# Überblick zu Kapitel 1 des Buches „Stochastische Signaltheorie”

Dieses erste Kapitel bringt eine kurze Zusammenfassung der **Wahrscheinlichkeitsrechnung**, die sicher viele von Ihnen bereits aus der Schulzeit kennen und die eine wichtige Voraussetzung für das Verständnis der nachfolgenden Kapitel darstellt.

Dieses Kapitel beinhaltet

- einige *Definitionen* wie Zufallsexperiment, Ergebnis, Ereignis und Wahrscheinlichkeit,
- die für die Wahrscheinlichkeitsrechnung relevanten *mengentheoretischen Grundlagen*,
- die Verdeutlichung von *Statistischer Abhängigkeit* bzw. *Statistischer Unabhängigkeit*,
- die mathematische Behandlung von statistischen Abhängigkeiten durch *Markovketten*.

Die theoretischen Grundlagen werden auf 25 Bildschirmseiten verdeutlicht. Außerdem beinhaltet dieses Kapitel noch 36 Grafiken, sieben Aufgaben und sieben Zusatzaufgaben mit insgesamt 67 Teilaufgaben, vier Lernvideos sowie ein interaktives Berechnungsmodul.

**Zusammenstellung der Lernvideos (LV) und Interaktionsmodule (IM) zu Kapitel 1:**

- **Klassische Definition der Wahrscheinlichkeit** (LV zu Kap. 1.1 – Dauer 5:19)
- **Würfelspiel „Mäxchen”** (Musterlösung zu A1.1(f) – Dauer 1:40)
- **Mengentheoretische Begriffe und Gesetze** (LV zu Kap. 1.2 – 2-teilig: Dauer je 6:10)
- **Statistische (Un–)Abhängigkeit** (LV zu Kap. 1.3 – 3–teilig: Dauer 4:17 – 3:48 – 3:48)
- **Ereigniswahrscheinlichkeiten einer Markovkette** (IM zu Kap. 1.4)

**Literaturhinweise**: **[Chu67] – [Cra74] – [Fis78] – [GT96] – [Hau03] – [HG73] – [Kol33] – [Kre85] – [Mül75] – [PP02] – [Söd88] – [Söd00] – [Spi76] – [Tho86]**

# Experiment – Ergebnis – Wahrscheinlichkeit

Der Ausgangspunkt einer jeden statistischen Untersuchung ist ein **Zufallsexperiment**. Darunter versteht man einen unter stets gleichen Bedingungen beliebig oft wiederholbaren Versuch mit ungewissem **Ergebnis** $E$, bei dem jedoch die Menge $\{E_\mu\}$ der möglichen Ergebnisse angebbar ist.

Die Anzahl der möglichen Ergebnisse bezeichnet man als den **Ergebnisumfang** $M$. Dann gilt:

$$E_\mu \in G = \{E_\mu\} = \{E_1, \ \dots \ , E_M\}.$$

Hierbei kann die Laufvariable $\mu$ alle ganzzahligen Werte zwischen 1 und $M$ annehmen. $G$ nennt man auch den Ereignisraum oder die **Grundmenge**.

**Beispiel:** Beim Experiment *Münzwurf* gibt es nur zwei mögliche Ergebnisse, nämlich *Zahl* und *Bild* ⇒ $M$ = 2. Dagegen sind beim Zufallsexperiment *Werfen einer Roulettekugel* insgesamt $M$ = 37 verschiedene Ergebnisse möglich, und es gilt hier für für die Grundmenge:

$$G = \{E_\mu\} = \{0, 1, 2, \dots, 36\}.$$

Wir setzen zunächst voraus, dass jeder Versuch genau ein einziges Ergebnis aus $G$ zur Folge hat und dass jedes dieser $M$ Ergebnisse in gleicher Weise (ohne Bevorzugung oder Benachteiligung) möglich ist.

**Definition:** Mit dieser Annahme gilt für die **Wahrscheinlichkeit** (englisch: *Probability*) eines jeden Ergebnisses $E_\mu$ gleichermaßen:

$$\Pr(E_\mu) = 1/M.$$

Dies ist die **klassische Definition** der Wahrscheinlichkeit. Pr( ... ) steht dabei für *Probability* und ist als eine mathematische Funktion zu verstehen.

**Beispiel**: Beim Zufallsexperiment *Münzwurf* gilt für die Wahrscheinlichkeiten der beiden möglichen Ergebnisse: Pr(*Zahl*) = Pr(*Bild*) = 1/2. Dies setzt voraus, dass jeder Versuch entweder mit *Zahl* oder mit *Bild* ausgeht und dass nicht bei einem Versuch die Münze auf ihrem Rand zu stehen kommen kann.

Auch beim Versuch *Werfen einer Roulettekugel* sind die Wahrscheinlichkeiten $\Pr(E_\mu)$ = 1/37 nur dann für alle Zahlen von 0 bis 36 gleich, wenn der Roulettetisch nicht manipuliert wurde.

**Anmerkung**: Die Wahrscheinlichkeitsrechnung – und die darauf aufbauende Statistik – kann nur dann fundierte Aussagen liefern, wenn alle implizit vereinbarten Voraussetzungen tatsächlich erfüllt sind. Diese Bedingungen zu überprüfen ist nicht Aufgabe der Statistik, sondern von denjenigen, die diese nutzen. Da gegen diese Grundregel oft verstoßen wird, hat die Statistik in der Gesellschaft einen viel schlechteren Ruf, als es ihr eigentlich zustehen würde.

# Ereignis und Ereignismenge

Unter einem **Ereignis** $A_i$ verstehen wir eine Menge bzw. die Zusammenfassung von Ergebnissen. Die Menge aller Ereignisse bezeichnen wir als die **Ereignismenge** $\{A_i\}$. Da die Anzahl $I$ der möglichen Ereignisse $\{A_i\}$ im Allgemeinen nicht mit der Anzahl $M$ der möglichen Ergebnisse – also der Elemente von $G = \{E_\mu\}$ – übereinstimmt, werden hier unterschiedliche Indizes gewählt.

**Klassische Definition der Wahrscheinlichkeit**: Setzt sich ein Ereignis $A_i$ aus $K$ (elementaren) Ergebnissen zusammen, so wird dessen Wahrscheinlichkeit wie folgt definiert:

$$\Pr(A_i) = \frac{K}{M} = \frac{\text{Anzahl der günstigen Ergebnisse}}{\text{Anzahl der möglichen Ergebnisse}}.$$

Diese Gleichung nennt man die Wahrscheinlichkeitsdefinition nach **Laplace.** Günstige Ergebnisse sind dabei solche, die zum zusammengesetzten Ereignis $A_i$ gehören. Aus dieser Gleichung geht bereits hervor, dass eine Wahrscheinlichkeit stets zwischen 0 und 1 liegen muss (einschließlich dieser beiden Grenzen).

Die Thematik von Kapitel 1.1 wird in einem Lernvideo anhand einfacher Beispiele behandelt:

**Klassische Definition der Wahrscheinlichkeit** (Dauer 5:19)

**Beispiel**: Wir betrachten wieder das Experiment *Werfen eines Würfels*. Die möglichen Ergebnisse sind $E_\mu \in G = \{1, 2, 3, 4, 5, 6\}$. Definieren wir nun zwei Ereignisse ($I = 2$), nämlich

- $A_1$ = [die Augenzahl ist geradzahlig] = $\{2, 4, 6\}$ und
- $A_2$ = [die Augenzahl ist ungeradzahlig] = $\{1, 3, 5\}$,

so ist die Ereignismenge $\{A_1, A_2\}$ gleich der Grundmenge $G$. Die Ereignisse $A_1$ und $A_2$ stellen für dieses Beispiel ein so genanntes **vollständiges System** dar.

Dagegen ist die weitere Ereignismenge $\{A_3, A_4\}$ ungleich der Grundmenge $G$, wenn man die beiden Einzelereignisse wie folgt definiert:

- $A_3$ = [die Augenzahl ist kleiner als 3] = $\{1, 2\}$,
- $A_4$ = [die Augenzahl ist größer als 3] = $\{4, 5, 6\}$.

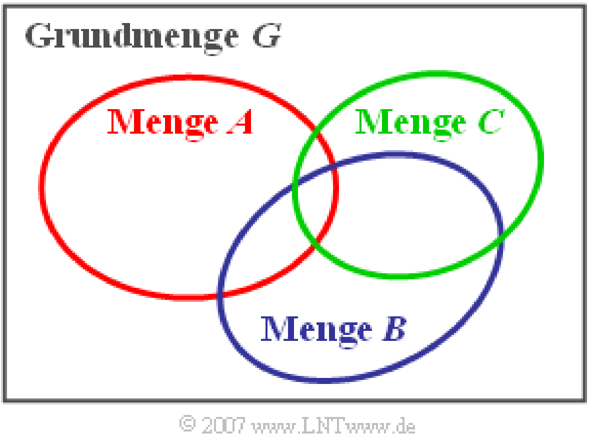
Hier beinhaltet die Ereignismenge $\{A_3, A_4\}$ nicht das Element „3”. Die Wahrscheinlichkeiten der obigen Ereignisse sind $\Pr(A_1) = \Pr(A_2) = \Pr(A_4) = 1/2$ und $\Pr(A_3) = 1/3$.

# Venndiagramm, Grundmenge und leere Menge

In späteren Kapitel wird manchmal auf die *Mengenlehre* Bezug genommen. Deshalb sollen hier die wichtigsten Grundlagen und Definitionen dieser Disziplin kurz zusammengefasst werden. Die Thematik wird auch im folgenden Lernvideo am Beispiel europäischer Staaten behandelt:

**Mengentheoretische Begriffe und Gesetzmäßigkeiten** (2-teilig: Dauer je 6:10).

Ein wichtiges Hilfsmittel der Mengenlehre ist das **Venndiagramm** gemäß dem folgenden Bild.




Angewandt auf die Wahrscheinlichkeitsrechnung sind hier die Ereignisse $A_i$ als Flächenbereiche dargestellt. Zur einfacheren Beschreibung bezeichnen wir hier die Ereignisse im Gegensatz zu Kapitel 1.1 nicht mit $A_1$, $A_2$, $A_3$ usw., sondern mit $A$, $B$ und $C$. Die Gesamtfläche entspricht der Grundmenge $G$.

Die Grundmenge $G$ beinhaltet alle möglichen Ergebnisse und steht für das **sichere Ereignis**, das definitionsgemäß mit der Wahrscheinlichkeit „Eins" eintritt: $\Pr(G) = 1$. Zum Beispiel ist beim Zufallsexperiment *Werfen eines Würfels* die Wahrscheinlichkeit für das Ereignis „die Augenzahl ist kleiner oder gleich 6" identisch 1.

Dagegen beinhaltet die **leere Menge** $\phi$ kein einziges Element. Bezogen auf Ereignisse gibt die leere Menge das **unmögliche Ereignis** mit der Wahrscheinlichkeit $\Pr(\phi) = 0$ an. Beispielsweise ist beim Experiment *Werfen eines Würfels* die Wahrscheinlichkeit für das Ereignis „die Augenzahl ist größer als 6" identisch 0.

Weiter ist anzumerken, dass nicht jedes Ereignis $A$ mit $\Pr(A) = 0$ wirklich nie eintreten kann. So ist die Wahrscheinlichkeit des Ereignisses „der Rauschwert $n$ ist identisch 0" zwar verschwindend klein und es gilt $\Pr(n = 0) = 0$, wenn $n$ durch eine kontinuierliche Zufallsgröße beschrieben wird. Trotzdem ist es natürlich möglich, dass irgendwann auch der exakte Rauschwert $n = 0$ auftritt.

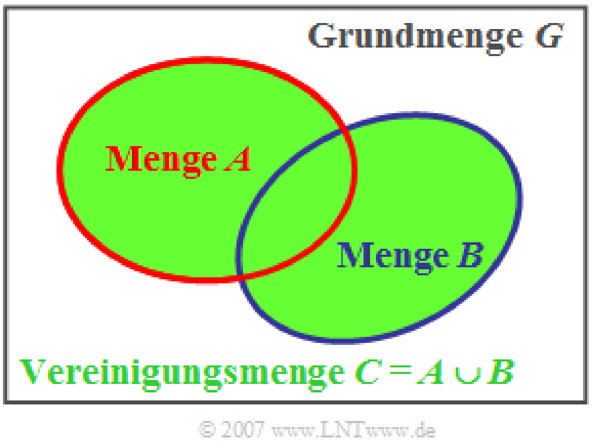
# Vereinigungsmenge

Anhand des Venndiagramms werden nun einige mengentheoretische Verknüpfungen erläutert.

**Definition:** Die **Vereinigungsmenge** $C$ zweier Mengen $A$ und $B$ beinhaltet alle die Elemente, die entweder in der Menge $A$ oder der Menge $B$ oder in beiden enthalten sind (englisch: *Set Union*). Formelmäßig wird dieser Zusammenhang wie folgt ausgedrückt:

$$C = A \cup B \; (= A + B).$$

In der Literatur ist auch die Bezeichnung *Summenmenge* gebräuchlich und es wird manchmal das Pluszeichen benutzt. In unserem Tutorial verwenden wir jedoch ausschließlich das ∪-Zeichen.




Anhand des Bildes sind die folgenden Gesetzmäßigkeiten der Mengenlehre leicht einzusehen:

$$A \cup \phi = A \qquad \text{(Vereinigung mit der leeren Menge)},$$

$$A \cup G = G \qquad \text{(Vereinigung mit der Grundmenge)},$$

$$A \cup A = A \qquad \text{(Tautologiegesetz)},$$

$$A \cup B = B \cup A \qquad \text{(Kommutativgesetz)},$$

$$(A \cup B) \cup C = A \cup (B \cup C) \qquad \text{(Assoziativgesetz)}.$$

Ist über die Ereignismengen $A$ und $B$ nichts weiter bekannt, so können für die Wahrscheinlichkeit der Vereinigungsmenge nur eine untere und eine obere Schranke angegeben werden:

$$\text{Max}(\Pr(A), \Pr(B)) \le \Pr(A \cup B) \le \Pr(A) + \Pr(B).$$

Die Wahrscheinlichkeit der Vereinigungsmenge ist gleich der unteren Schranke, wenn $A$ eine **Teilmenge** von $B$ ist oder umgekehrt. Die obere Schranke gilt für **disjunkte Mengen.**

**Beispiel**: Betrachtet man die beiden Ereignisse

- $A$ = „die Augenzahl ist größer oder gleich 5" = {5, 6},
- $B$ = „die Augenzahl ist geradzahlig" = {2, 4, 6},

so beinhaltet die Vereinigungsmenge die vier Elemente {2, 4, 5, 6}. Die Wahrscheinlichkeiten sind Pr($A$) = 2/6, Pr($B$) = 3/6 und Pr($A$ ∪ $B$) = 4/6. Die untere und die obere Schranke gemäß den hier angegebenen Ungleichungen ergeben sich zu 3/6 und 5/6.

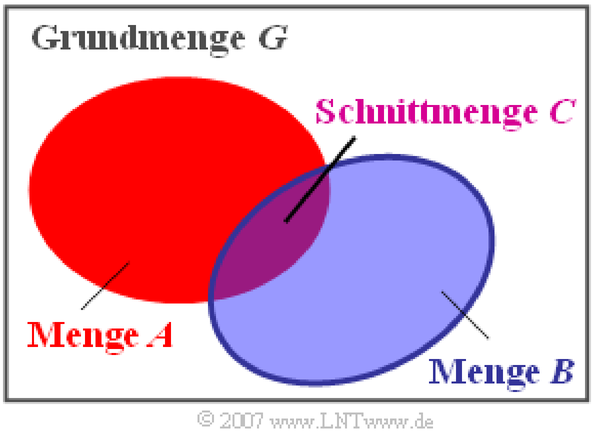
## Schnittmenge

Eine weitere wichtige mengentheoretische Verknüpfung stellt die Schnittmenge dar.

**Definition:** Die **Schnittmenge** $C$ zweier Mengen $A$ und $B$ beinhaltet alle diejenigen Elemente, die sowohl in der Menge $A$ als auch in der Menge $B$ enthalten sind (englisch: *Intersecting Set*). Formelmäßig wird dieser Zusammenhang wie folgt ausgedrückt:

$$C = A \cap B \ \ (= A \cdot B).$$

In der Literatur ist hierfür auch die Bezeichnung *Produktmenge* gebräuchlich und man verwendet das Multiplikationssymbol. Im nachfolgenden Bild ist die Schnittmenge violett dargestellt.




Analog zur Vereinigungsmenge sind hier folgende Gesetzmäßigkeiten zu nennen:

$$A \cap \phi = \phi \qquad \text{(Schnitt mit der leeren Menge)},$$

$$A \cap G = A \qquad \text{(Schnitt mit der Grundmenge)},$$

$$A \cap A = A \qquad \text{(Tautologiegesetz)},$$

$$A \cap B = B \cap A \qquad \text{(Kommutativgesetz)},$$

$$(A \cap B) \cap C = A \cap (B \cap C) \qquad \text{(Assoziativgesetz)}.$$

Ist über die Mengen $A$ und $B$ nichts weiter bekannt, so kann auch für die Wahrscheinlichkeit der Schnittmenge nur eine untere und eine obere Schranke angegeben werden:

$$0 \le \Pr(A \cap B) \le \text{Min}(\Pr(A), \Pr(B)).$$

$\Pr(A \cap B)$ wird auch **Verbundwahrscheinlichkeit** genannt und manchmal mit $\Pr(A, B)$ bezeichnet. Sie ist gleich der oberen Schranke, wenn $A$ eine **Teilmenge** von $B$ ist oder umgekehrt. Die untere Schranke ergibt sich für die Verbundwahrscheinlichkeit von **disjunkten Mengen.**

**Beispiel**: Die Ereignissen seien wieder $A$ = „die Augenzahl ist größer oder gleich 5" = {5, 6} sowie $B$ = „die Augenzahl ist geradzahlig" = {2, 4, 6}. Die Schnittmenge beinhaltet nur ein einziges Element: $A \cap B$ = {6}. Mit $\Pr(A) = 2/6$ und $\Pr(B) = 3/6$ ergibt sich $\Pr(A \cap B) = 1/6$.

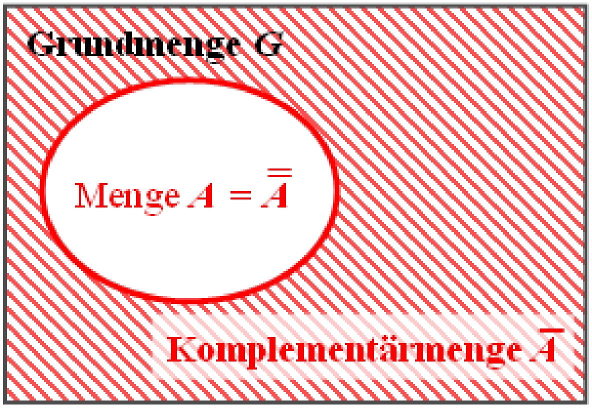
Die untere und obere Schranke entsprechend den angegebenen Ungleichungen sind 0 und 2/6.

# Komplementärmenge

**Definition:** Die **Komplementärmenge** (englisch: *Complementary Set*) von $A$ wird oft durch eine überstreichende Linie gekennzeichnet. Sie beinhaltet alle die Elemente, die in der Menge $A$ nicht enthalten sind, und es gilt für deren Wahrscheinlichkeit:

$$\Pr(\overline{A}) = 1 - \Pr(A).$$

Im nachfolgenden Venndiagramm ist die zu $A$ komplementäre Menge schraffiert dargestellt.




© 2008 www.LNTwww.de

Aus diesem Schaubild sind einige mengentheoretische Beziehungen zu erkennen:

- Die Komplementärmenge der komplementären Menge von $A$ ist die Menge $A$ selbst:

$$\overline{\overline{A}} = A.$$

- Die Vereinigungsmenge einer Menge $A$ mit der Komplentärmenge ergibt die Grundmenge:

$$\Pr(A \cup \overline{A}) = \Pr(G) = 1.$$

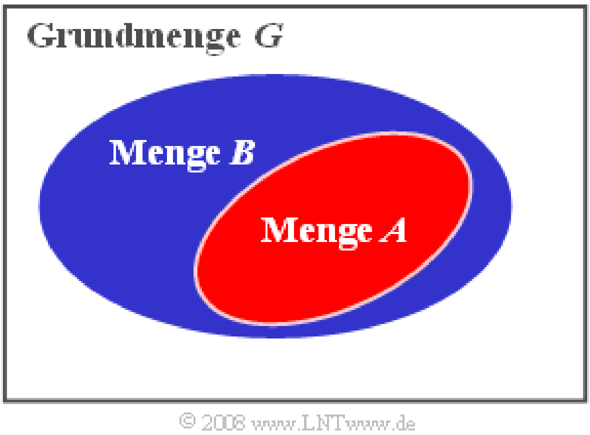
- Die Schnittmenge von $A$ mit der zugehörigen Komplementärmenge ergibt die leere Menge:

$$\Pr(A \cap \overline{A}) = \Pr(\phi) = 0.$$

**Beispiel**: Ausgehend von der Menge $A$ = „die Augenzahl ist kleiner als 5” lautet die zugehörige Komplentärmenge in Worten: „die Augenzahl ist größer oder gleich 5”. Die Wahrscheinlichkeit dieser Komplentärmenge berechnet sich zu 1 − Pr($A$) = 1 − 2/3 = 1/3.

# Echte Teilmenge – unechte Teilmenge

Diese mengentheoretische Relation wird durch das folgende Venndiagramm veranschaulicht.




**Definition:** Man bezeichnet $A$ als eine **echte Teilmenge** von $B$ und schreibt hierfür $A \subset B$, wenn alle Elemente von $A$ auch in $B$ enthalten sind, aber nicht gleichzeitig alle Elemente von $B$ auch in $A$ (englisch: *Strict Subset*). In diesem Fall gilt für die Wahrscheinlichkeiten:

$$\Pr(A) < \Pr(B).$$

Dagegen bezeichnet man $A$ als eine **unechte Teilmenge** von $B$ und verwendet die folgende Notation, wenn $A$ entweder eine echte Teilmenge von $B$ ist oder $A$ und $B$ gleiche Mengen sind:

$$A \subseteq B = (A \subset B) \cup (A = B).$$

Für die Wahrscheinlichkeiten gilt dann die Größenrelation $\Pr(A) \le \Pr(B)$. Das Gleichheitszeichen gilt nur für den Sonderfall identischer Mengen. Da

- die Schnittmenge $A \cap B$ stets eine Teilmenge von $A$ ist,
- $A$ aber auch gleichzeitig eine Teilmenge der Vereinigungsmenge $A \cup B$ ist,

gelten auch die beiden als **Absorptionsgesetze** bekannten Gleichungen:

$$(A \cap B) \cup A = A,$$

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

**Beispiel**: Die Menge $A$ = „die Augenzahl ist ungerade” = $\{1, 3, 5\}$ ist eine (echte) Teilmenge der Menge $B$ = „die Augenzahl ist eine Primzahl” = $\{1, 2, 3, 5\}$, wenn $G$ die Zahlen 1 bis 6 enthält. Die Wahrscheinlichkeit $\Pr(A) = 3/6$ ist deshalb kleiner als $\Pr(B) = 4/6$.

# Theoreme von de Morgan

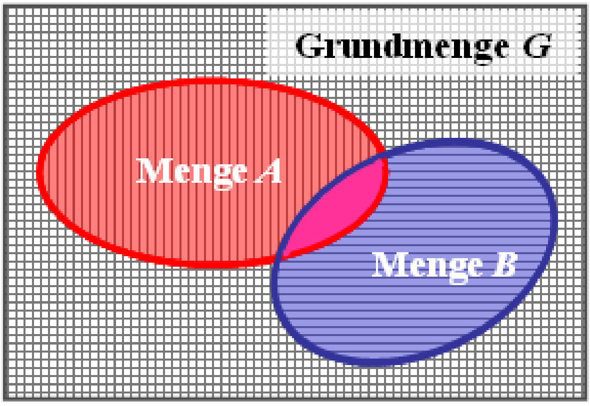
Bei vielen Aufgaben aus der Mengenlehre sind die beiden Theoreme von **de Morgan** äußerst nützlich. Diese lauten:

$$\overline{A \cup B} = \overline{A} \cap \overline{B},$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}.$$

Diese Gesetzmäßigkeiten sind im folgenden Schaubild veranschaulicht:

- Die Menge $A$ ist rot dargestellt und die Menge $B$ blau.
- Die Komplentärmenge von $A$ ist in horizontaler Richtung schraffiert.
- Die Komplentärmenge von $B$ ist in vertikaler Richtung schraffiert.
- Das Komplement der Vereinigungsmenge ist sowohl horizontal als auch vertikal schraffiert.
- Es ist damit gleich der Schnittmenge der beiden Komplentärmengen von $A$ und $B$.






Die Schnittmenge $A \cap B$ (im Bild violett dargestellt) ist weder horizontal noch vertikal schraffiert.

- Das Komplement der Schnittmenge ist dementsprechend entweder horizontal, vertikal oder in beiden Richtungen schraffiert.
- Nach dem zweiten Theorem von de Morgan ist das Komplement der Schnittmenge gleich der Vereinigungsmenge der beiden Komplentärmengen von $A$ und $B$.

**Beispiel**: Wir betrachten nun die beiden Mengen

- $A$ = „die Augenzahl ist ungeradzahlig" = $\{1, 3, 5\}$,
- $B$ = „die Augenzahl ist größer als 2" = $\{3, 4, 5, 6\}$.

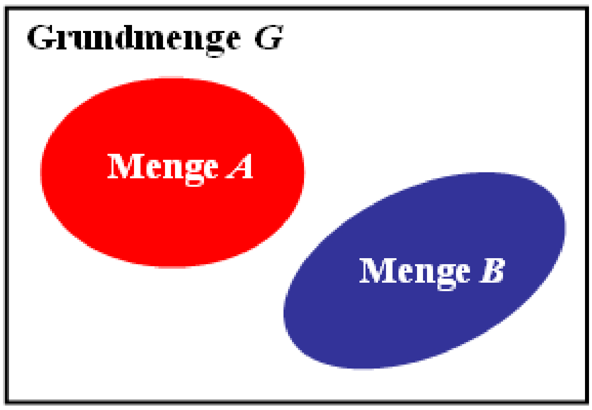
Daraus folgen die beiden komplementären Mengen „die Augenzahl ist geradzahlig" = $\{2, 4, 6\}$ bzw. „die Augenzahl ist kleiner als 3" = $\{1, 2\}$. Weiter erhält man mit den obigen Theoremen folgende Mengen:

$$\overline{A \cup B} = \overline{A} \cap \overline{B} = \{2\} \quad \text{und} \quad \overline{A \cap B} = \overline{A} \cup \overline{B} = \{1, 2, 4, 6\}.$$

# Disjunkte Mengen

**Definition:** Zwei Mengen $A$ und $B$ nennt man **disjunkt** (englisch: *disjoint*) oder miteinander **unvereinbar**, wenn es kein einziges Element gibt, das sowohl in $A$ als auch in $B$ enthalten ist.

Das Schaubild zeigt zwei disjunkte Mengen $A$ und $B$ im Venndiagramm.






In diesem Sonderfall gelten die folgenden Aussagen:

- Die Schnittmenge zweier disjunkter Mengen $A$ und $B$ ergibt stets die leere Menge:

$$\Pr(A \cap B) = \Pr(\phi) = 0.$$

- Die Wahrscheinlichkeit der Vereinigungsmenge zweier disjunkter Mengen $A$ und $B$ ist immer gleich der Summe der beiden Einzelwahrscheinlichkeiten:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B).$$

**Beispiel**: Die Mengen

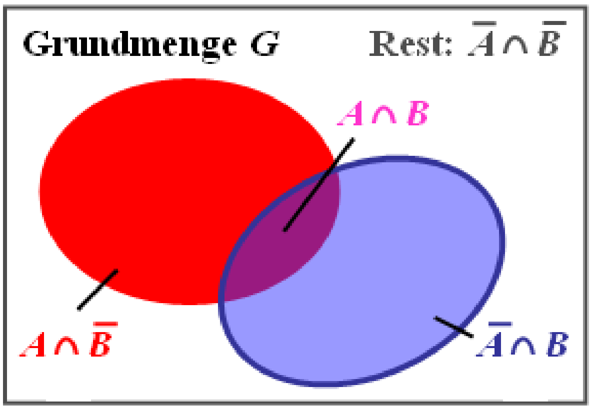
- $A$ = „die Augenzahl ist kleiner als 3” = {1, 2} ⇒ Pr($A$) = 2/6, und
- $B$ = „die Augenzahl ist größer als 3” = {4, 5, 6} ⇒ Pr($B$) = 3/6

sind zueinander disjunkt, da $A$ und $B$ kein einziges gemeinsames Element beinhalten. Deshalb ist die Wahrscheinlichkeit der Vereinigungsmenge {1, 2, 4, 5, 6} gleich Pr($A$) + Pr($B$) = 5/6.

# Additionstheorem

Nur im Sonderfall disjunkter Mengen $A$ und $B$ gilt der Zusammenhang $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ für die Wahrscheinlichkeit der Vereinigungsmenge. Wie errechnet sich diese Wahrscheinlichkeit aber bei allgemeinen, nicht notwendigerweise disjunkten Ereignissen?

Betrachten Sie hierzu das folgende Venndiagramm mit der violett dargestellten Schnittmenge $A \cap B$. Die rote Menge beinhaltet alle Elemente, die zu $A$ gehören, aber nicht zu $B$. Die Elemente von $B$, die nicht gleichzeitig in $A$ enthalten sind, erkennt man an der blauen Farbe. Alle roten, blauen und violetten Flächen zusammen ergeben die Vereinigungsmenge $A \cup B$.






Aus dieser mengentheoretischen Darstellung erkennt man folgende Zusammenhänge:

$$\Pr(A) = \Pr(A \cap B) + \Pr(A \cap \overline{B}),$$

$$\Pr(B) = \Pr(A \cap B) + \Pr(\overline{A} \cap B),$$

$$\Pr(A \cup B) = \Pr(A \cap B) + \Pr(A \cap \overline{B}) + \Pr(\overline{A} \cap B).$$

Addiert man die ersten beiden Gleichungen und subtrahiert davon die dritte, so erhält man:

$$\Pr(A) + \Pr(B) - \Pr(A \cup B) = \Pr(A \cap B).$$

Durch Umstellen dieser Gleichung kommt man zum sogenannten **Additionstheorem** (englisch: *Addition Rule*) für zwei beliebige, nicht notwendigerweise disjunkte Ereignisse:

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B).$$

**Beispiel**: Wir betrachten die beiden Mengen

- $A$ = „die Augenzahl ist ungeradzahlig" = {1, 3, 5} $\Rightarrow$ $\Pr(A) = 3/6$, und
- $B$ = „die Augenzahl ist größer 2" = {3, 4, 5, 6} $\Rightarrow$ $\Pr(B) = 4/6$.

Damit sind die Wahrscheinlichkeiten

- der Vereinigungsmenge $\Rightarrow$ $\Pr(A \cup B) = 5/6$, und
- der Schnittmenge $\Rightarrow$ $\Pr(A \cap B) = 2/6$.

Die Zahlenwerte zeigen die Gültigkeit des Additionstheorems: $5/6 = 3/6 + 4/6 - 2/6$.

# Vollständiges System

Nun betrachten wir wieder mehr als zwei mögliche Ereignisse, nämlich allgemein $I$. Diese Ereignisse werden im Folgenden mit $A_i$ bezeichnet, und es gilt für den Laufindex: $1 \le i \le I$ .

**Definition:** Eine Konstellation mit den Ereignissen $A_1, \ldots , A_i, \ldots , A_I$ bezeichnet man dann und nur dann als ein **vollständiges System**, wenn die beiden folgenden Bedingungen erfüllt sind:

- Alle Ereignisse sind paarweise disjunkt:

$$A_i \cap A_j = \phi \text{ für alle } i \neq j.$$

- Die Vereinigung aller Ereignismengen ergibt die Grundmenge:

$$\bigcup_{i=1}^{I} A_i = G.$$

Aufgrund dieser beiden Voraussetzungen gilt dann für die Summe aller Wahrscheinlichkeiten:

$$\sum_{i=1}^{I} \Pr(A_i) = 1.$$

**Beispiel**: Die Ereignismengen $A_1 = \{1,\ 5\}$ und $A_2 = \{2,\ 3\}$ ergeben beim Zufallsexperiment *Werfen eines Würfels* zusammen mit der Menge $A_3 = \{4,\ 6\}$ ein vollständiges System, nicht jedoch beim Experiment *Werfen einer Roulettekugel*.

Ein Beispiel für ein vollständiges System ist auch eine diskrete Zufallsgröße $X = \{x_1, x_2, \ldots , x_I\}$ mit den Auftrittswahrscheinlichkeiten entsprechend der **Wahrscheinlichkeitsfunktion:**

$$P_X(X) = [\, P_X(x_1), P_X(x_2), \ldots , P_X(x_I)\,] = [\, p_1, p_2, \ldots , p_I\,]$$
$$\Rightarrow \quad p_1 = P_X(x_1) = \Pr(X = x_1)\,, \;\; p_2 = \Pr(X = x_2)\,, \ldots\,, \;\; p_I = \Pr(X = x_I)\,.$$

Die möglichen Ergebnisse $x_i$ der Zufallsgröße $X$ sind paarweise zueinander disjunkt und die Summe aller Auftrittswahrscheinlichkeiten $p_1 + p_2 + \ldots + p_I$ liefert grundsätzlich das Ergebnis 1.

**Beispiel**: Es gelte $X = \{0, 1, 2\}$ und $P_X(X) = [0.2, 0.5, 0.3]$. Dann gilt:

$$\Pr(X = 0) = 0.2\,, \;\; \Pr(X = 1) = 0.5\,, \;\; \Pr(X = 2) = 0.3\,.$$

Bei der Zufallsgröße $X = \{1, \pi, \mathrm{e}\}$ und gleichem $P_X(X) = [0.2, 0.5, 0.3]$ lauten die Zuordnungen:

$$\Pr(X = 1) = 0.2\,, \;\; \Pr(X = \pi) = 0.5\,, \;\; \Pr(X = \mathrm{e}) = 0.3\,.$$

*Hinweis*: Die Wahrscheinlichkeitsfunktion $P_X(X)$ macht lediglich Aussagen über die Wahrscheinlichkeiten, nicht jedoch über den Wertevorrat $\{x_1, x_2, \ldots , x_I\}$ der Zufallsgröße $X$. Diese zusätzliche Information liefert die **Wahrscheinlichkeitsdichtefunktion** (WDF).

# Allgemeine Definition von statistischer Abhängigkeit (1)

Bisher haben wir die **statistische Abhängigkeit** zwischen Ereignissen nicht besonders beachtet, auch wenn wir sie wie im Fall zweier disjunkter Mengen bereits verwendet haben: Gehört ein Element zu $A$, so kann es mit Sicherheit nicht auch in der disjunkten Menge $B$ enthalten sein.

Die stärkste Form von Abhängigkeit überhaupt ist eine **deterministische Abhängigkeit** zwischen zwei Mengen bzw. zwei Ereignissen. Weniger ausgeprägt ist die statistische Abhängigkeit. Beginnen wir mit deren Komplement:

**Definition:** Zwei Ereignisse $A$ und $B$ bezeichnet man dann als **statistisch unabhängig** (englisch: *statistical independent*) , wenn die Wahrscheinlichkeit der Schnittmenge $A \cap B$ gleich dem Produkt der Einzelwahrscheinlichkeiten ist:

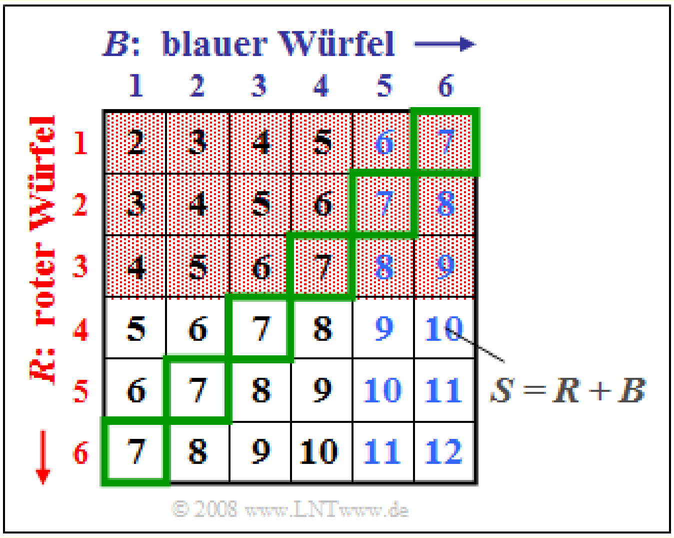
$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B).$$

In manchen Anwendungsfällen ist die statistische Unabhängigkeit offensichtlich, zum Beispiel beim Experiment *Münzwurf*. Die Wahrscheinlichkeit für *Zahl* oder *Bild* ist unabhängig davon, ob beim letzten Wurf *Zahl* oder *Bild* aufgetreten ist. Und auch die einzelnen Ergebnisse beim Zufallsexperiment *Werfen einer Roulettekugel* sind bei fairen Bedingungen stets statistisch unabhängig voneinander, auch wenn einzelne Systemspieler dies nicht wahrhaben wollen.

Bei anderen Anwendungen ist dagegen die Frage, ob zwei Ereignisse statistisch unabhängig sind oder nicht, gefühlsmäßig nicht oder nur sehr schwer zu beantworten. Hier kann man nur durch Überprüfung des oben angegebenen formalen Unabhängigkeitskriteriums zur richtigen Antwort gelangen, wie das Beispiel auf der nächsten Seite zeigen soll.

# Allgemeine Definition von statistischer Abhängigkeit (2)

**Beispiel**: Wir betrachten wieder das Zufallsexperiment *Werfen mit zwei Würfeln*, wobei die beiden Würfel an ihren Farben Rot ($R$) und Blau ($B$) unterschieden werden können. Die Grafik soll diesen Sachverhalt verdeutlichen, wobei in dem zweidimensionalen Feld ($R$, $B$) die Summe $S = R + B$ eingetragen ist.




Die Grafik kann wie folgt interpretiert werden:

- Die beiden Ereignisse $A_1$ = „$R < 4$” und $A_2$ = „$B > 4$” sind im Bild durch rötlichen Hintergrund bzw. blaue Zahlen hervorgehoben. $A_1$ und $A_2$ sind statistisch unabhängig, da die Wahrscheinlichkeit der Schnittmenge – also $\Pr(A_1 \cap A_2) = 1/6$ – gleich dem Produkt der beiden Einzelwahrscheinlichkeiten $\Pr(A_1) = 1/2$ und $\Pr(A_2) = 1/3$ ist. Aufgrund der Aufgabenstellung hätte auch jedes andere Ergebnis sehr überrascht.
- Aber auch die beiden Ereignisse $A_1$ = „$R < 4$” und $A_3$ = „$S = 7$” sind wegen $\Pr(A_1) = 1/2$, $\Pr(A_3) = 1/6$ und $\Pr(A_1 \cap A_3) = 1/12$ statistisch voneinander unabhängig. Das Ereignis „$S = 7$” ist in der Grafik durch grüne Umrahmungen gekennzeichnet.
- Dagegen bestehen wegen $\Pr(A_1) = 1/2$, $\Pr(A_4) = 5/36$ und $\Pr(A_1 \cap A_4) = 1/18$ statistische Bindungen zwischen $A_1$ = „$R < 4$” und dem im **Bild auf der nächsten Seite** markierten Ereignis $A_4$ = „$S = 8$”. Die beiden Ereignisse $A_1$ = „$R < 4$” und $A_5$ = „$S \geq 10$” sind sogar disjunkt. Dies zeigt, dass Disjunktivität eine besonders ausgeprägte Form von statistischer Abhängigkeit ist.

# Bedingte Wahrscheinlichkeit (1)

Bestehen zwischen den beiden Ereignissen $A$ und $B$ statistische Bindungen, so ist durch die (unbedingten) Wahrscheinlichkeiten Pr($A$) und Pr($B$) der Sachverhalt im statistischen Sinne nicht eindeutig beschrieben. Man benötigt dann noch so genannte bedingte Wahrscheinlichkeiten.

**Definition:** Die **bedingte Wahrscheinlichkeit** (englisch: *Conditional Probability*) von $A$ unter der Bedingung $B$ ist wie folgt berechenbar:

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

In gleicher Weise gilt für die bedingte Wahrscheinlichkeit von $B$ unter der Bedingung $A$:

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

Verknüpft man diese beiden Gleichungen, so ergibt sich der Satz von **Bayes:**

$$\Pr(B \mid A) = \frac{\Pr(A \mid B) \cdot \Pr(B)}{\Pr(A)}.$$

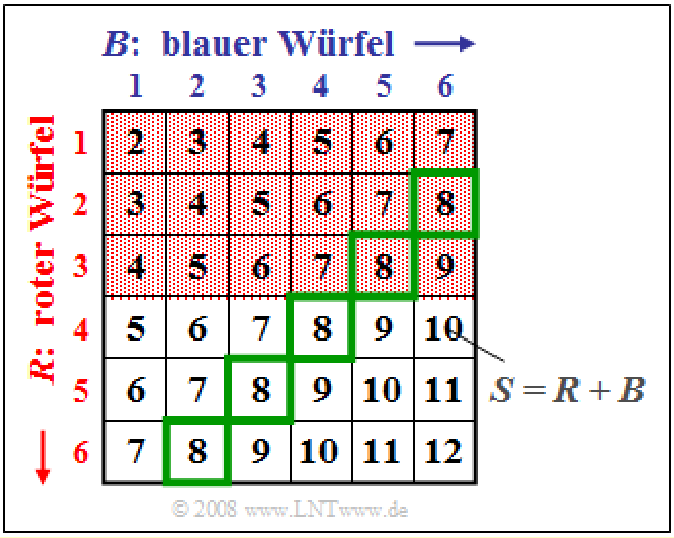
Nachfolgend sind einige Eigenschaften von bedingten Wahrscheinlichkeiten zusammengestellt:

- Auch eine bedingte Wahrscheinlichkeit liegt stets zwischen 0 und 1 einschließlich dieser beiden Grenzen: $0 \le \Pr(A \mid B) \le 1$.
- Kann die Bedingung $B$ als konstant angesehen werden, so gelten alle im Kapitel 1.2 für die unbedingten Wahrscheinlichkeiten Pr($A$) und Pr($B$) angegebenen Rechenregeln weiterhin.
- Sind die existierenden Ereignisse $A$ und $B$ disjunkt, so ist $\Pr(A \mid B) = \Pr(B \mid A) = 0$.
- Ist $B$ eine echte oder unechte Teilmenge von $A$, so ist $\Pr(A \mid B) = 1$.
- Sind zwei Ereignisse $A$ und $B$ statistisch voneinander unabhängig, so sind deren bedingte Wahrscheinlichkeiten gleich den unbedingten, wie die folgende Rechnung zeigt:

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A) \cdot \Pr(B)}{\Pr(B)} = \Pr(A).$$

# Bedingte Wahrscheinlichkeit (2)

**Beispiel**: Wir betrachten wieder das Zufallsexperiment *Werfen mit zwei Würfeln*, wobei wie beim letzten Beispiel $S = R + B$ die Summe des roten und des blauen Würfels bezeichnet. Im nachfolgenden Schema ist das Ereignis $A_1$ = „$R < 4$" wieder rötlich hinterlegt und das Ereignis $A_4$ = „$S = 8$" durch grüne Umrahmungen markiert.




Zu dieser Grafik ist anzumerken:

- Die bedingte Wahrscheinlichkeit $\Pr(A_1 \mid A_4) = 2/5$ (zwei der fünf grün umrandeten Felder sind rot hinterlegt) berechnet sich aus dem Quotienten der Verbundwahrscheinlichkeit $\Pr(A_1 \cap A_4) = 2/36$ und der Wahrscheinlichkeit $\Pr(A_4) = 5/36$.
- Da $A_1$ und $A_4$ statistisch abhängig sind, ist $\Pr(A_1 \mid A_4) = 2/5$ ungleich $\Pr(A_1) = 1/2$.
- Dieses letzte Ergebnis lässt sich zum Beispiel auch über den Satz von Bayes ableiten:

$$\Pr(A_4 \mid A_1) = \frac{\Pr(A_1 \mid A_4) \cdot \Pr(A_4)}{\Pr(A_1)} \quad \Rightarrow \quad 1/9 = \frac{2/5 \cdot 5/36}{1/2}.$$

- Dagegen gelten für $A_1$ und das hierzu statistisch unabhängige Ereignis $A_3$ = „$S = 7$" die folgenden bedingten Wahrscheinlichkeiten, siehe **Grafik zum letzten Beispiel:**

$$\Pr(A_1 \mid A_3) = \Pr(A_1) = 1/2 \quad \text{bzw.} \quad \Pr(A_3 \mid A_1) = \Pr(A_3) = 1/6.$$

# Allgemeines Multiplikationstheorem

Wir betrachten weiterhin mehrere Ereignisse $A_i$ mit $1 \le i \le I$. Diese Ereignisse $A_i$ stellen nun aber kein **vollständiges System** mehr dar, das heißt, sie sind nicht paarweise zueinander disjunkt, und es können zwischen den einzelnen Ereignissen auch statistische Bindungen bestehen.

Für die so genannte **Verbundwahrscheinlichkeit**, also für die Wahrscheinlichkeit der Schnittmenge aller $I$ Ereignisse $A_i$, gilt in diesem Fall:

$$\begin{aligned}\Pr(A_1 \cap \dots \cap A_I) = \quad & \dots \\ \dots \quad = \Pr(A_I) \cdot \Pr(A_{I-1} \,|\, A_I) \cdot \Pr(A_{I-2} \,|\, A_{I-1} \cap A_I) \cdot \ldots \cdot \Pr(A_1 \,|\, A_2 \cap \dots \cap A_I).\end{aligned}$$

In gleicher Weise gilt natürlich auch:

$$\begin{aligned}\Pr(A_1 \cap \dots \cap A_I) = \quad & \dots \\ \dots \quad = \Pr(A_1) \cdot \Pr(A_2 \,|\, A_1) \cdot \Pr(A_3 \,|\, A_1 \cap A_2) \cdot \ldots \cdot \Pr(A_I \,|\, A_1 \cap \dots \cap A_{I-1}).\end{aligned}$$

**Beispiel**: Eine Lostrommel enthält zehn Lose, darunter drei Treffer (Ereignis $T$). Dann ergibt sich für die Wahrscheinlichkeit, dass man mit zwei Losen zwei Treffer zieht:

$$\Pr(T_1 \cap T_2) = \Pr(T_1) \cdot \Pr(T_2 \,|\, T_1) = 3/10 \cdot 2/9 = 1/15 \approx 6.7\%.$$

Hierbei ist berücksichtigt, dass sich bei der zweiten Ziehung nur mehr neun Lose und zwei Treffer in der Urne befinden.

Würde man jedoch die Lose nach der Ziehung wieder in die Trommel zurücklegen, so wären die Ereignisse $T_1$ und $T_2$ statistisch unabhängig $\Rightarrow$ $\Pr(T_1 \cap T_2) = (3/10)^2 = 9\%$.

# Rückschlusswahrscheinlichkeit

Gegeben seien Ereignisse $A_i$ mit $1 \le i \le I$, die ein vollständiges System bilden. Das heißt: Alle Ereignisse sind paarweise disjunkt ($A_i \cap A_j = \phi$ für alle $i \ne j$) und die Vereinigungsmenge ergibt die Grundmenge:

$$\bigcup_{i=1}^{I} A_i = G.$$

Daneben betrachten wir noch das Ereignis $B$, von dem alle bedingten Wahrscheinlichkeiten $\Pr(B \mid A_i)$ mit den Indizes $1 \le i \le I$ bekannt sind.

**Satz von der totalen Wahrscheinlichkeit**: Unter den oben genannten Voraussetzungen gilt für die (unbedingte) Wahrscheinlichkeit des Ereignisses $B$:

$$\Pr(B) = \sum_{i=1}^{I} \Pr(B \cap A_i) = \sum_{i=1}^{I} \Pr(B \mid A_i) \cdot \Pr(A_i).$$

Aus dieser Gleichung folgt mit dem Satz von **Bayes** für die **Rückschlusswahrscheinlichkeit**:

$$\Pr(A_i \mid B) = \frac{\Pr(B \mid A_i) \cdot \Pr(A_i)}{\Pr(B)} = \frac{\Pr(B \mid A_i) \cdot \Pr(A_i)}{\sum_{k=1}^{I} \Pr(B \mid A_k) \cdot \Pr(A_k)}.$$

**Beispiel**: In Münchner Studentenheimen wohnen Studierende der LMU (Ereignis $L$; 70%) und der TUM (Ereignis $T$; 30%). Es ist weiterhin bekannt, dass an der LMU 60% aller Studierenden weiblich sind, an der TUM nur 10%. Der Anteil aller Studentinnen (Ereignis $W$) kann dann mit dem Satz von der totalen Wahrscheinlichkeit ermittelt werden:

$$\Pr(W) = \Pr(W \mid L) \cdot \Pr(L) + \Pr(W \mid T) \cdot \Pr(T) = 0.6 \cdot 0.7 + 0.1 \cdot 0.3 = 45\%.$$

Trifft man eine Studentin, so kann man mit der Rückschlusswahrscheinlichkeit

$$\Pr(L \mid W) = \frac{\Pr(W \mid L) \cdot \Pr(L)}{\Pr(W \mid L) \cdot \Pr(L) + \Pr(W \mid T) \cdot \Pr(T)} = \frac{0.6 \cdot 0.7}{0.6 \cdot 0.7 + 0.1 \cdot 0.3} = \frac{14}{15}$$

vorhersagen, dass sie an der LMU studieren wird. Ein durchaus realistisches Ergebnis.

Die Aussagen dieses Abschnitts sind im nachfolgenden Lernvideo zusammengefasst:

**Statistische (Un-)Abhängigkeit** (3-teilig: Dauer 4:20 – 3:40 – 3:40)

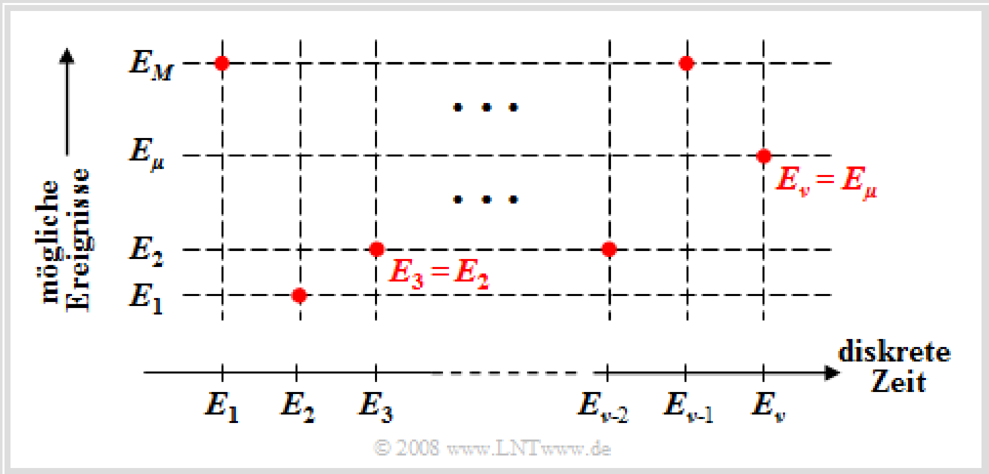
# Betrachtetes Szenario

Wir betrachten nun abschließend den Fall, dass man ein Experiment fortlaufend durchführt und zu jedem diskreten Zeitpunkt $\nu = 1, 2, 3, \ldots$.. ein Ereignis $E_\nu$ eintritt. Hierbei soll gelten:

$$E_\nu \in G = \{E_1, E_2, \ldots, E_\mu, \ldots, E_M\}.$$

Diese mathematisch nicht ganz saubere Nomenklatur bedeutet (siehe nachfolgende Grafik):

- Die $M$ möglichen Ereignisse werden mit dem Laufindex $\mu$ durchnummeriert.
- Der Index $\nu$ benennt die diskreten Zeitpunkte, zu denen Entscheidungen gefällt werden.




Zur einfacheren Darstellung beschränken wir uns im Folgenden auf den Fall $M = 2$ mit der Grundmenge $G = \{A, B\}$. Wir berücksichtigen, dass die Wahrscheinlichkeit des Ereignisses $E_\nu$ durchaus von allen vorherigen Ereignissen – also von $E_{\nu-1}$, $E_{\nu-2}$, $E_{\nu-3}$, . . . – abhängen kann. Das bedeutet auch, dass wir eine Ereignisfolge mit inneren statistischen Bindungen betrachten.

Dieses Szenario ist ein Sonderfall eines zeit- und wertdiskreten Zufallsprozesses. Solche Prozesse werden in **Kapitel 4.4** noch ausführlich behandelt.

**Beispiel:** Aus einem Kartenstapel mit 32 Karten (darunter 4 Asse) werden nacheinander Karten gezogen. Mit den Ereignissen $A$ = „die gezogene Karte ist ein Ass" und $B$ = „die gezogene Karte ist kein Ass" lauten die Wahrscheinlichkeiten zum Zeitpunkt $\nu = 1$:

$$\Pr(A_1) = \frac{4}{32} = \frac{1}{8}; \quad \Pr(B_1) = \frac{28}{32} = \frac{7}{8}.$$

Die Wahrscheinlichkeit $\Pr(A_2)$, dass zur Zeit $\nu = 2$ ein Ass gezogen wird, hängt nun davon ab,

- ob zum Zeitpunkt $\nu = 1$ ein Ass gezogen wurde $\Rightarrow$ $\Pr(A_2) = 3/31 < 1/8$, oder
- ob zum Zeitpunkt $\nu = 1$ kein Ass gezogen wurde $\Rightarrow$ $\Pr(A_2) = 4/31 > 1/8$.

Auch die Wahrscheinlichkeiten $\Pr(A_\nu)$ zu späteren Zeitpunkten $\nu$ hängen stets vom Eintreffen bzw. Nichteintreffen aller vorherigen Ereignisse $E_1$ ... $E_{\nu-1}$ ab.

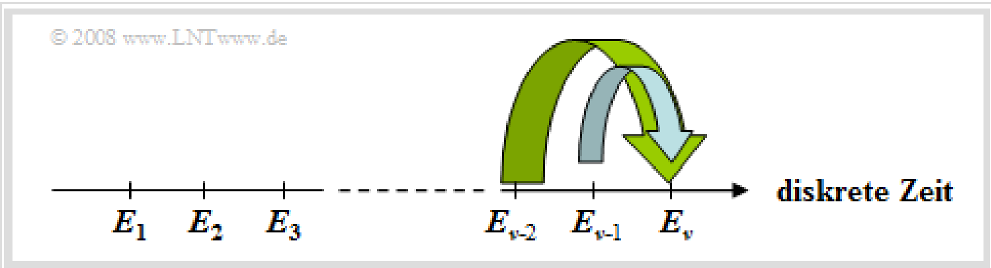
# Allgemeine Definition einer Markovkette

In Sonderfällen, die allerdings sehr häufig vorkommen, kann das oben beschriebene Szenario durch eine Markovkette beschrieben werden.

**Definition:** Eine **Markovkette $k$-ter Ordnung** (englisch: *Markov Chain*) dient als Modell für zeit- und wertdiskrete Vorgänge, bei denen die Ereigniswahrscheinlichkeiten zur Zeit $\nu$ von den vorherigen Ereignissen $E_{\nu-1}, \ \ldots \ , E_{\nu-k}$ abhängen und durch $M^{k+1}$ bedingte Wahrscheinlichkeiten ausgedrückt werden können. Für $M = 2$ gibt es deshalb $2^{k+1}$ solcher Wahrscheinlichkeiten:

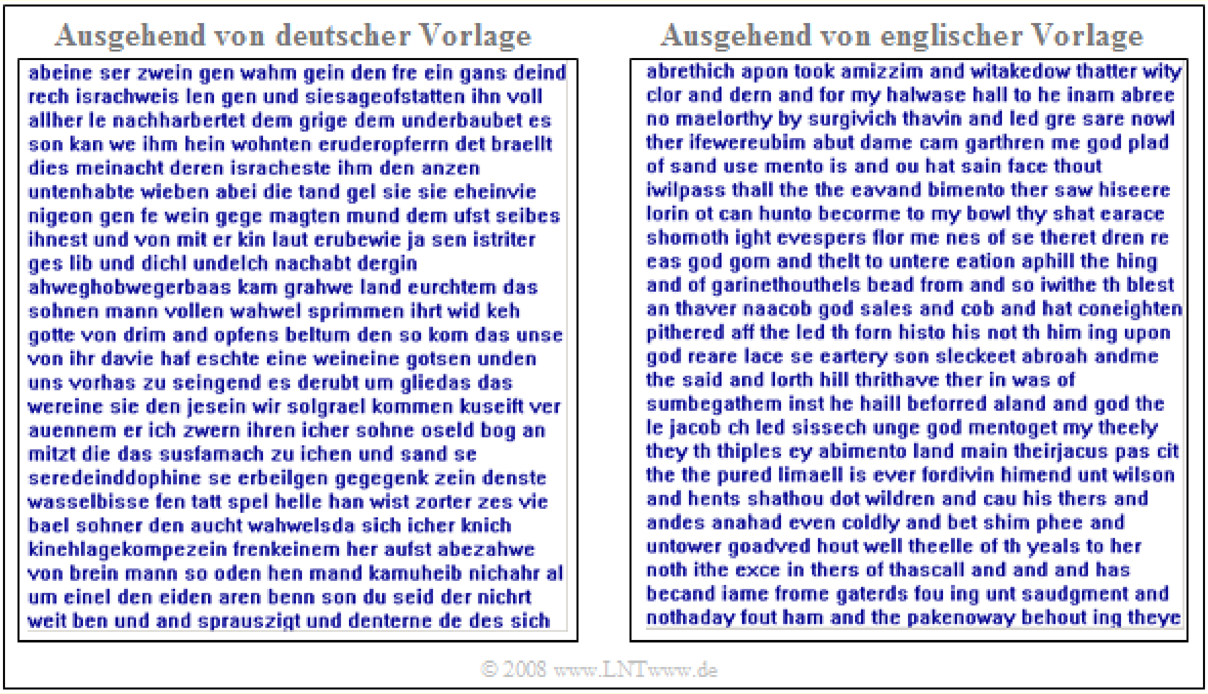
$$\mathrm{Pr}(E_\nu \,|\, E_{\nu-1}, \ \ldots \ , E_{\nu-k}) \quad \mathrm{mit} \quad E_\nu \in \{A, B\}, \ \ldots \ , E_{\nu-k} \in \{A, B\}.$$

Das nachfolgende Bild verdeutlicht diesen Sachverhalt am Beispiel $k = 2$.




**Beispiel:** Natürliche Sprachen sind oft durch Markovketten beschreibbar, wobei allerdings die Ordnung $k$ gegen Unendlich strebt. Nähert man einen Text durch eine Markovkette 2. Ordnung an, so ergibt sich zwar kein sinnvoller Inhalt, aber die Struktur der Sprache ist erkennbar.

Das untere Bild zeigt links einen Text, der ausgehend von einer deutschen Buchvorlage mit Bindungen bis zu zweiter Ordnung synthetisch erzeugt wurde. Beim rechten Text wurde eine englische Vorlage verwendet. Man erkennt trotz der Beschränkung $k = 2$ viele (kurze) deutsche bzw. englische Wörter und auch, dass deutsche Wörter im Mittel länger sind als englische.



# Markovkette erster Ordnung

Im Folgenden beschränken wir uns stets auf den Sonderfall $k = 1$.

**Definition:** Bei einer **Markovkette erster Ordnung** (englisch: *First Order Markov Chain*) werden lediglich die statistische Bindung zum letzten Ereignis berücksichtigt, die in der Praxis meist auch am stärksten ist.

Eine binäre Markovkette ⇒ Grundmenge $G = \{A, B\}$ weist folgende Wahrscheinlichkeiten auf:

$$\Pr(A_\nu) = \Pr(A_\nu\,|\,A_{\nu-1}) \cdot \Pr(A_{\nu-1}) \; + \; \Pr(A_\nu\,|\,B_{\nu-1}) \cdot \Pr(B_{\nu-1}),$$

$$\Pr(B_\nu) = \Pr(B_\nu\,|\,A_{\nu-1}) \cdot \Pr(A_{\nu-1}) \; + \; \Pr(B_\nu\,|\,B_{\nu-1}) \cdot \Pr(B_{\nu-1}).$$

Hierzu ist anzumerken:

- $\Pr(A_\nu)$ steht als Abkürzung für die Wahrscheinlichkeit, dass $E_\nu = A$ ist.
- Zu jedem beliebigen Zeitpunkt $\nu$ gilt: $\Pr(B_\nu) = 1 - \Pr(A_\nu)$.
- Es gibt zu jedem Zeitpunkt vier Übergangswahrscheinlichkeiten $\Pr(E_\nu \,|\, E_{\nu-1})$, von denen jedoch nur zwei unabhängig sind, denn es gilt:

  $$\Pr(B_\nu\,|\,A_{\nu-1}) = 1 - \Pr(A_\nu\,|\,A_{\nu-1}), \qquad \Pr(A_\nu\,|\,B_{\nu-1}) = 1 - \Pr(B_\nu\,|\,B_{\nu-1}).$$

- Durch Verallgemeinerung dieser letzten Aussage gelangt man zu dem Ergebnis, dass es bei einer Markovkette mit $M$ Ereignissen zu jedem Zeitpunkt $\nu$ genau $M \cdot (M-1)$ voneinander unabhängige Übergangswahrscheinlichkeiten gibt.

**Beispiel:** Mit den vorgegebenen Übergangswahrscheinlichkeiten

$$\Pr(A_\nu \,|\, A_{\nu-1}) = 0.2 \text{ und } \Pr(B_\nu \,|\, B_{\nu-1}) = 0.4$$

sind auch die beiden anderen Übergangswahrscheinlichkeiten festgelegt:

$$\Pr(B_\nu \,|\, A_{\nu-1}) = 0.8 \text{ und } \Pr(A_\nu \,|\, B_{\nu-1}) = 0.6.$$

## Homogene Markovketten

Eine Anwendbarkeit der Markovketten auf praktische Probleme ist meist nicht gegeben, wenn nicht weitere einschränkende Voraussetzungen getroffen werden.

**Definition:** Sind alle Übergangswahrscheinlichkeiten unabhängig vom betrachteten Zeitpunkt $\nu$, so bezeichnet man die Markovkette als **homogen** (englisch: *homogeneous*). Im Fall $M$ = 2 verwenden wir hierfür folgende Abkürzungen:

$$\mathrm{Pr}(A\,|\,A) = \mathrm{Pr}(A_\nu\,|\,A_{\nu-1}), \quad \mathrm{Pr}(A\,|\,B) = \mathrm{Pr}(A_\nu\,|\,B_{\nu-1}),$$

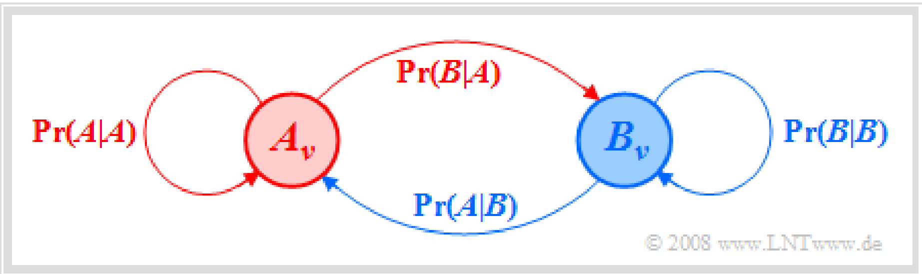
$$\mathrm{Pr}(B\,|\,A) = \mathrm{Pr}(B_\nu\,|\,A_{\nu-1}), \quad \mathrm{Pr}(B\,|\,B) = \mathrm{Pr}(B_\nu\,|\,B_{\nu-1}).$$

Damit lauten die beiden Ereigniswahrscheinlichkeiten einer binären homogenen Markovkette, die im Gegensatz zu den bedingten Übergangswahrscheinlichkeiten absolute Wahrscheinlichkeiten darstellen:

$$\mathrm{Pr}(A_\nu) = \mathrm{Pr}(A\,|\,A)\cdot\mathrm{Pr}(A_{\nu-1}) \;+\; \mathrm{Pr}(A\,|\,B)\cdot\mathrm{Pr}(B_{\nu-1}),$$

$$\mathrm{Pr}(B_\nu) = \mathrm{Pr}(B\,|\,A)\cdot\mathrm{Pr}(A_{\nu-1}) \;+\; \mathrm{Pr}(B\,|\,B)\cdot\mathrm{Pr}(B_{\nu-1}).$$

Diesen Zusammenhang kann man auch aus dem nachfolgend dargestellten **Markovdiagramm** ablesen. Die Summe der abgehenden Pfeile eines Ereignisses ($A_\nu$ bzw. $B_\nu$) ergibt sich stets zu 1.




Sie können das „Einschwingverhalten” der Ereigniswahrscheinlichkeiten einer solchen binären Markovkette mit dem folgenden Interaktionsmodul berechnen und anzeigen lassen:

**Ereigniswahrscheinlichkeiten bei einer Markovkette erster Ordnung**

# Stationäre Wahrscheinlichkeiten

Wichtige Eigenschaften von Zufallsprozessen sind **Stationarität** und **Ergodizität** (Näheres hierzu im Kapitel 4.4). Hier werden diese Begriffe vorausgreifend auf Markovketten angewandt.

**Definition:** Sind bei einer Markovkette neben den Übergangswahrscheinlichkeiten auch alle Ereigniswahrscheinlichkeiten unabhängig vom Zeitpunkt $\nu$, so bezeichnet man sie als **stationär** (englisch: *stationary*). Man verzichtet dann auf den Index $\nu$ und schreibt im binären Fall:

$$\Pr(A_\nu) = \Pr(A) \quad \text{bzw.} \quad \Pr(B_\nu) = \Pr(B).$$

Diese Größen nennt man auch die **ergodischen Wahrscheinlichkeiten** der Markovkette.

Stationäre Markovketten weisen die nachfolgend genannten Besonderheiten auf:

- Zur Berechnung der ergodischen Wahrscheinlichkeiten einer binären Markovkette ($M$ = 2) kann man folgende Gleichungen verwenden:

$$\Pr(A) = \Pr(A\,|\,A) \cdot \Pr(A) \;+\; \Pr(A\,|\,B) \cdot \Pr(B),$$

$$\Pr(B) = \Pr(B\,|\,A) \cdot \Pr(A) \;+\; \Pr(B\,|\,B) \cdot \Pr(B).$$

- Da diese Gleichungen linear voneinander abhängen, darf man nur eine davon benutzen. Als zweite Bestimmungsgleichung kann man beispielsweise $\Pr(A) + \Pr(B) = 1$ verwenden.
- Aus diesen Gleichungen ergeben sich die ergodischen Wahrscheinlichkeiten zu

$$\Pr(A) = \frac{\Pr(A\,|\,B)}{\Pr(A\,|\,B) + \Pr(B\,|\,A)}\,, \qquad \Pr(B) = \frac{\Pr(B\,|\,A)}{\Pr(A\,|\,B) + \Pr(B\,|\,A)}\,.$$

Bei einer stationären Markovkette treten sehr lange nach Einschalten der Kette ($\nu \to \infty$) stets die ergodischen Wahrscheinlichkeiten auf, unabhängig von den Startbedingungen $\Pr(A_0)$ und $\Pr(B_0)$.

**Beispiel:** Wir betrachten eine binäre Markovkette mit den beiden Ereignissen $A$ und $B$ und den Übergangswahrscheinlichkeiten $\Pr(A\,|\,A) = 0.4$ und $\Pr(B\,|\,B) = 0.8$. Weiterhin setzen wir voraus, dass jede Realisierung dieser Kette zum Startzeitpunkt $\nu = 0$ mit dem Ereignis $A$ beginnt. Man erhält dann die nachfolgend aufgelisteten Ereigniswahrscheinlichkeiten:

| $\nu$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| $\Pr(A_\nu)$ | 1.000 | 0.400 | 0.280 | 0.256 | 0.251 | 0.250 | 0.250 |
| $\Pr(B_\nu)$ | 0.000 | 0.600 | 0.720 | 0.744 | 0.749 | 0.750 | 0.750 |



Es handelt sich hier im strengen Sinne um eine nichtstationäre Markovkette, die jedoch bereits nach sehr kurzer Zeit (nahezu) eingeschwungen ist. Zu späteren Zeitpunkten ($\nu > 5$) werden die Ereigniswahrscheinlichkeiten $\Pr(A_\nu) \approx 1/4$ und $\Pr(B_\nu) \approx 3/4$ nicht mehr gravierend verändert.

*Anmerkung:* Aus der Angabe $\Pr(A_{\nu=5}) = 0.250$ bzw. $\Pr(B_{\nu=5}) = 0.750$ sollte nicht geschlossen werden, dass die Markovkette zum Zeitpunkt $\nu = 5$ schon vollkommen eingeschwungen ist. Die exakten Werte sind $\Pr(A_{\nu=5}) = 0.25024$ und $\Pr(B_{\nu=5}) = 0.74976$.

# Matrix-Vektordarstellung (1)

Homogene Markovketten können mit Vektoren und Matrizen sehr kompakt dargestellt werden. Dies empfiehlt sich insbesondere, wenn mehr als zwei Ereignisse betrachtet werden:

$$E_\nu \in G = \{E_1, E_2, \ldots, E_\mu, \ldots, E_M\}.$$

Für die Matrix-Vektorendarstellung verwenden wir folgende Nomenklatur:

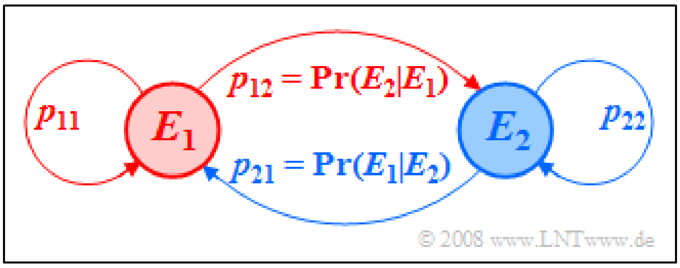
- Die $M$ Wahrscheinlichkeiten zum Zeitpunkt $\nu$ fasst man zu einem Spaltenvektor zusammen:

$$\mathbf{p}^{(\nu)} = \begin{bmatrix} p_1{}^{(\nu)} \\ \ldots\ldots \\ p_M{}^{(\nu)} \end{bmatrix} \quad \text{mit} \quad p_\mu{}^{(\nu)} = \Pr(E_\nu = E_\mu).$$

- Die Übergangswahrscheinlichkeiten werden durch eine $M$ x $M$-Matrix ausgedrückt:

$$\mathbf{P} = (p_{ij}) = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \ldots & \ldots & \ldots & \ldots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix} \quad \text{mit} \quad p_{ij} = \Pr(E_{\nu+1} = E_j \,|\, E_\nu = E_i).$$

Die nebenstehende Abbildung verdeutlicht diese Nomenklatur am Beispiel $M = 2$.




Der neue Ereigniswahrscheinlichkeitsvektor nach einem Schritt lautet:

$$\mathbf{p}^{(\nu+1)} = \mathbf{P}^{\rm T} \cdot \mathbf{p}^{(\nu)}.$$

$\mathbf{P}^{\rm T}$ bezeichnet hierbei die *transponierte Matrix* zu **P**. Nach $n$ Schritten ergibt sich somit

$$\mathbf{p}^{(\nu+n)} = \left(\mathbf{P}^{\rm T}\right)^n \cdot \mathbf{p}^{(\nu)}.$$

Im Grenzübergang ($n \to \infty$) erreicht man dann stets die Stationarität der Markovkette:

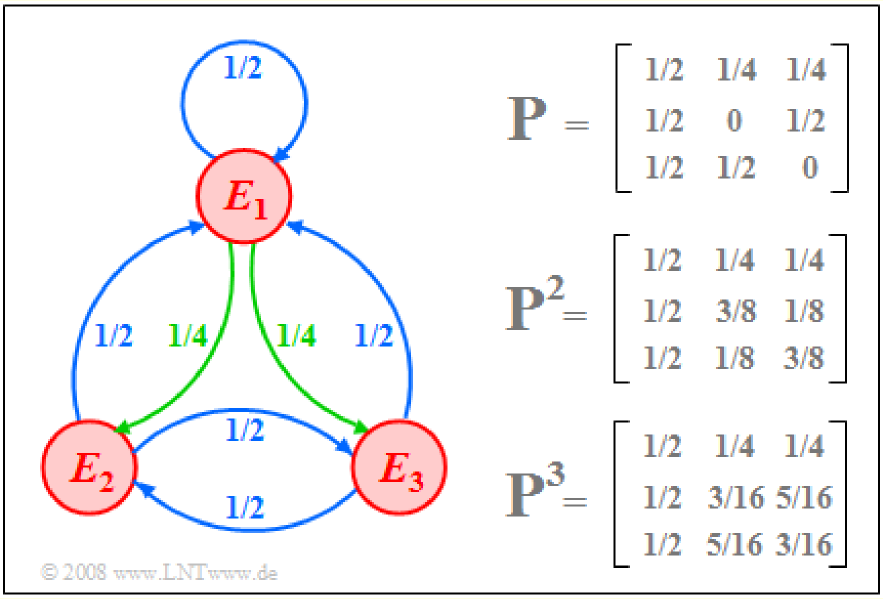
$$\lim_{n\to\infty} \mathbf{p}^{(\nu+n)} = \lim_{n\to\infty} \left(\mathbf{P}^{\rm T}\right)^n \cdot \mathbf{p}^{(\nu)} = \mathbf{p}_{\rm erg} = \begin{bmatrix} p_1 \\ \ldots \\ p_M \end{bmatrix}.$$

Die Wahrscheinlichkeiten $p_1, \ldots, p_M$ werden als die **ergodischen Wahrscheinlichkeiten** bezeichnet. Multipliziert man die Übergangsmatrix unendlich oft mit sich selbst und benennt das Ergebnis mit $\mathbf{P}_{\rm erg}$, so besteht die resultierende Matrix aus $M$ gleichen Spalten:

$$\mathbf{P}_{\rm erg} = \lim_{n\to\infty} \mathbf{P}^n = \begin{bmatrix} p_1 & p_2 & \cdots & p_M \\ p_1 & p_2 & \cdots & p_M \\ \ldots & \ldots & \ldots & \ldots \\ p_1 & p_2 & \cdots & p_M \end{bmatrix}.$$

# Matrix-Vektordarstellung (2)

**Beispiel:** Wir betrachten eine Markovkette mit den Ereignissen $E_1$, $E_2$ und $E_3$. Die Grafik zeigt das Markovdiagramm, die Übergangsmatrix **P** und deren Potenzen. Im Markovdiagramm sind alle Übergangswahrscheinlichkeiten „1/2” blau eingezeichnet; grün kennzeichnet „1/4”.




Beim Start ($\nu = 0$) sind alle Ereignisse gleichwahrscheinlich. Dann gilt für den Zeitpunkt $\nu = 1$:

$$\mathbf{p}^{(1)} = \mathbf{P}^{\mathrm{T}} \cdot \mathbf{p}^{(0)} = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 1/4 & 0 & 1/2 \\ 1/4 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/4 \\ 1/4 \end{bmatrix}.$$

Hieraus ist zu ersehen, dass man von der Matrix **P** durch den Austausch der Zeilen und Spalten zur transponierten Matrix $\mathbf{P}^{\mathrm{T}}$ kommt. Zum Zeitpunkt $\nu = 2$ (und auch zu allen späteren Zeiten) ergeben sich die gleichen Wahrscheinlichkeiten:

$$\mathbf{p}^{(2)} = \mathbf{P}^{\mathrm{T}} \cdot \mathbf{p}^{(1)} = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 1/4 & 0 & 1/2 \\ 1/4 & 1/2 & 0 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/4 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/4 \\ 1/4 \end{bmatrix}.$$

Das bedeutet: Die ergodischen Wahrscheinlichkeiten sind 1/2, 1/4 und 1/4. Dieses Ergebnis hätte man auch direkt aus der ergodischen Matrix ablesen können:

$$\mathbf{P}_{\mathrm{erg}} = \lim_{n \to \infty} \mathbf{P}^n = \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{bmatrix}.$$

Diese ergibt sich durch fortlaufende Multiplikation der Übergangsmatrix mit sich selbst. Im obigen Bild sind die Potenzen $\mathbf{P}^2$ und $\mathbf{P}^3$ angegeben, die sich der Matrix $\mathbf{P}_{\mathrm{erg}}$ annähern.