Penn State STAT 540
Homework #3, due Monday, Nov 2, 2015

Submit: (i) `R` code in a file titled PSUemailidHW3.R (e.g. muh10HW3.R), (ii) pdf file that contains a clear writeup for the questions below named PSUemailidHW3.pdf Note that your code should be readily usable. It should also be well commented.

(1) Use maximum likelihood to fit the GEV distribution to the following data set on storm surge heights from tide gauges `http://www.stat.psu.edu/~mharan/540/hwdir/AtlanticGEV.dat` Assume that the data are independent draws from the GEV distribution. Recall that the GEV(μ, σ, ξ) pdf is given by

$$f(x) = \exp\left(- \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right) \times \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1-1/\xi}, \ \mu, \xi \in \mathbb{R}, \sigma > 0.$$

Consider three methods available with the `R` function `optim`: Nelder-Mead, Newton-Raphson, and BFGS. Please turn in the following:

(a) Pseudocode for at least two algorithms you used. Even though you are using an `R` function, you should be able to summarize how the algorithm works by looking up the code and/or relevant papers from the documentation. Make sure you explain your approach for finding starting values.

(b) In a table with a different row for each method, report your ML estimates for each of the parameters along with standard error estimates for the parameters. Explain the method you used to obtain the standard error estimates.

(c) Overlay the fitted GEV density on a histogram of the data to get a sense for how the fitted model compares to the data.

(2) Carry out Bayesian inference for the above model using the same data set, using Markov chain Monte Carlo (MCMC). Please use the following "uniform" prior distribution for μ, σ, ξ, $p(\mu, \sigma, \xi) \propto \frac{1}{\sigma}, \sigma > 0, \mu, \xi \in \mathbb{R}$. Construct at least two algorithms and try at least 2 different starting values for each algorithm in order to get multiple confimations of your results. Please turn in:

(a) Pseudocode for each of your MCMC algorithms. You should indicate how you obtained starting values,

(b) Clearly explain how how you determined when to stop (how you determined chain length). Also provide standard diagnostics for your results, e.g. relevant plots.

(c) Provide your posterior mean and posterior standard deviation estimates in a table. Each row should correspond to one algorithm. Your posterior mean estimates should have MCMC standard errors associated with them.

(d) Provide in a table a comparison of the asymptotic 95% confidence intervals from the previous problem with the 95% credible intervals you obtain via Bayesian inference. How do they compare?

(e) Your well commented `R` code in a separate file.

(3) Pick your best optimization algorithm to compute MLEs and carry out a simulation study for ML inference for $\mu = 3, \sigma = 0.8, \xi = 0.4$.

(a) Approximate the mean squared error (MSE) for each parameter estimate. Provide Monte Carlo standard errors for these approximations.

(b) Approximate the coverage of your 95% confidence intervals for each parameter. Provide Monte Carlo standard errors for these approximations.

(c) Brief description of how you carried out your simulation study. Provide enough information that someone else should be able to replicate it completely, e.g. how you determined the Monte Carlo sample size for this study. Note in particular that the initial values should be selected automatically and you should not use the fact that you know the true values to select these! Come up with an approach that seems to work, perhaps based on the simulated data each time.

(d) Submit your `R` code for this exercise as well.