Katharina BÖCHERER-LINDER, Pädagogische Hochschule Freiburg, Andreas EICHLER, Universität Kassel, Markus VOGEL, Pädagogische Hochschule Heidelberg, DE

Der Einfluss von Frageformat und Visualisierung beim Satz von Bayes

1. Einleitung

Die Motivation, sich im Rahmen einer mathematikdidaktischen Forschungsarbeit mit dem Satz von Bayes zu beschäftigen, ist vielschichtig: Zum einen hat die Bestimmung von bedingten Wahrscheinlichkeiten mit Hilfe des Satzes von Bayes eine hohe Anwendungsrelevanz in vielfältigen Situationen des Alltags, sei es im medizinischen, juristischen oder wirtschaftlichen Bereich. Zum anderen hat die kognitionspsychologische Forschung festgestellt, dass Menschen allgemein, aber auch Fachleute wie Ärzte oder Juristen, erhebliche Schwierigkeiten haben, bedingte Wahrscheinlichkeiten auf der Grundlage des Satzes von Bayes richtig zu bestimmen (Eddy, 1982; Gigerenzer & Hoffrage, 1995).

Folgende Aufgabe aus dem medizinischen Bereich soll die Problemstellung beim Satz von Bayes erläutern:

Eine Frau lässt sich auf eine bestimmte Krankheit testen und erhält ein positives Testergebnis. Sie ist erschrocken und möchte wissen, mit welcher Wahrscheinlichkeit sie tatsächlich krank ist.

Der beratenden Ärztin stehen folgende Informationen zur Verfügung:

10% aller Frauen Anfang Vierzig haben die Krankheit (Basisrate). 80% der kranken Frauen reagieren positiv auf den Test (Sensitivität). 10% der gesunden Frauen reagieren ebenfalls positiv (Spezifität).

$$P(krank|positiv) = \frac{80\% \cdot 10\%}{80\% \cdot 10\% + 10\% \cdot 90\%} = 47\%$$

Abb.1: Information im Wahrscheinlichkeitsformat und Satz von Bayes

In Studien von Eddy (1982), Gigerenzer und Hoffrage (1995) und Binder et al. (2015) wurde bei ähnlichen Aufgaben, bei denen die bedingte Wahrscheinlichkeit im Rückschluss mit Hilfe des Satzes von Bayes bestimmt werden sollte, im Durchschnitt lediglich eine Lösungshäufigkeit von 10 bis 20% erreicht. Dies führt zu der Frage, wie die Anwendung des Satzes von Bayes erleichtert werden kann. Im Folgenden werden drei Methoden vorgestellt, wie Menschen durch eine geeignete Darstellung des Problems bei der Berechnung des Satzes von Bayes unterstützt werden können, nämlich erstens durch die Verwendung von natürlichen Häufigkeiten, zweitens durch den Einsatz von Visualisierung und drittens durch die Variation des Frageformates.

2. Methoden der Unterstützung beim Satz von Bayes

Cosmides und Tooby (1996) sowie Gigerenzer und Hoffrage (1995) schlugen vor, die statistische Information bei Bayesianischen Aufgaben nicht wie sonst üblich mit Hilfe von Wahrscheinlichkeiten auszudrücken, sondern mit Hilfe von natürlichen Häufigkeiten. Hierdurch erleichtert sich die Berechnung des Satzes von Bayes erheblich (vgl. Johnson & Tubau, 2015). Im eingangs beschriebenen Beispiel der medizinischen Diagnose kann nämlich die bedingte Wahrscheinlichkeit im Rückschluss dadurch bestimmt werden, dass der Anteil der Kranken unter den positiv Getesteten als Verhältnis absoluter Häufigkeiten berechnet wird:

100 von 1000 Frauen Anfang Vierzig haben die Krankheit. 80 von 100 kranken Frauen reagieren auf den Test positiv. 90 von 900 gesunden Frauen reagieren ebenfalls positiv.

$$P(krank|positiv) = \frac{80}{80+90} = 47\%$$

Abb. 2: Information und Berechnung im Häufigkeitsformat

Mehrere Studien belegen, dass durch eine Darstellung des Problems im Format der natürlichen Häufigkeiten die durchschnittliche Lösungshäufigkeit von den genannten 10 bis 20 % auf ca. 50% gesteigert werden kann (z.B. Gigerenzer & Hoffrage, 1995; Binder et al., 2015).

Eine weitere Möglichkeit, die Lösungshäufigkeit noch über den Effekt der natürlichen Häufigkeiten hinaus zu verbessern, ist die Visualisierung der statistischen Information. Hierbei ist es allerdings noch weitgehend ungeklärt, welche Eigenschaften einer Visualisierung für einen günstigen Effekt entscheidend sind (z.B. Böcherer-Linder & Eichler, 2017). In unserem Forschungsprojekt, das ein Teil des Forschungs- und Nachwuchskollegs VisDeM an der PH Freiburg ist, vergleichen wir die Effekte des Baumdiagrammes und des Einheitsquadrates. Da empirische Ergebnisse darauf hindeuten, dass besonders solche Visualisierungen günstig sind, die die statistische Information in Form von natürlichen Häufigkeiten enthalten (z. B. Binder et al., 2015), realisieren wir für unsere Studien das Baumdiagramm und das Einheitsquadrat mit natürlichen Häufigkeiten. In einer Studie mit 143 Studierenden war das Einheitsquadrat günstiger als das Baumdiagramm für die Berechnung des Satzes von Bayes (Böcherer-Linder & Eichler, 2017).

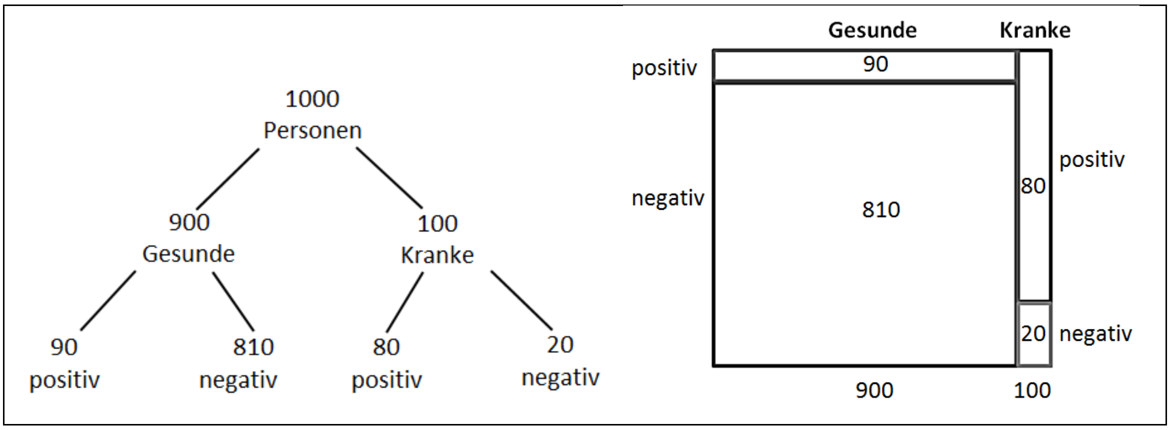

Abb. 3: Visualisierung der statistischen Information mit dem Baumdiagramm und dem Einheitsquadrat

Eine dritte, bisher noch wenig erforschte Möglichkeit der Unterstützung ist die Variation des Frageformates. Vertreter des sogenannten *nested-sets account* gehen davon aus, dass eine wesentliche Schwierigkeit bei der Anwendung des Satzes von Bayes das Erkennen der relevanten Teilmengenstruktur ist (z. B. Sloman et al., 2003). Ein Frageformat, das deutlich macht, welches Teilmengenverhältnis berechnet werden muss, sollte demnach günstig sein.

3. Ableitung der Forschungshypothese

Wird der Satz von Bayes mit Hilfe der natürlichen Häufigkeiten angewandt, so muss der Anteil der Kranken unter den positiv Getesteten bestimmt werden. Bei der Auswertung früherer Studien (Böcherer-Linder & Eichler, 2017) beobachteten wir häufig den Fehler, dass die Menge der Kranken mit der Menge aller positiv Getesteten ins Verhältnis gesetzt wurde. Hierbei wurde die für die Berechnung des Satzes von Bayes nötige Schnittmengenbildung (siehe Abb. 4) vernachlässigt.

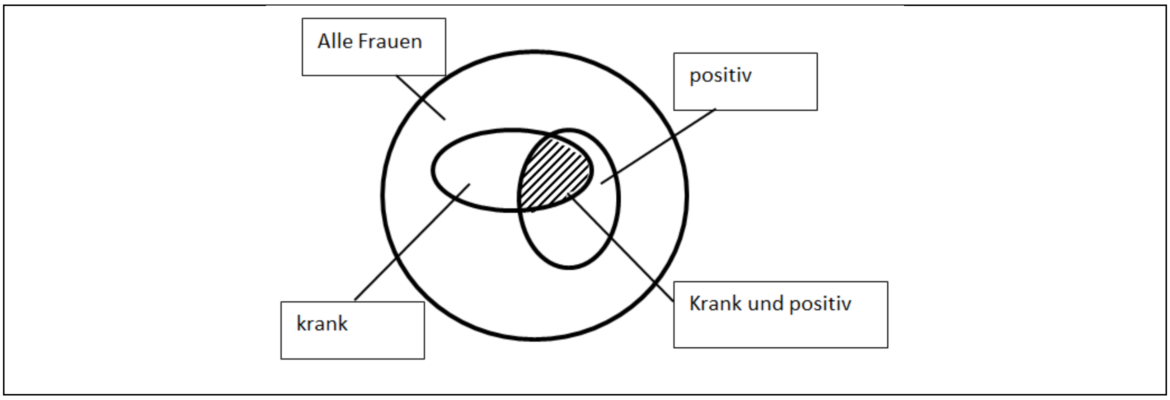

Abb. 4: Schnittmengenbildung beim Satz von Bayes

Dieser Fehler ist in der psychologischen Forschung unter dem Namen „Pre-Bayes“ bekannt (Zhu & Gigerenzer, 2006). Da eine wesentliche Bedingung für die Erleichterung der Anwendung des Satzes von Bayes die Transparenz von Teilmengenbeziehungen ist (Sloman et al, 2003), entwickelten wir ein Frageformat, dass die Schnittmengenbildung explizit macht. Wir unterscheiden zwischen explizitem und implizitem Frageformat:

Explizites Frageformat: Berechne den Anteil der kranken und positiv getesteten Frauen unter den positiv getesteten Frauen.

Implizites Frageformat: Berechne den Anteil der kranken unter den positiv getesteten Frauen.

Abb. 5: Explizites und implizites Frageformat

Entsprechend dem *nested-sets account* ist unsere Hypothese, dass ein explizites Frageformat günstiger als ein implizites Frageformat ist, sowohl für eine Visualisierung mit dem Baumdiagramm als auch für eine Visualisierung mit dem Einheitsquadrat.

Unsere Untersuchungen ergaben, dass im impliziten Frageformat das Einheitsquadrat signifikant besser geeignet ist als das Baumdiagramm (vgl. auch Böcherer-Linder & Eichler, 2017) und dass das explizite Frageformat signifikant dem impliziten überlegen ist.

Literatur

Binder, K., Krauss, S., & Bruckmaier, G. (2015). Effects of visualizing statistical information - an empirical study on tree diagrams and 2 × 2 tables. *Frontiers in psychology 6*, 1186. DOI: 10.3389/fpsyg.2015.01186.

Böcherer-Linder, K., & Eichler, A. (2017). The impact of visualizing nested sets. An empirical study on tree diagrams and unit squares. *Frontiers in psychology 7*, 2026 DOI: 10.3389/fpsyg.2016.02026.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition 58 (1)*, S. 1–73. DOI: 10.1016/0010-0277(95)00664-8.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: problems and opportunities. In: D. Kahneman, P. Slovic, & A. Tversky (Hrsg.), *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press, S. 249–267.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction. Frequency formats. *Psychological Review 102 (4)*, S. 684–704. DOI: 10.1037/0033-295X.102.4.684.

Johnson, E. D., & Tubau, E. (2015). Comprehension and computation in Bayesian problem solving. *Frontiers in psychology 6*, 938. DOI: 10.3389/fpsyg.2015.00938.

Sloman, S. A., Over, D., Slovak, L., & Stibel, J. M. (2003). Frequency illusions and other fallacies. *Organizational Behavior and Human Decision Processes 91 (2)*, S. 296–309. DOI: 10.1016/S0749-5978(03)00021-9.

Zhu, L., & Gigerenzer, G. (2006). Children can solve Bayesian problems: the role of representation in mental computation. *Cognition 98 (3)*, S. 287–308. DOI: 10.1016/j.cognition.2004.12.003.