Bayes'sche Statistik für Fußgänger

von

Oliver Passon

Inhaltsverzeichnis

1 Motivation

Die Bayes'sche Statistik ist ein alternativer Ansatz zur Gewinnung und Interpretation von Ergebnissen in der statistischer Datenanalyse. Von der herkömmlichen Statistik unterscheidet sie sich durch eine andere Definition von Wahrscheinlichkeit ("subjektivistisch" statt "frequentistisch"). Betrachten wir zunächst ein einfaches Beispiel um zu motivieren, warum die übliche Definition von Wahrscheinlichkeit kritisiert werden kann:

Gegeben sei das Ergebnis einer Messung des physikalischen Parameters μ, etwa der W Masse, $\mu = 80.35 \pm 0.02$ GeV. Üblich ist es, die experimentelle Standardabweichung s (etwa bei wiederholter Durchführung des Experiments) als "Standardfehler" anzugeben. Belasten wir uns nicht mit konkreten Zahlen und Einheiten, und schreiben:

$$\mu = \bar{x} \pm s$$

s bezeichnet den Schätzer der tatsächlichen Standardabweichung σ. Was ist die Interpretation dieser Aussage? Nehmen wir an die Größe sei Normalverteilt, und mithin im σ Intervall um den Mittelwert 68% der Fläche unter der Wahrscheinlichkeitsdichte. Bedeutet das Ergebnis also, dass mit 68% Wahrscheinlichkeit der tatsächliche Wert μ in diesem Intervall um die Messung $\bar{x}$ liegt? formal:

$$P(\bar{x} - \sigma \leq \mu \leq \bar{x} + \sigma) = 68\% \tag{1}$$

oder lautet die Interpretation von (1):

$$P(\mu - \sigma \leq \overline{X} \leq \mu + \sigma) = 68\% \tag{2}$$

In Worten: Die Wahrscheinlichkeit jedes weiteren Ausfalls der Messung $\overline{X}$ im σ Intervall um den wahren (und unbekannten) Wert μ zu liegen, ist 68%.

Spontan findet man die erste Aussage sinnvoller, da das Intervall in der zweiten Gleichung vom unbekannten Parameter μ abhängt, und man mithin dem Ziel der Messung, nämlich den tatsächlichen μ Wert einzugrenzen, nicht näher zu kommen scheint. Die herkömmliche Statistik belehrt einen jedoch, dass die erste Aussage (Gleichung 1) nicht nur *falsch*, sondern sogar *sinnlos* ist! "Herkömmlich" meint dabei, auf dem "frequentistischen" Wahrscheinlichkeitsbegriff (siehe Abschnitt 2.2) beruhend. In diesem Rahmen macht es aber keinen Sinn, über die Wahrscheinlichkeit von μ zu sprechen ("*eine Konstante von unbekanntem Wert*") , sondern lediglich über Zufallsvariablen [1] wie $\overline{X}$ können Wahrscheinlichkeitsaussagen getroffen werden.

Im Rahmen der konventionellen Statistik ist tatsächlich Gleichung 2 die korrekte Interpretation des Messergebnisses. Es wird also lediglich eine Aussage über den Ausgang weiterer *Messungen* getroffen. Dieses Problem stellt sich bei der Interpretation statistischer Aussagen in der Physik grundsätzlich. Eine Aussage vom Typ *"auf dem 90% CL ist die Higgsmasse über 96 GeV"* bedeutet lediglich die Wahrscheinlichkeit, in wiederholten identischen Messungen ein Ergebnis zu erzielen, dass mit besagter Higgs-Massenschranke konsistent ist. Es kann konventionell *nicht* als *Wahrscheinlichkeit für die Hypothese* gedeutet werden, dass die Higgsmasse größer als 96 GeV ist, denn die Higgsmasse ist keine Zufallsvariable und liegt (mit 100% Wahrscheinlichkeit) **entweder** innerhalb **oder** außerhalb des besagten Intervalls. Die Hypothesentests der konventionellen Statistik treffen lediglich Wahrscheinlichkeitsaussagen bei denen die Intervallgrenzen als Zufallsvariable gedeutet werden, was beachtliche sprachliche Verrenkungen bei der Interpretation der Ergebnisse provoziert.

In der Sprache bedingter Wahrscheinlichkeiten ausgedrückt sieht die konventionelle Statistik nur Aussagen vom Typ $P(\text{Daten} \mid \text{Theorie})$ vor, sprich: Wahrscheinlichkeit diese Daten zu messen, unter der Bedingung dass die Hypothese

[1] Die Zufallsvariable ist hier mit $\overline{X}$ bezeichnet, und $\overline{x}$ aus der Gleichung $\mu = \bar{x} \pm \sigma$ bedeutet einen Ausfall der Variable.

(hier symbolisch als Theorie bezeichnet) stimmt, wohingegen intuitiv die Wahrscheinlichkeit $P(\text{Theorie} \mid \text{Daten})$ gesucht wird, also eine Aussage vom Typ wie groß ist die Wahrscheinlichkeit, dass die Higgsmasse über 96 GeV liegt, oder die geladene Spur ein π^- und kein K^- ist, jeweils angesichts der genommenen Daten. Unbewusst werden viele Ergebnisse der konventionellen Statistik interpretiert, als wenn die Gleichung

$$P(\text{Theorie} \mid \text{Daten}) = P(\text{Daten} \mid \text{Theorie}) \tag{3}$$

gilt, aber zum einen ist die linke Seite in diesem Rahmen nicht definiert, zum anderen verrät einem der Satz von Bayes (siehe Abbschnitt 2.1), dass die Umwandlung bedingter Wahrscheinlichkeiten etwas komplizierter ist.

Um diese Problem zu lösen, wird man also erstens eine umfassendere Definition von Wahrscheinlichkeit brauchen, die auch Hypothesen wie die oben genannten erfasst, und desweiteren die Umwandlung von bedingten Wahrscheinlichkeiten untersuchen. Die "Bayes'sche Statistik", der die subjektivistische Definition von Wahrscheinlichkeit zugrunde liegt, verfolgt genau dieses Program.

Bevor dies in mehr Detail erklärt wird, sei ein kleiner Exkurs über die Definition von Wahrscheinlichkeit vorangestellt.

2 Wahrscheinlichkeit

2.1 Definition

Formal kann man Wahrscheinlichkeit als Abbildung P von beliebigen Teilmengen einer zunächst nicht weiter bestimmten Menge Ω ("Menge aller Ereignisse") auf das reelle Intervall von 0 bis 1 auffassen. Eine Abbildung nämlich, die die folgenden Eigenschaften erfüllt:

1. $P(A) \geq 0, \forall\ A \in \Omega$
2. A, B disjunkt $\Rightarrow P(A \cup B) = P(A) + P(B)$
3. $P(\Omega) = 1$

Aus diesen Axiomen können Eigenschaften der "P–Abbildung" abgeleitet werden, wie etwa: $A \subset B \Rightarrow P(A) \leq P(B)$. Man definiert weiter die "bedingte Wahrscheinlichkeit" $P(A \mid B)$ (sprich: Wahrscheinlichkeit für A, falls B gegeben)

$$P(A \mid B) = \frac{P(A \cap B)}{P(A)}$$

offensichtlich muss $P(A) \neq 0$ gelten. Unter der Bedingung $P(B) \neq 0$ kann man natürlich ebensogut

$$P(B \mid A) = \frac{P(B \cap A)}{P(B)}$$

schreiben, und da die Vereinigung von Mengen kommutativ ist, findet man:

$$P(A \mid B) = P(B \mid A) \cdot \frac{P(A)}{P(B)} \tag{4}$$

Diese Beziehung ist gerade der **Satz von Bayes** [1], und dieser wird (wie der Name schon verrät) im Folgenden eine wichtige Rolle spielen. Das Theorem von Bayes ist aber auch innerhalb der üblichen Statistik und Wahrscheinlichkeitsrechnung von großer Bedeutung. Die Beziehung 4 zeigt, worin der Fehler von Gleichung 3 besteht. Beim Vertauschen der Argumente A und B einer bedingten Wahrscheinlichkeit tritt zusätzlich der Faktor $P(A)/P(B)$ auf.

Mit Hilfe bedingter Wahrscheinlichkeiten lässt sich auch folgender Ausdruck für die Wahrscheinlichkeit $P(B)$ angeben: Sei eine disjunkte Zerlegung von Ω durch die i Mengen A_i gegeben, dann gilt:

$$P(B) = \sum_i P(B \mid A_i) P(A_i) \tag{5}$$

Damit kann der **Satz von Bayes** auch wie folgt aufgeschrieben werden:

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{\sum_i P(B \mid A_i) P(A_i)} \tag{6}$$

Natürlich ist die Frage der **Interpretation** und **Berechnung** von Wahrscheinlichkeiten durch diese formalen Definition noch unberührt, und gerade hier liegt die Differenz zwischen herkömmlicher und Bayes'scher Statistik!

2.2 Interpretation

Die Differenz zwischen konventioneller– und Bayes'scher Statistik liegt in der **Interpretation** von Wahrscheinlichkeit. Die konventionelle Statistik erklärt:

- $P(A) =$ Limes relativer Häufigkeit des Ereignisses A

Es liegt also ein Prozess vor, der verschiedene Ereignisse A, B, $\cdots$ produziert. Die relative Häufigkeit des Ausfalls von A ist der Schätzwert für seine Wahrscheinlichkeit. Die exakte Wahrscheinlichkeit gewinnt man im hypotetischen Fall einer unendlich oft wiederholten Messreihe (deshalb "Limes"). Das hübsche an dieser Definition ist, dass sie die Vorschrift zur näherungsweisen *Berechnung* der Wahrscheinlichkeit gleich mitliefert.

Für die Bayes'sche Statistik ist dieser Begriff von Wahrscheinlichkeit jedoch zu eng und unintuitiv, da er nur Aussagen über (zumindestens prinzipiell) beliebig oft identisch wiederholbare Ereignisse zulässt [2]. Eine Aussage darüber wie wahrscheinlich das Eintreffen einer Hypothese ist kann in diesem Rahmen nicht formuliert werden. Zum Beispiel ist die Behauptung einer *"80%igen Regenwahrscheinlichkeit am Wochenende"* ohne (frequentistische) Bedeutung, da das nächste Wochenende ein einmaliges Ereignis ist[3]. Deshalb führt die Bayes'sche Statistik Wahrscheinlichkeit als logisch einfachen Grundbegriff ein, der den Grad der eigenen (Un–)Kenntnis quantifiziert. In diesem Sinne kann man also etwa sprechen:

- $P(A) =$ Grad der Überzeugung, dass die Hypothese A richtig ist bzw. das Ereignis A eintritt

Diese *subjektivistische* Interpretation der Wahrscheinlichkeit bleibt natürlich akademisch, solange man keine Vorschrift zu ihrer Berechnung kennt. Tatsächlich kann in Abwesenheit irgendeiner Messung bzw. Beobachtung, über deren Ausgang die Hypothese eine Aussage trifft, ihr "Wahrheitsgrad" lediglich geraten werden. In Anwesenheit von Daten kann die Berechnung von $P(H)$ jedoch als Problem von bedingter Wahrscheinlichkeit $P(H \mid \text{Daten})$ aufgefasst werden.[4] Zu ihrer Bestimmung verwendet man den Satz von Bayes (Gleichung 4), nachdem

$$P(H \mid \text{Daten}) \propto P(\text{Daten} \mid H) \cdot P(H)$$

Die linke Seite der Gleichung wird Wahrscheinlichkeit aposteriori genannt, und setzt sich aus dem *Likelihood* der konventionellen Statistik $P(\text{Daten} \mid H)$ und dem sog. *Prior* $P(H)$ zusammen. Letzterer repräsentiert die Kenntnis *vor* der Messung. Sinngemäß kann *Prior* mit "Vorurteil" übersetzt werden. In diesem Konzept wird also formalisiert, wie neue Daten den vorherigen Kenntnisstand modifizieren.

[2] Ganz zu schweigen von der mathematischen Subtilität, ob obiger Limes überhaupt existiert. Dieser Einwand hat für Physiker, mit ihrer traditionell laxen Einstellung zu mathematischen Beweistechniken, jedoch nur geringes Gewicht.

[3] Der Autor ist sich bewusst, dass diese Aussage gerettet werden kann, wenn man nicht vom "nächsten Wochende", sondern von Tagen mit identischer (besser: ähnlicher) Wetterlage spricht.

[4] An dieser Stelle trifft die subjektivistische Wahrscheinlichkeitsdefinition häufig der Vorwurf der Unwissenschaftlichkeit. Wie kann ein klar denkender Mensch das "Raten" – später werden wir noch dem Begriff "Vorurteil" begegnen – zur Grundlage wissenschaftlicher Analyse erheben? Bei nüchterner Betrachtung stellt sich die Situaton jedoch sofort anders dar: In Abwesenheit irgendwelcher Daten oder Beobachtungen kann ???

Im Allgemeinen wird man es allerdings mit *kontinuierlichen* Zufallsgrößen zu tun haben, aber der Satz von Bayes (Gleichung 6) verallgemeinert sich leicht auf Wahrscheinlichkeitsdichten. Bezeichnet etwa $\overline{x}$ das Ergebnis einer Messung des Parameters μ, kann man schreiben:

$$f(\mu \mid \overline{x}) = \frac{f(\overline{x} \mid \mu) \cdot f_0(\mu)}{\int f(\overline{x} \mid \mu) f_0(\mu) d\mu}$$

Hier ist nun auch der Nenner ausgeschrieben worden, der aus der Normierungsbedingung für die Wahrscheinlichkeitsdichte gewonnen werden kann. Aus dem Prior P(Theorie) ist nun die Wahrscheinlichkeitsdichte $f_0(\mu)$ geworden – also eine Verteilungsfunktion. Der Quotient aus Likelihood $f(\overline{x} \mid \mu)$ und Normierung wird auch als Bayesfaktor bezeichnet.

Das Theorem von Bayes und die subjektivistische Interpretation von Wahrscheinlichkeit sind also unmittelbar miteinander verknüpft, da im Gegensatz zu frequentistischen Deutung keine "Berechnungsvorschrift" für Wahrscheinlichkeiten unmittelbar aus der Defition folgt. Es existiert jedoch eine (halb ernstgemeinte) Operationalisierung für die subjektivistische Wahrscheinlichkeit. Man kann sich nämlich fragen, welchen Betrag ein "rationaler" Wetter (will heißen: man selbst) bei einem Gewinnbetrag von 100 DM auf die Hypothese setzen würde. "Rational" meint dabei, dass negative Einsätze (bzw. solche über 100 DM) ausgeschlossen sind. Zudem bedeutet ein Einsatz von maximal N DM *für* die Hypothese, dass man auch bereit ist 100-N DM *gegen* sie zu setzen. Fast man den Geldbetrag als die prozentuale Wahrscheinlichkeit auf, folgt, dass diese die Axiome des letzten Kapitels erfüllt!

3 Beispiele und Anwendungen

Wir betrachten nun einige Beispiele die zum Einen die Grundbegriffe der Bayes'schen Statistik illustrieren (Beispiel 1 und 2) und ihre konkrete Anwendung deutlich machen (Beispiel 3).

3.1 Wahrscheinlichkeit ist immer "bedingt"

Nach dem Bayes'schen Schema hängt die Wahrscheinlichkeit immer auch von sog. Prior ab, also dem Kenntnisstand *vor* der Messung. Dieser Punkt erscheint zunächst unbefriedigend. Das folgende Beispiel kann helfen, dieses Unbehagen zu mindern: Wir planen ein Experiment mit einem "unverdächtigen" Würfel – will heißen er erscheint vollkommen symmetrisch und keinen Ausfall begünstigend. Die Wahrscheinlichkeit für den Ausfall "1" im ersten Wurf schätzen wir somit zu $\frac{1}{6}$. Ebenso natürlich die Wahrscheinlichkeit, eine “1” im 1000sten Wurf zu erzielen. Wenn wir jedoch nach 500 maligem Werfen 100 mal das Ereignis “1” gefunden haben, hegen wir einigen Zweifel daran, ob der Würfel wirklich symmetrisch ist, und wir modifizieren unsere Erwartung für die folgenden Würfe. Dies wird aber gerade im Konzept der Bayes'schen Statistik formalisiert! Dabei soll in diesem Beispiel nicht irritieren, dass die Wahrscheinlichkeit der ursprünglichen Hypothese (P(1)=1/6) mit der kombinatorischen Definition zusammenfällt, oder dass die beobachtete relative Häufigkeit von $\frac{1}{5}$ Anlass zum Zweifel gibt. Diesen Begriffen wird lediglich ihre herausragende Bedeutung bei der *Interpretation* von Wahrscheinlichkeit abgesprochen.

Es zeigt sich hier, dass *Wahrscheinlichkeit* immer *bedingte Wahrscheinlichkeit* ist, mithin unausweichlich das Theorem von Bayes eine Rolle spielt. Kehrt man dieses Argument um, findet man das Priorkonzept vielleicht unschön aber unausweichlich.

3.2 Ein konzeptionelles Beispiel

Betrachten wir ein weiteres konzeptionelles Beispiel zum Verhältnis und *Likelihood* und *Prior*. Gegeben sei ein Zufallsgenerator der standard–normalverteilte Zahlen liefert (also $\mu = 0$ und $\sigma = 1$). Wir erhalten den Ausfall 3.01. Die “Wahrscheinlichkeit” (in Bayes–Terminologie der *Likelihood*) hierfür beträgt 0.27% (bzw. richtiger: $P(\mid x \mid > 3) = 0.27\%$). Bedeutet dies nun eine Wahrscheinlichkeit von 0.27%, dass dieses Ereignis aus einer Standardnormalverteilung stammt? Sicherlich nicht, denn wir wissen in unserem Beispiel ja zu 100%, dass die Zahlen von einem Zufallsgenerator stammen. Für die Hypothese H_0 = *Der Generator liefert standard–normalverteilte Zufallszahlen*, und unser Datum $\mid x \mid > 3$ können wir also formal schreiben:

$$P(H_0 \mid x) = \frac{P(x \mid H_0) \cdot P(H_0)}{P(x)}$$

$P(x)$ ersetzen wir gemäß Gleichung 5 durch:

$$P(x) = P(x \mid H_0)P(H_0) + P(x \mid \overline{H_0})P(\overline{H_0})$$

Da in unserem Beispiel gilt[5] $P(H_0) = 1$, $P(\overline{H_0}) = 0$ finden wir:

$$P(H_0 \mid x) = \frac{0.027 \cdot 1}{0.027 \cdot 1 + P(x \mid \overline{H_0}) \cdot 0} = 1$$

Diese sicherlich etwas extreme Beispiel zeigt jedoch, dass beliebig geringe Likelihood Werte trotzdem zu einer hohen aposteriori Wahrscheinlichkeit (Hier sogar Wahrscheinlichkeit “1”) führen können.

[5] Wir *wissen* nach Voraussetzung, dass die Hypothese korrekt ist!

Diese Beispiel ist etwa lehrreich, wenn man an die berühmten HERA Ereignisse denkt, die Ende '97 für beträchtliche Furore sorgten. Die dort beobachteten Ereignisse hatten unter der Hypothese ein Standard Modell Prozess zu sein, eine Wahrscheinlichkeit von ca. 1%, also $P(\text{Daten} \mid SM) = 1\%$. Dieses Resultat wurde in der Presse (und unter zahlreichen Physikern) in die Aussage verwandelt, dass die Widerlegung des Standard Modells durch diese Daten zu 99% sicher ist, also $P(\overline{SM} \mid \text{Daten}) = 99\%$. Wie viele Physiker wären aber tatsächlich bereit gewesen, gegen 100 DM Preisgeld, 99 DM auf die Aussage *"Das Standard Modell ist durch die HERA Messung widerlegt"* zu setzen? In Bayes'scher Terminologie ausgedrückt illustriert dies, dass nicht der *Likelihood* alleine, sondern erst das Produnkt aus *Likelihood* und *Prior* die relevante Wahrscheinlichkeit liefert. Im Falle des experimentell glänzend bestätigten Standard Modells (und in Abwesenheit einer konkreten Erklärung der Ereignisse durch Physik jenseits des Standard Modells) hat der Prior für die Hypothese *"Das Standard Modell ist korrekt"* jedoch einen hohen Wert.

3.3 Wirkt Vitamin C erkältungslindernd?

Das nächste Beispiel ist dem Artikel [2] entlehnt, der in der Literatur gerne als Einführung in die Bayes'sche Statistik empfohlen wird. Es wird die Hypothese untersucht, ob Vitamin C erkältungslindernd wirkt. Bei 17 untersuchten Paaren findet man, dass in 13 Fällen die mit Vitamin C behandelten einen besseren Gesundheitszustand als die mit Plazebos behandelten aufweisen. Die Nullhypothese, dass Vitamin C keinen Effekt hat, entspricht einer Binomialverteilung mit $n = 17$ und $p = \frac{1}{2}$ für den zu erwartenden Ausgang des Experiments. Die Wahrscheinlichkeit für den Ausfall $k = 13$ ist dann lediglich 0.0182:

$$P(k = 13) = \frac{n!}{k!(n-k)!} \cdot p^k (1-p)^{n-k} = \frac{17!}{13! \cdot 4!} \cdot \left(\frac{1}{2}\right)^{17} = 0.0182$$

Üblich ist eine Aussage über die Hypothese auf der Grundlage des sog. P Wertes zu treffen, das heißt der summierten Wahrscheinlichkeit für Ereignisse die genauso oder noch unwahrscheinlicher als das beobachtete sind[6]. In unserem Fall wären das die Ausfälle 0 – 4 und 13 – 17. Diese Summe ergibt für unser Beispiel 0.049. Oft wird nun die Nullhypothese verworfen bzw. ihr Gegenteil für signifikant besser eingeschätzt, falls dieser P Wert kleiner als 0.05 ist. Als Resultat würde also die Hypothese, dass Vitamin C keinen Effekt hat, auf dem 95% Konfidenzniveau verworfen.

In der Bayes'schen Analyse verzichtet man darauf, seinen Schluss auf "nicht beobachtete Daten" zu stützen (d.h. die Wahrscheinlichkeit der Ausfälle 0 – 4 und

[6] Dies stellt im übrigen einen weiteren Kritikpunkt am herkömmlichen Vorgehen dar, da mit der Wahrscheinlichkeit *nicht* gemessener Daten argumentiert wird.

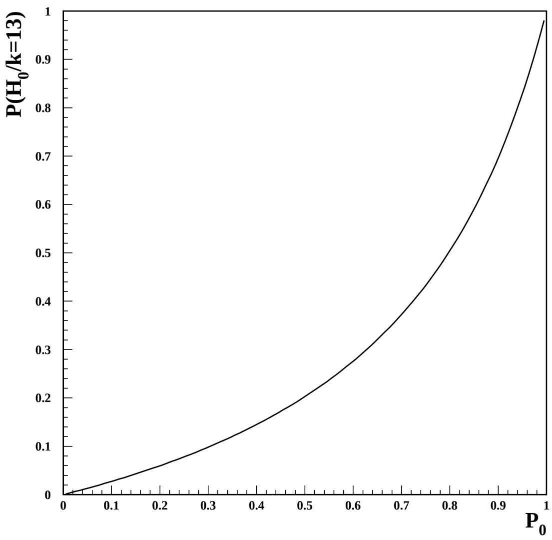

Abbildung 1: $P(H_0 \mid k = 13)$ als Funktion des Priorwertes $P(H_0)$. Bei $P(H_0) = 0.5$ kann man die Wkt. von ≈ 0.2 für die Nullhypothese ablesen.

13 – 17). Stattdessen wird man zusätzlich seine vorherige Überzeugung zum Ausgang des Experimentes angeben müssen – gestützt auf andere Untersuchungen oder Vorurteile. Naheliegend erscheint etwa den Prior $P_0(H_0) = 0.5$ zu wählen, also eine 50% Wahrscheinlichkeit *für* die Hypothese (Vitamin C ohne Effekt, die Daten können durch eine Binomialverteilung mit $p = \frac{1}{2}$ beschrieben werden). Offensichtlich gilt dann ebenfalls $P(\overline{H_0}) = 0.5$. Im Nenner des Satzes von Bayes geht jedoch der Ausdruck $P(k = 13 \mid \overline{H_0})$ ein. Es stellt sich also die Frage, wie wir die p Werte für die restlichen 50% verteilen. In [2] wird hierzu ein weitere Parameter p_0 eingeführt, der angibt welcher Wahrscheinlichkeit man der Vitamin C Wirksamkeit maximal zutraut und eine gleichförmige Verteilung von $1 - p_0$ bis p_0 angenommen. Mir erscheint es jedoch sinnvoller, von vornherein $p_0 = 1$ anzunehmen, also allen anderen Parametern $p \neq \frac{1}{2}$ die selbe Wahrscheinlichkeit einzuräumen. Man findet somit:

$$
\begin{aligned}
P(H_0 \mid k = 13) &= \frac{P(k = 13 \mid H_0) \cdot P(H_0)}{P(k = 13 \mid H_0) \cdot P(H_0) + P(k = 13 \mid \overline{H_0}) \cdot P(\overline{H_0})} \\
&= \frac{\frac{17!}{13!4!} \frac{1}{2}^{17} \cdot \frac{1}{2}}{\frac{17!}{13!4!} \frac{1}{2}^{17} \cdot \frac{1}{2} + \int_0^1 \frac{17!}{13!4!} p^{13}(1-p)^4 dp \cdot \frac{1}{2}} \\
&= \frac{\frac{1}{2}^{17}}{\frac{1}{2}^{17} + \int_0^1 p^{13}(1-p)^4 dp}
\end{aligned}
$$

$$= \left(1 + 2^{17} \int_0^1 p^{13}(1-p)^4 dp\right)^{-1} \approx 0.2$$

Mit anderen Worten erhält man in diesem Rahmen eine $\approx 20\%$ Wahrscheinlichkeit für die Hypothese (Vitamin C ohne Effekt), und umgekehrt $\approx 80\%$ für eine Wirksamkeit der Vitamin C Gabe. Die Abbildung 1 zeigt $P(H_0 \mid k = 13)$ als Funktion des Priorwertes $P(H_0)$. Falls man also einen anderen Priorwert bevorzugt, kann man aus dieser Abbildung den Einfluss auf die Hypothesenwahrscheinlichkeit sofort ablesen.

Auf den ersten Blick ist das Ergebnis der Bayes'schen Analyse also einfach nur konservativer, da die Hypothese der Vitamin C *Un*wirksamkeit nicht mehr auf dem 95% CL verworfen wird, sondern nun eine 20 %ige Wahrscheinlichkeit hat, richtig zu sein. Auf den zweiten Blick realisiert man jedoch, dass die Aussagen konzeptionell vollkommen verschieden sind: Während die herkömmliche Analyse ein unintuitives Konfidenzniveau definiert, erlaubt die Bayes'sche Statistik über das Zutreffens der Hypothese eine Wahrscheinlichkeitsaussage zu machen. Der Bayesianer darf also den Satz sprechen: Für denjenigen der bisher geschwankt hat ($P(\overline{H_0}) = 0.5$) ob Vitamin C erkältungslindernd wirkt, ist die Wirksamkeit nun mit 80%iger Wahrscheinlichkeit sicher!

Falls eine erneute Messung vorgenommen wird, kann nun das Ergebnis dieser Analyse als Prior verwendet werden. Dies ist ganz im Sinne der Bayes'schen Methode, bisherige Überzeugungen im Licht neuer Daten zu modifizieren.

Was bedeutet in diesem Beispiel nun die Abhängigkeit von der Prior Wahrscheinlichkeit? Falls das "Vorurteil" zugunsten der Nullhypothese ausfällt, etwa $P_0(H_0) = 0.8$ als Wahrscheinlichkeit für die Unwirksamkeit der Vitamingabe, wird die aposteriori Wahrscheinlichkeit zu $\approx 50\%$ (siehe Abb. 1). Hat sich diese skeptische Haltung bezüglich der Vitamin C Wirksamkeit etwa auf eine vorherige Untersuchung gestützt, ist im Licht der neuen Daten nun die Wahrscheinlichkeit wieder 50:50.

4 Das Problem der Prior Auswahl

Die Tatsache, dass in der Bayes'schen Statistik die Wahrscheinlichkeit der Hypothese *vor* der Messung eingeht, ist häufiger Anlass für Kritik. Selbst die Wahl einer konstanten Wahrscheinlichkeitsdichte im physikalischen Bereich ("Alle Werte gleichwahrscheinlich, kein Wert ausgezeichnet") ist unbefriedigend, da diese Verteilung für eine Funktion des Parameters natürlich *nicht* mehr konstant ist! Soll etwa auf die Neutrinomasse geschlossen werden, erscheint eine konstante Verteilung in m_ν sinnvoll. Oder ist die Messung vielleicht nur sensitiv auf m_ν^2? Allerdings kann der Prior auch dazu verwendet werden, Informationen über unphysikalische

Wertebereiche der gesuchten Größe in die Auswertung einzubeziehen. Im obigen Beispiel der Neutrinomassenbestimmung kann etwa die Prior–Wahrscheinlichkeit für *negative* Werte Null gesetzt werden.

Die Eigenschaft des Priors die Messung nicht zu beeinflussen ist im Konzept des sog. nicht–informativen Priors verallgemeinert worden. Jeffreys [3] argumentiert, dass für einen gesuchten Parameter θ die Wahrscheinlichkeitsdichte $1/\theta$ für den Prior ausgezeichnet ist. Das Argument beruht im wesentlichen darauf, dass $d\theta/\theta \sim d\theta^n/\theta^n$ gilt, also unter diesen speziellen Transformationen der Prior invariant ist. In [4] wird etwa als Prior in der Bestimmung des Higgsmassenlimits aus einer Anpassung an Standardmodellmessungen der Prior $1/M_H \times$ (Ausschlussgrenze direkter Higgssuche) gewählt.

Überzeugte Bayesianer haben sich jedoch dafür entschieden, aus der Not eine Tugend zu machen: Die unausweichliche Subjektivität die durch die Priorauswahl Einzug hält erklären sie zu einer Stärke des Konzeptes [5]. Gerade dadurch wird die Illusion der Objektivität statistischer Datenanalyse für den Konsumenten der Resultate zerstört. Realistischer Weise kann lediglich Intersubjektivität angestrebt werden, und die Auswahl eines Priors ist so wenig willkürlich wie andere Entscheidungen die bei der Durchführung des Experimentes getroffen werden müssen. Jeder Versuch ausgezeichnete "Standard Prior" einzuführen widerspricht ihnen zufolge dem Konzept der subjektivistischen Wahrscheinlichkeit, und läuft Gefahr, die Bayes'sche Statistik ebenso dogmatisch und mechanisch anwendbar wie die herkömmliche Statistik zu machen.

Für die praktische Anwendung ist es aber sicherlich sinnvoll, die Abhängigkeit der aposteriori Wahrscheinlichkeit von der Priorauswahl zu studieren. Falls diese sehr groß ist, kann lediglich des Bayes–Faktor sinnvoll angegeben werden. Auf diese Weise ist es dem Konsumenten der Daten überlassen, seinen Schluss zu ziehen.

5 Wann liefern beide Modelle die selben Ergebnisse?

Die Ergebnisse die in der Bayes'schen Statistik gewonnen werden unterscheiden sich in ihrer Interpretation natürlich *immer* von konventionellen Resultaten, numerisch zum Glück häufig nicht. Der Grund hierfür ist leicht einzusehen. Betrachten wir dazu noch einmal den Satz von Bayes für Wahrscheinlichkeitsdichten:

$$f(\mu \mid \overline{x}) = \frac{f(\overline{x} \mid \mu) \cdot f_0(\mu)}{\int f(\overline{x} \mid \mu) f_0(\mu) d\mu}$$

Falls die Prior Verteilung viel breiter als der Likelihood ist (geringe Vorkenntnis, bzw. hohe Datenstatistik), wird der Zähler vollkommen durch den Likelihood dominiert. Lediglich die Normierung ist betroffen. In vielen Anwendungen entscheidet jedoch das Maximum und Breite der Verteilung (etwa Maximum Likelihood Methode), sodass die Resultate identisch sind.

6 Zusammenfassung und Fazit

Die "herkömmliche" Statistik kann *keine Wahrscheinlichkeitsaussagen über das Zutreffen einer Hypothese treffen.* Der Grund hierfür liegt in ihrer Definition von Wahrscheinlichkeit als *Grenzwert relativer Häufigkeit* ("frequentistischer Wahrscheinlichkeitsbegriff"). Die Hypothesentests der herkömmlichen Statistik sprechen deshalb auch nicht etwa von einer 95% Wahrscheinlichkeit für eine Higgsmasse von mindestens so und so viel GeV, denn die Higgsmasse ist keine Zufallsvariable! Stattdessen geben sie Vertrauensintervalle ("confidence level") an. Damit ist gemeint, dass wiederholte Messungen in z.Bsp. 95% der Fälle mit der gegebenen Massenschranke konsistent sind. Als Zufallsvariable dienen die Grenzen dieses Vertrauensintervalls.

Die Bayes'sche Statistik findet diese Vorgehen unintuitiv, da schließlich im allgemeinen Verständnis Wissenschaft Aussagen über die "Dinge" und nicht nur ihre "Messbarkeit" treffen soll. Durch ihre subjektivistische Definition von Wahrscheinlichkeit erlaubt sie, Wahrscheinlichkeitsaussagen auch über Hypothesen treffen zu können. Der Preis den sie dafür zahlen muss, ist allerdings die Abhängigkeit vom sog. Prior, also der Wahrscheinlichkeit der Hypothese *vor* der Messung.

Etwas provokativ kann man formulieren:

> Die konventionelle Statistik gibt exakte Antworten auf **un**interessante Fragen. Die bayes'sche Statistik gibt **un**exakte Antworten auf interessante Fragen.

Numerische Abweichungen ergeben sich meist dann, wenn die Datenlage ohnehin zur Vorsicht rät. In zahlreichen Standardanwendungen in der Experimentalphysik bedeutet die Anwendung der Bayes'schen Statistik, numerisch identische Ergebnisse intuitiver interpretieren zu können.

Literatur

[1] T. Bayes, *An essay towards solving a problem in the doctrine of chances*, Philos. Trans. R. Soc. **53** (1763).

[2] J. O. Berger und D. A. Berry, *Statistical Analysis and the Illusion of Objectivity*, American Scientist, Vol.76 No.2 159 (1988).

[3] H. Jeffreys, *Theory of Probability*, Oxford at Clarendon, (1961).

[4] J. Erler, *Implications of Precision Elektroweak Measurements for the Standard Model Higgs Boson*, hep-ph/9904235 (1999).

[5] G. D'Agostini, *Overcoming Prior Anxiety*, ph/9906048 (1999).

[6] G. D'Agostini, *Bayesian Reasoning in High-Energy Physics: Principles and Applications*, CERN Report 99-03 (1999).