Abstract

The interest in Bayesian statistic has been increasing for the last two decades both among statisticians and researchers who cannot properly analyze their data using methods based on classical statistics. This course will provide a sound basis in Bayesian statistics for those who

• Want to understand what Bayesian statistics is about
• Want to use Bayesian statistics to build and evaluate statistical models
• Want to get hands on experience with Bayesian statistics in R and JAGS

Target Group

The course is aimed at researchers who not only work with statistical tools, but are also interested in the development and evaluation of statistical tools. Among these are psychometricians, sociometricians, epidemiologists, and statisticians. The only requirement is familiarity with the following concepts: the likelihood function, the p-value, analysis of variance, and multiple regression.

Course Data, Course Fee and Course Materials

The course will be given April 20-24, 2020, in Utrecht at Utrecht Science Park (Uithof). Each day will start at 9.00 and finish at 17.00.

Course materials will be made available one week in advance. The reading materials listed at the end of this document are freely available via the library of most universities and can be considered further reading.

Participants are required to bring their own laptop on which R (https://cran.r-project.org/) and JAGS (http://mcmc-jags.sourceforge.net/) should be installed. Participants should also install the R package bain.
Grading

Students who want to obtain a grade and credit points (Epidemiology Master Students 1.5 EC, IOPS PhD students 2 EC) for this course have to participate in the presentations on Friday afternoon.

Day 1: Introduction to the Key Concepts and Formulas of Bayesian Statistics: Introduction to R and JAGS.

Using simple models for binomial data a technical introduction of the key concepts in Bayesian statistics will be given. This day consists of three lectures and a lab meeting in which the students get acquainted with R and JAGS and practice what they have learned during the lectures.

Lecture 1: Using an example in which the data have a binomial distribution, key concepts in Bayesian statistics will be introduced: prior distribution, density of the data, posterior distribution, expected a posteriori estimates, credibility intervals and sampling from the posterior distribution.

Lecture 2: The posterior predictive distribution will be discussed. This distribution lies at the heart of the Bayesian equivalent of classical hypothesis testing: posterior predictive p-values. The pros and cons of posterior predictive compared to classical p-values will be discussed.

Lab Meeting: The students will learn to work with R and JAGS. This will be done by means of an application of the key concepts and formulas of Bayesian statistics on data that have a binomial distribution.

Day 2: Bayesian Estimation: the Gibbs sampler and the Metropolis-Hastings Algorithm. Bayesian Model Selection: the DIC.

On Day 1 it was shown that it is easy to sample from posterior distributions that are based on simple binomial densities for the data. However, for more complicated models/densities direct sampling is often not possible. In two lectures and one lab meeting it will be explained how this problem can be solved. Furthermore, a models selection criterion, the DIC, will be introduced.

Lecture 1: The Gibbs sampler: sampling from a multivariate posterior distribution if the conditional densities are known. The Metropolis Hastings algorithm: sampling from a multivariate posterior distribution if the conditional densities are unknown. Convergence issues when Markov chain Monte Carlo methods are used to sample from a posterior distribution.

Lecture 2: In this lecture the ideas underlying the deviance information criterion DIC will be discussed. Illustrations of the DIC will be provided.

Lab Meeting: R and JAGS will be used to construct a Gibbs sampler for the posterior distribution of a logistic regression model. The students will practice using the DIC in the context of the selection of predictors for the logistic regression model. Furthermore, for a person with known predictors but unknown outcome, a prediction of and credibility interval for the unknown outcome will be obtained using R and JAGS.

Day 3: Bayesian Model Selection: the Bayes Factor.

Today the Bayesian alternative for hypothesis testing will be discussed: Bayesian model selection. There will be lectures and a lab meeting.

Lecture 1: Informative hypotheses and the Bayes factor will be introduced. All from an applied perspective.

Lecture 2: Bayesian error probabilities, Bayesian updating, and the evaluation of replication studies using informative hypotheses will be discussed. All from an applied perspective.

Lecture 3: A technical elaboration of the Bayes factor in the context of a multiple regression model.
Lab Meeting: Using bain the participants will formulate and evaluate informative hypotheses in the context of an ANOVA model.

**Day 4: Prior Distributions**

A unique element of Bayesian statistics is that it makes use of prior distributions. Today’s material will focus on prior information and constructing prior distributions.

Lecture 1: Different types of prior distributions (informative, uninformative, skeptical, etc).

Lecture 2: Constructing priors based on expert opinions and empirical evidence.

Lecture 3: Prior Distributions: Risks and Good Practices

Lab meeting: Discussion on using informative prior distributions & Exercises with prior information

**Day 5: Discussion Meeting and Presentations**

Morning: Students will be divided into Frequentist and Bayesian groups and have to promote their respective convictions in a structured debate. Students may find inspiration for this debate in the assigned reading for the debate and in the preparatory assignments that will be executed before the debate.

Afternoon: Students who wish to be graded for their participation have to give a 5 minute presentation in which they use or elaborate on something they learned in this course in the context of their own work, research, or any other topic they find interesting.
**Reading**

**Recommended Reading for Day 1**


**Recommended Reading for Day 2**


*An application of Bayesian Statistics*


**Recommended Reading for Day 3**


*An application of informative hypotheses*


**Recommended Reading for Day 4**


**Recommended Reading for the debate on Day 5**


