Bayesian Data Analysis

Seminar Session 1

Matthew A. Fisher
School of Mathematics, Statistics and Physics
Newcastle University

February 3, 2022

Preliminary

Welcome to the first seminar of MAS8405 - Bayesian Data Analysis!

- The Thursday Seminars (14:30-16:30) will be lead by me, Matthew Fisher.
- My seminars will be focussed on the practical application of the course material.
- All code will be written in `R`.
- All the Thursday seminar materials can be found on GitHub. I will "push" the weeks material just before the session starts.
- If there is any particular aspect of the course you would like me to address in upcoming seminars, please let me know (time permitting).

Overview

1. A Quick Refresher
2. Problem 1: Conjugate Inference
3. Problem 2: Non-Conjugate Inference
4. Problem 3: JAGS Introduction

A Quick Refresher

Notation

- Data or observations will be represented as lower-case letters (e.g. y_i) and a collection of n data will be denoted as a vector (or matrix) as $\underline{y} = (y_1, \ldots, y_n)$.
- Data y_i are assumed to be an observation from a corresponding random-variable Y_i. Random variables will be denote by a capital letters (e.g. Y).
- We write $Y \sim \mathbb{P}$ to mean that the random variable Y follows the distribution $\mathbb{P}$.
- Probability distributions come equipped with parameters. Individual parameters will usually be denoted by lower-case letters or Greek letters. We will group all parameters together as a vector $\theta = (\theta_1, \ldots, \theta_m)$, where each $\theta_1, \ldots, \theta_m$ is a parameter. For example, the mean μ and variance σ^2 parameterise a normal distribution $\mathcal{N}(\mu, \sigma^2)$ and so we write $\theta = (\mu, \sigma^2)$.
- We define the set of “allowed” parameters (termed the *parameter space*) as Θ and we specify that $\theta \in \Theta$.

Statistical Inference

Statistical inference is a formalisation of *induction*. Given some observations, we seek to understand and describe the underlying random-process that generated our observations. In mathematics (and assuming I.I.D. data):

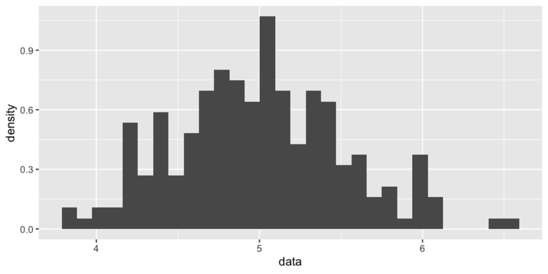
1. We have data $\underline{y} = (y_1, \ldots, y_n)$.
2. We assume each y_i is a realisation (or *observation*) from a random variable Y_i, which follows a distribution $Y_i \sim \mathbb{P}_\theta$, for some parameter $\theta \in \Theta$.
3. Using our given data $\underline{y}$, we seek the particular θ that generated $\underline{y}$ or some subset $\Theta_0 \subset \Theta$ that could likely have generated $\underline{y}$.

There are many ways of doing this...

Statistical Inference: Example

Example: Statistical Inference by Eye

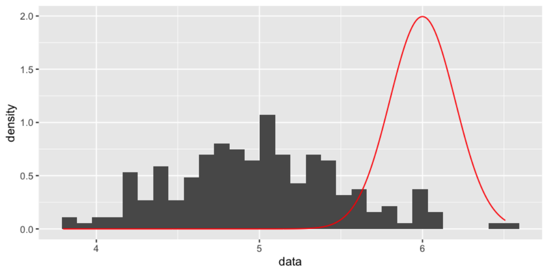
Suppose we have the following data $\underline{y} = (4.5, 4.8, 4.6, \ldots, 5.2)$, which we can visualise as a histogram:



Statistical Inference: Example

Example: Statistical Inference by Eye

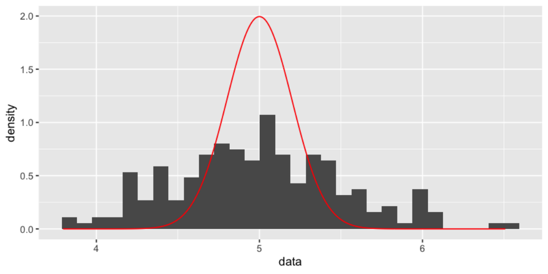
Suppose we have the following data $\underline{y} = (4.5, 4.8, 4.6, \ldots, 5.2)$, which we can visualise as a histogram:



Statistical Inference: Example

Example: Statistical Inference by Eye

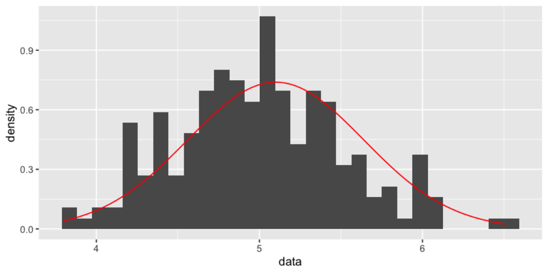
Suppose we have the following data $\underline{y} = (4.5, 4.8, 4.6, \ldots, 5.2)$, which we can visualise as a histogram:



Statistical Inference: Example

Example: Statistical Inference by Eye

Suppose we have the following data $\underline{y} = (4.5, 4.8, 4.6, \ldots, 5.2)$, which we can visualise with a histogram:



Bayesian Inference

Bayesian inference is a particular approach of statistical inference that is the focus of this course. Bayesian inference assumes the parameter θ is a random quantity to be inferred and relies on the Bayesian update.

The Bayesian Update

We place a prior distribution $p(\theta)$ on the parameter θ and, after observing data $\underline{y}$, we update our knowledge about θ based on the data $\underline{y}$ using Bayes rule:

$$p(\theta \mid \underline{y}) = \frac{p(\underline{y}|\theta)p(\theta)}{\int p(\underline{y}|\theta)p(\theta)\,\mathrm{d}\theta}.$$

This requires the specification of a probability model $p(\underline{y}|\theta)$ for the data $\underline{y}$. This is termed the likelihood.

Distribution Zoo

The notation $\mathbb{P}_\theta$ has been used as a placeholder for a general probability distribution with parameter $\theta \in \Theta$. Widely used probability distributions have special notation:

1. Normal distribution: continuous, $Y \sim \mathcal{N}(\mu, \sigma^2)$.
2. Uniform distribution: continuous, $Y \sim \mathcal{U}(a, b)$ or $Y \sim \text{Unif}(a, b)$.
3. Gamma distribution: continuous, $Y \sim \Gamma(\alpha, \beta)$ or $Y \sim \text{Gamma}(\alpha, \beta)$.
4. Beta distribution: continuous, $Y \sim \mathcal{B}(\alpha, \beta)$ or $Y \sim \text{Beta}(\alpha, \beta)$.
5. Binomial distribution: discrete, $Y \sim \text{Bin}(n, p)$.

Note on Parameterisation:

A distribution can be reparamaterised. For instance, you may be familiar with using the precision $\tau = 1/\sigma^2$ to parameterise the normal distribution instead of the variance.

Visit the Distribution Zoo to see a wide array of different distributions and their uses.

Problem 1: Conjugate Inference

Conjugate Inference

Conjugate inference is useful since Bayesian updating can be computed in closed-form. For this first problem, we consider implementing a conjugate inference problem from "scratch" using R. What is conjugate inference?

Bayesian Conjugate Inference

A likelihood $p(\underline{y}|\theta)$ and prior $p(\theta)$ are said to form a conjugate pair whenever the posterior $p(\theta|\underline{y})$ belongs to the same "distributional family" as the prior.

How do we know when prior is conjugate to our likelihood?

1. Check by hand using Bayes theorem.
2. Look up standard conjugate distributions. Good sources include Wikipedia, the Distribution Zoo or a book.

Problem 2: Non-Conjugate Inference

Non-Conjugate Inference

For our second "problem", we implement non-conjugate inference from "scratch" using `R`. For non-conjugate inference we have to resort to numerical methods to investigate our posterior due to the pesky normalisation constant:

$$p(\theta \mid \underline{y}) = \frac{p(\underline{y}|\theta)p(\theta)}{\int p(\underline{y}|\theta)p(\theta)\,\mathrm{d}\theta}.$$

In this course we will resort to Markov Chain Monte Carlo (MCMC) and use the MCMC software JAGS.

I thought it would be useful to implement a couple of alternative numerical methods.

Problem 3: JAGS Introduction

The End