Statistische Begriffe und Grundlagen

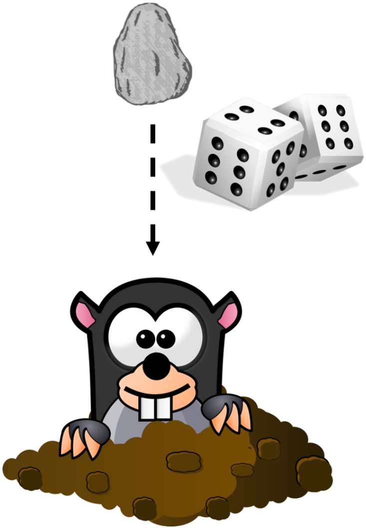
Prof. Dr. Jörg Frochte

Maschinelles Lernen

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

Einige Grundbegriffe der Statistik

- Grundlage ist ein **Zufallsexperiment** dessen Ausgang für einen Beobachter unbekannt oder zumindest nicht vollkommen sicher ist.
- Beispiel: Würfeln mit einem Spielwürfel (Werte 1 bis 6). Man kann vor dem Werfen nicht sagen, welche Zahl es sein wird, obwohl der Würfel den Gesetzen der klassischen Mechanik gehorcht.
- Eigenschaft die wir für ein solches Experiment fordern: Es muss zumindest theoretisch ein beliebig oft wiederholbarer Vorgang sein.

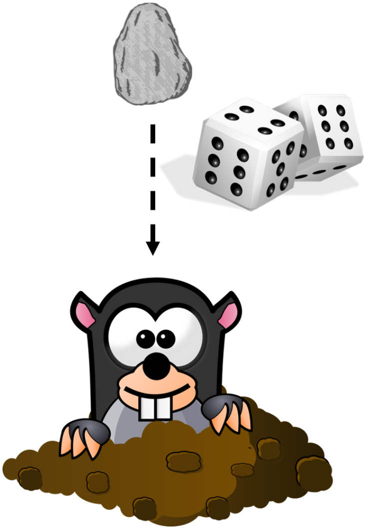


Einige Grundbegriffe der Statistik

- Die Menge der möglichen Ergebnisse eines solchen Experimentes ist der **Ergebnisraum**. Beim einmaligen Würfeln wären das die Ziffern 1 bis 6. Man notiert:

$$\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$$

- Die Elemente ω_i dieser Menge Ω sind die **Ergebnisse** und mögliche einzelne Ausgänge unseres Zufallsexperiments.
- Es gibt auch viele Spiele, in denen es z. B. die Regel gibt *Bei einer 5 oder 6 ....*



Einige Grundbegriffe der Statistik

- Daher geht es ganz allgemein um Teilmengen von Ω.
- Eine solche Teilmenge A von Ω heißt **Ereignis**$\neq$Ergebnisse.
- Ein solches Ereignis $A \subseteq \Omega$ tritt genau dann ein, wenn ein Ergebnis ω auftritt mit $\omega \in A$, z.B. $A = \{5, 6\}$.
- Die Menge aller Ereignisse, also aller möglichen Teilmengen von Ω wird mit $\mathcal{P}(\Omega)$ bezeichnet und heißt **Ereignisraum**.
- Die leere Menge $\emptyset$ hingegen ist ein unmögliches Ereignis.

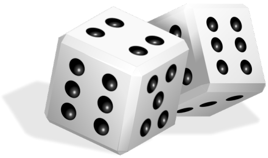
Bei einer 5 oder 6 wiederhole den Kurs, gehe nicht über Los, ziehe keine ECTS ein

- Wir definieren die **gemessene Wahrscheinlichkeit** P eines Ereignisses A durch den Quotienten

$$P(A) = \frac{\text{Anzahl der Fälle, in denen } A \text{ aufgetreten ist}}{\text{Anzahl aller Versuche}} .$$

- Die *wirkliche* Wahrscheinlichkeit würde sich jedoch erst ergeben, wenn wir die Anzahl der Versuche gegen unendlich laufen lassen.

Axiome von Andrei Nikolajewitsch Kolmogorow (1903 – 1987)



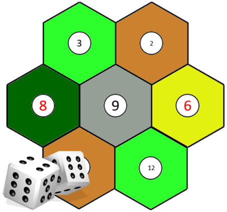
1. Axiom $P(A) \geq 0$
2. Axiom $P(\Omega) = 1$
3. Axiom $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

Beispiel für Punkt 3

Nehmen wir das Würfelbeispiel mit $A = \{5, 6\}$ und $B = \{1, 2\}$, womit $A \cap B = \emptyset$ gegeben ist. Nun können wir eine Aussage über die Wahrscheinlichkeit der Vereinigungsmenge $A \cup B = \{1, 2, 5, 6\}$ machen. Hier addieren wir die die Wahrscheinlichkeiten von A und B. Bei einem fairen Würfel gilt $P(A) = P(B) = 2/6 = 1/3$ und somit $P(A \cup B) = 4/6 = 2/3$. Dass die Schnittmenge leer ist, bildet eine wichtige Voraussetzung, wie man sich schnell klar macht, wenn man den Fall $A = \{3, 4, 5, 6\}$ und $B = \{1, 2, 3, 4\}$ betrachtet. Hier gilt $P(A) = P(B) = 2/3$ und würde man beide addieren, würde die 3 quasi doppelt gezählt.

Bedingte Wahrscheinlichkeit

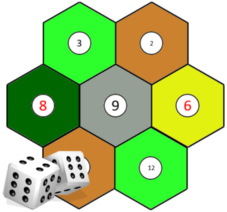
- Wir nehmen nun einen zweiten Würfel hinzu.
- Nehmen wir mal an, es gäbe einen roten und einen blauen Würfel.
- Geht man alle Kombinationen der beiden Würfel durch, erhält man folgende Tabelle:



2	3	4	5	6	7	8	9	10	11	12
1/36	1/18	1/12	1/9	5/36	1/6	5/36	1/9	1/12	1/18	1/36

Wahrscheinlichkeiten für die Summe eines Würfelergebnisse mit zwei Würfeln

- Das Ganze gilt, wenn man mit zwei Würfeln gleichzeitig würfelt.
- Die Lage ändert sich jedoch dramatisch, wenn man z. B. erst mit Rot und dann mit Blau würfelt.



- Denn würfelt man mit rot eine 1, so ergeben sich andere Wahrscheinlichkeiten für den Ausgang nach dem Wurf mit blau.
- Dieser steht als Ereignis noch aus und weiß nichts von dem Wurf mit rot.
- Wir erhalten die Werte aus der folgenden Tabelle als Wahrscheinlichkeiten.

2	3	4	5	6	7	8	9	10	11	12
1/6	1/6	1/6	1/6	1/6	1/6	0	0	0	0	0

Wahrscheinlichkeiten für die Summe eines Würfelergebnisse mit zwei Würfeln, wenn einer bereits eine 1 zeigt

Wichtig: Vorher und nachher unterscheiden

Es ist sehr unwahrscheinlich, nämlich $\left(\frac{1}{6}\right)^5 = \frac{1}{7776}$, 5 mal hintereinander eine 6 zuwürfen, wenn man dies vor dem ersten Wurf ankündigt. Hat man schon 4 mal eine 6 gewürfelt, ist die Chance beim letzten Wurf wieder bei $1/6$, denn der Würfel hat kein Gedächnis.

- Wahrscheinlichkeiten von Ereignissen verändern sich also, wenn bereits andere Ereignisse eingetreten sind.
- Man spricht hier von **bedingte Wahrscheinlichkeiten**.
- Eine solche bedingte Wahrscheinlichkeit dafür, dass A eintritt unter der Bedingung, dass B bereits eingetreten ist, notiert man als $P(A \mid B)$.
- Um mit dieser Art von Wahrscheinlichkeiten zu rechnen, ist der **Satz von Bayes** wesentlich:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \tag{1}$$

- Ein Nutzen dieses Satzes ist es, dass man mit einer bekannten bedingten Wahrscheinlichkeit $P(B \mid A)$, die einen nicht so sehr interessiert, eine andere bedingte Wahrscheinlichkeit $P(A \mid B)$ ausrechnen kann, die einem sehr wohl wichtig ist.

Beispiel

- Die Teilnehmer an einer regelmäßigen Veranstaltung kommen entweder mit dem Auto oder mit öffentlichen Verkehrsmitteln.
- Das Ereignis A steht für „Teilnehmer kommt mit dem Auto“ und das Ereignis B bedeutet, dass der Teilnehmer auch pünktlich ist.
- Ein Teilnehmer fährt zu 70% mit dem Auto ($P(\text{Auto}) = P(A) = 0.7$).
- In diesen Fällen ist er zu 80% pünktlich ($P(\text{pünktlich} \mid \text{Auto}) = P(B \mid A) = 0.8$).
- Generell is er zu 60% pünktlich ($P(\text{pünktlich}) = P(B) = 0.6$).
- Ziel: Die Wahrscheinlichkeit berechnen, dass er mit dem Auto gekommen ist, wenn er pünktlich ist, also $P(A \mid B)$ bzw. $P(\text{Auto} \mid \text{pünktlich})$.

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \rightsquigarrow P(\text{Auto} \mid \text{pünktlich}) = \frac{\frac{8}{10} \cdot \frac{7}{10}}{\frac{6}{10}} = \frac{14}{15} \approx 0.933$$

Wir sind uns zu $\approx 93\%$ sicher, dass er mit dem Auto gekommen ist, wenn wir beobachtet haben, dass er pünktlich ist.

- Um den Satzes von Bayes in eine für uns umfassendere und nützlichere Darstellung zu bringen, benötigen wir den **Satz von der totalen Wahrscheinlichkeit**.
- Diesen kann man anwenden, wenn man eine Menge von Ereignissen hat, die den ganzen Wahrscheinlichkeitsraum abdeckt, aber kein Ergebnis gemeinsam haben.

Satz von der totalen Wahrscheinlichkeit

Seien $B_1, B_2, \ldots$ paarweise disjunkte Ereignisse mit $P(B_j) > 0$ für alle j und

$$\bigcup_{j=1}^{\infty} B_j = \Omega,$$

also alle möglichen Ergebnisse – aufgepasst, es geht nicht um Ereignisse denn dann wäre von $\mathcal{P}(\Omega)$ die Rede – sind in den B_j enthalten, dann gilt:

$$P(A) = \sum_{j=1}^{\infty} P\left(A \mid B_j\right) \cdot P\left(B_j\right) \tag{2}$$

Beispiel für den Satz von der totalen Wahrscheinlichkeit

- Wir haben wiedermal einen sechsseitigen Würfel.
- Wir unterscheidenden haben drei Ereignisse $B_1 = \{1, 2\}$, $B_2 = \{3, 4\}$, $B_3 = \{5, 6\}$ die paarweise disjunkt sind und deren Vereinigung alle Ergebnisse abdecken.
- Wir wollen die Wahrscheinlichkeit von $A = \{1, 6\}$ berechnen.
- Wir wissen

$$P(A \mid B_1) = 0.5 = P(A \mid B_3) \quad P(A \mid B_2) = 0$$

- Dann erhalten wir mit den Satz von der totalen Wahrscheinlichkeit

$$P(A) = \sum_{j=1}^{\infty} P(A \mid B_j) \cdot P(B_j) = \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{3}$$

- Das hätten wir auch leichter haben können, aber wir brauchen den Satz an einige Stellen für Umformungen.

Anwendungsbeispiel 1 Der schlecht trainierte Roboter

- Robbi kann in einer Situation zwischen 4 Entscheidungen B_1, B_2, B_3, B_4 wählen.
- Je nachdem, welche Entscheidung er trifft, wird er mit einer unterschiedlichen Wahrscheinlichkeit beschädigt.

Entscheidung	B_1	B_2	B_3	B_4
Häufigkeit der Wahl	0.2	0.3	0.35	0.15
Wahrscheinlichkeit beschädigt zu werden	0.8	0.2	0.25	0.7

- Mit dem Satz von der totalen Wahrscheinlichkeit berechnen wir wie groß die Wahrscheinlichkeit ist, beschädigt $(P(A))$ zu werden:

$$P(A) = \sum_{j=1}^{4} P\left(A \mid B_j\right) \cdot P\left(B_j\right) = 0.2 \cdot 0.8 + 0.3 \cdot 0.2 + 0.35 \cdot 0.25 + 0.15 \cdot 0.7 = 0.4125$$

- Also können wir festhalten, dass unser Roboter Robbi aktuell noch in rund 40% der Fälle beschädigt werden wird.

Anwendungsbeispiel 2 Software zur Betrugserkennung (1 von 2)

- A ist „System zeigt Betrug an“ und das B ist „Es liegt ein Betrug vor“.
- Zwei relevante Fragen:
  1. Wie oft gibt es einen Fehlalarm?
  2. Wie effektiv können reale Betrugsfälle verhindert werden?
- Annahmen:
  1. Das System kann einen Betrug mit 96% Sicherheit erkennt, also $P(A \mid B) = 0.96$.
  2. Allerdings gibt das System leider auch bei 1% der Nicht-Betrugsfälle Alarm, also $P(A \mid B^C) = 0.01$.
- B^C steht hierbei für das Komplement von B, also die Menge aller Ereignisse ohne eben diejenigen, die schon in B enthalten waren. Es gilt:

$$P(B) + P(B^C) = 1 \text{ und } B \cup B^C = \emptyset$$

- Es fehlt noch anzunehmen wie häufig solche Betrugsfälle eigentlich vorkommen.

Anwendungsbeispiel 2 Software zur Betrugserkennung (2 von 2)

- Nehmen wir an, dass in 0.01% aller Fälle wirklich ein Betrug vorliegt, also $P(B) = 0.0001$ und damit automatisch $P(B^C) = 0.9999$.
- Wir wollen $P(B \mid A)$, haben hingegen nur $P(A \mid B)$ und nutzen den Satz von Bayes.
- Hier fehlt uns nur $P(A)$, diesmal allerdings im Zähler.
- Wir helfen und durch den **Satz von der totalen Wahrscheinlichkeit** und erhalten:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} = \frac{P(B \mid A) \cdot \left(P(A \mid B) \cdot P(B) + P(A \mid B^C) \cdot P(B^C)\right)}{P(B)}$$

$$\Rightarrow 0.96 = \frac{P(B \mid A) \cdot (0.96 \cdot 0.0001 + 0.01 \cdot 0.9999)}{0.0001}$$

$$\Rightarrow P(B \mid A) = 0.00950966 \approx 0.01$$

- Fazit: Das Ergebnis ist für Betrugserkennung sicherlich noch zu ungenau.
- Wenn es anschlägt, ist es in ca. 1% der Fälle wirklich ein Betrug und in ca. 99% der Fälle ein Fehlalarm.