Seminararbeit:

Grundlagen der Bayes - Inferenz. Das Bertrand - Paradoxon

Autorin: Mihaela Hanea

Leitung: PD. Dr. Dr. Christina Schneider

Betreuer: Paul Fink

10. Januar 2015

Inhaltsverzeichnis

1 Grundlagen der Bayes-Inferenz

1.1 Was ist die Bayes-Inferenz?

Die Bayes-Inferenz ist eine Alternative zu den klassischen Methoden der Statistik. Die Bayes'schen Methoden sind mit dem Konzept der A-Priori-Information verbunden, d.h. dass in den Datenanalysen Informationen einbezogen werden, die vor der Erhebung der Daten vorliegen. Das Ziel der Bayes'schen Methoden ist die Kombination der beiden Informationsquellen A-Priori-Informationen und Daten aus der Stichprobe. Der Begriff 'Bayesianismus' stammt von dem englischen Mathematiker und presbyterianischen Pfarrer Thomas Bayes (1701 - 1761). Sein wichtigstes Werk , 'An essay towards solving a problem in the doctrine of chance', wurde erst posthum 1763 durch seinen Freund Richard Price in den Philosophical Transactions der Royal Society publiziert. Darin wird zum ersten Mal die 'Bedingte Wahrscheinlichkeit' eingeführt und über folgendes Problem diskutiert:

> „Gegeben die Anzahl Male, die ein unbekanntes Ereignis eingetreten und ausgeblieben ist. Gesucht die Chance, dass die Wahrscheinlichkeit seines Eintretens bei einem einzelnen Versuch irgendwo zwischen zwei angebbaren Geraden der Wahrscheinlichkeit liegt.“(Bayes 1763, S. 1)

Man kann sich also das Problem folgender Maßen vorstellen (Schmid 2014, S. 5):

- Eine Folge von 0-1-Experimenten wird gegeben (0=Misserfolg, 1=Erfolg)
- Die Erfolgswahrscheinlichkeit π ist eine unbekannte Zufallsgröße
- Bedingt auf π, sind die Experimente unabhängig
- Gesucht wird die Verteilung von $\pi|x$ (x=Anzahl der Erfolge im Experiment)

In der Bayes-Inferenz wird eine subjektive Position in Bezug auf die Interpretation der Wahrscheinlichkeiten angenommen. Deshalb wird die Wahrscheinlichkeit als Maß für die Glaubwürdigkeit einer Aussage eingeführt. Man kann also von

$$P(A) = \text{Grad der Plausibilität, dass Ereignis A eintritt}$$

sprechen (Cox 1946, S. 2).

1.2 Bayes-Formel

Die Bayes-Formel, oder auch Theorem von Bayes genannt, ist das Kernstück der Bayes-Inferenz, das die Berechnung der bedingten Wahrscheinlichkeit beschreibt. *Das Bayes Theorem für Ereignisse* lautet:

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)} \tag{1}$$

Wobei A und B zwei unterschiedliche Ereignisse sind, mit $P(A) \neq 0$ und $P(B|A)$ = die Wahrscheinlichkeit, dass B wahr ist unter der Bedingung, dass

A eintritt. Die Wahrscheinlichkeit $P(B)$ wird mit Hilfe von dem Satz der totalen Wahrscheinlichkeit ausgerechnet: $P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$, wobei $A_1, \ldots A_k$ eine disjunkte Zerlegung der Ergebnismenge Ω mit $P(B_i) \neq 0$ ist.

Nach dem selben Muster wird auch *die Bayes-Formel für Dichten* aufgebaut:

$$f_{Y|X}(y|x) = \frac{1}{f_X(x)} f_{X|Y}(x|y) f_Y(y) \tag{2}$$

Wobei X und Y zwei Zufallsgrößen auf dem Ergebnisraum Ω sind und

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y) f_Y(y) dy$$

Da der Vektor $\mathbf{x}$ feste Werte enthält, ist $f_X(x)$ konstant bzgl y. Das Bayes-Theorem kann also noch in der Form

$$f_{Y|X}(y|x) \propto f_{X|Y}(x|y) f_Y(y) \tag{3}$$

geschrieben werden, wobei $\propto$ das Proportionalitätszeichen bedeutet.

Der Satz von Bayes erlaubt in gewissem Sinn eine umgekehrte Interpretation der Schlussfolgerungen, die sogenannte 'inverse probability' (De Morgan 1837, S. 239). Man stellt aus der Gleichung (1) fest, dass die Beziehung von A und B aus $P(A|B)$ (linke Seite) zu $P(B|A)$ (rechte Seite) umgestellt wurde.

1.2.1 Beispiel zur Bayes-Formel

Es wird ein geeigneter Test durchgeführt, der über das Vorliegen einer Erkrankung entscheidet. Der Test fällt zu 99.5 % positiv aus, falls der Patient krank ist. Bei Nichtvorliegen der Krankheit beträgt die Wahrscheinlichkeit für einen positiven Test 1 %. Aus Studien schätzt man, dass eine von vier Personen an der betreffenden Krankheit leidet. Die Frage ist nun: wie groß ist die Wahrscheinlichkeit, dass jemand an der Krankheit leidet, obwohl der Test ein negatives Ergebnis zeigt?

Bezeichnungen:
K = untersuchte Person ist krank
T = Test fällt positiv aus

Gegeben werden also
$P(T|K) = 0.995$
$P(T|\bar{K}) = 0.01$
$P(K) = 0.25$
Gesucht wird die Wahrscheinlichkeit $P(K|\bar{T})$

$$P(K|\bar{T}) \overset{(1)}{=} \frac{P(\bar{T}|K)P(K)}{P(\bar{T})} = \frac{P(\bar{T}|K)P(K)}{P(\bar{T}|K)P(K) + P(\bar{T}|\bar{K})P(\bar{K})}$$
$$= \frac{[1 - P(T|K)]P(K)}{[1 - P(T|K)]P(K) + [1 - P(T|\bar{K})][1 - P(K)]}$$
$$= \frac{(1 - 0.995)0.25}{(1 - 0.995)0.25 + (1 - 0.01)0.75}$$
$$= 0.00185$$

Die Wahrscheinlichkeit, dass die untersuchte Person krank ist, obwohl der Test negativ ausfällt, beträgt 0.185%. Die gesuchte Wahrscheinlichkeit ist sehr niedrig, was bestätigt, dass der durchgeführte Test, präzise Ergebnisse ausgibt.

1.3 Bayes-Postulaten

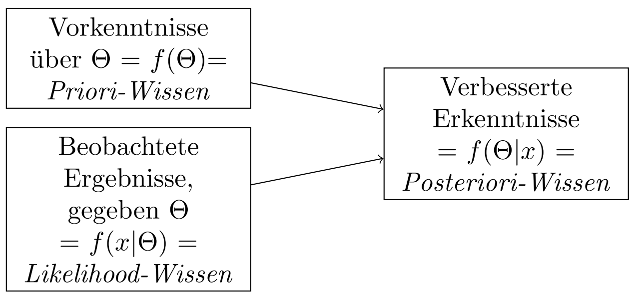

Abbildung 1: Vorgehensweise der Bayes-Inferenz (Mock 1995, S. 59)

Die Abbildung 1 zeigt den Zusammenhang der drei Begriffe: Priori, Likelihood und Posteriori. Nach Rüger (1999, S. 186) werden sie mit drei Postulaten motiviert.

1.3.1 Postulat 1 (Priori-Wissen)

Nach dem ersten Postulat wird eine Annahme über die Verteilung des Parameters Θ getroffen, die als A-Priori-Dichte $f(\Theta)$ bezeichnet wird. Zu diesem Zeitpunkt interessiert man sich darüber, wie gut sich die Parameterwerte an das Vorwissen anpassen. Die Frage ist aber, wie wählt man die Prioris? Bei der Wahl der Priori-Verteilung gibt es mehrere Möglichkeiten, wie zum Beispiel: konjugierte A-Priori-Verteilung, subjektive Prioris, Jeffreys-Priori, Referenz-Priori, Prioris auf Prioris, Regularisierungsprioris, Monte-Carlo-Simulation, u.a.

Am häufigsten benutzt man *die konjugierte A-Priori-Verteilung*, d.h. dass die A-Priori-Verteilung zu der gleichen Verteilungsfamilie wie die Posteriori-Verteilung gehört. Ein Beispiel hierfür ist das Beta-Binomial-Modell, welches später genauer im Abschnitt 1.4 vorgelegt wird. Für

$$X \sim Bin(n, \pi)$$

wird die konjugierte A-Priori-Verteilung $Beta(\alpha, \beta)$ angenommen. Unter diese Voraussetzungen ist die A-Posteriori-Verteilung auch eine

$$Beta(x + \alpha, n - x + \beta)\text{-Verteilung}$$

Weiterhin kann man *die Jeffreys-Priori* als A-Priori-Verteilung annehmen. Sie ist eine objektive Verteilung, die als

$$f(\Theta) \propto \sqrt{detI(\Theta)}$$

berechnet wird, wobei $I(\Theta) = -\mathbb{E}_\Theta \left[\frac{\partial^2 l(X, \Theta)}{\partial \Theta^2} \right]$ die Fischer-Information ist, mit $l(X, \Theta)$ als logarithmiertes Likelihood. Also zum Beispiel die Jeffreys-Priori für $X \sim Bin(n, \pi)$ ist

$$f(\pi) \propto \sqrt{\frac{1}{\pi(1 - \pi)}}$$

Der Vorteil der Jeffreys-Priori ist die Invarianz bezüglich Anzahl und Transformationen der Parameter. Der Nachteil liegt darin, dass die Jeffreys-Priori das starke Likelihood-Prinzip verletzt, d.h. dass nicht alle Beobachtungen unterschiedlicher Modelle zu identischen Schlüssen führen, wenn die Likelihood gleich ist.

Ein subjektivistischer Bayesianer beschreibt die A-Priori-Verteilung durch *'Expertenwissen'*. Bei einem Münzwurf z.B. kann man davon ausgehen, dass 'Kopf' und 'Zahl' gleich wahrscheinlich sind, falls die Münze fair ist. Somit wird angenommen, dass die Priori-Dichte normalverteilt auf dem Intervall $[0, 1]$ mit bekannten $\mu = 0.5$ ist.

Die Wahl von Prioris ist meist ein Kritikpunkt der Frequentisten und gilt als unwissenschaftlich. Z.B Gelman (2008, S. 447) fragt sich: „Where do prior distributions come from, anyway? I don't trust them ". Und Little (2006, S. 218) stützt die Idee, dass die Bayes-Inferenz zu viele Antworten liefert: „once the likelihood is nailed down, every prior distribution leads to a different answer ".

1.3.2 Postulat 2 (Likelihood-Wissen)

Nach dem zweiten Postulat wird die Wirkung des Parameters auf die Stichprobe bestimmt, basierend auf der bedingten Wahrscheinlichkeit

$$f(x|\Theta) = f(\text{Zufallsvariablen}|\text{fixer Parameter})$$

Die Likelihood-Funktion nimmt nur die in den Daten erhaltenen Informationen wahr (Saint-Mont 2011, S. 267), d.h. dass sie auf dem *Likelihood-Prinzip* aufgebaut wird. Nach Birnbaum (1962, S. 271), ist das Likelihood-Prinzip äquivalet zum *Konditionalprinzip* (alles was tatsächlich passiert ist, muss auch beobachtet werden) und *Suffizienzprinzip* (alle für die Fragestellung relevanten Daten werden beobachtet). Das bedeutet weiterhin, dass „man logisch gezwungen ist, nur die Likelihood-Funktion als einzige Quelle der Evidenz zu akzeptieren "(Saint-Mont 2011, S. 267).

Die A-Priori-Verteilung $f(\Theta)$ hat keinen Einfluss auf die Likelihood-Funktion

(Bjørnstad 1996, S. 797), die Art der Erhebung der Daten beeinflusst aber die Auswertung und Interpretation der Likelihood (siehe Beispiel aus dem Abschnitt 1.4).

1.3.3 Postulat 3 (Posteriori-Wissen)

Nach dem dritten Postulat wird, basierend auf den ersten zwei Postulaten, die A-Posteriori-Verteilung $f(\Theta|x)$ festgelegt. Wie auch der Namen sagt, ist die Posteriori-Verteilung eine Verteilung, die nach der Ausführung des Experiments aufgebaut wird. Sie ist eine aktualisierte Version der A-Priori-Verteilung durch die Beobachtungen aus der Stichprobe (Rüger 1999, S. 188). Damit ist die A-Posteriori-Verteilung die Grundlage zur Berechnung aller Punktschätzer und Konfidenzintervalle in der bayesianischen Statistik.

1.4 Beispiel zur Bayes-Inferenz

Folgendes Beispiel ist eine Erweiterung des Beispiels aus dem Abschnitt 1.2.1 und wird die Vorgehensweise der Bayes-Inferenz durch die obengenannten Postulaten verfolgen.

Sei nun $x_1, \ldots, x_n$ eine iid-Stichprobe von $X \sim Bin(n, \pi)$ mit $n = 100$ (Anzahl der untersuchten Patienten) und $\pi = 0.25$ (Wahrscheinlichkeit, mit der die Krankheit in der Gesamtpopulation auftritt).

Der erste Schritt der Bayes-Inferenz ist nach dem ersten Postulat die Bestimmung der A-Priori-Verteilung. In diesem Beispiel wird für π die konjugierte Verteilung $Beta(\alpha, \beta)$ mit Dichte

$$f(\pi) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \pi^{\alpha-1}(1-\pi)^{\beta-1}$$

angenommen, wobei α und β bekannt sind:

$$(\alpha, \beta) \in \{(0.3, 0.5), (5, 1), (1, 6), (2, 2), (6, 2), (2, 10)\}$$

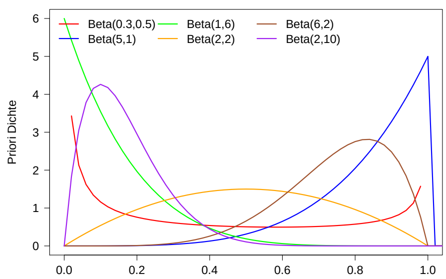

Abbildung 2: $Beta(\alpha, \beta)$ - Verteilungen als A-Priori-Verteilungen

Wie das zweite Postulat betont, muss im nächsten Schritt die Likelihood-Funktion bestimmt werden. Sie stimmt in diesem Fall mit der Dichte der Binomialverteilung (für eine Beobachtung) überein:

$$f(x|\pi) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$$

Mit Hilfe der A-Priori-Verteilung und der Likelihood-Funktion, wird nun nach dem dritten Postulat die A-Posteriori-Verteilung, wie folgt, ausgerechnet:

$$\begin{aligned}
f(\pi|x) &\overset{(3)}{\propto} f(x|\pi)f(\pi) \\
&\overset{iid}{\propto} \prod_{i=1}^{n}\binom{n}{x}\pi^{x_i}(1-\pi)^{n-x_i}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\pi^{\alpha-1}(1-\pi)^{\beta-1} \\
&\propto \underbrace{\binom{n}{x}^n}_{konstant}\pi^{\sum_{i=1}^n x_i}(1-\pi)^{n-\sum_{i=1}^n x_i}\underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{konstant}\pi^{\alpha-1}(1-\pi)^{\beta-1} \\
&\propto \pi^{\sum_{i=1}^n x_i+\alpha-1}(1-\pi)^{n-\sum_{i=1}^n x_i+\beta-1}
\end{aligned}$$

$\Rightarrow \pi|x \sim Beta(\sum_{i=1}^n x_i+\alpha, n-\sum_{i=1}^n x_i+\beta)$ Die Posteriori-Verteilung ist wieder eine Beta-Verteilung (außer Normierungskonstante), das heißt

$$\pi|x \sim Beta(\sum_{i=1}^{n} x_i + \alpha, n - \sum_{i=1}^{n} x_i + \beta)$$

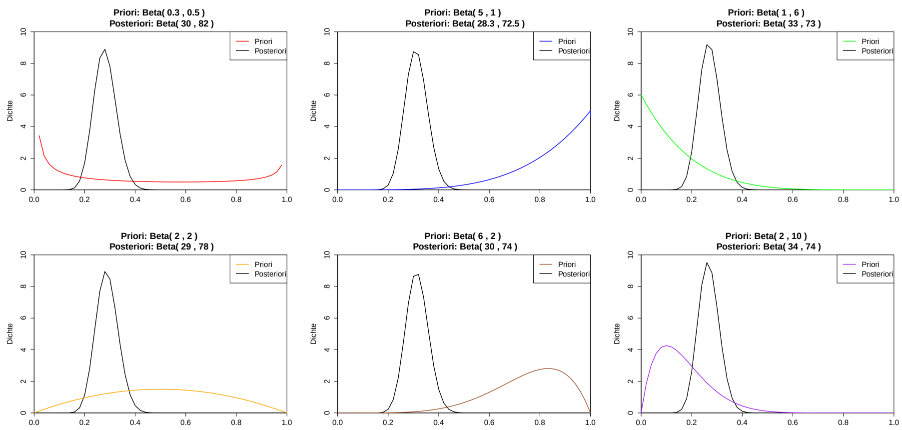

Abbildung 3: Priori und Posteriori Verteilungen nach der Beobachtung von 28 Erfolgen unter n = 100 Versuchen

Die Abbildung 3 stellt beispielhaft dar, wie sich die A-Posteriori-Verteilung eines Parameters von der A-Priori-Verteilung unterscheidet. In allen sechs Fällen haben sich die Varianzen verringert, da sich die Unsicherheit durch die neuen Informationen verkleinert hat. Jedoch hat die Priori in diesem Beispiel keinen großen Einfluss auf die Posteriori-Verteilung. Obwohl die Priori-Dichten sehr unterschiedlich ablaufen, kommen bei der Posterioris keine dramatischen Änderungen vor.

Was eine große Einwirkung auf die Posteriori hat, ist der Umfang der Stichprobe. Dazu wurde eine A-Priori-Verteilung des Parameters $\pi \sim Beta(2, 10)$ und $n = (1, 50, 100, 500, 1000, 10000)$ angenommen. Folgende Abbildung zeigt, dass für wachsendes n, die Posteriori eine immer kleinere Varianz bekommt, also einen höheren Peak, und wird rund um den wahren Wert $\pi = 0.25$ maximal.

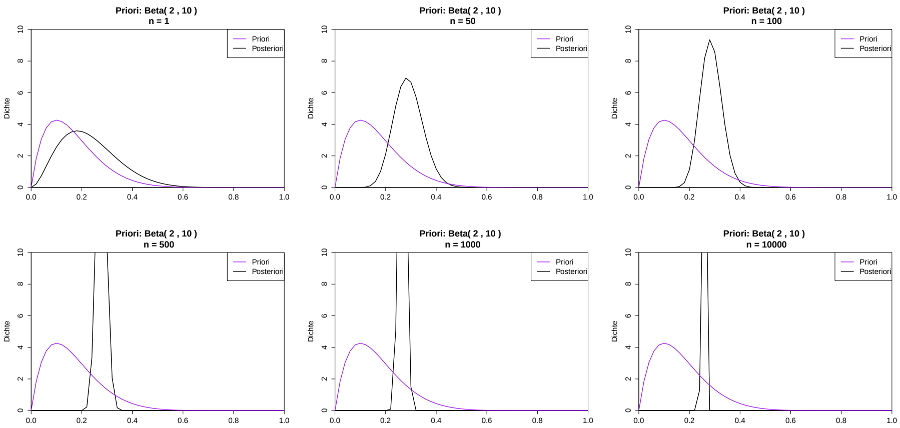

Abbildung 4: Priori und Posteriori Verteilungen

1.5 Vergleich zwischen Frequentisten und Bayesianern

Die folgende Tabelle fasst die Unterschiede zwischen die frequentistische und bayesianische Denkrichtung zusammen.

	Frequentisten	Bayesianer
Begründer	Neymann, Pearson, Wald	Bayes, Laplace, de Finetti
Interpretation $P(A)$	Limes relativer Häufigkeit des Ereignisses A (Gelman 2008, S. 467)	Grad der Überzeugung, dass Ereignis A auftritt (Cox 1946, S. 2)
Daten	wiederholbare, unabhängige Zufallsexperimente	fixe, beobachtete Stichprobe
Parameter	bleibt konstant während der Prozess	ist unbekannt und wird als Zufallsvariable behandelt (Lindley 1975, S. 106)

Selbst wenn die Bayes-Statistik mehr Genauigkeit und Verlässlichkeit ausliefert, setzen sich die frequentistischen Methoden in der Statistik als prägende Techniken gegenüber bayesianischen durch, da bayesianische Verfahren oft rechnerisch aufwändig sind.

Bei der Modellierung des Vorwissens von Parametern kann das Problem auftreten, dass kein Vorwissen vorhanden ist. Im nächsten Kapitel wird gezeigt, dass unterschiedliche Behandlungen des Nichtwissens zu unterschiedlichen Lösungen führen können. Dazu wird beispielhaft das Bertrandsche Paradoxon vorgestellt.

2 Das Bertrand-Paradoxon

In der Theorie der geometrischen Wahrscheinlichkeit sind die Zufallselemente geometrische Objekte wie Punkte, Linien und Rotationen. Weil die Interpretation der Wahrscheinlichkeit unter solchen Umständen schwer ist, wurde eine Reihe von Paradoxa entwickelt (Marinoff 1994, S. 1). Das berühmteste davon ist das Bertrand-Paradoxon, das erstmals im Jahre 1889 von dem Mathematiker Joseph Bertrand in seinem Buch 'Calcul des probabilités' veranschaulicht wurde. Bertrand (1889, S. 5) hat das Problem folgendermaßen formuliert: Es wird ein Kreis mit Radius 1 betrachtet. Die Seite eines darin eingetragenen gleichseitigen Dreiecks y hat eine Länge von $\sqrt{3}$. Wie groß ist die Wahrscheinlichkeit dafür, dass die Länge Y einer zufällig ausgewählten Sehne im Kreis die Länge der Seite des eingeschriebenen Dreiecks übertrifft? Gesucht wird hier also die Wahrscheinlichkeit $P(Y > \sqrt{3})$. Bertrand gab drei Möglichkeiten an, das Problem zu lösen, die alle gültig scheinen, aber zu unterschiedlichen Ergebnissen für $P(Y > \sqrt{3})$ führen.

2.1 Bertrands Lösungen des Wahrscheinlichkeitsproblems

Die kommenden drei Ansätze im Bertrandschen Paradoxon werden auf Basis einer Gleichverteilungsannahme konstruiert, die aus *dem Prinzip vom unzureichenden Grund* ergibt. Dies wird von Carnap und Stegmüller (1959, S. 3) folgendermaßen formuliert: „Wenn keine Gründe dafür bekannt sind, um eines von verschiedenen möglichen Ereignissen zu begünstigen, dann sind die Ereignisse als gleich wahrscheinlich anzusehen.“Wir werden sehen, dass die Gleichwahrscheinlichkeitsannahmen zu je einer anderen Antwort auf die gestellte Frage führen.

2.1.1 Die 'zufällige Endpunkte' - Methode

Es werden zufällig zwei Punkte auf dem Kreisumfang gewählt, welche zu einer Sehne verbunden werden. Das gleichseitige Dreieck wird so gedreht, dass einer der Endpunkte und ein Eckpunkt des Dreiecks übereinstimmen. Dieser Punkt wird also als fester Punkt $P1$ fixiert. Es werden nur die Sehnen betrachtet, die von Punkt $P1$ ausgehen. Die Sehnen, deren Endpunkt $P2$ auf dem Segment des Kreisumfangs zwischen den anderen beiden Eckpunkten des Dreiecks liegen, sind länger als die Dreiecksseite. Die Länge dieses Segments beträgt 1/3 des Kreisumfangs, darum ist die Wahrscheinlichkeit $P(Y > \sqrt{3}) = 1/3$.

α: Öffnungswinkel zwischen $P1$ und $P2$

$$\begin{aligned} \text{Das Ereigniss } 'Y \leq y' \text{ trifft zu} &\Leftrightarrow Y = 2\sin(\alpha/2) \leq y \\ &\Leftrightarrow \alpha \leq 2\arcsin(y/2) \end{aligned}$$

Angenommen $P1$ und $P2$ sind gleichverteilt, daraus folgt α ist gleichverteilt auf $[0;\pi]$. $\Rightarrow P(Y \leq y) = \frac{2}{\pi}\arcsin(y/2)$
$\Rightarrow P(Y > \sqrt{3}) = 1 - \frac{2}{\pi}\arcsin(\sqrt{3}/2) = 1/3$

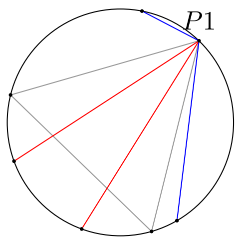

Abbildung 5: Die 'zufällige Endpunkte' - Methode
● = länger als Dreiecksseite
● = kürzer als Dreiecksseite

2.1.2 Die 'zufälliger Radius' - Methode

Es werden ein Radius (die Strecke von Kreismittelpunkt zu Kreisumfang) und ein zufälliger Punkt auf dem Radius gewählt. Es wird durch den Punkt eine zum Radius orthogonale Sehne gezogen. Das Dreieck wird jetzt so gedreht, dass eine Seite auch orthogonal zum Radius steht. Die Sehne ist länger als die Dreiecksseite, falls der zufällig ausgewählte Punkt näher am Mittelpunkt liegt, als der Schnittpunkt der Dreiecksseite mit dem Radius. Da die Dreiecksseite den Radius halbiert, ist die Wahrscheinlichkeit $P(Y > \sqrt{3}) = 1/2$.

r: Abstand zwischen Kreismittelpunkt und Sehnenmittelpunkt

Das Ereigniss $'Y \leq y'$ trifft zu $\Leftrightarrow Y = 2\sqrt{1-r^2} \leq y$

$$\Leftrightarrow r \geq \sqrt{1 - y^2/4}$$

Angenommen die Punkte auf r sind gleichverteilt.
$\Rightarrow P(Y \leq y) = 1 - \sqrt{1 - y^2/4}$
$\Rightarrow P(Y > \sqrt{3}) = 1 - (1 - \sqrt{1/4}) = 1/2$

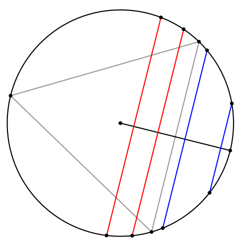

Abbildung 6: Die 'zufälliger Radius' - Methode
● = länger als Dreiecksseite
● = kürzer als Dreiecksseite

2.1.3 Die 'zufälliger Mittelpunkt' - Methode

Ein Kreis mit Hälfte des Radius und gleichem Mittelpunkt M wie der Kreis wird erzeugt. Ein zufälliger Punkt im Inneren des größeren Kreises wird gewählt und die Sehne mit diesem Punkt als Mittelpunkt konstruiert. Die Sehne ist länger als die Dreiecksseite, falls der Punkt innerhalb des kleineren Kreises liegt. Die Fläche des inneren Kreises beträgt ein Viertel der Fläche des großen Kreises, also die Wahrscheinlichkeit $P(Y > \sqrt{3}) = 1/4$.

ρ: Radius des inneren Kreises
f: Fläche des inneren Kreises

$$\begin{aligned} \text{Das Ereigniss } 'Y \leq y' \text{ trifft zu} &\Leftrightarrow Y = 2\sqrt{1-r^2} \leq y \\ &\Leftrightarrow r^2 \geq 1 - y^2/4 \\ &\Rightarrow \text{M liegt außerhalb des inneren Kreises} \end{aligned}$$

Gegeben $\rho = \sqrt{1 - y^2/4}$ und angenommen M ist gleichverteilt, folgt $f = 1 - \rho^2 = y^2/4$.
$\Rightarrow P(Y \leq y) = y^2/4$
$\Rightarrow P(Y > \sqrt{3}) = 1 - 3/4 = 1/4$

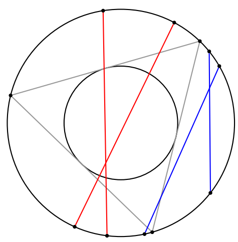

Abbildung 7: Die 'zufälliger Mittelpunkt' - Methode
● = länger als Dreiecksseite
● = kürzer als Dreiecksseite

2.2 Interpretation des Bertrand-Paradoxon

Man stellt sich jetzt die Frage: „Unter diesen drei Lösungen, welche ist die richtige Antwort?“. Bertrand's (1889, S. 5) Antwort lautet: keine der drei ist korrekt und keine der drei ist inkorrekt. Alle Versionen von Bertrands Problem sind lösbar, und jede Lösung basiert auf den selben Verfahren, nämlich auf dem *Prinzip der vom unzureichenden Grund.*

Offensichtlich ist es nicht glaubwürdig zu behaupten, dass das Problem keine eindeutige Lösung hat. Darum haben viele Mathematiker und Philosophen versucht, eine klare Antwort zu finden. Die wichtigsten Versuche sind die Lösungen von Marinoff und Jaynes, die sehr unterschiedliche Strategien anwenden. Wir werden aber sehen, dass beide Strategien scheitern und das Paradoxon ungelöst bleibt.

2.2.1 Marinoff's Lösung des Problems

Marinoff's Behauptung ist, dass das Paradoxon ein Problem darstellt, dessen Identität unbestimmt ist. Er nimmt an, dass das Problem nur dann eine Lösung hat, wenn die Methode spezifiziert ist, mit der eine Sehne zufällig gewählt wird. Zum Beispiel kann die Sehne zufällig im Kreis gezeichnet werden oder es können zwei Punkte zufällig auf dem Kreisumfang gewählt werden und zu einer Sehne verbunden werden oder es kann ein Punkt innerhalb des Kreises gewählt werden und eine Sehne, mit dem Punkt als Mittelpunkt, gezeichnet werden. Nach Marinoff's Einsicht, gibt es keinen Grund eine Methode der anderen zu vorziehen, in Abwesenheit von weiteren Information.

> „There exists a multiplicity, if not an infinite number, of procedures for generating random chords of a circle. The answers that one finds to Bertrand's generic question Q vary according to the way in which the question is interpreted"(Marinoff 1994, S. 17).

Das Antwort kann nicht eine Lösung des Problems sein, weil es das Indifferenzprinzip ablehnt. Wie auch van Fraassen behauptet, würden wir das Indifferenzprinzip gar nicht benötigen, falls wir von Anfang an wissen, welcher Parameter gleichverteilt sein soll („In other words, this response rejects the Principle of Indifference altogether. After all, if we were told as part of the problem which parameter should receive a uniform distribution, no such Principle would be needed"(van Fraassen 1989, S. 305)).

2.2.2 Jaynes's Lösung des Problems

Jaynes weist darauf hin, dass Bertrand's Problem sich nicht mit einem bestimmten Kreis beschäftigt, also jede zugelassene Wahrscheinlichkeit darf nicht von der Lage und Umfang des Kreises abhängen. Darum motiviert er die Symmetrieanforderungen gemäß folgender Problemformulierung: Ein langer Strohhalm wird zufällig auf einen Kreis geworfen. Gegeben, dass er den Kreis schneidet, wie ist die Wahrscheinlichkeit dafür, dass die Sehne die Länge der Seite eines eingeschriebenen Dreiecks übertrifft? (Jaynes 1973, S. 477) Die einzige Methode, die sowohl skaleninvariant (hängt nicht von Skalierung ab), als auch translationsinvariant (Volumen des Kreises ändert sich durch eine Verschiebung im Raum nicht) ist, ist *die 'zufälliger Radius' - Methode* (der Kreis in *der 'zufälliger Mittelpunkt' - Methode* ist nur skaleninvariant und in *der 'zufällige Endpunkte' - Methode* ist weder skaleninvariant, noch translationsinvariant). Darum wird die zweite Methode mit der Lösung 1/2 von Janeys als wahr angenommen.

> „Neither Bertrand's original statement nor our restatement in terms of straws specified the exact size of the circle, or its exact location. If, therefore, the problem is to have any definite solution at all, it must be 'indifferent' to these circumstances; *i.e.*, it must be unchanged by a small change in the size or position of the circle"(Jaynes 1973, S. 480).

Das Problem, das Jaynes vorstellt, ist nicht gleich mit Bertrand's Problem, sondern nur eine Beschränkung davon (Shakel 2007, S. 172). Darum kann man behaupten, dass das generelle Problem nicht gelöst wurde, sondern nur 'die Strohhalme-werfen Variante'. Es stellt sich also heraus, dass Jaynes's Annahme „Es kommen keine Änderungen des Problems vor, falls wir Strohhalme werfen"(„we do no violence to the problem if we suppose we are tossing straws"(Jaynes 1973, S. 478)) falsch ist.

2.3 Das Bertrand-Paradoxon und die Bayes-Inferenz

Es wurde in den Abschnitten 2.1 und 2.2 bereits gezeigt, dass das Paradoxon keine eindeutige Antwort hat. Das heißt, dass im Fall des Paradoxons die Wahrscheinlichkeiten nicht wohldefiniert sind und der Parameterraum unendlich viele mögliche Parameterwerte hat. Weil wenig über den Parameter bekannt ist, möchte man den Einfluss der Priori auf die Posteriori so gering wie möglich halten. Die Bayes-Inferenz betreibt die Posterori-Verteilung trotz fehlendem Vorwissen, nämlich durch den sogenannten *nichtinformativen Priori-Verteilungen.*

Zur Erstellung einer nichtinformativen Priori-Verteilung stehen unterschiedliche Motivationsansätze zur Verfügung. Im Fall von dem Bertrandschen Paradoxon ist das 'Prinzip vom unzureichenden Grund' zu beachten. Dies besagt nach Neumaier (2015, S. 3), dass kein Grund dazu bestehe, einen Parameterwert gegenüber einem anderen als wahrscheinlicher zu betrachten, falls von einer vollkommenen Unkenntnis über den interessierenden Parameter Θ ausgegangen wird. Mit anderen Worten lässt sich eine Annahme über die Wahl einer bestimmten Sehne genau so rechtfertigen wie die Wahl einer beliebigen anderen Annahme (Walter 2006, S. 5). Das Prinzip weist aber Schwachpunkte auf (invariant gegenüber eineindeutigen Transformationen; nicht widerspruchsfrei anwendbar wenn Θ aus unendlich vielen Elementen besteht) und darum ist es nicht genügend, nichtinformativen Verteilungen zu ermitteln (Neumaier 2015, S. 3-4)).

Im Fall von unendlich vielen möglichen Parameterwerten sind auch Prioris akzeptiert, die nicht mehr integrierbar sind, also deren Normierungskonstante nicht bestimmt werden kann, wenn die damit gewonnene A-Posteriori-Verteilung existiert (Robert 2001, S. 27). Eine solche A-Priori-Verteilung heißt *uneigentliche A-Priori-Verteilung.* Das Bayes-Prinzip mit den drei Postulaten muss jetzt auf uneigentliche A-Priori-Verteilungen erweitert werden (Rüger 1999, S. 216). Obwohl eine uneigentliche A-Priori-Verteilung auf Θ (mit einer λ-Dichte $p(\Theta)$) kein gemeinsames stochastisches Modell $p(x;\Theta)$ definiert, ist die Posteriori-Verteilung eine eigentliche Verteilung mit der Dichte

$$p(\Theta|x) = c(x)f(x|\Theta)p(\Theta) \tag{4}$$

und Normierungsfaktor $c(x) = [\int_\Theta f(x|\Theta)p(\Theta)d\Theta]^{-1}$ wenn und nur wenn für jede Beobachtung x die Ungleichung

$$\int_\Theta f(x|\Theta)p(\Theta)d\lambda < \infty \tag{5}$$

gilt.

Die Beziehung (4) stimmt mit dem Bayes-Theorem für uneigentliche A-Priori-Verteilungen überein. Wenn man sie mit dem Bayes-Theorem für eigentliche Prioris vergleicht, merkt man keinen Unterschied zwischen den beiden. Unter der Bedingung (5) wird die Posteriori-Verteilung im Fall von uneigentlichen Priori genau so umgerechnet wie im Fall von eigentlichen Priori (Rüger 1999, S. 217).

3 Zusammenfassung

Als Zusammenfassung stelle ich die Vor- und Nachteile der Bayes-Inferenz vor. Die Vorteile sind folgende:

- Bayes-Inferenz kombiniert Priori-Information mit die Beobachtung. Mann kann frühere Informationen über den Parameter benutzen, um eine Priori-Verteilung zu bilden. Wenn neue Beobachtungen zur Verfügung stehen, kann man die bisherige Posteriori-Verteilung als Priori verwenden.
- Bayes-Inferenz ist genau und zuverlässig, ohne auf asymptotische Approximation zu greifen. Kleine Stichproben werden genauso wie große behandelt.
- Die Bayes-Inferenz verfolgt das Likelihood-Prinzip. Das heißt, dass falls zwei Stichprobendesigns proportionale Likelihood-Funktionen ausliefern, dann sind alle Folgerungen der Stichproben identisch.
- es ist möglich auch ohne oder mit wenig Information über den Parameter die Posteriori-Verteilung zu berechnen. Als Beispiel wurde das Bertrand-Paradoxon gezeigt.

Die Nachteile sind folgende:

- Es gibt keine konkrete Methode, um eine A-Priori-Verteilung zu wählen. Falls diese aber nicht korrekt gewählt wird, kann man zu falschen Schlüssen und Interpretationen gelangen.
- Die Bayes-Inferenz ist meist rechnerisch aufwändig, besonders bei Verteilungen mit mehreren Parametern.

Obwohl die Bayes-Inferenz auch Nachteile aufweist, ist sie eine sinvolle Alternative zu die klassischen Methoden der Statistik.

Literatur

Bayes, T. (1763). An Essay towards solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society of London* **53**: S. 370–418.

Bertrand, J. L. F. (1889). *Calcul des probabilités*, Gauthier-Villars, Paris.

Birnbaum, A. (1962). On the Foundations of Statistical Inference, *Journal of the American Statistical Association* **57**(298): S. 269–306.

Bjørnstad, J. F. (1996). On the Generalization of the Likelihood Function and the Likelihood Principle, *Journal of the American Statistical Association* **91**(434): S. 791–806.

Carnap, R. und Stegmüller, W. (1959). *Induktive Logik und Wahrscheinlichkeit*, Springer-Verlag, Wien.

Cox, R. T. (1946). Probability, Frequency and Reasonable Expectation, *American Journal of Physics* **14**(1): S. 1–13.

De Morgan, A. (1837). Review of Laplace's Théorie Analytique des Probabilités, *Dublin Review* **2**(3): S. 338–354, 237–248.

Dennis, B. (1996). Discussion: Should Ecologists Become Bayesians?, *Ecological Applications* **6**(4): S. 1095–1103.

Efron, B. (1986). Why Isn't Everyone a Bayesian?, *The American Statistician* **40**(1): S. 1–5.

Efron, B. (2013). Bayes Theorem in the 21st Century, *Science* **340**: S. 1177–1178.

Fienberg, S. E. (2006). When Did Bayesian Inference Become „Bayesian"?, *Bayesian Analysis* **1**(1): S. 1–40.

Gelman, A. (2008). Objections to Bayesian statistics, *Bayesian Analysis* **3**(3): S. 445–450.

Jaynes, E. T. (1973). The Well-Posed Problem, *Foundations of Physics* **3**: S. 477–493.

Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*, Cambridge University Press, Cambridge.

Lindley, D. V. (1975). The Future of Statistics - a Bayesian 21st Century, *Suppl. Adv. Appl. Prob.* **7**: S. 106–115.

Little, R. (2006). Calibrated Bayes: A Bayes/Frequentist Roadmap, *The American Statistician* **60**(3): S. 213–223.

Marinoff, L. (1994). A Resolution of Bertrand's Paradox, *Philosophy of Science* **61**(1): S. 1–24.

Mock, R. (1995). *Methoden zur Datenhandhabung in Zuverlässigkeitsanalysen*, vdf Hochschulverlag AG an der ETH Zürich, Zürich.

Neumaier, A. (2015). Konzepte einer nichtinformativen Priori. Seminararbeit WS 2014/2015, Institut für Statistik.

Robert, C. P. (2001). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*, Springer-Verlag, New York.

Rüger, B. (1999). *Test- und Schätztheorie*, Band I: Grundlagen, Oldenbourg Wissenschaftsverlag, Wien.

Saint-Mont, U. (2011). *Statistik im Forschungsprozess: Eine Philosophie der Statistik dls Baustein einer integrativen Wissenschaftstheorie*, Springer-Verlag, Heidelberg.

Schmid, V. (2014). Einführung in die Bayes-Statistik. Vorlesung SS 2014, Institut für Statistik.

Shackel, N. (2007). Bertrands Paradox and the Principle of Indifference, *Philosophy of Science* **74**: S. 150–175.

van Fraassen, B. C. (1989). *Laws and Symmetry*, Clarendon, Oxford.

Walter, G. (2006). Robuste Bayes-Regression mit Mengen vonPrioris - Ein Beitrag zur Statistik unter komplexer Unsicherheit. Diplomarbeit, Institut für Statistik.