STA 941: Bayesian Nonparametrics
HW Set 3

Download the file www.stat.duke.edu/~st118/sta941/Datasets/us-mortality.csv which contains US mortality records by county between 1999 and 2012. The variables measured are:

- `Year`: measurement year
- `FIPS_Code`: unique FIPS code of the county
- `County`: name of the county
- `State`: state
- `Deaths`: actual number of deaths
- `Population`: total population size
- `Mortality`: crude mortality rate in 100,000
- `Tmax90`: #days in the year with max daily temp over 90°F
- `Tmax32`: #days in the year with max daily temp below 32°F
- `Tmin32`: #days in the year with min daily temp below 32°F
- `Tmin0`: #days in the year with min daily temp below 0°F
- `Precip1`: #days in the year with 0.01 inch rainfall or more
- `Precip10`: #days in the year with 0.1 inch rainfall or more
- `Snow`: #days in the year with 0.1 inch snowfall or more
- `Density`: population density
- `Income`: median household income
- `Hospital`: #number of hospitals (averaged across years)

The weather records vary across years and are not available for year 1999. The records on income and population density are static. So is hospital count, which is calculated by averaging sporadic information available for each county during the observation time period.

Consider a mixed effects linear model regression of crude mortality rate as follows. Below, c indexes counties and t denotes `Year` - 1999.

$$\begin{aligned}
\text{Mortality}_{ct} &\sim \beta_0(c) + t \cdot \beta_1(c) + \beta_1^w \text{Tmax90}_{ct} + \beta_2^w \text{Snow}_{ct} + \beta_3^w \text{Precip10}_{ct} + \text{err}_{ct}, \\
\beta_0(c) &= \delta_0(c) + \gamma_{01} \cdot \text{Hospital}_c + \gamma_{02} \cdot \text{Density}_c + \gamma_{03} \cdot \text{Income}_c \\
\beta_1(c) &= \delta_1(c) + \gamma_{11} \cdot \text{Hospital}_c + \gamma_{12} \cdot \text{Density}_c + \gamma_{13} \cdot \text{Income}_c
\end{aligned}$$

which can be re-expressed in the more conventional "fixed effects + random effects + error" form as

$$\text{Mortality}_{ct} = x_{ct}^T \beta + z_{ct}^T \theta_c + \text{err}_{ct}$$

where

$$x_{ct} = \begin{pmatrix} \text{Tmax90}_{ct} \\ \text{Snow}_{ct} \\ \text{Precip10}_{ct} \\ \text{Hospital}_c \\ \text{Density}_c \\ \text{Income}_c \\ t \cdot \text{Hospital}_c \\ t \cdot \text{Density}_c \\ t \cdot \text{Income}_c \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1^w \\ \beta_2^w \\ \beta_c^w \\ \gamma_{01} \\ \gamma_{02} \\ \gamma_{03} \\ \gamma_{11} \\ \gamma_{12} \\ \gamma_{13} \end{pmatrix}, \quad z_{ct} = \begin{pmatrix} 1 \\ t \end{pmatrix}, \text{ and, } \quad \theta_c = \begin{pmatrix} \delta_0(c) \\ \delta_1(c) \end{pmatrix}$$

Analyze the model under the following assumptions:

- $\text{err}_{ct} \sim N(0, \sigma^2)$, independently across c and t [For now, ignore the fact that this is a bad assumption! One should at least have $\text{err}_{ct} \sim N(0, \sigma^2/P_c)$ where P_c is the county population count.]
- $\theta_c \sim P$, independently across c, where P is an unknown bivariate distribution to be estimated from data.

Assume the reference prior on $(\beta, \sigma^2) \propto 1/\sigma^2$. For specifying a prior on P, consider three possibilities[1]:

Gaussian: $P = N(\mu, \Sigma)$, $\mu \sim N(m, S)$, $\Sigma \sim IW_2(\nu, T)$

DP: $P \sim \text{DP}(a, N(\mu, \Sigma))$, $\mu \sim N(m, S)$, $\Sigma \sim IW_2(\nu, T)$, $a \sim Ga(1, 1)$

DPM: $P = \int N(\eta, \Omega) Q(d\eta)$, $Q \sim \text{DP}(a, N(\mu, \Sigma))$, $\mu \sim N(m, S)$, $\Sigma \sim IW_2(\nu, T)$, $a \sim Ga(1, 1)$, $\Omega \sim IW(\kappa, W)$

where the hyper-parameters m, S, ν, T, κ and W are fixed as follows. First run an ordinary regression of mortality on the fixed effects regressors (i.e., x_{ct}) to get an estimate $\hat{\beta}$ of β and corresponding residuals $r_{ct} = \text{Mortality}_{ct} - x_{ct}^T \hat{\beta}$. Then for each county c, run a separate regression of r_{ct} on z_{ct}, $t = 1, \ldots, T$, to obtain an initial estimate $\hat{\theta}_c$ of θ_c. Let m_0 and V_0 be the mean and variance of the estimated $\hat{\theta}_c$, $c = 1, \ldots, C$. Set: $m = m_0$, $S = 4V_0$, $\nu = 4$, $T = V_0$, $\kappa = 4$, $W = V_0$.

Each model can be fitted with a Gibbs sampler. Here are some clues as to how to design such samplers.

- For the Gaussian model, a Gibbs sampler cycles through making draws of (β, σ^2), $\theta_{1:C}$, μ and Σ from their respective conditional posterior distributions given everything else. The conditional posterior of (β, σ^2) is easily derived from the reduced model $Y_{ct} = x_{ct}^T \beta + \text{err}_{ct}$, $\text{err}_{ct} \overset{\text{IID}}{\sim} N(0, \sigma^2)$, with $[\beta, \sigma^2] \propto 1/\sigma^2$, where, $Y_{ct} = \text{Mortality}_{ct} - z_{ct}^T \theta_c$. Similarly, the conditional posterior of $\theta_{1:C}$ comes from the reduced model, $Y_{ct} = z_{ct}^T \theta_c + \text{err}_{ct}$, $\text{err}_{ct} \sim N(0, \sigma^2)$, with, $\theta_c \overset{\text{IID}}{\sim} N(\mu, \Sigma)$ where, now, $Y_{ct} = \text{Mortality}_{ct} - x_{ct}^T \beta$. Notice that the conditional posterior of $\theta_{1:C}$ factors into a product over the individual θ_c, i.e., draws can be made in parallel separately for each county. Finally, updating μ and Σ depends only on the piece: $\theta_c \overset{\text{IID}}{\sim} N(\mu, \Sigma)$, $\mu \sim N(m, S)$ and $\Sigma \sim IW(\nu, T)$.

[1] Here the d-dimensional inverse Wishart distribution $IW_d(r, S)$ is parametrized such that $V \sim IW_d(r, S)$ implies $\mathbb{E}V = S/(r - d - 1)$

- For DPM, you can do a Neal Algorithm 2 like extension of the above by introducing latent parameters η_c, $c = 1, \ldots, C$ and rewriting the model as $\theta_c \overset{\text{IND}}{\sim} N(\eta_c, \Omega)$ and $\eta_c \overset{\text{IID}}{\sim} Q$. To implement Algorithm 2, you will have to keep track of the clustering labels [call them ℓ_c] and unique values of η_c.
- For the DP implementation, you can implement Neal Algorithm 2 into a Gibbs sampler. You can also try a version of the DPM model with a very large value of κ so that Ω is effectively a zero matrix.

Implement MCMC model fitting for each prior specification and compare results. To make the computing time more manageable, you may want to restrict your analysis to a single state at a time (e.g., NC which has 100 counties). Hold out 2012 data to compare the three specifications on their accuracy in mortality forecasting. To measure accuracy, you may want to use the average absolute errors from the counties, weighted by county population sizes. Also, report a short comparative analysis of the estimates/credible bands of the fixed effects and random effects.