Bayessches Theorem

Das Bayessche Theorem ist ein Ergebnis aus der Wahrscheinlichkeitstheorie und liefert einen Zusammenhang zwischen bedingten Wahrscheinlichkeiten.

1. Bayessches Theorem für 2 Ereignisse

Seinen A und B zwei beliebige Ereignisse.
Die bedingte Wahrscheinlichkeit von A vorausgesetzt B oder auch die Wahrscheinlichkeit von A unter der Bedingung B ist definiert als:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad (1)$$

Die bedingte Wahrscheinlichkeit A vorausgesetzt B gibt sozusagen einen Grad an, wie abhängig das Eintreten des Ereignisses A vom Ereignis B ist.

Das diese Definition sinnvoll ist wird klar, wenn man für die Ereignisse A und B stochastische Unabhängigkeit voraussetzt.

$$P(A \cap B) = P(A) \times P(B) \qquad (2)$$

(2) in die Definition von (1) eingesetzt ergibt nun:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A) \times P(B)}{P(B)} = P(A) \qquad (3)$$

Da die Ereignisse A und B nicht voneinander abhängen (=stochastisch unabhängig) ergibt sich, dass auch die bedingte Wahrscheinlichkeit von A nicht von B abhängt und damit dieselbe Wahrscheinlichkeit wie von A ist.

Ebenso gilt nach (1) für die bedingte Wahrscheinlichkeit von B vorausgesetzt A:

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)} \qquad (4)$$

Daraus folgt für die bedingte Wahrscheinlichkeit von A vorausgesetzt B aus (1) das Bayessche Theorem:

$$P(A \mid B) = \frac{P(A) \times \dfrac{P(A \cap B)}{P(A)}}{P(B)} = \frac{P(A) \times P(B \mid A)}{P(B)} \qquad (5)$$

2. Verallgemeinerung des Bayesschen Theorems

Gegeben sein nun ein Stichprobenraum Ω, der aus disjunkten Ereignissen $A_1 \ldots A_k$ zusammengesetzt ist.
Disjunkte Ereignisse haben die Eigenschaft $A_i \cap A_j = \{\ \} \quad \forall i, j \quad i \neq j$
Das heißt die A_i (i=1…k) liefern eine Zerlegung des Stichprobenraums:

$$\Omega = \bigcup_{i=1\ldots k} A_i \qquad (6)$$

Des weitern sei nun B ein Ereignis aus Ω also $B \subseteq \Omega$ gegeben.
Dann gilt für B nach (4):

$$P(B \cap A_i) = P(B \mid A_i) \times P(A_i) \qquad (7)$$

Und da sich (wegen disjunkt) die Wahrscheinlichkeiten der A_i ausschließen tun das auch die Schnitte mit B:

$$P(B \cap A_i) \cap P(B \cap A_j) = 0 \quad \forall i, j \quad i \neq j \qquad (8)$$

Aus (8) und (7) folgt für die Wahrscheinlichkeit von B:

$$P(B) = P((B \cap A_1) \cup (B \cap A_2) \cup \ldots. \cup (B \cap A_k)) = P(B \cap A_1) + P(B \cap A_2) + \ldots. + P(B \cap A_k)$$
$$= \sum_{i=1}^{k} P(B \cap A\ ) = \sum_{i=1}^{k} P(B \mid A_i) \times P(A_i) \qquad (9)$$

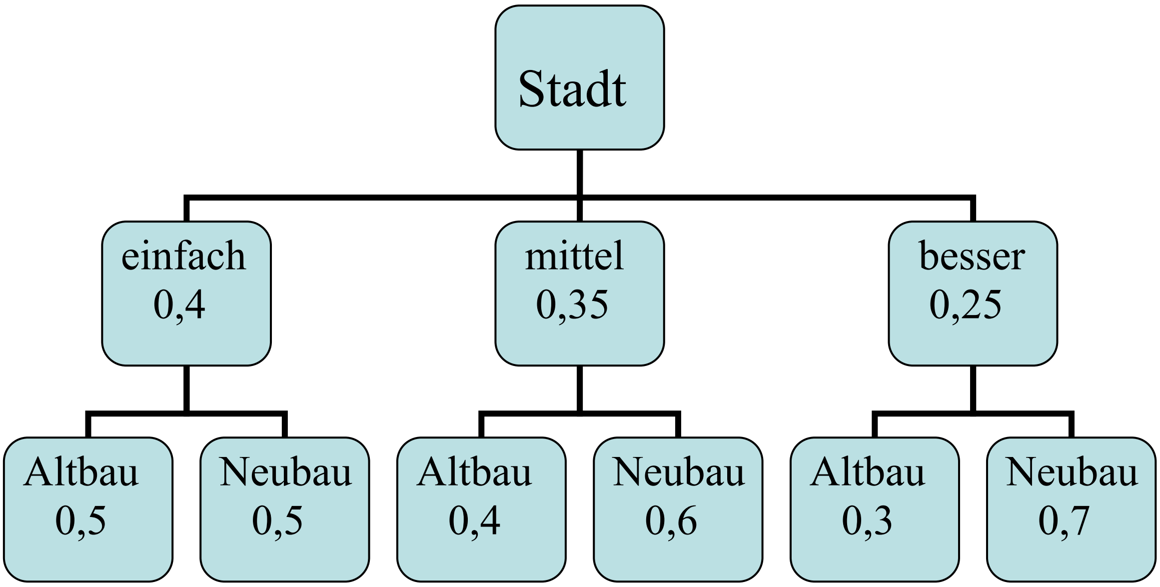
Die Wahrscheinlichkeit von B eingesetzt in (5) ergibt:

$$P(B \mid A_i) = \frac{P(B) \times P(A_i \mid B)}{\sum_{j=1}^{k} P(B \mid A_j) \times P(A_j)} \qquad (10)$$ Bayessches Theorem verallgemeinert

3.Beispiele

3.1 Beispiel zu (9)

In einer Stadt gibt es drei verschiedene Wohngegenden, einfach, mittel, besser.
Darüber hinaus kann man jedes Haus noch unterscheiden in Neubau oder Altbau.
Gesucht ist die Wahrscheinlichkeit aus allen Wohnungen einen Altbau zu „ziehen“.



Die Wahrscheinlichkeiten kann man dem obigen Baumdiagramm entnehmen.

Die Wahrscheinlichkeiten der Wohngegenden:

$$P(W_e) = 0{,}4$$
$$P(W_m) = 0{,}35$$
$$P(W_b) = 0{,}25$$

Die bedingte Wahrscheinlichkeiten dass die Wohnung ein Altbau ist unter der Bedingung der Wohngegenden ist gegeben durch:

$$P(Alt \mid W_e) = 0{,}5$$
$$P(Alt \mid W_m) = 0{,}4$$
$$P(Alt \mid W_b) = 0{,}3$$

Die gesuchte Wahrscheinlichkeit ergibt sich mit der Formel aus (9).

$$P(Alt) = \sum_{i=e,m,b} P(Alt \mid W_i) \times P(W_i) = P(Alt \mid W_e) \times P(W_e) + P(Alt \mid W_m) \times P(W_m) + P(Alt \mid W_b) \times P(W_b)$$

$$= 0{,}5 \times 0{,}4 + 0{,}4 \times 0{,}35 + 0{,}3 \times 0{,}25 = 0{,}415$$

3.2 Das Ziegenproblem

Ein Kandidat ist bei einer Quizshow und darf sich zwischen drei gleichwahrscheinlichen Toren entscheiden. Hinter einem ist der Hauptgewinn zwischen den zwei übrigen ist jeweils eine Ziege.
Der Kandidat entscheidet sich nun für eine Tor und der Moderat deckt darauf eines der übrigen zwei Toren auf hinter dem sich eine Ziege befindet und gibt nun dem Kandidaten noch einmal die Wahl ob er nicht doch lieber das andere der zwei verbleibenden Tore wählen will.

$G_i =$ das Ereignis der Gewinn befindet sich hinter Tor i, (i=1,2,3)
$M_j =$ das Ereignis der Moderator öffnet Tor j, (j=1,2,3)

In diesem Beispiel wählt nun der Kandidat Tor 1 und der Moderator deckt Tor 3 auf.
(das Ergebnis ist für jeden Fall gleich)

Die Frage ist nun lohnt es sich für den Kandidaten von Tor 1 zu Tor 2 zu wechseln.
Mathematischer ausgedrückt:
Es wird die bedingte Wahrscheinlichkeit gesucht, dass der Gewinn hinter Tor 2 ist unter der Bedingung der Moderator hat Tor 3 geöffnet: $P(G_2 \mid M_3)$
Diese ist durch (10) leicht auszurechnen.

$$P(G_2 \mid M_3) = \frac{P(M_3 \mid G_2) \times P(G_2)}{P(M_3)} = \frac{P(M_3 \mid G_2) \times P(G_2)}{\sum_{j=1}^{3} P(M_3 \mid G_j) P(G_j)}$$

$$= \frac{P(M_3 \mid G_2) \times P(G_2)}{P(M_3 \mid G_1) \times P(G_1) + P(M_3 \mid G_2) \times P(G_2) + P(M_3 \mid G_3) \times P(G_3)} \qquad (11)$$

Da alle Tore gleich wahrscheinlich sind folgt $P(G_1) = P(G_2) = P(G_3) = \frac{1}{3}$. (12)

Die bedingten Wahrscheinlichkeiten sind auch durch einfaches Überlegen zu bestimmen.

Die Wahrscheinlichkeit, dass der Moderator Tor 3 öffnet unter der Bedingung, dass der Gewinn hinter Tor 1 ist, ist 0,5. Da er sowohl Tor 2 als auch Tor 3 aufmachen könnte, hinter der jeweils eine Ziege steht.

$$P(M_3 \mid G_1) = \frac{1}{2} \qquad (13)$$

Die Wahrscheinlichkeit, dass der Moderator Tor 3 öffnet unter der Bedingung, dass der Gewinn hinter Tor 2 ist, ist 1. Er hat keine andere Wahl also Tor 3. Hinter Tor 2 ist der Gewinn, den er nicht aufdecken darf und Tor 1 hat bereits der Kandidat ausgewählt. Ergo bleibt nur Tor 3.

$$P(M_3 \mid G_2) = 1 \qquad (14)$$

Die Wahrscheinlichkeit, dass der Moderator Tor 3 öffnet unter der Bedingung, dass der Gewinn hinter Tor 3 ist, ist 0. Laut Spielregeln deckt der Moderator ein Tor auf, hinter der sich eine Zeige befindet. D. h. es ist ihm nicht erlaubt Tor 3 zu wählen unter der Bedingung der Gewinn liegt hinter Tor 3.

$$P(M_3 \mid G_3) = 0 \quad (15)$$

Die Ergebnisse (12), (13), (14) und (15) eingesetzt in (11) liefert:

$$P(G_2 \mid M_3) = \frac{1 \times \frac{1}{3}}{\frac{1}{2} \times \frac{1}{3} + 1 \times \frac{1}{3} + 0 \times \frac{1}{3}} = \frac{2}{3}$$

Das heißt wenn der Moderator ein Tor geöffnet hat, lohnt es sich für den Kandidaten immer zu wechseln, da sich seine Chancen zu gewinnen dadurch verdoppeln.
(wenn er sein Tor behält hat er ein drittel Chance zu gewinnen)

4.Bayessches Theorem zum Testen von Theorien

Das Bayessche Theorem kann man auch verwenden um die Wahrscheinlichkeit zu testen wie „gut“ eine Theorie ist. Gemessen an den Messergebnissen und der subjektiven Einschätzung wie „richtig“ das Ergebnis ist.

Die so genannte A-posteriori-Wahrscheinlichkeit wird mit $P(\Theta = \Theta_0 \mid x)$ bezeichnet und kann mit Hilfe des Bayestheorems berechnet werden. Im Spezialfall einer diskreten A-priori-Verteilung $P(\Theta = \Theta^*)$ mit $\Theta^* \in \Theta$ erhält man:

$$P(\Theta = \Theta_0 \mid x) = \frac{f(x \mid \Theta_0) \times P(\Theta = \Theta_0)}{\sum_{\Theta^* \in \Theta} f(x \mid \Theta^*) \times P(\Theta = \Theta^*)} \quad (16)$$

Θ sein hierbei ein unbekannter Umweltzustand (z.B. Hubbelkonstante), der auf der Basis von Beobachtungen *x* einer Zufallsgröße X geschätzt werden soll.

Die bedingte Verteilung von X unter der Bedingung dass Θ den Wert Θ_0 annimmt, wird mit $f(x \mid \Theta_0)$ bezeichnet (Likelihood). Diese Wahrscheinlichkeitsverteilung kann nach Beobachtung der Stichprobe bestimmt werden. Dieser Wert ist sozusagen aus der Messungen abzulesen.

Was man nun leider nicht kennt, ist die A-priori-Verteilung des Umweltzustandes Θ. Man nimmt nun an, dass diese Unwissenheit sein persönliches Problem ist, und wird daher subjektive Wahrscheinlichkeiten dafür angeben, inwieweit man das Ergebnis für richtig oder falsch hält. Damit hat man eine bessere Vorstellung davon, inwieweit man das Ergebnis für richtig oder falsch hält.

Mit Hilfe des Bayestheorems kann nun die A-posteriori-Verteilung des Umweltzustands Θ bestimmt werden. Falls die Menge aller möglichen Umweltzustände endlich ist, lässt sich die A-posteriori-Verteilung im Wert Θ_0 als die Wahrscheinlichkeit interpretieren, mit der man nach Beobachtung der Stichprobe und unter Einbeziehung des Vorwissens bzw. der subjektiven Schätzung den Umweltzustand Θ_0 erwartet.

Quellen:

Wikipedia: http://de.wikipedia.org/wiki/Bayestheorem
http://de.wikipedia.org/wiki/Wahrscheinlichkeitstheorie#Bayestheorem
http://de.wikipedia.org/wiki/Ziegenproblem

http://web.neuestatistik.de/demo/Demo_DE/MOD_97823/html/comp_98229.html
http://www.raunvis.hi.is/~sksi/v_referate/franz_skulise1.html

gehalten am Mittwoch den 30. November 2011 von Korbinian Groh im Fach Kosmologie