Das Bayes-Theorem

Herleitung und Erläuterung für Schriftsachverständige

1 Axiomatische Definitionen

Wahrscheinlichkeit

Es sei S = $\{s_1;...;s_n\}$ die Ergebnismenge eines Zufallsexperiments, wobei i $\leq$ n natürliche Zahlen sind.

Wenn gilt:

$0 < p(s_i)$ ist reelle Zahl $\forall$ i und	Axiom der Nichtnegativität
$\sum_{i=1}^{n} p(s_i) = 1$	Axiom der Normiertheit

dann heißen die Zahlen $p(s_i)$ ***Wahrscheinlichkeiten*** [1]).

Ist H = $\{h_1;...;h_n\}$ mit $h_i \in S$ $\forall$ i ein Ereignis[2], dann ist

$P(H) = \sum_{i=1}^{n} p(h_i)$	Axiom der Additivität

die Wahrscheinlichkeit des Ereignisses H [3]).

Hypothese

Unter einer statistischen ***Hypothese*** versteht man eine Annahme (Vermutung) über die unbekannte ***Wahrscheinlichkeit eines Ereignisses*** H... [4]). (Demzufolge werden die Begriffe „Ereignis“ und „Hypothese“ synonym verwendet.)

Additionssatz

Für die Zwecke der forensischen Schriftvergleichung gibt es die Forderung, daß der Hypothesenraum einerseits vollständig erfaßt werden sollte. Andererseits sollen die Einzelhypothesen sich gegenseitig ausschließen.

Von insgesamt k möglichen paarweise sich ***gegenseitig ausschließenden Hypothesen*** des Hypothesen-Gesamtraumes H soll zunächst nur der Fall zweier Teilhypothesen H_1 und H_2 betrachtet werden. Die Hypothese H_1 sei eine Menge aus r Einzelergebnissen. Die Hypothese H_2 sei eine Menge aus n-r Einzelergebnissen, also:

$$H_1 = \{h_1;...;h_r\} \quad \text{bzw.} \quad H_2 = \{h_{r+1};...;h_n\}.$$

---

[1]) nach: Lambacher Schweizer, „Leistungskurs Stochastik“, S. 21
[2] Für Schriftanalysen: Ein „Ereignis“ ist eine Befundkonstellation.
[3]) ebd.
[4]) Duden, Rechnen und Mathematik, S. 278

Dann lauten nach (3) definitionsgemäß die Hypothesenwahrscheinlichkeiten:

$$P(H_1) = \sum_{i=1}^{r} p(h_i) \quad \text{und} \quad P(H_2) = \sum_{i=r+1}^{n} p(h_i),$$

Der aus den beiden Hypothesen sich zusammensetzende Hypothesenteilraum $H_1 \cup H_2$ besteht aus der Menge aller n zu den Einzelhypothesen H_1 und H_2 führenden Ergebnismenge

$$H_1 \cup H_2 = \{h_1;\ldots;h_r;h_{r+1};\ldots;h_n\}$$

Die Wahrscheinlichkeit für einen aus n Ergebnissen bestehenden Hypothesenteilraum H ist definitionsgemäß

$$P(H_1 \cup H_2) = \sum_{i=1}^{n} p(h_i)$$

Eine Zerlegung der Summe in Teilsummen führt wieder zurück zu

$$P(H_1 \cup H_2) = \sum_{i=1}^{r} p(h_i) + \sum_{i=r+1}^{n} p(h_i),$$

also insgesamt:

$$P(H_1 \cup H_2) = P(H_1) + P(H_2).$$

Es leuchtet ein, daß auf die gleiche Art und Weise jede disjunkte Zerlegung des vollständigen Hypothesenraumes in k einander ausschließende Einzelhypothesen zum Additionssatz führt:

$$P(H_1 \cup H_2 \cup \ldots \cup H_k) = P(H_1) + P(H_2) + \ldots + P(H_k)$$

Additionssatz für sich ausschließende Hypothesen

Multiplikationssatz

Der Multiplikationssatz für bedingte Wahrscheinlichkeiten zweier **voneinander abhängiger Ereignisse H_i und K** lautet[5]:

$$P(H_i \cap K) = P(H_i) * P(K \mid H_i)$$

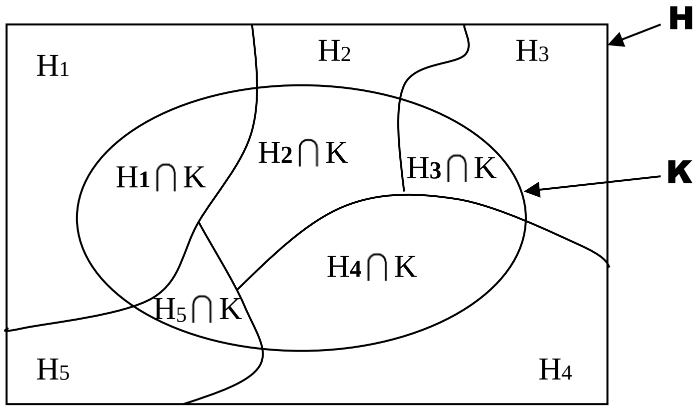
Multiplikationssatz für abhängige Ereignisse

(Die gewählten Abkürzungen sollen auf die avisierten Anwendungen durch „H_i“ für „Hypothese Nr. i“ und durch „K“ für „Konstellationen bestimmter Befunde“ hindeuten.)

---

[5]) Lambacher Schweizer, „Leistungskurs Stochastik“, S. 75

2 Herleitung des Theorems



Gegeben sei ein aus k disjunkten Hypothesen bestehender Hypothesenraum

$$H = \{H_1; \dots; H_k,\} = H_1 \cup H_2 \cup \dots \cup H_k.$$

K gebe die Menge aller möglichen Befundkonstellationen an und ist in obiger Skizze als Ellipse dargestellt. Unter bestimmten Hypothesen H_i ist eine gewisse Auswahl an Befunden möglich, andere Befunde nicht. Umgekehrt ist bei bestimmten Befunden nur eine eingegrenzte Auswahl an Hypothesen sinnvoll, der Rest nicht. Dieser Sachverhalt läßt sich gemäß Skizze durch Schnittmengen ($H_i \cap K$) darzustellen, die K restlos zusammensetzen:

$$K = (H_1 \cap K) \cup \dots \cup (H_k \cap K)$$

Der zugehörige Wahrscheinlichkeitsausdruck

$$P(K) = P((H_1 \cap K) \cup \dots \cup (H_k \cap K))$$

ist gemäß (4) äquivalent zu

$$P(K) = P(H_1 \cap K) + \dots + P(H_k \cap K)$$

Hierin ersetzt man jeden einzelnen Term ($H_i \cap K$) gemäß (5) und hat dann

$$P(K) = \sum_{i=1}^{n} P(H_i) * P(K \mid H_i)$$

Weiter läßt sich ($H_1 \cap K$) kommutieren und formal durch den entsprechenden Ausdruck von (5) ersetzen:

$$P(H_i \cap K) = P(K \cap H_i)$$

$$P(H_i) * P(K \mid H_i) = \qquad = P(K) * P(H_i \mid K)$$

Aus der letzten Zeile hat man:

$$P(H_i \mid K) = P(K \mid H_i) * \frac{P(H_i)}{P(K)} .$$

Indem man hierin P(K) gemäß (6) ersetzt, erhält man das Bayes-Theorem in der folgenden Schreibweise:

$$\boxed{P(H_i \mid K) = \frac{P(H_i) * P(K \mid H_i)}{\sum_{j=1}^{n} P(H_j) * P(K \mid H_j)}}$$

3 Erläuterungen

Im Zusammenhang mit schriftvergleichenden Schriftaltersbestimmungen können die oben verwendeten Bezeichnungen die folgenden Bedeutungen haben:

Bezeichnung	**Interpretation**	**Beispiel**
$\mathbf{H_i}$	$\mathbf{H_i}$ ist die i-te von insgesamt N möglichen sich einander gegenseitig ausschließenden Annahmen zur Schriftentstehung.	Bei der Schriftaltersbestimmung könnten einander sich ausschließende Annahmen etwa wie folgt lauten: $\mathbf{H_1}$: **Entstehung in 1998** $\mathbf{H_2}$: **Entstehung in 1999** ►$\mathbf{H_i}$: **Entstehung in 1997+i**◄ $\mathbf{H_N}$: **Entstehung in 2006** wobei der Laufindex i eine ganze Zahl zwischen 1 und N darstellt.
K	Die Befundkonstellation ist gegeben durch eine umgrenzbare Menge von möglichst wertstarken Merkmalen $\mathbf{K} = \{k_1, k_2, \ldots k_N\}$.	Günstig sind oft solche Merkmale der fraglichen Schrift, die binär quantifizierbar sind, z. B.: k_1: „b“ in Druckschrift/in Kurrentschrift k_2: „A“ im First arkadig/winklig … k_N: „t“ mit/ohne Anstrich Dabei ist es vorteilhaft, wenn die jeweilige Merkmalsausprägung k_j über die gesamte fragliche Schrift konstant erhalten bleibt.
$\mathbf{P(H_i)}$	Dieser Wert gibt eine Anfangs-wahrscheinlichkeit für das Zutreffen einer bestimmten Hypothese an.	In Ermangelung besseren Wissens geht man davon aus, daß jede der möglichen Hypothesen gegenüber allen anderen gleichberechtigt sei. Hat man insgesamt bspw. 9 Jahresintervalle, denen die Entstehung der fraglichen Schrift zugeordnet werden kann, so nimmt man zunächst an, daß prinzipiell jede Zuordnung möglich ist. Daher wird rechnerisch $P(H_1) = P(H_2) = \ldots = P(H_9) = 1/9$ gesetzt.

$P(K \mid H_i)$	*lies: „Befundwahrscheinlichkeit bei (der Hypothese) H_i“* Dieser Wert wird auch im deutschen Sprachgebrauch *Likelihood* genannt (Korrekt übersetzt: „Wahrscheinlichkeit“, zutreffender jedoch wäre jedoch „Ähnlichkeit“). Das Likelihood ist die frequentistische Wahrscheinlichkeit dafür, daß eine bestimmte Befundkonstellation **K** unter den Bedingungen der Hypothese **H_i** vorgefunden wurde.	Betrachtet man die in obigem Beispiel zur Hypothese H_3 gehörigen Vergleichsschriften des Jahres 2000, so mag man die Häufigkeiten der einzelnen Befunde sowohl in ihrer positiven (zur fraglichen übereinstimmenden) wie auch in ihrer negativen (unterschiedlichen) Ausprägung abzählen. Die Wahrscheinlichkeit dafür, daß *ein bestimmtes* Merkmal in den Vergleichsschriften des Jahres 2000 mit der fraglichen Ausprägung k_i übereinstimmend aufgefunden werden kann, ist gleich der relativen Häufigkeit j/(j + n), (wobei j: Anzahl der Übereinstimmungen n: Anzahl der Unterschiede im i-ten Merkmal) mit der es abgezählt worden ist. Die Wahrscheinlichkeit, genau die *Gesamtkonstellation* **K** aus den fraglichen Merkmalsausprägungen in den Vergleichsschriften vorzufinden, ist gleich dem *Produkt* aus den relativen Häufigkeiten der übereinstimmenden Einzelmerkmale.
Nenner	Der Nenner gewährleistet die Normierung. Er stellt sicher, daß das Ergebnis zwischen 0 und 1 liegt.	
$P(H_i \mid K)$	*lies: „Hypothesenwahrscheinlichkeit bei (Befundkonstellation) K“* Dies ist das erwünschte Ergebnis. Es gibt die Wahrscheinlichkeitszahl **P** dafür an, daß bei einer bestimmten Befundkonstellation **K** der fraglichen Schrift eine bestimmte Hypothese **H_i** erfüllt ist.	Der Zahlenwert liegt zwischen Null und Eins. Ist er bspw. 0,931, bedeutet dies, daß die Annahme, daß die fragliche Schrift, welche durch die Merkmalsmenge K beschrieben ist, in einem bestimmten Jahr geschrieben worden sei, mit 93,1 prozentiger Wahrscheinlichkeit zutrifft.