



HOCHSCHULE FÜR TECHNIK UND WIRTSCHAFT

DOKUMENTATION

Projektseminar
Optimierung und
Unsicherheitsquantifizierung mit
Bayesianischer Statistik und
MCMC-Methoden
(Prof. Schwarzenberger)

Clemens Näther, s85426

Jakub Kliemann, s85515

Contents

1	Einleitung	2
2	Theoretischer Teil	3
2.1	Grundlagen der bayesianischen Statistik und das Bayes'sche Theorem .	3
2.1.1	Einführung in die bayesianische Statistik	3
2.1.2	Das Bayes'sche Theorem und seine Bestandteile	3
2.1.3	Beispiele und praktische Anwendungen	5
2.1.4	Punktschätzer, Konfidenzintervalle und Hypothesenprüfung in der bayesianischen Statistik	6
2.2	Markov Chain Monte Carlo (MCMC) Methoden	9
2.2.1	Einführung in MCMC-Methoden	9
2.2.2	Der Metropolis-Hastings-Algorithmus	10
2.2.3	Konvergenzkriterien für MCMC-Simulationen	10
3	Praktischer Teil	11
3.1	Implementierung des Metropolis-Hastings-Algorithmus in Python . . .	11
3.2	Anwendung der bayesianischen Statistik am Beispiel	12
3.2.1	Überblick und Aufbereitung des Datensatzes	13
3.2.2	Anwendung der bayesianischen Statistik	14
3.2.3	Rekursives Einbeziehen von neuen Daten	15
3.2.4	Vergleich mit traditioneller Statistik	16
3.3	Vergleich mit klassischen Methoden am Beispiel	16
3.3.1	Einleitung	16
3.3.2	Datensatz	16
3.3.3	Berechnung mit klassischer Statistik	16
3.3.4	Berechnung mit bayesianischer Statistik	17
3.3.5	Vergleich der Ergebnisse	19
3.3.6	Vergleich der klassischen und bayesianischen Statistik	20
3.4	Vergleich anhand gewählter Datensätze - Fussball WM	21
3.4.1	Berechnung mit klassischer Statistik	22
3.4.2	Berechnung mit bayesianischer Statistik	23
4	Zusammenfassung und Ausblick	26
5	Literaturverzeichnis	27
6	Selbstständigkeitserklärung	28

1 Einleitung

Die Statistik ist ein wesentlicher Bestandteil der Wissenschaft. Sie ermöglicht es, aus Daten Informationen zu gewinnen und aus diesen Schlussfolgerungen zu ziehen. Es existieren dabei aber zwei verschiedene Ansätze, die klassische und die bayesianische Statistik. Beide bieten unterschiedliche Perspektiven und Werkzeuge zur Analyse und Interpretation von Unsicherheiten, und die Entscheidung für eine Methode kann erhebliche Auswirkungen auf die Schlussfolgerungen einer Studie haben.

In dieser Arbeit liegt der Fokus auf der bayesianischen Statistik, einem Ansatz, der die Wahrscheinlichkeit als subjektives Maß für die Unsicherheit interpretiert und es ermöglicht, Vorwissen systematisch in den Analyseprozess einzubringen. Der Vergleich mit der klassischen Statistik, die auf der relativen Häufigkeit basiert, zeigt, wie unterschiedlich die Ansätze sowohl in der Theorie als auch in der Praxis sind.

Ein zentraler Schwerpunkt der Arbeit liegt darin die theoretischen Grundlagen der bayesianischen Statistik zu erläutern und den Prinzipien der klassischen Statistik gegenüberzustellen. Dabei soll insbesondere auf die Markov Chain Monte Carlo (MCMC) Methoden eingegangen werden, die es ermöglichen, komplexe Modelle zu schätzen und zu simulieren. Insbesondere wird der Metropolis-Hastings-Algorithmus, einer der bekanntesten Algorithmen zur Erstellung von Markovketten vorgestellt.

Anschließend wird im zweiten Teil der Arbeit die praktische Anwendung der bayesianischen Statistik und der MCMC-Methoden anhand von Beispielen und Datensätzen in Python demonstriert. Dabei wird gezeigt, wie bayesianische Modelle implementiert und auf verschiedene Datensätze angewendet werden können. Die Ergebnisse werden interpretiert und mit klassischen Methoden verglichen.

2 Theoretischer Teil

2.1 Grundlagen der bayesianischen Statistik und das Bayes'sche Theorem

2.1.1 Einführung in die bayesianische Statistik

Die bayesianische Statistik ist ein Zweig der Statistik. Sie unterscheidet sich im wesentlichen in der Interpretation der Wahrscheinlichkeit von der klassischen Statistik. Die klassische Statistik definiert die Wahrscheinlichkeit als die **relative Häufigkeit** in einem Zufallsexperiment [4, p. 2]. In der bayesianischen Statistik hingegen wird die Wahrscheinlichkeit als Grad des Glaubens respektiv als **Plausibilität** eines Ereignisses oder einer Aussage interpretiert [1, p. 1].

Kern der bayesianischen Statistik ist es Wissen über ein Ereignis zu verfeinern, sobald neue Informationen vorliegen. Dazu nutzt man hauptsächlich das **Bayes'sche Theorem**, welches erlaubt das Vorwissen (Prior) mit neuen Daten (Likelihood) zu kombinieren und daraus eine aktualisierte Wahrscheinlichkeit (Posterior) zu berechnen.

Mit Hilfe des Bayes'schen Theorems kann man unbekannte Parameter schätzen, ein Konfidenzintervall für diese Parameter angeben und Hypothesen prüfen. Die klassische Statistik benötigt hingegen dafür Testgrößen, weshalb die bayesianische Statistik als flexibler und intuitiver gilt. [1, p. 1].

Problem der bayesianischen Statistik ist jedoch, dass die Berechnung der Posteriorverteilung analytisch oft nicht möglich ist. Da es nun aber gute numerische Methoden wie die **Markov Chain Monte Carlo (MCMC)** Methoden gibt, findet die bayesianische Statistik immer mehr Anwendungen. So zum Beispiel in der Medizin oder für künstliche Intelligenzen. [4, p. 1].

2.1.2 Das Bayes'sche Theorem und seine Bestandteile

Das Bayes'sche Theorem ist ein fundamentales Konzept der bayesianischen Statistik. Es beschreibt, wie man vorhandenes Vorwissen durch neue Daten aktualisiert.

Die **Priorverteilung** beschreibt die anfänglichen Annahmen oder das Vorwissen über einen Parameter oder ein Ereignis, bevor neue Daten berücksichtigt werden. Dabei "enthält die Priorverteilung eines Parameters θ , ausgedrückt durch $f(\theta)$, was man vor Auswertung der Stichprobe über θ weiß." [4, p. 90].

Als Priori-Wahrscheinlichkeit wird somit die Wahrscheinlichkeit $P(A)$ bezeichnet.

Die **Posteriorverteilung** beschreibt das Wissen über einen Parameter oder ein Ereignis, nachdem alle vorhandenen Daten berücksichtigt wurden. Durch die neuen Daten, meist einer Stichprobe, wird die anfängliche Annahme, die durch die Priorverteilung ausgedrückt wird, aktualisiert. Dies führt zu einer neuen Verteilung die widerspiegelt, wie wahrscheinlich verschiedene Werte des Parameters auf Grundlage sowohl des Vorwissens als auch der neuen Informationen sind. [4, p. 109]

Die Posteriori-Wahrscheinlichkeit wird somit als $P(A|B)$ bezeichnet.

Die **Likelihood-Funktion** enthält die Informationen, die die Daten über den Parameter oder das Ereignis liefern. Dabei beschreibt die Likelihood die Informationen aus den neuen Daten, die zur Aktualisierung der Prioriverteilung beitragen. [4, p. 88] Die Likelihood-Wahrscheinlichkeit wird somit als $P(B|A)$ bezeichnet.

Die Wahrscheinlichkeit $P(B)$ wird als Normierungskonstante bezeichnet. Sie sorgt dafür, dass die Posterioriverteilung korrekt normiert ist, das heißt, dass die Summe der Wahrscheinlichkeiten aller möglichen Werte des Parameters 1 ergibt. [4, p. 109]

Das Bayes'sche Theorem lässt sich somit wie folgt darstellen:

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

Das Bayes'sche Theorem lässt sich auch rekursiv anwenden [1, p. 17].

Gegeben sei das Ereignis A sowie die Teilergebnisse B_1, B_2, \dots, B_n . Dann ergibt sich die Wahrscheinlichkeit $P(A|B_1)$ zu:

$$P(A|B_1) = \frac{P(A) \cdot P(B_1|A)}{P(B_1)} \quad (2)$$

Nun wird die Information B_2 hinzugefügt. Die Wahrscheinlichkeit $P(A|B_1, B_2)$ ergibt sich bei Unabhängigkeit von den Teilereignissen B_1, B_2, \dots, B_n zu:

$$P(A|B_1, B_2) = \frac{P(A) \cdot P(B_1|A) \cdot P(B_2|A)}{P(B_1) \cdot P(B_2)} \quad (3)$$

Weiterhin lässt sich diese Formel umstellen, wodurch deutlich wird, dass beim Hinzufügen von neuen Informationen die Posterioriverteilung aktualisiert wird:

$$P(A|B_1, B_2) = \frac{P(A) \cdot P(B_1|A) \cdot P(B_2|A)}{P(B_1) \cdot P(B_2)} = P(A|B_1) \cdot \frac{P(B_2|A)}{P(B_2)} \quad (4)$$

Dies lässt sich allgemein formulieren für:

$$P(A|B_1, B_2, \dots, B_n) = P(A|B_1, B_2, \dots, B_{n-1}) \cdot \frac{P(B_n|A)}{P(B_n)} \quad (5)$$

Die Wahl der Prioriverteilung ist ein wichtiger Aspekt der bayesianischen Statistik. Sie wird immer so gewählt, dass die Entropie maximal ist. Die Entropie ist ein Maß für die Unsicherheit, was bedeutet, dass nur Informationen enthalten sind, die vor der Beobachtung bekannt sind. [1, p. 57]. Unter folgenden Bedingungen ist die Prioriverteilung optimal [1, p. 59]:

- Zufallsvariablen, die in $[a, b]$ definiert sind, sind **gleichverteilt**
- Zufallsvariablen mit gegebenen Mittelwert und Varianz sind **normalverteilt**
- Zufallsvariablen mit gegebenem Mittelwert sind **exponentialverteilt**
- Zufallsvariablen mit gegebenem Mittelwert und Varianz im Intervall $[0, \infty]$ besitzen eine **abgeschnittene Normalverteilung**

Wenn keine Informationen über den Parameter vorliegen, wird eine **uninformative** Prioriverteilung gewählt. Es handelt sich dabei um eine uneigentliche Verteilung. [1, p. 57].

2.1.3 Beispiele und praktische Anwendungen

Beispiel 1: m gleichgeformte Kugeln, unter denen sich k rote Kugeln und $m - k$ schwarze Kugeln befinden. Eine Kugel wird zufällig gezogen. Die Wahrscheinlichkeit, dass die gezogene Kugel rot ist, beträgt

$$P(A) = \frac{k}{m} = p \quad (6)$$

Der Versuch wird erweitert, sodass n -mal eine Kugel mit Zurücklegen gezogen wird. Die Wahrscheinlichkeit, dass x -mal eine rote Kugel bei n -maligem Ziehen gezogen wird, beträgt

$$P(x|n, p) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x} \quad (7)$$

Sei nun p unbekannt. Dieses p ist nun zu schätzen. Die Binomialverteilung wird nun als Likelihood-Funktion verwendet:

$$P(n, x|p) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n-x} \quad (8)$$

wobei $0 \leq p \leq 1$. Als Prioridichte wird die Gleichverteilung verwendet, da es keine Informationen über p gibt.

$$P(p) = \begin{cases} 1, & \text{für } 0 \leq p \leq 1 \\ 0, & \text{sonst} \end{cases} \quad (9)$$

Die Posterioridichte ergibt sich somit zu:

$$P(p|n, x) = \frac{\binom{n}{x} p^x \cdot (1 - p)^{n-x}}{P(n, x)} \quad (10)$$

Vergleicht man dies mit der Dichtefunktion der Beta-Verteilung, so erkennt man dass die Posterioridichte einer Beta-Verteilung entspricht.

$$P(p|n, x) = \frac{(n+1)!}{x! \cdot (n-x)!} \cdot p^x \cdot (1-p)^{n-x} = \frac{\Gamma(n+1)}{\Gamma(x+1) \cdot \Gamma(n-x+1)} \cdot p^x \cdot (1-p)^{n-x} \quad (11)$$

Somit suche nach Maximum der Posterioridichte, um den Schätzer für p zu finden.

$$\frac{d}{dp} P(p|n, x) = x p^{x-1} \cdot (1-p)^{n-x} - (n-x) p^x \cdot (1-p)^{n-x-1} = 0 \quad (12)$$

$$\Rightarrow x p^{x-1} \cdot (1-p)^{n-x} = (n-x) p^x \cdot (1-p)^{n-x-1} \quad (13)$$

Vereinfacht ergibt sich:

$$\Rightarrow x(1-p) = (n-x)p \quad (14)$$

$$\Rightarrow \frac{x}{p} - \frac{n-x}{1-p} = 0 \quad (15)$$

$$\Rightarrow p = \frac{x}{n} \quad (16)$$

Der Schätzer für p ist somit der relative Anteil der roten Kugeln an der Gesamtanzahl der Kugeln.

Beispiel 2: Beispiel 1 wird erweitert. Es wird nun eine zweite Stichprobe gezogen. Stichprobe 1: $n_1 = 10$, $x_1 = 4$, Stichprobe 2: $n_2 = 20$, $x_2 = 6$. Daten sind unabhängig voneinander. Die Daten (Posterioriverteilung) der 1. Stichprobe dienen nun als Prioridichte für die 2. Stichprobe. Man erhält somit:

$$P(p|n_1, x_1, n_2, x_2) = \frac{P(p|n_1, x_1) \cdot P(p|n_2, x_2)}{P(n_1, x_1, n_2, x_2)} \quad (17)$$

Dabei ist die Prioridichte identisch zur Posterioridichte der 1. Stichprobe, siehe Gleichung (10). Die Posterioridichte der 2. Stichprobe ergibt sich somit zu:

$$P(p|n_1, x_1, n_2, x_2) = \frac{\binom{n_1+n_2}{x_1+x_2} \cdot p^{x_1} \cdot (1-p)^{n_1-x_1} \cdot p^{x_2} \cdot (1-p)^{n_2-x_2}}{P(n_1, x_1, n_2, x_2)} \quad (18)$$

Oder mithilfe der Beta-Verteilung:

$$P(p|n_1, x_1, n_2, x_2) = \frac{\Gamma(n_1 + n_2 + 1)}{\Gamma(x_1 + x_2 + 1) \cdot \Gamma(n_1 + n_2 - x_1 - x_2 + 1)} \cdot p^{x_1+x_2} \cdot (1-p)^{n_1+n_2-x_1-x_2} \quad (19)$$

Die Daten der 1. und 2. Stichprobe können somit kombiniert werden. Für die Daten $n_1 = 10$, $x_1 = 4$, $n_2 = 20$, $x_2 = 6$:

$$P(p|10, 4, 20, 6) = 931395465p^{10} \cdot (1-p)^{20} \quad (20)$$

2.1.4 Punktschätzer, Konfidenzintervalle und Hypothesenprüfung in der bayesianischen Statistik

Punktschätzer

Im folgenden wird die Schätzung eines Parameters mithilfe der Bayes-Strategie erläutert. Die möglichen Schätzwerte der Parameter x werden als \hat{x} bezeichnet. Die wahren Parameter werden als x bezeichnet.

Es wird eine Kostenfunktion $L(\hat{x}, x)$ definiert, die die Kosten für die Schätzung \hat{x} des wahren Parameters x angibt. Dies bedeutet, dass die Kostenfunktion die Differenz zwischen dem wahren Parameter x und der Schätzung \hat{x} angibt. Dabei gibt es verschiedene Kostenfunktionen, die verwendet werden können. [1, p. 65]

Die **quadratische Kostenfunktion** ist definiert als: $L(x - \hat{x}) = (x - \hat{x})\Sigma^{-1}(x - \hat{x})$. Diese gibt den quadratischen Abstand zwischen dem wahren Parameter x und der Schätzung \hat{x} an. Die zu erwartenden Kosten werden berechnet mit dem Erwartungswert der Kostenfunktion. Diese Schätzung führt zu dem Erwartungswert von x , das heißt $\hat{x} = E(x)$. [1, pp. 65–66]

Die **Kostenfunktion der absoluten Fehler** ist definiert als: $L(x, \hat{x}) = |x - \hat{x}|$. Diese gibt den absoluten Abstand zwischen x und \hat{x} an. Die Schätzung mit dem absoluten Fehler ergibt den Median der Verteilung, das heißt $F(\hat{x}_{med}) = 0.5$ [1, pp. 67–68]

Die **Null-Eins-Kostenfunktion** bedeutet, dass es entweder Kosten oder keine Kosten gibt. Diese ist definiert durch: $L(x - \hat{x}) = \begin{cases} 0 & \text{für } |x - \hat{x}| < b \\ a & \text{für } |x - \hat{x}| \geq b \end{cases}$, wobei a und b als Konstanten angenommen werden. Wenn der Fall $b \rightarrow 0$ betrachtet wird, ergibt sich als Schätzer das Argument des Maximums der Posterioriverteilung, das heißt $\hat{x}_M = \arg \max p(x|y)$. [1, pp. 68–69]

Es ist so ersichtlich, dass es verschiedene Punktschätzer in der bayesianischen Statistik gibt:

- **Erwartungswert** $\hat{x} = E(x)$
- **Median** $\hat{x} = x_{0.5}$
- **Maximum der Posterioriverteilung** $\hat{x} = \arg \max p(x|y)$

Die Wahl zwischen den Schätzern hängt so von der Problemstellung ab.

Konfidenzintervalle

Um Unsicherheiten bei Schätzungen zu quantifizieren werden oftmals **Konfidenzintervall** verwendet. Konfidenzintervalle geben einen Bereich an, in dem ein unbekannter Parameter mit einer bestimmten Wahrscheinlichkeit liegt. In der Bayes Statistik kann eine Bereichsschätzung unmittelbar aus der Posterioriverteilung abgeleitet werden. [1, p. 71]

Wenn $P(x|y)$ die Posterioridichte für den Parameter x ist, ist das Konfidenzintervall mit Konfidenzniveau $1-\alpha$ definiert als:

$$P(x \in X_u|y) = \int_{X_u} P(x|y)dx = 1 - \alpha \quad (21)$$

Wobei für alle $x_1 \in X_u, x_2 \notin X_u$ gilt $P(x_1|y) \geq P(x_2|y)$; X_u ist dabei ein Unterraum auf dem Parameterraum [1, p. 71]. Meist wird für α 0.1, 0.05 oder 0.01 gewählt. Höhere Konfidenzniveaus führen zu größeren Intervallen, da mehr Unsicherheit abgedeckt wird, während niedrigere Konfidenzniveaus kleinere Intervalle liefern, die jedoch ein höheres Risiko haben, den wahren Parameter nicht einzuschließen.

Falls die Posterioriverteilung $P(x|y)$ beispielsweise der Normalverteilung $N(\mu, \sigma^2)$ entspricht, ist das Konfidenzintervall symmetrisch um den Mittelwert und die Grenzen lassen sich einfach durch die Quantile der Standardnormalverteilung z über $\mu - z_{1-\alpha/2} * \sigma$ und $\mu + z_{1-\alpha/2} * \sigma$ bestimmen.

Da die Konfidenzintervalle in der bayesianischen Statistik auf der Posterioriverteilung basieren, unterscheiden sich die Ergebnisse von denen der klassischen Statistik, da diese auf einer anderen Grundlage definiert sind.

Hypothesenprüfung

Hypothesenprüfung ist eine Methode, mit der man Annahmen über unbekannte Parameter überprüft. Dabei wird entschieden, ob solche Annahmen akzeptiert oder abgelehnt werden sollen. [1, p. 74]. Man unterscheidet zwischen der Nullhypothese H_0 und der Alternativhypothese H_1 . Die Nullhypothese ist die Annahme, die überprüft wird, während die Alternativhypothese die Annahme ist, die akzeptiert wird, wenn die Nullhypothese abgelehnt wird.

In der bayesianischen Statistik werden Hypothesen durch Unterräume des Parameterraums definiert. Seien X_1 und X_2 disjunkte Unterräume des Parameterraums, so lauten die Hypothesen $H_0 : x \in X_1$ und $H_1 : x \in X_2$. Die Wahrscheinlichkeiten für diese Hypothesen ergeben sich direkt aus der Posterior-Dichte. Die Wahrscheinlichkeit, dass die Nullhypothese wahr ist, lautet:

$$P(H_0|y) = \int_{X_1} P(x|y)dx \quad (22)$$

Analog gilt dies für $P(H_1|y)$ [1, p. 74]. Ist jetzt $P(H_0|y) > P(H_1|y)$ resp. $\frac{P(H_0|y)}{P(H_1|y)} > 1$, so wird die Nullhypothese akzeptiert, andernfalls die Alternativhypothese. [1, p. 77].

Bei einer Punkthypothese $H_0 : x = x_0$ ergibt sich $P(H_0|y) = 0$, da die Posterior-Dichte kontinuierlich ist und ein einzelner Punkt daher keine Wahrscheinlichkeit besitzt. In solchen Fällen greift man beispielsweise auf Konfidenzintervalle zurück: Liegt der Wert x_0 innerhalb des Konfidenzintervalls, wird die Nullhypothese akzeptiert, andernfalls abgelehnt [1, p. 84].

Ein wesentlicher Unterschied zur klassischen Statistik besteht darin, dass in der traditionellen Hypothesenprüfung die Nullhypothese solange als gültig angesehen wird, bis genügend Beweise dagegen vorliegen (Prinzip des Zweifels). In der bayesianischen Statistik hingegen werden die Wahrscheinlichkeiten beider Hypothesen direkt geschätzt, wodurch beide Hypothesen von Anfang an gleichberechtigt betrachtet werden. [1, p. 83].

2.2 Markov Chain Monte Carlo (MCMC) Methoden

2.2.1 Einführung in MCMC-Methoden

Bei einer direkten Simulation wird vorausgesetzt, dass die Verteilung der Zufallsvariablen bekannt ist. Dies ist jedoch in der Praxis nicht immer gegeben. Die Berechnung der Posterioriverteilung ist analytisch oft nicht möglich, vor allem bei komplexen Modellen oder hohen Dimensionen.

Die Markov Chain Monte Carlo (MCMC) Methoden sind eine Klasse von Algorithmen, die es ermöglichen, eine Stichprobe aus einer Verteilung zu ziehen, ohne die Verteilung zu kennen. [3, p. 179]

Diese Methoden verwenden zwei Konzepte: Markov-Ketten und Monte Carlo-Methoden. Eine Markov-Kette ist eine Folge von Zufallsvariablen, die die Markov-Eigenschaft erfüllen. Die Markov-Eigenschaft sagt aus, dass die nächste Zufallsvariable nicht von den vorherigen Zufallsvariablen, sondern nur von der letzten Zufallsvariable abhängt. Das bedeutet, dass die Wahrscheinlichkeit, im nächsten Zustand $X_n + 1$ zu landen, nur von X_n abhängt. Die Übergangswahrscheinlichkeit zwischen den Zuständen kann in einer Übergangsmatrix dargestellt werden. [3, 188f.]

Die Monte Carlo-Methoden sind eine Gruppe von Algorithmen, die es ermöglichen, Zufallsvariablen zu schätzen, indem Zufallszahlen generiert werden. Sie erzeugen zufällige Stichproben, um eine Näherung der Verteilung zu erhalten. [3, 14f.]

Die MCMC-Methoden nutzen die Monte Carlo-Methoden, um eine Markov-Kette zu simulieren. Diese Technik ist besonders nützlich, um eine Posterioriverteilung zu schätzen, wenn direkte Berechnungen nicht möglich sind. [3, p. 179]

Im folgenden wird der Metropolis-Hastings-Algorithmus erläutert, der eine der bekanntesten MCMC-Methoden ist.

2.2.2 Der Metropolis-Hastings-Algorithmus

Der Metropolis-Hastings-Algorithmus erstellt eine Markov-Kette, die eine Posterioriverteilung $f(x)$ approximiert. $f(x)$ besteht aus einer bekannten Funktion $p(x)$, die sich aus der Likelihood-Funktion und der Prioriverteilung zusammensetzt und aus einer unbekannten Normierungskonstante. Für die Berechnung wird eine Hilfsfunktion $q(y|x)$ benötigt, welche die Übergangswahrscheinlichkeit von einem Zustand x_t zum nächsten Zustand x_{t+1} angibt und ähnlich wie die Posterioriverteilung ist. Oftmals wird dafür die Normalverteilung gewählt sodass $q(y|x) = N(x, \sigma^2)$, wobei σ beeinflusst dabei die Schrittweite, die die Markov-Kette macht. [2, 226f.]

Der Algorithmus besteht aus folgenden Schritten:

1. Wähle einen Startwert x_0
2. Wähle einen neuen Wert y aus der Hilfsfunktion $q(y|x_t)$ im Fall einer Normalverteilung $N(x_t, \sigma^2)$ wird ein zufälliger Wert aus der Normalverteilung gezogen
3. Berechne die Akzeptanzwahrscheinlichkeit α :

$$\alpha = \min\left\{1, \frac{p(y) \cdot q(x_t|y)}{p(x_t) \cdot q(y|x_t)}\right\} \quad (23)$$

Diese gibt an, wie wahrscheinlich der neue Wert x_{t+1} akzeptiert wird.

4. Ziehe eine Zufallszahl u aus einer Gleichverteilung $U(0,1)$
5. Wähle den nächsten Wert x_{t+1} :

$$x_{t+1} = \begin{cases} y, & \text{wenn } u < \alpha \\ x_t, & \text{sonst} \end{cases} \quad (24)$$

6. Wiederhole die Schritte 2-5 bis die Stichprobe die gewünschte Verteilung genau genug approximiert

Der Metropolis-Hastings-Algorithmus ist eine Verallgemeinerung des Metropolis-Algorithmus, der nur für symmetrische Hilfsfunktionen $q(y|x)$ funktioniert. Bei einer symmetrischen Hilfsfunktion ist $q(y|x_t) = q(x_t|y)$ und deswegen kürzt sich die Akzeptanzwahrscheinlichkeit zu $\alpha = \min\left\{1, \frac{p(y)}{p(x_t)}\right\}$. [2, 226f.]

Der erste Teil der Markov-Kette wird als **Burn-in-Phase** bezeichnet, welche verworfen wird, weil sich die Kette erst an die Zielverteilung anpassen muss. Der restliche Teil wird als **Sampling-Phase** bezeichnet, welche dann die Verteilung simuliert. [2, 226f.]

2.2.3 Konvergenzkriterien für MCMC-Simulationen

3 Praktischer Teil

3.1 Implementierung des Metropolis-Hastings-Algorithmus in Python

3.2 Anwendung der bayesianischen Statistik am Beispiel

3.2.1 Überblick und Aufbereitung des Datensatzes

3.2.2 Anwendung der bayesianischen Statistik

3.2.3 Rekursives Einbeziehen von neuen Daten

3.2.4 Vergleich mit traditioneller Statistik

3.3 Vergleich mit klassischen Methoden am Beispiel

3.3.1 Einleitung

Semmelweis' Vermutung zum Kindbettfieber basiert auf seiner Beobachtung, dass der Sterberate von Frauen bei der Geburt in einer Klinik stark variierte, je nachdem, ob die Geburtshelfer ihre Hände vor der Entbindung in einer Chlorlösung wuschen. Er vermutete, dass das Kindbettfieber, das häufig tödlich endete, durch eine Infektion verursacht wurde, die durch nicht desinfizierte Hände von Ärzten übertragen wurde. Diese Vermutung führte zu einer drastischen Senkung der Sterblichkeit in den Kliniken, in denen sie umgesetzt wurde. [4, p. 3]

3.3.2 Datensatz

Der Datensatz, den Semmelweis zur Untermauerung seiner Theorie verwendete, umfasst zwei Zeiträume:

- Vor Einführung der Chlorwaschung: In den 6 Jahren vor der Einführung der Chlorwaschung in der Klinik starben 1.989 von 20.042 Frauen.
- Nach Einführung der Chlorwaschung: In den 14 Jahren nach der Einführung starben nur 1.883 von 56.104 Frauen.
- $x = 1989, n_x = 20.042, y = 1883, n_y = 56.104$

3.3.3 Berechnung mit klassischer Statistik

Punktschätzer:

$$\hat{\delta} = \hat{\pi}_X - \hat{\pi}_Y \quad (25)$$

$$= \frac{x}{n_x} - \frac{y}{n_y} \quad (26)$$

$$= \frac{1989}{20.042} - \frac{1883}{56.104} \quad (27)$$

$$= 0.0992 - 0.0336 \quad (28)$$

$$= 0.0656 \quad (29)$$

Konfidenzintervall:

Für $\alpha = 0.01$ beidseitige Konfidenzintervalle für π_X und π_Y :
 $x = 1989, n_x = 20.042, y = 1883, n_y = 56.104, z_{\alpha/2} = 2.576$

$$I_X = \left[\hat{\pi}_X - z_{\alpha/2} \cdot \sqrt{\frac{\hat{\pi}_X \cdot (1 - \hat{\pi}_X)}{n_x}}, \hat{\pi}_X + z_{\alpha/2} \cdot \sqrt{\frac{\hat{\pi}_X \cdot (1 - \hat{\pi}_X)}{n_x}} \right] \quad (30)$$

$$I_X = \left[0.0992 - 2.576 \cdot \sqrt{\frac{0.0992 \cdot (1 - 0.0992)}{20.042}}, 0.0992 + 2.576 \cdot \sqrt{\frac{0.0992 \cdot (1 - 0.0992)}{20.042}} \right] \quad (31)$$

$$I_X = [0.0938, 0.1046]$$

$$I_Y = \left[\hat{\pi}_Y - z_{\alpha/2} \cdot \sqrt{\frac{\hat{\pi}_Y \cdot (1 - \hat{\pi}_Y)}{n_y}}, \hat{\pi}_Y + z_{\alpha/2} \cdot \sqrt{\frac{\hat{\pi}_Y \cdot (1 - \hat{\pi}_Y)}{n_y}} \right] \quad (32)$$

$$I_Y = \left[0.0336 - 2.576 \cdot \sqrt{\frac{0.0336 \cdot (1 - 0.0336)}{56.104}}, 0.0336 + 2.576 \cdot \sqrt{\frac{0.0336 \cdot (1 - 0.0336)}{56.104}} \right] \quad (33)$$

$$I_Y = [0.0316, 0.0356]$$

Hypothesentest:

Es wird $H_0 : \pi_X \leq \pi_Y$ oder auch $H_0 : \pi_X - \pi_Y \leq 0$ gegen $H_1 : \pi_X - \pi_Y > 0$ getestet. Dafür wird ein einseitiger z-Test der Differenz der Anteile durchgeführt.

$$\delta_+ = z_\alpha \cdot \sqrt{\frac{\hat{\pi}_X \cdot (1 - \hat{\pi}_X)}{n_x} + \frac{\hat{\pi}_Y \cdot (1 - \hat{\pi}_Y)}{n_y}} \quad (34)$$

Für $\alpha = 0.01$ ergibt sich $z_\alpha = 2.33$.

$$\delta_+ = 2.33 \cdot \sqrt{\frac{0.0992 \cdot (1 - 0.0992)}{20042} + \frac{0.0336 \cdot (1 - 0.0336)}{56104}} \quad (35)$$

$$\delta_+ = 0.0052 \quad (36)$$

Da $\hat{\delta} = 0.0656 > \delta_+ = 0.0052$ können wir die Nullhypothese H_0 ablehnen und die Vermutung von Semmelweis annehmen.

3.3.4 Berechnung mit bayesianischer Statistik

Bei der bayesianischen Statistik können wir Priors definieren und die Wahrscheinlichkeit der Sterblichkeitsraten unter Berücksichtigung von Vorwissen berechnen.

Es werden die Wahrscheinlichkeiten für $\Pi_X > \Pi_Y$ und $\Pi_X \leq \Pi_Y$ ermittelt.

Vorwissen für Sterblichkeitsrate mit und ohne Chlorwaschung:

- wahrscheinlichste Sterblichkeitsrate: 0.021
- halb so wahrscheinlich: 0.030

Parameter a' und b' für die Beta-Verteilung:

$$m = 0.021 \quad (37)$$

$$\pi_v = 0.03 \quad (38)$$

$$\frac{f(v)}{f(m)} = 0.5 \quad (39)$$

Die Beta-Verteilung hat die Dichtefunktion:

$$f(p; a', b') = \frac{1}{B(a', b')} p^{a'-1} (1-p)^{b'-1},$$

wobei $B(a', b')$ die Beta-Funktion ist:

$$B(a', b') = \frac{\Gamma(a')\Gamma(b')}{\Gamma(a' + b')}.$$

Der Modus m der Verteilung ist definiert als:

$$m = \frac{a' - 1}{a' + b' - 2}.$$

Das Verhältnis der Wahrscheinlichkeiten bei π_v und m ist:

$$\frac{f(\pi_v)}{f(m)} = \frac{\frac{1}{B(a', b')} \pi_v^{a'-1} (1 - \pi_v)^{b'-1}}{\frac{1}{B(a', b')} m^{a'-1} (1 - m)^{b'-1}} = \left(\frac{\pi_v}{m}\right)^{a'-1} \left(\frac{1 - \pi_v}{1 - m}\right)^{b'-1}.$$

Daraus ergibt sich:

$$\frac{f(\pi_v)}{f(m)} = \left(\frac{\pi_v}{m}\right)^{a'-1} \left(\frac{1 - \pi_v}{1 - m}\right)^{b'-1}.$$

Umstellen führt zu den folgenden Formeln für die Parameter a' und b' :

$$a' = \frac{\ln \frac{f(\pi_v)}{f(m)}}{\ln \frac{\pi_v}{m} + \frac{1-m}{m} \ln \frac{1-\pi_v}{1-m}} + 1,$$

$$b' = \frac{\ln \frac{f(\pi_v)}{f(m)}}{\ln \frac{\pi_v}{m} + \frac{1-m}{m} \ln \frac{1-\pi_v}{1-m}} \cdot \frac{1-m}{m} + 1.$$

Setzen wir die bekannten Werte ein:

$$m = 0.021, \quad \pi_v = 0.03, \quad \ln \frac{f(\pi_v)}{f(m)} = \ln(0.5) \approx -0.693,$$

erhalten wir:

$$a' = \frac{-0.693}{\ln \frac{0.03}{0.021} + \frac{1-0.021}{0.021} \ln \frac{1-0.03}{1-0.021}} + 1 = 9.38 + 1 \approx 10,$$

$$b' = \frac{-0.693}{\ln \frac{0.03}{0.021} + \frac{1-0.021}{0.021} \ln \frac{1-0.03}{1-0.021}} \cdot \frac{1-0.021}{0.021} + 1 = 9.38 \cdot 46.62 + 1 \approx 438.$$

Parameter der Posterioriverteilungen von Π_X und Π_Y :

$$a_X = a' + x = 10 + 1989 = 1999 \tag{40}$$

$$b_X = b' + n_x - x = 438 + 20042 - 1989 = 18491 \tag{41}$$

$$a_Y = a' + y = 10 + 1883 = 1893 \tag{42}$$

$$b_Y = b' + n_y - y = 438 + 56104 - 1883 = 54659 \quad (43)$$

Median der Posterioriverteilung:

$$\hat{\pi}_X = \frac{a_X - 1/3}{a_X + b_X - 2/3} = \frac{1999 - 1/3}{1999 + 18491 - 2/3} = 0.0975 \quad (44)$$

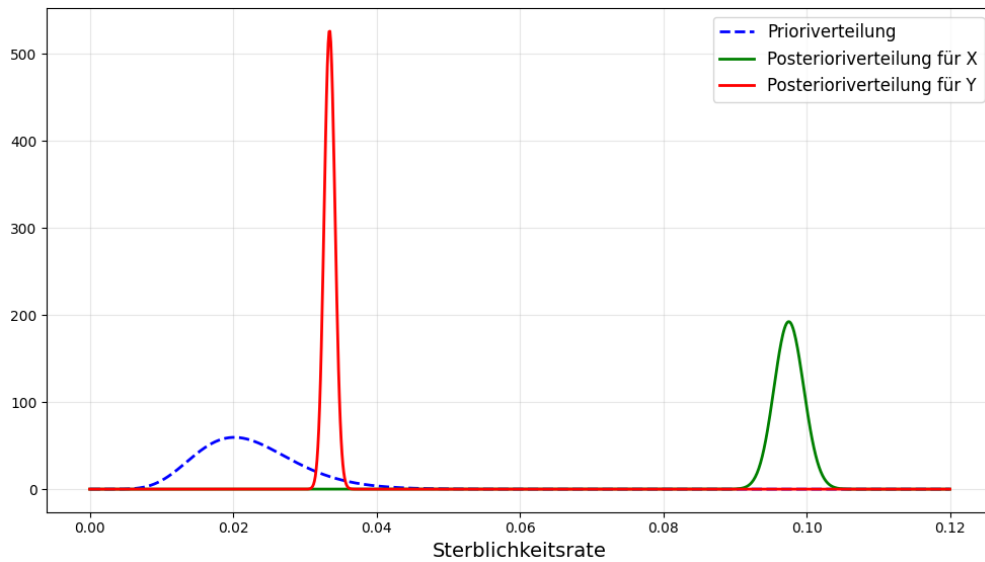
$$\hat{\pi}_Y = \frac{a_Y - 1/3}{a_Y + b_Y - 2/3} = \frac{1893 - 1/3}{1893 + 54659 - 2/3} = 0.0335 \quad (45)$$

Intervallschätzung:

Die Posteriors für Π_X und Π_Y sind Beta-verteilte Zufallsvariablen:

$$\Pi_X \sim \text{Beta}(1999, 18491) \quad (46)$$

$$\Pi_Y \sim \text{Beta}(1893, 54659) \quad (47)$$



Für $\alpha = 0.01$ beidseitige Konfidenzintervalle für Π_X und Π_Y :

$$I_X = [0.0927, 0.1034] \quad I_Y = [0.0316, 0.0355] \quad (48)$$

3.3.5 Vergleich der Ergebnisse

Der Vergleich der klassischen Statistik mit der bayesianischen Statistik liefert uns unterschiedliche Perspektiven:

- Mit der klassischen Statistik haben wir die Sterblichkeitsraten für beide Perioden direkt berechnet und den Unterschied gemessen.
- Die bayesianische Statistik berücksichtigt zusätzlich Vorwissen und liefert eine schätzungsweise Wahrscheinlichkeitsverteilung der Sterblichkeitsraten, was eine flexiblere Modellierung und Unsicherheitsabschätzung ermöglicht.

3.3.6 Vergleich der klassischen und bayesianischen Statistik

	Klassische Statistik	Bayesianische Statistik
Bevorzugte Verwendung	Human- und Sozialwissenschaften, Wirtschaftswissenschaften, Biologie	Technik und Künstliche Intelligenz
Schätzen von Parametern, Testen von Hypothesen	nur Stichprobe wird betrachtet	Stichprobe und Vorwissen wird betrachtet
Ergebnis beim Schätzen	Punkt- oder Intervallschätzer, z. B. Konfidenzintervall	Wahrscheinlichkeitsverteilung für den Parameter (Posterior)
Ergebnis beim Testen	p -Wert; Entscheidung basierend auf Signifikanzniveau, angenommen oder abgelehnt	Posterior-Wahrscheinlichkeit für Hypothesen; Bayes-Faktor

Table 1: Vergleich der klassischen und bayesianischen Statistik

Statistikmethode	Vorteile	Nachteile
Klassische Statistik	<ul style="list-style-type: none"> • Einfachheit und breite Akzeptanz in der Wissenschaft. • Gut für große Datensätze und standardisierte Testverfahren. • Benötigt keine subjektiven Annahmen über Wahrscheinlichkeiten. 	<ul style="list-style-type: none"> • Abhängig von großen Stichproben. • Unflexibel gegenüber neuen Daten und Modellen. • Ergebnisse oft schwer zu interpretieren bei komplexen Datenstrukturen.
Bayesianische Statistik	<ul style="list-style-type: none"> • Flexibel und anpassbar an komplexe Modelle. • Ermöglicht die Einbeziehung von Vorwissen (Priors). • Liefert direkte Wahrscheinlichkeiten für Parameter und Modelle. 	<ul style="list-style-type: none"> • Subjektivität bei der Wahl der Priors. • Rechenintensiv, besonders bei großen Datensätzen. • Oft weniger bekannt und akzeptiert in konservativen Disziplinen.

Table 2: Vergleich der Vor- und Nachteile der klassischen und bayesianischen Statistik.

3.4 Vergleich anhand gewählter Datensätze - Fussball WM

Alle WM-Spiele und Siege der deutschen Nationalmannschaft wurden analysiert. (siehe oben) Besonderes Augenmerk lag auf der Siegrate gegen europäische und nicht-europäische Gegner.

Daten:

- Gegen europäische Teams wurden 65 Spiele gespielt, davon 36 gewonnen.
- Gegen nicht-europäische Teamsn wurden 47 Spiele gespielt, davon 36 gewonnen.
- Insgesamt wurden 112 Spiele gespielt, davon 72 gewonnen.

3.4.1 Berechnung mit klassischer Statistik

Punktschätzer:

$$\hat{\pi}_E = \frac{36}{65} = 0.5538$$

$$\hat{\pi}_{\neg E} = \frac{36}{47} = 0.7659$$

$$\hat{\pi}_G = \frac{72}{112} = 0.6429$$

Konfidenzintervall: Für $\alpha = 0.05$ beidseitige Konfidenzintervalle:

$$I_E = \left[0.5538 - 1.96 \cdot \sqrt{\frac{0.5538 \cdot (1 - 0.5538)}{65}}, 0.5538 + 1.96 \cdot \sqrt{\frac{0.5538 \cdot (1 - 0.5538)}{65}} \right]$$

$$I_E = [0.4330, 0.6746]$$

$$I_{\neg E} = \left[0.7659 - 1.96 \cdot \sqrt{\frac{0.7659 \cdot (1 - 0.7659)}{47}}, 0.7659 + 1.96 \cdot \sqrt{\frac{0.7659 \cdot (1 - 0.7659)}{47}} \right]$$

$$I_{\neg E} = [0.6448, 0.8870]$$

$$I_G = \left[0.6429 - 1.96 \cdot \sqrt{\frac{0.6429 \cdot (1 - 0.6429)}{112}}, 0.6429 + 1.96 \cdot \sqrt{\frac{0.6429 \cdot (1 - 0.6429)}{112}} \right]$$

$$I_G = [0.5542, 0.7316]$$

Hypothesentest:

Es wird $H_0 : \pi_E \geq \pi_{\neg E}$ oder auch $H_0 : \pi_E - \pi_{\neg E} \geq 0$ gegen $H_1 : \pi_E - \pi_{\neg E} < 0$ getestet. Dafür wird ein einseitiger z-Test der Differenz der Anteile durchgeführt.

Wir berechnen den Standardfehler für den Unterschied der Anteile:

$$SE = \sqrt{\hat{\pi}_G(1 - \hat{\pi}_G) \left(\frac{1}{65} + \frac{1}{47} \right)} = \sqrt{0.6429 \cdot (1 - 0.6429) \left(\frac{1}{65} + \frac{1}{47} \right)} \approx 0.0917$$

Berechnung des z-Werts:

$$z = \frac{\hat{\pi}_E - \hat{\pi}_{\neg E}}{SE} = \frac{0.5538 - 0.7659}{0.0917} \approx -2.31$$

Für ein Signifikanzniveau von $\alpha = 0.05$ und einen einseitigen Test ist der kritische z-Wert:

$$z_{\text{kritisch}} = -1.645$$

Da der berechnete z-Wert von -2.31 kleiner als der kritische z-Wert von -1.645 ist, können wir die Nullhypothese $H_0 : \pi_E \geq \pi_{\neg E}$ ablehnen. Interpretation: Die Siegrate der deutschen Nationalmannschaft gegen europäische Teams ist signifikant niedriger als gegen nicht-europäische Teams.

3.4.2 Berechnung mit bayesianischer Statistik

Vorwissen:

Für die bayesianische Analyse nehmen wir als Prior-Verteilungen Beta-Verteilungen für die Siegquote der deutschen Nationalmannschaft gegen europäische und nicht-europäische Teams an. Die Parameter dieser Verteilungen basieren auf den beobachteten Daten.

- Für europäische Teams:

$$\alpha_E = 36 + 1 = 37, \quad \beta_E = 29 + 1 = 30$$

Somit ist die Prior-Verteilung: Beta(37, 30).

- Für nicht-europäische Teams:

$$\alpha_{\neg E} = 36 + 1 = 37, \quad \beta_{\neg E} = 11 + 1 = 12$$

Somit ist die Prior-Verteilung: Beta(37, 12).

Likelihood-Verteilung:

Für die gesamte Siegquote (Likelihood) wird ebenfalls eine Beta-Verteilung verwendet, basierend auf den gesamten Spielen der deutschen Nationalmannschaft:

$$\alpha_{\text{total}} = 72 + 1 = 73, \quad \beta_{\text{total}} = 40 + 1 = 41$$

Die Likelihood-Verteilung ist daher: Beta(73, 41).

Die Posterior-Verteilung kann als Produkt von Prior und Likelihood berechnet werden. Sie wird durch Normierung auf den gesamten Bereich der Siegquote p bestimmt.

Posterior-Verteilung:

Der Posterior wird durch die folgende Gleichung berechnet:

$$\text{Posterior}(p) \propto \text{Prior}(p) \cdot \text{Likelihood}(p)$$

Die Verteilungen für die Prior-, Likelihood- und Posterior-Verteilungen werden wie folgt dargestellt:

- Prior für europäische Teams: Beta(37, 30)
- Prior für nicht-europäische Teams: Beta(37, 12)
- Likelihood für alle Spiele: Beta(73, 41)

Die Posterior-Verteilungen für europäische und nicht-europäische Teams werden durch das Produkt der Prior- und Likelihood-Verteilungen berechnet.

Punktschätzer:

Der Erwartungswert der Posterior-Verteilung ist der Mittelwert der Beta-Verteilung, der mit folgender Formel berechnet wird:

$$E = \frac{\alpha_{\text{posterior}}}{\alpha_{\text{posterior}} + \beta_{\text{posterior}}}$$

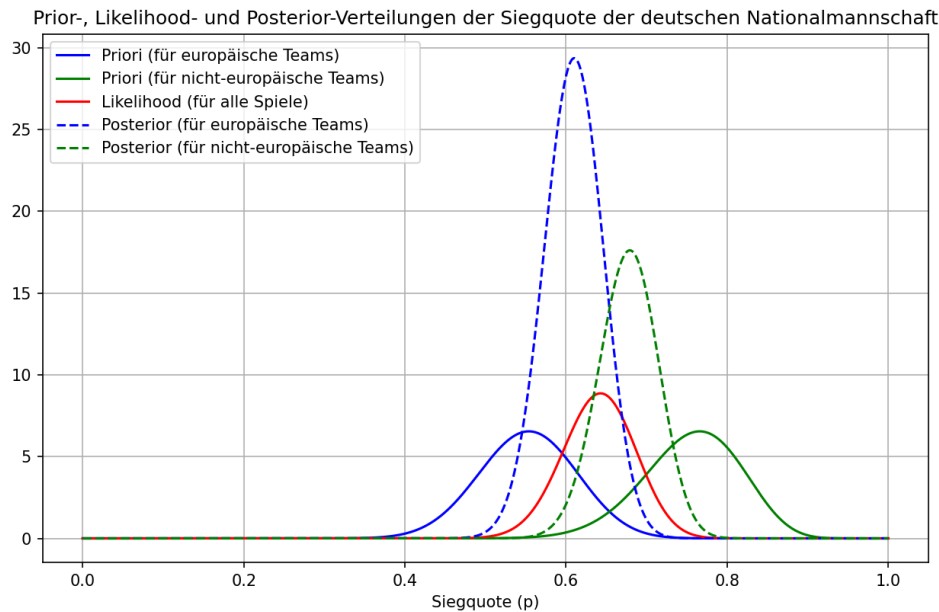


Figure 1: Prior-, Likelihood- und Posterior-Verteilungen der Siegquote der deutschen Nationalmannschaft

Für europäische Teams ist der Erwartungswert:

$$E_E = \frac{37 + 73}{37 + 73 + 30 + 41} = \frac{110}{181} \approx 0.6077$$

Für nicht-europäische Teams ist der Erwartungswert:

$$E_{\neg E} = \frac{37 + 73}{37 + 36 + 12 + 41} = \frac{110}{163} \approx 0.6748$$

Konfidenzintervall:

Das 95%-Konfidenzintervall für die Posterior-Verteilung wird durch die Quantilfunktion der Beta-Verteilung berechnet. Wir verwenden das 2.5%- und 97.5%-Quantil. Dies wurde mithilfe des Programms `world_cup_bayes.py` durchgeführt.

Für europäische Teams:

$$I_E = [0.5358, 0.6775]$$

Für nicht-europäische Teams:

$$I_{\neg E} = [0.6012, 0.7444]$$

Hypothesentest:

Wir testen die Hypothese, dass die Siegquote der deutschen Nationalmannschaft gegen europäische Teams niedriger ist als gegen nicht-europäische Teams. Dafür definieren wir die Nullhypothese $H_0 : \pi_E \geq \pi_{\neg E}$ und die Alternative $H_1 : \pi_E < \pi_{\neg E}$. Um die Hypothese zu testen, dass die Siegquote gegen europäische Teams niedriger ist als gegen nicht-europäische Teams, wurden Monte-Carlo-Simulationen verwendet:

- Aus den Posterior-Verteilungen wurden 100.000 Zufallswerte (Samples) gezogen.

- Die Wahrscheinlichkeit $P(H_1) = P(\pi_E < \pi_{\neg E})$ wurde als der Anteil der Simulationen berechnet, bei denen $\pi_E < \pi_{\neg E}$.
- Die Wahrscheinlichkeit $P(H_0) = 1 - P(H_1)$ wurde ebenfalls berechnet.
- Je nachdem, welche Wahrscheinlichkeit größer ist, wurde entschieden, ob H_0 (Nullhypothese) oder H_1 (Alternativhypothese) akzeptiert wird.

Ergebnisse:

$P(H_0) = 0.0972$ und $P(H_1) = 0.9028$

Da $P(H_1) > P(H_0)$, wird die Nullhypothese $H_0 : \pi_E \geq \pi_{\neg E}$ abgelehnt.

Interpretation: Die Siegquote der deutschen Nationalmannschaft gegen europäische Teams ist signifikant niedriger als gegen nicht-europäische Teams.

4 Zusammenfassung und Ausblick

5 Literaturverzeichnis

References

- [1] Karl-Rudolf Koch. *Einführung in die Bayes-Statistik*. Berlin [u.a.]: Springer, 2000. ISBN: 3540666702. URL: <https://katalog.slub-dresden.de/id/0-306244284>.
- [2] Dirk P. Kroese, Thomas Taimre, and Zdravko I. Botev. *Handbook of Monte Carlo Methods*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley, 2011. ISBN: 9780470177938. URL: <https://www.wiley.com/en-us/Handbook+of+Monte+Carlo+Methods-p-9780470177938>.
- [3] Thomas Müller-Gronbach, Erich Novak, and Klaus Ritter. *Monte Carlo-Algorithmen*. Berlin: Springer, 2012. ISBN: 9783540891406. URL: <https://katalog.slub-dresden.de/id/0-618339728>.
- [4] Wolfgang Tschirk. *Statistik: Klassisch oder Bayes zwei Wege im Vergleich*. Berlin , , © 2014. ISBN: 3642543847. URL: <https://katalog.slub-dresden.de/id/0-160866449X>.

6 Selbstständigkeitserklärung