

# Frequentistische und Bayes'sche Statistik

Karsten Kirchgessner

In den Naturwissenschaften herrscht ein wahrer Glaubenskrieg, ob die frequentistische oder Bayes'sche Statistik als Grundlage zur Auswertung von Experimenten dienen soll. Obwohl die Resultate oftmals sehr ähnlich sind und die Anwendung beider Statistiken gleichermaßen von Erfolg gekrönt ist, weswegen sie überhaupt erst nebeneinander bestehen konnten, hält der Konflikt über die unterschiedliche Anschauung bis in unsere heutige Zeit an.

Bereits bei der **Definition des Wahrscheinlichkeitsbegriffs** werden von beiden Fraktionen unterschiedliche Ansichten vertreten: Die Frequentisten interpretieren die Wahrscheinlichkeit für das Eintreten eines Ereignisses als **relative Häufigkeit**, was für sie den Charakter der Objektivität impliziert. Auf Grund dieser Definition kann die Anwendung jedoch nur bei beliebig wiederholbaren Ereignissen, bei Symmetrien oder unter weiteren äquivalenten Umständen erfolgen – anders als die subjektive Interpretation der Wahrscheinlichkeit von den Bayesianern als „**Grad persönlicher Überzeugung**“, welche auch auf einmalige Ereignisse oder zur Bewertung von Hypothesen herangezogen werden kann.

Wenn wir eine große Anzahl gleicher, unabhängiger Zufallsexperimente durchführen, dann lässt sich die als relative Häufigkeit interpretierte Wahrscheinlichkeit für ein bestimmtes Ereignis mathematisch durch den Limes des Quotienten der absoluten Häufigkeit dieses Ereignisses und der Anzahl der wiederholten Experimente gegen unendlich viele Experimente darstellen:

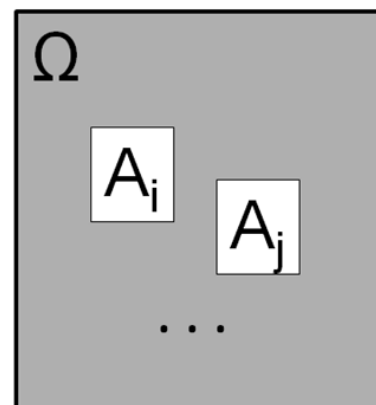
$$P(x) := \lim_{n \rightarrow \infty} \frac{n_x}{n}$$

Kritik an dieser Formulierung findet primär die Tatsache, dass es in der Praxis ohnehin unmöglich ist ein gewisses Experiment unendliche Male zu wiederholen, sowie dass es bereits im Allgemeinen fraglich erscheint, ob sich ein Experiment überhaupt exakt wiederholen lässt, d.h. ob sich die Präparation tatsächlich reproduzieren lässt.

Damit man mit den Wahrscheinlichkeiten wie gewohnt rechnen kann, müssen diese die sogenannten **Kolmogorov-Axiome** erfüllen. Sie fordern, dass die Wahrscheinlichkeit für das Eintreten eines elementaren Ereignisses stets größer oder gleich Null ist (**Positivität**), dass die Wahrscheinlichkeit für das Eintreten von dem einen oder dem anderen von zwei elementaren Ereignissen gleich der Summe der Wahrscheinlichkeiten für das Eintreten der einzelnen Ereignisse ist (**Additivität**), und dass die Summe aller Wahrscheinlichkeiten für das Eintreten der elementaren Ereignisse gleich Eins sein muss (**Normiertheit**).

- $A_i$  – elementares Ereignis
- $\Omega$  – Gesamtheit aller elementaren Ereignisse
- $P(A_i)$  – Wahrscheinlichkeit von  $A_i$

- $P(A_i) \geq 0 \quad \forall i$  – Positivität
- $P(A_i \text{ oder } A_j) = P(A_i) + P(A_j)$  \* – Additivität
- $\sum_{\Omega} P(A_i) = 1$  – Normiertheit



\* falls  $A \cap B = \emptyset$

Aus den Kolmogorov-Axiomen lassen sich Sätze ableiten, die eine Aussage über die Wahrscheinlichkeit kombinierter Ereignisse treffen, z.B. über die Wahrscheinlichkeit der Vereinigung oder des Schnittes zweier Ereignisse, d.h. über die Wahrscheinlichkeit, dass mindestens eines von beiden oder beide zusammen geschehen:

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cap B) = P(A) + P(B) - P(A \cup B)$

Geometrisch lassen sich diese Sätze einfach nachvollziehen (siehe Grafik rechts!):

Die wichtigste Kombination von zwei Wahrscheinlichkeiten ist für die Bayesianer jedoch die sogenannte **bedingte Wahrscheinlichkeit**, welche die Wahrscheinlichkeit für ein Ereignis A unter der Voraussetzung, dass ein Ereignis B bereits eingetreten ist, angibt.

Sie ist wie folgt definiert:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) \neq 0$$

Durch einfache Termumformungen können wir diesen Ausdruck auch auf die Form  $P(A \cap B) = P(B) \cdot P(A|B)$  bzw. unter Berücksichtigung der Tatsache, dass  $P(A \cap B) = P(B \cap A)$  gilt, auf die Form  $P(A \cap B) = P(A) \cdot P(B|A)$  bringen.

Sind A und B unabhängig voneinander, so vereinfacht sich die bedingte Wahrscheinlichkeit für ein Ereignis auf die einfache Wahrscheinlichkeit für dieses Ereignis, da es durch die Unabhängigkeit irrelevant ist, ob das andere Ereignis vorher eingetreten ist oder nicht. Wir können dann  $P(A \cap B) = P(A) \cdot P(B)$  schreiben.

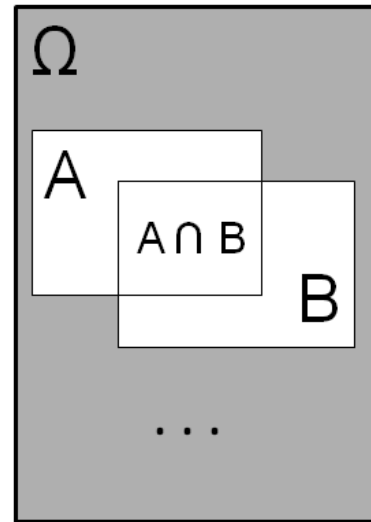
Wenn wir die bedingte Wahrscheinlichkeit  $P(B|A)$  analog zur obigen Definition von  $P(A|B)$  bilden und nach  $P(B \cap A)$  auflösen, können wir unter der Feststellung, dass  $P(A \cap B) = P(B \cap A)$  gilt, den Ausdruck für  $P(B \cap A)$  in die Bestimmungsgleichung von  $P(A|B)$  einsetzen und erhalten damit das **Bayes-Theorem**:

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Der Satz von Bayes stammt von dem Mathematiker und presbyterianischen Pfarrer **Reverend Thomas Bayes** (1702-1761), der Wahrscheinlichkeit als den Grad des Glaubens interpretiert, dass ein Experiment ein bestimmtes Ereignis haben wird.

Den großen Vorteil seiner Interpretation bietet die Tatsache, dass er auch Aussagen über Wahrscheinlichkeiten treffen kann, wenn keine Frequenz-Aussagen möglich sind, z.B. in den folgenden Fällen:

- Morgen wird die Welt untergehen.
- Deutschland wird 2010 Fußball-Weltmeister.



- Die erste Mondlandung war kein Fake.
- Das Teilchen in diesem Experiment ist ein Neutrino.
- Die Natur ist supersymmetrisch.

Dies ist dadurch möglich, dass vorher getroffene Annahmen in die Berechnung der Wahrscheinlichkeiten integriert werden. Diese **Anfangswahrscheinlichkeit** oder **A-priori-Wahrscheinlichkeit** (entspricht dem  $P(A)$  in der obigen Definition des Bayes-Theorems) kann unterschiedlichen Ursprungs sein, z.B.:

- Symmetrieeigenschaften (z.B. Würfel, Münze, Homogenität des Raumes)
- Naturgesetze (z.B. Thermodynamik, Quantenmechanik)
- Erfahrungen, empirische Studien (z.B. Sterbetafeln)
- Vermutungen, Expertenmeinungen

Gegner kritisieren jedoch gerade diesen Punkt, dass im Vorfeld getroffene Annahmen das Resultat der Berechnungen massiv beeinflussen (können). Diese Vorgehensweise sei sehr subjektiv und unwissenschaftlich, da der Glaube auf Kosten der Belegbarkeit in den Vordergrund tritt.

Verfechter der Praxis argumentieren hingegen wiederum, dass der Satz von Bayes auf der einfachen, allseits akzeptierten Wahrscheinlichkeitsrechnung basiert und durch die Erfüllung der Kolmogorov-Axiome ein solides mathematisches Fundament aufweisen kann. Außerdem gebe es bei korrekter Anwendung keine Widersprüche zu dem Ansatz der Frequentisten.

Weitere Vorteile bietet der Bayes-Ansatz durch sein großes Spektrum an Anwendungsmöglichkeiten: Zum einen ist er in der Lage, Aussagen über die Wahrscheinlichkeit von Hypothesen zu treffen, und zum anderen erschließen sich ihm komplett neue Gebiete wie die Mustererkennung oder die Spieltheorie.

Der Preis dafür ist jedoch die Einführung zusätzlicher, nicht beweisbarer Annahmen, was insbesondere bei wissenschaftlichen Untersuchungen ein striktes Reglement erfordert, das nur sinnvolle Annahmen erlaubt und voraussetzt, dass diese auch kommuniziert werden.

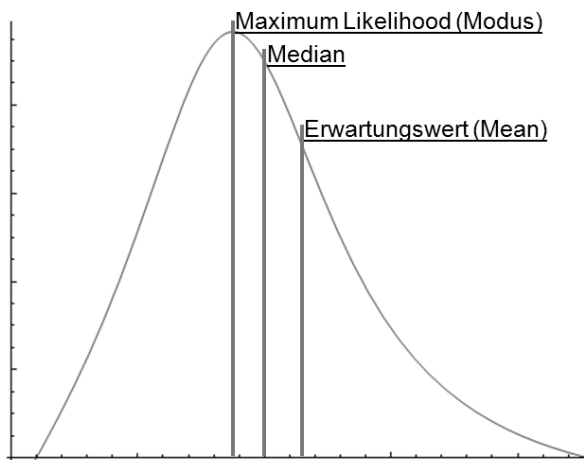
Durch das Treffen von sinnvollen Annahmen können die Resultate unmittelbar verbessert werden. Sogar wenn keinerlei Informationen vorliegen, kann dieser Sachverhalt durch den sogenannten **nicht-informativen Prior** (**Jeffreys Prior**) eingebunden werden. Dieser berechnet sich aus der Quadratwurzel der Determinante der Fisher-Information, welche sich wiederum als Varianz der partiellen Ableitung des Logarithmus der Likelihood berechnen lässt.

Als Beispiel suchen wir die Verteilung der Prior-Wahrscheinlichkeit für eine dimensionsbehaftete Größe  $x$ , von der nicht einmal die Größenordnung bekannt ist. Hierfür erhalten wir nicht etwa eine Gleichverteilung der Prior-Wahrscheinlichkeit, sondern eine Gleichverteilung für den Logarithmus von dieser Größe, da alle Größenordnungen gleich wahrscheinlich sind:

$$f(\ln x) = \text{const.} \Leftrightarrow f(x) \sim 1/x$$

Ein weiterer Streitpunkt zwischen Frequentisten und Bayesianern ist die Angabe der Unsicherheit von Messergebnissen bzw. der Messungenauigkeit von Experimenten.

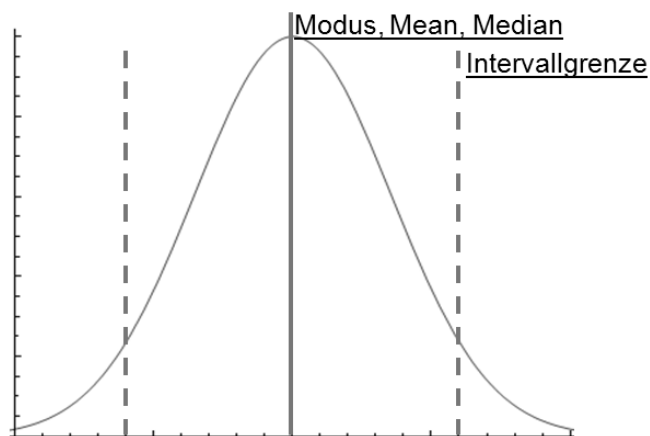
**Punktschätzer** sind eine einzige Zahl, mit der man eine möglichst gute Approximation erreichen möchte, wie z.B. hier:



Anders als bei einer Gauß-Kurve fallen Mittelwert, Median und das Maximum der Likelihood bei diesem Beispiel nicht auf einen Wert, da hier eine asymmetrische Verteilung vorliegt.

Die unterschiedlichen Punktschätzer bieten alle ihre Vor- und Nachteile; beispielsweise ist die Bestimmung des Medians relativ robust gegenüber Abweichungen, aber nicht effizient in der Durchführung. Die Maximum Likelihood ist mit weniger Aufwand zu bestimmen, aber dafür weniger robust.

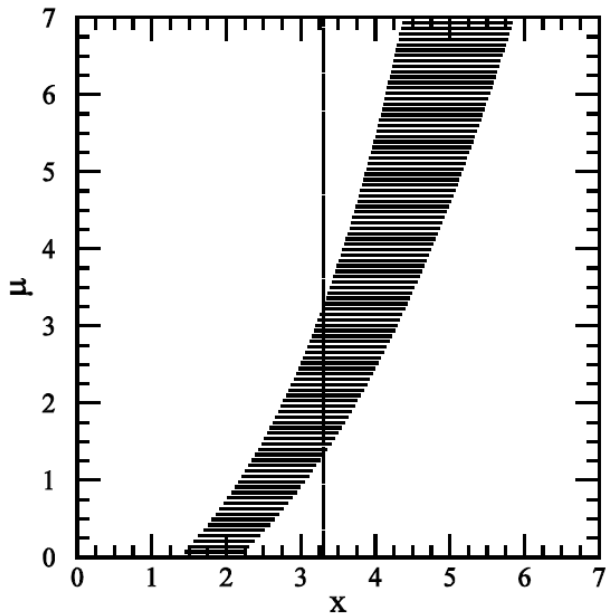
Eine andere Methode um Aussagen über die Messungenauigkeit zu treffen ist die der **Intervallschätzung**, mit dem prominenten Vertreter des **Konfidenzintervalls** (Vertrauensbereich). Dieses schließt einen Bereich um den geschätzten Wert ein, in dem zu einer bestimmten, vorher festgelegten Wahrscheinlichkeit der wahre Wert liegt. Meistens werden hierfür  $n\sigma$ -Grenzen oder hohe Prozentwerte gewählt.



Vorteil dieser Darstellung ist die Möglichkeit, unmittelbar Aussagen über die Präzision und Signifikanz treffen zu können. Ein sehr breites Intervall deutet beispielsweise auf einen zu geringen Umfang an Stichproben oder ein sehr variables Phänomen hin.

Die Frequentisten betrachten den Parameter der Theorie  $\theta$  als fest, aber unbekannt ( $f(x|\theta)$ ). Das Konfidenzintervall als Zufallsvariable überdeckt den wahren Wert nun zu einer bestimmten Wahrscheinlichkeit (Coverage).

Zur Bestimmung des Konfidenzintervalls berechnet man für jeden möglichen wahren Wert die tolerierte Diskrepanz und trägt diese als Akzeptanzintervall in Form von horizontalen Linien in ein Diagramm (siehe unten!), das auf den Achsen den wahren und den Messwert enthält. Durch das ermittelte Messergebnis trägt man nun eine vertikale Linie auf. Das Konfidenzintervall besteht nun aus den Werten für  $\mu$ , bei denen die vertikale Linie das Akzeptanzintervall schneidet



Die Bayesianer betrachten den Parameter der Theorie  $\theta$  als Zufallsvariable ( $f(\theta|x) = f(x|\theta) f(\theta) / f(x)$ ), der sich zu bestimmter Wahrscheinlichkeit im Kreditabilitätsintervall  $[\theta_1, \theta_2]$  befindet.

In den Naturwissenschaften ist der Satz von Bayes ein wichtiges Instrument zur Verifikation oder Falsifizierung von Theorien. Dies wird mit seiner Interpretation mit den Argumenten „Theorie“ und „Daten“ praktiziert. Statt zu überprüfen, wie gut gemessene Daten zu einer Theorie passen, bestimmt man direkt explizit, inwieweit eine Theorie mit aufgenommenen Messungen korrespondiert.

$$\underbrace{P(\textit{Theorie}|\textit{Daten})}_{\text{Posterior}} = \frac{\overbrace{P(\textit{Daten}|\textit{Theorie})}^{\text{Likelihood}} \overbrace{P(\textit{Theorie})}^{\text{Prior}}}{\underbrace{P(\textit{Daten})}_{\text{Evidenz}}}$$

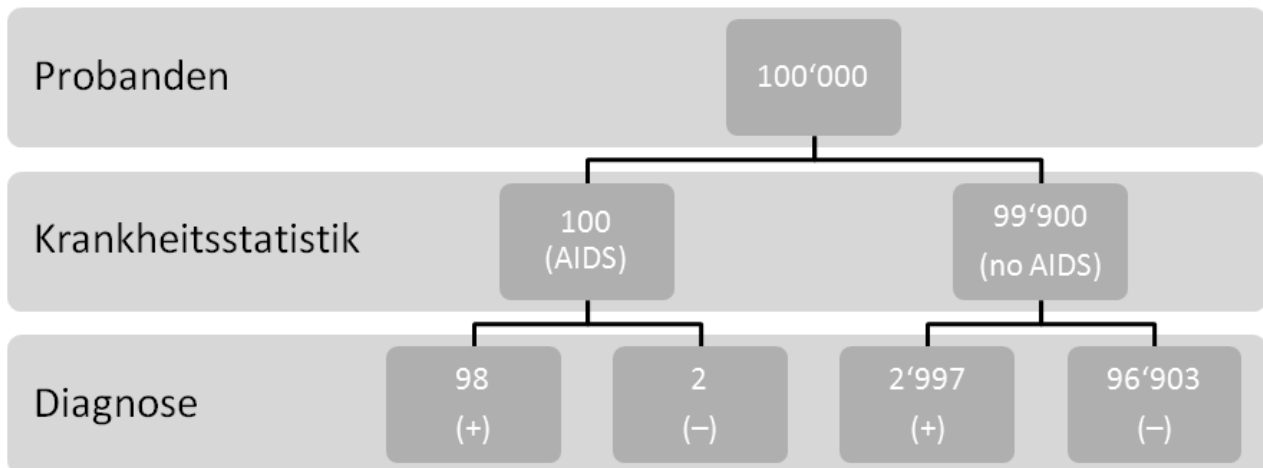
Ein Paradebeispiel für die Anwendung des Theorems von Bayes ist ein Screening-Test zur Diagnose einer Krankheit, z.B. AIDS. Als A-priori-Wissen lassen wir die Tatsache und damit verbundene Anfangswahrscheinlichkeit einfließen, dass 1‰ der allgemeinen Bevölkerung infiziert ist und der Rest nicht. Weitere zur Berechnung der bedingten Wahrscheinlichkeit notwendige Informationen fließen durch unser Wissen über die Güte des Tests ein, nämlich folgende Wahrscheinlichkeiten für positive und negative Diagnosen bei infizierten und nichtinfizierten Probanden:

$$P(+|AIDS) = 0.98 \quad P(-|AIDS) = 0.02 \quad P(+|no AIDS) = 0.03 \quad P(-|no AIDS) = 0.97$$

Bereits hier fällt auf, dass die Wahrscheinlichkeit als gesunde Person eine positive Diagnose zu erhalten höher ist, als die Verbreitung in der allgemeinen Bevölkerung. Die Berechnung der Posterior-Wahrscheinlichkeit deckt auf, dass man als Nicht-Angehöriger einer Risikogruppe nur mit einer Wahrscheinlichkeit von 3,2% tatsächlich infiziert ist:

$$\begin{aligned}
 P(AIDS|+) &= \frac{P(+|AIDS) P(AIDS)}{P(+|AIDS) P(AIDS) + P(+|no AIDS) P(no AIDS)} \\
 &= \frac{0.98 \times 0.001}{0.98 \times 0.001 + 0.03 \times 0.999} \\
 &= 0.032
 \end{aligned}$$

Die Überprüfung mit dem Wahrscheinlichkeitsbaum macht die Korrektheit der Berechnung schnell plausibel:



Auch außerhalb der Naturwissenschaften basieren auf dem Bayes-Theorem etliche Anwendungen, die in unserem Alltag allgegenwärtig sind, z.B. in Form des Bayes'schen-Filters, einem SPAM-Filter, der an Hand des Auftretens bestimmter Begriffe entscheidet, ob eine unerwünschte Mail vorliegt oder nicht. Durch das Training des Benutzers lernt der Filter immer besser korrekte Urteile zu fällen, sodass das Risiko von falsch-positiven Fällen bei gutem Training geringer ist als bei manueller Filterung. Spammer versuchen mit Gegenmaßnahmen den Filter zu überlisten, z.B. indem sie die kritischen Wörter als Bilddateien versenden, modifizierte Schreibweisen verwenden oder unverwerfliche Texte anhängen, um die Statistik zu manipulieren. Letzteres ist aber keine besonders erfolgreiche Methode, da hierbei ohnehin nur „harmlose“ Ausdrücke künstlich eingefügt werden, die bei der Analyse und Klassifikation der Mails keine wesentliche Rolle spielen.