

# Kurze Einführung in Bayes-Statistik mit R für Ornithologen

Fränzi Korner-Nievergelt & Ommo Hüppop

---

Korner-Nievergelt F & Hüppop O 2016: A short introduction to Bayes statistics with R for ornithologists.  
Vogelwarte 54: 181-194.

We give a short introduction to the principles of Bayesian data analyses and guide the reader through two examples using R. Differences and similarities to frequentist statistics are explained. Bayesian analysis is combining prior information (e. g. from the literature) with information in the data to obtain a posterior distribution of the parameter of interest. The prior and the posterior distributions are probability density functions that describe mathematically what is known of a parameter prior to and after having collected the data. In the first example, we explore the posterior distribution of a proportion of males in a population. In the second example, we estimate the difference in mean wing length between the sexes of a songbird species once without prior information and once using information from the literature as prior information. For both cases, we calculate the posterior probability of the hypothesis that the average wing length in males is at least 1 mm longer than in females. For these examples, we use the R-packages “arm” and “rstanarm”. We explain why the results from a Bayesian data analyses are often easier to interpret and more useful than the result from a classical null hypothesis test. Finally, we recommend some introductory Bayesian text books. The purpose of this article is to help ornithologists understand Bayesian analyses and to provide an easy start-up for doing own Bayesian analyses with R.

✉ FK-N: oikostat GmbH, Ausserdorf 43, CH-6218 Ettiswil, Schweiz. E-Mail: [fraenzi.korner@oikostat.ch](mailto:fraenzi.korner@oikostat.ch) und Schweizerische Vogelwarte, Seerose 1, CH-6204 Sempach, Schweiz  
OH: Institut für Vogelforschung „Vogelwarte Helgoland“, An der Vogelwarte 21, D-26386 Wilhelmshaven

---

## Einleitung

Immer häufiger werden in ornithologischen Arbeiten Daten mittels Bayes-Statistik ausgewertet. Der Hauptgrund für diese Wende von der uns vertrauten klassischen („frequentistischen“) Statistik zur Bayes-Statistik ist einerseits in der hohen Flexibilität der Methoden der Bayes-Statistik zu suchen. Andererseits liefert sie im Vergleich zu frequentistischen Tests einfacher zu interpretierende Resultate und erlaubt, früheres Wissen in die Analysen einfließen zu lassen. Biologen, welche den Zugang zur Bayes-Statistik gefunden haben, schätzen die Flexibilität, unterschiedlichste biologisch relevante Hypothesen untersuchen und bereits vorhandenes Wissen formalisiert in die Analysen aufnehmen zu können. Viele mit der klassischen frequentistischen Statistik aufgewachsene Biologen tun sich jedoch mit der Bayes-Statistik schwer, meist weil sie ihnen zu kompliziert erscheint oder weil die vertraute Statistiksoftware keine Bayes-Methoden implementiert hat. Im Zeitalter von R (R Development Core Team 2016) wird der zweite Grund aber immer unbedeutender, da beständig neue R-Pakete geschrieben werden, welche die Anwendung von Bayes-Methoden erleichtern. Hier geben wir eine kurze, (fast) formelfreie Einführung in die Prinzipien der Bayes-Statistik (Tipps für weiterführende Literatur gibt es ganz am Ende). Danach zeigen wir, wie man einen einfachen Mittelwertvergleich auf Bayesisch mit R durchführt. Unser R-Code ist so geschrieben, dass die Beispiele gleich am eigenen Rechner durchgeführt,

ausprobiert und nach eigenen Vorstellungen abgewandelt werden können. Am Ende geben wir einen kurzen Ausblick, wie publiziertes Wissen aus der Literatur mit aktuellen Daten kombiniert werden kann.

Wir setzen nur voraus, dass der Leser weiß, wie R gestartet wird, wie die zusätzlich benötigten Pakete „arm“ und „rstanarm“ installiert werden (ein vorheriges Update auf die neueste stabile Version von R wird empfohlen), wie eine Funktion in R ausgeführt wird und wie Objekte in R betrachtet werden. Wir gaben bereits früher eine Einführung in die inzwischen sehr weit verbreitete Statistiksoftware R (Korner-Nievergelt and Hüppop 2010).

Der R-Code zu diesem Artikel ist auf den Internetseiten der „Vogelwarte“ unter folgender Adresse verfügbar: [www.do-g.de/publikationen/vogelwarte/inhalte-online](http://www.do-g.de/publikationen/vogelwarte/inhalte-online). Klicken Sie dort bitte auf den Link für Heft 3/2016.

## Der Satz von Bayes

Der Geistliche Thomas Bayes (1702 - 1761) beschrieb, wie basierend auf aktuellem Wissen und neuen Beobachtungen die Wahrscheinlichkeit berechnet werden kann, dass eine Hypothese zutrifft (Bayes 1763). Das ist der berühmte Satz von Bayes,  $P(H|X) = P(X|H) \times P(H) / P(X)$ : Die Wahrscheinlichkeit der Hypothese, nachdem wir die Daten betrachtet haben ( $P(H|X)$ ), ist gleich der Wahrscheinlichkeit der Daten, angenommen die Hypothese trifft zu ( $P(X|H)$ ), mal der Wahrscheinlichkeit

der Hypothese, bevor wir die Daten betrachtet haben ( $P(H)$ ), geteilt durch die Wahrscheinlichkeit der Daten ( $P(X)$ ). Zum Beispiel kann mit Hilfe dieses Satzes die Wahrscheinlichkeit berechnet werden, dass die Hypothese „Der Männchenanteil einer Population der Amsel *Turdus merula* liegt über 0,6.“ zutrifft, gegeben das bereits vorhandene Wissen (aus der Literatur) sowie Daten aus einer aktuellen Zählung in der untersuchten Amselpopulation.

Thomas Bayes zweifelte allerdings so stark an seiner Erfindung, dass sein Satz erst nach seinem Tode publiziert wurde. Ohne Computer war der Satz von Bayes zunächst nur für ganz einfache Beispiele anwendbar. Erst die sprunghaften Fortschritte der Computertechnik sowie die Entwicklung von Simulationstechniken, insbesondere „Markov chain Monte Carlo“ (MCMC) Simulationen (Smith et al. 1985; Gelfand & Smith 1990; Gilks et al. 1996), ermöglichen es uns heute, bayessche Methoden einfach auf jedem PC durchzuführen. Wir möchten mit dieser Einführung den Lesern der „Vogelwarte“ den Zugang zu diesen Methoden schmackhaft machen und ihn erleichtern.

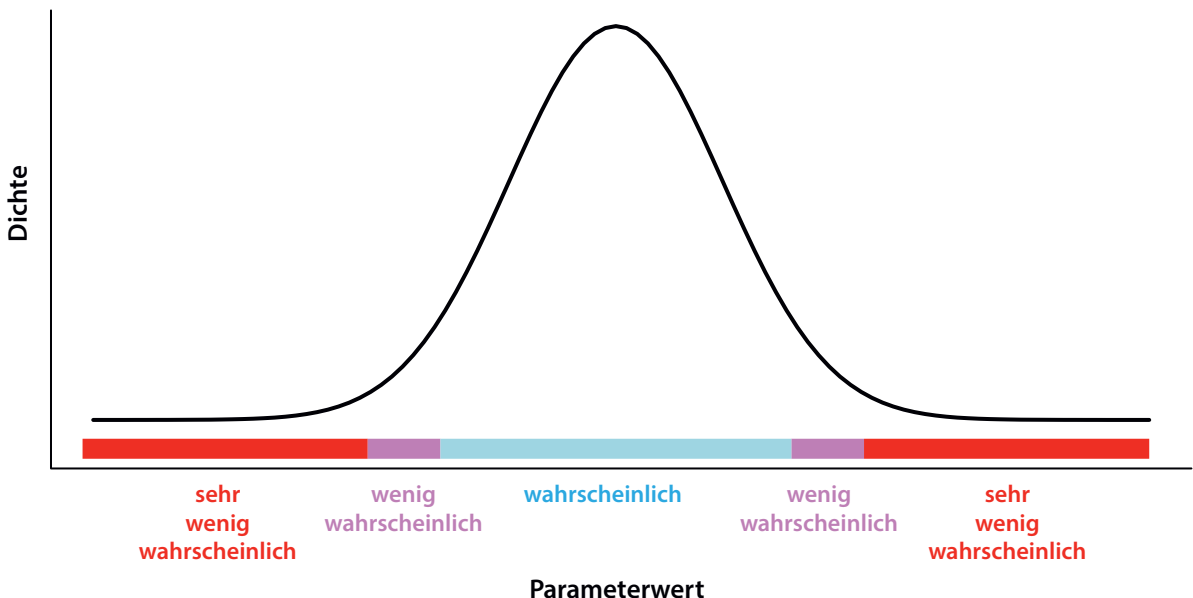
### Das Prinzip der Bayes-Statistik

Bayes-Statistik besteht darin, den Satz von Bayes systematisch zu verwenden, um Schlüsse aus Daten zu ziehen. Dabei wird das bereits vor der Datenerhebung vorhandene Wissen - das A-priori-Wissen - mit der Information in den erfassten Daten zum A-posteriori-Wissen kombiniert.

Das Wissen über einen Parameter (a-priori und a-posteriori) wird mathematisch als Wahrscheinlichkeitsverteilung ausgedrückt (A-priori-Verteilung und A-posteriori-Verteilung). Eine solche Wahrscheinlichkeitsverteilung besitzt für jeden möglichen Parameterwert eine Dichte (Abb. 1). Je höher die Dichtefunktion ist, desto wahrscheinlicher ist der betreffende Parameterwert. Die Fläche unter der Kurve einer Wahrscheinlichkeitsverteilung ist eins, d. h. wir sind hundertprozentig sicher, dass der wahre Parameterwert irgendeinen Wert entlang des Wertebereichs der x-Achse annimmt. Diese Aussage ist natürlich nicht sehr nützlich. Dass die Fläche unter der Kurve als Wahrscheinlichkeit interpretiert werden kann, ermöglicht es uns aber, Wahrscheinlichkeiten für viele mögliche Hypothesen zu berechnen. Zum Beispiel ist die Wahrscheinlichkeit der Hypothese, dass der Männchenanteil in einer Population größer als 0,6 ist, gleich der Fläche unter der Kurve der Wahrscheinlichkeitsverteilung des Männchenanteils für Werte größer als 0,6.

Eine bayessche Datenanalyse besteht aus folgenden Schritten:

1. Definition der A-priori-Verteilung anhand vorhandener Daten aus der Literatur oder von früheren Studien.
2. Sammeln und Beschreiben der eigenen Daten, Erstellen eines Datenmodells.
3. Kombination der A-priori-Verteilung mit den eigenen Daten, um die A-posteriori-Verteilung zu erhalten.



**Abb. 1:** Das Wissen über einen Parameter wird in der Bayes-Statistik mit Wahrscheinlichkeitsverteilungen beschrieben. Wertebereiche mit kleinen Dichtewerten sind eher unwahrscheinlich, während Wertebereiche mit hohen Dichtewerten mögliche Parameterwerte sind. - *In Bayes statistics, the knowledge of a parameter value is expressed by probability density functions. Ranges of parameter values with low density values are unlikely whereas it is more likely that the true parameter has a value within the range of high density values.*

4. Überprüfung der Modellanpassung.
5. Schlussfolgerungen aus der A-posteriori-Verteilung ziehen.

Zuerst erstellen wir also eine A-priori-Verteilung für den interessierenden Parameter, in unserem Beispiel den Männchenanteil. Dazu tragen wir zusammen, was wir über den Männchenanteil in Amselpopulationen wissen, bevor wir die eigenen Daten aufnehmen. Danach suchen wir nach einer Wahrscheinlichkeitsverteilung, welche dieses Wissen am zutreffendsten beschreibt. Wir vermuten, dass der Anteil männlicher Amseln um 0,5 herum betragen muss. Aus der Literatur entnehmen wir, dass gemessene Männchenanteile bei adulten Vögeln zwischen 0,4 und 0,75 liegen (Liker et al. 2013). Wir müssen aber davon ausgehen, dass bei vielen Vogelarten Männchen einfacher beobachtet werden können als Weibchen (z. B. Amrhein et al. 2012). Deshalb gehen wir a-priori trotz publizierter Männchenanteile von über 0,5 real von einem ausgeglichenen Geschlechterverhältnis aus. Zudem möchten wir a-priori nicht annehmen, dass der Männchenanteil hundertprozentig zwischen 0,4 und 0,75 liegt. Wir sind aber hundertprozentig sicher, dass der Männchenanteil zwischen 0 und 1 liegt, und wir können basierend auf unserem Vorwissen annehmen, dass Werte unter 0,15 und über 0,85 eher unwahrscheinlich sind. Wir suchen also nach einer A-priori-Verteilung mit einem Wertebereich von 0 bis 1, einem Mittelwert von 0,5 und einer so breiten Streuung, dass 95 % der möglichen Werte zwischen 0,15 und 0,85 liegen (punktierte Linie in Abb. 2).

Die Suche nach der Verteilungsfunktion für die A-priori-Verteilung ist vergleichbar mit dem Auswählen des passenden Tests in einer frequentistischen Datenanalyse: Für sie braucht man einen Überblick über die häufigsten Tests und muss ihre Eigenschaften und Voraussetzungen kennen. Für die bayessche Datenanalyse ist ein Überblick über die verschiedenen Wahrscheinlichkeitsverteilungen hilfreich. Lehrbücher der Bayes-Statistik enthalten meistens tabellarische Übersichten über eine Auswahl von häufig verwendeten Wahrscheinlichkeitsverteilungen und ihren Eigenschaften (z. B. Gelman et al. 2014). Für die A-priori-Verteilung des Männchenanteils wählen wir beispielsweise eine Betaverteilung, weil diese Verteilung einen Wertebereich von 0 bis 1 aufweist und mit zwei Parametern,  $a$  und  $b$ , Mittelwert und Streuung festgelegt werden können. Der Mittelwert ist  $a/(a+b)$ . Die Varianz nimmt mit zunehmenden Werten von  $a$  und  $b$  ab. Haben  $a$  und  $b$  die gleichen Werte, ist der Mittelwert 0,5. Je höher  $a$  und  $b$  sind, desto enger ist die Wahrscheinlichkeitsverteilung. In unserem Beispiel finden wir relativ schnell mittels „Versuch und Irrtum“ die Parameterwer-

te  $a$  und  $b$ , welche eine Betaverteilung mit 95 % Intervall von 0,15 - 0,85 ergeben. Wir haben dazu den R-Code `qbeta(0.025, a, b)` und `qbeta(0.975, a,b)` verwendet und  $a$  und  $b$  solange variiert, bis der Code 0,15 bzw. 0,85 ausgab. Alternativ könnte man die Parameter  $a$  und  $b$  auch aus den Quantilen oder aus Varianz und Mittelwert berechnen, z. B. mit der Funktion `get.beta.par` aus dem package „riskDistributions“. Als Resultat verwenden wir die A-priori-Verteilung  $Beta(3,3)$ . Mit dieser A-priori-Verteilung drücken wir mathematisch aus, dass wir zu 95 % sicher sind, dass der Männchenanteil in der Amselpopulation zwischen 0,15 und 0,85 liegt und dass der Wert zwischen 0 und 1 liegen muss.

Nun sind wir bereit, eigene Daten aufzunehmen. Wir beobachten 10 Weibchen und 14 Männchen. Der Männchenanteil in der Stichprobe beträgt also  $14/(14+10) = 0,58$ . Der Informationsgehalt in den Daten wird durch die sog. Likelihood-Funktion gemessen. Diese Funktion beschreibt für jeden möglichen Parameterwert die Wahrscheinlichkeit, genau die beobachteten Daten zu erhalten. Dazu müssen wir uns überlegen, wie die Daten zustande kommen (welche Werte können sie annehmen und wie streuen sie?). Wir suchen eine Wahrscheinlichkeitsverteilung, welche die Verteilung der Daten am besten beschreibt. Dieses Vorgehen ist uns aus der frequentistischen Statistik bereits vertraut. Im Beispiel der Anzahl Männchen in einer Population („14 von 24“) nehmen wir eine Binomialverteilung an. Basierend auf der Binomialverteilung berechnen wir die Wahrscheinlichkeit, in Abhängigkeit vom Männchenanteil in der Population

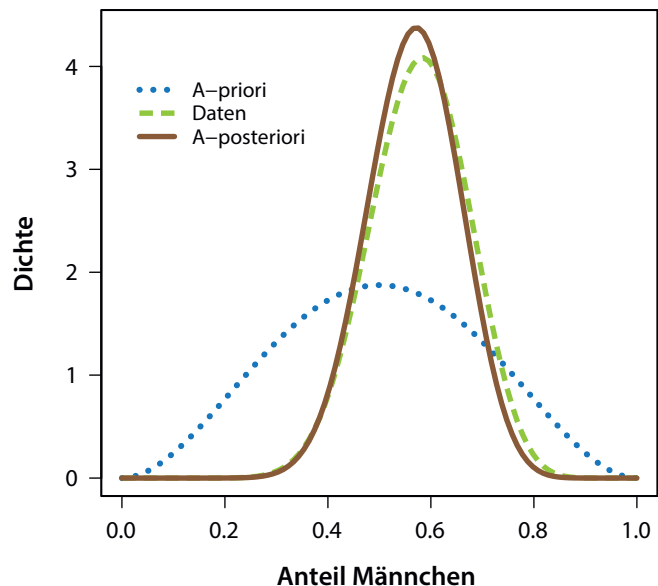


Abb. 2: Die A-priori-Verteilung, die Likelihood-Funktion der Daten sowie die A-posteriori-Verteilung für den Anteil Männchen im Beispiel der Amseln (siehe Text). - *The prior distribution (A-priori), the likelihood function (Daten) and the posterior distribution (A-posteriori) for the proportion of males in the Blackbird example. See box for R-code to produce this figure.*

genau 14 Männchen und 10 Weibchen zu beobachten (Abb. 2). In Abb. 2 sind diese Wahrscheinlichkeitswerte so skaliert, dass die Fläche unter der Likelihood-Funktion (gestrichelte Linie) genau 1 ist, damit sie als Wahrscheinlichkeitsverteilung verwendet werden kann.

Nun werden die Daten und die A-priori-Verteilung mittels Satz von Bayes kombiniert, um die A-posteriori-Verteilung zu erhalten. Um diese Kombination technisch zu bewältigen, sind meistens moderne, komplizierte Simulationstechniken nötig. Zum Glück gibt es heute ein-

fach zu benutzende Funktionen und Software, welche dies für uns übernehmen. Im vergleichsweise einfachen Beispiel des Männchenanteils in der Amselpopulation sind die Berechnungen sogar ohne Simulation möglich (siehe Box). Die resultierende A-posteriori-Verteilung (durchgezogene Linie in Abb. 2) hat einen Mittelwert von 0,57 und ein 95 % Intervall von 0,39 bis 0,74. Wir können also aufgrund des Vorwissens und der Daten zu 95 % sicher sein, dass der wahre Männchenanteil in der Population zwischen 0,39 und 0,74 liegt.

#### R-Code für die Berechnung der A-posteriori-Verteilung des Männchenanteils einer Amselpopulation:

Als A-priori-Verteilung wählen wir eine Betaverteilung. Die Betaverteilung ist eine kontinuierliche Wahrscheinlichkeitsverteilung für den Wertebereich zwischen 0 und 1. Deshalb eignet sie sich, um das Wissen über eine Proportion mathematisch auszudrücken. Die Betaverteilung hat zwei Parameterwerte, a und b. Setzen wir beide Parameter auf 3, resultiert die in Abb. 2 dargestellte A-priori-Verteilung mit Mittelwert 0,5 und den von uns postulierten 2,5 %- und 97,5 %-Quantilen von 0,15 und 0,85.

```
qbeta(c(0.5,0.025,0.975), 3, 3)
[1] 0.5000000 0.1466328 0.8533672
```

In diesem einfachen Fall kann die A-priori-Verteilung mit den Daten kombiniert werden, indem die Anzahl Männchen zum Parameter a der A-priori-Verteilung und die Anzahl Weibchen zum Parameter b addiert werden (siehe Text). Daraus erhalten wir die neuen Parameter a und b der A-posteriori-Verteilung, welche einen Mittelwert von 0,56 und 2,5 %- und 97,5 %-Quantilen von 0,39 und 0,74 ergeben (fette Linie in Abb. 2).

```
qbeta(c(0.5,0.025,0.975), 3+14, 3+10)
[1] 0.5681675 0.3893628 0.7355447
```

#### Der R-Code für die Erstellung von Abb. 2 lautet:

```
theta <- seq(0, 1, length=100)
priora <- 3
priorb <- 3
prior <- dbeta(theta, priora,priorb)
nmales <- 14
nfemales <- 10
```

Die Likelihood des Parameterwertes theta ist die Wahrscheinlichkeit, genau 14 Männchen und 10 Weibchen zu zählen, falls der Parameterwert theta wahr ist. Die Likelihoodfunktion beschreibt die Likelihood über alle möglichen Parameterwerte. (Die „Maximum Likelihood“-Schätzung eines Parameters entspricht dem Parameterwert am Gipfel der Likelihoodfunktion). Zur Berechnung der Likelihoodfunktion verwenden wir die R-Funktion `dbinom`. Damit die Fläche unter der resultierenden Kurve genau 1 ist, d. h. die Kurve eine Wahrscheinlichkeitsverteilung beschreibt, teilen wir diese Wahrscheinlichkeit durch die Fläche unter der Kurve (Nenner im folgenden R-Code):

```
likelihood <- dbinom(prob=theta, x=nmales, size=nmales+nfemales)/(sum(dbinom(prob=theta, x=nmales, size=nmales+nfemales))*(theta[2]-theta[1]))
```

```
posterior <- dbeta(theta, nmales+priora, nfemales+priorb)
```

```
plot(0:1, range(posterior), type="n", xlab="Anteil Männchen", ylab="Dichte", las=1, cex.lab=1.4, cex.axis=1.2)
lines(theta, prior, lty=3, lwd=3)
lines(theta, likelihood, lty=2, lwd=3)
lines(theta, posterior, lty=1, lwd=3)
legend(0, 4, lty=c(3,2,1), lwd=3, legend=c("A-priori", "Daten", "A-posteriori"), bty="n")
```

Zur Übung können Sie ausprobieren, wie A-priori-Wissen und die erhobenen Daten das Ergebnis beeinflussen, indem Sie die Größe der beobachteten Stichprobe oder die beiden Parameter der A-priori-Verteilung verändern.

Damit haben wir schon wichtige Aspekte der Bayes-Statistik kennengelernt:

Die A-posteriori-Verteilung entspricht einem gewichteten Mittel aus der A-priori-Verteilung und der Likelihoodfunktion der Daten. Ihre Gewichte entsprechen der jeweiligen Präzision, d. h. dem Informationsgehalt des A-priori-Wissens und der Daten. Ist das A-priori-Wissen hoch, d. h. sind wir uns also sehr sicher, wo der wahre Parameterwert liegen soll, und ist der Informationsgehalt in den Daten z. B. wegen kleiner Stichprobengröße oder hoher Streuung gering, dann wird die A-posteriori-Verteilung näher an der A-priori-Verteilung sein als an den Daten. Das ist gut so, denn niemand möchte bestehendes Wissen wegen ein paar weniger unerwarteter neuer Beobachtungen umkrempeln oder gar verwerfen. Ist aber das A-priori-Wissen schwach und die Daten weisen eine hohe Stichprobengröße auf, dann wird die A-posteriori-Verteilung nahe bei den Daten liegen.

Werden nicht-informative A-priori-Verteilungen, d. h. kein früheres Wissen, verwendet (im Amselbeispiel die Uniformverteilung, d. h. eine horizontale Dichtefunktion zwischen 0 und 1, Parameter  $priora = 1$  und  $priorb = 1$ ), dann entspricht das Resultat der bayesschen Analyse jenem einer frequentistischen Analyse.

Werden zwei Datensätze gepoolt (also rechnerisch der Stichprobenumfang erhöht) und für die Analyse eine nicht-informative A-priori-Verteilung verwendet, entspricht dies rechnerisch dem Resultat, das sich ergäbe, wenn die A-posteriori-Verteilung der Analyse des ersten Datensatzes als A-priori-Verteilung für die Analyse des zweiten Datensatzes verwendet würde. Die oben verwendete A-priori-Verteilung Beta(3,3) entspricht also der Information über den wahren Männchenanteil in der Population, den wir aus einer (zusätzlichen) Stichprobe von sechs Individuen mit drei Männchen und drei Weibchen erhalten hätten.

### Praktisches Beispiel in R: Vergleich zweier Mittelwerte

In einer Pilotstudie haben wir drei männliche und sechs weibliche adulte Schneesperlinge *Montifringilla nivalis* gefangen. Von jedem Individuum wurde die Flügellänge gemessen. Wir fragen uns, ob ein Geschlechtsdimorphismus in der Flügellänge vorhanden ist und wie groß der Unterschied gegebenenfalls ist. Die Daten bestehen aus zwei Variablen, die wir in R in einem „data frame“ bündeln.

```
geschlecht <- c("M", "M", "M", "W", "W", "W", "W", "W", "W")
fluegel <- c(125, 122, 121, 119, 119.5, 116, 121, 121, 117.5)
dat <- data.frame(geschlecht=geschlecht, fluegel=fluegel)
dat
```

```
  geschlecht fluegel
1          M  125.0
2          M  122.0
3          M  121.0
4          W  119.0
5          W  119.5
6          W  116.0
7          W  121.0
8          W  121.0
9          W  117.5
```

Die Modelle und Modellannahmen, welche wir treffen müssen, unterscheiden sich nicht zwischen der Bayes- und der frequentistischen Statistik. In diesem Fall nehmen wir an, dass die individuellen Flügellängen um den (geschlechtsspezifischen) Mittelwert herum normalverteilt sind. In der Bayes-Statistik wird dieses Modell als Datenmodell bezeichnet, um es von den Modellen für die Parameter (A-priori-Verteilungen und A-posteriori-Verteilungen) zu unterscheiden. Wenn wir nicht-informative A-priori-Verteilungen für die Modellparameter verwenden, können wir zur Anpassung des Modells an die Daten die aus der frequentistischen Statistik bekannten R-Funktionen `lm`, `glm`, `lmer`, `glmer` verwenden. Unser konkretes Beispiel entspricht einem zweiseitigen t-Test für unabhängige Stichproben bzw. einer Ein-Weg-ANOVA über zwei Gruppen, die wir allgemein mit der Funktion `lm` anpassen können:

```
mod <- lm(fluegel~geschlecht, data=dat)
```

Wie nach jeder Modellanpassung werden als erstes die Modellannahmen überprüft. Hier betrachten wir die Residuen grafisch (z. B. Stahel 2002).

```
par(mfrow=c(2,2))
plot(mod)
```

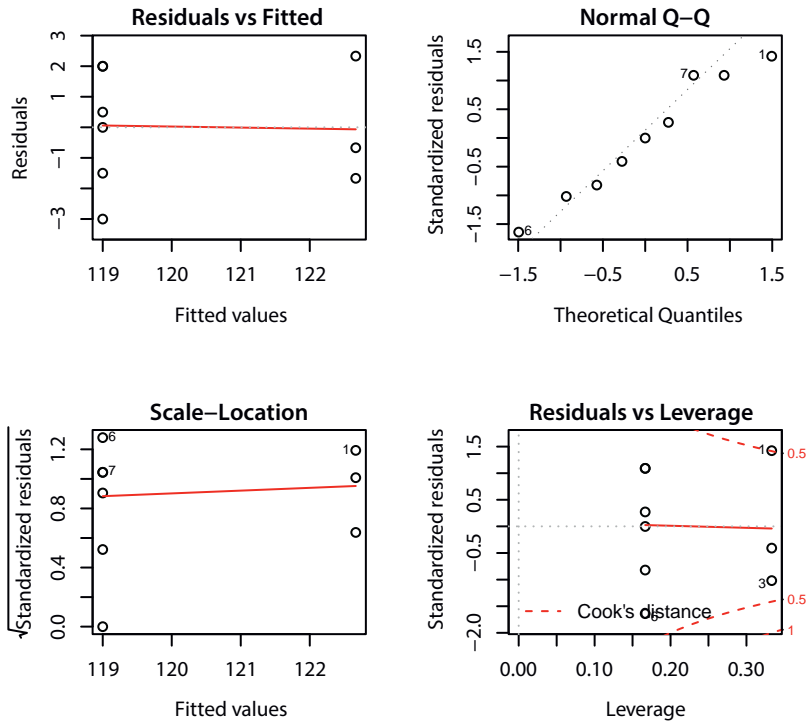


Abb. 3: Die grafische Analyse der Residuen zeigt keine gravierenden Verletzungen der Modellannahmen. Oben links: Die Mittelwerte der Residuen sind bei beiden Gruppen Null; oben rechts: Die Quantilen der Residuen entsprechen den Quantilen einer Normalverteilung; unten links: Die Varianz ist für alle Erwartungswerte ungefähr gleich groß; unten rechts: es gibt keine Beobachtungen mit besonders starkem Einfluss. - *Graphical analyses of the residuals do not indicate any serious violation of model assumptions. Upper left: the means of the residuals are close to zero for both groups; upper right: the quantiles of the residuals correspond to the quantiles of a normal distribution; lower left: the residual variance does not depend on the fitted value; lower right: there are no excessively influential observations.*

Nachdem wir uns vergewissert haben, dass das Modell gut zu den Daten passt (Abb. 3), betrachten wir das Resultat:

`mod`

Call:

`lm(formula = fluegel ~ geschlecht, data = dat)`

Coefficients:

(Intercept) geschlechtW  
122.667      -3.667

Der Achsenabschnitt (Intercept) von 122,7 entspricht dem Mittelwert der Flügelängen für die Männchen. Der zweite Modellparameter (geschlechtW) ist die Differenz zwischen dem Mittelwert der Männchen und jenem der Weibchen. Die Weibchen haben also im Durchschnitt einen um 3,7 mm kürzeren Flügel. Die Streuung der individuellen Messwerte um die geschlechtsspezifischen Mittelwerte wird mit der Standardabweichung der Residuen gemessen. Sie beträgt hier 2,01 mm (im Modell haben wir angenommen, dass diese Streuung in beiden Geschlechtern gleich ist):

`summary(mod)$sigma`

[1] 2.005944

Bisher unterscheidet sich das Vorgehen nicht von den herkömmlichen frequentistischen Methoden. Erst jetzt, um die Unsicherheit der geschätzten Modellparameter zu quantifizieren, benutzen wir bayessche Methoden. Dazu müssen wir die A-posteriori-Verteilungen der Modellparameter erstellen. Die mathematische Berechnung der A-posteriori-Verteilung ist etwas, das wir Biologen lieber Mathematikern überlassen. Dankenswerterweise wurde das für die gängigen linearen Modelle bereits gemacht und die Prozeduren in einfach zu verwendende R-Funktionen verpackt. So enthält das Paket „arm“ (Gelman & Hill 2007) die Funktion `sim`. Diese Funktion ermöglicht die Berechnungen der A-posteriori-Verteilungen der Modellparameter von gängigen linearen Modellen. Sie nimmt dabei nicht-informative A-priori-Verteilungen (s. o.) an. Die Funktion `sim` zieht Zufallswerte aus der A-posteriori-Verteilung der Modellparameter. Die technischen Details des Verfahrens sind zu kompliziert, um hier behandelt zu werden, sind aber ausführlich beschrieben in Gelman et al. (2014). Wir können viele solcher Zufallswerte aus der A-posteriori-Verteilung ziehen und dann von diesen Zufallswerten ein Histogramm erstellen, um die A-posteriori-Verteilung anzusehen. Hier ziehen wir 5.000 zufällige Werte aus der A-posteriori-Verteilung:

```
library(arm)
bsim <- sim(mod, n.sim=5000)
str(bsim)
Formal class 'sim' [package "arm"] with 2 slots
..@ coef: num [1:5000, 1:2] 117 123 122 124 123 ...
..- attr(*, "dimnames")=List of 2
..- ..$: NULL
..- ..$: chr [1:2] "(Intercept)" "geschlechtW"
..@ sigma: num [1:5000] 4.7 2.02 1.77 2.05 1.73 ...
```

Mit der Funktion `str` können wir die Struktur des durch die Funktion `sim` erstellten Objektes (`bsim`) betrachten. Das Objekt enthält zwei Elemente (slots), „coef“ und „sigma“. Das Element „coef“ ist eine Matrix mit 5.000 Zeilen und 2 Spalten. Die erste Spalte enthält 5.000 Zufallswerte aus der A-posteriori-Verteilung des Achsenabschnitts. Die zweite Spalte enthält Zufallswerte aus der A-posteriori-Verteilung der Differenz zwischen den geschlechtsspezifischen Mittelwerten. Das zweite Element „sigma“ enthält 5.000 Zufallswerte aus der A-posteriori-Verteilung der Residuen-Standardabweichung. Nun können wir die A-posteriori-Verteilungen visualisieren, indem wir die Zufallswerte in Histogrammen darstellen (Abb. 4). Da es sich um Zufallswerte handelt, werden die Histogramme (und auch die folgenden Berechnungen) bei Ihnen geringfügig anders aussehen als hier.

```
par(mfrow=c(1,2))
for(i in 1:2) hist(bsim@coef[,i], main=names(coef(mod))[i], freq=FALSE)
```

Der Mittelwert der 5.000 Ziehungen aus der A-posteriori-Verteilung sollte nahe den durch die Funktion `lm` erhaltenen bestpassenden Parameterwerten liegen. Abweichungen kommen dadurch zustande, dass hier Mittelwerte aus 5.000 Zufallsziehungen berechnet werden, während die Funktion `lm` die Parameterwerte, welche mit der kleinste Summe der quadrierten Abweichungen zu den Daten einhergehen (bestpassend), mittels Matrixalgebra genau berechnet (siehe z. B. Aitkin et al. 2009). Es ist deshalb genauer, als Resultat die durch `lm` berechneten Werte anzugeben als die Mittelwerte der simulierten Werte aus der A-posteriori-Verteilung. Letztere erhält man z. B. mit der Funktion `apply`, welche eine Funktion (hier `mean`, spezifiziert im dritten Argument der Funktion) entweder zeilen- oder spaltenweise (hier spaltenweise, spezifiziert mit der „2“ im zweiten Argument) auf eine Matrix anwendet (erstes Argument):

```
apply(bsim@coef, 2, mean)
(Intercept) geschlechtW
122.633609   -3.668933
```

Die Standardabweichungen der Zufallswerte aus den A-posteriori-Verteilungen entsprechen den Standardfehlern der Schätzwerte für die Modellparameter:

```
apply(bsim@coef, 2, sd)
(Intercept) geschlechtW
1.372626    1.692902
```

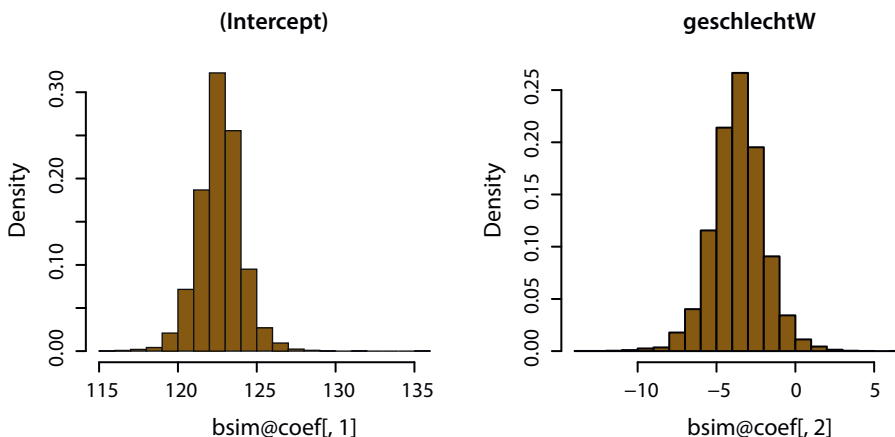


Abb. 4: Histogramme der 5.000 simulierten Werte aus den A-posteriori-Verteilungen für die beiden Modellparameter Achsenabschnitt (Intercept) und Mittelwertsdifferenz. – Histogram of 5,000 simulated values from the a-posteriori distributions for both parameters of the model (intercept and difference of arithmetic means).

Das Kreditabilitätsintervall ist das bayessche Pendant zum Vertrauensbereich. Es ist jenes Intervall, in dem wir, wenn wir das Modell korrekt spezifiziert haben, mit einer bestimmten (oft 95 %) Wahrscheinlichkeit den wahren Parameterwert vermuten, nachdem wir die Daten betrachtet haben. Ein mögliches 95 % Kreditabilitätsintervall ist jenes zwischen den 2,5 % und 97,5 % Quantilen der simulierten Werte aus der A-posteriori-Verteilung:

```
apply(bsim@coef, 2, quantile, prob=c(0.025, 0.975))
      (Intercept) geschlechtW
2.5%   119.8496   -7.0067436
97.5%  125.3028   -0.3465695
```

Basierend auf den gemessenen Flügellängen der drei männlichen und sechs weiblichen Schneesperlinge (und der Annahme, dass die individuellen Messwerte gemäß einer Normalverteilung um die geschlechtsspezifischen Mittelwerte streuen) können wir also zu 95 % sicher sein, dass der Mittelwert der Flügellängen für die Männchen zwischen 119,8 und 125,3 mm liegt. Ebenso sicher sind wir, dass die Differenz der Mittelwerte zwischen Männchen und Weibchen zwischen -7,0 und -0,3 mm liegt. Wie groß ist nun die mittlere Flügellänge der Weibchen? Wenn wir zum Mittelwert der Männchen aus dem linearen Modell (Intercept) die Differenz hinzuzählen, erhalten wir den Mittelwert für die Weibchen:

```
coef(mod)[1] + coef(mod)[2]
(Intercept)
      119
```

Um die A-posteriori-Verteilung des Mittelwerts für die Weibchen zu erhalten, können wir einfach diese Summe für alle 5.000 simulierten Wertepaare aus der gemeinsamen A-posteriori-Verteilung der beiden Modellparameter berechnen:

```
mw <- bsim@coef[,1] + bsim@coef[,2]
```

Hier zählen wir also 5.000 Mal die Mittelwertsdifferenz zum Mittelwert der Männchen und erhalten 5000 Mittelwerte für die Weibchen. Aus diesen 5000 Mittelwerten können wir wiederum ein 95 % Kreditabilitätsintervall berechnen:

```
quantile(mw, prob=c(0.025, 0.975))
      2.5%   97.5%
116.8717 120.9598
```

Wir sind also zu 95 % sicher, dass der Mittelwert der Flügellängen bei den Weibchen zwischen 116,9 und 121,0 mm liegt. Wie sicher sind wir aber beispielsweise, dass die Flügel der Weibchen um mindestens einen Millimeter kürzer sind als jene der Männchen? In der Bayes-Statistik ist es einfach möglich, für solche konkreten Hypothesen, wie „die mittlere Flügellänge der Weibchen ist um mindestens 1 mm kürzer als jener der Männchen“, Wahrscheinlichkeiten zu berechnen. Diese Wahrscheinlichkeit kann man auch schon im Histogramm der simulierten Werte aus der A-posteriori-Verteilung der Differenz (Abb. 4) ablesen. Sie entspricht dem Anteil der Fläche unter der Dichtefunktion der A-posteriori-Verteilung für den Wertebereich kleiner als -1 mm. Näherungsweise kann diese Wahrscheinlichkeit durch den Anteil Werte, die kleiner als -1 sind, an den 5.000 simulierten Parameterwerten berechnet werden:

```
sum(bsim@coef[,2] < -1)/5000
[1] 0.951
```

Aufgrund unserer Daten sind wir also zu 95 % sicher, dass die Weibchen im Durchschnitt einen um mindestens 1 mm kürzeren Flügel haben.

In dieser Pilotstudie ist die Stichprobengröße natürlich sehr klein und entsprechend sind Aussagen, die wir aufgrund der wenigen Daten machen können, sehr unsicher. Hier zeigt sich aber ein großer Vorteil der Bayes-Statistik: Wir können in die Analysen bereits vorhandenes Wissen einfließen lassen.

#### Kombination von früherem Wissen mit aktuellen Daten: informative A-priori-Verteilungen

Sind wir wirklich nur 95 % sicher, dass die Weibchen im Durchschnitt einen um mindestens 1 mm kürzeren Flügel haben als die Männchen? Wir wissen aus der Literatur, dass einige Ornithologen vor uns bereits Flügelmessungen an Schneesperlingen vorgenommen haben. Aus dem Handbuch (Glutz von Blotzheim & Bauer 1997) entnehmen wir folgende Angaben aus einer Untersuchung in den Alpen (Lang 1946):



**Männchen:** Spannweite 121-128 mm (Mittelwert 123,8 mm; n=21)

**Weibchen:** Spannweite 114-120 mm (Mittelwert 116,7 mm; n = 15)

Ein Viertel der Spannweite entspricht bei einer Normalverteilung ungefähr der Standardabweichung. Aus der Standardabweichung erhalten wir den Standardfehler, indem wir sie durch die Wurzel der Stichprobengröße teilen, deshalb sind

```
mm <- 123.8
semm <- (128-121)/4/sqrt(21)
```

Mittelwert und Standardfehler der Flügellänge der Männchen und

```
mw <- 116.7
semw <- (120-114)/4/sqrt(15)
```

Mittelwert und Standardfehler der Flügellängen der Weibchen. Diese beiden Mittelwerte und Standardfehler beschreiben zwei Normalverteilungen, die wir nun als A-priori-Verteilungen für die beiden Mittelwerte im Modell verwenden wollen. Im Moment werden laufend neue Software und R-Pakete entwickelt, um bayessche Analysen durchzuführen. Hier verwenden wir die Funktion `stan_glm` aus dem R-Paket „rstanarm“. Diese Funktion ruft die mit dem Paket installierte Software „Stan“ (<http://mc-stan.org/>) auf und führt „Hamiltonian Monte Carlo“ Simulationen durch. Diese produzieren Ketten von simulierten Werten. Die Ketten wandern solange im Wertebereich herum, bis sie die A-posteriori-Verteilung gefunden haben und danach entsprechen die Werte der Ketten Zufallswerten aus der A-posteriori-Verteilung. So wird das Modell an die Daten angepasst und Zufallswerte aus der A-posteriori-Verteilung der Modellparameter gezogen. Letzteres ist das Ergebnis der bayesschen Datenanalyse. Hamiltonian Monte Carlo ist eine hochkomplexe Simulationstechnik, die sehr effizient Zufallswerte aus der A-posteriori-Verteilung zieht (Details siehe Betancourt 2013; Gelman et al. 2014). Um sicher zu sein, dass korrekte A-posteriori-Verteilungen gefunden wurden, sollte immer mehr als eine Simulationskette ausgehend von verschiedenen Startwerten durchgeführt werden. Die Funktion `stan_glm` simuliert automatisch vier solche Ketten. Der Vergleich der Lösungen der verschiedenen Ketten zeigt, ob die A-posteriori-Verteilungen gefunden wurden (ob die Ketten konvergiert sind).

Spezifizieren wir das obige Modell nun mit informativen A-priori-Verteilungen für die beiden Mittelwerte (über das Argument „prior“), um die Information aus dem Handbuch in die Analysen einfließen zu lassen. Wir erinnern uns, dass das obige Modell zwei Koeffizienten besaß: Den Mittelwert für die Männchen (Intercept) und die Differenz zwischen den Mittelwerten für Männchen und Weibchen. Die Literatur lieferte uns jedoch nicht die Differenz, sondern die beiden Mittelwerte selbst. Wir können das Modell leicht anders parameterisieren, so dass die beiden Koeffizienten den beiden Mittelwerten entsprechen. Das geschieht, indem wir den Intercept in der Modellformel mit der Notation „-1“ entfernen.

```
mod <- lm(fluegel~geschlecht-1, data=dat)
mod
```

Call:

```
lm(formula = fluegel ~ geschlecht - 1, data = dat)
```

Coefficients:

```
geschlechtM geschlechtW
      122.7       119.0
```

Wir können genau diese Notation auch in der Funktion `stan_glm` verwenden, die bayessche Alternative zur Funktion `glm`:

```
library(rstanarm)
stanmod <- stan_glm(fluegel~geschlecht-1, data=dat,
  prior = normal(c(mm, mw), scale=c(semm, semw)),
  prior_intercept = NULL, iter=5000)
```

Während Stan simuliert, sehen wir den rasend schnellen Fortschritt der Simulationen in der R-Console. Bevor wir das Resultat interpretieren, kontrollieren wir, ob die Simulationsketten konvergiert sind. Dazu können wir uns den Monte Carlo Standardfehler (mcse), den R-hat-Wert und die effektive Anzahl simulierter Werte (`n_eff`) anschauen.

```
summary(stanmod)
```

```
[...]
```

```
Diagnostics:
```

	mcse	Rhat	n_eff
geschlechtM	0.0	1.0	4390
geschlechtW	0.0	1.0	4094
sigma	0.0	1.0	3787

```
[...]
```

Die Modellresultate sind nicht exakt berechnet, sondern wurden mittels Simulationen angenähert. Wie nahe eine solche Annäherung an der exakten Lösung ist, sagt uns der Monte Carlo Standardfehler. Er sollte nahe bei Null sein, was in unserem Falle für alle Modellparameter der Fall ist. Wenn der Monte Carlo Standardfehler groß ist, kann er einfach durch Erhöhung der Anzahl Simulationen (im Argument „iter“ in der Funktion `stan_glm`) verringert werden. Der R-hat-Wert vergleicht die Varianz innerhalb einer Simulationskette mit der Varianz zwischen den Simulationsketten. Wenn alle Ketten in die gleiche Verteilung konvergiert sind, dann ist die Varianz zwischen den Ketten gleich der Varianz innerhalb der Ketten und der R-hat-Wert ist 1. Sind die R-hat-Werte größer als 1, müssen längere Ketten simuliert werden.

In der Spalte „n\_eff“ wird die Anzahl als voneinander unabhängig zu betrachtender simulierter Werte (effektive Anzahl simulierter Werte) angegeben. Sie entspricht der Anzahl simulierter Werte korrigiert dafür, dass in den Simulationsketten eine Simulation vom vorherigen Wert in der Kette abhängig ist (Autokorrelation). Deshalb werden sie auch „Markov-Ketten“ genannt. Je nach Stärke der Autokorrelation kann die effektive Anzahl simulierter Werte viel kleiner sein als die Kettenlänge. Die effektive Anzahl simulierter Werte zeigt an, auf wie vielen unabhängigen Werten die A-posteriori-Verteilung der Modellparameter beschrieben wird. Sie sollte groß genug sein (z. B. größer als 2000). Damit die effektive Anzahl simulierter Werte größer als 2.000 ist, haben wir im obigen Beispiel das Argument „iter“ vom Default-Wert (von 2000) auf 5.000 heraufgesetzt. Weitere Methoden, vor allem grafische Methoden, können als Ergänzung zur Überprüfung der Konvergenz verwendet werden, siehe z. B. Lunn et al. (2013).

In unserem Beispiel gibt es keine Hinweise, dass die Simulationsketten nicht konvergiert sind, und wir können die simulierten Werte zur Beschreibung der A-posteriori-Verteilungen verwenden. Im summary-Output (oben unterdrückt) finden wir den Mittelwert, die Standardabweichung (zu interpretieren als Standardfehler!), sowie einige Quantilen der A-posteriori-Verteilung wieder.

```
[...]
```

```
Estimates:
```

	mean	sd	2.5%	25%	50%	75%	97.5%
geschlechtM	123.0	1.1	120.8	122.3	123.0	123.6	125.1
geschlechtW	118.6	0.9	116.8	118.1	118.6	119.1	120.2
sigma	2.3	0.7	1.4	1.8	2.2	2.6	4.0

```
[...]
```

Wir schließen also, dass basierend auf unseren Daten und der im Handbuch zitierten Arbeit mit 95 % Sicherheit die mittlere Flügellänge bei Männchen zwischen 120,8 und 125,1 mm und bei Weibchen zwischen 116,8 und 120,2 mm liegt.

Aus den beiden Mittelwerten erhalten wir auch die Mittelwertsdifferenz (-4,4 mm) und die Wahrscheinlichkeit, dass die mittlere Flügellänge bei den Weibchen um mindestens einen Millimeter kürzer ist als bei den Männchen:

```
bsim <- as.data.frame(stanmod)
```

```
d <- bsim$geschlechtW - bsim$geschlechtM
```

```
mean(d)
```

```
[1] -4.429309
```

```
sum(d < -1)/nrow(bsim)
```

```
[1] 0.992
```

Unter den 10.000 Zufallswerten aus der A-posteriori-Verteilung für die Mittelwertsdifferenz sind 99,2 % kleiner als -1. Wenn wir unsere eigenen Daten mit den Angaben in der Literatur kombinieren, sind wir also zu 99,2 % sicher, dass die Weibchen im Durchschnitt einen um mindestens 1 mm kürzeren Flügel haben als die Männchen. Die Kombination von früherem Wissen und aktuellen Daten hat im Beispiel bewirkt, dass wir eine viel genauere Angabe über die geschlechtsspezifischen Mittelwerte der Flügellängen machen können und dass wir viel sicherer sind, dass tatsächlich ein stark ausgeprägter Geschlechtsdimorphismus in der Flügellänge vorhanden ist.

### Worin unterscheidet sich das Resultat einer bayesschen Analyse von jenem einer herkömmlichen statistischen Analyse?

Mittels herkömmlicher Methoden berechnen wir für männliche und weibliche Schneesperlinge mittlere Flügelängen z. B. jeweils mit dem 95 % Vertrauensintervall. Dieses sagt uns, in welchem Intervall wir 95 % der Mittelwerte von weiteren Zufallsstichproben gleicher Stichprobengröße wie unsere Daten erwarten. Für den Praktiker unterscheidet sich ein solches Resultat nicht wesentlich von Mittelwerten mit je einem Kreditabilitätsintervall. Letzteres sagt uns, in welchem Wertebereich mit einer Wahrscheinlichkeit von 95 % der wahre Mittelwert der Population liegt. Sowohl Vertrauensintervall wie auch Kreditabilitätsintervall beschreiben die stichprobenbedingte Unsicherheit. Solange wir bei der Schätzung von Parametern mit Unsicherheitsangaben bleiben und solange wir kein Vorwissen formell in die Analyse einbeziehen wollen, zeigt die Bayes-Statistik, abgesehen von der intuitiveren Interpretation des Kreditabilitätsintervalls gegenüber dem Vertrauensintervall, keine wirklichen Vorteile gegenüber der herkömmlichen Statistik. Im Gegenteil, wenn die Resultate alleine auf den aktuellen Daten basiert sein sollen, dann birgt die Bayes-Statistik die Gefahr, dass scheinbar nicht-informative A-priori-Verteilungen die Resultate beeinflussen können.

Die Vorteile der Bayes-Statistik zeigen sich vor allem dort, wo früheres Wissen formell in den aktuellen Analysen verwendet werden soll, und wo Hypothesen getestet werden sollen. Die meisten wissenschaftlichen Arbeiten beinhalten Hypothesentests.

Als Wissenschaftler formulieren wir eine (oder mehrere) Hypothesen, sammeln Daten, welche Information über die Hypothese enthalten, und möchten danach wissen, mit welcher Wahrscheinlichkeit die Hypothese zutrifft. Die Bayes-Statistik setzt genau dieses Vorgehen mathematisch um.

In der herkömmlichen Statistik ist eine solche direkte Vorgehensweise unmöglich. Stattdessen werden eine Nullhypothese und eine Alternativhypothese formuliert und berechnet, mit welcher Wahrscheinlichkeit die Daten (oder noch extremere Daten) beobachtet würden, falls die Nullhypothese wahr wäre. Diese Wahrscheinlichkeit wird p-Wert genannt. Die Nullhypothese wird verworfen und die Alternativhypothese angenommen, wenn der p-Wert klein genug ist, meist kleiner 5 %. Dieses Vorgehen erlaubt keine Aussage darüber, wie wahrscheinlich die Nullhypothese oder die Alternativhypothese ist, sondern liefert als Resultat lediglich eine ja/nein-Antwort: Der Geschlechtsunterschied in der Flügelänge ist „signifikant“ oder er ist „nicht signifikant“, eine Information, die ohne Angaben der Effektgröße mit Unsicherheitsintervall, so gut wie nichtssagend ist (Abb. 5).

Zum Beispiel wird ein Nullhypothesentest in einem klassischen zweiseitigen t-Test (mit Nullhypothese: „Effekt = 0“) signifikant, wenn der Effekt ziemlich sicher

biologisch relevant ist (Fall A in Abb. 5) aber auch, wenn er ziemlich sicher biologisch unbedeutend ist (Fall E in Abb. 5). Umgekehrt ist ein nicht signifikantes Resultat keinen Beweis dafür, dass der Effekt biologisch nicht relevant ist (Fall D), sondern es könnte genauso gut bedeuten, dass der Effekt sehr wohl biologisch relevant sein könnte, die Daten aber einfach nur schwache Information enthalten (Fall B). Die Unterscheidung der Fälle C (keine Information in den Daten) und D (eindeutiger Hinweis auf einen Effekt, der nahe Null, d.h. unbedeutend, ist) macht der herkömmliche Nullhypothesentest nicht. Das Problem des Nullhypothesentests ist nicht, dass er falsch wäre, sondern dass die meisten Anwender meinen, er unterscheide zwischen Fall A und D und vergessen, dass die Fälle B, C und E auch existieren können.

Eine Alternative zum oben beschriebenen klassischen t-Test wäre ein einseitiger t-Test mit Nullhypothese „Differenz = 1“. Dieser Test vereinfacht die Interpretation signifikanter Resultate, da er nur im Fall A signifikant wird, er erschwert jedoch die Interpretation von nicht signifikanten Resultaten, da diese alle anderen vier Fälle beinhalten. Das Resultat der bayesschen Analyse besagt direkt, wie sicher wir sind, dass die Männchen im Mittel um mindestens 1 mm längere Flügel haben.

Das Resultat einer bayesschen Datenanalyse ist keine ja/nein-Antwort mit wenig biologischer Relevanz („signifikant“ oder „nicht signifikant“), sondern es ergibt vielmehr Wahrscheinlichkeiten für biologisch meist sinnvolle Hypothesen. Diese Wahrscheinlichkeiten können je nach Kontext weiter interpretiert und für Entscheidungen verwendet werden.

Ob ein Effekt „signifikant/relevant“ ist oder nicht, sollte nicht alleine von den Daten abhängen, sondern auch davon, welche Entscheidungen basierend auf den gemessenen Effekten getätigt werden und welche Konsequenzen diese haben. Wenn z. B. eine Schutzmaßnahme für eine nicht sonderlich bedrohte Art sehr aufwändig und teuer ist, möchten wir sie erst anwenden, wenn wir wirklich ganz sicher sind, dass die Population ohne diese Schutzmaßnahme abnimmt. Wenn es sich hingegen um eine weltweit bedrohte Art handelt, möchten wir die Schutzmaßnahme einsetzen, sobald wir schwache Hinweise auf einen Rückgang haben. Eigentlich ist es einleuchtend, dass wir das Signifikanzniveau für die beiden Fälle ganz unterschiedlich ansetzen sollten (wie das übrigens auch schon Fisher 1956 postulierte). Da der p-Wert jedoch nicht einfach zu interpretieren ist (z. B. Burnham & Anderson 2014), ist es extrem schwierig, in spezifischen Fällen das Signifikanzniveau selbst vorzugeben. Häufig wird deshalb einfach das verwendet, was andere schon vorher verwendet hatten, nämlich 5 %, ohne sich zu überlegen, ob dieses Signifikanzniveau für den spezifischen Fall sinnvoll ist.

Das Resultat aus einer bayesschen Analyse ist dagegen eine intuitive und relativ einfach zu interpretie-

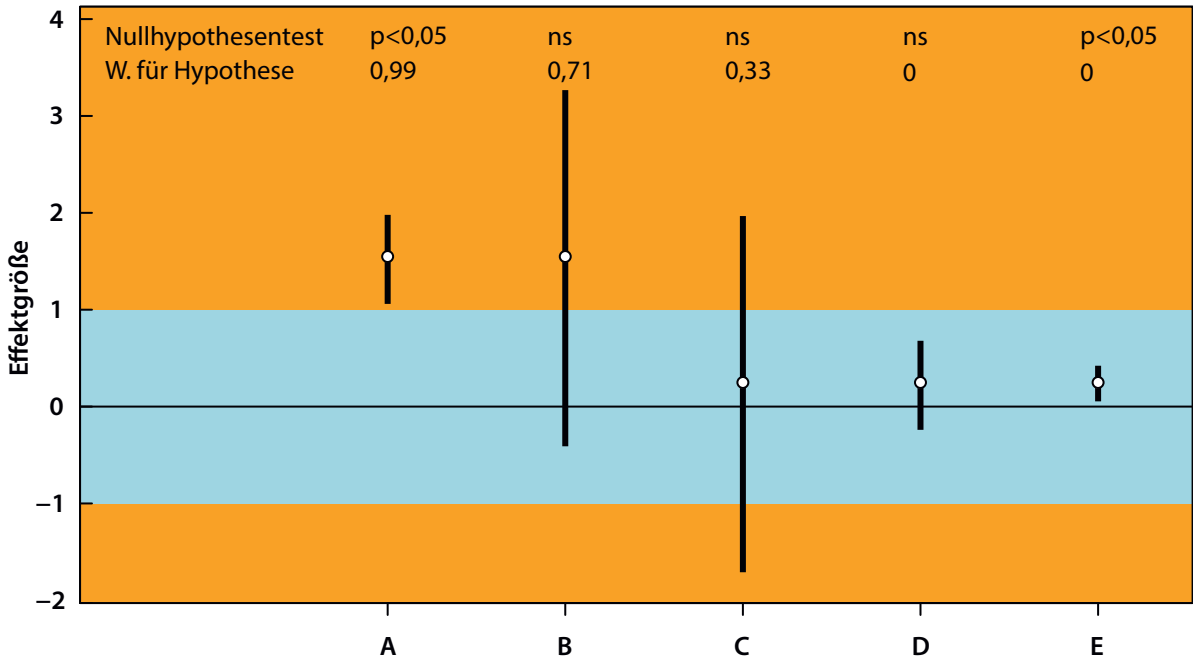


Abb. 5: Fünf mögliche Resultate der Schätzung eines Effektes, z. B. des Unterschieds in der mittleren Flügellänge zwischen den Geschlechtern. Die Punkte geben die geschätzten Differenzen, die vertikalen Balken die 95 % Unsicherheitsintervalle (Vertrauensintervall bzw. Kreditintervall) an. Die Resultate des zugehörigen Nullhypotesentest (klassischer zweiseitiger t-Test mit Nullhypothese: „Effekt = 0“; „ $p < 0,05$ “ = signifikant, „ns“ = nicht signifikant) sind in der ersten Zeile aufgeführt. In der zweiten Zeile finden sich die (gerundeten) posterior Wahrscheinlichkeit für die Hypothese, dass der Effekt „biologisch relevant“ ist. Die Hintergrundfarbe unterscheidet schematisch „biologisch relevante“ (orange) von „biologisch nicht relevanten“ (hellblau) Effektgrößen. Der Übergang von biologisch nicht relevanten zu relevanten Effekten muss durch den Biologen festgelegt werden. In den meisten Fällen ist der Übergang nicht scharf, sondern fließend. Hier wurde der Einfachheit zuliebe eine scharfe Grenze bei 1 angenommen. – *Five different possible results of an estimated effect such as a difference. Circles are the point estimates, vertical bars are 95 % uncertainty intervals such as confidence or credible intervals. Results from the corresponding null hypothesis tests are given in the first row („ $p < 0,05$ “ = significant, „ns“ = non-significant). Posterior probabilities of the hypothesis that the difference is biologically relevant (orange shading) are given on the second row. A sharp transition between biologically negligible (light blue) and relevant (orange) effects is used here for simplicity.*

rende Wahrscheinlichkeit für eine sinnvolle Hypothese, z. B. die Wahrscheinlichkeit für die Hypothese, dass eine Population abnimmt. Daraus lassen sich spezifisch sinnvolle Signifikanzniveaus einfacher ableiten, weil wir oft eine Vorstellung davon haben, wie sicher wir sein sollten, dass z. B. eine Population abnimmt, um die Abnahme als relevant (oder eben „signifikant“) einzustufen.

Ob wir A-priori-Information in einer Datenanalyse verwenden möchten oder nicht, hängt vom Zweck einer Studie ab. Wenn diese Basis für wichtige Entscheidungen in der Praxis liefern soll, ist es meist sehr wertvoll, wenn wir neben der Information in den aktuellen Daten auch Information aus der Literatur berücksichtigen können. In der Grundlagenforschung hingegen ist es manchmal wünschenswert, dass das Resultat hauptsächlich durch die aktuellen Daten bestimmt wird, um jegliche Subjektivität durch die Wahl der A-priori-Verteilungen auszuschließen. In solchen Fällen ist es

üblich, die Sensitivität der Resultate gegenüber der A-priori-Verteilungen zu messen.

#### Was sollte in einer Publikation angegeben werden?

Die Bayes-Statistik erfordert meist etwas umfangreichere Angaben als die herkömmliche Statistik. In den Methodenteil gehören folgende Beschreibungen der Datenauswertung:

1. Datenmodell (z. B. „Normalverteilung für die Flügellängen der Schneesperlinge mit unterschiedlichen Mittelwerten für die Geschlechter, jedoch gleicher Varianz“).
2. Beschreibung der A-priori-Verteilungen der Modellparameter: Welche Verteilungen werden verwendet? Woher stammt die Information? Wie stark beeinflussen die A-priori-Verteilungen das Resultat?
3. Erwähnung der Technik, mit der die A-posteriori-Verteilung berechnet wurde (z. B. Software, Simulationsalgorithmen) sowie, wenn nötig, Angaben über

die Überprüfung des Erfolgs und der Genauigkeit dieser Techniken (z. B. Überprüfung von Konvergenzen bei Simulationsketten, Zahl der verwendeten Ziehungen).

#### 4. Art der Überprüfung der Modellgüte sowie eine Beurteilung der Modellgüte.

Der Punkt 1 und die Prinzipien von Punkt 4 unterscheiden sich nicht von der herkömmlichen Statistik. Die Punkte 2 und 3 sind spezifisch für bayessche Datenanalysen.

Im Resultateteil einer Publikation präsentieren wir Wahrscheinlichkeiten für spezifische Hypothesen. So können wir im ersten Beispiel angeben, dass mit 95 % Wahrscheinlichkeit der Männchenanteil der Amsel zwischen 0,39 und 0,74 liegt. Im zweiten Beispiel können wir schreiben, wie „sicher“ es ist, dass der Unterschied in der Flügellänge zwischen den Geschlechtern einen bestimmten biologisch relevanten Betrag überschreitet. Zudem geben wir den geschätzten Unterschied und/oder die beiden Mittelwerte inklusive der 95 % Kreditabilitätsintervalle an.

#### Weiterführende Literatur

Falls Sie nun Lust auf mehr Bayes-Statistik erhalten haben, dann empfehlen wir Ihnen eines der folgenden Bücher. Als Einstieg in die bayessche Datenanalyse eignet sich das Lehrbuch von Kruschke (2015). Es beschreibt ausführlich und anschaulich die Philosophie der Bayes-Statistik und führt durch alle wichtigen Modelle mit R-, BUGS- und Stan-Code hindurch. BUGS ist eine weitere Software zur Durchführung von Markov chain Monte Carlo Simulationen ([www.mrcbsu.cam.ac.uk/software/bugs](http://www.mrcbsu.cam.ac.uk/software/bugs)). Eine ähnliche, jedoch kürzere Einführung bietet das Buch von Korner-Nievergelt et al. (2015), welches sich im ersten Teil auf Beispiele mit nicht-informativen A-priori-Verteilungen beschränkt, die ganz in R durchgeführt werden können. Erst im zweiten Teil werden Markov chain Monte Carlo Simulationen eingeführt und einige Beispiele für etwas komplexere ökologische Modelle durchgeführt mit BUGS oder Stan (<http://mc-stan.org>). Für etwas fortgeschrittenere Leser empfehlen wir gerne unser Lieblingsbuch von Gelman & Hill (2007), welches als sehr kreatives Lehrbuch für lineare Modelle verwendet werden kann. Es beschreibt mit sehr präzisen knappen Worten das Essenzielle der verschiedenen Modelle, die jeweils für spezifische praktische Beispiele zugeschnitten sind. Die hier vorgestellte Funktion `sim` stammt aus dem Paket „arm“, das als Begleitung zu diesem Buch geschrieben wurde.

Eine sehr gut verständliche Einführung in WinBUGS gibt Kéry (2010), ein Lehrbuch, das von einem Biologen für Biologen geschrieben wurde. Etwas technischer, aber auch gut verständlich und vor allem sehr kurz und bündig ist McCarthy (2007). Für die mathematischen Hintergründe der bayesschen Statistik in der angewandten

Statistik ist nach wie vor Gelman et al. (2014) der Klassiker. Allerdings ist dieses Buch wegen der vielen Formeln eher für Leute mit viel Freude an Mathematik. Wer nicht so viel Freude an Mathematik aufbringt, sich aber trotzdem die Prinzipien und Theorien der Bayes-Statistik aneignen möchte, dem können wir Hobbs & Hooten (2015) empfehlen.

#### Zum Schluss

Die Entwicklung benutzerfreundlicher Software hat Bayes-Statistik auch Biologen zugänglich gemacht. Trotzdem wird sie noch längst nicht standardmäßig für Datenauswertungen verwendet, vermutlich weil Universitäten nach wie vor klassische Nullhypotheseentests lehren, weil der Mensch ein „Gewohnheitstier“ ist und weil viele Anwender Angst haben, die statistischen Methoden würden zu kompliziert, wenn bayessche Methoden verwendet werden. Mit dieser Kurzeinführung möchten wir zeigen, dass Bayes-Statistik nicht schwieriger ist als frequentistische Statistik. Im Gegenteil, die Interpretation von Resultaten aus der Bayes-Statistik ist viel einfacher als die von Resultaten aus frequentistischen Analysen, weil interpretierbare Wahrscheinlichkeiten für konkrete, biologisch sinnvolle Hypothesen berechnet werden. Wir wünschen Ihnen einen guten Start beim Erlernen der bayesschen Statistik!

#### Dank

Wolfgang Fiedler und Holger Schielzeth danken wir herzlich für ihre wertvollen Kommentare und Anregungen zu einer früheren Version.

#### Zusammenfassung

In dieser kurzen praktischen Einführung führen wir den Leser anhand von zwei Beispielen aus der Ornithologie in die Grundzüge der bayesschen Datenanalyse ein. Wir erläutern einige Prinzipien der bayesschen Datenanalyse und erklären Unterschiede und Parallelen zur frequentistischen („klassischen“) Statistik. In der bayesschen Datenanalyse wird früheres Wissen (a-priori) mit der Information in den Daten mittels des Satzes von Bayes kombiniert. Als Resultat erhält man die A-posteriori-Verteilung. Das Wissen über einen Parameter (vor und nach der Datenaufnahme) wird mathematisch mittels Wahrscheinlichkeitsverteilungen beschrieben. Im ersten Beispiel schätzen wir den Männchenanteil einer Population. Im zweiten Beispiel berechnen wir die Differenz in der mittleren Flügellänge zwischen den Geschlechtern einer Singvogelart und zwar einmal ohne A-priori-Wissen und einmal mit A-priori-Information aus der Literatur. In beiden Fällen berechnen wir die Wahrscheinlichkeit für die simple Hypothese, dass die mittlere Flügellänge der Männchen um mindestens 1 mm länger ist als jene der Weibchen. Für die praktischen Beispiele verwenden wir die R-Pakete „arm“ und „rstanarm“. Wir erklären, weshalb das Resultat einer bayesschen Datenanalyse einfacher zu interpretieren ist als jenes eines Nullhypotheseentests. Zuletzt empfehlen wir einige aus

unserer Sicht besonders hilfreiche Lehrbücher zur Bayes-Statistik. Der Artikel soll Ornithologen den Zugang zur Bayes-Statistik erleichtern und als Einstiegshilfe für eigene bayessche Datenanalysen mit R dienen.

---

## Literatur

- Aitkin M, Francis B, Hinde J & Darnell R 2009: Statistical Modelling in R. Oxford University Press, Oxford.
- Amrhein V, Scaar B, Baumann M, Minéry N, Binnert J-P & Korner-Nievergelt F 2012: Estimating adult sex ratios from bird mist netting data. *Meth. Ecol. Evol.* 3: 713-720.
- Bayes RT 1763: An essay towards solving a problem in the doctrine of chances. *Phil. Trans. R. Soc. London* 53: 370-418.
- Betancourt M 2013: Generalizing the no-u-turn sampler to Riemannian manifolds. arXiv 1304.
- Burnham KP & Anderson, DR 2014: P values are only an index to evidence: 20th- vs. 21st-century statistical science. *Ecology* 95: 627-630.
- Fisher RA 1956: Statistical methods and scientific inference. Oliver and Boyd, University of California.
- Gelfand AE & Smith AFM 1990: Sampling-based approach to calculating marginal densities. *J. Am. Stat. Assoc.* 85: 398-409.
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A & Rubin DB 2014: Bayesian Data Analysis. 3. Aufl., CRC Press, New York.
- Gelman A & Hill J 2007: Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge.
- Gilks WR, Richardson S & Spiegelhalter DJ 1996: Markov Chain Monte Carlo in practice. Chapman& Hall/CRC, Boca Raton.
- Glutz von Blotzheim U & Bauer K 1997. Handbuch der Vögel Mitteleuropas. Band 14 Passeriformes (5. Teil). Aula-Verlag, Wiesbaden.
- Hobbs NT & Hooten MB 2015: Bayesian Models - A Statistical Primer for Ecologists. Princeton University Press, New Jersey.
- Kéry M 2010: Introduction to WinBUGS for ecologists. Academic Press, London.
- Korner-Nievergelt F & Hüppop O 2010: Das freie Statistikpaket „R“: Eine Einführung für Ornithologen. *Vogelwarte* 48: 119-135.
- Korner-Nievergelt F, Roth T, von Felten S, Guélat J, Almasi B & Korner-Nievergelt P 2015: Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan. Elsevier, New York.
- Kruschke JK 2015: Doing Bayesian Data Analysis, A Tutorial with R, JAGS, and Stan. Academic Press, Burlington, USA.
- Lang EM 1946: Über die Brutgewohnheiten des Schneefinken *Montifringilla n. nivalis* (L.). *Ornithol. Beob.* 43: 33-43.
- Liker A, Freckleton RP & Székely T 2013: The evolution of sex roles in birds is related to adult sex ratio. *Nature Comm.* 4: 1587.
- Lunn D, Jackson C, Best N, Thomas A & Spiegelhalter D 2013: The BUGS Book - A practical introduction to Bayesian Analysis. Taylor & Francis Group, Boca Raton.
- McCarthy MA 2007: Bayesian Methods for Ecology. Cambridge University Press, Cambridge.
- R Development Core Team 2016: R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Smith AFM, Skene AM, Shaw JEH, Naylor JC & Dransfield M 1985: Progress with numerical and graphical methods for Bayesian statistics. *Commun. Stat. A - Theor. Meth.* 14: 1079-1102.
- Stahel WA 2002: Statistische Datenanalyse: Eine Einführung für Naturwissenschaftler. 4. Aufl., Vieweg, Wiesbaden.

**Elektronischer Anhang:** Der R-Code zu diesem Artikel ist auf den Internetseiten der „Vogelwarte“ unter folgender Adresse verfügbar: [www.do-g.de/publikationen/vogelwarte/inhalte-online](http://www.do-g.de/publikationen/vogelwarte/inhalte-online). Klicken Sie auf dort bitte auf den Link für Heft 3/2016.

# ZOBODAT - [www.zobodat.at](http://www.zobodat.at)

Zoologisch-Botanische Datenbank/Zoological-Botanical Database

Digitale Literatur/Digital Literature

Zeitschrift/Journal: [Vogelwarte - Zeitschrift für Vogelkunde](#)

Jahr/Year: 2016

Band/Volume: [54\\_2016](#)

Autor(en)/Author(s): Korner-Nievergelt Fränzi, Hüppop Ommo

Artikel/Article: [Kurze Einführung in Bayes-Statistik mit R für Ornithologen 181-194](#)