

Zur Abhängigkeit von stochastischen Ereignissen

Einleitung und Motivation

In Döller und Götz (2024) ist das sogenannte Assoziationsmaß als Maß für die Abhängigkeit von zwei nominalskalierten Merkmalen mit je zwei Ausprägungen vorgestellt worden. Legt man die Daten einer Vierfeldertafel zugrunde und konstruiert daraus ein Einheitsquadrat, dann ist das Assoziationsmaß am Höhenunterschied der unteren (oder oberen) beiden Rechtecke visuell ablesbar und kann über deren Seitenlängen auch berechnet werden: $\mathcal{A}_E = h(B_1|A_1) - h(B_1|A_2)$, A_1 und A_2 bzw. B_1 und B_2 sind dabei die möglichen Merkmalsausprägungen der Merkmale A und B , wobei A das dominierende Merkmal ist (Eichler & Vogel, 2013, S. 84). Mit h werden die relativen Häufigkeiten bezeichnet. Betrachtet man das transponierte Einheitsquadrat, so ist $\mathcal{A}_{\hat{E}} = h(A_1|B_1) - h(A_1|B_2)$ das entsprechende Assoziationsmaß, das im Allgemeinen nicht gleich dem ursprünglichen Assoziationsmaß \mathcal{A}_E ist (Döller & Götz, 2024). Ein Wechsel des dominierenden Merkmals ändert also den auf diese Weise quantifizierten Grad der Abhängigkeit. Unabhängige Merkmale bleiben allerdings unabhängig, es gilt nämlich $\mathcal{A}_E = 0 \Leftrightarrow \mathcal{A}_{\hat{E}} = 0$.

Ein elementares, *symmetrisches*, quantitatives Maß für die Abhängigkeit von zwei nominalskalierten Merkmalen wäre wünschenswert. Ersetzt man die relativen Häufigkeiten h durch Wahrscheinlichkeiten $P(\cdot)$, dann kann ein solches Maß auch für den Grad von Abhängigkeiten von zwei stochastischen Ereignissen verwendet werden.

Ein elementares, symmetrisches und normiertes Abhängigkeitsmaß

Für die folgende Herleitung werden stochastische Ereignisse A und B und ihre positiven Wahrscheinlichkeiten P betrachtet. Ausgehend vom Ausdruck $P(A|B) - P(A|B^c)$, der dem Assoziationsmaß \mathcal{A}_E von eben entspricht, wobei B^c das Komplement von B meint, gewinnen wir mit Hilfe der Formel von Bayes $P(A|B^c) = \frac{P(A) - P(A \cap B)}{1 - P(B)}$. Die Differenz $P(A|B) - P(A|B^c)$ ist dann gleich $\frac{P(A \cap B) - P(A) \cdot P(B)}{P(B) \cdot (1 - P(B))}$. Division durch $P(A) \cdot P(A^c)$ liefert dann den in A und B *symmetrischen* Quotienten $C := \frac{P(A \cap B) - P(A) \cdot P(B)}{P(A) \cdot P(A^c) \cdot P(B) \cdot P(B^c)}$ für $\frac{P(A|B) - P(A|B^c)}{P(A) \cdot P(A^c)}$ bzw. für $\frac{P(B|A) - P(B|A^c)}{P(B) \cdot P(B^c)}$. Er ist aber *nicht normiert*: Für $B \subseteq A$ ist der Quotient $C = \frac{1}{P(A) \cdot (1 - P(B))}$, für $P(A) \rightarrow 0$ divergiert also C .

In: L. Schick, M. Platz & A. Lambert (Hrsg.),
Beiträge zum Mathematikunterricht 2025.

58. Jahrestagung der Gesellschaft für Didaktik der Mathematik. WTM.

<https://doi.org/10.37626/GA9783959873307.0>

Der Pearson'sche Korrelationskoeffizient $\frac{E[(X-E(X)) \cdot (Y-E(Y))]}{\sqrt{E[(X-E(X))^2] \cdot E[(Y-E(Y))^2]}}$ für Zu-

fallsvariablen X und Y und ihre Erwartungswerte E als Maß für lineare Abhängigkeiten kann mittels der Indikatorfunktionen $X(\omega) = 1_A(\omega)$ und $Y(\omega) = 1_B(\omega)$ für Ereignisse A und B uminterpretiert werden und hilft so weiter. Wegen $E(X) = P(A)$, $E(X^2) = P(A)$ und $E(X \cdot Y) = P(A \cap B)$ ergibt sich für die Erwartungswerte

- $E[(X - E(X))^2] = P(A) \cdot (1 - P(A))$, dementsprechend
- $E[(Y - E(Y))^2] = P(B) \cdot (1 - P(B))$ und schließlich
- $E[(X - E(X)) \cdot (Y - E(Y))] = P(A \cap B) - P(A) \cdot P(B)$.

Damit ist der Pearson'sche Korrelationskoeffizient für Ereignisse A und B gleich

$$\frac{P(A \cap B) - P(A) \cdot P(B)}{\sqrt{P(A) \cdot (1 - P(A)) \cdot P(B) \cdot (1 - P(B))}} =: D.$$

Es gilt $-1 \leq D \leq 1$ (Stepniak, 2015) und D ist ebenfalls symmetrisch in A und B . Damit ist D ein *elementares, symmetrisches und normiertes* Abhängigkeitsmaß für die stochastischen Ereignisse A und B bzw. für zwei nominalskalierte Merkmale mit je zwei Ausprägungen.

Bemerkung: Die Gleichung $D^2 = C \cdot [P(A \cap B) - P(A) \cdot P(B)]$ drückt einen Zusammenhang zwischen C und D aus. Durch Nachrechnen erkennt man $P(A \cap B) - P(A) \cdot P(B) = P(A^c \cap B^c) - P(A^c) \cdot P(B^c)$, damit kann $D^2 = \frac{P(A \cap B) - P(A) \cdot P(B)}{P(A) \cdot P(B)} \cdot \frac{P(A^c \cap B^c) - P(A^c) \cdot P(B^c)}{P(A^c) \cdot P(B^c)}$ gefolgert werden (Braza, 2024).

Alle vier Eintragungen einer Vierfeldertafel (siehe Einleitung und Motivation) können also in D^2 gefunden werden.

Zur Normierung von D

Die Eigenschaft $-1 \leq D \leq 1$ kann auch direkt bewiesen werden. Dazu werden die Variablen $P(A) =: a$, $P(B) =: b$, $P(A \cap B) =: c$ eingeführt, um den Term für D formal einfacher darzustellen: $D = \frac{c-ab}{\sqrt{ab(1-a)(1-b)}}$. Zu beachten

ist, dass D nur von c abhängt, da a und b in der Regel gegeben sind und daher feststehen. Ohne Beschränkung der Allgemeinheit sei $b \leq a$. Weil weiterhin $c \leq b$ gilt, erhalten wir die Ungleichungskette $0 \leq c \leq b \leq a \leq 1$. Das Abhängigkeitsmaß D wird minimal, wenn $c = 0$ der Fall ist. Es ergibt sich dann

$$D^2 = \frac{a^2 b^2}{ab(1-a)(1-b)} = \frac{ab}{(1-a)(1-b)} = \frac{1-(a+b)+ab-[1-(a+b)]}{1-(a+b)+ab} = 1 - \frac{1-(a+b)}{1-(a+b)+ab}.$$

Wegen $c = P(A \cap B) = 0$ ist $0 \leq P(A \cup B) = P(A) + P(B) = a + b \leq 1$ in diesem Fall. Damit folgern wir $0 \leq \frac{1-(a+b)}{1-(a+b)+ab} \leq 1$ und daraus $|D| \leq 1$.

Das Abhängigkeitsmaß D wird maximal, wenn $c = b$ gilt, also $B \subseteq A$. Dann ist $D^2 = \frac{(b-ab)^2}{ab(1-a)(1-b)} = \frac{b^2 \cdot (1-a)^2}{ab(1-a)(1-b)} = \frac{b}{1-b} \cdot \frac{1-a}{a}$. Die zweistellige Funktion $D^2 = D^2(a, b)$ ist auf $0 \leq b \leq a \leq 1$ definiert, also auf einer „Dreiecksregion“ R . Mit GeoGebra 3D kann der Graph von D^2 geplottet werden: mit Abbildung 1 wird deutlich, dass D^2 auf R für $a = b$ maximal wird und dort den Wert 1 annimmt. Damit gilt auch in diesem Fall $|D| \leq 1$.

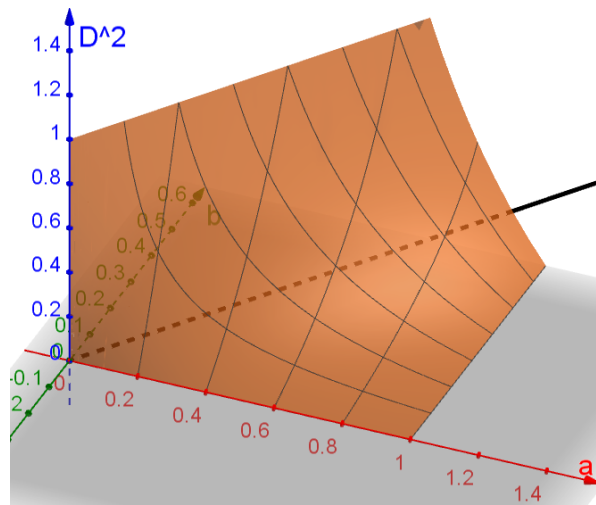


Abb. 1: Graph von D^2 auf R

Beispiele

Für $A = B$ ist $D = 1$, die Abhängigkeit von A und B ist maximal und positiv. Für $B = A^c$ ist $D = -1$, die Abhängigkeit von A und B ist wieder maximal, aber negativ. Beide Resultate sind nachvollziehbar, für $B = A^c$ überdecken A und B den gesamten Ereignisraum, haben aber kein Element gemeinsam.

Für den zweimaligen Wurf einer fairen Münze betrachten wir die beiden Ereignisse A „Es wird mindestens einmal Wappen (W) geworfen“ und B „Es wird zweimal Wappen geworfen“. Damit ist $A = \{WW, WZ, ZW\}$ als Menge geschrieben und $B = \{WW\}$, dabei steht „ Z “ für „Zahl“. (Der erste Eintrag benennt das Ergebnis des ersten Wurfs, der zweite das des zweiten.) Mit $P(A) = \frac{3}{4}$, $P(B) = \frac{1}{4}$ und $P(A \cap B) = \frac{1}{4} = P(B)$ ergibt sich $D = \frac{1}{3}$. Das Ergebnis gewinnt erst an Bedeutung, wenn man es mit einem anderen vergleicht. Wird A zu A_1 „Beim ersten Wurf erscheint Wappen“ abgeändert und B bleibt unverändert, dann ist $D = \frac{\sqrt{3}}{3} \approx 0,577$ wegen $A_1 = \{WW, WZ\}$, $P(A_1) = \frac{1}{2}$ und $P(A_1 \cap B) = \frac{1}{4}$. Die Abhängigkeit ist größer geworden, denn

das Ereignis A_1 unterscheidet sich vom Ereignis B um „weniger“ als das Ereignis A vom Ereignis B (Braza, 2024).

Für das klassische Urnenbeispiel betrachten wir eine Urne mit r roten und g grünen Kugeln, aus der wir zweimal *ohne* Zurücklegen ziehen. Das Ereignis A trete ein, wenn beim ersten Zug eine rote Kugel gezogen wird, und B , wenn beim zweiten Zug eine grüne Kugel gezogen wird. Mit $n = r + g$ für die Gesamtzahl der Kugeln erhalten wir $P(A) = \frac{r}{n}$ und mit der Formel von Bayes ist $P(B) = P(B|A) \cdot P(A) + P(B|A^c) \cdot P(A^c) = \frac{g}{n-1} \cdot \frac{r}{n} + \frac{g-1}{n-1} \cdot \frac{g}{n} = \frac{g}{n}$. Für den Durchschnitt von A und B ergibt sich die Wahrscheinlichkeit $\frac{r}{n} \cdot \frac{g}{n-1}$. Damit ist $D = \frac{1}{n-1}$. Je größer die Anzahl n der Kugeln ist, desto geringer ist die Abhängigkeit der Ereignisse A und B voneinander (Braza, 2024).

Eine Familie hat $n \geq 2$ Kinder. Das Ereignis A trete ein, wenn es höchstens ein Mädchen gibt, das Ereignis B , wenn es in der Familie Kinder beiderlei Geschlechts gibt. Es gilt $P(A) = \left(\frac{1}{2}\right)^n \cdot (n+1)$, $P(B) = 1 - \frac{1}{2^{n-1}}$ und $P(A \cap B) = \left(\frac{1}{2}\right)^n \cdot n$. Der Zähler von D ist $\left(\frac{1}{2}\right)^n \cdot \left[\left(\frac{1}{2}\right)^{n-1} \cdot (n+1) - 1\right]$ und das Quadrat des Nenners $(n+1) \cdot (2^{n-1} - 1) \cdot (2^n - n - 1) \cdot \frac{1}{4^{2n-1}}$ (mittels CAS). Eine Auswertung mit Excel ergibt für $n = 2$ den positiven Wert $D = \frac{\sqrt{3}}{3} \approx 0,577$, für $n = 3$ ist $D = 0$, und für $n > 3$ ist $D < 0$ (mittels vollständiger Induktion) und D konvergiert für $n \rightarrow \infty$ gegen 0 (Betrachten der Zweierpotenzen von D). Das heißt, wenn die Familie drei Kinder hat, sind die Ereignisse A und B unabhängig voneinander, eine positive Abhängigkeit gibt es nur bei zwei Kindern, in allen anderen Fällen ist die Abhängigkeit negativ und wird betragsmäßig für größere Familien kleiner.

Literatur

- Braza, P. (2024). Dependence of events, revisited. *The Mathematical Gazette*, 108(572), 257–261. <https://doi.org/10.1017/mag.2024.64>
- Döller, V. & Götz, S. (2024). Bedingte relative Häufigkeiten in Einheitsquadraten. P. Ebers, F. Rösken, B. Barzel, A. Büchter, F. Schacht & P. Scherer (Hrsg.), *Beiträge zum Mathematikunterricht 2024. 57. Jahrestagung der Gesellschaft für Didaktik der Mathematik* (S. 721–724). WTM. <http://dx.doi.org/10.17877/DE290R-24729>
- Eichler, A. & Vogel, M. (2013). *Leitidee Daten und Zufall: Von konkreten Beispielen zur Didaktik der Stochastik* (2., aktualisierte Auflage). Springer. <https://doi.org/10.1007/978-3-658-00118-6>
- Stepniak, C. (2015). A note on correlation coefficient between random events. *Discussiones Mathematicae, Probability and Statistics*, 35, 57–60. <http://dx.doi.org/10.7151/dmps.1177>