



Bayesianische Inferenz als Alternative zur klassischen Statistik

Jeff Gill und Richard Traummüller

Inhalt

1	Einleitung	672
2	Zwei gegensätzliche Philosophien	673
3	Grundlagen der bayesianischen Inferenz	676
4	Bayesianische stochastische Simulation	683
5	Ein empirisches Beispiel aus der empirischen Demokratieforschung	686
6	Fazit	693
7	Weiterführende Literatur	695
	Literatur	695

Zusammenfassung

Dieser Beitrag stellt die Grundzüge der bayesianischen Inferenz vor und argumentiert, dass es sich dabei um das ideale statistische Paradigma für die empirische Politikwissenschaft handelt. Die Politikwissenschaft ist in der Regel mit methodischen Herausforderungen und Daten konfrontiert, die mit den Vorstellungen der klassischen „frequentistischen“ Statistik nur unzureichend vereinbar sind. Bayesianische Methoden dagegen kombinieren Priori-Annahmen über interessierende Phänomene mit empirischer Evidenz, um dadurch zu informierten Wahrscheinlichkeitsaussagen zu gelangen. Darüber hinaus steht der moderne bayesianische Ansatz in enger Verbindung mit *Markov Chain Monte Carlo* (MCMC) Simulationsalgorithmen. Diese ermöglichen es, komplexere Modelle

J. Gill (✉)

Department of Government, American University, Washington DC, USA

E-Mail: jgill@american.edu

R. Traummüller

Fakultät für Sozialwissenschaften, Universität Mannheim, Mannheim, Deutschland

zu schätzen, als dies für herkömmliche Schätzverfahren der Fall ist. Schließlich überzeugt die bayesianische Herangehensweise durch die intuitive Form und Interpretierbarkeit der durch sie erzielten Ergebnisse. Wir demonstrieren die Nützlichkeit des bayesianischen Ansatzes anhand eines Beispiels aus der empirischen Demokratieforschung: der Frage, welchen Einfluss die staatliche Unterstützung von Religion für religiöses Sozialkapital im europäischen Vergleich besitzt.

Schlüsselwörter

Bayesianische Statistik · Hierarchische Modelle · Markov Chain Monte Carlo · A-priori Verteilung · Stochastische Simulation

1 Einleitung

Dieser Beitrag stellt die Grundzüge der bayesianischen Inferenz vor. Wir argumentieren, dass es sich dabei um das ideale statistische Paradigma für die empirische Politikwissenschaft handelt. Tatsächlich sind wir in unserer alltäglichen Arbeit mit methodischen Herausforderungen und Daten konfrontiert, die mit den Vorstellungen der klassischen frequentistischen Statistik kaum vereinbar sind. In der Politikwissenschaft haben wir es nur in den allerseltensten Fällen mit Daten zu tun, die sich in experimentellen Settings unter stets gleichen Bedingungen wiederholen und beliebig häufig erzeugen ließen. Vielmehr stellen sozialwissenschaftliche Datenerhebungen in der Regel singuläre Ereignisse dar oder umfassen oftmals sogar die gesamte interessierende Grundgesamtheit. Selbst für Umfragedaten bleibt die Vorstellung unendlich wiederholbarer Stichprobenziehungen, wie sie in der frequentistischen Statistik vorherrscht, ein reines Gedankenexperiment, das mit den tatsächlich verfügbaren Daten schlicht nicht übereinstimmt. Diese nicht-stochastische Natur der verwendeten Daten bedeutet, dass zentrale Konzepte der herkömmlichen Statistik wie etwa Standardfehler oder statistische Signifikanz, die jeweils von wiederholbaren Zufallsstichproben ausgehen, keine sinnvolle Interpretation besitzen und damit streng genommen bedeutungslos sind.¹ Zum anderen führt die in vielen ländervergleichenden Studien typischerweise kleine Fallzahl dazu, dass die asymptotischen Annahmen herkömmlicher statistischer Inferenz – auf welchen wiederum so wichtige Eigenschaften wie die Unverzerrtheit von Schätzungen beruhen – nicht plausibel sind.

Mit der *bayesianischen Inferenz* steht dagegen ein statistischer Ansatz bereit, der diesen für die Politikwissenschaft symptomatischen methodologischen Gege-

¹Manchmal wird die Ansicht vertreten, Signifikanztests hätten aufgrund von zufälligen Mess- oder Codierfehlern auch eine Berechtigung bei der Analyse nicht-stochastischer Daten. Doch ist hier weder klar, auf welche Grundgesamtheit sich die „Stichprobe“ dann beziehen soll, noch über welchen konkreten datengenerierenden Prozess sich diese Fehler beschreiben lassen. Um es kurz und deutlich zu sagen, „adhering to a frequentist conception of probability in the face of non-repeatable data risks intellectual embarrassment“ (Jackman 2009, S. xxxii).

benheiten in angemessener Weise begegnet. Es handelt sich dabei um eine wahrscheinlichkeitstheoretische Herangehensweise, durch die Priori-Annahmen über interessierende Parameter (z. B. Anteils- oder Mittelwerte, Regressionskoeffizienten) mit empirischer Evidenz kombiniert werden, um dadurch zu informierten Wahrscheinlichkeitsaussagen zu gelangen – ohne dabei Voraussetzungen über die Art der Datengenerierung oder die Mindestfallzahl erfüllen zu müssen, die in der Mehrzahl der sozialwissenschaftlichen Anwendungen ohnehin nicht zu erfüllen sind. Darüber hinaus steht der moderne bayesianische Ansatz in enger Verbindung mit inferentiellen Simulationstechniken, sogenannten *Markov Chain Monte Carlo* (MCMC) Algorithmen. Diese ermöglichen es, Modelle von höherer Komplexität zu schätzen, als dies mit herkömmlichen Schätzverfahren der Fall ist. Zuletzt besticht die bayesianische Herangehensweise durch die intuitive Form und Interpretierbarkeit der durch sie erzielten Ergebnisse als einfache Wahrscheinlichkeitsaussagen.

Der rasante Anstieg an bayesianischen Methoden in den Sozialwissenschaften geht mit Anwendungen auf generalisierte lineare Modelle, Zeitreihenmodelle, Idealpunktschätzung, kausale Inferenz, Imputation fehlender Daten, Textanalyse, ökologische Inferenz, neuronale Netzwerke, Strukturgleichungsmodelle, nicht-parametrische Methoden und Faktorenanalyse einher. Diese lange Liste zeigt, dass bayesianische Inferenz nicht einfach ein weiteres „Werkzeug“ im Repertoire empirischer Forscher, sondern vielmehr eine generelle Herangehensweise an die Analyse empirischer Daten und die Schätzung statistischer Modelle darstellt (Samaniego 2010). Wir sind davon überzeugt, dass der bayesianische Ansatz zukünftig weiter an Popularität gewinnen wird, da er perfekt zu denjenigen Daten passt, die uns als Politikwissenschaftler interessieren.

Im Folgenden beschreiben wir die einzelnen Schritte des bayesianischen Inferenzprozesses und demonstrieren dessen Nützlichkeit anhand eines Beispiels aus der empirischen Demokratieforschung. Wir beginnen mit den philosophischen Unterschieden zwischen herkömmlicher und bayesianischer Statistik in Abschn. 2. Danach diskutieren wir in Abschn. 3 die Elemente des bayesianischen Inferenzprozesses und gehen auf Möglichkeiten zur Ergebnisdarstellung ein. Abschn. 4 widmet sich kurz den für die moderne bayesianische Statistik zentralen Simulationsalgorithmen, während Abschn. 5 sich dann dem ausführlichen Anwendungsbeispiel zuwendet, namentlich der Frage welchen Einfluss die staatliche Unterstützung von Religion auf religiöses Sozialkapital hat. In Abschn. 6 wagen wir einen Ausblick.

2 Zwei gegensätzliche Philosophien

Die meisten Nicht-Statistiker sind vermutlich verblüfft, dass es überhaupt so etwas wie verschiedene *philosophische Ansätze* innerhalb der Statistik gibt. Die akademische Ausbildung in den Sozialwissenschaften konzentriert sich verständlicherweise auf die Vermittlung von praktischem Handwerkzeug und verwendet nur wenig Zeit darauf, die theoretischen Grundlagen statistischer Analyse kritisch zu reflektieren. Das Problem ist freilich, dass dieser Fokus oftmals in praktische Anwendungen und Analysen mündet, die ‚suboptimal‘ sind. Als Konsequenz sind Teile der sozialwis-

senschaftlichen Forschungsliteratur irreführend und nur von geringem Nutzen für die Weiterentwicklung der Disziplin als Ganzes – und dies, obwohl sich diese Praxis leicht verbessern ließe. Überraschenderweise ist diesem Problem bislang nur wenig Aufmerksamkeit geschenkt worden. Bayesianische Datenanalyse ist für die meisten empirischen Politikwissenschaftler nach wie vor ein wenig bekanntes, wenn nicht völlig obskures Unterfangen. Wir hoffen dies mit dem vorliegenden Beitrag zu ändern.

Das Ziel statistischer Inferenz besteht ganz allgemein gesprochen darin, aus empirischen Daten zu lernen und auf diese Weise zu fundierten Aussagen über interessierende Sachverhalte zu gelangen. Bayesianische Inferenz beruht dabei auf grundlegend anderen Annahmen bezüglich der Daten und Parameter als die klassische Statistik. Der entscheidende Unterschied – aus dem sich wichtige Implikationen für die inferentielle Praxis ergeben – liegt in ihrem entgegengesetzten Verständnis von Wahrscheinlichkeit. Aus frequentistischer Sicht handelt es sich bei der Wahrscheinlichkeit eines Ereignisses oder Phänomens schlicht um die *relative Häufigkeit* mit der es unter wiederholter Betrachtung eintritt oder beobachtet wird. Wahrscheinlichkeit ist damit die objektive Eigenschaft des betrachteten Ereignisses oder Phänomens selbst. Bayesianische Inferenz basiert demgegenüber auf einem subjektiven Verständnis von Wahrscheinlichkeit. Diese gibt den *Grad der Unsicherheit* an, mit der der subjektive Glaube eines Beobachters hinsichtlich eines Ereignisses oder Phänomens behaftet ist.

Für Bayesianer lassen sich alle Größen in zwei Gruppen einteilen: jene, die beobachtet und daher gegeben sind, und jene, die unbeobachtet sind und daher geschätzt werden müssen. In der Regel handelt es sich bei den beobachteten Größen um die konkret vorliegenden Daten, deren Werte fix gegeben sind. Bei den unbeobachteten Größen handelt es sich meist um interessierende Modellparameter oder fehlende Datenwerte. In der Welt der Bayesianer werden sämtliche unbeobachteten Größen mit Verteilungsannahmen versehen und in der Analyse als Zufallsvariablen behandelt. Diese Zufallsverteilungen lassen sich wiederum in zwei verschiedene Arten unterteilen. Wenn die Verteilungsannahme bezüglich einer unbekanntem Größe nicht durch gegebene Daten bedingt ist, wird sie als *Priori-Verteilung* bezeichnet, da sie das Wissen *vor* der Betrachtung der Daten beschreibt. Wenn die Verteilungsannahme demgegenüber durch die beobachteten Daten bedingt ist, stellt sie gegenüber der unbedingten eine *aktualisierte* und damit informativere Verteilung dar. Diese Verteilung wird als *Posteriori-Verteilung* bezeichnet.

Der Unterschied zwischen dem, was als variabel, und dem, was als gegeben angesehen wird, stellt den eigentlichen Kern der Debatte zwischen Frequentisten und Bayesianern dar. Bayesianer betrachten die beobachteten Daten als fix gegeben, während unbekannte Parameter als variable Größen behandelt werden. Diesen werden Verteilungsannahmen zugeordnet, die den Grad der subjektiven Unsicherheit formalisieren. Frequentisten betrachten Daten dagegen als stochastisch und durch ein und denselben, prinzipiell unendlich wiederholbaren Datengenerierungsprozess erzeugt. Für sie sind es die zugrunde liegenden Parameter, die durch die Natur fix gegeben und daher unwandelbar sind.

Diese zweite Sichtweise eignet sich zweifellos hervorragend für die Qualitätskontrolle maschineller Produktionsprozesse oder die experimentellen Versuchsaufbauten einiger Naturwissenschaften, in denen Forscher große Mengen an unabhängigen und identisch verteilten Daten generieren können. Leider sind dies jedoch nicht die Daten, mit denen wir als Politikwissenschaftler in der Regel zu tun haben. Unsere Datensätze beziehen sich vielmehr auf einen ganz bestimmten Zeitpunkt und ganz bestimmte Umstände und werden daher niemals in dieser Form repliziert werden können. Wir können nicht in die Zeit zurückreisen und Menschen unter den exakt gleichen Voraussetzungen befragen oder auch nur annehmen, dass sich politische Einstellungen, Erfahrungen und Ereignisse bei wiederholter Erhebung nicht geändert hätten. Mit anderen Worten stellen unsere Daten einen fix gegebenen, einmaligen Blick auf das interessierende politische Phänomen dar. Was wir dagegen nicht kennen, sind die Modellparameter, die uns helfen, mögliche Beziehungen innerhalb dieser Daten zu beschreiben und zu verstehen.

Bezüglich des statistischen Frequentismus besteht einige Verwirrung. Der Begriff entstammt der Jugendzeit der Statistik, namentlich den Arbeiten von Neyman und Pearson (1928a, b, 1933a, b, 1936). Diese postulierten einen fixen, vor einem Experiment festzulegenden α -Wert, der nach vielen Wiederholungen desselben Experiments die Wahrscheinlichkeit eines Fehlers erster Art beschreibt. Da die Daten als prinzipiell unendlich sowie unabhängig und identisch verteilt angenommen werden, ist es dabei möglich, sehr präzise Hypothesen zu testen. Frequentistische Hypothesentests beziehen sich stets auf zwei mögliche Naturzustände, die den gesamten Möglichkeitsraum umfassen. Das *Verwerfen* der einen Hypothese bedeutet daher das *Akzeptieren* der anderen. Keine von beiden wird dabei als *Nullhypothese* bezeichnet.

Im Rahmen unserer akademischen Ausbildung wurde uns allerdings (richtigerweise) beigebracht, dass wir eine Hypothese niemals akzeptieren können, da stets eine unendliche Zahl an Alternativhypothesen existiert, die nicht überprüft wurden. Das bedeutet aber, dass wir unsere Hypothesentests und unser inferentielles Vorgehen eigentlich gar nicht nach frequentistischem Vorbild gestalten. Tatsächlich wäre es in den Sozialwissenschaften außerordentlich schwierig, einen wirklich frequentistischen Test durchzuführen, da unsere Untersuchungsobjekte in der Regel nicht sehr kooperativ sind, ihre Eigenschaften im Laufe der (auch sehr kurzen) Zeit ändern und somit eine Wiederholung unter exakt identischen Bedingungen äußerst unwahrscheinlich ist.

Wenn der Großteil der empirisch arbeitenden Politikwissenschaftler keine Frequentisten sind, was sind sie dann? In der sozialwissenschaftlichen Forschungspraxis ist vor allem der problematische *Nullhypothesen-Signifikanztest* (NHST) tief verwurzelt. Dieser stellt eine inkorrekte und unzulässige Mischung aus Frequentismus und *Likelihoodismus* dar. Fishers Likelihood-basierter Hypothesentest bedarf weder einer vorab erfolgenden Festlegung eines α -Niveaus, noch zweier komplementärer, sich gegenseitig ausschließender Hypothesen. Stattdessen wird eine Nullhypothese aufgestellt, die in der Fisher'schen Konstruktion lediglich dazu dient, verworfen zu werden. Aber generell wird sie als die Aussage betrachtet, dass keine spezifische Variablenbeziehung innerhalb der Daten vorliegt. Fisher betrachtete das

Vorabfestlegen von α als übertrieben rigide und auch als nicht sehr wissenschaftlich. Der NHST stellt nun eine unglückliche, weil inkonsistente Kombination dieser beiden gegensätzlichen Testlogiken dar (Gill 1999).

Glücklicherweise halten empirische Politikwissenschaftler meist nicht starr an der Logik des NHST fest, sondern gründen ihre Urteile bezüglich der in den Daten vorfindbaren Beziehungen vor allem auf *Maximum Likelihood*-basierten Inferenzschlüssen. Das macht sie zu *Likelihoodisten*. Die Pointe, auf die es uns ankommt ist nun, dass alle Likelihood-basierten Modelle nichts anderes als bayesianische Modelle sind, in denen die Priori-Verteilung einer passenden uniformen Verteilung folgt. Asymptotisch sind beide Modelle sogar bei jeder ‚properen‘ Priori-Verteilung identisch. Mit anderen Worten sind die meisten empirischen Politikwissenschaftler bereits Bayesianer – ob sie es wissen oder nicht.

3 Grundlagen der bayesianischen Inferenz

Der bayesianische Inferenzprozess besteht in der Spezifizierung von Priori-Verteilungen für alle unbekannt Parameter und der durch beobachtete Daten bedingten Aktualisierung zu Posteriori-Verteilungen. Dieser Prozess lässt sich dann mit neuen Daten wiederholen, was den bayesianischen Inferenzprozess zu einem prinzipiengeleiteten Vorgang wissenschaftlicher Wissensakkumulation macht. Im Folgenden stellen wir die einzelnen Schritte bayesianischer Inferenz genauer vor.

3.1 Die Priori-Verteilung

Der erste Schritt im bayesianischen Inferenzprozess besteht in der Formulierung von Priori-Verteilungen. Priori-Verteilungen sind Wahrscheinlichkeitsaussagen über einen interessierenden Parameter θ , die nicht von den beobachteten Daten abhängen und als $p(\theta)$ geschrieben werden. Wir könnten beispielsweise mit $p(\theta) = k$, $a < \theta < b$ eine uniforme Verteilung festlegen, die die konstante Wahrscheinlichkeit, dass θ einen Wert zwischen a und b annimmt ausdrückt. Wenn wir behaupten, dass $\theta \sim N(\mu, \sigma^2)$, dann drückt dies den Glauben aus, dass, ohne die Daten zu berücksichtigen, θ normal um μ und mit Varianz σ^2 verteilt ist. Häufig wird dabei $\mu = 0$ und σ^2 auf einen großen Wert gesetzt, um eine vage Aussage über die Bedeutsamkeit von θ auszudrücken. Es ist außerdem wichtig, dass die gewählte Priori-Verteilung denselben Wertebereich wie der Modellparameter annimmt. Soll beispielsweise im Rahmen einer Mehrebenenanalyse eine Varianzkomponente modelliert werden, würde eine zwischen $(0, \infty)$ beschränkte Priori-Verteilung gewählt werden, da Varianzen nicht negativ sein können. Dies macht etwa die (inverse) Gamma-Verteilung zu einer beliebigen Wahl in bayesianischen Spezifikationen von Varianztermen (siehe dazu Gelman 2006).

Priori-Verteilungen reichen von sehr informativen Beschreibungen früherer Forschung bis hin zu vagen Formen, die nur ein geringes Vorwissen über einen interessierenden Zusammenhang reflektieren. Dieser Teil des inferentiellen Prozes-

ses sollte jedoch nicht als mühsames Ärgernis, sondern vielmehr als Chance begriffen werden, qualitatives, narratives und intuitives Wissen auf systematische Weise in unsere statistischen Modelle zu integrieren. Neben vielen anderen Vorzügen besitzt der bayesianische Ansatz damit das Potenzial, bestehende Gräben zwischen quantitativen und qualitativen Forschern zu überbrücken (siehe hierzu den Beitrag von Beach zu Process Tracing in diesem Band; weiterführend u. a. Bennett 2008; Fairfield und Charman 2017; Humphreys und Jacobs 2015).

Historische Kontroversen bezüglich der Wahl von Priori-Verteilungen haben viele angewandte Bayesianer in den Sozialwissenschaften dazu veranlasst, nur sehr diffuse Formen zu verwenden. Aber informative Priori-Verteilungen sind äußerst nützlich, um nicht-quantitative Information in statistische Modelle einzuführen. Dabei sind zwei Dinge wichtig: a) die Verteidigung der verwendeten Informationsquelle sowie b) das Offenlegen des Einflusses dieser Priori-Verteilung im Vergleich zu einer uniformativen Referenzverteilung.

3.2 Die Likelihood-Funktion

Der zweite Schritt im bayesianischen Inferenzprozess besteht im Aufstellen der Likelihood-Funktion. Dabei wird zunächst eine parametrische Form für die Verteilung der Daten angenommen und dann die beobachteten Daten in diese eingesetzt. Dieser Schritt ist *exakt* der gleiche Vorgang wie bei Likelihood-basierten Modellen, die in den Sozialwissenschaften so häufig anzutreffen sind (Eliason 1993; King 1998). Hier sei außerdem betont, dass es sich dabei um eine ebenso subjektive wie zu rechtfertigende Entscheidung handelt, was ebenso für die Wahl der Priori-Verteilung zutrifft.

Im bayesianischen Ansatz werden die Daten \mathbf{X} als fixe Grösse behandelt und Annahmen über deren Verteilung getroffen, um den datengenerierenden Prozess in Abhängigkeit eines einzelnen oder mehrerer unbekannter Parameter θ zu beschreiben. Der bekannte *Maximum Likelihood* (ML) Schätzvorgang substituiert die begrenzte Definition von Wahrscheinlichkeit mit dem unbegrenzten Konzept der Likelihood, indem zunächst die gemeinsame Verteilung der beobachteten Daten betrachtet wird: (Gl. 1)

$$p(\mathbf{X}|\theta) = p(\mathbf{X}_1|\theta) p(\mathbf{X}_2|\theta) \dots p(\mathbf{X}_n|\theta), \quad (1)$$

wobei $p(\mathbf{X}|\theta)$ die gewählte Wahrscheinlichkeits- oder Dichtefunktion darstellt. Fisher's (1922, 1925a, b, 1930, 1934) Idee war es, diesen Term logisch umzukehren, da θ den unbekanntem Parameter darstellt, und \mathbf{X} bekannte Werte sind. Gl. (1) wird daher als $L(\theta|\mathbf{X})$ geschrieben und ist nichts Anderes als die Likelihood-Funktion. Aufmerksamen Lesern wird auffallen, dass dieser Prozess in sehr subtiler Weise der bayesianischen Logik entspricht.

Der nächste Schritt im ML-Schätzvorgang besteht darin, jene Werte von θ zu finden, die die beobachteten Daten am ehesten (*most likely*) erzeugt haben könnten. Dies ist einfach zu bewerkstelligen, da die üblicherweise verwendete Wahrschein-

lichkeits- oder Dichtefunktion in eine unimodale, zur X-Achse konkave Verteilung münden. Damit existiert ein eindeutiger Wert $\hat{\theta}$, an der Spitze der Verteilung, der die Funktion $L(\theta|\mathbf{X})$ maximiert. Dieser Wert lässt sich leicht finden: Man nehme die erste Ableitung von $L(\theta|\mathbf{X})$, um die Steigungen der Tangenten für jeden Punkt θ zu erhalten, und bestimme, wo die Tangente mit Steigung null liegt – dies ist das Maximum der Funktion.

3.3 Die Posteriori-Verteilung

Im Gegensatz zum ML-Schätzvorgang, der die Likelihoodfunktion maximiert, besteht der dritte Schritt im bayesianischen Inferenzprozess darin, die Likelihoodfunktion mit der Priori-Verteilung zu multiplizieren. Auf diese Weise wird die Posteriori-Verteilung gebildet. Dies ist nichts anderes als der Zähler des bekannten *Bayestheorems*: (Gl. 2)

$$p(\theta|\mathbf{X}) = p(\mathbf{X}|\theta) p(\theta) / p(\mathbf{X}). \quad (2)$$

Die Logik der bayesianischen Inferenz besteht also darin, subjektive Wahrscheinlichkeiten mit empirischer Evidenz zu aktualisieren, um auf diese Weise zu revidierten Wahrscheinlichkeitsaussagen zu kommen. Anders ausgedrückt: Die auf der Grundlage der Daten \mathbf{X} gewonnene Information bezüglich θ ist eine Kombination aus Vorinformation und neuer, in den Daten enthaltener Information. Die Posteriori-Verteilung stellt dabei das aktuellste Wissen über ein interessierendes Phänomen dar. Man beachte, dass in keinem Schritt Annahmen über die Größe oder Wiederholbarkeit von Stichproben gemacht werden.

Es ist wichtig zu betonen, dass die Posteriori-Verteilung alles, was wir über den interessierenden Parameter wissen, in Form einer Wahrscheinlichkeits- bzw. Dichtefunktion wiedergibt. Diese beinhaltet mehr Information als typische Zusammenfassungen von Likelihood-basierten Analysen, die nur einen Punktschätzer und die Schätzunsicherheit liefern. Tatsächlich erlaubt die Zusammenfassung in Form einer gesamten Verteilung eine weitaus vielfältigere Beschreibung. Neben dem Modus, Mittelwert oder Median könnte ebenso jedes beliebige Quantil der Verteilung von Interesse sein. Gleichzeitig müssen dafür keinerlei asymptotischen Annahmen bemüht werden, die in der politikwissenschaftlichen Forschungspraxis bei geringer Fallzahl oft nicht gegeben sind.

3.4 Zwischenzusammenfassung und ein einfaches Beispiel

Nehmen wir einmal an, wir hätten Interesse an einem einzelnen Parameter, θ , und eine Datenmatrix, \mathbf{X} . Die Schritte der bayesianischen Inferenz lassen sich dann folgendermaßen zusammenfassen: (Gl. 3)

$$\text{Posteriori-Wahrscheinlichkeit} \propto \text{Likelihoodfunktion} \times \text{Priori-Wahrscheinlichkeit}$$

$$\pi(\theta|\mathbf{X}) \propto L(\theta|\mathbf{X}) p(\theta), \quad (3)$$

wobei $\pi(\theta|\mathbf{X})$ die resultierende Posteriori-Verteilung darstellt, die sich aus dem Produkt zwischen der Priori-Verteilung, $p(\theta)$ und der Likelihoodfunktion, $L(\theta|\mathbf{X})$ ergibt. Die Schreibweise $\pi(\)$ wird hier zur besseren Abgrenzung der Posteriori-Verteilung von der Priori-Verteilung $p(\)$ verwendet. Die Verwendung des Proportionalitätszeichens (\propto) mag zunächst seltsam erscheinen. Es resultiert aus dem Weglassen des Nenners in Gl. (2). Der einzige Zweck des Nenners besteht darin, die Posteriori-Verteilung vollständig zu normalisieren, d. h. sicherstellen, dass sich das Integral über den gesamten Wertebereich auf eins summiert und die Verteilung damit eine richtige Wahrscheinlichkeitsverteilung darstellt. Da es in der Regel einfach ist, den entsprechenden Wert im Nachhinein zu ermitteln, ist es angenehmer, diese in der Berechnung zunächst zu ignorieren und erst im letzten Schritt zu *renormalisieren*.

Sobald im Forschungsprozess zusätzliche Daten beobachtet werden, kann die Posteriori-Verteilung als neue Priori-Verteilung behandelt werden. Auf diese Weise wird das Wissen über den interessierenden Parameter aktualisiert und über die Zeit akkumuliert. Nehmen wir an, $\pi_1(\theta|\mathbf{X}_1)$ ist die Posteriori-Verteilung mit Daten \mathbf{X}_1 . Später wird ein neuer Datensatz \mathbf{X}_2 gesammelt. Dann können wir $\pi_1(\theta|\mathbf{X}_1)$ als neue Priori-Verteilung für eine zweite Aktualisierung unseres Wissens nutzen. Die neue Posteriori-Verteilung ergibt sich dann mit $\pi_2(\theta|\mathbf{X}_2) \propto L(\theta|\mathbf{X}_2)\pi_1(\theta|\mathbf{X}_1)$. Interessanterweise ist dies die gleiche Verteilung, die wir erhalten würden, wenn \mathbf{X}_1 und \mathbf{X}_2 gleichzeitig vorliegen würden, und wenn die Posteriori-Verteilung auf beiden Datensätzen basieren würde, $\pi_{1,2}(\theta|\mathbf{X}_1, \mathbf{X}_2)$.

Wir können die eben beschriebene Logik an einem simplen Anschauungsbeispiel verdeutlichen. Nehmen wir an, uns interessiert der Bevölkerungsanteil in Deutschland, der sich freiwillig in religiösen Organisationen engagiert, etwa um den Armen und Kranken zu helfen oder sich anderweitig aktiv im Gemeindeleben einzubringen. Um uns darüber zu informieren, ziehen wir die Umfragedaten des *European Values Survey* (EVS) 2008 heran (siehe Abschn. 5 für eine ausführlichere und realistischere Anwendung). Da wir uns unsicher darüber sind, wie hoch der genaue Bevölkerungsanteil ist, können wir unser Wissen probabilistisch beschreiben: Die Posteriori-Verteilung $\pi(\theta|\mathbf{X})$ drückt formal aus, was wir über die mögliche Größe des Bevölkerungsanteils θ , gegeben der EVS Umfragedaten \mathbf{X} , vermuten. Um zu dieser Posteriori-Verteilung zu gelangen haben wir laut Gl. (3) nichts weiter zu tun, als eine passende Priori-Verteilung mit einer ebenfalls passenden Likelihood-Funktion zu multiplizieren und auf diese Weise unser Vorwissen mit neuen Daten zu aktualisieren.

Tatsächlich gehen wir nicht völlig ahnungslos an diese Frage heran, sondern können auf bestehendes Vorwissen aus der einschlägigen Forschungsliteratur aufbauen. Der Studie von Traunmüller und Freitag (2011, S. 255) können wir beispielsweise entnehmen, dass sich im Jahr 2004 etwa 6,5 Prozent der Deutschen aktiv an den Aktivitäten religiöser Organisationen beteiligt haben. Natürlich gehen wir nicht davon aus, dass dieser Wert *genau* unserem interessierenden Anteil entspricht. Tatsächlich ist es denkbar, dass sich die Engagementrate in den letzten Jahren

verändert hat. Zudem ist die Frageformulierung in dieser Studie so gewählt worden, dass sie auch Aktivitäten umfasst, die nicht als freiwilliges Engagement im engeren Sinne gelten. Und nicht zuletzt handelt es sich bei diesem Prozentsatz nur um einen einzelnen aus einer Stichprobe ermittelten Wert. Aus bayesianischer Perspektive erscheint es daher sinnvoll, dieses Vorwissen probabilistisch und zwar in Form einer Priori-Verteilung auszudrücken.

Eine angemessene Verteilung für Anteilswerte θ ist die Beta-Verteilung. Diese nimmt einen Wertebereich zwischen 0 und 1 an und lässt sich anhand der Parameter a und b beschreiben, wobei $a/(a + b)$ den Erwartungswert der Verteilung darstellt. Die Parameter a und b können dabei als erfolgreiche und nicht-erfolgreiche Ereignisse interpretiert werden. Gehen wir davon aus, dass die in der Studie von Traummüller und Freitag (2011) für Deutschland aufgeführte Engagementrate von 6,5 Prozent auf etwa 1000 Befragten beruhen (und lassen wir die für unsere Zwecke unerhebliche normalisierende Konstante der Beta-Verteilung weg), so lässt sich diese Priori-Informationen folgendermaßen ausdrücken: (Gl. 4)

$$p(\theta) \propto \theta^{65-1}(1 - \theta)^{935-1} \quad (4)$$

Abb. 1 oben stellt die gewählte Priori-Verteilung mit dem erwarteten Anteilswert von $65/(65 + 935) = 0,065$ grafisch dar. Es sei angemerkt, dass wir unserem Vorwissen in diesem illustrativen Beispiel eine sicherlich viel zu hohe Bedeutung beimessen, wie sich an der begrenzten Spannweite zwischen 0,04 und 0,09 für den vermuteten Wert ablesen lässt.

Als nächstes müssen wir eine passende Likelihood-Funktion wählen, um die neu vorliegende empirische Evidenz zu beschreiben. Da wir an einem binären Merkmal (engagiert vs. nicht engagiert) interessiert sind, eignet sich eine Binomial-Verteilung. Diese gibt abhängig von der Erfolgswahrscheinlichkeit – diese ist zugleich unser interessierender Parameter θ – die Anzahl erfolgreicher Ereignisse aus einer gewissen Anzahl an Versuchen wieder. Für Deutschland liegen im EVS 2008 Informationen zu insgesamt $N = 2075$ Befragten – d. h. Versuchen – vor, wovon 75 angeben, sich freiwillig für eine religiöse Organisation zu engagieren – was den erfolgreichen Ereignissen entspricht. Damit ergibt sich – abermals ohne normalisierende Konstante – folgende Likelihood-Funktion: (Gl. 5)

$$L(\theta|\mathbf{X}) \propto \theta^{75}(1 - \theta)^{2075-75} \quad (5)$$

die in der Mitte von Abb. 1 dargestellt ist und ihr Maximum bei 0,036 hat. Mit anderen Worten gehen laut im EVS 2008 enthaltener empirischer Evidenz rund 3,6 Prozent aller Deutschen einem freiwilligen Engagement in einer religiösen Organisation nach.

Die Posteriori-Verteilung, die sich bis zu einer normalisierenden Konstanten aus dem Produkt zwischen Priori-Verteilung und Likelihood-Funktion ergibt (Gl. 6)

$$\pi(\theta|\mathbf{X}) \propto \theta^{75}(1 - \theta)^{2075-75} \theta^{65-1}(1 - \theta)^{935-1} = \theta^{139}(1 - \theta)^{2934} \quad (6)$$

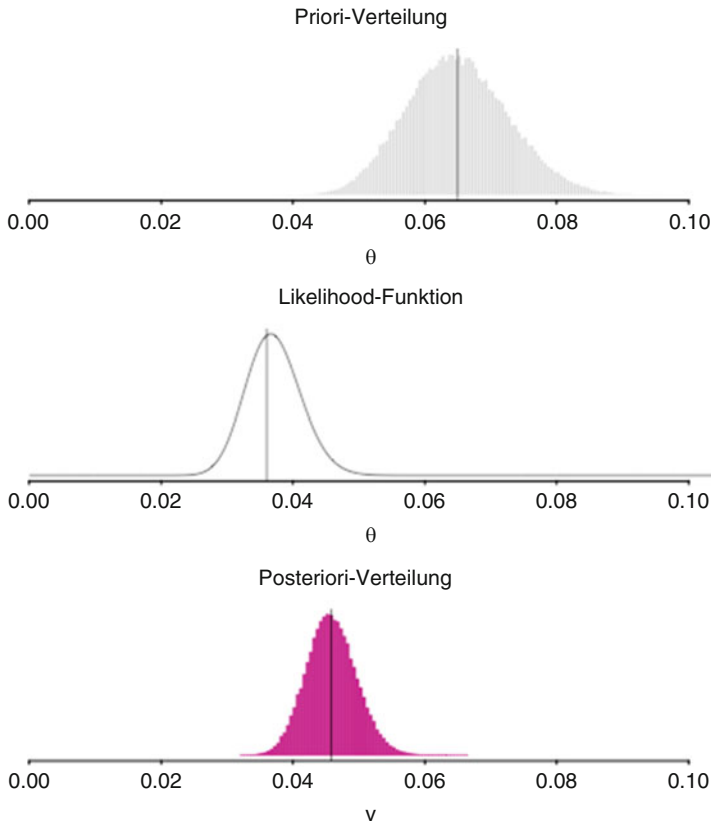


Abb. 1 Der bayesianische Inferenzprozess am Beispiel des Bevölkerungsanteils in Deutschland, der sich freiwillig in religiösen Organisationen engagiert. Die Posteriori-Verteilung kann als (gewichteter) Kompromiss zwischen Priori-Verteilung und Likelihood betrachtet werden

ist ebenfalls eine Beta-Verteilung. Sie besitzt die Parameter $a = 139$ und $b = 2934$. Aktualisieren wir unser Priori-Wissen mit der neuen empirischen Evidenz kommen wir durch den bayesianischen Inferenzprozess also zu der Einschätzung, dass sich rund 4,5 Prozent der deutschen Bevölkerung freiwillig in einer religiösen Organisation engagiert ($139/(139 + 2934) = 0,045$). Dieser Anteil liegt zwischen dem Priori-Anteil und dem Anteil, der sich aus den neuen Daten ergibt. Das macht deutlich, dass es sich bei der Posteriori-Verteilung um einen gewichteten Kompromiss zwischen unserem Vorwissen und der gegebenen empirischen Evidenz handelt.

3.5 Die Beschreibung bayesianischer Ergebnisse

Im Gegensatz zum problematischen NHST mit seinen p-Werten, Sternchen und willkürlichen Signifikanzniveaus werden die Ergebnisse einer bayesianischen Ana-

lyse am besten dargestellt, indem man die Posteriori-Verteilung auf verschiedene Weisen zusammenfasst. In der Regel geschieht dies mittels Quantilen und entsprechenden Wahrscheinlichkeitsaussagen. Diese werden dann als Wahrscheinlichkeit interpretiert, dass sich der interessierende Parameter innerhalb eines bestimmten Wertebereichs befindet, oder als Wahrscheinlichkeit, dass er kleiner oder größer als eine bestimmte Konstante ist, beispielsweise $p(\theta > 0)$. Dies führt zu sehr intuitiv verständlichen und nützlichen Aussagen, die frei von statistischem Jargon sind. Wie etwa der folgenden: „Die Wahrscheinlichkeit, dass Kirchensteuern die Spendenbereitschaft für religiöse Organisationen in der Bevölkerung verringern, liegt bei 93 Prozent.“²

Es gibt aber keinen Grund, die Posteriori-Verteilung nicht einfach anhand ihres Mittelwerts und ihrer Standardabweichung – diese entspricht dem Standardfehler in konventionellen Analysen – zu beschreiben. Autoren können also in gewohnter Weise einfache Regressionstabellen erstellen, außer natürlich mit der Ausnahme, dass Sternchen für Signifikanzniveaus jetzt keinen Sinn mehr ergeben. Die Ergebnisse einer bayesianischen Analyse können also in zugänglicher Form präsentiert werden. Es ist nicht notwendig, neue oder exotische Darstellungskriterien zu erlernen.

Es gibt auch ein bayesianisches Analog zum klassischen Konfidenzintervall, das sogenannte *Kredibilitätsintervall*. Dieses wird auf die exakt gleiche Art und Weise wie ein Konfidenzintervall konstruiert – also Punktschätzer plus/minus kritischer Wert mal Standardfehler. Unterschiedlich ist jedoch die Interpretation. Diese entspricht nämlich genau jener Interpretation, die statistische Anfänger dem klassischen Konfidenzintervall geben, und die ihm viele angewandte Forscher gerne geben würden: die Wahrscheinlichkeit, dass ein Parameter innerhalb zweier Grenzen liegt. Das frequentistische Konfidenzintervall ist aber vielmehr ein Intervall, das bei mehreren *Replikationen des exakt gleichen Experiments* den wahren Parameterwert mit einer durchschnittlichen Wahrscheinlichkeit von $1-\alpha$ enthält.

Es existiert auch ein noch flexibleres bayesianisches Konfidenzintervall, das als *highest posterior density* (HPD) bezeichnet wird. Das ist jene Region im Wertebereich von θ , die die größte $1-\alpha$ Posteriori-Wahrscheinlichkeitsdichte enthält – egal ob diese kontinuierlich ist oder nicht. Das HPD kann daher bei multimodalen Posteriori-Verteilungen aus mehreren Intervallen bestehen. Genau wie bei frequentistischen Konfidenzintervallen impliziert ein HPD, das nicht den Wert Null enthält, dass der geschätzte Koeffizient als statistisch reliabel angesehen werden kann. Aber anstatt konfident zu sein, dass $(1-\alpha)$ Prozent der zu wiederholenden Intervalle den wahren Parameterwert enthalten, liefert das HPD die $(1-\alpha)$ -prozentige Wahrscheinlichkeit, dass der wahre Parameterwert innerhalb dieses konkret vorliegenden Intervalls liegt.

²Dieser Wert ist hypothetisch und dürfte laut der Studie von Traummüller und Freitag (2011) sogar höher liegen. Die Autoren berichten neben dem Koeffizienten mit negativen Vorzeichen jedoch nur zwei Sternchen **, die einem p-Wert $< 0,01$ entsprechen.

3.6 Wir alle sind Bayesianer

So andersartig der soeben beschriebene Inferenzprozess des bayesianischen Paradigmas zunächst auch erscheinen mag, so beinhaltet er Likelihood-basierte Inferenz doch als Sonderfall. Wie eingangs erwähnt, können wir in der Politikwissenschaft aufgrund der Eigenheiten unseres Untersuchungsgegenstandes keine Frequentisten sein. Der eigentlich interessierende Vergleich betrifft demnach den standardmäßigen Likelihood-basierten Ansatz auf der einen und den bayesianischen Ansatz auf der anderen Seite. Das Ergebnis einer ML-Schätzung und das Ergebnis aus einem bayesianischen Modell mit uniformen Priori-Verteilungen sind jedoch äquivalent. Asymptotisch sind die beiden sogar für *jede beliebige* Priori-Verteilung äquivalent: Beide nähern sich bei steigender Samplegröße einer Normalverteilung an. Außerdem ist in vielen Fällen die konkrete Wahl der Priori-Verteilung nicht von Bedeutung, da mit steigender Samplegröße die Likelihood zunehmend an Gewicht gewinnt und letztendlich jeden vernünftigen Prior dominiert.

Obwohl die Wahl einer Priori-Verteilung als sehr subjektive Entscheidung angesehen werden kann, gibt es oftmals gute Argumente für eine spezifische Form der Priori-Verteilung. Geringes oder kein Vorwissen führt zu diffusen oder uniformen Priors, bestimmte Wahrscheinlichkeitsmodelle legen ganz logisch bestimmte Formen (konjugierte Priors) nahe, und andere Priori-Verteilungen erlauben es dem Forscher, zusätzliche Information jenseits der vorliegenden Studie mit in die Analyse einfließen zu lassen. Ganz unabhängig davon sind freilich *alle* statistischen Modelle letztlich subjektiv.

4 Bayesianische stochastische Simulation

Wie bisher aus theoretischer Sicht dargestellt ist bayesianische Inferenz an sich ein relativ einfaches Unterfangen. Eine Herausforderung in der praktischen Anwendung besteht darin, das erzielte Ergebnis – die Posteriori-Verteilung $\pi(\theta|\mathbf{X})$ – zusammenzufassen und zu beschreiben. Dies erfordert, das Integral dieser konditionalen Wahrscheinlichkeitsverteilung zu berechnen. Bis auf die allertrivialsten Modelle ist dies analytisch unmöglich. Für viele komplexe, d. h. realistische sozialwissenschaftliche Modelle ist $\pi(\theta|\mathbf{X})$ hoch mehrdimensional und folgt keiner bekannten Form. Man kann dann zur Berechnung von Erwartungswerten und Varianzen nicht auf bestehende Formeln zurückgreifen.

Die Lösung für dieses Problem besteht in der Anwendung von Simulationstechniken. Diese erlauben es, Posteriori-Verteilungen zu beschreiben, indem der analytische Integrationsprozess durch iterative Rechenschritte ersetzt wird. Zu diesem Zweck wird in der modernen bayesianischen Statistik insbesondere auf *Markov Chain Monte Carlo* (MCMC) Algorithmen zurückgegriffen. Diese Schätzmaschinerie ist zweifelsohne eines der mächtigsten Werkzeuge der gegenwärtigen Statistik. MCMC wurde durch einen Überblicksartikel von Gelfand und Smith (1990) in die allgemeine statistische Literatur eingeführt – nachdem es über vierzig Jahre unbeachtet im Bereich der statistischen Physik vergraben lag. Bayesianische stochasti-

sche Simulation – so die allgemeine Bezeichnung für MCMC – ersetzt analytische Berechnungen mit Stift und Papier sowie standardmäßige numerische Optimierungsverfahren mit einem iterativen Rechenprozess, der multidimensionale Posteriori-Verteilungen zu explorieren und beschreiben imstande ist, deren Integration sich als schlicht unmöglich erweisen würde.

Die grundlegende Idee des *Monte Carlo* Prinzips besteht darin, dass man alles über eine Wahrscheinlichkeitsverteilung lernen kann, indem man wiederholt zufällig generierte Werte aus ihr zieht („simuliert“), und das Sample von Werten zu ihrer Beschreibung verwendet. Je mehr solcher Werte gezogen werden, desto genauer lässt sich eine Verteilung beschreiben. *Markov Chain* Algorithmen wiederum generieren spezielle Sequenzen von Werten – sogenannte Markovketten – die es ermöglichen, auch komplexe hochdimensionale Verteilungen zu beschreiben, indem diese in ihre Bestandteile aus einfacheren Verteilungen zerlegt werden. Um also multidimensionale Wahrscheinlichkeitsstrukturen zu schätzen, wird eine Markovkette in Gang gesetzt. Diese wandert durch den Parameterraum und besucht dabei alle Wertebereiche mit einer Häufigkeit proportional zu den ihnen zugrunde liegenden Wahrscheinlichkeitsdichten. Wenn sie dann einige Zeit in dieser spezifischen Verteilung verbringt, können wir aus den simulierten Werten Statistiken wie Mittelwerte, Varianzen und Quantile berechnen. Nachdem jede Iteration der Kette eine multidimensionale Position darstellt, ist eine Marginalisierung der gemeinsamen Posteriori-Verteilung äquivalent zur Einzelbetrachtung jeder der Einzeldimensionen. Dabei bestehen einige Herausforderungen darin, die Konvergenz der Kette zu beurteilen, eine effiziente Exploration des gesamten Parameterraums zu erzielen, sowie die bedingten Wahrscheinlichkeiten abzuleiten (Gill 2002, 2008a, b).

Es gibt zwei grundlegende MCMC-Algorithmen. *Gibbs Sampling* zieht Werte für jeden Modellparameter iterativ aus dem gesamten Set bedingter Wahrscheinlichkeiten (Geman und Geman 1984). Der Nutzer stellt für jeden Parameter eine bedingte Wahrscheinlichkeitsfunktion auf, in der dieser auf der linken Seite der Bedingung steht und alle anderen Parameter sowie die Daten auf der rechten Seite der Bedingung. Für Parameter θ_k und einen Parametervektor ohne diesen Parameter θ_{-k} sieht dies folgendermassen aus: $\pi(\theta_k | \theta_{-k}, \mathbf{X})$. Der *Metropolis-Hastings*-Algorithmus vollzieht dagegen bei jeder Iteration eine einzige multidimensionale Bewegung, indem eine Kandidatenposition aus einer *Proposal Distribution* gezogen und je nach Qualität der Ziehung eine Annahme- oder Ablehnungsentscheidung getroffen wird (Metropolis und Ulam 1949; Metropolis et al. 1953; Hastings 1970). Immer wenn die posteriori Wahrscheinlichkeit der Kandidatenposition höher ist als jene der gegenwärtigen Position der Markovkette, dann wird der Zug angenommen. Wenn die Posteriori-Wahrscheinlichkeit allerdings kleiner ist als die gegenwärtige Position, basiert die Annahme des Zuges auf dem Wahrscheinlichkeitsverhältnis $\pi(\theta_{\text{Kandidatenposition}}) / \pi(\theta_{\text{gegenwärtige Position}})$. Anders als der *Gibbs Sampler* kann der *Metropolis-Hastings*-Algorithmus also die Bewegung zu einer neuen Position ablehnen und die gegenwärtige Position als neuen Schritt der Markovkette wählen.

Abb. 2 hilft, den MCMC Prozess besser zu verdeutlichen. Hier wird der Samplepfad für die ersten beiden Parameter des in Teil 5 geschätzten Modells verfolgt. Nachdem die Posteriori-Verteilung hochdimensional ist, lässt sich die Gesamtstruk-

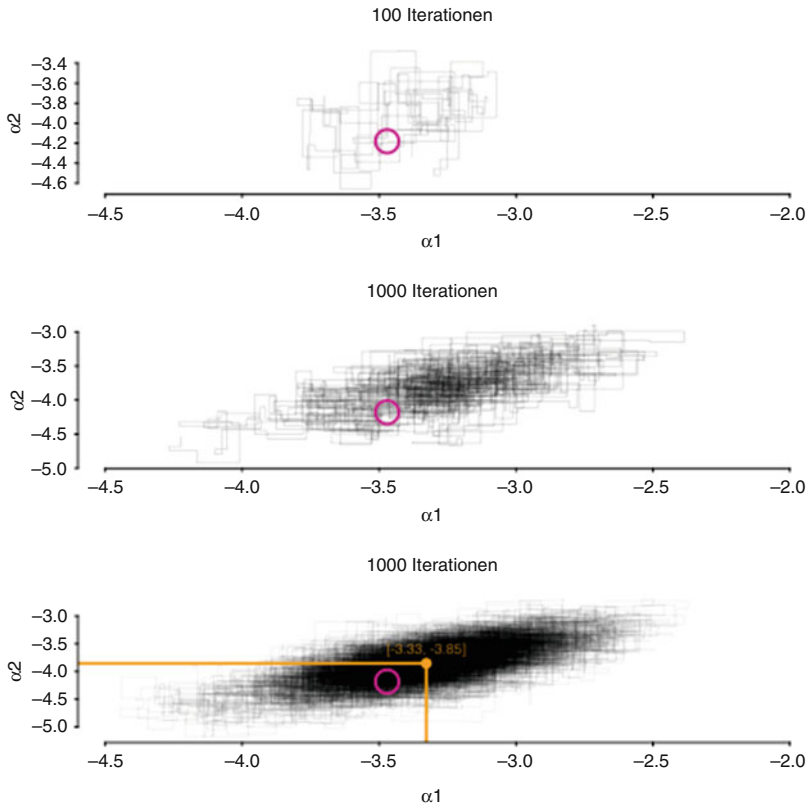


Abb. 2 Die ersten 100, 1000 und 10.000 Schritte des *Gibbs Samplers* für zwei ausgewählte Parameter. Die Markovkette wurde mit den zufällig generierten Startwerten $[-3,47; -3,22]$ in Gang gesetzt (Kreis). Schon nach 10.000 Iterationen scheint die zweidimensionale Posteriori-Verteilung recht gut exploriert. Die Posteriori-Mittelwerte für die beiden Parameter liegen bei $\alpha_1 = -3,33$ und $\alpha_2 = -3,85$

tur nicht darstellen. Die Auswahl zweier Dimensionen zeigt jedoch das grundlegende Verhalten der Markovkette und kann auf mehr Dimensionen übertragen werden. Dargestellt sind die ersten 100, 1000 und 10.000 Schritte der Markovkette, die mit den zufällig generierten Startwerten $\alpha_1 = -3,47$ und $\alpha_2 = -3,22$ in Gang gesetzt wurde – was in der Darstellung mit einem Kreis eingezeichnet ist. Der Algorithmus wandert im zweidimensionalen Parameterraum umher, wobei er stets orthogonale Bewegungen vollzieht. Dies ergibt sich aus den getrennten Zügen aus den konditionalen Verteilungen des Gibbs Samplers. Jeder Wert wird in einem iterativen Prozess einzeln gezogen. In diesem Beispiel wird erst α_1 gezogen, dann α_2 , dann wieder α_1 , dann wieder α_2 und immer so weiter. Schon nach 1000 Iterationen hält sich die Kette am häufigsten in der Gegend oben rechts des Startwerts (etwa bei $\alpha_1 = -3,25$ und $\alpha_2 = -4,00$) auf und scheint bereits nach 10.000

Schritten die zweidimensionale Posteriori-Verteilung gut exploriert zu haben. Die Posteriori-Mittelwerte für die beiden Parameter liegen bei $\alpha_1 = -3,33$ und $\alpha_2 = -3,85$

All dies mag programmiertechnisch eher aufwendig erscheinen. Allerdings existiert inzwischen sehr gute kostenlose Software, die dem Nutzer erhebliche Lasten abnimmt. Sowohl bei WinBUGS („*Windows Bayesian inference Using Gibbs Sampling*“) (Spiegelhalter et al. 1999) als auch JAGS („*Just Another Gibbs Sampler*“) (Plummer 2010a) muss der Nutzer lediglich die Modellsyntax definieren. Die Programme übersetzen diese dann in die für Gibbs Sampling benötigten bedingten Wahrscheinlichkeiten und ziehen aus ihnen. Beide Programme wechseln auch automatisch zu Metropolis-Hastings, sollte sich der Gibbs Sampler für einige Parameter als ungeeignet herausstellen.

MCMC ist weitaus bedeutender als ML, auch wenn wir das bislang noch nicht wissen – von Fishers wichtigen Schriften zur ML-Schätzung bis hin zur vollständigen Beschreibung ihrer Eigenschaften durch Birnbaum (1962) dauerte es immerhin 40 Jahre. Die Gründe dafür liegen auf der Hand: MCMC liefert die gleiche Information wie ML, nämlich den Modus sowie die Wölbung um den Modus, es liefert Information über die gesamte Posteriori-Verteilung, so dass interessierende Größen wie Quantile und *Bayesfactors* bestimmt werden können und der *Prozess* liefert unterwegs zusätzliche Information, vor allem bei bayesianischen nicht-parametrischen Verfahren (siehe hierzu Gill und Casella 2009 oder Traummüller et al. 2015). Vor allem aber lassen sich mit MCMC Modelle schätzen, bei denen herkömmliche Schätzverfahren scheitern (Gill 2008b; Jackman 2009).

5 Ein empirisches Beispiel aus der empirischen Demokratieforschung

Wir demonstrieren die Nützlichkeit des bayesianischen Ansatzes an einem Beispiel aus der empirischen Demokratieforschung. Wir lassen *a priori* Informationen aus früheren Studien in eine neue Analyse zum Einfluss staatlicher Unterstützung von Religion auf die Engagementbereitschaft in religiösen Organisationen einfließen. Konkret soll untersucht werden, inwieweit staatliche Unterstützung von Religionsgemeinschaften in der Form von Privilegien, Subventionen und Finanzierungen einen förderlichen oder hinderlichen Einfluss auf das religiöse Sozialkapital einer Gesellschaft, d. h. auf ehrenamtliches Engagement in religiösen oder kirchlichen Organisationen, ausübt.

Bislang bestehen zum Zusammenhang von Staat-Kirche-Beziehungen und sozialem Kapital erst sehr wenige empirische Studien, deren Ergebnisse sich auch noch widersprechen. Während etwa Roßteutscher (2009) in einer Studie von lokalen Zivilgesellschaften in acht europäischen Ländern zu dem Schluss kommt, dass vom Staat gewährte Privilegien und Subventionen die Vitalität religiöser Zivilgesellschaften fördern, finden Traummüller und Freitag (2011) in einem Vergleich von 24 EU-Mitgliedstaaten Hinweise für einen negativen Einfluss des Staat-Kirche-Verhältnisses. Wie wir zeigen werden, lässt sich diese widersprechende Information

aus früheren Studien im Rahmen des bayesianischen Ansatzes in eine neue Analyse mit neuen Daten inkorporieren wobei diese Information dazu genutzt wird, ein Modell zu konstruieren und Priori-Verteilungen festzulegen.

5.1 Datengrundlage

Bei den neuen Daten, mit welchen wir unser Wissen über den Zusammenhang von staatlicher Förderung von Religion und religiösem Sozialkapital aktualisieren, handelt es sich um Umfragedaten für $N = 36.355$ Befragte der neuesten Welle des EVS 2008. Diese werden mit Makroindikatoren für die $J = 24$ europäischen Länder verknüpft, die auch Traummüller und Freitag (2011) betrachtet haben.³ Als abhängige Variable zur Messung der Engagementbereitschaft im religiösen Zivilgesellschaftssektor verwenden wir einen binären Indikator, der angibt, ob der Befragte gegenwärtig ehrenamtlich, das heißt ohne Bezahlung, für eine religiöse oder kirchliche Organisation tätig ist. Die zentrale erklärende Größe der Analyse – staatliche Unterstützung von Religion – setzt sich aus einer Vielzahl von verschiedenen gesetzlichen Regelungen sowie aus einer Reihe direkter und indirekter Formen der Finanzierung und Subvention zusammen. Um diese in einer kompakten Maßzahl zusammenzufassen, greifen wir auf den IRI *Government Favoritism Index* 2008 von Grim und Finke (2006) zurück. Selbstverständlich können neben den Staat-Kirche-Beziehungen noch weitere Faktoren einen Einfluss auf die Vitalität des religiösen Zivilgesellschaftssektors haben. Wie Traummüller und Freitag (2011) kontrollieren wir daher für potenziell bedeutsame Größen auf Individual- und Länderebene. Neben dem Geschlecht, dem Alter, der Bildung, dem Erwerbstatus sowie dem Wohnort berücksichtigen wir insbesondere die individuelle Religiosität der Befragten. Auf Kontextebene kontrollieren wir das BIP pro Kopf sowie die religiöse Kulturtradition, namentlich den Protestantenanteil eines Landes.

5.2 Das Modell

Wir schätzen eine bayesianische Version der Modellspezifikation von Traummüller und Freitag (2011), namentlich ein einfaches bayesianisches logistisches Mehrebenenmodell: (Gl. 7)

$$\Pr(y_i = 1) = \text{logit}^{-1}(\alpha_{j[i]} + \beta \mathbf{X}_i), \quad \text{für } i = 1, \dots, \text{Befragte} \quad (7)$$

$$\alpha_j \sim N(\boldsymbol{\gamma} \mathbf{Z}_j, \sigma^2), \quad \text{für } j = 1, \dots, 31 \text{ Länder}$$

³Angaben zu einzelnen Variablen fehlen bei rund 15 Prozent aller Befragten. Wir haben wir fehlende Werte mittels multipler Imputation (Little und Rubin 2002) unter Zuhilfenahme des *mice* packages in R imputiert.

wobei die Wahrscheinlichkeit ehrenamtlich in einer religiösen Organisation tätig zu sein, $y_i = 1$, eine logistische Funktion der individuellen Merkmale \mathbf{X}_i und den korrespondierenden Koeffizienten $\boldsymbol{\beta}$ darstellt. Wir lehnen uns bei den länderspezifischen Konstanten $\alpha_{[j]}$ an die Notation von Gelman und Hill (2007) an, um zu verdeutlichen, dass sich Individuum i in Land j befindet. Auf der zweiten Ebene werden diese länderspezifischen Konstanten nun ihrerseits als normalverteilte Variable mit einem sich aus den Ländermerkmalen \mathbf{Z}_j und Koeffizienten $\boldsymbol{\gamma}$ ergebenden Mittelwert sowie einer Varianz von σ^2 modelliert.

Das bayesianische Mehrebenenmodell wird durch Priori-Verteilungsannahmen für alle unbekannt Parameter vervollständigt, wobei wir uns auf sogenannte konjugierte Formen von Priori-Verteilungen beschränken, d. h. Normalverteilungen für die Koeffizienten (Gl. 8)

$$\boldsymbol{\beta} \sim N(b, S_\beta) \quad (8)$$

$$\boldsymbol{\gamma} \sim N(g, S_\gamma)$$

und eine inverse Gamma-Verteilung für die Varianz auf Länderebene (Gl. 9):

$$\sigma^2 \sim IG(s_1, s_2). \quad (9)$$

Hier wird die natürliche Affinität des bayesianischen Ansatzes zu hierarchischen Modellstrukturen deutlich, bei denen Parameter der Individualebene ein eigenes Modell auf Gruppenebene erhalten. Tatsächlich ist hierarchische Modellierung ein fester Bestandteil der bayesianischen Herangehensweise, da hier Parameter stets als variabel aufgefasst und mit einer Priori-Verteilung versehen werden.⁴

Um unser Vorwissen bezüglich der unbekannt Parameter in die Analyse einfließen zu lassen, wählen wir informierte Priors, die wir aus den beiden vorangegangenen Studien von Roßteutscher (2009) sowie Traummüller und Freitag (2011) ableiten. Nachdem sich diese beiden Studien in ihren Aussagen bezüglich des Koeffizienten für die staatliche Unterstützung von Religion widersprechen, formulieren wir konkurrierende Priori-Annahmen und aktualisieren diese mit den neuen Daten. Für das erste Set an Priori-Annahmen wählen wir Normalverteilungen, in die wir die Punktschätzungen aus den Ergebnissen von Traummüller und Freitag (2011) als Mittelwerte sowie die dazugehörigen Standardfehler (mal zwei multipliziert) als Standardabweichungen einsetzen. Für den Koeffizienten des GFI 2008 wäre dies $\gamma \sim N(-0,124, 0,04x2)$. Da sich das Untersuchungsdesign von Roßteutscher (2009) stärker von dem hier gewählten unterscheidet, formalisieren wir hier die

⁴Die bayesianische Herangehensweise bietet auch eine Lösung für Probleme der ML Schätzung von Mehrebenenmodellen. Das Hauptproblem ist, dass die Schätzung zentraler Parameter von Punktschätzungen konditionierender Parameter abhängt, deren Schätzunsicherheit nicht angemessen berücksichtigt wird (Raudenbush und Bryk 2002; Seltzer et al. 1996). Die Folge sind zu enge Konfidenzintervalle oder sogar verzerrte Koeffizienten. Im bayesianischen Ansatz wird die Unsicherheit bezüglich aller Parameter berücksichtigt.

Annahme eines positiven Einfluss staatlicher Unterstützung von Religion mit einer diffusen Normalverteilung mit Mittelwert von 0,5, was einem sehr großen positiven Koeffizienten entspricht: $\gamma \sim N(0, 5, 1)$. Für die Koeffizienten der Kontrollvariablen und die Varianzkomponente wählen wir dagegen uninformative Priors. Zudem kontrastieren wir die von verschiedenen Priori-Annahmen ausgehenden Modelle noch mit einem Referenzmodell, das kein Vorwissen bezüglich des Koeffizienten für staatliche Unterstützung ausdrückt: $\gamma \sim N(0, 1)$.

5.3 Modellschätzung und Ergebnisse

Im Folgenden generieren wir die marginalisierten Posteriori-Verteilungen der uns interessierenden Koeffizienten. Wir verwenden dafür die in JAGS implementierten MCMC Algorithmen, die wir über das `rjags` Paket der statistischen Programmiersprache R steuern (Plummer 2010a, b; R Development Core Team 2009). Der Vorteil von JAGS besteht darin, dass wir die in Abschn. 4 beschriebenen einzelnen konditionalen Wahrscheinlichkeitsverteilungen nicht selbst herleiten müssen, sondern dies das Programm erledigen lassen.

Die gemeinsame Posteriori-Verteilung ergibt sich, wie oben in Gl. (3) dargestellt, aus der Kombination der Likelihoodfunktion für das logistische Mehrebenenmodell mit den verschiedenen Priori-Verteilungen: (Gl. 10)

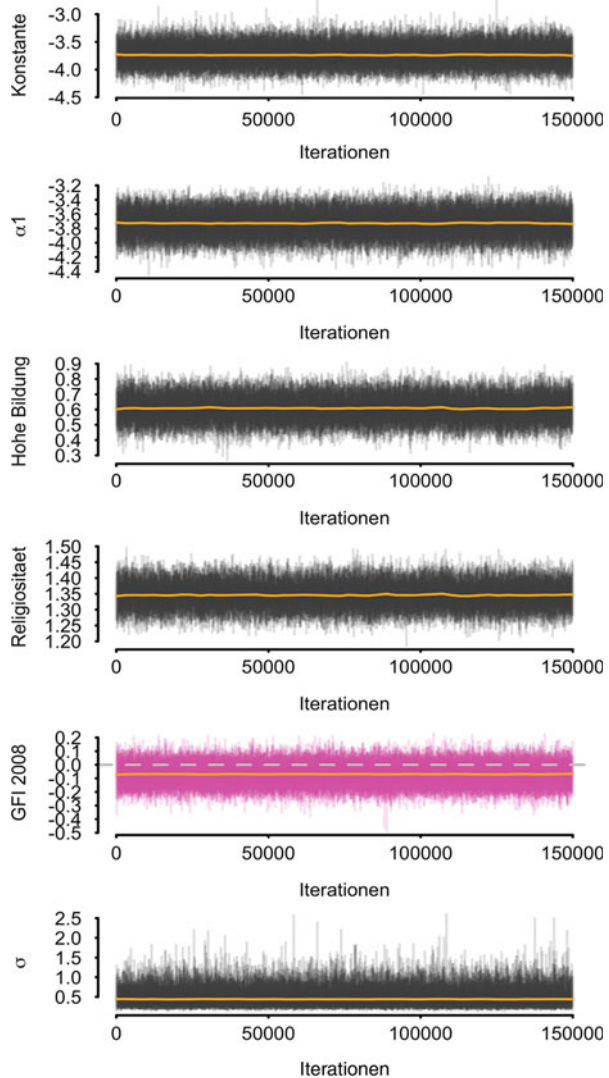
$$\pi(\beta, \gamma, \alpha, \sigma^2 | y, X, Z) \propto L(y, X, Z | \beta, \gamma, \alpha, \sigma^2) p(\alpha | \gamma, \sigma^2) p(\beta, \gamma, \sigma^2). \quad (10)$$

Es handelt sich dabei um eine hochdimensionale Verteilung – mit acht Koeffizienten auf Individualebene, drei Koeffizienten und einer Varianzkomponente auf Länderebene sowie 24 länderspezifischen Konstanten – so dass eine analytische Beschreibung der einzelnen marginalen Verteilungen eine Integration über 36 Dimensionen erfordern würde. Dies ist ganz offensichtlich unpraktisch, und die oben beschriebenen MCMC Simulationstechniken helfen uns dies zu vermeiden.

Für jede Modellspezifikation lassen wir eine einzige Markovkette für insgesamt 250.000 Iterationen laufen. Obwohl es theoretisch unerheblich ist, mit welchen Parameterwerten die Kette startet, und die Startwerte zufällig gewählt werden können, haben wir die Schätzungen aus einem herkömmlichen ML-Modell als Startwerte für den Simulationsalgorithmus gewählt, um den Konvergenzprozess zu beschleunigen. Da es eine Weile dauert, bis sich der Algorithmus auf die Parameterwerte der Posteriori-Verteilung einpendelt, verwerfen wir die ersten 100.000 Iterationen als „burn-in“. Die verbleibenden Iterationen unterziehen wir einer Konvergenzdiagnostik mittels des R Pakets `superdiag` (Tsai und Gill 2012), das alle gängigen MCMC Diagnostiken implementiert (z. B. Geweke 1992; Raftery und Lewis 1992; Gelman und Rubin 1992; Heidelberger und Welch 1983; siehe Gill 2008a für eine ausführliche Diskussion von MCMC Diagnostiken). Es gibt keine Anzeichen dafür, dass der Simulationsalgorithmus nicht konvergiert wäre.

Dies legt auch die in Abb. 3 dargestellte grafische Inspektion der Samplepfade für sechs ausgewählte Modellparameter nahe. Hier sehen wir das typische „Raupen-

Abb. 3 Trace Plots für die letzten 150.000 Iterationen des Samplepfads sechs ausgewählter Modellparameter. Die Simulationen stammen aus der bayesianischen Modellspezifikation mit uninformativen Priori-Annahmen aus Tab. 1. Das typische „Raupenmuster“ und die Stationarität legen nahe, dass der MCMC-Algorithmus keinerlei Konvergenzprobleme aufweist



muster“ sowie die klare Stationarität der Simulationsschritte. Dies deutet darauf hin, dass sich der MCMC Algorithmus in der gewünschten Posteriori-Verteilung eingefunden hat und diese effektiv exploriert. Das gibt uns die Zuversicht, dass wir die uns interessierenden Modellparameter mit den verbleibenden 150.000 Iterationen in angemessener Weise beschreiben können.

Die in Tab. 1 präsentierten Schätzgrößen sind die marginalisierten a posteriori Mittelwerte, Standardabweichungen und 95 % Kreditabilitätsintervalle für die einzelnen Modellparameter. Diese ergeben sich aus der einfachen Zusammenfassung der letzten 150.000 simulierten Werte des Gibbs Samplers. Die Resultate

Tab. 1 Ergebnisse der bayesianischen Hierarchischen Logistischen Modelle mit verschiedenen A-priori-Annahmen

	Uninformative Priors			Informierte Priors aus Traum Müller und Freitag 2011			Informierte Priors aus Roßteutscher 2009		
	Mittelwert	SA	0,025 % 0,975 %	Mittelwert	SA	0,025 % 0,975 %	Mittelwert	SA	0,025 % 0,975 %
Weiblich	-0,001	0,052	[-0,087 0,085]	0,000	0,052	[-0,086 0,086]	-0,001	0,052	[-0,087 0,085]
Alter/10	0,005	0,015	[-0,020 0,030]	0,005	0,015	[-0,020 0,030]	0,005	0,015	[-0,020 0,030]
Mittlere Bildung	0,178	0,065	[0,071 0,285]	0,179	0,066	[0,070 0,288]	0,177	0,066	[0,068 0,286]
Hohe Bildung	0,608	0,070	[0,492 0,724]	0,611	0,071	[0,494 0,728]	0,608	0,071	[0,491 0,725]
Arbeitslos	-0,271	0,131	[-0,487 -0,055]	-0,276	0,132	[-0,494 -0,058]	-0,271	0,131	[-0,487 -0,055]
Stadt	-0,283	0,053	[-0,370 -0,196]	-0,283	0,053	[-0,370 -0,196]	-0,283	0,053	[-0,370 -0,196]
Religiosität	1,346	0,034	[1,290 1,402]	1,346	0,034	[1,290 1,402]	1,345	0,034	[1,289 1,401]
GFI 2008	-0,069	0,062	[-0,171 0,033]	-0,069	0,053	[-0,156 0,018]	-0,067	0,062	[-0,169 0,035]
BIP 2008	0,011	0,007	[-0,000 0,023]	0,013	0,006	[0,002 0,024]	0,011	0,007	[-0,000 0,023]
Protestantenanteil	0,008	0,005	[0,000 0,016]	0,008	0,005	[0,000 0,016]	0,008	0,005	[0,000 0,016]
Konstante	-3,735	0,155	[-3,991 -3,479]	-3,754	0,137	[-3,980 -3,528]	-3,733	0,155	[-3,989 -3,477]
σ^2	0,467	0,167	[0,191 0,743]	0,331	0,101	[0,164 0,498]	0,466	0,165	[0,194 0,738]
J Länder	24			24			24		
N Individuen	36.355			36.355			36.355		

Anmerkung: Ergebnisse aus drei bayesianischen Logistischen Mehrebenenmodellen für freiwilliges Engagement in religiösen Organisationen mit jeweils drei verschiedenen Prior-Annahmen für die Modellparameter. Dargestellt sind Mittelwerte, Standardabweichungen sowie 95 % Kreditibilitätsintervalle der Posteriori-Verteilungen. Inferenz basiert jeweils auf 150.000 MCMC Iterationen. Herkömmliche Diagnostiken geben keine Hinweise auf Nichtkonvergenz der Simulationsalgorithmen

bayesianischer Datenanalysen werden häufig in Form von Wahrscheinlichkeitsaussagen berichtet oder grafisch dargestellt (siehe unten). Doch die meisten Leser erwarten in der Regel eine standardmäßige Regressionstabelle. Es ist wichtig, daran zu erinnern, dass der Mittelwert einer Posteriori-Verteilung nicht das gleiche wie der Modus einer Likelihood-Funktion ist. Allerdings ist die Differenz bei nur geringfügig informativen Priors und großen Fallzahlen vernachlässigbar. Das gleiche gilt auch für die posteriori Standardabweichung und den Standardfehler.

Tab. 1 lässt sich genauso wie eine Ergebnistabelle einer ML-Schätzung lesen, außer in Fällen, in denen der Prior einflussreich oder die Samplegröße gering ist. In einem solchen Fall spielen die getroffenen Annahmen eine wichtige Rolle und müssen ausführlich gerechtfertigt werden. Mit insgesamt $N = 36.355$ Fällen und diffusen konjugierten Priors wie in unserem Beispiel ist dies jedoch nicht notwendig, da die Daten die Priori-Annahmen klar dominieren. Tatsächlich unterscheiden sich die Ergebnisse fast überhaupt nicht zwischen den verschiedenen Priori-Annahmen. Lediglich die Varianzkomponente für die Länderunterschiede erweist sich als sensibel für die gewählte Priori-Verteilung. Während der posteriori Mittelwert für σ^2 in den beiden uniformierten Fällen auf 0,47 geschätzt wird und das 95-Prozent-Kreditintervall zwischen 0,19 und 0,74 liegt, beträgt die Ländervarianz unter Berücksichtigung der Priori-Information aus Traummüller und Freitag (2011) nur 0,33. Dieser Posteriori-Wert liegt ziemlich genau zwischen der aus den früheren Ergebnissen stammenden Priori-Annahme von 0,22 und dem durch die Daten bestimmten Wert von 0,47. Er stellt somit einen klaren Kompromiss aus Vorwissen und neuer empirischer Evidenz dar. Auch das Kreditintervall fällt enger aus ([0,16, 0,50]).

Ein Blick auf die individuellen Erklärungsgrößen zeigt, dass weder das Geschlecht noch das Alter der Befragten für freiwilliges Engagement in religiösen Organisationen bedeutsam ist. Dies unterscheidet unsere Befunde von Traummüller und Freitag (2011), die für das Jahr 2004 sowohl einen Geschlechts- als auch Alterseffekt ausmachen konnten. Diese gegensätzliche Information, die wir in der Form von Priori-Annahmen in die Analyse der neuen Daten für 2008 haben einfließen lassen, macht sich jedoch aufgrund der hohen Fallzahl und damit der Stärke der neuen empirischen Evidenz kaum bemerkbar. Umgekehrt finden wir im Gegensatz zu Traummüller und Freitag (2011) einen Stadt-Land-Unterschied. Städter weisen geringeres religiöses Sozialkapital auf als Landbewohner, wobei der entsprechende Koeffizient mit 95 prozentiger Wahrscheinlichkeit zwischen $-0,37$ und $-0,19$ liegt. In Einklang mit den früheren Befunden steigt das Engagement mit dem Bildungsgrad und sinkt mit individueller Arbeitslosigkeit. Nicht überraschend ist zudem, dass Religiosität in einem starken Zusammenhang mit Freiwilligentätigkeit im religiösen Kontext steht.

Auch für die Erklärungsfaktoren auf Länderebene erweisen sich die unterschiedlichen Priori-Spezifikationen als wenig folgenreich. Die entsprechenden marginalen Posteriori-Verteilungen sind nahezu identisch. Wie schon in Traummüller und Freitag (2011) beschrieben, neigen die Menschen in wirtschaftlich stärkeren Ländern zu mehr freiwilligem Engagement. Zusätzlich finden wir entgegen der früheren Ergeb-

nisse nun auch einen positiven Zusammenhang zwischen dem Protestantenanteil und dem glaubensbasierten Sozialkapital.

Interessant ist der Befund für den uns primär interessierenden Erklärungsfaktor der staatlichen Unterstützung von Religion. Auch hier unterscheiden sich die posteriori Mittelwerte trotz diametral entgegengesetzter Priori-Annahmen nicht voneinander, sondern liegen bei $-0,07$, was auf einen negativen Zusammenhang hinweist. Je stärker organisierte Religion vom Staat unterstützt und privilegiert wird, desto geringer ist die Bereitschaft der Bürger, sich freiwillig für religiöse Organisationen zu engagieren. Dies war auch das zentrale Ergebnis von Traummüller und Freitag (2011). Allerdings beinhalten die hier berichteten 95-Prozent-Kreditintervalle alle den Wert Null, so dass nach konventionellen Kriterien von keinem statistisch reliablen Zusammenhang ausgegangen werden kann – von statistischer Signifikanz zu sprechen hat aus bayesianischer Sicht ohnehin keinen Sinn.

Heißt das nun, dass die staatliche Involvierung in religiöse Belange bedeutungslos für den religiösen Sektor der Zivilgesellschaft ist? Dies ist wohl der Schluss, der in weiten Teilen der gegenwärtigen sozialwissenschaftlichen Forschungspraxis gezogen werden würde. Der bayesianische Ansatz erlaubt es uns aber, zu einer differenzierteren Sichtweise zu gelangen. Wie in Abb. 4 illustriert, können wir die gesamte Posteriori-Verteilung betrachten und probabilistische Aussagen darüber treffen, wie wahrscheinlich es ist, dass der Koeffizient des *Government Favoritism Index* kleiner als Null ist – und folglich einen negativen Zusammenhang anzeigt. Wie wir sehen, ist unabhängig von den gewählten Priori-Annahmen die Wahrscheinlichkeit sehr hoch, dass staatliche Unterstützung von Religion mit geringerem freiwilligem Engagement in religiösen Organisationen einhergeht. Tatsächlich liegt sie, wenn man die Priori-Information aus Traummüller und Freitag (2011) berücksichtigt, bei immerhin 91 Prozent. Selbst wenn man diesen früheren Befunden eher skeptisch eingestellt ist und von einem positiven Effekt staatlicher Unterstützung ausgeht, ist der Zusammenhang immer noch mit 87 prozentiger Wahrscheinlichkeit negativ.

6 Fazit

Bayesianische Datenanalyse lässt sich anhand von drei Eigenschaften charakterisieren: a) der Bereitschaft, unbekannte Parameter mit Priori-Verteilungen zu versehen, b) der Anwendung des Bayestheorems, um zu Posteriori-Verteilungen für unbekannte Parameter zu gelangen, und c) die Beschreibung von erzielten Inferenzschlüssen mittels probabilistischer Aussagen. Die wichtigste philosophische Grundlage bayesianischer Inferenz ist es, sowohl unbeobachtete Größen als auch Modellparameter als Zufallsvariablen zu betrachten. Der Hauptvorteil dieses Ansatzes besteht in der prinzipiell unbegrenzten Modellierung komplexer Modelle mit mehreren Ebenen und einer Vielzahl von unbekanntem Parametern. Da die Modellannahmen im bayesianischen Ansatz sehr viel deutlicher offengelegt werden, kann auch die Modellqualität besser beurteilt werden.

Bayesianische Modelle lassen sich allerdings nicht nach einem Rezept entwickeln. Stattdessen verlangt jeder einzelne Schritt – von der Formulierung der Priori-

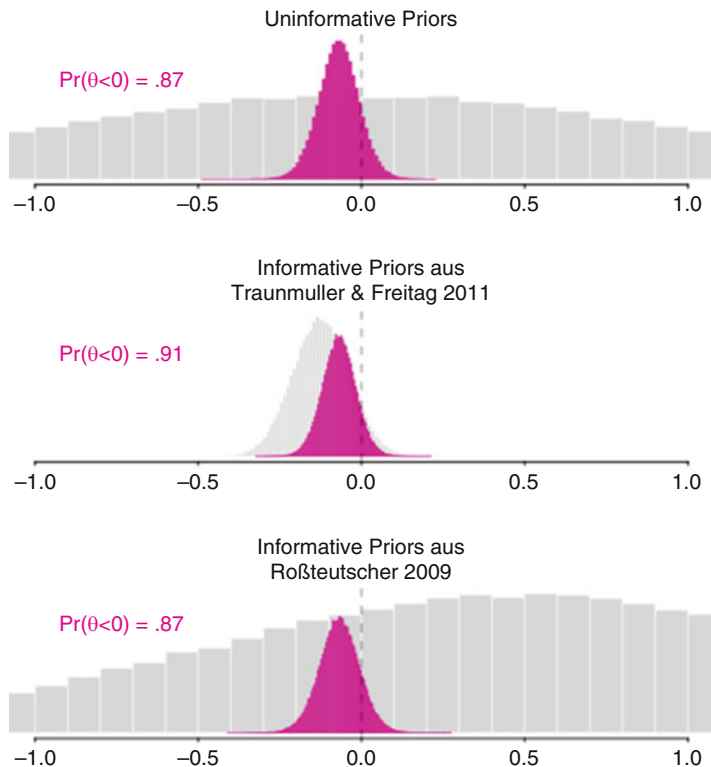


Abb. 4 Posteriori-Verteilung (purpur) für den Koeffizienten des GFI 2008 für verschiedene Priori-Annahmen (grau) sowie die Wahrscheinlichkeit, dass dieser Koeffizient negativ ist ($\Pr(\theta < 0)$). Aufgrund der hohen Fallzahl fallen die unterschiedlichen Priori-Annahmen kaum ins Gewicht und die Posteriori-Verteilung wird vor allem durch die Daten bestimmt. Obwohl nach konventionellen Kriterien nicht von einem statistisch reliablen Zusammenhang zwischen staatlicher Unterstützung von Religion und religiösem Sozialkapital ausgegangen werden kann, ist die Posteriori-Wahrscheinlichkeit für einen negativen Einfluss mit 87 bzw. 91 Prozent doch sehr hoch

Annahmen und der Likelihood-Funktion bis hin zur Modellschätzung und Ergebnisbeschreibung – wohl überlegte und begründete Entscheidungen, die für den Konsumenten der Analyse klar nachvollziehbar sein müssen. Bayesianer lassen eine solche Sorgfalt schon allein aus historischen Gründen walten, da sie für ihren Ansatz etwa hundert Jahre lang von Vertretern anderer Ansätze angefeindet wurden. Dies führte zu einer ausgeprägten Defensivhaltung, die sich in einer detaillierten Verteidigung sämtlicher Modellannahmen manifestiert. Natürlich wird ein solches Vorgehen prinzipiell von allen Entwicklern statistischer Modelle erwartet, und die weitverbreitete, unreflektierte Routinisierung und Ritualisierung statistischer Modellbildung schadet vielen Disziplinen, auch der Politikwissenschaft. Wir sind überzeugt, dass die Einführung der bayesianischen Inferenzkultur in die empirische Politikwissenschaft zu besseren wissenschaftlichen Erkenntnisprozessen und weg von gleichschrittartigen, lediglich auf Konventionen beruhenden Prozeduren führt, die in den Sozialwissenschaften so häufig anzutreffen sind.

Natürlich geben wir uns nicht der Illusion hin, dass sich alle Politikwissenschaftler in absehbarer Zeit in angewandte Statistiker verwandeln werden. Wir glauben auch nicht, dass sie das sollten. Unsere Kollegen sind berechtigterweise vornehmlich an Theorien politischer Strukturen und Prozesse interessiert, die sie mittels gesammelter Daten empirisch untermauern und in bestehende Diskussionszusammenhänge einordnen möchten. Der Schritt der statistischen Modellierung sollte aber auch nicht als mühsames Pflichtprogramm betrachtet, sondern vor allem als Chance zur kreativen Exploration verstanden werden. Gegenwärtig wird ein solcher Erkenntnisprozess aber durch überkommene Praktiken eingeengt, die die Entwicklung unserer Disziplin bremsen.

Politikwissenschaftler wenden zunehmend bayesianische Methoden an, um ihre vielfältigen empirischen und methodologischen Probleme auf nützliche Weise zu lösen. Im Feld der Statistik sind die Kontroversen um den bayesianischen Ansatz im Verlauf der letzten Jahrzehnte stark zurückgegangen, wie ein Blick in führende Statistikjournale zeigt. Eine Reihe an frei verfügbaren MCMC-Werkzeugen lässt den bayesianischen Schätzvorgang mittlerweile auch unter scheinbar schwierigen Bedingungen zu einer einfach zu bewerkstelligen Aufgabe werden. Dies führt uns in eine Welt, in der Politikwissenschaftler kaum daran gehindert werden, nützliche und prinzipiengeleitete bayesianische Modelle für ihre empirischen Forschungsfragen zu entwickeln.

Bayesianische Methoden sind natürlich kein Allheilmittel für die quantitative Politikwissenschaft. Es ist genauso möglich, fehlerhafte bayesianische Modelle zu konstruieren, wie dies bei nicht-bayesianischen Spezifikationen der Fall ist. Forscher müssen sich im Rahmen der Modellkonstruktion immer noch sorgfältige Gedanken über die Messung der Daten, die Beziehungen zwischen den Parametern und die Beschreibung der diesbezüglichen Unsicherheit machen. Das bayesianische Paradigma bietet dabei aber einen von klaren Prinzipien geleiteten Ansatz, diese bestehende Unsicherheit auszudrücken. Aus unserer Sicht sind vielleicht nicht alle bayesianischen Prozeduren gut, aber alle guten Prozeduren bayesianisch.

7 Weiterführende Literatur

Zu Einführungen in Buchlänge, welche sich explizit an ein sozialwissenschaftliches Publikum richten, zählen Gill (2002), Lynch (2007) und Jackman (2009). Ein guter Überblicksartikel, welcher insbesondere auch den Nutzen von informierten Prioriverteilungen verdeutlicht, ist Western und Jackman (1994).

Literatur

Bennett, Andrew. 2008. Process tracing: A Bayesian Perspective. In *The Oxford Handbook of political methodology*, Hrsg. Janet M. Box-Steffensmeier, Henry E. Brady und David Collier, 702–721. Oxford: Oxford University Press.

- Birnbaum, Allan. 1962. On the foundations of statistical inference. *Journal of the American Statistical Association* 57:269–306.
- Eliason, Scott R. 1993. *Maximum likelihood estimation: logic and practice*. Newbury Park: Sage.
- Fairfield, Tasha, and Andrew E. Charman. 2017. Explicit Bayesian analysis for process tracing: Guidelines, opportunities, and caveats. *Political Analysis* 25(3): 363–380.
- Fisher, Ronald A. 1922. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Statistical Society of London A* 222:309–360.
- Fisher, Ronald A. 1925a. *Statistical methods for research workers*. Edinburgh: Oliver and Boyd.
- Fisher, Ronald A. 1925b. Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society* 22:700–725.
- Fisher, Ronald A. 1930. Inverse probability. *Proceedings of the Cambridge Philosophical Society* 26:528–535.
- Fisher, Ronald A. 1934. *The Design of Experiments*, 1. Aufl. Edinburgh: Oliver and Boyd.
- Gelfand, Alan E., und Adrian F. M. Smith. 1990. Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85:398–409.
- Gelman, Andrew. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1:514–534.
- Gelman, Andrew, und Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.
- Gelman, Andrew, und Donald Rubin. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457–511.
- Geman, Stuart, und Donald Geman. 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:721–741.
- Geweke, John. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Bayesian statistics*, Hrsg. J. M. Bernardo, A. F. M. Smith, A. P. Dawid und J. O. Berger, Bd. 4, 169–193. Oxford: Oxford University Press.
- Gill, Jeff. 1999. The insignificance of null hypothesis significance testing. *Political Research Quarterly* 52(3): 647–674.
- Gill, Jeff. 2002. *Bayesian methods. A social and behavioral sciences approach*. Boca Raton: Chapman/Hall/CRC.
- Gill, Jeff. 2008a. Is partial convergence a problem for inferences from MCMC algorithms? *Political Analysis* 16(2): 153–178.
- Gill, Jeff. 2008b. *Bayesian methods. A social and behavioral sciences approach*, 2. Aufl. Boca Raton: Chapman/Hall/CRC.
- Gill, Jeff, und George Casella. 2009. Nonparametric priors for ordinal Bayesian social science models: Specification and estimation. *Journal of the American Statistical Association* 104: 453–464.
- Grim, Brian J., und Roger Finke. 2006. International religion indexes: Government regulation, government favoritism, and social regulation of religion. *Interdisciplinary Journal of Research on Religion* 2:2–40.
- Hastings, W. Keith. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.
- Heidelberger, Philip, und Peter D. Welch. 1983. Simulation run length control in the presence of an initial transient. *Operations Research* 31:1109–1144.
- Humphreys, Macartan, und Alan Jacobs. 2015. Mixing methods: A Bayesian approach. *American Political Science Review* 109(4): 653–673.
- Jackman, Simon. 2009. *Bayesian analysis for the social sciences*. Chichester: Wiley.
- King, Gary. 1998. *Unifying political methodology. The likelihood theory of statistical inference*. Ann Arbor: University of Michigan Press.
- Little, Roderick J. A., und Donald B. Rubin. 2002. *Statistical analysis with missing data*, 2. Aufl. New York: John Wiley & Sons.

- Lynch, Scott M. 2007. *Introduction to applied Bayesian statistics and estimation for social scientists*. New York: Springer.
- Metropolis, Nicholas, and Stanislaw Ulam. 1949. The Monte Carlo method. *Journal of the American Statistical Association* 44:335–341.
- Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, und E. Teller. 1953. Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 21:1087–1091.
- Neyman, Jerzy, und Egon S. Pearson. 1928a. On the use and interpretation of certain test criteria for purposes of statistical inference. Part I. *Biometrika* 20A:175–240.
- Neyman, Jerzy, und Egon S. Pearson. 1928b. On the use and interpretation of certain test criteria for purposes of statistical inference. Part II. *Biometrika* 20A:263–294.
- Neyman, Jerzy, und Egon S. Pearson. 1933a. On the problem of the most efficient test of statistical hypotheses. *Philosophical Transactions of the Royal Statistical Society, Series A* 231:289–337.
- Neyman, Jerzy, und Egon S. Pearson. 1933b. The testing of statistical hypotheses in relation to probabilities a priori. *Proceedings of the Cambridge Philosophical Society* 24:492–510.
- Neyman, Jerzy, und Egon S. Pearson. 1936. Sufficient statistics and uniformly most powerful tests of statistical hypotheses. *Statistical Research Memorandum* 1:113–137.
- Plummer, Martyn. 2010a. *JAGS Version 2.2.0 User Manual*.
- Plummer, Martyn. 2010b. *rjags: Bayesian graphical models using MCMC*. R package version 2.1.0-2.
- R Development Core Team. 2009. *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing.
- Raftery, Adrian, und Steven Lewis. 1992. How many iterations in the Gibbs sampler? In *Bayesian statistics*, Hrsg. J. M. Bernardo, A. F. M. Smith, A. P. Dawid und J. O. Berger, Bd. 4, 763–773. Oxford: Oxford University Press.
- Raudenbush, Stephen W., und Anthony S. Bryk. 2002. *Hierarchical linear models*. Thousand Oaks: Sage.
- Roßteutscher, Sigrid. 2009. *Religion, Zivilgesellschaft, Demokratie. Eine internationale vergleichende Studie zur Natur religiöser Märkte und der demokratischen Rolle religiöser Zivilgesellschaften*. Baden-Baden: Nomos.
- Samaniego, Francisco J. 2010. *A comparison of the Bayesian and frequentist approaches to estimation*. New York: Springer.
- Seltzer, Michael H., Wing Hung Wong, und Anthony S. Bryk. 1996. Bayesian analysis in applications of hierarchical models: Issues and methods. *Journal of Educational and Behavioral Statistics* 21:131–167.
- Spiegelhalter, David J, Andrew Thomas, und Nicky G, Best. 1999. *WinBUGS Version 1.2 User Manual*. MRC Biostatistics Unit.
- Traunmüller, Richard, und Markus Freitag. 2011. State support of religion: Making or breaking faith-based social capital? *Comparative Politics* 43(3): 253–269.
- Traunmüller, Richard, Andreas Murr, und Jeff Gill. 2015. Modeling latent information in voting data with Dirichlet process priors. *Political Analysis* 23(1): 1–20.
- Tsai, Tsung-han, und Jeff Gill. 2012. superdiag: A comprehensive test suite for Markov chain non-convergence. *The Political Methodologist* 19:12–18.
- Western, Bruce, und Simon Jackman. 1994. Bayesian inference for comparative research. *American Political Science Review* 88:412–423.