In-class exercise: Logistic regression

Names: (signatures only please, printed names will not be counted)

1.) 4.)

2.) 5.)

3.) 6.)

Overview

In this exercise we simultaneously model the probability of measurable rainfall on a given day and the amount of rainfall on days with measurable rain in Amherst, MA from 1900 through 1992 in `taunton_rain.csv`

- The probability of measurable rain on a given day is modeled as a Bernoulli distribution.
- If there is measurable rainfall, the amount is modeled by a gamma distribution.
- If no rainfall is measured, with probability 1 the rainfall is zero.

The function of the model block of the .stan file is to total the log of the likelihood function over the observed values of y.

On days with no measurable rainfall, the likelihood is just $1 - \theta$, where θ is the probability of rain on a given day. The log likelihood is then:

$$\ln(L(y)) = \ln(1 - \theta)$$

On days with measurable rainfall, the likelihood is the product of the probability of measurable rainfall, θ, and the gamma density function with parameters α and β:

$$L(y) = \theta \cdot f_\gamma(\alpha, \beta)$$

In this case the log likelihood is

$$\ln(L(y)) = \ln\theta + \ln(f_\gamma(\alpha, \beta))$$

Look at the code in the .stan file model block, which performs these computations.

Stan is very powerful because it allows you to implement any probability distribution that you can code the likelihood for.

Instructions

As usual, start by bringing your copy of the `MTH225_Spring2017` archive up to date.

Open a command prompt or terminal window, and use the `cd` command to change to the `MTH225_Fall2016` subdirectory. Then type the command:

```
git pull origin master
```

The pull operation should download the following files:

- The R-knitr code: `MTH225-11_logistic_mis.Rnw`
- The data in Rdata format: `logistic1.csv`
- The STAN model file: `logistic1.stan`

Questions

Use the *Compile PDF* button to run the model, and use the output to answer the following questions. Because some numbers are rather small, you should extract the posterior draw from the stanfit object and compute means and quantiles from it.

1) What is the point estimate and 95% confidence interval for the probability of rainfall θ?

2) What is the point estimate and 95% confidence interval for the paraeter α?

3) What is the point estimate and 95% confidence interval for the parameter β?