

Bayesualize: Visualisierung bedingter Wahrscheinlichkeiten

Barbara Messing, Michael Seubert

Lehrgebiet Wissensbasierte Systeme
FernUniversität in Hagen
58084 Hagen

barbara.messing@fernuni-hagen.de
michael.seubert@gmx.de

Zusammenfassung: *Bayesualize* ermöglicht das komfortable Berechnen von Übergangs- und Randverteilungen und die Darstellung in Baumdiagrammen, im Einheitsquadrat sowie als Paling-Palette. Die Darstellung erfolgt wahlweise mit Wahrscheinlichkeiten, mit Prozentangaben oder mit natürlichen Häufigkeiten. Statistisches Material, das zum Beispiel aus Studien gewonnen wurde, kann mit *Bayesualize* aufbereitet und visualisiert werden. Das Programm ist für tutorielle Zwecke und zur Veranschaulichung von Risiken konzipiert.

1 Einleitung

Der Umgang mit Risiken ist ein vieldiskutiertes und komplexes Thema, zuletzt eindrucksvoll zu studieren am Umgang mit dem H1N1-Virus („Schweinegrippe“) im Winter 2009/2010. Unterschiedliche Informationen und Einschätzungen führten zu tiefer Verunsicherung unter der Bevölkerung.

Gerd Antes, Direktor des Deutschen Cochrane Zentrums an der Universität Freiburg, schrieb am 23.10.2009 in der FAZ im Zusammenhang mit der Schweinegrippe-Diskussion über den Eindruck, „dass die meisten, die sich zu Wort melden, beim Dreisatz in der Schule gefehlt haben“. Er klagt, „selbst einfache Quotientenbildung scheint nicht jedermanns Sache zu sein [...] Dabei ist die quantitative Betrachtung ganz einfach. Alle Zahlen sind nur die eine Hälfte der Geschichte – nämlich der Zähler. Zu jedem Zähler gehört aber auch ein Nenner, damit es ein richtiger Bruch wird. Und nur Brüche sind geeignet, die Risikoverhältnisse richtig zu bewerten. Diese Feststellung ist banal. Umso erstaunlicher ist es, wie konsequent sie ignoriert wird.“

Impfungen, Screenings (Reihenuntersuchungen an gesunden Patienten) und die Reduktion von Risiken sind typische Problemstellungen, bei denen mit bedingten Wahrscheinlichkeiten und dem Satz von Bayes gerechnet werden muss. Dies bedeutet im Grunde nichts anderes, als die korrekte Betrachtung von Brüchen, was aber selbst für ausgebildete Ärzte oftmals alles andere als eine Banalität ist, wie Gigerenzers Experimente zeigen [Gi02].

Hilfreich sowohl für das Verständnis als auch für die Darstellung und Kommunikation von Risiken ist nach Gigerenzers Resultaten die Verwendung natürlicher Häufigkeiten anstelle von Wahrscheinlichkeiten oder Prozentangaben. Dies bedeutet lediglich, die oben erwähnte Quotientenbildung sprachlich anders umzusetzen, sorgt aber für einen beträchtlichen Zuwachs an Verständnis.

Bayesualize soll diese Veranschaulichung unterstützen. Neben dem Umrechnen von Wahrscheinlichkeiten in natürliche Häufigkeiten können mit dem Programm verschiedene graphische Darstellungen aufgerufen werden, die die Größenverhältnisse deutlich machen. Dies soll im Folgenden an zwei Anwendungsbeispielen gezeigt werden.

2 Diagnostische Tests

Bei diagnostischen Tests sind in der Regel die Sensitivität (die Sicherheit, mit der eine Erkrankung erkannt wird) und die Spezifität (die Sicherheit, dass ein Test negativ ausfällt, sofern die Krankheit nicht vorliegt) bekannt. Auf Basis von diesen Daten kann ein einzelnes Testergebnis allerdings noch nicht beurteilt werden, der Grundanteil der Krankheit in der betrachteten Population muss ebenso bekannt sein. Doch auch bei gegebener Prävalenz (Häufigkeit des Auftretens insgesamt) kann der positive oder negative Vorhersagewert intuitiv kaum eingeschätzt werden. Eine genaue Berechnung ist nur mit der Bayes-Formel möglich.

In *Bayesualize* steht speziell für diagnostische Tests eine Eingabehilfe zur Verfügung, bei der Sensitivität, Spezifität und Prävalenz einer Erkrankung eingegeben werden. Nach diesen Angaben werden ein Projekt für diese gemeinsame Wahrscheinlichkeitsverteilung erstellt und entsprechende natürliche Häufigkeiten berechnet (siehe Abbildung 1).

Die Projektdaten können nun in *Bayesualize* in verschiedenen Darstellungsformen visualisiert werden. Das Baumdiagramm mit der Reihenfolge „*Testergebnis => Krankheit*“ würde den positiven bzw. negativen Vorhersagewert direkt anzeigen. Daneben ist die Paling-Palette eine gute Kombination aus natürlichen Häufigkeiten und flächenartiger Darstellung. Es können ein oder zwei Merkmale mit verschiedenen Farben und unterschiedlichen geometrischen Figuren abgebildet werden. Die Anzahl der Figuren und damit die Summe der Häufigkeitsverteilung sind hier in der Paling-Palette stets 1000. In der Paling-Palette in Abbildung 2 mit den eingegebenen Daten von Abbildung 1 lässt sich auf einfache Weise erkennen, dass 69 von 1000 Patienten ein positives Testergebnis zu erwarten haben, aber nur bei 20 von diesen 69 die Krankheit auch wirklich vorliegt.

Diese Information kann in dieser Form und mit dieser Visualisierung mitgeteilt werden und ermöglicht so eine realistische Einschätzung der Situation. Meistens wird übersehen, dass ein sehr sensibler Test vermehrt zu falsch-positiven Resultaten führt, die zur Grundrate jedoch in Beziehung gesetzt werden müssen.



Abbildung 1: Eingabehilfe für diagnostische Tests in *Bayesualize*

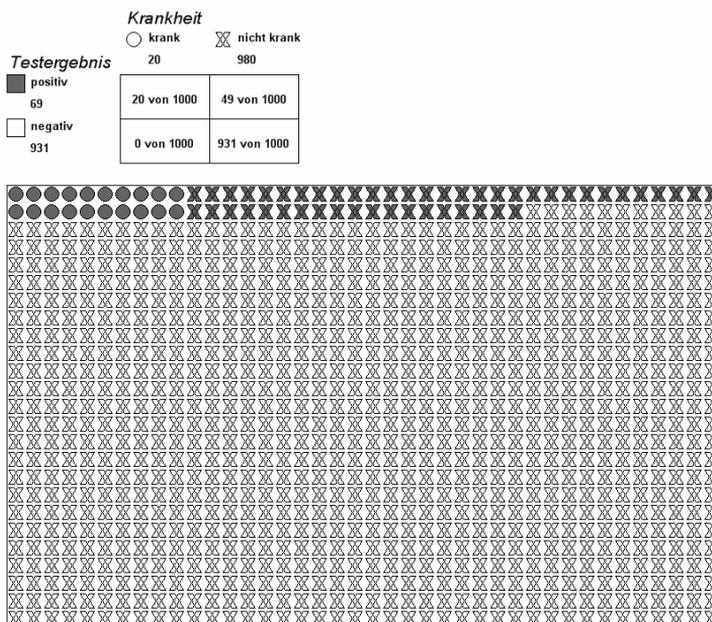


Abbildung 2: Paling-Palette in *Bayesualize*

3 Bedingte Unabhängigkeit

Als Beispiel betrachten wir folgendes Szenario mit fiktiven Daten: Ein Vitaminpräparat, das insbesondere vor Atemwegserkrankungen schützen soll, wird an 95 Personen getestet. 50 Personen erhalten das Präparat, während die restlichen 45 Personen als Kontrollgruppe das Präparat nicht bekommen. Nach einem Jahr haben 10 von den 50 Testpersonen (20%) eine Atemwegserkrankung durchgemacht. Dagegen sind in der Kontrollgruppe 17 von den 45 Personen (38%) erkrankt. Scheinbar wirkt die Vitamingabe vorbeugend.

Angenommen, bei einer näheren Analyse der Daten fiel auf, dass von den 50 Testpersonen, die das Präparat erhalten haben, zehn Personen starke Raucher waren, während in der Kontrollgruppe der Anteil der Raucher deutlich größer war (25 von 45 Personen). In einer Baumdarstellung, die die Raucher von den Nichtrauchern trennt, wird dies sichtbar (Abbildung 3): Bewegt man sich im Teilbaum unterhalb der Raucher, sieht man, dass das Risiko für die Erkrankung bei 60 Prozent liegt, unabhängig davon, ob das Vitaminpräparat genommen wurde oder nicht. Im Teilbaum unterhalb der Nichtraucher liegt Anteil der Erkrankten bei zehn Prozent, ebenfalls unabhängig von der Vitamingabe. Für die Bestimmung des Risikos der Erkrankung ist es unerheblich zu wissen, ob die Person das Vitaminpräparat genommen hat, entscheidend ist das Rauchverhalten. Für einen aussagekräftigen Test hätten die Raucher gleichmäßig auf beide Gruppen verteilt werden müssen.

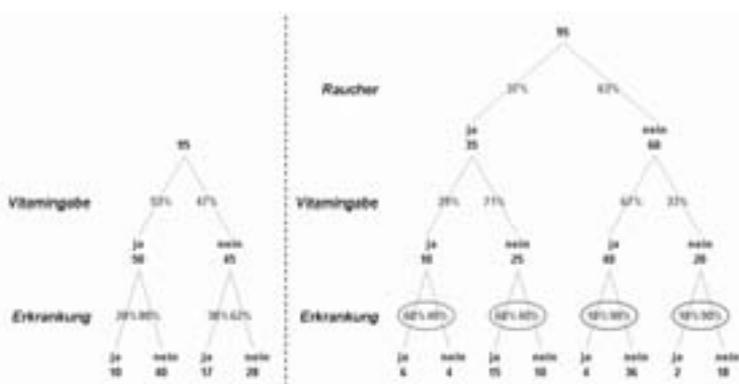


Abbildung 3: Darstellung bedingter Unabhängigkeit in *Bayesialize*

Mit *Bayesialize* lassen sich alle Baumstrukturen und Visualisierungen, die diese Verhältnisse darstellen, ohne weiteren Rechenaufwand darstellen, nachdem die Daten als Tabelle eingelesen oder direkt eingegeben wurden.

4 Implementierung und Verfügbarkeit

Bayesialize wurde in C# auf Basis des .NET 3.5 implementiert; das Microsoft .NET-Framework ist kostenlos im Internet erhältlich. Das Programm kann bei der Erstautorin angefragt werden. Die Systemdemonstration zeigt die Möglichkeiten des Programms anhand der angeführten und an weiteren Beispielen; interessierte Zuhörer können das Programm an eigenen Szenarien testen.

Literaturverzeichnis

- [Gi02] Gigerenzer, G.: Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken. Berlin Verlag, Berlin, 2002.