ECL

Wahrscheinlichkeitswiederholung

Katja Markert
Folien adaptiert von Frank Keller, University of Edinburgh

Computerlinguistik
Universität Heidelberg

November 10, 2021

- Wahrscheinliche/häufige Worte werden schneller erkannt
- Bei ambigen Wörtern, wird die häufigere Bedeutung schneller abgerufen
- Bei ambigen Sätzen, präferieren wir eine Lesart als wahrscheinlicher
- **Sequenzmodellierung:** Wenn der erste Buchstabe eines Wortes *k* ist, wie wahrscheinlich ist es, dass der nächste *w* ist?
- **Sequenzmodellierung:** Wenn man das Wort *amok* hört, wie wahrscheinlich ist das nächste *laufen*?

- Wie wahrscheinlich ist eine Email Spam, wenn das Wort *Gewinn* darin vorkommt? Was ist die wahrscheinlichste Kategorie eines Textes?
- Welches englische Wort ist die wahrscheinlichste Übersetzung für *Hund*?
- Was ist die wahrscheinlichste Korrektur für einen Tippfehler?

Terminologie

- Experiment: Beobachtung oder Messen
- Elementarergebnis/Outcome: Resultat eines Experiments
- Ergebnismenge/Sample Space: Menge aller möglichen Elementarergebnisse
- Ereignis/event: Jede Teilmenge der Ergebnismenge.

- Dreimal eine Münze werfen ist ein Experiment.
- Die Serie von Kopf/Zahl ist ein Elementarergebnis.
- Die Menge aller möglichen Kopf/Zahlsequenzen ist die Ergebnismenge.
- Die Menge der Elementarergebnisse mit genau zweimal Kopf ist ein Ereignis.

Wir beschäftigen uns vorerst nur mit diskreten /abzählbaren Ergebnismengen. Ereignisraum ist dann die Potenzmenge von S.

Terminologie

- Experiment: Beobachtung oder Messen
- Elementarergebnis/Outcome: Resultat eines Experiments
- Ergebnismenge/Sample Space: Menge aller möglichen Elementarergebnisse
- Ereignis/event: Jede Teilmenge der Ergebnismenge.

- Dreimal eine Münze werfen ist ein Experiment.
- Die Serie von Kopf/Zahl ist ein Elementarergebnis.
- Die Menge aller möglichen Kopf/Zahlsequenzen ist die Ergebnismenge.
- Die Menge der Elementarergebnisse mit genau zweimal Kopf ist ein Ereignis.

Wir beschäftigen uns vorerst nur mit diskreten /abzählbaren Ergebnismengen. Ereignisraum ist dann die Potenzmenge von S.

Terminologie

- Experiment: Beobachtung oder Messen
- Elementarergebnis/Outcome: Resultat eines Experiments
- Ergebnismenge/Sample Space: Menge aller möglichen Elementarergebnisse
- Ereignis/event: Jede Teilmenge der Ergebnismenge.

- Dreimal eine Münze werfen ist ein Experiment.
- Die Serie von Kopf/Zahl ist ein Elementarergebnis.
- Die Menge aller möglichen Kopf/Zahlsequenzen ist die Ergebnismenge.
- Die Menge der Elementarergebnisse mit genau zweimal Kopf ist ein Ereignis.

Wir beschäftigen uns vorerst nur mit diskreten /abzählbaren Ergebnismengen. Ereignisraum ist dann die Potenzmenge von S.

Beispiel: Endliche Ergebnismenge

Zweimal Würfeln. Ergebnismenge:

$$S_1 = \{(x, y) | x = 1, 2, \ldots, 6; y = 1, 2, \ldots, 6\}$$

Alternative Ergebnismenge: die Summe der beiden Würfel.

$$S_2 = \{x | x = 2, 3, \ldots, 12\}$$

Beispiel: Unendliche Ergebnismenge

Werfe eine Münze, solange bis Kopf vorkommt. Was ist die Ergebnismenge?

Beispiel

Mit Bezug auf S_1, beschreibe das Ereignis B, dass die Summe der beiden Würfel 7 ist?

Terminologie

Beispiel: Endliche Ergebnismenge

Zweimal Würfeln. Ergebnismenge:

$$S_1 = \{(x, y)|x = 1, 2, \ldots, 6; y = 1, 2, \ldots, 6\}$$

Alternative Ergebnismenge: die Summe der beiden Würfel.

$$S_2 = \{x|x = 2, 3, \ldots, 12\}$$

Beispiel: Unendliche Ergebnismenge

Werfe eine Münze, solange bis Kopf vorkommt. Was ist die Ergebnismenge?

Beispiel

Mit Bezug auf S_1, beschreibe das Ereignis B, dass die Summe der beiden Würfel 7 ist?

Ereignisse

Geg. eine Ergebnismenge S und zwei Ereignisse A und B:

- **Komplement $\bar{A}$ (auch A'):** alle Elemente von S, die nicht in A sind;
- **Teilmenge $A \subset B$:** alle Elemente von A sind auch Elemente von B;
- **Vereinigung $A \cup B$:** alle Elemente von S, die in A oder B sind;
- **Schnittmenge $A \cap B$:** alle Elemente von S, die in A und B sind.

Axiome

Geg. sei eine diskrete Ergebnismenge S. Wir schreiben Ereignisse mit Großbuchstaben A, B, C, etc. und die **Wahrscheinlichkeit** eines Ereignisses A mit $P(A)$.

Axiome für eine Wahrscheinlichkeitsmaß

Wahrscheinlichkeitsmaß ordnet Ereignissen reelle Zahlen zu mit den folgenden Eigenschaften

1. $P(A) \geq 0$ für jedes $A \subset S$.
2. $P(S) = 1$.
3. Seien $A_1, A_2, A_3, \ldots,$ Ereignisse mit $A_i \cap A_j = \emptyset$, falls $i \neq j$. Dann gilt:

$$P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$$

Wahrscheinlichkeitsraum: Ergebnismenge, Ereignisraum (= Potenzmenge von S) und Wahrscheinlichkeitsmaß.

Theorem: Wahrscheinlichkeit eines Ereignisses

Sei A ein Ereignis in einer abzählbaren Ergebnismenge S und $O_1, O_2, O_3, \ldots,$ die Elementarergebnisse in A, dann $P(A) = P(O_1) + P(O_2) + P(O_3) + \ldots$

Spezialfall

Theorem: Gleichwahrscheinliche Elementarergebnisse

Wenn ein Experiment zu N gleichwahrscheinlichen Elementarergebnissen führt, und n dieser Ergebnisse das Ereignis A bilden , dann $P(A) = \frac{n}{N}$.

Wahrscheinlichkeit eines Ereignisses

Theorem: Wahrscheinlichkeit eines Ereignisses

Sei A ein Ereignis in einer abzählbaren Ergebnismenge S und $O_1, O_2, O_3, \ldots$, die Elementarergebnisse in A, dann $P(A) = P(O_1) + P(O_2) + P(O_3) + \ldots$

Spezialfall

Theorem: Gleichwahrscheinliche Elementarergebnisse

Wenn ein Experiment zu N gleichwahrscheinlichen Elementarergebnissen führt, und n dieser Ergebnisse das Ereignis A bilden , dann $P(A) = \frac{n}{N}$.

Wahrscheinlichkeit eines Ereignisses

Beispiel

Man nehme an alle Buchstaben sind gleichhäufig im Englischen (und voneinander unabhängig, wenn man ein Wort formt). Was ist dann die Wahrscheinlichkeit eines Wortes, das exakt aus drei Vokalen und keinen Konsonanten besteht?

Es gibt $N = 26^3$ 3-B Wörter. Vokale = $\{a, e, i, o, u\}$. Damit ist die Größe von A (= drei Buchstaben, nur Vokale) $n = 5^3$. Damit folgt $P(A) = \frac{n}{N} = \frac{5^3}{26^3} = 0.00711$.

Beispiel

Man nehme an alle Buchstaben sind gleichhäufig im Englischen (und voneinander unabhängig, wenn man ein Wort formt). Was ist dann die Wahrscheinlichkeit eines Wortes, das exakt aus drei Vokalen und keinen Konsonanten besteht?

Es gibt $N = 26^3$ 3-B Wörter. Vokale = $\{a, e, i, o, u\}$. Damit ist die Größe von A (= drei Buchstaben, nur Vokale) $n = 5^3$. Damit folgt $P(A) = \frac{n}{N} = \frac{5^3}{26^3} = 0.00711$.

Wahrscheinlichkeitsregeln

Theoreme: Wahrscheinlichkeitsregeln (abzählbare Ergebnismenge)

1. $P(\bar{A}) = 1 - P(A)$.
2. $P(\emptyset) = 0$ für jede Ergebnismenge S.
3. Wenn $A \subset B$, dann $P(A) \leq P(B)$.
4. $0 \leq P(A) \leq 1$ für jedes Ereignis A.

Zu zeigen: $P(\bar{A}) = 1 - P(A)$
Beweis:

$$
\begin{aligned}
P(\bar{A}) \stackrel{(3)}{=} \textstyle\sum_{x \in \bar{A}} P(x) &= \textstyle\sum_{x \in \bar{A}} P(x) + \sum_{x \in A} P(x) - \sum_{x \in A} P(x) \\
&\stackrel{(3)}{=} \textstyle\sum_{x \in S} P(x) - \sum_{x \in A} P(x) \stackrel{(2)}{=} 1 - P(A)
\end{aligned}
$$

Minibeweise II

Zu zeigen: $P(\emptyset) = 0$ für jede Ergebnismenge S.

Beweis:

$$P(\emptyset) + 1 \stackrel{(2)}{=} P(\emptyset) + P(S) \stackrel{(3)}{=} P(\emptyset \cup S) = P(S) \stackrel{(2)}{=} 1$$

Minibeweise III

Zu zeigen: $A \subset B$, dann $P(A) \leq P(B)$.

Beweis:

$$P(B) = P(A \cup B \setminus A) \stackrel{(3)}{=} P(A) + P(B \setminus A)$$

Hieraus folgt mit Axiom 1 direkt das Gewünschte.

Zu zeigen:
$0 \leq P(A) \leq 1$ für jedes Ereignis A.

Beweis: Folgt direkt aus
$\emptyset \subseteq A \subseteq S$ sowie dem eben Gezeigtem

Additionsregeln

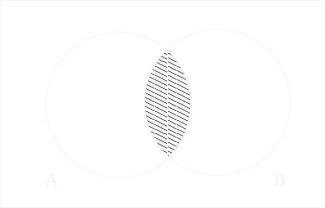
Axiom 3 erlaubt uns, die Wahrscheinlichkeit von Ereignissen, die paarweise disjunkt sind, aufzuaddieren. Was ist mit Ereignissen, die nicht paarweise disjunkt sind?

Theorem: Generelle Additionsregel

Seien A, B zwei Ereignisse in Ergebnismenge S, dann

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Schraffierte Fläche kommt zweimal vor, und muss substrahiert werden.



Additionsregeln

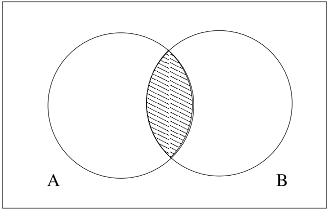
Axiom 3 erlaubt uns, die Wahrscheinlichkeit von Ereignissen, die paarweise disjunkt sind, aufzuaddieren. Was ist mit Ereignissen, die nicht paarweise disjunkt sind?

Theorem: Generelle Additionsregel

Seien A, B zwei Ereignisse in Ergebnismenge S, dann

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Schraffierte Fläche kommt zweimal vor, und muss substrahiert werden.



Additionsregel

Beispiel

Sprache ist in Gehirnhälften lokalisiert. Für die meisten Menschen, hauptsächlich in der linken Hemisphäre. In einigen in der rechten.

Man nehme an, die Wahrscheinlichkeit, Linkshänder zu sein sei $P(A) = 0.15$, und die Wahrscheinlichkeit rechts lateralisiert zu sein $P(B) = 0.05$.

Wenn A und B paarweise disjunkt wären, dann wäre die Wahrscheinlichkeit Linkshänder oder rechts lateralisiert zu sein $P(A \cup B) = P(A) + P(B) = 0.2$.

Dies ist aber nicht der Fall. Es gibt Linkshänder, die rechts lateralisiert sind. Wir wissen, dass $P(A \cap B) = 0.04$.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.16$.

Additionsregel

Beispiel

Sprache ist in Gehirnhälften lokalisiert. Für die meisten Menschen, hauptsächlich in der linken Hemisphäre. In einigen in der rechten.

Man nehme an, die Wahrscheinlichkeit, Linkshänder zu sein sei $P(A) = 0.15$, und die Wahrscheinlichkeit rechts lateralisiert zu sein $P(B) = 0.05$.

Wenn A und B paarweise disjunkt wären, dann wäre die Wahrscheinlichkeit Linkshänder oder rechts lateralisiert zu sein $P(A \cup B) = P(A) + P(B) = 0.2$.

Dies ist aber nicht der Fall. Es gibt Linkshänder, die rechts lateralisiert sind. Wir wissen, dass $P(A \cap B) = 0.04$.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.16$.

Additionsregel

Beispiel

Sprache ist in Gehirnhälften lokalisiert. Für die meisten Menschen, hauptsächlich in der linken Hemisphäre. In einigen in der rechten.

Man nehme an, die Wahrscheinlichkeit, Linkshänder zu sein sei $P(A) = 0.15$, und die Wahrscheinlichkeit rechts lateralisiert zu sein $P(B) = 0.05$.

Wenn A und B paarweise disjunkt wären, dann wäre die Wahrscheinlichkeit Linkshänder oder rechts lateralisiert zu sein $P(A \cup B) = P(A) + P(B) = 0.2$.

Dies ist aber nicht der Fall. Es gibt Linkshänder, die rechts lateralisiert sind. Wir wissen, dass $P(A \cap B) = 0.04$.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.16$.

Zwischenfazit

- Ergebnismenge S enthält alle möglichen Ergebnisse eines Experiments. Ereignisse sind Teilmengen von S.
- für gleichwahrscheinliche Elementarereingnisse: $P(A) = \frac{n}{N}$;
- Rechnen mit Wahrscheinlichkeiten:
  - $P(\bar{A}) = 1 - P(A)$;
  - $P(\emptyset) = 0$;
  - if $A \subset B$, then $P(A) \leq P(B)$;
  - $0 \leq P(B) \leq 1$;
- Additionsregel: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Bedingte Wahrscheinlichkeit

Definition

Seien A und B zwei Ereignisse in Ergebnismenge S und $P(A) \neq 0$ dann ist die bedingte Wahrscheinlichkeit B gegeben A definiert als:

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

Examples

Die Wahrscheinlichkeit eines Unfalls bei Schnee: $P(\text{Unfall}|\text{Schnee})$.

Die Wahrscheinlichkeit *laufen* zu lesen, wenn das vorherige Wort *amok* war. $P(\text{laufen}|\text{amok})$.

Bedingte Wahrscheinlichkeit

Definition

Seien A und B zwei Ereignisse in Ergebnismenge S und $P(A) \neq 0$ dann ist die bedingte Wahrscheinlichkeit B gegeben A definiert als:

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

Examples

Die Wahrscheinlichkeit eines Unfalls bei Schnee: $P(\text{Unfall}|\text{Schnee})$.

Die Wahrscheinlichkeit *laufen* zu lesen, wenn das vorherige Wort *amok* war. $P(\text{laufen}|\text{amok})$.

Bedingte Wahrscheinlichkeit

Multiplikationsregel

Seien A und B zwei Ereignisse in einer Ergebnismenge S und $P(A) \neq 0$, dann gilt:

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

Da $A \cap B = B \cap A$, gilt:

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

Multiplikationsregel

Seien A und B zwei Ereignisse in einer Ergebnismenge S und $P(A) \neq 0$, dann gilt:

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

Da $A \cap B = B \cap A$, gilt:

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

Beispiel

Lateralisierung. Sei $P(A) = 0.15$ die Wahrscheinlichkeit für Linkshänder $P(B) = 0.05$ die Wahrscheinlichkeit für rechts lateralisiert sowie $P(A \cap B) = 0.04$.

Die Wahrscheinlichkeit rechts lateralisiert zu sein, gegeben, dass man Linkshänder ist

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.04}{0.15} = 0.267$$

Die Wahrscheinlichkeit Linkshänder zu sein, gegeben dass man rechts lateralisiert ist.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.04}{0.05} = 0.80$$

Beispiel

Lateralisierung. Sei $P(A) = 0.15$ die Wahrscheinlichkeit für Linkshänder $P(B) = 0.05$ die Wahrscheinlichkeit für rechts lateralisiert sowie $P(A \cap B) = 0.04$.

Die Wahrscheinlichkeit rechts lateralisiert zu sein, gegeben, dass man Linkshänder ist

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.04}{0.15} = 0.267$$

Die Wahrscheinlichkeit Linkshänder zu sein, gegeben dass man rechts lateralisiert ist.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.04}{0.05} = 0.80$$

Beispiel

Lateralisierung. Sei $P(A) = 0.15$ die Wahrscheinlichkeit für Linkshänder $P(B) = 0.05$ die Wahrscheinlichkeit für rechts lateralisiert sowie $P(A \cap B) = 0.04$.

Die Wahrscheinlichkeit rechts lateralisiert zu sein, gegeben, dass man Linkshänder ist

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.04}{0.15} = 0.267$$

Die Wahrscheinlichkeit Linkshänder zu sein, gegeben dass man rechts lateralisiert ist.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.04}{0.05} = 0.80$$

Die Kettenregel

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

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

Kann generalisiert werden zu Ereignissequenzen:

$$P(A_1 \cap A_2 \cap \ldots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\ldots P(A_n|A_1 \ldots A_{n-1})$$

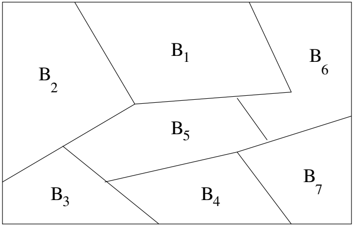
Totale Wahrscheinlichkeit

Theorem

Sind die Ereignisse $B_1, B_2, \ldots, B_k$ eine Partition der Ergebnismenge S und $P(B_i) \neq 0$ für $i = 1, 2, \ldots, k$, dann folgt für jedes Ereignis A in S:

$$P(A) = \sum_{i=1}^{k} P(B_i) P(A|B_i)$$

$B_1, B_2, \ldots, B_k$ sind eine **Partition** von S wenn sie paarweise disjunkt sind und $B_1 \cup B_2 \cup \ldots \cup B_k = S$.



Totale Wahrscheinlichkeit

Beispiel

In einem Gedächtnisexperiment müssen Teilnehmer sich Wörter (B_1), Zahlen (B_2), und Bilder (B_3) merken. Die Wahrscheinlichkeiten der Vorkommnisse sind $P(B_1) = 0.5$, $P(B_2) = 0.4$, $P(B_3) = 0.1$.

Sei A das Erinnerungsereignis. Die Resultate zeigen $P(A|B_1) = 0.4$, $P(A|B_2) = 0.2$, $P(A|B_3) = 0.1$. Berechne $P(A)$.

$$\begin{aligned} P(A) &= \sum_{i=1}^{k} P(B_i)P(A|B_i) \\ &= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) \\ &= 0.5 \cdot 0.4 + 0.4 \cdot 0.2 + 0.1 \cdot 0.1 = 0.29 \end{aligned}$$

Beispiel

In einem Gedächtnisexperiment müssen Teilnehmer sich Wörter (B_1), Zahlen (B_2), und Bilder (B_3) merken. Die Wahrscheinlichkeiten der Vorkommnisse sind $P(B_1) = 0.5$, $P(B_2) = 0.4$, $P(B_3) = 0.1$.

Sei A das Erinnerungsereignis. Die Resultate zeigen $P(A|B_1) = 0.4$, $P(A|B_2) = 0.2$, $P(A|B_3) = 0.1$. Berechne $P(A)$.

$$\begin{aligned} P(A) &= \textstyle\sum_{i=1}^{k} P(B_i)P(A|B_i) \\ &= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3) \\ &= 0.5 \cdot 0.4 + 0.4 \cdot 0.2 + 0.1 \cdot 0.1 = 0.29 \end{aligned}$$

Unabhängigkeit

Definition

Zwei Ereignisse A und B sind unabhängig, genau dann wenn

$$P(B \cap A) = P(A)P(B) \; bzw; P(A|B) = P(A)$$

Unabhängigkeit

1. reduziert die Information, die wir für gemeinsame Verteilungen spezifizieren müssen
2. selten in Sprache

Beispiel

Eine faire Münze wird dreimal geworfen. Sei A: zweimal Kopf am Anfang, B: Zahl am Schluss, C: genau zweimal Zahl in drei Würfen. Zeige, dass A und B unabhängig sind, aber B und C abhängig.
$P(A \cap B) = P(A)P(B)$ aber $P(B)P(C) \neq P(B \cap C)$

Bayes' Theorem

Bayes' Theorem

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

- $P(B|A)$ heißt **posterior probability**.
- $P(A)$ heißt **prior probability**.
- $P(B|A)$ aus $P(A|B)$ zu berechnen ist **Bayes Inversion.**

Beweis:

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

Bayes' Theorem mit totaler Wahrscheinlichkeit

Bayes' Theorem

Sei $B_1, B_2, \ldots, B_k$ eine Partition S und $P(B_i) \neq 0$ für $i = 1, 2, \ldots, k$, dann gilt für jedes Ereignis A mit $P(A) \neq 0$:

$$P(B_r|A) = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)}$$

Diagnose

Example

Die Prävalenz einer Krankheit ist 1/1000. Ein Test, um die Krankheit zu entdecken, wird entwickelt und hat eine 5% Rate von falsch-positiv Ergebnissen. Falsche Negativa hat der Test keine. Wie hoch ist die Wahrscheinlichkeit, dass eine Person, die den Test macht und ein positives Resultat hat, wirklich die Krankheit hat? Man nehme an, wann weiss nichts von Symptomen der Person.

Beispiel

Häufigste Antwort: 95%

Richtige Antwort: 2% Casscells:ea:78

Nur 12% von Teilnehmern geben die richtige Antwort.

Bayes Theorem

- D: Person krank;
- $\bar{D}$: Person nicht krank;
- T: positiver Test.

Bekannt aus der Aufgabenstellung ist:

$$P(D) = 1/1000 = 0.001 \qquad P(\bar{D}) = 1 - P(D) = 0.999$$
$$P(T|\bar{D}) = 5\% = 0.05 \qquad P(T|D) = 1$$

Und damit:

$$P(D|T) = \frac{P(D)P(T|D)}{P(T)} = \frac{0.001 \cdot 1}{P(T)}$$

$$P(\bar{D}|T) = \frac{P(\bar{D})P(T|\bar{D})}{P(T)} = \frac{0.999 \cdot 0.05}{P(T)} = \frac{0.04995}{P(T)}$$

$$P(T) = P(D)P(T|D) + P(\bar{D})P(T|\bar{D}) = 0.001 \cdot 1 + 0.999 \cdot 0.05 = 0.05095$$

$$P(D|T) = \frac{P(D)P(T|D)}{P(T)} = \frac{0.001 \cdot 1}{0.05095} = 0.019$$

Bayes Theorem

- D: Person krank;
- $\bar{D}$: Person nicht krank;
- T: positiver Test.

Bekannt aus der Aufgabenstellung ist:

$$P(D) = 1/1000 = 0.001 \qquad P(\bar{D}) = 1 - P(D) = 0.999$$
$$P(T|\bar{D}) = 5\% = 0.05 \qquad P(T|D) = 1$$

Und damit:

$$P(D|T) = \frac{P(D)P(T|D)}{P(T)} = \frac{0.001 \cdot 1}{P(T)}$$

$$P(\bar{D}|T) = \frac{P(\bar{D})P(T|\bar{D})}{P(T)} = \frac{0.999 \cdot 0.05}{P(T)} = \frac{0.04995}{P(T)}$$

$$P(T) = P(D)P(T|D) + P(\bar{D})P(T|\bar{D}) = 0.001 \cdot 1 + 0.999 \cdot 0.05 = 0.05095$$

$$P(D|T) = \frac{P(D)P(T|D)}{P(T)} = \frac{0.001 \cdot 1}{0.05095} = 0.019$$

- D: Person krank;
- $\bar{D}$: Person nicht krank;
- T: positiver Test.

Bekannt aus der Aufgabenstellung ist:

$$P(D) = 1/1000 = 0.001 \qquad P(\bar{D}) = 1 - P(D) = 0.999$$
$$P(T|\bar{D}) = 5\% = 0.05 \qquad P(T|D) = 1$$

Und damit:

$$P(D|T) = \frac{P(D)P(T|D)}{P(T)} = \frac{0.001 \cdot 1}{P(T)}$$

$$P(\bar{D}|T) = \frac{P(\bar{D})P(T|\bar{D})}{P(T)} = \frac{0.999 \cdot 0.05}{P\,(T)} = \frac{0.04995}{P(T)}$$

$$P(T) = P(D)P(T|D) + P(\bar{D})P(T|\bar{D}) = 0.001 \cdot 1 + 0.999 \cdot 0.05 = 0.05095$$

$$P(D|T) = \frac{P(D)P(T|D)}{P(T)} = \frac{0.001 \cdot 1}{0.05095} = 0.019$$

Bayes Regel als Inversion

Kann Ursache und Folge umdrehen:

$$P(\mathit{Ursache}|\mathit{Effekt}) = \frac{P(\mathit{Ursache})P(\mathit{Effekt}|\mathit{Ursache})}{P(\mathit{Effekt})}$$

- Erlaubt Berechnung, wenn man die eine, aber nicht die andere bedingte Wahrscheinlichkeit weiss.
- Oft ist Ursachenwissen robuster.

Zusammenfassung

- Bedingte Wahrscheinlichkeiten
- Totale Wahrscheinlichkeit
- Satz von Bayes
- Aufgabenblatt 4

Definition: Zufallsvariable

Sei S diskrete Ergebnismenge mit Wahrscheinlichkeitsmaß und X eine reelle Funktion über den Elementen von S, dann heißt X eine Zufallsvariable.

Wir benutzen Großbuchstaben für Zufallsvariablen (e.g., X), und schreiben ihre Werte klein (e.g., x).

Diskrete Zufallsvariablen: Beispiel

Dreifacher Münzwurf. Sei X die Zufallsvariable, die die Anzahl von Kopf denotiert.

Ergebnis	Wahscheinlichkeit	x
HHH	$\frac{1}{8}$	3
HHT	$\frac{1}{8}$	2
HTH	$\frac{1}{8}$	2
THH	$\frac{1}{8}$	2
Ergebnis	Wahrscheinlichkeit	x
TTH	$\frac{1}{8}$	1
THT	$\frac{1}{8}$	1
HTT	$\frac{1}{8}$	1
TTT	$\frac{1}{8}$	0

Daher $P(X = 0) = \frac{1}{8}$, $P(X = 1) = P(X = 2) = \frac{3}{8}$, $P(X = 3) = \frac{1}{8}$.

Wahrscheinlichkeitsverteilungen

Definition

Sei X eine diskrete Zufallsvariable. Dann ist die Funktion $f(x) = P(X = x)$ für alle x im Bild von X die Wahrscheinlichkeitsverteilung von X.

Theorem

Eine Funktion kann nur eine Wahrscheinlichkeitsverteilung einer diskreten Zufallsvariable sein, genau dann wenn für alle $f(x)$ gilt:

1. $f(x) \geq 0$;
2. $\sum_x f(x) = 1$

Beispiel

Beispiel vorher

x	$f(x) = P(X = x)$
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

Kann präziser geschreiben werden als:

$$f(x) = \frac{4 - |3 - 2x|}{8}$$

oder als Vektor $\mathbf{P}(X) = \langle \frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8} \rangle$

Gemeinsame Verteilungen

Seien zwei Ereignisse A und B durch die Zufallsvariablen X mit Wert x und Y mit Wert y beschrieben. Dann können wir schreiben:

$$P(A \cap B) = P(X = x \cap Y = y) = P(X = x, Y = y)$$

Dies bezeichnet man als die gemeinsame Wahrscheinlichkeit von $X = x$ und $Y = y$.

NB: oft wird der Ausdruck gemeinsame Wahrscheinlichkeit und $P(A, B)$ auch für die Wahrscheinlichkeit des Schnitts zweier Ereignisse verwendet.

Definition: Gemeinsame Wahrscheinlichkeitsverteilungen

Seien X und Y zwei diskrete Zufallsvariablen. Dann ist die Funktion $f(x, y) = P(X = x, Y = y)$ für alle Wertepaare (x, y) die gemeinsame Wahrscheinlichkeitsverteilung von X und Y.

Beispiel: Korpusdaten

Korpus mit 100 Wörtern. Man tabelliert die Wörter, ihre Wahrscheinlichkeiten und Häufigkeiten:

w	$c(w)$	$P(w)$	x	y
the	30	0.30	3	1
to	18	0.18	2	1
will	16	0.16	4	1
of	10	0.10	2	1
Earth	7	0.07	5	2
on	6	0.06	2	1
probe	4	0.04	5	2
some	3	0.03	4	2
Comet	3	0.03	5	2
BBC	3	0.03	3	0

Beispiel: Korpusdaten

Wir definieren folgende Zufallsvariablen:

- X: Wortlänge;
- Y: Anzahl Vokale.

Beispiele:

- $f_X(5) = P(\text{Earth}) + P(\text{probe}) + P(\text{Comet}) = 0.14$;
- $f_Y(2) = P(\text{Earth}) + P(\text{probe}) + P(\text{some}) + P(\text{Comet}) = 0.17$.

Beispiel: Korpusdaten

Gemeinsame Verteilung von X und Y

Beispiele:

- $f(2,1) = P(\text{to}) + P(\text{of}) + P(\text{on}) = 0.18 + 0.10 + 0.06 = 0.34;$
- $f(3,0) = P(\text{BBC}) = 0.03;$
- $f(4,3) = 0.$

Gesamtverteilung:

		x			
		2	3	4	5
	0	0	0.03	0	0
y	1	0.34	0.30	0.16	0
	2	0	0	0.03	0.14

Marginalverteilungen

Wenn wir eine der zwei Dimensionen der gemeinsamen Verteilung projizieren, so erhalten wir die Marginalverteilung:

Definition: Marginalverteilung

Seien X und Y diskrete Zufallsvariablen und $f(x, y)$ der Wert der gemeinsamen Verteilung bei (x, y), dann sind die Funktionen

$$g(x) = \sum_y f(x, y) \quad \text{und} \quad h(y) = \sum_x f(x, y)$$

die Marginalverteilungen von X bzw Y.

- X: Wortlänge
- Y: Anzahl Vokale.

Gemeinsame Verteilung X and Y:

		x				
		2	3	4	5	$\sum_x f(x,y)$
	0	0	0.03	0	0	0.03
y	1	0.34	0.30	0.16	0	0.80
	2	0	0	0.03	0.14	0.17
$\sum_y f(x,y)$		0.34	0.33	0.19	0.14	

Marginalverteilung von Y. Marginalverteilung von X.

Beispiel: Korpusdaten

- X: Wortlänge
- Y: Anzahl Vokale.

Gemeinsame Verteilung X and Y:

		x				
		2	3	4	5	$\sum_x f(x,y)$
	0	0	0.03	0	0	**0.03**
y	1	0.34	0.30	0.16	0	**0.80**
	2	0	0	0.03	0.14	**0.17**
$\sum_y f(x,y)$		**0.34**	**0.33**	**0.19**	**0.14**	

Marginalverteilung von Y. **Marginalverteilung von X.**

Beispiel: Korpusdaten

- X: Wortlänge
- Y: Anzahl Vokale.

Gemeinsame Verteilung X and Y:

		x				
		2	3	4	5	$\sum_x f(x,y)$
	0	0	0.03	0	0	**0.03**
y	1	0.34	0.30	0.16	0	**0.80**
	2	0	0	0.03	0.14	**0.17**
$\sum_y f(x,y)$		**0.34**	**0.33**	**0.19**	**0.14**	

Marginalverteilung von Y. Marginalverteilung von X.

Bedingte Verteilungen

Bedingte Wahrscheinlichkeit:

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

Dies kann man auf Verteilungen verallgemeinern.

$$P(X = x | Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x, y)}{h(y)}$$

wobei $f(x, y)$ gemeinsame Verteilung von X and Y und $h(y)$ Marginalverteilung von y.

Bedingte Verteilungen

Definition

Sei $f(x, y)$ der Wert der gemeinsame Wahrscheinlichkeitsverteilung der diskreten Zufallsvariablen X und Y bei (x, y) und $h(y)$ der Wert der Marginalverteilung Y bei y, und $g(x)$ der Wert der Marginalverteilung X bei x, dann sind:

$$f(x|y) = \frac{f(x,y)}{h(y)} \quad \text{and} \quad w(y|x) = \frac{f(x,y)}{g(x)}$$

die bedingten Verteilungen von X gegeben $Y = y$, und von Y gegeben $X = x$ (für $h(y) \neq 0$ und $g(x) \neq 0$).

Beispiel: Korpusdaten

Basierend auf der gemeinsamen Verteilung $f(x,y)$ und den Marginalverteilungen $h(y)$ und $g(x)$ können wir die bedingten Verteilungen von X gegeben $Y = 1$ berechnen:

		x			
		2	3	4	5
y	1	$\frac{f(2,1)}{h(1)} = \frac{0.34}{0.80} = 0.43$	$\frac{f(3,1)}{h(1)} = \frac{0.30}{0.80} = 0.38$	$\frac{f(4,1)}{h(1)} = \frac{0.16}{0.80} = 0.20$	$\frac{f(5,1)}{h(1)} = \frac{0}{0.80} = 0$

Unabhängigkeit

Definition: Unabhängigkeit

Sei $f(x, y)$ der Wert der gemeinsamen Verteilung der diskreten Zufallsvariablen X und Y bei (x, y), und $g(x)$ und $h(y)$ die Werte der Marginalverteilungen von X bei x und Y bei y, dann sind X and Y unabhängig genau dann wenn:

$$f(x, y) = g(x)h(y)$$

für alle (x, y)

Beispiel: Korpusdatem

Marginalverteilungen:

		x				
		2	3	4	5	$h(y)$
	0	0	0.03	0	0	0.03
y	1	0.34	0.30	0.16	0	0.80
	2	0	0	0.03	0.14	0.17
$g(x)$		0.34	0.33	0.19	0.14	

Jetzt berechne $g(x)h(y)$ für jede Zelle:

		x			
		2	3	4	5
	0	0.01	0.01	0.01	0.00
y	1	0.27	0.26	0.15	0.12
	2	0.06	0.06	0.03	0.02

X und Y sind
nicht unabhängig.

Zusammenfassung

- Eine gemeinsame Wahrscheinlichkeitsverteilung gibt eine Wahrscheinlichkeitsverteilung für jedes Wertepaar von zwei Zufallsvariablen zurück.
- Marginalverteilungen projizieren eine Dimension der gemeinsamen Verteilung
- Bedingte Verteilung ist gemeinsame Verteilung geteilt durch Marginalverteilung
- Zwei Verteilungen sind unabhängig, wenn die gemeinsame Verteilung gleich dem Produkt der Marginalverteilungen ist.
- Aufgabenblatt 4.