CSSS/STAT 564: Assignment 2

Sheridan Grant

April 17, 2017

0.1 Instructions

1. Fork this repository to your account
2. Edit the file `solutions.Rmd` with your solutions to the problems.
3. Submit a pull request to have it graded. Include either or both a HTML and PDF file.

For updates and questions follow the Slack channel: #assignment2.

This assignment will require the following R packages:

```r
library("rstan")
```

0.2 How much does the prior influence the posterior?

In this problem, we aim for a greater understanding of the influence of a normal prior on inference about the mean of a normal distribution.

Suppose you observe n single data points which you model as coming from a normal distribution. Consider the case in which the mean of the distribution is μ (unknown), but the variance, σ^2, is known:

$$x|\mu \sim N(\mu, \sigma^2)$$

You provide a prior distribution for μ,

$$\mu \sim N(\mu_0, \sigma_0^2)$$

A normal prior for μ is conjugate to a normal likelihood, meaning that the posterior distribution is also normal:

$$\mu|x \sim N(\frac{\frac{1}{\sigma_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}}, \frac{1}{\sigma_0^2} + \frac{n}{\sigma^2})$$

Q1 In your own words, what is a conjugate prior? What are the advantages and disadvantages of such a prior? Suppose that instead of a normal prior for μ, a non-conjugate distribution such as the Laplace distribution were used. Write down the posterior density function in this case. Why might a normal prior be preferred?

Q2 Describe the effects of the following on inference about μ:

- Lots of data vs. not much data
- Strong prior information vs. weak prior information
- Lots vs. little noise in the data
- Data that is incompatible with the prior vs. data that is compatible with the prior

A normal prior with finite variance necessarily contributes information to the model; the question is how much. We now investigate a way of understanding the information in a normal prior in terms of implicit additional observed data.

Q3 The data in `data/pr2data.csv` were generated from a normal distribution with a certain mean and variance. First, fit a linear model with just an intercept and report the estimated mean and variance, as well as a centered 97% confidence interval for μ. Interpret the confidence interval appropriately.

Q4 Now you're told that the variance of the normal distribution that generated these data is 100 (this should be no surprise if you've done the previous question). Fit a Bayesian model for the mean assuming a normal likelihood with standard deviation 10, and with prior $N(10, 0.1^2)$. (In the usual notation for the normal model, the second parameter is the variance; but in *Stan*, the second parameter is the standard deviation, because that is unfortunately the convention in *R*. So here, the standard deviation is 0.1 and the variance is 0.01.) Would you characterize this prior as "highly informative," "modestly informative," or "uninformative," and why? Provide the same output (estimate, variance/standard deviation of estimate, confidence interval) you did in **Q3**, and interpret the 97% confidence interval appropriately (hint: use the *print* function for Stan models).

Q5 Now, generate 10^4 independent observations from a $N(10, 10^2)$ distribution, center them to have a mean of 10 (this is cheating a bit, we fully admit), concatenate them to the data from `data/pr2data.csv`, and fit both a standard linear model with just a mean intercept and a Bayesian model with normal likelihood and an improper flat prior. Provide the same outputs as before (no need to interpret the confidence intervals again, though).

Q6 What do you think is going on here? How can we interpret the prior in terms of a number of additional data points drawn from a distribution? Relate this distribution back to the formula for the posterior distribution $\mu|x$, given before **Q1**.

Q7 What simplifying assumptions were made in this problem for illustrative purposes, and what do you think might happen to inference about μ (the estimation of the mean and the standard error of that mean) if those assumptions were relaxed?