Statistics 601, Spring 2026
Assignment 2

1 Background

In Chapter 9.10 of the Stat 520 notes, we considered a case study that consisted of formulating an additive error nonlinear regression for relating weight (as responses) to length (as covarites) in Walleye (*Sander vitreus*) taken from several lakes in Minnesota over the years 1991 to 1993. In that case study we eventually arrived at a model to relate random variables $\{Y_i : i = q, \ldots, n\}$ connected with the weight (in pounds) of fish i, to covariates $\{x_i : i = 1, \ldots, n\}$ being the length of those same fish (in inches),

$$\begin{aligned} Y_i &= \mu_i + \sigma \mu_i^{\theta} \epsilon_i, \\ \mu_i &= \beta x_i^{\alpha} \\ \epsilon_i &\sim \text{iid N}(0, 1). \end{aligned} \tag{1}$$

We conducted both likelihood and Bayesian analyses. In the Bayesian approach we used prior distributions,

$$\begin{aligned} \beta &\sim N(\lambda, \tau^2) \\ \alpha &\sim \text{Gamma}(\psi_1, \psi_2) \\ \sigma^2 &\sim \text{IG}(\xi_1, \xi_2) \end{aligned} \tag{2}$$

Information was obtained from the web pages of state natural resource agencies such as the Wisconsin Department of Natural Resources to guide the selection of parameter values in these prior distributions. We ended up with $\lambda = 0.0004$, $\tau^2 = 0.01$, $\psi_1 = 3$, $\psi_2 = 1$, $\xi_1 = 0.1$, and $\xi_2 = 0.1$. The power θ was chosen on the basis of exploratory analysis as $\theta = 1.0$. Using an overall Gibbs Sampling algorithm we arrived at posterior means of $E(\beta|\boldsymbol{y}) = 0.00027$, $E(\alpha|\boldsymbol{y}) = 3.0880$ and $E(\sigma^2|\boldsymbol{y}) = 0.0097$ (See Table 9.6 of the Stat 520 notes).

2 Our Current Objective

For this assignment, we wish to extend (1) to a hierarchical model across multiple lakes in Iowa for a different species, Large Mouth Bass (*Micropterus salmoides*). Like Walleye, Large Mouth Bass (LMB) are a highly sought after game fish, although they do not typically grow as large as Walleye. The Iowa Department of Natural Resources (IDNR) manages lakes and reservoirs in the state for a whole host of fish species, including LMB. The data available to us are observations of lengths and weights for $5,454$ individual LMB sampled from 75 different lakes in Iowa. These data are available on the course web page in the Data module and file `LMBdat_for601.txt`.

Of concern to biologists at IDNR are the health of populations of LMB in individual lakes. Fish, in general, exhibit indeterminate growth, which means they continue to grow over their entire life span (like trees). But if fish grow without the proper levels of nutrition they may become long and slender. So health, or what fisheries scientists sometimes refer to as *condition*, involves the ratio of weight to length. In populations of fish, an imbalance in the distribution of fish sizes can result in what is called a *stunted* population in which there are many, many small individuals, but few or no large individuals. This may result from the population over-eating food resources combined with a lack of predators that eat smaller fish. Under these conditions there isn't sufficient food for a large population of smaller individuals to grow bigger and most of the fish in the population are in poor condition. In the model of (1) the parameter α is believed to be a characteristic of morphology, which varies from species to species, but should be essentially constant over populations of the same species of fish. The parameter β, in contrast, reflects the proportional relation of weight Y_i to the expression of length represented as x_i^α, and is therefore related to condition. Although there is some variability among individuals in a population, condition or population health is largely governed by environmental factors and the available resource base of a given lake. So condition is generally interpreted as a population-level or lake-level phenomenon. While this characteristic of a fish population can certainly change over time, it does not tend to do so rapidly unless there is some intervention on the part of biologists or a major natural disturbance such as a tornado or major drought. Even then, a shift of any real meaning would probably require a period of several years.

One way to judge health is to compare the weight of fish of a given length to what are called *standard weights*. Exactly what the source or sources of standard weights are is difficult to determine, but the same table of standard weights for LMB may be found on the web sites of multiple state resource agencies (e.g., Georgia, New York, Wisconsin). These standard weights indicate that a healthy LMB of length 10 inches should weigh 0.50 pounds, one of length 15 inches should weigh 1.83 pounds, and a 20 inch LMB in good condition should weigh 4.59 pounds. The Iowa DNR would like to assess the health of individual populations of LMB in the particular 92 lakes sampled.

3 A Hierarchical Model

Given the information discussed previously, we will take the parameter α, a characteristic of species, to have the same value for all lakes in our data set. The parameters β and σ^2 will be modeled as lake-specific but following a common distribution across the 75 situations (lakes) in our data. We will fix $\theta = 1.0$ following the analysis of Walleye in Chapter 9 of the Stat 520 notes. We will choose distributions and priors based on previous experience with the Walleye example and information on LMB length-weight relations found on resource agency web sites. The model given in (1) can be extended to a hierarchical version as follows. Let $\{Y_{i,j} : i = 1, \ldots, n;\ j = 1, \ldots, m_i\}$ denote random variables connected with the weight of fish j in lake i, let $\{x_{i,j} : i = 1, \ldots, n;\ j = 1, \ldots, m_i\}$ be the corresponding lengths. Then model,

$$\begin{aligned} Y_{i,j} &= \mu_{i,j} + \sigma_i \mu_{i,j}^{\theta} \epsilon_{i,j} \\ \mu_{i,j} &= \beta_i x_{i,j}^{\alpha} \\ \epsilon_{i,j} &\sim \text{iid N}(0,1). \end{aligned} \tag{3}$$

Distributions of the data model parameters $\{\beta_i : i = 1, \ldots, n\}$ and $\{\sigma_i^2 : i = 1, \ldots, n\}$ were chosen based on what had been used as prior distributions in the one lake model for Walleye,

$$\begin{aligned} \beta_i &\sim \text{iid N}(\lambda, \tau^2) \\ \sigma_i^2 &\sim \text{iid IG}(\xi_1, \xi_2) \end{aligned} \tag{4}$$

To choose priors and parameter values for those priors, we rely on a combination of conditional conjugacy and general recommendations in the literature. The prior distributions chosen for this problem were,

$$\begin{aligned} \alpha &\sim \text{Gamma}(\psi_1, \psi_2) \\ \lambda &\sim \text{N}(\lambda_0, \tau_0^2) \\ \tau &\sim \text{Unif}(0, T) \\ \xi_1 &\sim \text{Unif}(0, S_1) \\ \xi_2 &\sim \text{Unif}(0, S_2) \end{aligned} \tag{5}$$

Note here that a uniform prior is assigned to τ, rather than τ^2, which follows the advice of Gelman (2006, Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**: 515-533). This uniform prior results in a prior for τ^2

of $\pi(\tau^2) = (1/(2T))(\tau^2)^{1/2};\ 0 < \tau^2 < T^2$. To select parameter values for the priors in (5) we rely on a combination of information available on the web sites of natural resource agencies concerning LMB, and assessment of how various parameters cause the prior distributions to place greater and lesser probabilities. In the previous Walleye example, the formula given for relating weight (W) to length (L) was $W = L^3/2700$. Several agency web sites give a similar formula for LMB of $W = (L^2 * G)/1200$, where G is girth (circumference at thickest portion of body). Our data do not include girth, but information on average bass morphology implies that, roughly, $G = 0.75L$ meaning that we could plan on $W = L^3/1600$, giving values for α and β in (1) of about 3 and 0.0006, respectively. This resulted in choices of $\lambda_0 = 0.0006$, $\tau_0^2 = 0.001$ (a coefficient of variation of about 160%), $\psi_1 = 3$, and $\psi_2 = 1$, which gives 95% probability to the interval $(0.8, 6)$. The value of T was set arbitrarily to 3 and we took $S_1 = 0.25$ and $S_2 = 0.1$, resulting in Inverse Gamma distributions that placed at least 90% probability in the interval $(0, 10)$.
The joint posterior $p(\lambda, \tau, \xi_1, \xi_2, \alpha, \{\beta_i : i = 1, \ldots, n\}, \{\sigma_i^2 : i = 1, \ldots, n\}|\boldsymbol{y})$ was approximated using an overall Gibbs Sampling algorithm. The conditional posteriors for ξ_1, ξ_2, α, and $\{\beta_i : i = 1, \ldots, n\}$ were sampled using Metropolis within Gibbs. Those for λ, τ^2 and $\{\sigma_i^2 : i = 1, \ldots, n\}$ were available in closed form as a result of conditional conjugacy. Based on autocorrelation and visual examination of trace plots, the chain appeared to mix quite rapidly. A burn-in of $B = 1,000$ followed by collection of the subsequent $M = 10,000$ values was used in the MCMC procedure.

Output from this MCMC is available on the course web page in the Data module in the files `LMBMCMC1.txt` and `LMBMCMC2.txt`. The first file contains a data frame with variables `lam`, `tau2`, `xi1` and `xi2`. Rows correspond to iterations in the Gibbs algorithm. The second file contains a matrix of size $10,000 \times 152$. The columns of this matrix have no names but the first 75 give values of β_i; $i = 1, \ldots, 75$. The next set of 75 columns give values of σ_i^2; $i = 1, \ldots, 75$. The last 2 columns are values of

α and θ (which are all 1s as θ was fixed in this analysis).