

Prävalenzfehler

Als **Prävalenzfehler** bezeichnet man den Fehler, der entsteht, wenn die Bestimmung der bedingten Wahrscheinlichkeit einer statistischen Variable *A* unter einer Bedingung *B* ohne Rücksicht auf die Prävalenz oder A-priori-Wahrscheinlichkeit von *A* vorgenommen wird. Die Prävalenz bezeichnet die Verteilung von *A* über die in Frage stehende Grundgesamtheit und wird auch als *Basisrate* bezeichnet. Der *Prävalenzfehler* wird daher auch als *Basisratenfehler*, als *Basisratenmissachtung* oder als *Base Rate Fallacy* bezeichnet. Für *A* und *B* kommen Ereignisse, aber auch Eigenschaften in Frage, der Fehler ist ein allgemeines Phänomen bei der Interpretation von statistischer Korrelation.

Inhaltsverzeichnis

Rechenbeispiel

Annahmen

Folgen

Auswertung

Psychologische Ergebnisse

Siehe auch

Literatur und Weblinks

Einzelnachweise

Rechenbeispiel

In Anlehnung an Lindsey/Hertwig/Gigerenzer.^[1]

Die Vermutung, dass durch einen DNA-Test zweifelsfrei eine Person anhand von Spuren identifiziert werden kann, beruht auf zwei Prävalenzfehlern.

Annahmen

Bei einem Verbrechen wird eine DNA-Spur am Opfer gefunden. Es liegen nur wenige weitere Hinweise vor:

- 10 Millionen Personen kommen prinzipiell als Urheber der Spur in Frage. Die Wahrscheinlichkeit, dass eine zufällig aus dieser Gesamtheit herausgegriffene Person der Urheber ist, ist $P(U) = 1/10^7$.
- Bei einer natürlichen Zufallsstreuung des DNA-Profiles weisen ungefähr 10 der 10 Millionen in Frage kommenden Personen einen genetischen Fingerabdruck auf, der mit dem DNA-Profil der Spur am Opfer identisch ist, dieses bestimmte DNA-Profil hat also eine Prävalenz von $P(D) = 1/10^6$.
- Der verwendete DNA-Test soll keine falsch negativen Ergebnisse produzieren. Wenn eine Person das DNA-Profil der Spur aufweist, so stellt der Test auch eine Übereinstimmung fest.

Sei $P(A|B)$ die bedingte Wahrscheinlichkeit von A unter der Bedingung B , so ist $P(T|D) = 1$, das heißt, unter der Bedingung, dass das DNA-Profil vorliegt, ist der Test positiv.

4. Aber auch bei Personen, die dieses DNA-Profil nicht aufweisen, wird aufgrund einer kleinen, aber unvermeidlichen Testungenauigkeit von $P(T|\bar{D}) = 1/10^5$ (das ist eine Prävalenz von 1 in 100.000) trotzdem eine Übereinstimmung festgestellt.

Wie ist es zu bewerten, wenn im Rahmen einer DNA-Rasterfahndung eine zufällig aus den 10 Millionen herausgegriffene Person positiv getestet wird?

Folgen

- Der Test liefert, wenn alle 10 Millionen Personen getestet werden, in durchschnittlich 100 Fällen ein falsch positives Ergebnis. Dies ergibt sich aus der bedingten Wahrscheinlichkeit bzw. dem Multiplikationssatz der Wahrscheinlichkeitstheorie:

$$P(T \cap \bar{D}) = P(\bar{D}) \cdot P(T|\bar{D}) = (1 - P(D)) \cdot P(T|\bar{D}) = 0,00000999999$$

$$10\,000\,000 \cdot P(T \cap \bar{D}) \approx 100$$

- Von den 10 Millionen Personen würden bei einer DNA-Rasterfahndung insgesamt 110 positiv getestet. Die Prävalenz für den positiven Test $P(T) = 110/10^7 = 11 \cdot 1/10^6$ ist somit größer als die Prävalenz des DNA-Profiles $P(D) = 1/10^6$.
- Aber das Ergebnis wäre nur in rund einem Elftel der Fälle korrekt. Obwohl die Wahrscheinlichkeit, dass der genetische Fingerabdruck einen positiven Test zur Folge hat, $P(T|D)$, sogar 1 beträgt, ist die Wahrscheinlichkeit, dass hinter einem positiven Test auch das DNA-Profil steht ($P(D|T)$) nach dem Satz von Bayes und den berechneten relativen Häufigkeiten $P(D|T) = \frac{P(T|D) \cdot P(D)}{P(T)} = 1/11$. Wird also angenommen, dass ein positiver Test an einer zufällig ausgewählten Person eine gute Vorhersage für eine Übereinstimmung mit dem DNA-Profil der Spur ist, wird ein erster Prävalenzfehler begangen, bei dem $P(T|D)$ und $P(D|T)$ verwechselt wurden.
- Der Urheber der Spur ist einer der 10 Träger des DNA-Profiles. Es gilt also $P(D|U) = 1$, aber zugleich $P(U|D) = 1/10$. Obwohl der Urheber der Spur mit Sicherheit das DNA-Profil aufweist und er also mit Sicherheit positiv getestet würde, ist unter den gegebenen Voraussetzungen die Wahrscheinlichkeit, dass ein beliebiger positiver Testfall den Urheber bezeichnet, recht klein: Von 110 positiven Testfällen haben nur 10 das DNA-Profil $P(D|T) = 1/11$, und von diesen ist nur einer der Urheber:

$$P(U|T) = P(D|T) \cdot P(U|D) = 1/110 = 0,0090 = 0,90 \%$$

- Im Vergleich dazu ist die Wahrscheinlichkeit, dass eine zufällig positiv getestete Person nicht Urheber der Spur ist, sehr groß: Von 110 Testfällen sind 100 nicht Träger des Profils und 9 Träger, aber nicht die Urheber:

$$P(\bar{U}|T) = P(\bar{D}|T) + \frac{P(\bar{U}|D) \cdot P(D)}{P(T)} = 0,990 = 99,09 \%$$

Auswertung

Wird also behauptet, dass einer Person, die ohne Anfangsverdacht positiv getestet wurde, sicher auch der Urheber der DNA-Spur wäre, so liegt ein doppelter Prävalenzfehler vor, weil sowohl die Prävalenz des DNA-Profiles als auch die des Markers, der zu positiven Testergebnissen führt, übersehen werden. Es werden also $P(U|T)$ und $P(T|U)$ miteinander verwechselt. Obwohl es nahezu sicher ist, dass der Urheber der Spur positiv getestet wird, ist es unwahrscheinlich, dass ein beliebiger positiver Testfall der Urheber ist. Durch die Vertauschung werden die hohen Wahrscheinlichkeiten von $P(T|\bar{U})$ und $P(T|\bar{D})$ unterschlagen.

Der DNA-Test ist im Beispiel ungeeignet, eine ansonsten unverdächtige Person zu belasten. Liegt bereits eine Verdächtigung aufgrund anderer, von der Spur unabhängiger Umstände vor, so kann der Test aber den Verdacht erhärten oder zerstreuen. Seine Aussagekraft steigert sich, je kleiner die Grundgesamtheit der in Frage kommenden Urheber wird – im Beispiel ist diese mit 10 Millionen sehr groß –, jedoch nur so lange, wie sichergestellt werden kann, dass der Urheber der Spur noch in der Grundgesamtheit enthalten ist. Damit von einer Übereinstimmung auf eine Urheberschaft geschlossen werden kann, muss zuerst ein Kreis von Menschen gefunden werden, der objektiv in Frage kommt. Außerdem muss geprüft werden, ob der Täterkreis Personen mit DNA-Markern enthält, die dasselbe Vergleichsergebnis erzielen. Um die Sicherheit weiter zu erhöhen, kann versucht werden, die Fehlerrate des Tests zu senken.

Psychologische Ergebnisse

Psychologische Experimente haben gezeigt, dass die Einschätzung der Wahrscheinlichkeit einer bestimmten statistischen Variable stark verzerrt wird und von der Prävalenz abweicht, sofern zuvor andere Eigenschaften des zu beurteilenden Falls bekannt waren, auch wenn diese keinen Vorhersagewert oder erklärenden Wert für das Auftreten von A besitzt.

Nach Daniel Kahneman und Amos Tversky ist dieser Befund durch die Repräsentativitätsheuristik zu erklären.^[2] Richard Nisbett hat dafür argumentiert, dass Attributionsfehler wie z. B. der *correspondence bias* auf dem Prävalenzfehler beruhen: Die komplexe Prävalenzwahrscheinlichkeit eines Verhaltens in einer Situation wird zugunsten der einfacheren dispositionalen Attribution ignoriert.

Siehe auch

- Bayesscher Wahrscheinlichkeitsbegriff
- Induktion in den Sozialwissenschaften
- Fehlschluss

Literatur und Weblinks

- Max-Planck-Institut für Bildungsforschung (<http://www-abc.mpib-berlin.mpg.de/users/wassner/art1.html>), abgerufen im Dezember 2011
- M. Bar-Hillel: *The base-rate fallacy in probability judgments*. In: *Acta Psychologica*. Band 44, Nr. 3, Mai 1980, S. 211–233. doi:10.1016/0001-6918(80)90046-3
- D. Kahneman, A. Tversky: *On the psychology of prediction*. In: *Psychological Review*. Band 80, 1973, S. 237–251. doi:10.1037/h0034747 (Zusammenfassung (<http://faculty.babson.edu/krolla>

[g/org_site/soc_psych/kan_tver_pred.html](http://org_site/soc_psych/kan_tver_pred.html))

- R. E. Nisbett, E. Borgida, R. Crandall, H. Reed: *Popular induction: Information is not always informative*. In: J. S. Carroll, J. W. Payne (Hrsg.): *Cognition and social behavior*. Band 2, 1976, S. 227–236.
- Gerd Gigerenzer: *Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken*. Berlin 2002, ISBN 3-8270-0079-3.
- The Base Rate Fallacy. (<http://www.fallacyfiles.org/baserate.html>) The Fallacy Files
- Psychology of Intelligence Analysis: Base Rate Fallacy (<https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/psychology-of-intelligence-analysis/art15.html#ft145>)

Einzelnachweise

1. Samuel Lindsey, Ralph Hertwig, Gerd Gigerenzer: *Communicating Statistical DNA Evidence*. In: *Jurimetrics*. Band 43, 2003, S. 147–163, JSTOR 29762803 (<https://www.jstor.org/stable/29762803>).
 2. A. Tversky, D. Kahneman: *Availability: A heuristic for judging frequency and probability*. In: *Cognitive Psychology*. Band 42, 1973, S. 207–232, doi:10.1016/0010-0285(73)90033-9.
-

Abgerufen von „<https://de.wikipedia.org/w/index.php?title=Prävalenzfehler&oldid=207950524>“

Diese Seite wurde zuletzt am 22. Januar 2021 um 20:01 Uhr bearbeitet.

Der Text ist unter der Lizenz „Creative Commons Attribution/Share Alike“ verfügbar; Informationen zu den Urhebern und zum Lizenzstatus eingebundener Mediendateien (etwa Bilder oder Videos) können im Regelfall durch Anklicken dieser abgerufen werden. Möglicherweise unterliegen die Inhalte jeweils zusätzlichen Bedingungen. Durch die Nutzung dieser Website erklären Sie sich mit den Nutzungsbedingungen und der Datenschutzrichtlinie einverstanden. Wikipedia® ist eine eingetragene Marke der Wikimedia Foundation Inc.