Mathematische Grundlagen der Computerlinguistik

Wahrscheinlichkeit

Dozentin: Wiebke Petersen

8. Foliensatz

Motivation

- In vielen Bereichen der CL kommt Wahrscheinlichkeitstheorie zur Anwendung, da es oft unmöglich ist, mit rein symbolischen Ansätzen ein vollständiges Bild aller möglichen Strukturen einschließlich Präferenzen bei Ambiguitäten zu gewinnen.
- Wir haben es meist mit einer endlichen oder abzählbar unendlichen Menge von sogenannten Ergebnissen zu tun, deren Wahrscheinlichkeit irgendwie abgeschätzt werden muss.

Bsp.:

- Wahrscheinlichkeit dafür, dass VP → VP PP verwendet wird, vorausgesetzt, man möchte eine VP generieren.
- Wahrscheinlichkeit dafür, dass *chair* eine Nomen ist.

ideales Zufallsexperiment (Modell)

Anforderungen an ein ideales Zufallsexperiment:

- Das Experiment wird unter genau festgelegten Versuchsbedingungen durchgeführt.
- Die Menge der möglichen Ergebnisse ist vor der Durchführung des Experiments bekannt.
- Das Experiment kann zumindest prinzipiell beliebig oft unter gleichen Bedingungen wiederholt werden.

ideales Zufallsexperiment (Modell)

Anforderungen an ein ideales Zufallsexperiment:

- Das Experiment wird unter genau festgelegten Versuchsbedingungen durchgeführt.
- Die Menge der möglichen Ergebnisse ist vor der Durchführung des Experiments bekannt.
- Das Experiment kann zumindest prinzipiell beliebig oft unter gleichen Bedingungen wiederholt werden.

Ergebnisraum

Die Menge der möglichen Ergebnisse eines idealen Zufallsexperiments bildet den Ergebnisraum und wird mit Ω ('Omega') bezeichnet.
Ω wird auch der Stichprobenraum genannt.
Ist der Ergebnisraum nicht leer und abzählbar, dann heißt er diskret.

Zufallsexperiment und Ereignisse

Wir unterscheiden einzelne Ergebnisse und Ereignisse, die Mengen von Ergebnissen sind.

- Ein Ereignis bildet eine Teilmenge von Ω.
- $\emptyset$ ist das unmögliche Ereignis.
- Ω ist das sichere Ereignis.
- Zwei Ereignisse E_1 und E_2 heißen unvereinbar, wenn $E_1 \cap E_2 = \emptyset$.
- Die Einermengen $\{e\}$ $(e \in \Omega)$ heißen Elementarereignisse.
- Das Komplement eines Ereignisses E, also $\overline{E}$, heißt Gegenereignis zu E.

Beispiel Zufallsexperiment: Würfeln mit einem Würfel

- $\Omega =$ {⚀, ⚁, ⚂, ⚃, ⚄, ⚅}
- Der Wurf einer 3 ist das Elementarereignis {⚂} des Zufallexperiments.
- {⚁, ⚃, ⚅} ist das Ereignis ‚Wurf einer geraden Augenzahl‘
- Das Gegenereignis von ‚Wurf einer geraden Augenzahl‘ist ‚Wurf einer ungeraden Augenzahl‘

Beispiel Zufallexperiment: Augensumme bei zweimaligem Würfeln

Summe 2	entspricht dem Ereignis	{⚀⚀}
Summe 3	entspricht dem Ereignis	{⚀⚁, ⚁⚀}
Summe 4	entspricht dem Ereignis	{⚀⚂, ⚂⚀, ⚁⚁}
Summe 5	entspricht dem Ereignis	{⚀⚃, ⚃⚀, ⚁⚂, ⚂⚁}
Summe 6	entspricht dem Ereignis	{⚀⚄, ⚄⚀, ⚁⚃, ⚃⚁, ⚂⚂}
Summe 7	entspricht dem Ereignis	{⚀⚅, ⚅⚀, ⚁⚄, ⚄⚁, ⚂⚃, ⚃⚂}
Summe 8	entspricht dem Ereignis	{⚁⚅, ⚅⚁, ⚂⚄, ⚄⚂, ⚃⚃}
Summe 9	entspricht dem Ereignis	{⚂⚅, ⚅⚂, ⚃⚄, ⚄⚃}
Summe 10	entspricht dem Ereignis	{⚃⚅, ⚅⚃, ⚄⚄}
Summe 11	entspricht dem Ereignis	{⚄⚅, ⚅⚄}
Summe 12	entspricht dem Ereignis	{⚅⚅}

Diskrete Wahrscheinlichkeitsräume

Ein diskreter Wahrscheinlichkeitsraum ist ein Paar $\langle \Omega, P \rangle$, bestehend aus

1. einer nicht leeren, abzählbaren Menge Ω von Ergebnissen (diskreter Ergebnisraum) und
2. einem Wahrscheinlichkeitsmaß $P : \mathscr{POT}(\Omega) \to \mathbb{R}$, so dass
    1. $P(A) \geq 0$ für alle $A \in \mathscr{POT}(\Omega)$;
    2. $P(\Omega) = 1$;
    3. für paarweise disjunkte Mengen $A_n \in \mathscr{POT}(\Omega)$, $n \in \mathbb{N}$ gilt

$$P\Big(\bigcup_{n \in \mathbb{N}} A_n\Big) = \sum_{n \in \mathbb{N}} P(A_n)$$

Diskrete Wahrscheinlichkeitsräume

Es ergeben sich folgende Eigenschaften für Wahrscheinlichkeitsmaße:

1. $P(\emptyset) = 0$
2. Für Ereignisse A, B mit $A \cap B = \emptyset$ gilt $P(A \cup B) = P(A) + P(B)$.
3. $P(A) + P(\overline{A}) = 1$ für alle $A \subseteq \Omega$ (Tertium non datur)
4. Impliziert Ereignis A das Ereignis B (d.h. $A \subseteq B$), dann gilt $P(B \setminus A) = P(B) - P(A)$.
5. Kein Ereignis kann eine Wahrscheinlichkeit über 1 haben.

Diskrete Wahrscheinlichkeitsräume

Bsp.: $\Omega = \{$is-noun, has-plural-s, is-adjective, is-verb$\}$.
Frage: Kann die Funktion f mit

$$\begin{aligned} f(\text{is-noun}) &= 0.45 \\ f(\text{has-plural-s}) &= 0.2 \\ f(\text{is-adjective}) &= 0.25 \\ f(\text{is-verb}) &= 0.3 \end{aligned}$$

zu einem Wahrscheinlichkeitsmaß $f : \mathscr{P}(\Omega) \to \mathbb{R}$ ergänzt werden?

Diskrete Wahrscheinlichkeitsräume

Bsp.: $\Omega = \{\text{is-noun, has-plural-s, is-adjective, is-verb}\}$.
Frage: Kann die Funktion f mit

$$\begin{array}{rcl} f(\text{is-noun}) & = & 0.45 \\ f(\text{has-plural-s}) & = & 0.2 \\ f(\text{is-adjective}) & = & 0.25 \\ f(\text{is-verb}) & = & 0.3 \end{array}$$

zu einem Wahrscheinlichkeitsmaß $f : \mathscr{P}(\Omega) \to \mathbb{R}$ ergänzt werden?
Nein, da dann $f(\Omega) = 0.45 + 0.2 + 0.25 + 0.3 = 1.2 > 1$ wäre.

Diskrete Wahrscheinlichkeitsräume

Besser: $\Omega = \{\text{is-noun-with-plural-s, is-noun-without-plural-s, is-adjective, is-verb}\}$.

$$\begin{aligned} f(\text{is-noun-with-plural-s}) &= 0.09 \\ f(\text{is-noun-without-plural-s}) &= 0.36 \\ f(\text{is-adjective}) &= 0.25 \\ f(\text{is-verb}) &= 0.3 \end{aligned}$$

Laplace-Räume

Laplace-Räume sind diskrete Wahrscheinlichkeitsräume, in denen alle Ergebnisse gleich wahrscheinlich sind.
Bsp.: Würfelexperiment. $\Omega = \{1,2,3,4,5,6\}$.
Jedes Ergebnis hat die Wahrscheinlichkeit $\frac{1}{|\Omega|} = \frac{1}{6}$.

In Laplace-Räumen gilt also

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

.

Beispiel Laplace-Raum: zweimaliges Würfeln und Augensumme

Augensumme	Ereignis	Wahrscheinlichkeit
2	{⚀⚀}	$\frac{1}{36}$
3	{⚀⚁, ⚁⚀}	$\frac{2}{36}$
4	{⚀⚂, ⚂⚀, ⚁⚁}	$\frac{3}{36}$
5	{⚀⚃, ⚃⚀, ⚁⚂, ⚂⚁}	$\frac{4}{36}$
6	{⚀⚄, ⚄⚀, ⚁⚃, ⚃⚁, ⚂⚂}	$\frac{5}{36}$
7	{⚀⚅, ⚅⚀, ⚁⚄, ⚄⚁, ⚂⚃, ⚃⚂}	$\frac{6}{36}$
8	{⚁⚅, ⚅⚁, ⚂⚄, ⚄⚂, ⚃⚃}	$\frac{5}{36}$
9	{⚂⚅, ⚅⚂, ⚃⚄, ⚄⚃}	$\frac{4}{36}$
10	{⚃⚅, ⚅⚃, ⚄⚄}	$\frac{3}{36}$
11	{⚄⚅, ⚅⚄}	$\frac{2}{36}$
12	{⚅⚅}	$\frac{1}{36}$

Beispiel Laplace-Raum: Geburtstagsproblem

Bsp.: Wahrscheinlichkeit dafür, dass in einer Gruppe von 30 Personen mindestens zwei am gleichen Tag Geburtstag haben.

Beispiel Laplace-Raum: Geburtstagsproblem

Bsp.: Wahrscheinlichkeit dafür, dass in einer Gruppe von 30 Personen mindestens zwei am gleichen Tag Geburtstag haben.
Vereinfachung: Wir ignorieren Schaltjahre und saisonale Variationen.
D.h., Wahrscheinlichkeit dafür, an einem bestimmten Tag Geburtstag zu haben, ist $\frac{1}{365}$.

Beispiel Laplace-Raum: Geburtstagsproblem

Bsp.: Wahrscheinlichkeit dafür, dass in einer Gruppe von 30 Personen mindestens zwei am gleichen Tag Geburtstag haben.
Vereinfachung: Wir ignorieren Schaltjahre und saisonale Variationen.
D.h., Wahrscheinlichkeit dafür, an einem bestimmten Tag Geburtstag zu haben, ist $\frac{1}{365}$.
Wahrscheinlichkeitsraum:

- $\Omega = \{1, \ldots, 365\}^{30}$, also alle Folgen von 30 Zahlen aus $\{1, \ldots, 365\}$.
- $|\Omega| = 365^{30}$. Alle Folgen sind gleichwahrscheinlich (Laplace-Raum).

Hinweis

Für die Modellierung als Laplace-Raum ist es unerläßlich, die Geburtstagsverteilung als Urnenproblem mit Beachtung der Reihenfolge zu betrachten.
Würde die Reihenfolge vernachlässigt und Ω als die Menge aller ungeordneten Kombinationen möglicher Geburtstagsverteilungen betrachtet (also $|\Omega| = \binom{n+k-1}{k}$), so wären die Ereignisse in dem Ereignisraum nicht gleichwahrscheinlich.

Hinweis

Für die Modellierung als Laplace-Raum ist es unerläßlich, die Geburtstagsverteilung als Urnenproblem mit Beachtung der Reihenfolge zu betrachten.
Würde die Reihenfolge vernachlässigt und Ω als die Menge aller ungeordneten Kombinationen möglicher Geburtstagsverteilungen betrachtet (also $|\Omega| = \binom{n+k-1}{k}$), so wären die Ereignisse in dem Ereignisraum nicht gleichwahrscheinlich.
Beispiel:
Die Wahrscheinlichkeit für das Ereignis, dass alle am 1. Januar Geburtstag haben ist $\left(\frac{1}{365}\right)^{30}$ während die Wahrscheinlichkeit dafür, dass die Geburtstage genau auf die ersten 30 Tage des Jahres fallen $\left(\frac{1}{365}\right)^{30} * 30!$ ist.

Beispiel Laplace-Raum: Geburtstagsproblem

Ziel: Wahrscheinlichkeit dafür, dass eine Folge eintritt, in der sich mindestens ein Element wiederholt.

Beispiel Laplace-Raum: Geburtstagsproblem

Ziel: Wahrscheinlichkeit dafür, dass eine Folge eintritt, in der sich mindestens ein Element wiederholt.
Einfacher: Wahrscheinlichkeitsermittlung über das Komplement.
Wieviel Folgen gibt es, in denen sich kein Element wiederholt?

Beispiel Laplace-Raum: Geburtstagsproblem

Ziel: Wahrscheinlichkeit dafür, dass eine Folge eintritt, in der sich mindestens ein Element wiederholt.
Einfacher: Wahrscheinlichkeitsermittlung über das Komplement.
Wieviel Folgen gibt es, in denen sich kein Element wiederholt?

$$365 \times 364 \times \cdots \times (365 - 29) = \frac{365!}{(365-30)!}$$

Beispiel Laplace-Raum: Geburtstagsproblem

Ziel: Wahrscheinlichkeit dafür, dass eine Folge eintritt, in der sich mindestens ein Element wiederholt.
Einfacher: Wahrscheinlichkeitsermittlung über das Komplement.
Wieviel Folgen gibt es, in denen sich kein Element wiederholt?

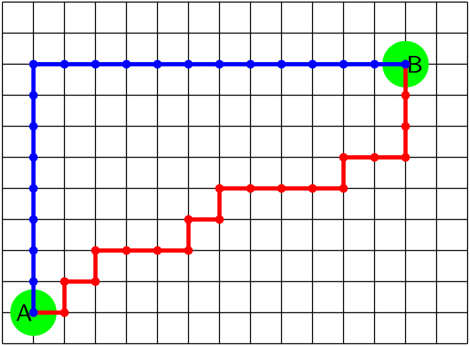
$$365 \times 364 \times \cdots \times (365 - 29) = \frac{365!}{(365-30)!}$$

$\Rightarrow$ Wahrscheinlichkeit dafür, dass zwei am gleichen Tag Geburtstag haben ist

$$1 - \frac{365!}{365^{30}(365-30)!} \approx 1 - 0.29 = 0.71$$

Beispiel Laplace-Raum: Wege im Gitter

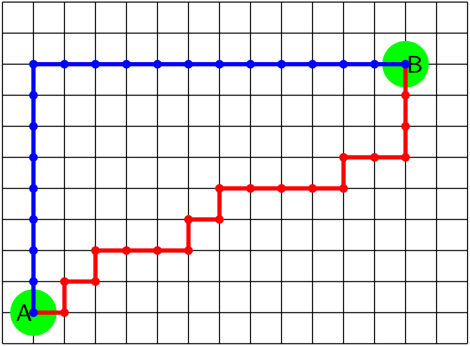
Wege von A nach B. Regel: Schritt nach oben oder Schritt nach rechts.
Beide Richtungen gleichwahrscheinlich.



Wieviele Wege gibt es von A nach B?

Beispiel Laplace-Raum: Wege im Gitter

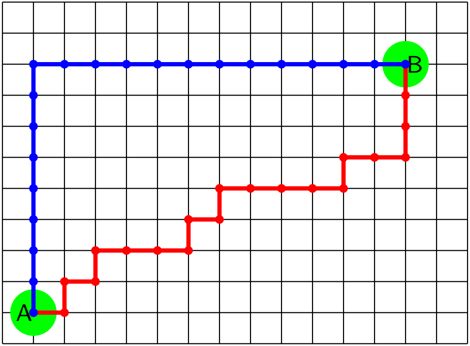
Wege von A nach B. Regel: Schritt nach oben oder Schritt nach rechts.
Beide Richtungen gleichwahrscheinlich.



Wieviele Wege gibt es von A nach B?
Jeder erfolgreiche Weg besteht aus 20 Schritten (8 nach oben, 12 nach rechts).

Beispiel Laplace-Raum: Wege im Gitter

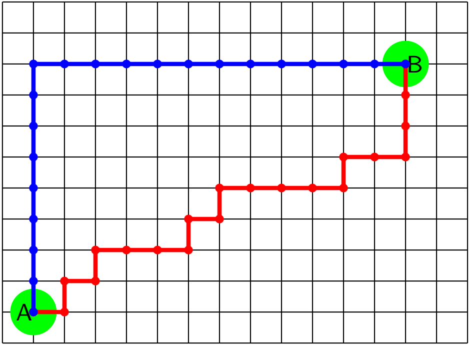
Wege von A nach B. Regel: Schritt nach oben oder Schritt nach rechts.
Beide Richtungen gleichwahrscheinlich.



Wieviele Wege gibt es von A nach B?
Jeder erfolgreiche Weg besteht aus 20 Schritten (8 nach oben, 12 nach rechts).
Zahl der möglichen Wege: $|A \triangleright^{20} B| = \binom{20}{8} = 125970$

Beispiel Laplace-Raum: Wege im Gitter

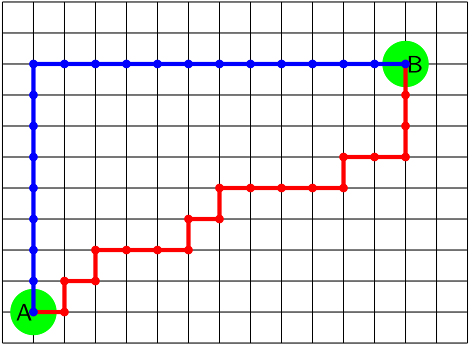
Wege von A nach B. Regel: Schritt nach oben oder Schritt nach rechts.
Beide Richtungen gleichwahrscheinlich.



Wie hoch ist die Wahrscheinlichkeit für das Ereignis nach 20 Schritten bei Punkt B anzukommen?

Beispiel Laplace-Raum: Wege im Gitter

Wege von A nach B. Regel: Schritt nach oben oder Schritt nach rechts. Beide Richtungen gleichwahrscheinlich.

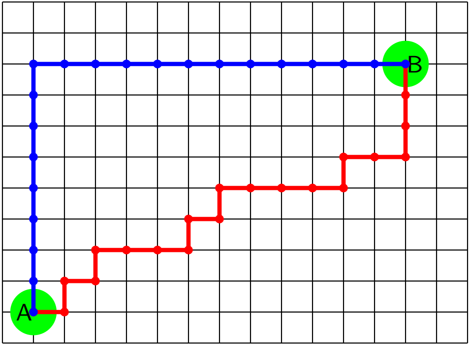


Wie hoch ist die Wahrscheinlichkeit für das Ereignis nach 20 Schritten bei Punkt B anzukommen?

$$P(A \triangleright^{20} B) =$$

Beispiel Laplace-Raum: Wege im Gitter

Wege von A nach B. Regel: Schritt nach oben oder Schritt nach rechts.
Beide Richtungen gleichwahrscheinlich.

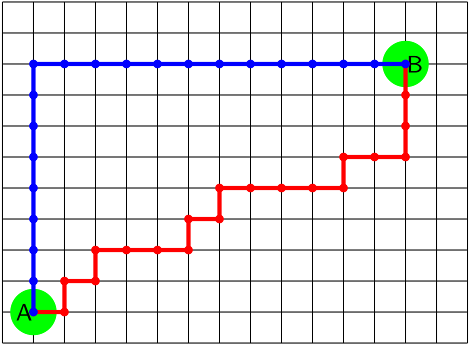


Wie hoch ist die Wahrscheinlichkeit für das Ereignis nach 20 Schritten bei Punkt B anzukommen?

$$P(A \triangleright^{20} B) = \frac{|A \triangleright^{20} B|}{\Omega} =$$

Beispiel Laplace-Raum: Wege im Gitter

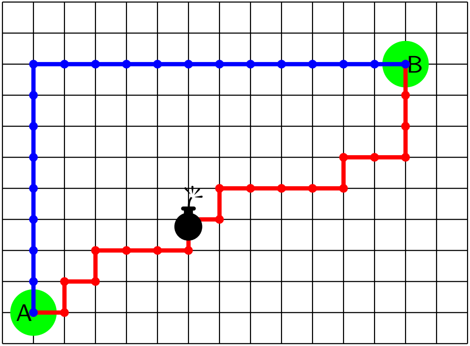
Wege von A nach B. Regel: Schritt nach oben oder Schritt nach rechts.
Beide Richtungen gleichwahrscheinlich.



Wie hoch ist die Wahrscheinlichkeit für das Ereignis nach 20 Schritten bei Punkt B anzukommen?

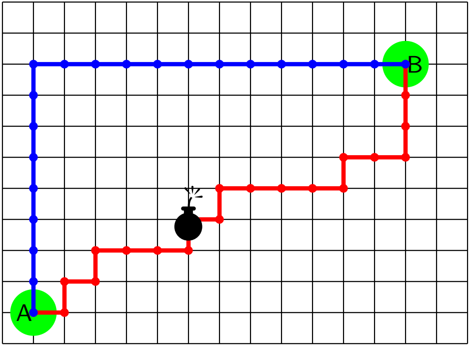
$$P(A \triangleright^{20} B) = \frac{|A \triangleright^{20} B|}{\Omega} = \frac{\binom{20}{8}}{2^{20}} = \frac{125970}{1048576} \approx 0,12$$

Beispiel Laplace-Raum: Gitter mit Bombe



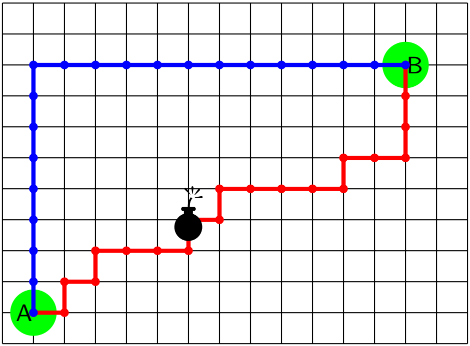
Wie hoch ist die Wahrscheinlichkeit für das Ereignis auf einem Weg von A nach B nicht auf die Bombe zu stoßen? Also die Wahrscheinlichkeit unter allen Wegen von A nach B einen zu wählen, der nicht auf die Bombe stößt.

Beispiel Laplace-Raum: Gitter mit Bombe



Wie hoch ist die Wahrscheinlichkeit für das Ereignis auf einem Weg von A nach B nicht auf die Bombe zu stoßen? Also die Wahrscheinlichkeit unter allen Wegen von A nach B einen zu wählen, der nicht auf die Bombe stößt.
Wege von A nach B, die die Bombe treffen: $\binom{8}{3} \cdot \binom{12}{5} = 56 \cdot 792 = 44352$

Beispiel Laplace-Raum: Gitter mit Bombe

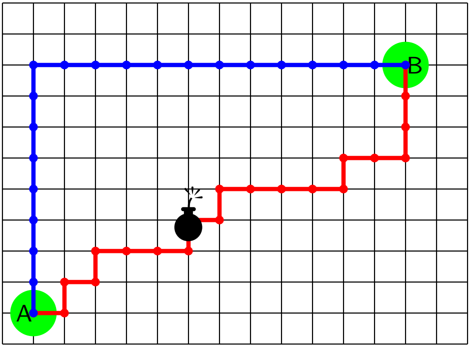


Wie hoch ist die Wahrscheinlichkeit für das Ereignis auf einem Weg von A nach B nicht auf die Bombe zu stoßen? Also die Wahrscheinlichkeit unter allen Wegen von A nach B einen zu wählen, der nicht auf die Bombe stößt.

Wege von A nach B, die die Bombe treffen: $\binom{8}{3} \cdot \binom{12}{5} = 56 \cdot 792 = 44352$

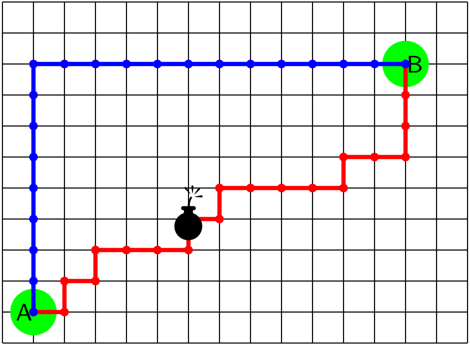
Wahrscheinlichkeit Weg von A nach B ohne Bombe: $\frac{125970-44352}{125970} \approx 0,65$

Beispiel Laplace-Raum: Gitter mit Bombe



Wie hoch ist die Wahrscheinlichkeit für das Ereignis nach 20 Schritten sowohl bei Punkt B anzukommen, als auch auf dem Weg nicht auf die Bombe zu treffen?

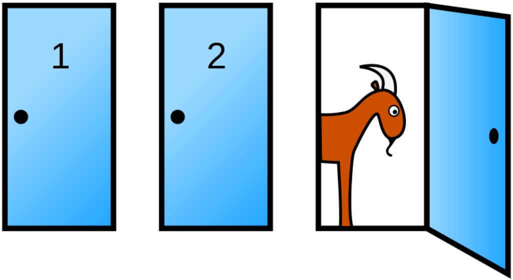
Beispiel Laplace-Raum: Gitter mit Bombe



Wie hoch ist die Wahrscheinlichkeit für das Ereignis nach 20 Schritten sowohl bei Punkt B anzukommen, als auch auf dem Weg nicht auf die Bombe zu treffen?

$$\frac{\binom{20}{8}-\binom{8}{3}\cdot\binom{12}{5}}{2^{20}}=\frac{125970-44352}{1048576}\approx 0{,}078$$

Ziegenproblem



Situation: 3 verschlossene Türen, hinter einer der Türen befindet sich ein Gewinn, hinter zwei Türen befinden sich Nieten (Ziegen).

1. der Kandidat wählt eine Tür
2. der Moderator öffnet von den verbleibenden beiden Türen eine Ziegentür
3. der Kandidat darf die Tür wechseln

Frage: Lohnt sich ein Wechsel?

Bedingte Wahrscheinlichkeiten

Bsp.:

- Wahrscheinlichkeit für eine Produktion VP $\rightarrow$ V NP für die Generierung einer VP, gegeben, dass es sich um das Verb *kisses* (bzw. *sleeps*) handelt.
- Wahrscheinlichkeit dafür, dass *chairs* ein Nomen ist, gegeben die Tatsache, dass das vorangehende Wort ein Artikel ist.
- Wahrscheinlichkeit dafür, dass *chairs* ein Nomen ist, gegeben die Tatsache, dass das nachfolgende Wort ein Artikel ist.

Bedingte Wahrscheinlichkeiten

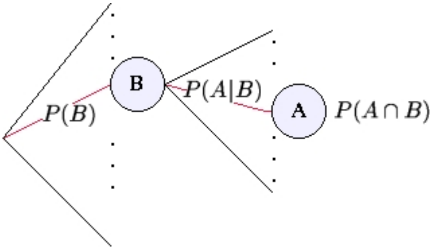
In einem diskreten Wahrscheinlichkeitsraum $\langle \Omega, P \rangle$, gegeben ein Ereignis $A \subseteq \Omega$ mit $P(A) > 0$, ist durch

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

das durch A bedingte Wahrscheinlichkeitsmaß $P(\cdot|A) : \Omega \to \mathbb{R}$ auf $\mathscr{P}(\Omega)$ definiert.

$\langle \mathscr{P}(\Omega), P(\cdot|A) \rangle$ ist ein diskreter Wahrscheinlichkeitsraum.

bedingte Wahrscheinlichkeiten: Produktregel



Produktregel

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

Unabhängigkeit von Ereignissen

Zwei Ereignisse A und B heißen unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$. Das heißt $P(A|B) = P(A)$.

Bsp. Würfelexperiment.

- Die Ereignisse, dass (A) eine gerade Zahl gewürfelt wird und (B) eine Zahl ≤ 2,

Unabhängigkeit von Ereignissen

Zwei Ereignisse A und B heißen unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$. Das heißt $P(A|B) = P(A)$.

Bsp. Würfelexperiment.

- Die Ereignisse, dass (A) eine gerade Zahl gewürfelt wird und (B) eine Zahl ≤ 2, sind unabhängig:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{1,2\})} = 0.5 = P(A)$$

Unabhängigkeit von Ereignissen

Zwei Ereignisse A und B heißen unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$. Das heißt $P(A|B) = P(A)$.

Bsp. Würfelexperiment.

- Die Ereignisse, dass (A) eine gerade Zahl gewürfelt wird und (B) eine Zahl ≤ 2, sind unabhängig:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{1,2\})} = 0.5 = P(A)$$

- Die Ereignisse A wie oben und B, dass genau die 2 gewürfelt wird,

Unabhängigkeit von Ereignissen

Zwei Ereignisse A und B heißen unabhängig, falls $P(A \cap B) = P(A) \cdot P(B)$. Das heißt $P(A|B) = P(A)$.

Bsp. Würfelexperiment.

- Die Ereignisse, dass (A) eine gerade Zahl gewürfelt wird und (B) eine Zahl ≤ 2, sind unabhängig:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{1,2\})} = 0.5 = P(A)$$

- Die Ereignisse A wie oben und B, dass genau die 2 gewürfelt wird, sind nicht unabhängig:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\{2\})}{P(\{2\})} = 1 \neq P(A)$$

Die Formel von Bayes

Thomas Bayes (1701-1761)

Die Formel von Bayes

Ziel: $P(A|B)$ berechnen auf der Grundlage von $P(B|A)$, $P(A)$ und $P(B)$.

Laut Definition gilt

$$P(A \cap B) = P(A|B) \cdot P(B) \text{ und } P(B \cap A) = P(B|A) \cdot (P(A)$$

Daraus ergibt sich

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

Die Formel von Bayes

Man kann das Theorem von Bayes noch verallgemeinern:
Angenommen, es gibt eine endliche oder abzählbar unendliche Folge $(A_i)_{i\in\mathbb{N}}$ von paarweise disjunkten Ereignissen mit $A_i \subseteq \Omega$ und $P(A_i) > 0$ für alle $i \in \mathbb{N}$, die eine Zerlegung von Ω bilden, dann gilt für jedes Ereignis $B \subseteq \Omega$: $(B \cap A_i)_{i\in\mathbb{N}}$ bildet eine disjunkte Zerlegung von B, und daher

$$P(B) = \sum_{i\in\mathbb{N}} P(B|A_i)P(A_i)$$

Die Formel von Bayes

Man kann das Theorem von Bayes noch verallgemeinern:
Angenommen, es gibt eine endliche oder abzählbar unendliche Folge $(A_i)_{i\in\mathbb{N}}$ von paarweise disjunkten Ereignissen mit $A_i \subseteq \Omega$ und $P(A_i) > 0$ für alle $i \in \mathbb{N}$, die eine Zerlegung von Ω bilden, dann gilt für jedes Ereignis $B \subseteq \Omega$: $(B \cap A_i)_{i\in\mathbb{N}}$ bildet eine disjunkte Zerlegung von B, und daher

$$P(B) = \sum_{i\in\mathbb{N}} P(B|A_i)P(A_i)$$

Spezialfall: Zerlegung in A und $\overline{A}$:

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A})$$

$$P(B) = \frac{P(B \cap A)}{P(A)}P(A) + \frac{P(B \cap \overline{A})}{P(\overline{A})}P(\overline{A})$$

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

Die Formel von Bayes

Aus

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

und

$$P(B) = \sum_{i \in \mathbb{N}} P(B|A_i)P(A_i)$$

ergibt sich dann für die Folge $(A_i)_{i \in \mathbb{N}}$ und das Ereignis B wie oben die verallgemeinerte Formel von Bayes:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i \in \mathbb{N}} P(B|A_i)P(A_i)}$$

Die Formel von Bayes

Bsp.: Angenommen, wir interessieren uns für eine relativ seltene Konstruktion, z.B. *Parasitic Gaps*, die ungefähr alle 100.000 Sätze einmal vorkommt.[1] Joe Linguist hat einen Pattern-Matching Algorithmus zur Erkennung von Parasitic Gaps implementiert, der, falls ein Satz ein Parasitic Gap enthält, dies mit einer Wahrscheinlichkeit von 0.95 auch erkennt. Enthält ein Satz kein Parasitic Gap, liefert der Algorithmus mit einer Wahrscheinlichkeit von 0.005 das falsche Ergebnis, dass ein Parasitic Gap in dem Satz vorhanden ist.

Frage: Angenommen, der Test meldet ein Parasitic Gap in einem Satz. Wie wahrscheinlich ist es, dass es sich wirklich um eines handelt?

---

[1] Z.B. *which book did she review _ without reading _?*.

Die Formel von Bayes

Sei G das Ereignis eines parasitic gaps, T das eines positiven Tests.
Wir kennen die Werte $P(G) = \frac{1}{100.000} = 0,00001$, $P(\overline{G}) = 1 - P(G) = 0,99999$, $P(T|G) = 0,95$ und $P(T|\overline{G}) = 0,005$.

Wir wollen $P(G|T)$ berechnen.

Die Formel von Bayes

Sei G das Ereignis eines parasitic gaps, T das eines positiven Tests.
Wir kennen die Werte $P(G) = \frac{1}{100.000} = 0{,}00001$, $P(\overline{G}) = 1 - P(G) = 0{,}99999$, $P(T|G) = 0{,}95$ und $P(T|\overline{G}) = 0{,}005$.

Wir wollen $P(G|T)$ berechnen.

$$P(G|T) = \frac{P(G \cap T)}{P(T)} = \frac{P(T|G) \cdot P(G)}{P(T)}$$

Die Formel von Bayes

Sei G das Ereignis eines parasitic gaps, T das eines positiven Tests.
Wir kennen die Werte $P(G) = \frac{1}{100.000} = 0{,}00001$, $P(\overline{G}) = 1 - P(G) = 0{,}99999$, $P(T|G) = 0{,}95$ und $P(T|\overline{G}) = 0{,}005$.

Wir wollen $P(G|T)$ berechnen.

$$P(G|T) = \frac{P(G \cap T)}{P(T)} = \frac{P(T|G) \cdot P(G)}{P(T)}$$

$P(T)$ lässt sich über $P(T|G)$ und $P(T|\overline{G})$ berechnen:

$$P(T) = P(G \cap G) + P(T \cap \overline{G}) = P(T|G) \cdot P(G) + P(T|\overline{G}) \cdot P(\overline{G})$$

Die Formel von Bayes

Sei G das Ereignis eines parasitic gaps, T das eines positiven Tests.
Wir kennen die Werte $P(G) = \frac{1}{100.000} = 0{,}00001$, $P(\overline{G}) = 1 - P(G) = 0{,}99999$, $P(T|G) = 0{,}95$ und $P(T|\overline{G}) = 0{,}005$.

Wir wollen $P(G|T)$ berechnen.

$$P(G|T) = \frac{P(G \cap T)}{P(T)} = \frac{P(T|G) \cdot P(G)}{P(T)}$$

$P(T)$ lässt sich über $P(T|G)$ und $P(T|\overline{G})$ berechnen:

$$P(T) = P(G \cap G) + P(T \cap \overline{G}) = P(T|G) \cdot P(G) + P(T|\overline{G}) \cdot P(\overline{G})$$

Also erhalten wir

$$P(G|T) = \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|\overline{G})P(\overline{G})}$$

Die Formel von Bayes

Sei G das Ereignis eines parasitic gaps, T das eines positiven Tests.
Wir kennen die Werte $P(G) = \frac{1}{100.000} = 0{,}00001$, $P(\overline{G}) = 1 - P(G) = 0{,}99999$, $P(T|G) = 0{,}95$ und $P(T|\overline{G}) = 0{,}005$.

Wir wollen $P(G|T)$ berechnen.

$$P(G|T) = \frac{P(G \cap T)}{P(T)} = \frac{P(T|G) \cdot P(G)}{P(T)}$$

$P(T)$ lässt sich über $P(T|G)$ und $P(T|\overline{G})$ berechnen:

$$P(T) = P(G \cap G) + P(T \cap \overline{G}) = P(T|G) \cdot P(G) + P(T|\overline{G}) \cdot P(\overline{G})$$

Also erhalten wir

$$P(G|T) = \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|\overline{G})P(\overline{G})}$$

$$= \frac{0.95 \times 0.00001}{0.95 \times 0.00001 + 0.005 \times 0.99999}$$

Die Formel von Bayes

Sei G das Ereignis eines parasitic gaps, T das eines positiven Tests.
Wir kennen die Werte $P(G) = \frac{1}{100.000} = 0{,}00001$, $P(\overline{G}) = 1 - P(G) = 0{,}99999$, $P(T|G) = 0{,}95$ und $P(T|\overline{G}) = 0{,}005$.

Wir wollen $P(G|T)$ berechnen.

$$P(G|T) = \frac{P(G \cap T)}{P(T)} = \frac{P(T|G) \cdot P(G)}{P(T)}$$

$P(T)$ lässt sich über $P(T|G)$ und $P(T|\overline{G})$ berechnen:

$$P(T) = P(G \cap G) + P(T \cap \overline{G}) = P(T|G) \cdot P(G) + P(T|\overline{G}) \cdot P(\overline{G})$$

Also erhalten wir

$$P(G|T) = \frac{P(T|G)P(G)}{P(T|G)P(G) + P(T|\overline{G})P(\overline{G})}$$

$$= \frac{0.95 \times 0.00001}{0.95 \times 0.00001 + 0.005 \times 0.99999}$$

$$\approx 0.002$$

Gefangenenparadoxon

Aus drei zum Tode verurteilten Gefangene (Anton, Bernd und Clemens) wird einer zur Begnadigung ausgewählt. Anton erfährt, dass Bernd hingerichtet wird. Anton erzählt dies Clemens weiter.

- Anton: entweder Clemens wird begnadigt oder er selbst, so dass seine Überlebenschance von $1/3$ auf $1/2$ gestiegen sei.
- Clemens: Überlebenschance von $1/3$ auf $2/3$ gestiegen

Wer von beiden Gefangenen schätzt seine Chancen korrekt ein?

Berkeley 1973 (Simpsonsche Paradoxon)

- Annahmequote Universität
  für Männer: 44,3% für Frauen: 34,6%
- liegt hier eine Benachteiligung der Frauen vor?

Nicht notwendig, siehe: Statistikmodul Mathe-Prisma (Link)

Hinweis

Arbeiten Sie bitte das Mathe-Prisma Modul zur bedingten Wahrscheinlichkeit durch (Link)