

Grundlagen und Anwendungen Bayesianischer Statistik

Julius M. Pfadt

27.09.2023

Inside every non-Bayesian, there is a Bayesian struggling to get out.

Dennis V. Lindley

Inhalt

- Einführung
- Grundlagen
- Parameter schätzen
- Hypothesen testen
- Exkurs: JASP
- Modelle auswählen
- Zusammenfassung

Vorstellung: Julius M. Pfadt

- Bisher: Dr. rer. nat. 2023 an der Universität Ulm:
 - Thema: Die Entwicklung und Implementierung Bayesianischer Versionen populärer Reliabilitätskoeffizienten
- Zurzeit: Post-Doc an der Universität Tilburg:
 - Thema: Die Implementierung von dynamischen sozialen Netzwerkmodellen in Open-Source Software
- Bald: Post-Doc an der Universität Amsterdam:
 - Thema: Bayesian Model Averaging für Strukturgleichungsmodelle



universität
uulm



UNIVERSITY
OF AMSTERDAM

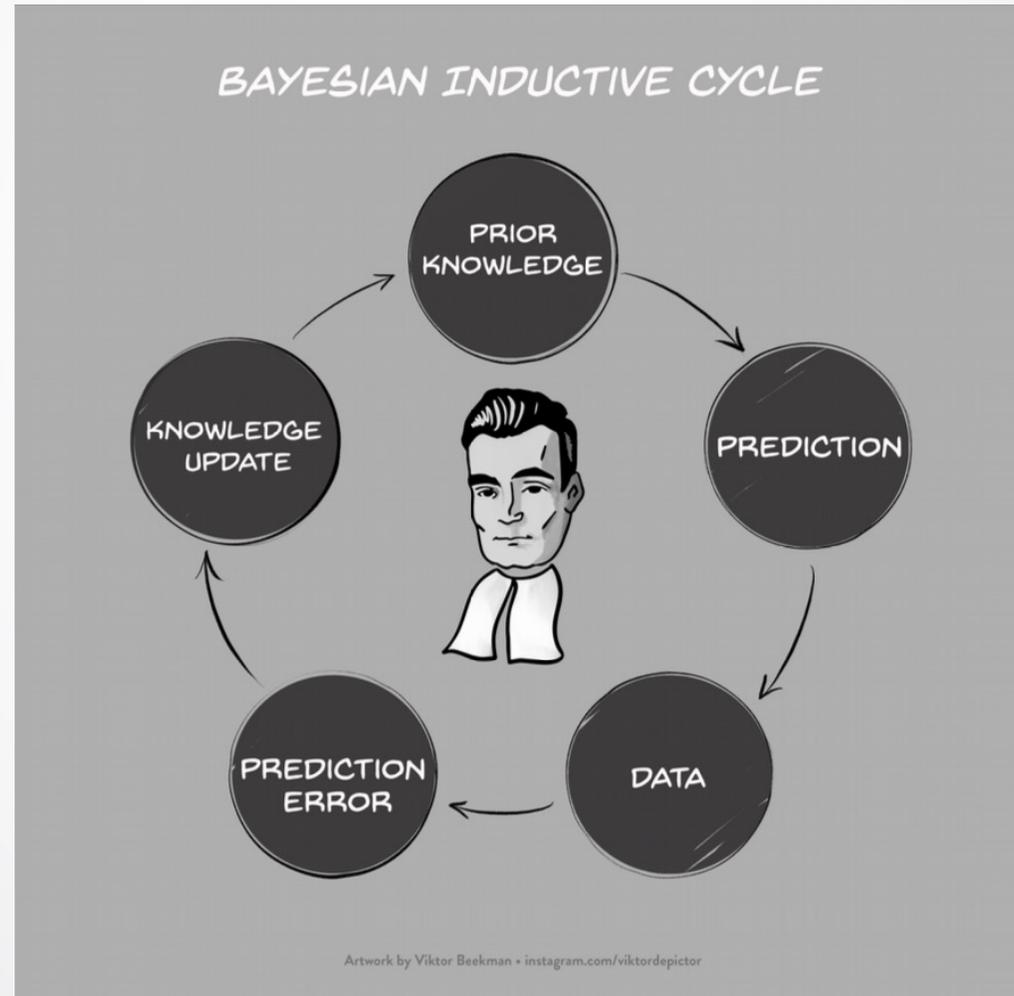
Ziele

- Sie sollen nicht:
 - Überzeugte Bayesians werden
 - Alles verlernen was man Ihnen beigebracht hat
 - Sehen wie man Bayesianische Verfahren in R oder SPSS anwendet
- Sie sollen:
 - Die wichtigsten Bayesianischen Konzepte kennenlernen:
 - Prior
 - Posterior
 - Bayes Factor
 - Anwendungen und Vorteile Bayesianischer Verfahren verstehen (Bayesian benefits)
 - Beispielhaft sehen wie man Bayesianische Verfahren in JASP anwenden kann

Warum Statistik?

- Benutzen der Daten, um Aussagen über die Population zu treffen
- z.B.: Wie wirksam ist eine neue Therapieform in der Population?
- Verschiedene statistische Philosophien um Schlüsse aus den Daten zu ziehen:
 - Klassische (frequentistische) Statistik
 - Bayesianische Statistik:
 - andere Inferenzen als in klassischer Statistik
- In Bayesianischer Statistik wird Unsicherheit modelliert
- Bayesianische Statistik ist kumulativ

Bayesian Cycle

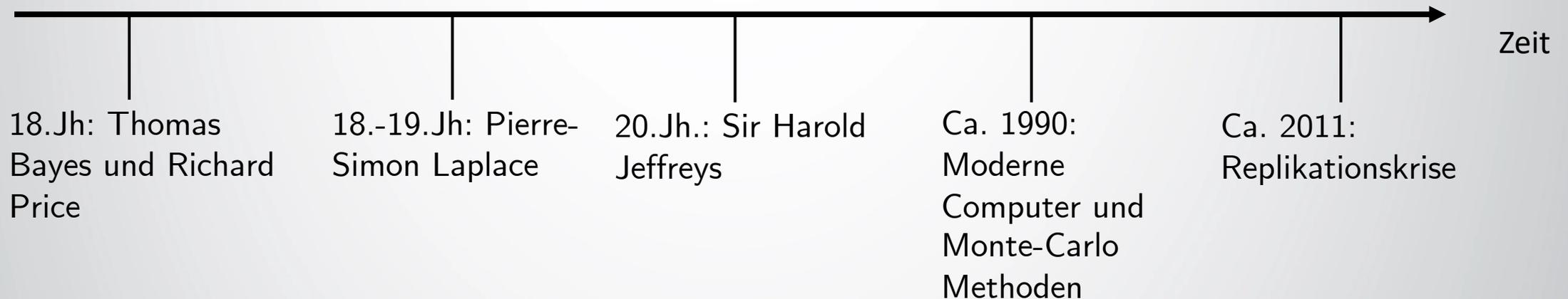


Bayesian Benefits

- Vorwissen optimal nutzen (häufiger Kritikpunkt)
- Wahrscheinlichkeit, dass ein Parameter größer ist als Wert X , z.B.:
Wahrscheinlichkeit, dass IQ in der Population größer als 100 ist?
- 95% Wahrscheinlichkeit, dass ein Parameter innerhalb bestimmter Intervallgrenzen liegt
- Wahrscheinlichkeiten zu Hypothesen: Wie wahrscheinlich ist es, dass die Null-/Alternativhypothese die Daten generiert hat?
- Wie groß ist ein Parameter, z.B. Regressionskoeffizient, über mehrere konkurrierende Modelle hinweg?

Eine Bayesianische Revolution?

The image shows three overlapping document covers. The leftmost cover is from Elsevier, titled "Another failure to replicate" by Jelte M. Wicherts and Conor V. Dolan. The middle cover is from Memory & Cognition, titled "A failure to replicate rapid" by Caoimhe M. Harrington Stack and Ariel N. The rightmost cover is from the Journal of Personality and Social Psychology, titled "Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011)" by Eric-Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, and Han L. J. van der Maas.



Eine Bayesianische Revolution!

- Erhebliche Zunahme Bayesianischer Methoden in den letzten 20 Jahren
- Softwarepakete: R, JASP, Jamovi, Stan, JAGS
- Anwendungen in vielen Disziplinen:
 - Forschung
 - Marketing
 - Finanzmärkte
 - Wettervorhersage

Inhalt

- Einführung
- **Grundlagen**
- Parameter schätzen
- Hypothesen testen
- Exkurs: JASP
- Modelle auswählen
- Zusammenfassung

Wahrscheinlichkeit

- Ist eine Münze fair (50% für Kopf)?
- Was ist die Wahrscheinlichkeit, dass die Münze auf Kopf landet?
- Wie bestimmen wir die Wahrscheinlichkeit?

Wahrscheinlichkeit – Klassisch

- Wahrscheinlichkeit eines Ereignisses ist der Grenzwert der relativen Häufigkeit in vielen Wiederholungen
- Ich werfe die Münze sehr oft und die relative Häufigkeit für Kopf entspricht der Wahrscheinlichkeit
- Nebenbemerkung: Die Wahrscheinlichkeit, dass die Münze auf derselben Seite landet ist größer als 50%

Wahrscheinlichkeit – Klassisch

- Grundlage: Stichprobenverteilung der Teststatistik:
 - Verteilung einer Teststatistik (bspw. Mittelwert) in vielen Wiederholungen
- Nullhypothese und Stichprobenverteilung:
 - Unter der Nullhypothese folgt die Teststatistik einer bestimmten Stichprobenverteilung
 - p-Wert: Wahrscheinlichkeit die Teststatistik (oder extremere) zu beobachten unter der Nullhypothese
 - Die Wahrscheinlichkeit können wir nur berechnen, da wir die relative Häufigkeit eines Ereignisses (Beobachtung der Teststatistik) als Wahrscheinlichkeit interpretieren
- Konfidenzintervall:
 - Intervall, welches den Parameter (z.B. Mittelwert) in X% der Fälle umfasst, wenn wir das Intervall viele Male auf die gleiche Art bilden, sprich Daten erheben und das X% Intervall berechnen

Wahrscheinlichkeit – Bayesianisch

- Ist eine Münze fair?
- Ich glaube Ja!
- „Subjektive“ Wahrscheinlichkeit
- Jedoch, ist die Wahrscheinlichkeit repräsentiert durch eine Verteilung, wir implizieren also Unsicherheit bei der „subjektiven“ Beurteilung der Wahrscheinlichkeit

Wahrscheinlichkeitsverteilungen

- Bei Wahrscheinlichkeitsverteilungen entspricht das Integral der Verteilung der Wahrscheinlichkeit
- Das Integral ist die Fläche unter der Funktionskurve
- Das Integral ist eine mathematische Größe, die wir berechnen oder approximieren können
- z.B.: Die Wahrscheinlichkeit, dass der IQ zwischen -1 und $+1$ SD um den Mittelwert liegt ist 68%

Bayes Theorem

- Bayes Theorem benutzt Gesetze der Wahrscheinlichkeitstheorie (bedingte Wahrscheinlichkeit) um Quantitäten abzuleiten, die uns interessieren:

$$\begin{array}{c} \text{Ws von A} \\ \text{gegeben B} \\ \\ p(A|B) \end{array} = \frac{\begin{array}{c} \text{Ws von B} \\ \text{gegeben A} \\ \\ p(B|A) \end{array} \cdot \begin{array}{c} \text{Ws von A} \\ \\ \\ p(A) \end{array}}{\begin{array}{c} p(B) \\ \\ \text{Ws von B} \end{array}}$$

Bayes Theorem

- M := Modell; θ := Parameter; y := Daten;

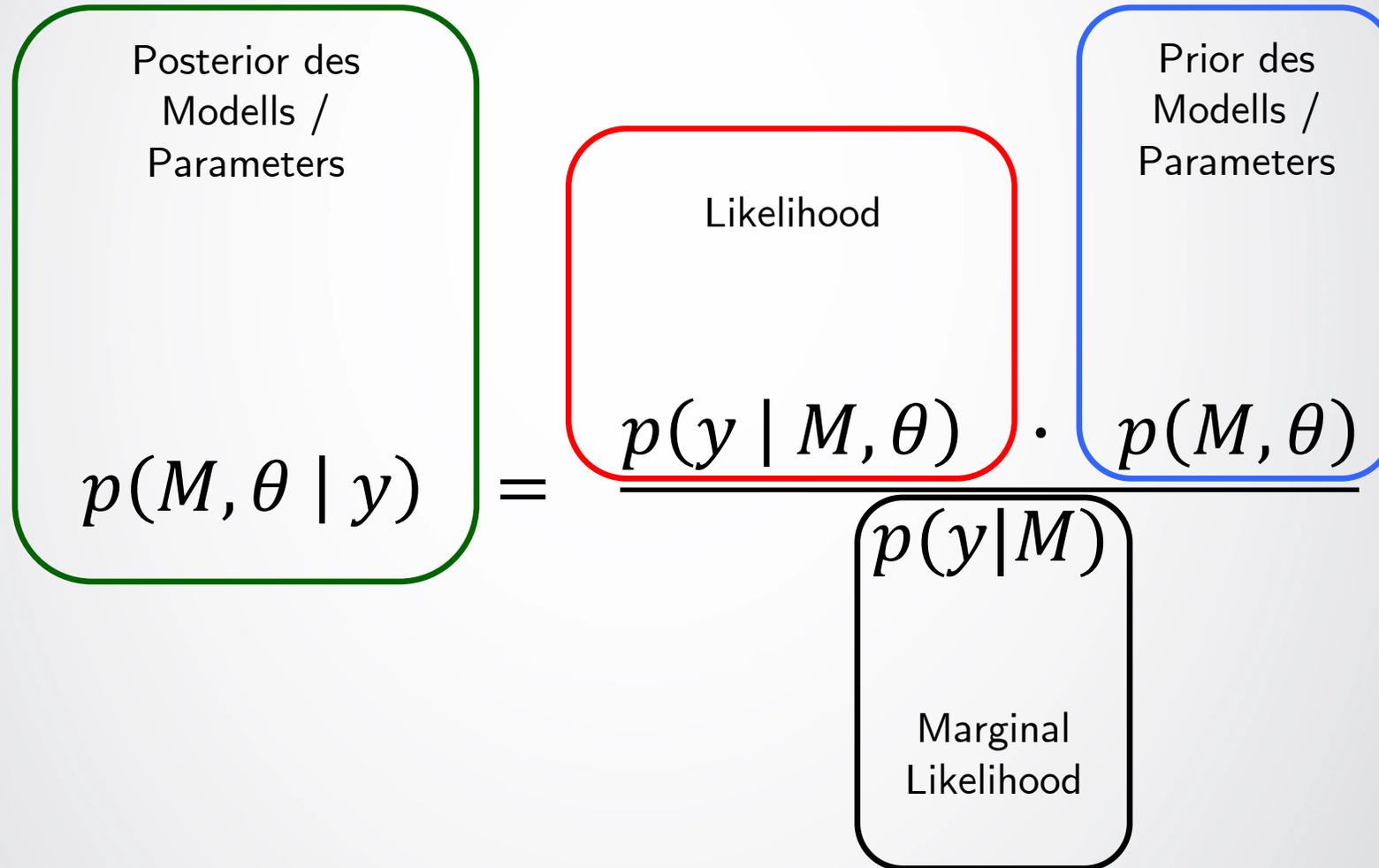
The diagram illustrates Bayes Theorem with three main components in rounded rectangular boxes:

- Green box (left):** Contains the text "Ws des Modells / Parameters gegeben die Daten" and the mathematical expression $p(M, \theta | y)$.
- Red box (middle):** Contains the text "Ws der Daten gegeben das Modell und Parameter" and the mathematical expression $p(y | M, \theta)$.
- Blue box (right):** Contains the text "Ws des Modells / Parameters" and the mathematical expression $p(M, \theta)$.

The green box is connected to the red and blue boxes by an equals sign (=). The red and blue boxes are connected by a multiplication dot (\cdot). A horizontal line is drawn under the product of the red and blue boxes, with a vertical line extending downwards to a black box containing the expression $p(y|M)$ and the text "Ws der Daten gegeben das Modell".

Grundlagen

Bayes Theorem



Inhalt

- Einführung
- Grundlagen
- **Parameter schätzen**
- Hypothesen testen
- Exkurs: JASP
- Modelle auswählen
- Zusammenfassung

Parameterschätzung

- Ziel ist es einen Parameter θ zu schätzen, z.B. einen Mittelwert, eine Varianz oder einen Regressionskoeffizienten
- Wie groß ist der Parameter in der Population?
- In Bayes schätzen wir deshalb die Posterior Verteilung des Parameters: $p(\theta|y) = p(\theta) \cdot p(y|\theta)$
- Mit der Posterior können wir berechnen:
 - Welcher Parameterwert der wahrscheinlichste ist
 - Wie groß die Wahrscheinlichkeit ist, dass der Parameter größer als ein bestimmter Wert ist
 - Was die Intervallgrenzen sind, die den Parameter mit 95% Wahrscheinlichkeit einschließen (Credible Interval)

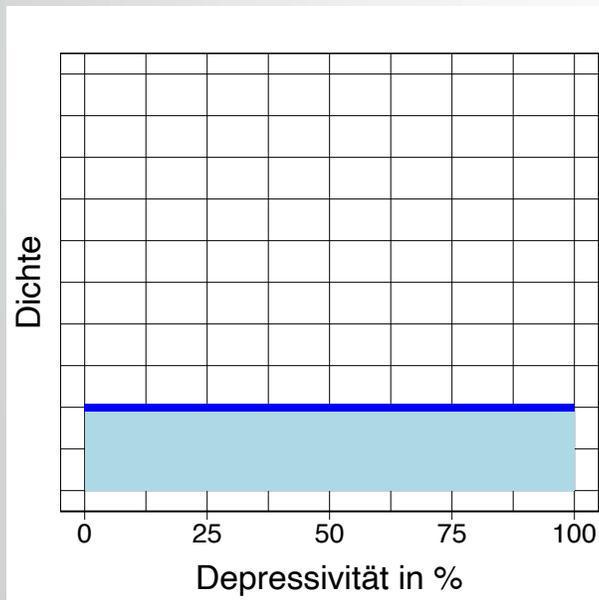
Beispiel: Depressivitätsrate

- Wie hoch ist der Anteil der Menschen mit Depression in der Gesamtbevölkerung?
- Ziehung einer Stichprobe, Diagnostik zu depressiver Symptomatik
- Was denken Sie: Wie hoch ist der Anteil und wie unsicher sind Sie sich bei Ihrer Schätzung? 5%? 10%? 15%? $SD=5\%$? $SD=20\%$?
- Nichts anderes ist eine Prior:
 - Wir haben Annahmen über die Depressivitätsrate bevor wir die Daten anschauen
 - Diese Annahmen formalisieren wir mit einer Wahrscheinlichkeitsverteilung
 - Wir brauchen diese Annahmen um die Posterior schätzen zu können

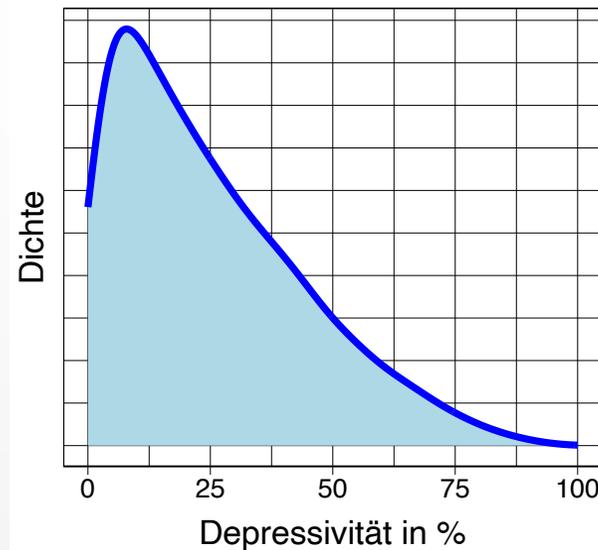
Depressivität – Prior Terminologie

Was sind die wahrscheinlichsten Werte für die Depressivitätsrate und welche Werte sind weniger wahrscheinlich?

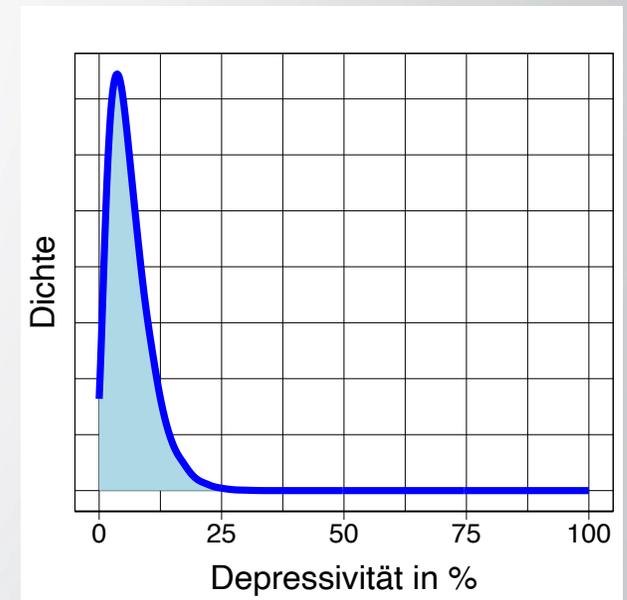
Uninformativ/diffus/flach/objektiv



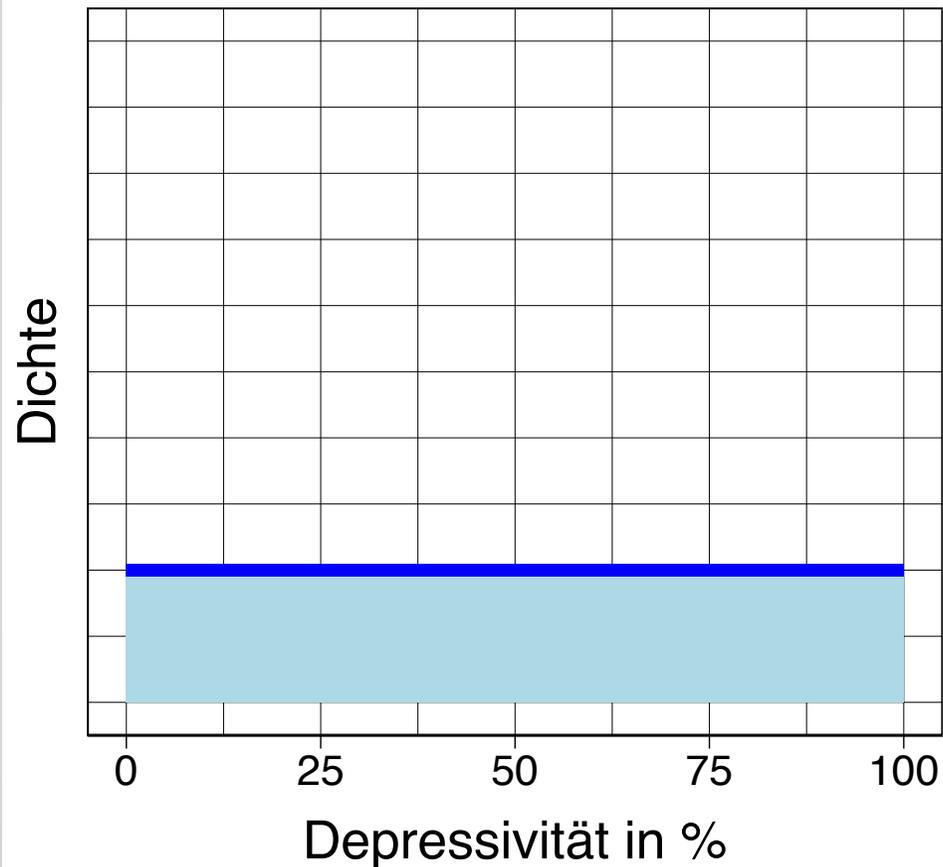
Relativ uninformativ / Schwach informativ



Informativ/Subjektiv

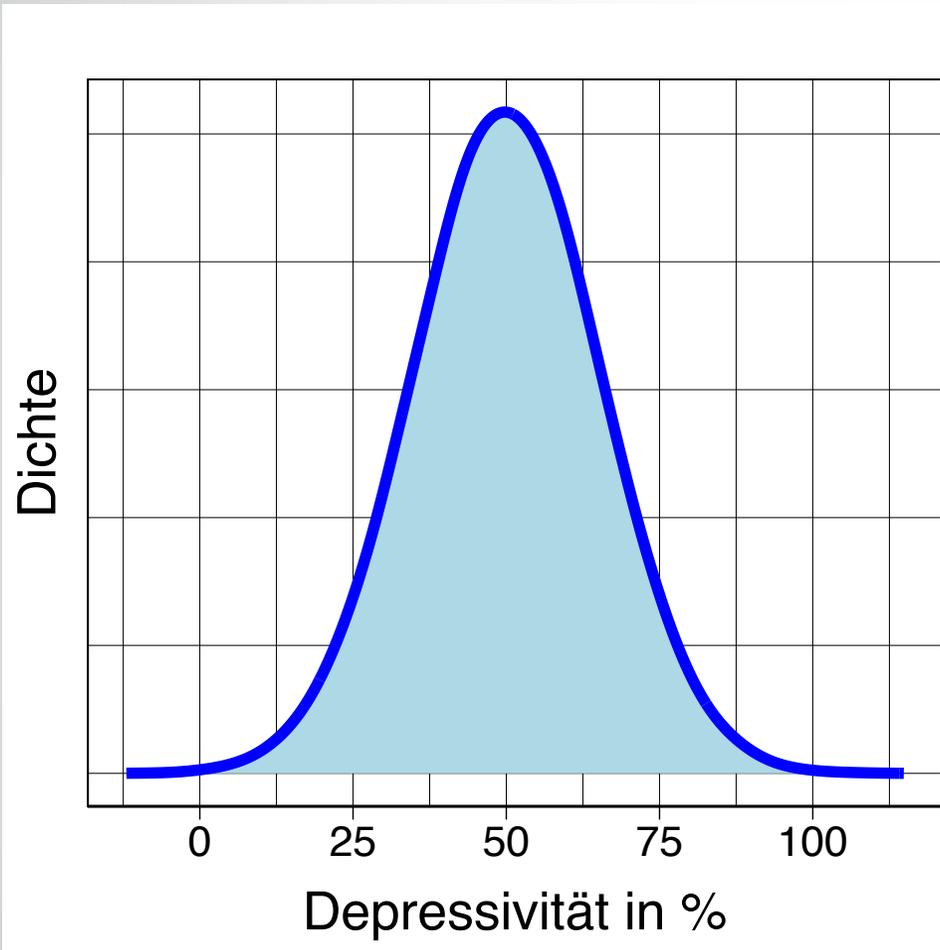


Depressivität – Uniforme Prior



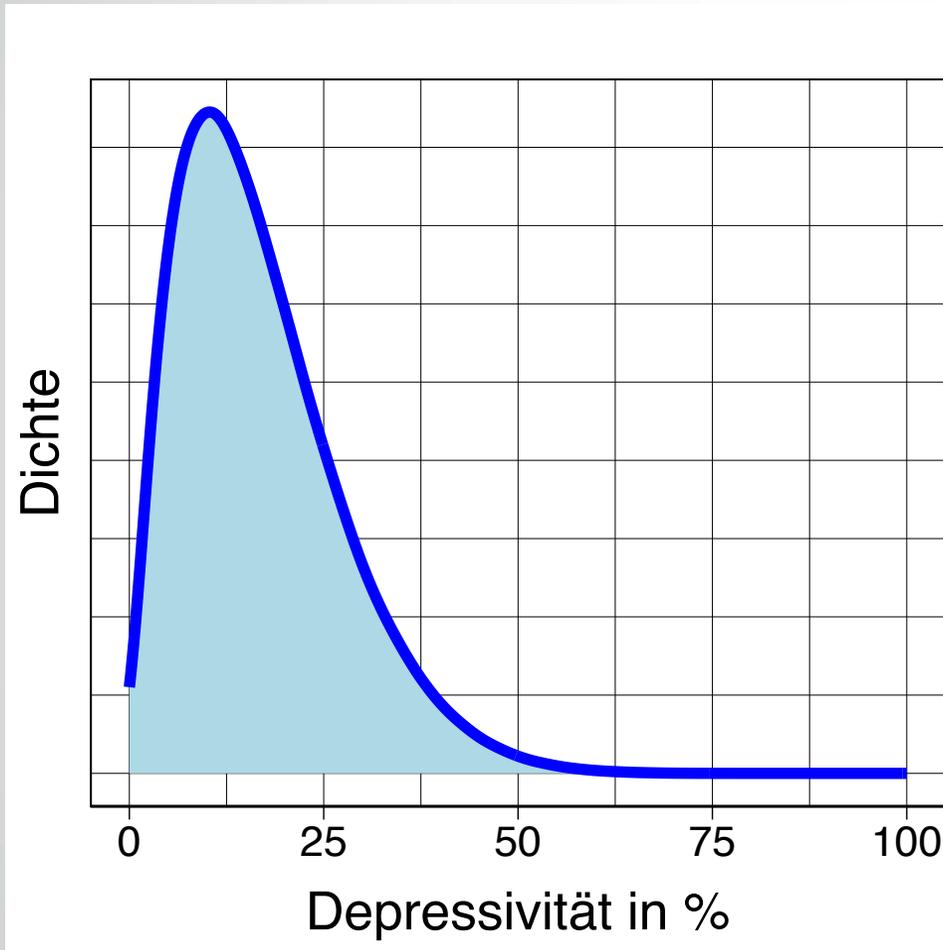
- Fläche unter der Kurve ist 1
- Uniforme Verteilung
- Die Werte für die Depressivität sind gleich wahrscheinlich a-priori

Depressivität – Normal Prior



- Normalverteilung: $N(50, 15)$
- Wahrscheinlichste Depressivitätsrate: 50%
- Extremere Werte sind weniger wahrscheinlich

Depressivität – Beta Prior

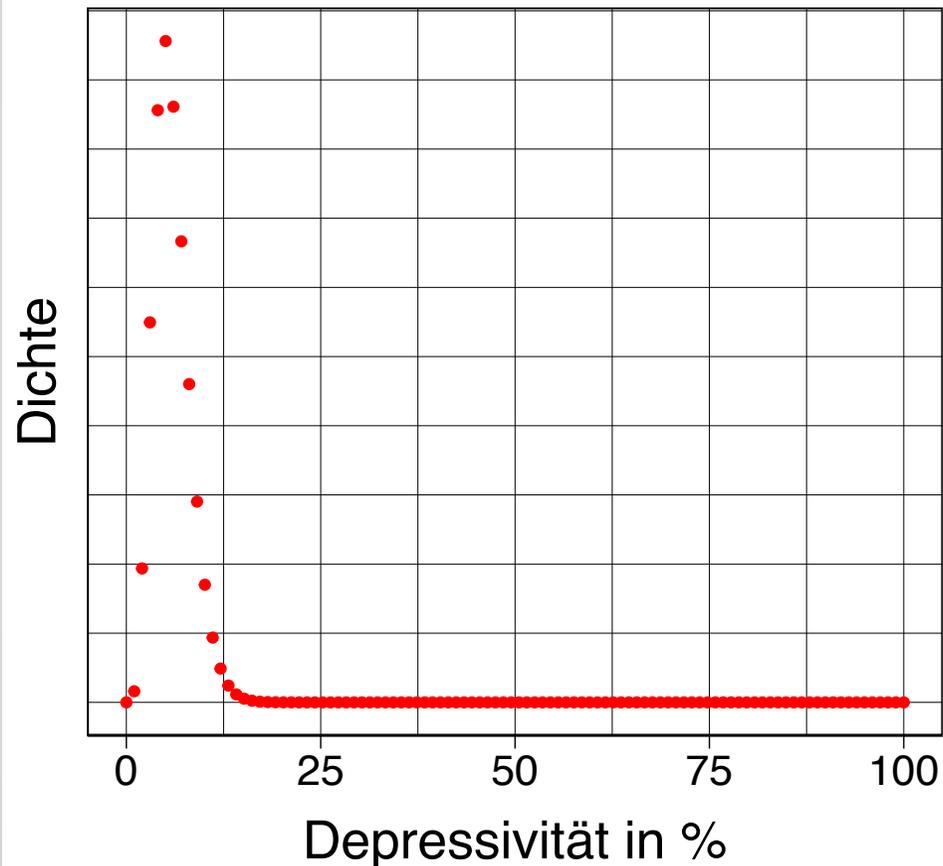


- Fläche unter der Kurve ist 1
- Beta-Verteilung für binomialverteilte Zufallsvariable
- $B(\alpha = 2, \beta = 10)$
- Wahrscheinlichster Wert für Depressivitätsrate a-priori ist $\sim 16\%$
- Wahrscheinlichkeit für Werte $> 30\%$ nimmt weiter ab
- Da die Beta-Verteilung unseren Annahmen über die Depressivitätsrate am ehesten entspricht, nehmen wir nun diese als Prior

Depressivität – Likelihood (1)

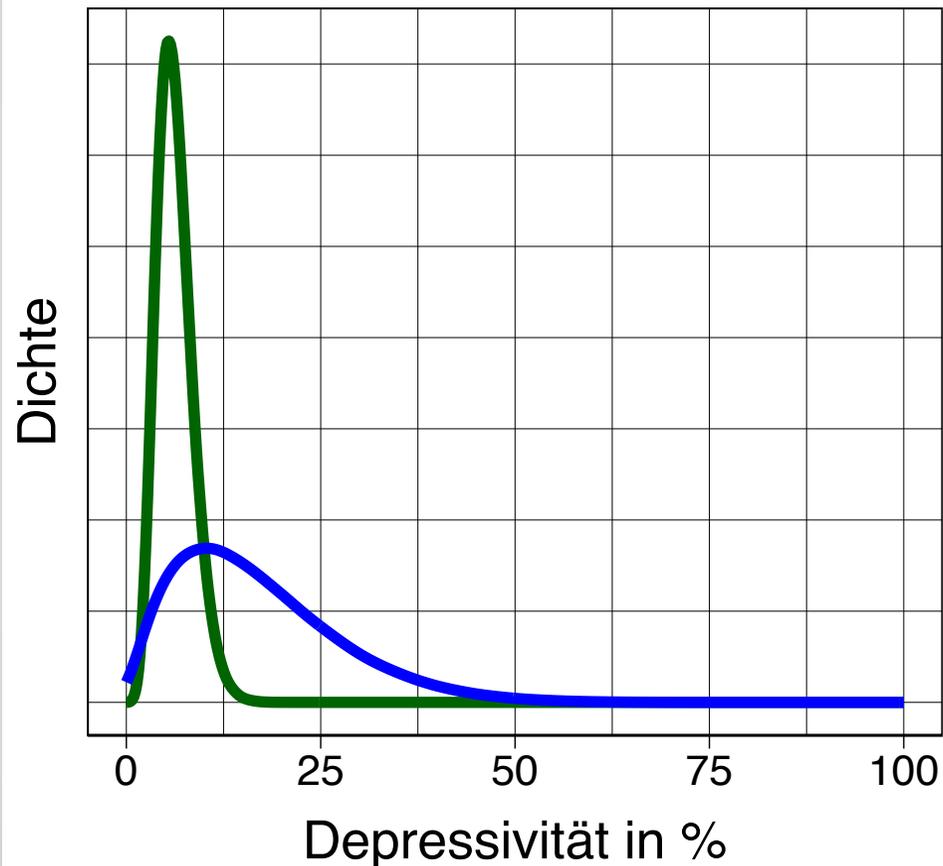
- Dann betrachten wir die Daten
- Die Prior wird mit den Daten „ge-updated“: $p(\theta) \cdot p(y|\theta)$
- Wie groß ist die Evidenz, die die Daten bringen?

Depressivität – Likelihood (2)



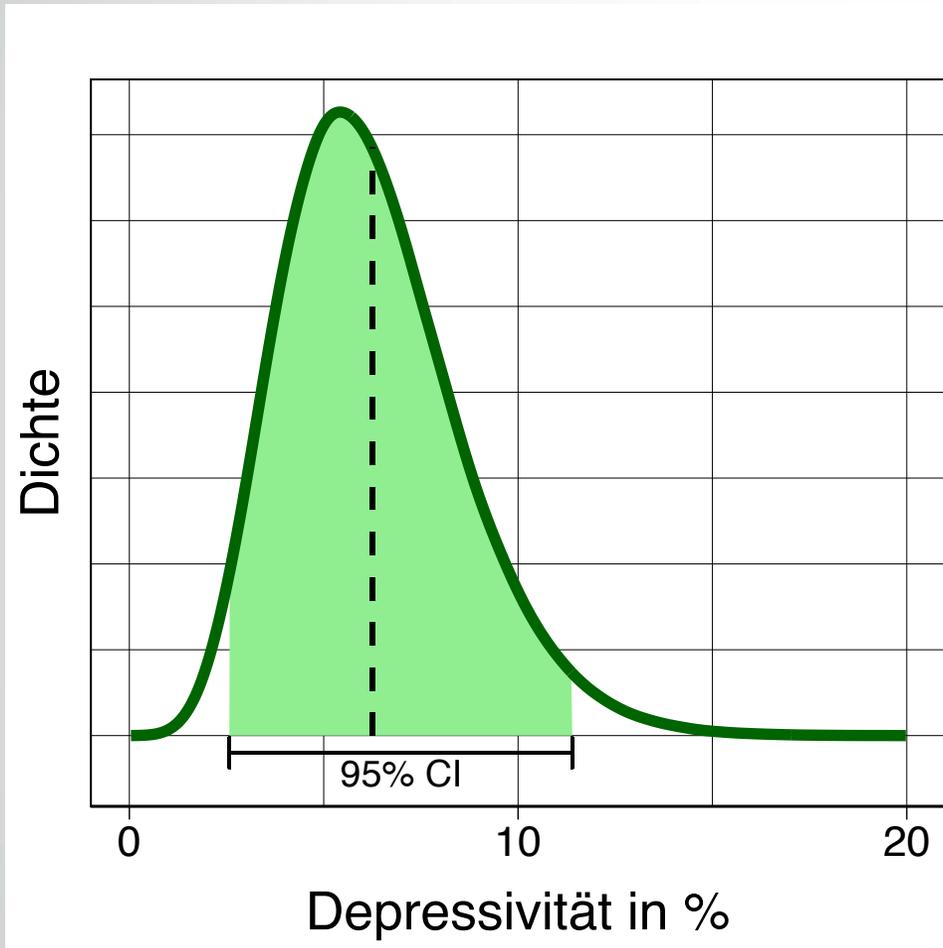
- Wir ziehen 100 Personen und beobachten 5 Depressive
- Wir können die Likelihood berechnen, da die „Ziehung“ von 100 Personen mit 5 „Erfolgen“ ein Binomialesperiment ist
- $N = 100, x = 5$
- $p(y|\theta) = \theta^5(1 - \theta)^{95}$

Depressivität – Posterior (1)



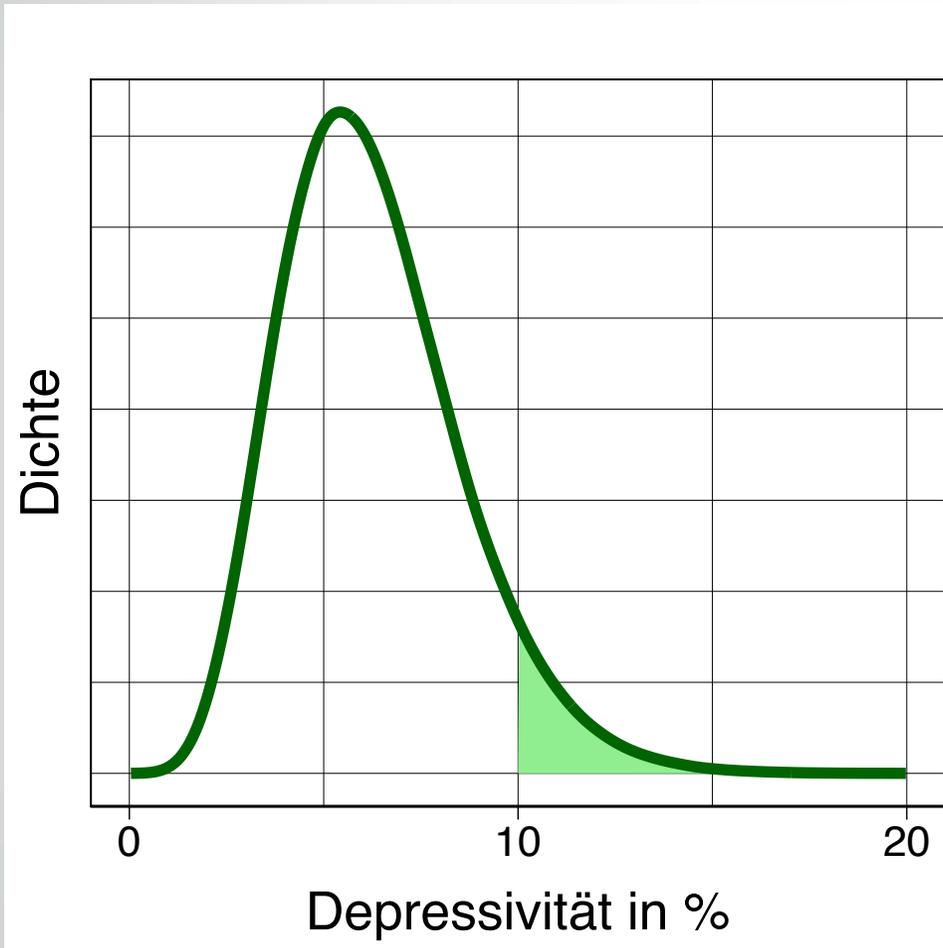
- In grün, die Posterior
- In blau, die Beta-Prior
- Bei binomialverteiletem Parameter und Wahl der Beta-Verteilung als Prior können wir die Posterior analytisch bestimmen (siehe Wikipedia)

Depressivität – Posterior (2)



- Das 95% Credible Interval enthält die wahre Depressivitätsrate mit 95% Wahrscheinlichkeit
- 95% CI [.026, .114]
- Der wahrscheinlichste Wert für die Depressivitätsrate (mean) liegt bei 6.25%

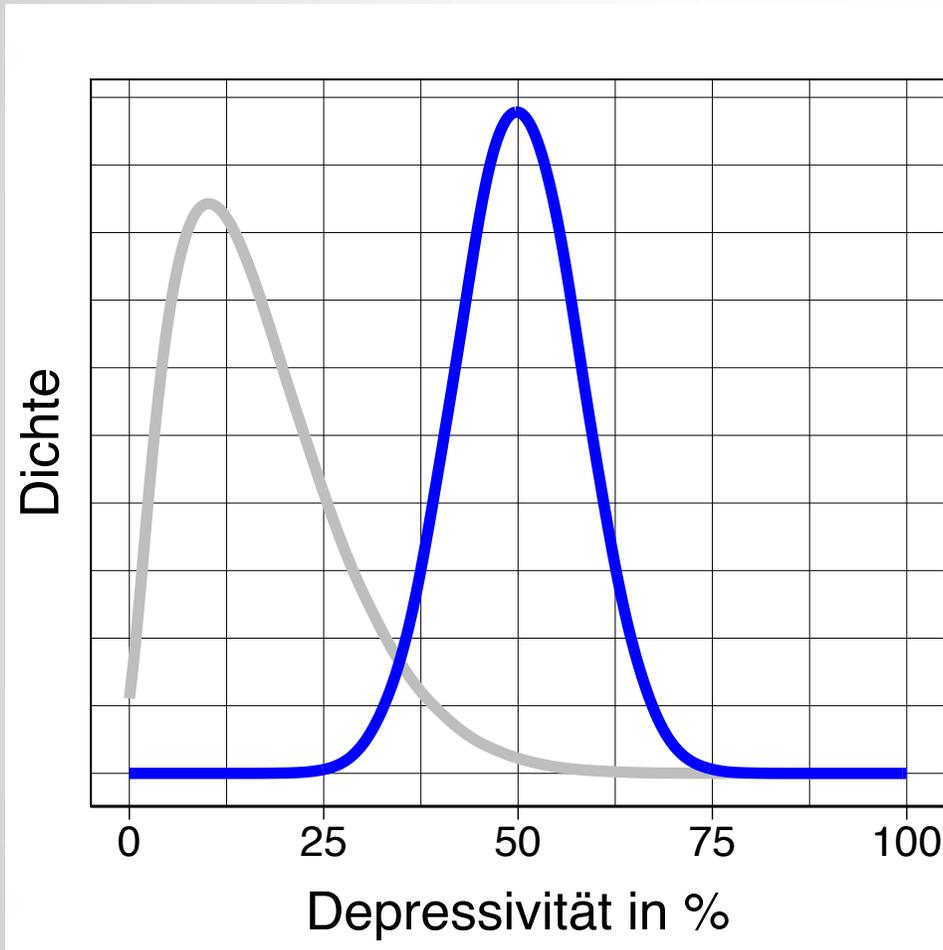
Depressivität – Posterior (3)



- Wie hoch ist die Wahrscheinlichkeit, dass die wahre Depressivitätsrate über, z.B., 10% liegt?

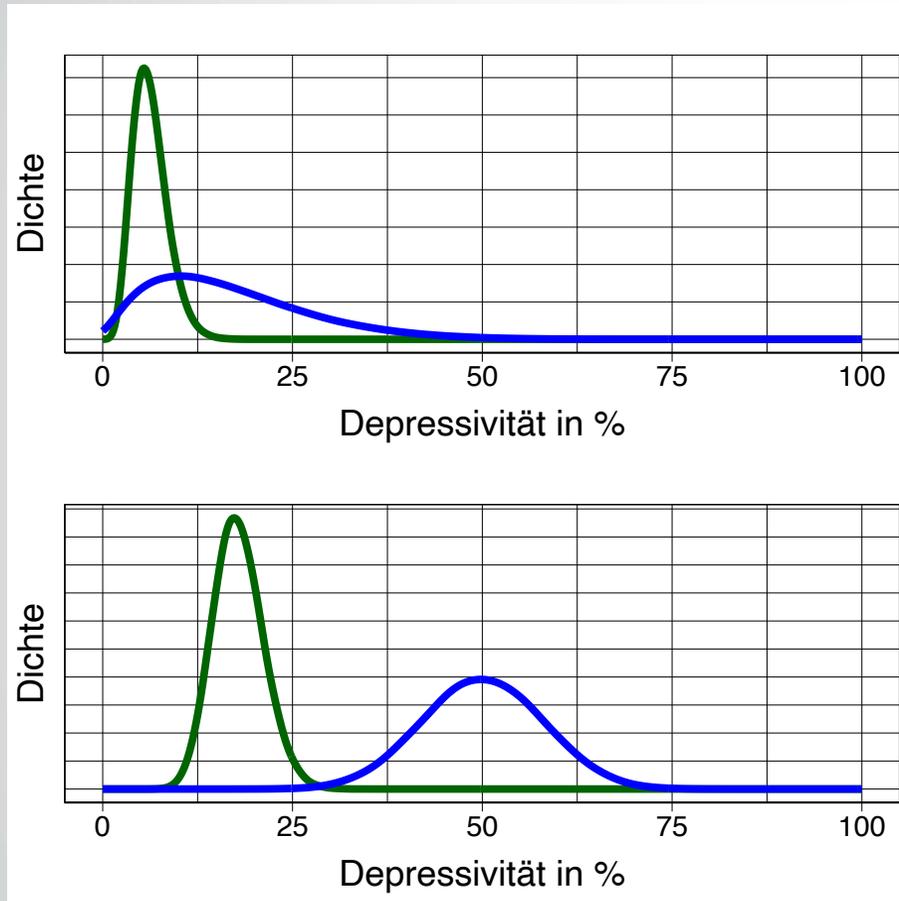
$$p(\theta > .10 | y) = 0.065$$

Depressivität – Einfluss der Prior (1)



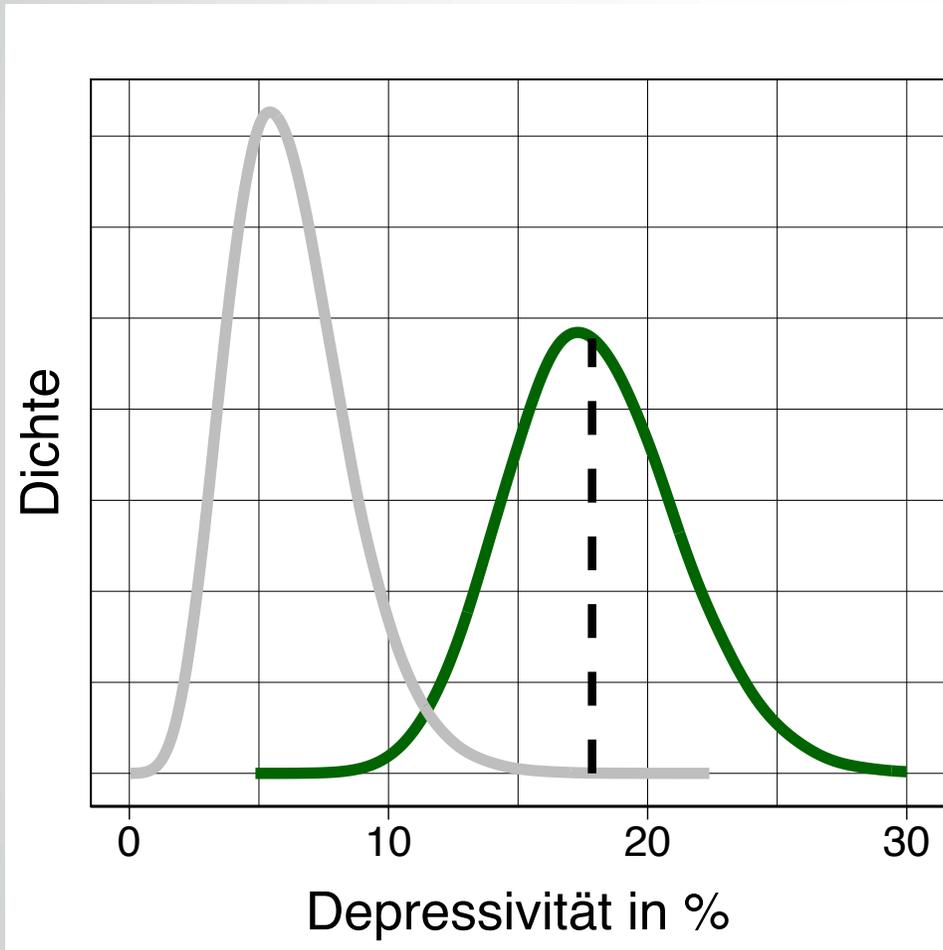
- Was passiert wenn wir eine andere Prior Verteilung benutzen?
- Wir nehmen nun zur Veranschaulichung die Normalverteilung: $N(50, 15)$
- In blau, die normalverteilte Prior
- In grau, die Beta-verteilte Prior von vorher

Depressivität – Einfluss der Prior (2)



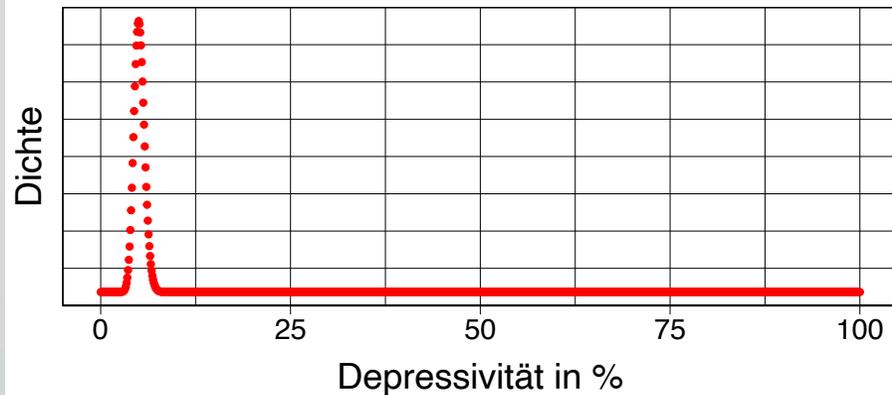
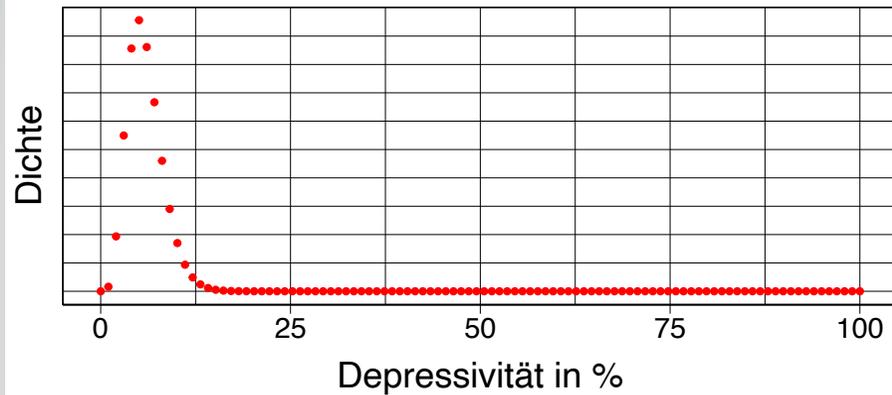
- Die Daten bleiben gleich, also auch die Likelihood: 5 Depressive in einer Stichprobe von 100
- Oben, die Verteilungen mit Beta-Prior
- Unten, die Verteilungen mit Normal-Prior
- In grün, die Posterior
- In blau, die Normal-Prior

Depressivität – Einfluss der Prior (3)



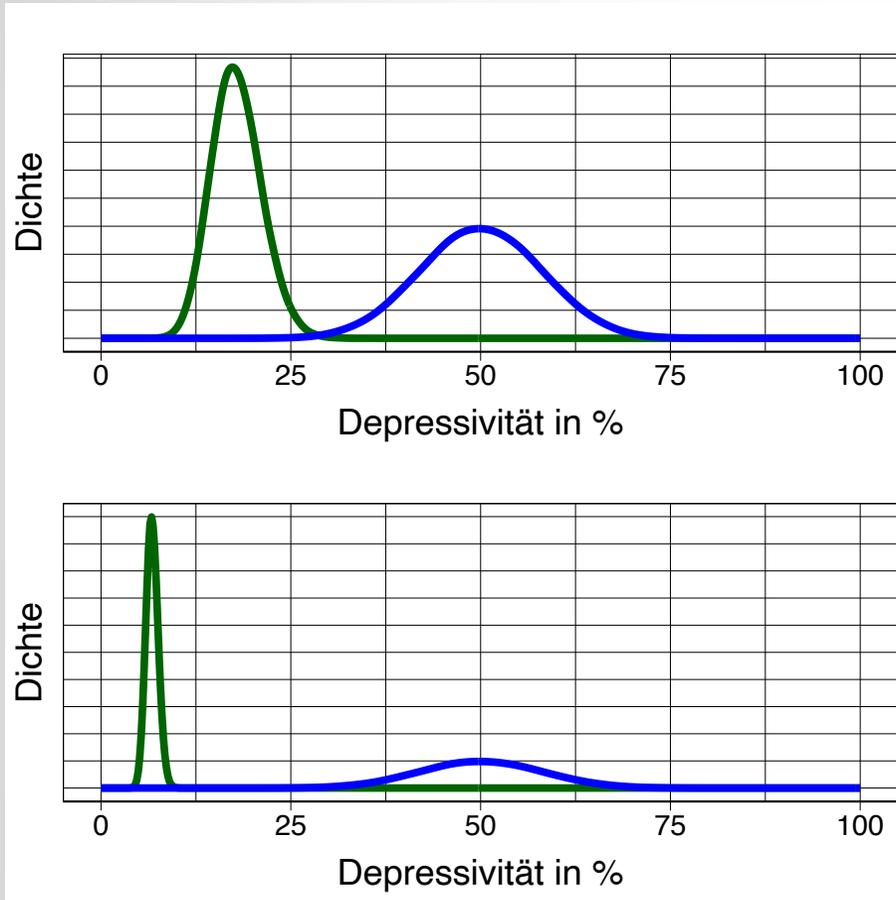
- In grau, die Posterior mit der vorherigen Beta-Prior
- In grün, die Posterior mit der jetzigen Normal-Prior
- Der Mittelwert der Posterior liegt nun bei 17.86% (vorher bei 6.25%)
- Die Prior hat Einfluss!

Depressivität – Einfluss der Daten (1)



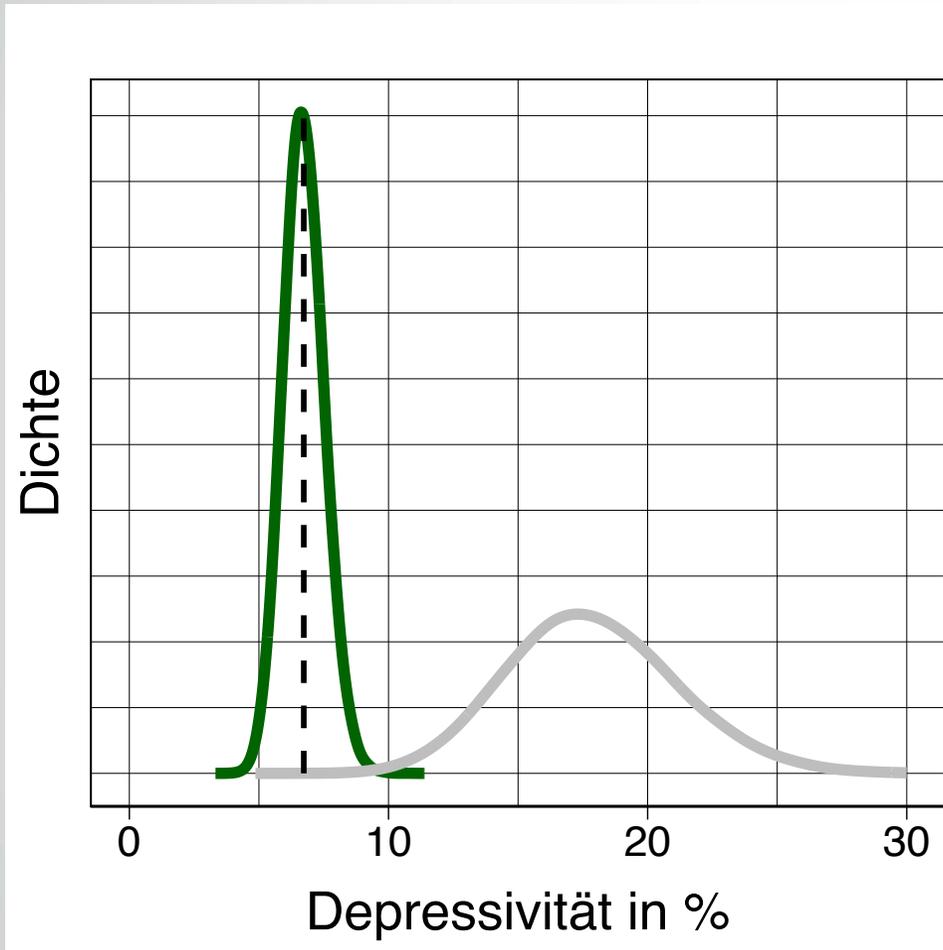
- Oben, Likelihood aus $N = 100, x = 5$
- Unten, Likelihood von Erhebung: $N = 1000, x = 50$
- Die Prior bleibt gleich
- Die Likelihood ist nun viel „informativer“

Depressivität – Einfluss der Daten (2)



- Oben, $N = 100, x = 5$
- Unten, $N = 1000, x = 50$
- In grün, die Posterior aus 1000 Beobachtungen ist nun viel spitzer als mit nur 100 Beobachtungen, und viel weiter von der Prior entfernt
- Viele Daten (Evidenz) verringern den Einfluss der Prior

Depressivität – Einfluss der Daten (3)



- In grün, die Posterior aus 1000 Beobachtungen
- In grau, die Posterior aus 100 Beobachtungen
- Die Unsicherheit über verschiedene Depressivitätsraten ist sehr viel kleiner geworden (engere und spitzere Verteilung)
- Die wahrscheinlichste Depressivitätsrate liegt nun bei 6.73%
- Würden wir die Beobachtungen auf $N = 10000$, $x = 500$ erhöhen wäre der Mittelwert bei 5.18%
- „Die Likelihood dominiert die Prior“

Prior

- Wie wir die Prior wählen, beeinflusst die Posterior und dadurch die Schätzung der Depressivität
- Standard- (Default) Priors gibt es für die meisten Parameter
- Uninformative Priors sind ein guter Startpunkt, jedoch sollte keine Analyse mit nur einer einzigen Prior gerechnet werden
- Liegen aber genug Informationen über den Parameter a-priori vor, so ist eine uninformative Prior eine schlechte Wahl

Schätzung

- Im Depressivitätsbeispiel konnten wir die Posterior analytisch bestimmen
- Dieses Privileg haben wir in vielen Situationen nicht
- Statistiker haben deswegen Methoden entwickelt, die uns erlauben die Posterior durch sukzessives Sampling zu approximieren
- Dies nennt sich Markov-Chain Monte-Carlo Sampling (MCMC)
- MCMC ist in R-Paketen und in JASP implementiert
- Als anwendungsorientierte Wissenschaftler können Sie MCMC benutzen, ohne es zu verstehen

Parameterschätzung - Recap

- Die Prior drückt unsere Ansicht über den Parameter vor der Datenerhebung aus
- Wir erhalten die Posterior indem wir die Prior mit den Daten (Likelihood) updaten
- Wir schätzen einen Parameter auf Basis der Posterior Verteilung:
 - Wahrscheinlichster Wert (Mittelwert, Median)
 - Credible Interval
 - Andere Inferenzen (größer als Cutoff)
- Die Posterior enthält alle Information über verschiedene Parameterwerte nachdem wir die Daten betrachtet haben
- Viele Daten verringern den Einfluss der Prior

Inhalt

- Einführung
- Grundlagen
- Parameter schätzen
- **Hypothesen testen**
- Exkurs: JASP
- Modelle auswählen
- Zusammenfassung

Hypothesen Testen

- Hypothesen testen ist nichts anderes als Modelle vergleichen:
 - Modell 1: Der Anteil der Depressiven liegt bei 5%
 - Modell 2: Der Anteil der Depressiven ist verschieden von 5%
- Für ein einziges Modell entspricht Bayes Theorem:

$$p(M | y) = \frac{p(M) \cdot p(y|M)}{p(y)}$$

Hypothesen Testen

- Wir betrachten jeweils die Wahrscheinlichkeiten eines Modells relativ zum anderen
- Nach Bayes ergibt sich:

$$\frac{p(M_1 | y)}{p(M_2 | y)} = \frac{p(y|M_1)}{p(y|M_2)} \cdot \frac{p(M_1)}{p(M_2)}$$

Posterior model odds = Bayes Factor · Prior model odds

Hypothesen Testen

$$\frac{p(M_1)}{p(M_2)}$$

- Prior Odds: Verhältnis der prior Wahrscheinlichkeit von Modell 1 und Modell 2
- Wie viel wahrscheinlicher ist Modell 1 als Modell 2 bevor wir die Daten betrachten?

$$\frac{p(M_1 | y)}{p(M_2 | y)}$$

- Posterior Odds: Verhältnis der posterior Wahrscheinlichkeit von Modell 1 und Modell 2
- Wie viel wahrscheinlicher ist Modell 1 als Modell 2 nachdem wir die Daten betrachtet haben?

$$\frac{p(y|M_1)}{p(y|M_2)}$$

- Bayes Factor: Verhältnis der Marginal Likelihoods von Modell 1 und Modell 2
- Wie viel wahrscheinlicher sind die Daten unter Modell 1 als unter Modell 2?

Bayes Factor

- ist für Bayesians, was für Frequentists der p-Wert ist
- drückt aus, wie wahrscheinlich die Daten unter Modell 1 im Vergleich zu Modell 2 sind
- ist der Grad relativer Evidenz in den Daten für Modell 1 gegenüber Modell 2
- ist ein Maß für die relative prädiktive Performanz der Modelle: Wie viel wahrscheinlicher ist es, dass Modell 1 die Daten produziert hat, als Modell 2?
- bevorzugt weniger komplexe Modelle im Vergleich zu komplexeren Modellen (Sparsamkeit)
- hängt (anders als bei der reinen Parameterschätzung) auch bei einer großen Stichprobe von den verwendeten Priors ab
- wird oft als BF_{10} angegeben (Hypothese 1 im Vergleich zu Hypothese 0)

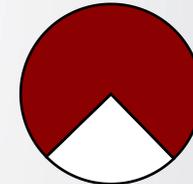
Bayes Factor - Interpretation

- $BF_{10} > 1$: Evidenz für H1
- $BF_{10} = 1$: keine Evidenz
- $BF_{10} < 1$: Evidenz für H0

- $BF_{10} = 3$: Es ist 3-mal wahrscheinlicher, dass H1 die Daten generiert hat, als H0

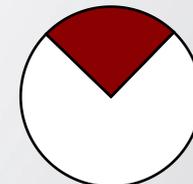
- $BF_{10} = 1/3$: Es ist 3-mal wahrscheinlicher, dass H0 die Daten generiert hat als H1

$p(y|H1)$



$p(y|H0)$

$p(y|H1)$



$p(y|H0)$

Depressivität – Hypothesen Testen

- Wir beobachten 5 Depressive in einer Stichprobe von 100
- H_0 : Anteil der Depressiven liegt bei 5%
- H_1 : Anteil der Depressiven liegt nicht bei 5%

- Klassisch:
 - Angenommen die H_0 gilt, dann ist die Wahrscheinlichkeit unter 100 Personen 5 Depressive (oder extremere Werte) zu beobachten nicht signifikant (sehr nah an 1)
 - Aber: Wie wahrscheinlich ist es, dass die H_1 die Daten produziert hat?

Depressivität – Bayes Factor

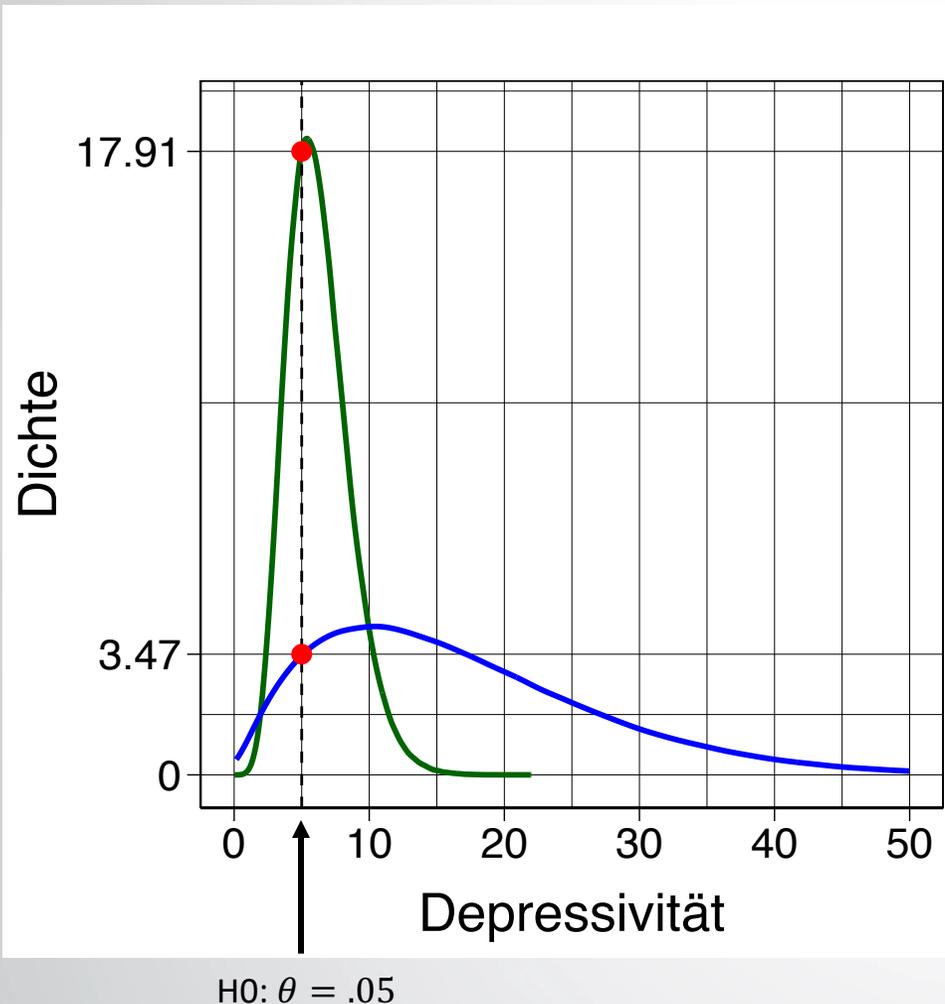
- Wie erhalten wir den Bayes Factor für die Depressivitätsrate?
- Für einfache genestete Hypothesen (wie im Beispiel) kann der BF mit einer Methode bestimmt werden, die die Verteilungen der Prior und Posterior benutzt

Depressivität – Bestimmung Bayes Factor

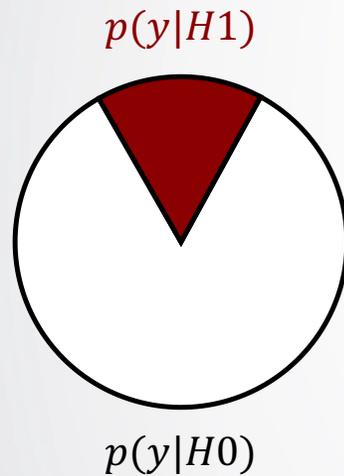
- Bestimmung des Bayes Factor mithilfe der Savage-Dickey Density Ratio:
- Hierbei wird die Höhe der Dichte von Prior und Posterior für die Depressivitätsrate von H1 (Kurven links) an der Stelle der H0 ($\theta = .05$) verglichen

$$BF_{10} = \frac{3.47}{17.91} = 0.19$$

$$BF_{01} = \frac{1}{BF_{10}} = 5.17$$



Depressivität – Bayes Factor



- $BF_{01} = 5.17$
- Die Daten sind etwa 5-mal wahrscheinlicher unter der $H0$ als unter der $H1$
- $BF_{10} = 0.19$
- Die Daten sind 0.19-mal wahrscheinlicher unter $H1$ als unter $H0$

Bayes Factor - Schätzung

- Wie auch schon für die Posterior, ist es oft nicht möglich die Marginal Likelihoods, die man für den Bayes Factor braucht, analytisch zu bestimmen
- Savage-Dickey Density Methode funktioniert nur bei einfachen Fragestellungen
- Sampling Methoden (MCMC) und Approximationen helfen (z.B. Laplace Approximation)
- Es ist Aufgabe von Statistikern Ihnen diese Methoden zugänglich zu machen

Hypothesen Testen - Recap

- Man benutzt den Bayes Factor um Hypothesen zu testen
- Der Bayes Factor vergleicht, wie wahrscheinlich es ist, dass verschiedene Hypothesen die Daten produziert haben
- $BF_{10} = 3$ bedeutet, dass die Daten unter der H_1 3-mal wahrscheinlicher sind als unter der H_0
- Der Bayes Factor kann sowohl die Evidenz für die Null- als auch die Alternativhypothese quantifizieren
- Die Prior ist wichtig
- Im Zweifelsfall bevorzugt der Bayes Factor die sparsamere Hypothese

Inhalt

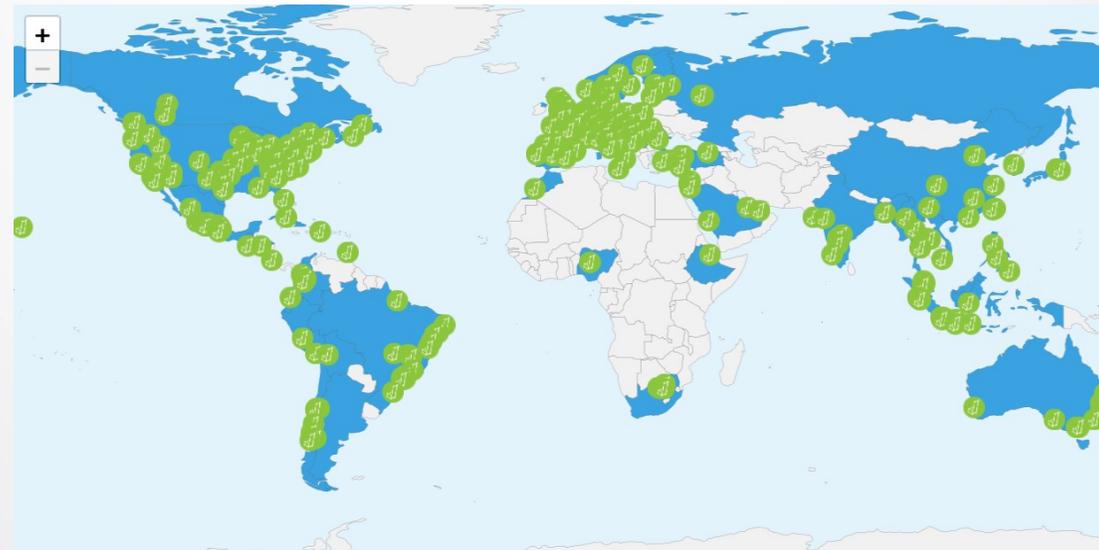
- Einführung
- Grundlagen
- Parameter schätzen
- Hypothesen testen
- **Exkurs: JASP**
- Modelle auswählen
- Zusammenfassung

Software: JASP



A Fresh Way to Do Statistics

- Gratis und Open-Source
- Wird an der Universität Amsterdam von Wissenschaftlern für Wissenschaftler entwickelt
- Basiert auf R-Paketen



JASP

Software: JASP



A Fresh Way to Do Statistics

GUI Features	JASP 0.18		SPSS 29	SPSS 29
	Classic	Bayesian	Classic	Bayesian
Analysis				
(repeated) (M)AN(C)OVA and non-parametrics	✓	(✓)	✓	(✓)
Audit - Bayesian Tools for the Auditing of Organisations	✓	✓	modeler	X
Bain - Bayesian informative hypotheses evaluation		✓		X
BSTS - Bayesian structural time series		✓		X
Circular / Directional Statistics - analysis of directions, often angles	✓	X	X	X
Cochrane Meta-Analyses	✓	✓	X	X
Descriptives	✓		✓	
Distributions	✓	✓	X	X
Equivalence T-Tests (TOST): Independent, Paired, One-Sample	✓	✓	X	X
Factor Analysis (PCA, EFA, CFA)	✓		✓ / AMOS	
Frequencies (Binomial, Multinomial, Contingency, Chi ² , log-linear regression)	✓	✓	✓	(✓)
JAGS (Bayesian black-box Markov chain Monte Carlo (MCMC) sampler)		✓		(AMOS)
Learn Bayes		✓		X
Machine Learning (incl Cluster & Discriminant Analyses)	✓	X	✓	X
(Cochrane) Meta-Analysis (PET-PEESE, WAAP-WLS for publication bias correction)	✓	✓	✓	X
(Generalized or Linear) Mixed Models	✓	✓	✓	X
Network	✓	✓	✓	X
Sample Size Planning & Power Analysis	(✓)	X	(✓)	X
Prophet - Time Series Forecasting	X	✓	✓	X
Quality Control	✓	X	(✓)	X
Regression / Correlation (Pearson, Spearman, Kendall etc, (log)linear, multinomial, ordinal, firth logistic, Analysis of Residuals etc.	✓	(✓)	✓	(✓)
Reliability	✓	(✓)	✓	X
Structural Equation Modeling (SEM) (module based on R Lavaan) incl. Mediation, Partial Least Squares, Latent Growth & MIMIC	✓		AMOS	
Summary Statistics		✓		X
T-Tests: Independent, Paired, One-Sample (incl. z, Welch, non-parametrics)	✓	✓	✓	(✓)
Visual Modeling: Automated Plotting, (Non-)Linear, Mixed, Generalized Linear	✓	X	✓	X
Acceptance Sampling	✓		X	

Sources <https://jasp-stats.org/features/> and official IBM SPSS documentation

Empty field suggests, that this may not be possible

This table is made and maintained by Thomas Langkamp.

Depressivitätsbeispiel in JASP

- Wir können das Beispiel mit dem Depressivitätsrate sehr einfach in JASP reproduzieren
- Für gewöhnlich würden wir einen Datensatz in JASP laden, der unsere Beobachtungen enthält: Für das Depressivitätsbeispiel wäre das ein Datensatz mit 95 Nullen und 5 Einsen ($N = 100, x = 5$)
- Für einfache Daten gibt es in JASP das Modul "Summary Statistics" das auch ohne Daten funktioniert
- Wir wählen „Bayesian Binomial Test“ und müssen lediglich die Zahl der Erfolge ($x = 5$) und Misserfolge ($N - x = 95$) kennen

Depressivitätsbeispiel in JASP



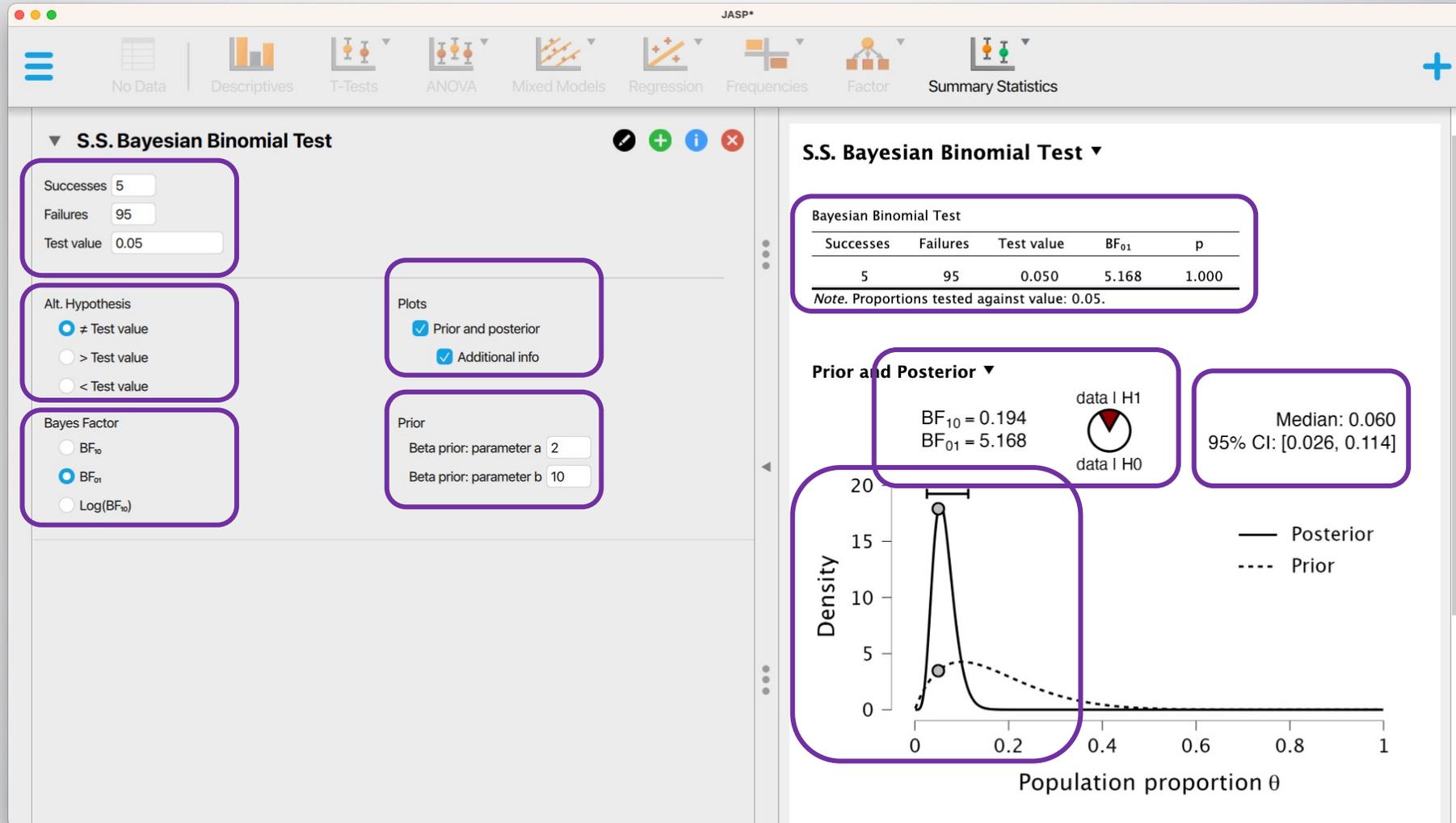
The screenshot shows the JASP software interface. At the top, there is a menu bar with icons for 'New Data', 'Descriptives', 'T-Tests', 'ANOVA', 'Mixed Models', 'Regression', 'Frequencies', and 'Factor'. A purple circle highlights a '+' icon in the top right corner of the menu bar. Below the menu bar, the main area has a blue header with the JASP logo and the text 'Welcome to JASP' and 'A Fresh Way to Do Statistics: Free, Friendly, and Flexible'. The version 'JASP 0.18 (Intel)' is displayed in the top right. Below the header, there is a list of features:

- **Free:** JASP is an open-source project with structural support from the University of Amsterdam.
- **Friendly:** JASP has an intuitive interface that was designed with the user in mind.
- **Flexible:** JASP offers standard analysis procedures in both their classical and Bayesian manifestations.

So open a data file and take JASP for a spin!

Please keep in mind that this is a preview release and a number of features are still missing.
If JASP doesn't do all you want today, then check back tomorrow: JASP is being developed at break-neck speed!

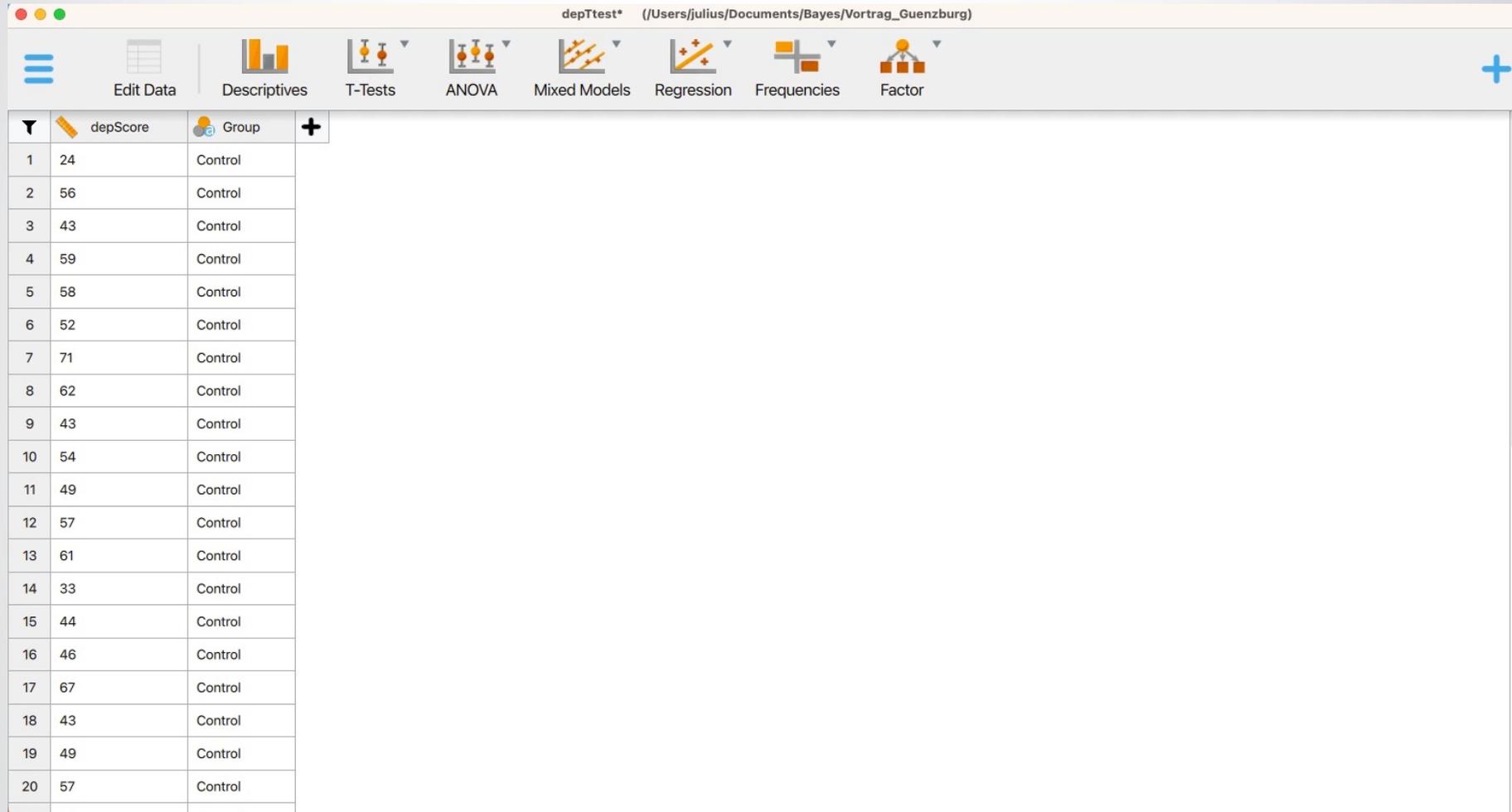
Depressivitätsbeispiel in JASP



T-Test in JASP

- Angenommen wir haben zwei Gruppen von Patienten ($N = 44$)
- Die eine Hälfte erhält ein Treatment (23), die andere ist Kontrolle (21)
- Wir erheben anschließend einen Depressivitäts-Score
- Der Mittelwert in der Treatmentgruppe ist ca. 42, und in der Kontrollgruppe ca. 51
- Ist der Mittelwertsunterschied von Bedeutung?
- H_0 : Es gibt keinen Unterschied zwischen der Treatment- und der Kontrollgruppe, $\delta = 0$
- H_1 : Es gibt einen zwischen den Gruppen, $\delta \neq 0$

T-Test in JASP – Bayesian



The screenshot shows the JASP software interface with a dataset loaded. The window title is "depTtest*" and the file path is "/Users/julius/Documents/Bayes/Vortrag_Guenzburg". The top toolbar includes icons for Edit Data, Descriptives, T-Tests, ANOVA, Mixed Models, Regression, Frequencies, and Factor. The main area displays a table with two columns: "depScore" and "Group". The "depScore" column contains numerical values, and the "Group" column contains the value "Control" for all 20 rows.

	depScore	Group
1	24	Control
2	56	Control
3	43	Control
4	59	Control
5	58	Control
6	52	Control
7	71	Control
8	62	Control
9	43	Control
10	54	Control
11	49	Control
12	57	Control
13	61	Control
14	33	Control
15	44	Control
16	46	Control
17	67	Control
18	43	Control
19	49	Control
20	57	Control

T-Test in JASP – Bayesian

The screenshot displays the JASP interface for a Bayesian Independent Samples T-Test. The left panel shows the configuration options, and the right panel shows the results and inferential plots.

Bayesian Independent Samples T-Test Configuration:

- Dependent Variables: depScore
- Grouping Variable: Group
- Alternative Hypothesis: Group 1 ≠ Group 2
- Bayes Factor: BF₁₀
- Tests: Student
- Additional Statistics: Descriptives
- Repeatability: Set seed: 1
- Plots: Prior and posterior, Additional info, Credible interval: 95.0 %
- Descriptives: Descriptives, Credible interval: 95.0 %
- Bar plots: Bar plots, Credible interval: 95.0 %, Standard error, Fix horizontal axis to 0
- Raincloud plots: Raincloud plots, Horizontal display
- Missing Values: Exclude cases per dependent variable, Exclude cases listwise

Results:

Bayesian Independent Samples T-Test

	BF ₁₀	error %
depScore	2.217	0.008

Inferential Plots:

depScore

Prior and Posterior

BF₁₀ = 2.217
BF₀₁ = 0.451

data | H1
data | H0

Median: 0.568
95% CrI: [0.006, 1.173]

Density

Effect size δ

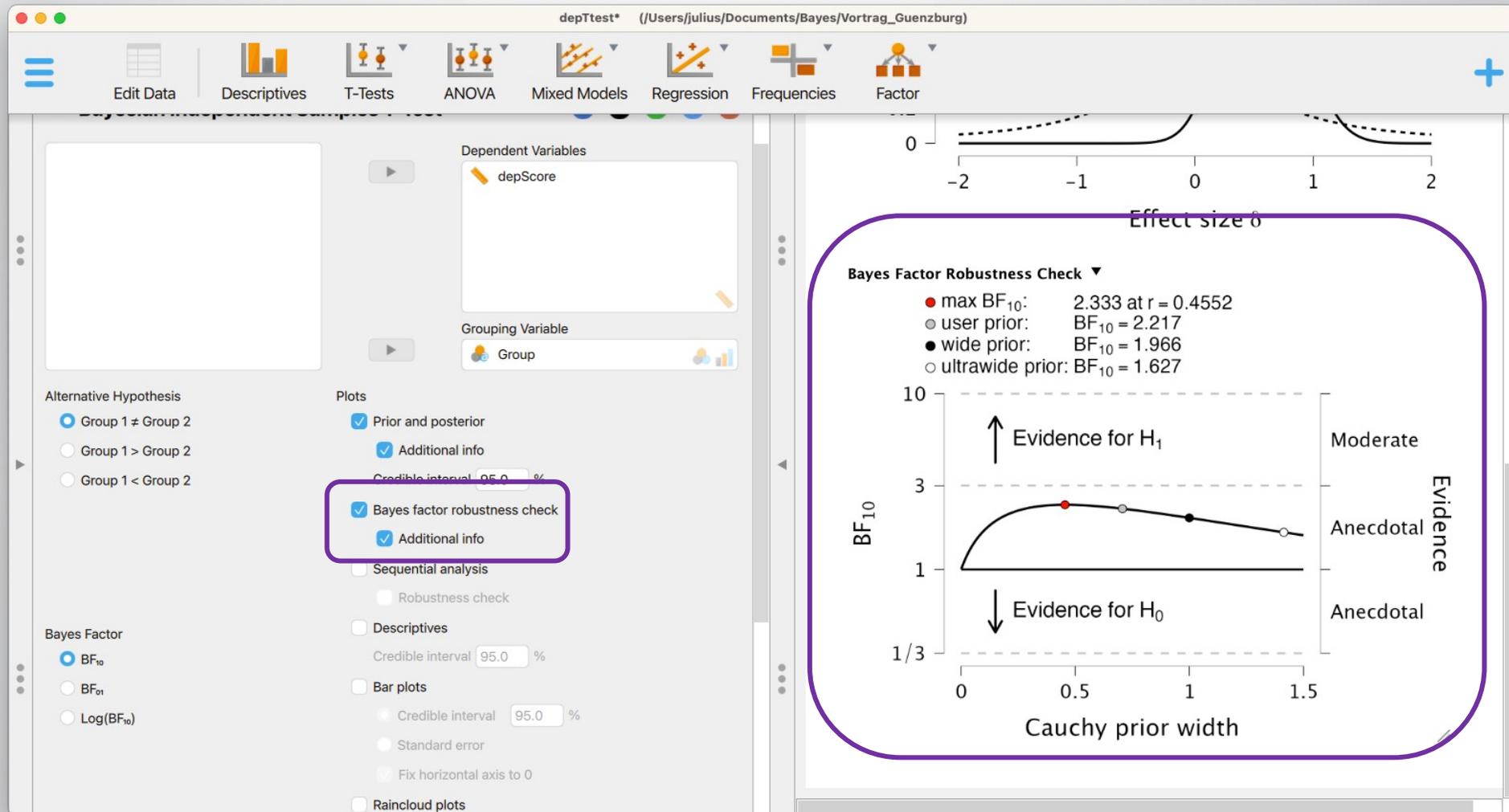
— Posterior
- - - Prior

Descriptives:

Raincloud Plots

depScore

T-Test in JASP – Bayesian



T-Test in JASP

- Einen $BF_{10} = 2.217$ würde man als anekdotische Evidenz für H_1 bezeichnen (siehe bspw. Klassifikation auf Wikipedia)
- Die Wahrscheinlichkeit, dass die H_1 die Daten produziert hat ist 2.2 mal größer als, dass die H_0 die Daten produziert hat
- Die Robustheitsanalyse ergibt, dass eine Veränderung der Form der Prior wenig an der Schlussfolgerung ändert: Die Evidenz für H_1 ist anekdotisch

Inhalt

- Einführung
- Grundlagen
- Parameter schätzen
- Hypothesen testen
- Exkurs: JASP
- Modelle auswählen
- Zusammenfassung

Modelle Auswählen – Single Model Fit

- Bayes ermöglicht post-hoc zu überprüfen ob ein Modell zu den Daten passt:
 - Man zieht Werte aus der Posterior und schaut ob diese den Daten ähnlich sind
 - Posterior predictive checking (PPC)
- Beispiel: Wir hatten zuvor die Posterior der Depressivitätsrate bestimmt:
 - Ziehen wir nun wiederholt Werte aus der Posterior-Verteilung, können wir diese numerisch und mit graphischen Methoden mit den eigentlichen Daten vergleichen
 - Je eher die gezogenen Werte die Daten reproduzieren, desto passender das Modell

Modelle Auswählen und Schätzen

- In vielen Szenarien gibt es mehr als ein mögliches Modell und der Vergleich von „nur“ zwei Modellen ist unzureichend, z.B. Regression oder Strukturgleichungsmodellierung
- Es gibt oft viele Kandidatenmodelle
- Klassisch:
 - Durch viele Vergleiche erhöht sich die Wahrscheinlichkeit, eine Nullhypothese fälschlicherweise abzulehnen (Alphafehler-Kumulierung)
 - Wir wählen ein Modell aus, und schätzen die Parameter (z.B. Regressionskoeffizienten) für dieses
- Bayes:
 - Wie gut passen verschiedene Modelle zu den Daten relativ zu den möglichen Modellen?
 - Wie wahrscheinlich ist es, dass ein Parameter im „wahren“ Modell vorliegt?
 - Wie können wir die Unsicherheit bei der Modellauswahl in die Parameterschätzung übernehmen, aka, einen Parameter nicht nur auf Basis eines potenziell ungeeigneten Modells schätzen?

Beispiel - Regression

- World Happiness Data:
 - $N = 155$ Länder
 - Happiness vorhergesagt durch:
 - GDP per capita, Family, Life Expectancy, Freedom, Generosity, Government Corruption
 - Wir erhalten $2^6 = 64$ mögliche Regressionsmodelle (ohne Interaktionen)
- Wie vergleichen wir die Kandidatenmodelle?
- Wählen wir ein Modell aus und betrachten die Koeffizienten?
- Wie können wir die Koeffizienten über mehrere Modelle hinweg betrachten?

Bayesian Model Averaging (BMA)

- Im Bayesianischen Ansatz sind wir im Allgemeinen nicht auf der Suche nach dem „besten“ Modell, sondern gewichten die Modelle mit ihrer Wahrscheinlichkeit
- In BMA kommen Modellauswahl und Parameterschätzung zusammen:
 - Wie wahrscheinlich sind die Modelle in Relation zueinander: Wie wahrscheinlich ist ein Modell relativ zu den anderen Modellen?
 - Wir betrachten einen Prädiktor über alle Kandidatenmodelle, die den Prädiktor einschließen:
 - Wie wahrscheinlich ist es, dass ein Prädiktor im (wahren) Modell ist?
 - Wie viel Einfluss hat ein Prädiktor gemittelt über die Modelle?

BMA in JASP

- BMA ist in JASP für Bayesian Linear Regression verfügbar
- Wir betrachten exemplarisch das Happiness Beispiel und ausgewählte Teile des Outputs (die 10 besten Modelle)

BMA im Regressionsbeispiel (1)

Models	P(M)	P(M data)
GDP + LifeExpect + Fam + Free	0.016	0.498
GDP + LifeExpect + GovernCorrupt + Fam + Free	0.016	0.297
GDP + LifeExpect + Fam + Free + Gen	0.016	0.137
GDP + LifeExpect + GovernCorrupt + Fam + Free + Gen	0.016	0.053
LifeExpect + GovernCorrupt + Fam + Free	0.016	0.006
GDP + Fam + Free	0.016	0.003
LifeExpect + Fam + Free	0.016	0.002
GDP + Fam + Free + Gen	0.016	0.002
GDP + GovernCorrupt + Fam + Free	0.016	8.481×10^{-4}
LifeExpect + GovernCorrupt + Fam + Free + Gen	0.016	5.494×10^{-4}

- $P(M)$: **Prior Probability:**

Wahrscheinlichkeit für die einzelnen Modelle, wird hier als gleich für alle Modelle angenommen

- $P(M|data)$: **Posterior Model Probability:**

- Wahrscheinlichkeit für ein Modell nach Betrachten der Daten

- Das Modell welches GDP, Life Expectancy, Family und Freedom ist nach Betrachten der Daten am wahrscheinlichsten (.498)

BMA im Regressionsbeispiel (2)

Posterior Summary

Posterior Summaries of Coefficients

Coefficient	P(incl)	P(excl)	P(incl data)	P(excl data)	Mean	SD	95% Credible Interval	
							Lower	Upper
Intercept	1.000	0.000	1.000	0.000	5.354	0.040	5.281	5.434
GDP	0.500	0.500	0.991	0.009	0.787	0.220	0.350	1.168
LifeExpect	0.500	0.500	0.995	0.005	1.295	0.344	0.632	1.953
GovernCorrupt	0.500	0.500	0.358	0.642	0.332	0.529	0.000	1.483
Fam	0.500	0.500	1.000	2.864×10^{-5}	1.081	0.205	0.682	1.465
Free	0.500	0.500	1.000	2.667×10^{-4}	1.711	0.349	0.966	2.328
Gen	0.500	0.500	0.193	0.807	0.090	0.234	0.000	0.745

- $P(\text{incl}|\text{data})$: Posterior Wahrscheinlichkeit, dass der Prädiktor im (wahren) Modell ist:
 - Wir addieren die Posterior Wahrscheinlichkeiten aller Modelle, die den Prädiktor einschließen
 - Die Wahrscheinlichkeit, dass GDP im Modell ist nachdem wir die Daten gesehen haben ist 99%
- *Mean*:
 - Der mittlere Koeffizient gemittelt über alle Modelle in denen er vorkommt, gewichtet mit deren Posterior Wahrscheinlichkeiten
 - Der Regressionskoeffizient für GDP ist im Mittel 0.787
- 95% Credible Interval:
 - Die Intervalgrenzen sind auch hier gemittelt über alle Modelle die den Prädiktor einschließen gewichtet mit ihren Wahrscheinlichkeiten
 - Wir können mit 95% Wahrscheinlichkeit sagen, dass der Koeffizient für GDP zwischen 0.35 und 1.168 liegt, gegeben GDP ist im wahren Modell

BMA - Recap

- Wir beachten nicht nur die Unsicherheit bei Parameterwerten, sondern auch die Unsicherheit bei der Modellauswahl
- Die Parameterschätzung basiert so nicht nur auf einem Modell, das mehr oder weniger plausibel ist, sondern auf allen Modellen, die wir betrachten

Inhalt

- Einführung
- Grundlagen
- Parameter schätzen
- Hypothesen testen
- Exkurs: JASP
- Modelle auswählen
- **Zusammenfassung**

Zusammenfassung

- Als Bayesianer modellieren wir Unsicherheit mithilfe von Wahrscheinlichkeitsverteilungen
- Mithilfe von Bayes Theorem lassen wir unsere a-priori Überzeugungen durch die Daten updaten, um zu neuen Schlüssen zu gelangen
- Bayesianische Analysen lassen fundamental andere Schlüsse zu als klassische:
 - Wie groß ist die Wahrscheinlichkeit, dass ein Parameter in einem bestimmten Intervall liegt?
 - Wie viel wahrscheinlicher ist die Alternativhypothese?
 - Wie groß ist ein Parameter über verschiedene plausible Modelle hinweg?

Praktische Empfehlungen – Prior

- Für Anwender mag die Auswahl der Prior abschreckend wirken, aber:
 - Sie müssen in den wenigsten Fällen von den Standard-Priors wie sie in der Literatur beschrieben und in Software implementiert sind, abweichen
 - Sie können deren Einstellungen ändern (z.B. andere Varianz in Normalverteilung)
 - Ausprobieren welche Einstellungen Ihren Annahmen entsprechen
- Sollte eine informative Prior benutzt werden, sollte dies begründet werden
- Eine Sensitivitätsanalyse, bzw. Robustheitsanalyse ist unerlässlich: Dabei wird dieselbe Analyse mit verschiedenen Prior Einstellungen gerechnet und die Ergebnisse verglichen
- Transparenz ist wichtig
- Die Prior ist eine Chance:
 - Warum sollten wir Vorwissen nicht in die Analyse einfließen lassen: kumulative Wissenschaft
 - Bei wenigen Daten kann uns die Prior vor voreiligen Schlüssen bewahren
 - Ohne Prior keine Posterior

Praktische Empfehlungen – Software

- Probieren Sie JASP:
 - Selbst wenn Sie nur klassische Analysen rechnen wollen
 - Es gibt viel Online-Material zu verschiedensten Analysen
 - In JASP selbst gibt es kommentierte Datenbeispiele und Hilfeseiten
 - Die Analysen werden kontinuierlich erweitert und verbessert

Leseempfehlungen

- Bayesian statistics:

- van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child Development, 85*, 842–860. <https://doi.org/10.1111/cdev.12169>
- Wagenmakers, E. J., Morey, R. D., & Lee, M. D. (2016). Bayesian benefits for the pragmatic researcher. *Current Directions in Psychological Science, 25*(3), 169-176. <https://doi.org/10.1177/0963721416643289>
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139087759>

- JASP:

- Wagenmakers, E.J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q.F., Dropmann, D., Boutin, B. & Meerhoff, F. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review, 25*, 58-76. <https://doi.org/10.3758/s13423-017-1323-7>

- MCMC:

- van Ravenzwaaij, D., Cassey, P., & Brown, S.D. (2018). A simple introduction to Markov chain Monte–Carlo sampling. *Psychonomic Bulletin & Review, 25*, 143–154. <https://doi.org/10.3758/s13423-016-1015-8>

Schlusswort

Anhang

MCMC – Basics für Anwender

- MCMC ist ein Oberbegriff für verschiedene Sampling-Methoden die Posterior zu approximieren
- Dabei werden Werte (=Samples) sukzessive aus Wahrscheinlichkeitsverteilungen gezogen bis man sicher ist, dass die gezogenen Werte die Zielverteilung (=Posterior) gut genug approximieren
- MCMC folgt einigen grundlegenden Prinzipien:
 - Es werden sehr viele Samples gezogen, genannt *Iterationen*
 - Das Sampling wird mehrmals mit neuen Ausgangswerten neu gestartet, jede Reihe stellt dann eine *Chain* dar
 - Der erste Teil jeder Chain wird verworfen, da diese Werte oft von den Startwerten abhängen, genannt *Burnin*
 - Oft korrelieren die Werte nah beieinander liegender Iterationen, deswegen verwirft man bspw. jeden dritten oder vierten Wert, genannt *Thinning*
- Oft stellt man die Samples als Liniendiagramm (rechts) dar
- Wenn die Approximierung gut funktioniert, sieht das Liniendiagramm aus wie eine haarige Raupe
- Wenn die Sampling-Verteilung die Zielverteilung gut approximiert spricht man von *Konvergenz*

