



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	4
2.2	Binomiale Verteilung und deren bayesianische Interpretation	5
2.3	Markov Chain Monte Carlo (MCMC) Methoden	6
2.4	Konvergenzkriterien und Diagnosewerkzeuge für MCMC-Simulationen .	7
3	Praktischer Teil	8
3.1	Implementierung bayesianischer Modelle unter Verwendung in Python .	8
3.2	Anwendung der Modelle auf verschiedene Datensätze	9
3.3	Durchführung von MCMC-Simulationen	10
3.4	Interpretation der Ergebnisse	11
3.5	Vergleich mit klassischen Methoden	12
4	Zusammenfassung und Ausblick	13
5	Literaturverzeichnis	14
6	Selbstständigkeitserklärung	15

1 Einleitung

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 [2, 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. [2, 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." [2, 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. [2, 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. [2, 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. [2, p. 109]

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

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

Alternativ kann das Bayes-Theorem ohne die Normierungskonstante $P(B)$ in folgender proportionaler Form dargestellt werden [3, p. 15]:

$$P(A|B) \propto P(B|A) \cdot P(A) \quad (2)$$

2.1.3 Beispiele und praktische Anwendungen

2.2 Binomiale Verteilung und deren bayesianische Interpretation

2.3 Markov Chain Monte Carlo (MCMC) Methoden

2.4 Konvergenzkriterien und Diagnosewerkzeuge für MCMC-Simulationen

3 Praktischer Teil

3.1 Implementierung bayesianischer Modelle unter Verwendung in Python

3.2 Anwendung der Modelle auf verschiedene Datensätze

3.3 Durchführung von MCMC-Simulationen

3.4 Interpretation der Ergebnisse

3.5 Vergleich mit klassischen Methoden

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] Wolfgang Tschirk. *Statistik: Klassisch oder Bayes zwei Wege im Vergleich*. Berlin , , © 2014. ISBN: 3642543847. URL: <https://katalog.slub-dresden.de/id/0-160866449X>.
- [3] Dieter Wickmann. *Bayes-Statistik Einsicht gewinnen und entscheiden bei Unsicherheit*. Mannheim: BI-Wiss.-Verl., 1990. ISBN: 3411146710. URL: <https://katalog.slub-dresden.de/id/0-276492471>.

6 Selbstständigkeitserklärung