4,5,6\}$ und $\Pr(1) = \Pr(2) = \ldots = \Pr(6) = \frac{1}{6}$ eine Gleichverteilung.

2) Der diskrete Wahrscheinlichkeitsraum für einen Münzwurf besteht aus $\Omega = \{0,1\}$ mit $\Pr(0) = \Pr(1) = 1/2$ (Gleichverteilung), wobei 0 den Kopf und 1 die Zahl der Münze bezeichnet.

3) Der diskrete Wahrscheinlichkeitsraum für zwei Münzwürfe besteht aus
$\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$ mit der Gleichverteilung $\Pr((i,j)) = 1/4$. Dabei geht man davon aus, daß die beiden Würfe voneinander unabhängig sind und entweder hintereinander oder gleichzeitig mit zwei unterscheidbaren Münzen erfolgen.

4) Beim gleichzeitigen Werfen von zwei ununterscheidbaren Münzen so ist $\Omega = \{\{0,0\},\{0,1\},\{1,1\}\}$. In diesem Fall ist Pr (faire Münzen vorausgesetzt) keine Gleichverteilung, denn $\Pr(\{0,0\}) = \Pr(\{1,1\}) = 1/4$ und $\Pr(\{0,1\}) = 1/2$. Dieses Modell ist eher als Gedankenspiel zu betrachten, denn in der Realität sind es zwei verschiedene Münzen, auch wenn sie äußerlich nicht unterscheidbar sind.

5) Definiert man eine Folge von Münzwürfen, bis das erste Mal Zahl (also 1) fällt, als elementares Ereignis, dann gibt es offensichtlich für jedes $n \in \mathbb{N}^+$ ein elementares Ereignis, nämlich die Folge $\omega_n = \underbrace{0,0,\ldots,0}_{(n-1)\text{ mal}},1$. Bei einer fairen Münze ist $\Pr(\omega_n) = \left(\frac{1}{2}\right)^n$ und damit ist auch für diesen unendlichen Raum die Summenbedingung erfüllt:

$$\sum_{n=1}^{\infty} \Pr(\omega_n) = \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n = 1$$

Definition: Eine beliebige Untermenge $A \subseteq \Omega$ wird *Ereignis* genannt, und man definiert $\Pr(A) = \sum_{\omega\in A} \Pr(\omega)$. Dadurch wird die Verteilungsfunktion zu einem Wahrscheinlichkeitsmaß über der Menge $\mathcal{P}(\Omega)$ aller Ereignisse erweitert. Die Verwendung der gleichen Bezeichnung Pr für die Verteilung und das Wahrscheinlichkeitsmaß ist eine kleine technische Unsauberkeit, die aber in der Literatur weit verbreitet ist.

Beispiele:

- Die Wahrscheinlichkeit, eine gerade Zahl zu würfeln, ist gleich der Summe der Wahrscheinlichkeiten, eine $2, 4$ oder 6 zu würfeln, also $1/2$.

- Das Ereignis A, bei drei Münzwürfen genau zweimal Zahl zu erhalten, setzt sich aus den drei elementaren Ereignissen $(0,1,1),(1,0,1)$ und $(1,1,0)$ zusammen und somit ist $\Pr(A) = 3 \cdot \frac{1}{8} = \frac{3}{8}$.

Aus der Definition und dem kombinatorischen Basiswissen folgt:

1. $\Pr(\Omega) = 1$ und $\Pr(\emptyset) = 0$
2. $A \subseteq B \Longrightarrow \Pr(A) \leq \Pr(B)$
3. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
4. $\Pr(A) = \sum_{i=1}^{k} \Pr(A_i)$ für jede Partition $A = \bigcup_{i=1}^{k} A_i$
5. $\Pr(\Omega \setminus A) = 1 - \Pr(A)$
6. $\Pr\left(\bigcup_{i=1}^{k} A_i\right) \leq \sum_{i=1}^{k} \Pr(A_i)$.

Oft taucht das Problem auf, die Wahrscheinlichkeit eines Ereignisses A unter der zusätzlichen Voraussetzung zu bestimmen, daß ein anderes Ereignis B eingetreten ist. So ist offensichlich unter der Vorausetzung, daß eine gerade Zahl gewürfelt wurde, die Wahrscheinlichkeit der 1 gleich Null und die Wahrscheinlichkeit der 2 gleich 1/3.

Definition: Seien A und B Ereignisse und $\Pr(B) > 0$. Dann ist die *bedingte Wahrscheinlichkeit von A unter B* definiert als

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

Beispiel: Würfelt man mit zwei unabhängigen und unterscheidbaren Würfeln, dann ist $\Omega = \{(i,j) \mid 1 \leq i,j \leq 6\}$ und $\Pr((i,j)) = 1/36$ für jedes Paar (i,j) aus Ω. Bezeichne $A_1 = \{(i,j) \in \Omega \mid i = 1\}$ das Ereignis, daß der erste Würfel eine 1 zeigt und der zweite beliebig ist. Analog kann man die Ereignisse $A_2, \ldots, A_6$ definieren. Ähnlich können durch Festlegung der Augenzahl des zweiten Würfels Ereignisse $B_1 = \{(i,j) \in \Omega \mid j = 1\}$, $B_2, \ldots, B_6$ definiert werden. Offensichlich ist die Wahrscheinlichkeit dieser Ereignisse jeweils 1/6 und Durchschnitte wie $A_1 \cap B_3 = \{(1,3)\}$ haben die Wahrscheinlichkeit 1/36. Folglich ist

$$\Pr(A_1|B_3) = \frac{\Pr(A_1 \cap B_3)}{\Pr(B_3)} = \frac{1/36}{1/6} = \frac{1}{6}$$

und analog ist die bedingte Wahrscheinlichkeit von jedem A_i unter einem beliebigen B_j gleich 1/6, also gleich der Wahrscheinlichkeit von A_i selbst. Diese Eigenschaft ist durch die Unabhängikeit der beiden Würfel begründet.

Definition: Zwei Ereignisse A und B werden *unabhängig* genannt, falls $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$.

Lemma: Sind A und B zwei unabhängige Ereignisse mit
$\Pr(A \cap B) > 0$, dann gilt $\Pr(A|B) = \Pr(A)$ und $\Pr(B|A) = \Pr(B)$.

Ereignisse mit positver Wahrscheinlichkeit, die einander ausschließen, wie zum Beispiel A_1 und A_2 aus dem letzten Beispiel sind voneinander nicht unabhängig, also abhängig. Ein Beispiel für abhängige Ereignisse beim Spiel mit einem Würfel ist der Wurf einer geraden Zahl und der Wurf einer 2 oder auch der Wurf einer geraden Zahl und der Wurf einer Zahl ≤ 3. Auch im folgenden Beispiel geht es um die Abhängigkeit von Ereignissen.

Beispiel: Ein Krebstest fällt mit 96% Wahrscheinlichkeit positiv aus, wenn der Patient Krebs hat und mit 94% Wahrscheinlichkeit negativ, wenn der Patient keinen Krebs hat. Bei einem Patienten in dessen Altersgruppe 0.5% aller Personen Krebs haben, fällt der Test positiv aus. Wie wahrscheinlich ist es, dass er tatsächlich Krebs hat?
Wir betrachten die folgenden Ereignisse

K	zufälliger Patient hat Krebs	$\Pr(K) = 0.005$
N	zufälliger Patient hat keinen Krebs	$\Pr(N) = 0.995$
T	Test positiv bei zufälligen Patienten	$\Pr(T) =??$

Gesucht ist die bedingte Wahrscheinlichkeit $\Pr(K|T) = \dfrac{\Pr(K \cap T)}{\Pr(T)}$.

Aus $\Pr(T|K) = \dfrac{\Pr(K \cap T)}{\Pr(K)} = 0.96$ kann man $\Pr(K \cap T) = 0.005 \cdot 0.96$ ausrechnen.
Dagegen ist $\Pr(T|N) = 1 - \Pr\big(\overline{T}|N\big) = 1 - 0.94 = 0.06$. Zur Bestimmung von $\Pr(T)$ überlegt man sich, dass das Ereignis T die disjunkte Vereinigung $(T \cap K) \cup (T \cap N)$ ist und sich die Wahrscheinlichkeiten der beiden Komponenten wieder durch bedingte Wahrscheinlichkeiten ausdrücken lassen:
$\Pr(T) = \Pr(T|K) \cdot \Pr(K) + \Pr(T|N) \cdot \Pr(N) = 0.96 \cdot 0.005 + 0.06 \cdot 0.995$. Letzlich erhalten wir

$$\Pr(K|T) = \frac{\Pr(K \cap T)}{\Pr(T)} = \frac{0.005 \cdot 0.96}{0.96 \cdot 0.005 + 0.06 \cdot 0.995} = 0.0744\ldots$$

Damit ist die Wahrscheinlichkeit, dass der Patient mit positiven Test wirklich Krebs hat, kleiner als 7.5%.

Definition: Sind $(\Omega_1, \Pr_1)$ und $(\Omega_2, \Pr_2)$ zwei Wahrscheinlichkeitsräume, dann ist der Produktraum $(\Omega_1 \times \Omega_2, p)$ definiert durch
$\Pr((\omega_1, \omega_2)) = \Pr_1(\omega_1) \cdot \Pr_2(\omega_2)$.

Diese Definition impliziert Unabhägigkeiten der folgenden Form: für alle $A \subseteq \Omega_1$ und $B \subseteq \Omega_2$ sind die Ereignisse $A \times \Omega_2$ und $\Omega_1 \times B$ voneinander unabhängig. Ein Beispiel dafür wurde bereits behandelt, denn der Wahrscheinlichkeitsraum für zwei Würfel ist das Produkt des Raums eines Würfel mit sich selbst. Im folgenden werden endliche Produkte des Münzwurfraums genauer analysiert:

- Sei eine faire Münze gegeben, die sechsmal geworfen wird. Der zugehörige Wahrscheinlichkeitsraum besteht aus allen 0/1–Folgen der Länge 6, wobei die

Wahrscheinlichkeit jeder Folge gleich $\frac{1}{2^6}$ ist. Wie groß ist die Wahrscheinlichkeit des Ereignisses, daß genau zwei Einsen in der Folge auftreten? Die Anzahl solcher Folgen stimmt offensichtlich mit der Anzahl der 2–Kombinationen einer 6–Menge überein. Folglich hat das Ereignis die Wahrscheinlichkeit $\binom{6}{2}/2^6$.

- Sei eine "unfaire" Münze gegeben, die mit Wahrscheinlichkeit p $(0 < p < 1)$ Zahl (d.h. 1) und mit Wahrscheinlichkeit $q = 1 - p$ Kopf (d.h. 0) liefert. Betrachtet man den Wahrscheinlichkeitsraum für n Münzwürfe, dann ist die Wahrscheinlichkeit einer festen Folge der Länge n mit genau k Einsen gleich $p^k(1-p)^{n-k}$. Die Wahrscheinlichkeit des Ereignisses, daß eine Folge genau k Einsen enthält ist $\binom{n}{k}p^k(1-p)^{n-k}$. Deshalb werden Verteilungen dieser Art auch *Binomialverteilungen* genannt.

4.2 Zufallsvariable und Erwartungswert

Definition: Sei (Ω, p) ein Wahrscheinlichkeitsraum. Eine *Zufallsvariable* oder *Zufallsgröße* ist eine Funktion $X : \Omega \longrightarrow \mathbb{R}$.

Im Fall eines Würfels oder eine zufällig gezogenen Zahl aus einer Urne fällt der Unterschied zwischen einem elementaren Ereignis und einer Zufallsvariable nicht auf, weil das Ereignis selbst schon eine Zahl ist, d.h. man kann die identische Abbildung als eine Zufallsvariable ansehen. Die folgenden Beispiele zeigen, dass man viele Zufallsexperimente auf sehr natürliche Weise mit Zufallsvariablen beschreiben kann.

Beispiele:

1) Der diskrete Wahrscheinlichkeitsraum $\Omega = \{1, 2, \ldots, 6\} \times \{1, 2, \ldots, 6\}$ für zwei Münzwürfe sind Zufallsvariablen für die Summe und die Differenz der beiden Werte interessant: $X(a, b) = a + b \qquad Y(a, b) = |a - b|$.

2) Auf dem Wahrscheinlichkeitsraum für n Münzwürfe $\Omega = \{0, 1\}^n$ interessiert man sich vor allen für die Zufallsvariablen, welche die Anhzahl der Zahl- bzw. Kopfergebnisse in einer solchen Serie Zählen, d.h.

$$\begin{aligned} X(a_1, a_2, \ldots, a_n) &= |\{i \mid 1 \leq i \leq n \text{ und } a_i = 1\}| \\ Y(a_1, a_2, \ldots, a_n) &= |\{i \mid 1 \leq i \leq n \text{ und } a_i = 0\}| = n - X(a_1, a_2, \ldots, a_n) \end{aligned}$$

3) Bei der Wiederholung von Münzwürfen bis zur ersten Zahl (oder bis zum ersten Kopf) ist offensichtlich die Länge des Experiments eine interessante zufälige Größe:

$$X(\underbrace{0, 0, \ldots, 0}_{(n-1) \text{ mal}}, 1) = n$$

Definition: Der *Erwartungswert* der Zufallsvariablen X ist definiert als $E(X) = \sum_{\omega \in \Omega} X(\omega) \Pr(\omega)$.

Im ersten und dritten Beispiel kann man Erwartungswerte relativ leicht ausrechen:

1) $E(X) = 7$ und $E(Y) = \frac{70}{36}$;

3) $E(X) = \sum_{n=1}^{\infty} n \cdot \frac{1}{2^n} = 2$ (ist etwas schwerer);
Im zweiten Beispiel ist $E(X) = \frac{n}{2}$ die richtige Antwort. Intuitiv ist das klar, aber eine direkte Berechnung durch konkrete Auswertung des Defintion ist relativ aufwändig. Der folgende Fakt macht die Bestimmung des Erwartungswerts sehr einfach.

Lemma: Der Erwartungswert ist additiv, d.h. für beliebige Zufallsvariable $X, Y : \Omega \longrightarrow \mathbb{R}$ gilt $E(X+Y) = E(X) + E(Y)$ wobei die neue Variable $X+Y$ in naheliegender Weise durch $(X+Y)(w) = X(w) + Y(w)$ definiert ist.

Anwendungen:

- Wir betrachten die binomialverteilte Zufallsvariable X aus Beispiel 2, die jedem 0/1–Tupel der Länge n die Anzahl seiner Einsen zuordnet. Um das Beispiel noch etwas allgemeiner zu machen, betrachten wir eine unfaire Münze, die mit Wahrscheinlichkeit p eine 1 (Zahl) und mit Wahrscheinlichkeit $1-q$ eine 0 (Kopf) liefert. Gruppiert man die Folgen derart, daß alle Folgen mit genau k Einsen einer Gruppe angehören, so erhält man $E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$ und es ist relativ aufwändig daraus $E(X) = np$ abzuleiten. Durch Nutzung der Additivität der Erwartungswerte kann man diese Formel einfacher erhalten: Offensichtlich ist $X = X_1 + X_2 + \ldots + X_n$ wobei die Variable X_i angewendet auf eine Folge von Münzwurfergebnissen $(a_1, \ldots, a_n)$ das Ergebnis des i-ten Wurfs auswertet, also $X_i((a_1, \ldots, a_n)) = a_i$. Da die Erwartungswerte dieser Summanden jeweils q sind folgt $E(X) = E(X_1) + \ldots + E(X_n) = nq$.

- An einem Tisch sitzen 12 Personen. Von jeder Person wurde ein Foto gemacht. Diese 12 Bilder liegen verdeckt auf dem Tisch. Jede Person zieht zufällig ein Foto (zufällig bedeutet hier, dass alle möglichen Zuordnungen der Bilder gleich wahrscheinlich sein sollen). Wie hoch ist die erwartete Anzahl der Personen, die Ihr eigenes Foto gezogen haben?

  Wieder kann man die Zufallsvariable X, welche die Anzahl dieser Personen zählt, als Summe von zwölf Variablen $X_1, \ldots, X_{12}$ darstellen, wobei X_i den Wert 1 hat, wenn Person i ihr eigenes Foto gezogen hat, und 0 sonst. Offensichtlich ist $E(X_i) = \frac{1}{12}$ für alle i und folglich ist $E(X) = 12 \cdot \frac{1}{12} = 1$.
  Aus dieser Betrachtung ergibt sich außerdem die (vielleicht etwas überraschende) Tatsache, das der Erwatungswert nicht von der Anzahl der Personen abhängt - auch bei 100 Personen zieht im Erwartungswert eine Person ihr eigenes Foto.

- Eine Versicherung macht folgende Kalkulation: Jeder Versicherte zahlt 100 Euro ein. Mit Wahrscheinlichkeit von 1/20 (also 5 %) muß eine Leistung von 1000 Euro und mit Wahrscheinlichkeit 1/10 eine muß eine Leistung von 400 Euro erbracht werden. Wie hoch ist der Erwartungswert für den Gewinn (pro Versicherungsnehmer)?

  Der zu betrachtende Wahrscheinlichkeitsraum hat drei Elementarereignisse ω_1, ω_2 und ω_3 mit $p(\omega_1) = 1/20$, $p(\omega_2) = 1/10$ und $p(\omega_3) = 1 - 1/20 - 1/10 = 17/20$. Die Zufallsvariable X, die den Gewinn der Versicherung beschreibt, ist die Differenz aus einer Zufallsvariablen Y für die Einnahmen und einer Zufallsvariablen

Z für die Ausgaben in den einzelnen Ereignissen: Dabei hat Y in jedem Fall den Wert 100 und $Z(\omega_1) = 1000$, $Z(\omega_2) = 400$ und $Z(\omega_3) = 0$. Nach Definition erhalten wir für den Erwartungswert

$$E(X) = E(Y) - E(Z) = 100 - (1000 \cdot \frac{1}{20} + 400 \cdot \frac{1}{10} + 0) = 100 - 90 = 10$$

In vielen praktischen Anwendungen entstehen Zufallsvariable durch Messung von zufälligen Größen, die auf sehr schwer zu analysierende Wahrscheinlichkeitsräume zurückgehen. Beispiele dafür sind die Zeit, die man für seinen täglichen Arbeitsweg braucht oder die Körpergröße einer Person. Deshalb spricht man häufig auch von der Verteilung einer Variablen, d.h. man interessiert sich für die Wahrscheinlichkeit, dass die Variable einen bestimmten Wert (oder einen Wert aus einem bestimmeten Intervall) annimmt und weniger für die Wahrscheinlichkeiten der einzelnen elementaren Ereignisse. Die Wahrscheinlichkeit, dass eine Zufallsvariable X den Wert x_0 bzw. einen Wert kleiner oder gleich x_0 annimmt wird mit $\Pr(X = x_0)$ bzw. $\Pr(X \leq x_0)$ notiert. In den gerade genannten Beispielen sind das sogenannte Normalverteilungen, die man nicht mit einem diskreten Modell beschreiben kann. Es gibt aber auch eine Reihe von diskreten Zufallsvariablen, die häufig auftreten und deshalb in bestimmte Typen eingeteilt werden.

- Eine *Gleichverteilung* liegt vor, wenn eine Zufallsvariable k verschiedene Werte annehmen kann und für jeden Wert die Wahrscheinlichkeit $1/k$ hat.

- Eine *Bernoulli–Verteilung* mit dem Parameter p liegt vor, wenn eine Zufallsvariable X nur die Werte 0 und 1 annimmt und $\Pr(X = 1) = p$, $\Pr(X = 0) = 1 - p$ gilt.

- Eine Zufallsvariable X ist *binomialverteilt* mit den Parametern n und p, wenn sie die Werte $0, 1, 2, \ldots, n$ annimmt und $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ für alle k aus dem Wertebereich gilt. Mit anderen Worten entsteht eine Binomialverteilung immer dann, wenn man ein Bernoulli–Experiment n–mal wiederholt und die Zufallsvariable die Anzahl der 1–Ergebnisse zählt.

- Eine Zufallsvariable X ist *geometrisch verteilt* mit den Parameter p, wenn sie alle Werte aus $\mathbb{N}+$ annehmen kann und $\Pr(X = k) = (1-p)^{k-1} p$ für alle $k \in \mathbb{N}^+$ gilt. Mit anderen Worten entsteht eine geometrische Verteilung immer dann, wenn man ein Bernoulli–Experiment so lange wiederholt, bis zum ersten Mal eine 1 fällt und die Zufallsvariable die Anzahl der Versuche zählt.

Wichtige Anwendungen von Erwartungswerten in der Informatik sind die *Analyse der durchschnittlichen Laufzeit* von Algorithmen (Ω besteht aus den Eingaben und $X(\omega)$ ist die Laufzeit für die Eingabe ω) und die Analyse *randomisierter Algorithmen* (d.h. die Arbeit des Algorithmus wird durch eingebaute Zufälle gesteuert).