ST740 – Data Analysis Project 1 – Due 10/12

THIS IS AN EXAM, YOU MUST WORK INDEPENDENTLY! Do not discuss the exam with anyone other than the professor, including other students or the TA.

For this assignment, you will analyze the streamflow data at

```
file  <- "https://www4.stat.ncsu.edu/~bjreich/ST740/HCDN_annual_max.RData"
load(url(file))
```

These data are collected as part of the USGS' Hydro-Climatic Data Network of sites that have not been affected by human development for the purpose of studying the effects of climate change on flood risk. This workspace includes annual maximum streamflow (cubic feet per second) at 702 locations from 1950-2021 in the matrix `Y`. The spatial coordinates of the stations are given in the matrix `s` and the years are given in the vector `year`. Let Y_{st} be the observation at location $s \in \{1, ..., 702\}$ and time (year) $t \in \{1, ..., 72\}$ and $X_t = (year_t - 1985)/10$ be a linear trend covariate, where $year_t$ is the year for observation t. Discard sites with missing observations, which leaves $n = 236$ sites. The goal is to determine if and where the distribution of extreme streamflow has changed over the past 72 years.

We will analyze these data using the Generalized Extreme Value (GEV) distribution. Fit the model $Y_{st}|\theta_s \sim \text{GEV}(\mu_{st}, \sigma_s, \xi_s)$, independent across s and t, with $\mu_{st} = \theta_{1s} + \theta_{2s}X_t$, $\log(\sigma_s) = \theta_{3s}$ and $\xi_s = \theta_{4s}$. The vector of parameters at site s is $\theta_s = (\theta_{1s}, \theta_{2s}, \theta_{3s}, \theta_{4s})^T$ and has random effect distribution $\theta_s|\mu, \Sigma \sim \text{Normal}(\mu, \Sigma)$ for the 4×1 mean vector μ and 4×4 covariance matrix Σ.

1. Select and justify uninformative priors for μ and Σ, and check whether results are sensitive to the prior choice.

2. Construct an MCMC algorthm (step-by-step in `R`, not `JAGS` or similar program) to sample from the joint posterior of $\theta_1, ..., \theta_n, \mu, \Sigma$. Describe the steps of the algorithm including your approach to tuning the sampler, and include plots/tables to verify that the algorithm has converged.

3. Summarize the posterior of the slopes $\{\theta_{21}, ..., \theta_{2n}\}$ in a 2-3 figures/tables.

4. Based on your analysis, would you say the distribution of extreme streamflow has changed over time in the US? If so, where?

5. Idenfity 2-3 limitations of your analysis and discuss how these could be addressed in future research.

Your write-up should be written in manuscript form with complete sentences and paragraphs and carefully labelled figures and tables. Do not submit a markdown document. Include sufficient detail of your model and results so that another student in the class could reproduce your results. It should be a single PDF document. The main text is limited to 5 pages (double-spaced) including tables and figures but excluding code, which should be given as an appendix that does not count towards the 5 page limit. A large part of your grade will be on presentation (clear writting, clean figures and tables, etc) so take time to polish your work. Submit your report in class on Oct 12.

HAVE FUN!